Intelligent Methods for Condition Monitoring of Rolling Bearings Using Vibration Data

A thesis submitted for the degree of

Doctor of Philosophy

By

Hosameldin Ahmed

Department of Electronic and Computer Engineering

College of Engineering, Design and Physical Sciences

Brunel University London

August 2018

Abstract

Owing to the importance of rolling bearings in rotating machines, there has been great interest in the development of computational methods for rolling bearings condition monitoring over the last few decades. The aim of these methods is to determine early and automatically the occurrence of a fault condition in rolling bearings to avoid machine breakdowns that may lead to downtime, possibly safety incidents, production lost, and higher costs of repairs. Vibration-based methods are commonly used and have become well-accepted techniques of many condition-based maintenance (CBM) management. However, the vast amount of the collected vibration data requires large storage and time for signal processing and this also may limit the number of machines that can be monitored remotely across wireless sensor networks (WSNs) due to bandwidth and power constraints. To avoid the burden of much storage requirements and processing time of a tremendously large amount of vibration data, the scope of this thesis is the development and application of vibration analysis methods of a large amount of acquired vibration dataset for condition monitoring of rolling bearings. The main aim of this analysis is to obtain compressively-sampled dataset which possesses the quality of the original vibration dataset and then learn fewer features from these compressively-sampled signals directly without reconstructing the original signal. Regardless of its importance in addressing the challenge of learning from a large amount of vibration dataset, no existing method possesses the ability to learn directly from low-rate of compressed measurements.

This thesis presents new and tested vibration condition monitoring methods for roller bearings that can greatly compress the amount of the original vibration and further learn features from this compressed amount of data to increase the identifying power of the compressed measurements in rolling bearing faults diagnosis. There are two most important contributions in this thesis. The first one is the formulation of a three-stage method, Compressive Sampling with Correlated Principal and Discriminant Components (CS-CPDC) for classification of bearing faults. This method applies CS to obtain compressively-sampled signals from the raw vibration data and then adopts a multi-step feature learning algorithm to learn fewer features from the compressively-sampled signals. Finally, it employs a multi-class support vector machine (SVM) to classify bearing health conditions using these learned features. The second one is the design of intelligent condition monitoring method for bearing faults from highly compressed measurements using sparse over-complete features. This method applies CS to produce highly compressed measurements of the original bearing vibration dataset and then it uses an effective deep neural network (DNN) with unsupervised feature learning algorithm based on sparse autoencoder to learn over-complete sparse representations of these compressed measurements. Finally, it employs two techniques to deal with the classification problem, namely, pre-training classification based on stacked autoencoder and softmax regression layer, and re-training classification based on backpropagation algorithm. These methods and other less performance methods proposed in this thesis have been validated and applied to several real vibration datasets of roller bearings. The experimental validation demonstrates improved bearing health condition classification accuracy, with highly reduced feature dimension, and much lower computational complexity, compared to state-of-the-art methods.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACO	Ant Colony optimisation
AE	Autoencoder
AIC	Akaike's Information Criterion
ANC	Adaptive noise cancellation
ANN	Artificial neural network
AR	Autoregressive model
ARIMA	Autoregressive integrated moving average model
ARMA	Autoregressive moving average model
BFDF	Bearing fundamental defect frequency
BP	Back-propagation
BPFI	Bearing pass frequency of inner race
BPFO	Bearing pass frequency of outer race
BPNN	Backpropagation neural network
BSF	Ball spin frequency
BSS	Blind Source Separation
CA	Cage fault condition
CART	Classification and regression tree
CBM	Condition-based maintenance
CCA	Canonical correlation analysis
CDBN	Convolutional deep belief network
CF	Crest factor
CFT	Continuous Fourier Transform
CG	Conjugate Gradient
Chi-2	Chi-square
CM	Condition monitoring
CNNs	Convolution deep neural networks
CoSaMP	Compressive sampling matched pursuit
CS	Compressive sampling
CS-Chi-2	Compressive sampling and Chi-square feature selection
CS-CPDC	Compressive sampling with correlated principal and discriminant components

CSFR	Compressive sampling and feature ranking framework
CS-FS	Compressive sampling and Fisher score feature selection
CS-LDA	Compressive sampling with linear discriminant analysis
CS-LS	Compressive sampling and Laplacian feature selection
CSLSL	Compressive sampling and linear subspace learning
CS-PCA	Compressive sampling with principal component analysis
CS-PCC	Compressive sampling and Pearson correlation coefficients
CS-Relief-F	Compressive sampling and Relief-F feature selection
CS-SAE-DNN	Compressive sampling with sparse autoencoder based deep neural network
CWRU	Case Western Reserve University
CWT	Continuous wavelet transform
DBNs	deep belief networks
DFT	Discrete Fourier Transform
DNN	Deep neural network
DWT	Discrete wavelet transform
ECOC	Error-correcting output codes
EMA	Exponential moving average
EMD	Empirical mode decomposition
FCM	Fuzzy C-means
FFT	Fast Fourier transform
FMM	Fuzzy min-max
FR	Feature ranking
FS	Fisher score
FSAR	Feature selection by Adjunct rand index standard deviation ratio
FT	Fourier transform
FTF	Fundamental train frequency
GA	Genetic algorithm
GP	Genetic programming
GR	Gain ratio
HHT	Hilbert-Huang Transform
HIST	Histogram
HOM	Higher order moment
ICA	Independent component analysis

IG	Information gain
IR	Inner race fault condition
KL	Kullback-Leibler divergence
KNN	k-nearest neighbor
KUR	Kurtogram
LASSO	Least absolute sum of squares operator
LB	Lower bound of the histogram
LFDA	Local fisher discriminant analysis
LMD	Local mean decomposition
LRC	Logistic regression classifier
LS	Laplacian score
LSL	Linear subspace learning
LSSVM	Least squares support vector machine
MA	Moving average
MBSE	Multi-band spectrum entropy
MCM	Machine condition monitoring
MD	Mahalanobis distance
MFE	Multi-scale fuzzy entropy
MI	Mutual information
MLP	Multi-layer perceptron
MMV	Multiple measurement vector
MMV-CS	Multiple measurement vectors based compressive sampling
MPE	Multi-scale permutation entropy
mRmR	Minimum redundancy maximum relevance
MSE	Multi-scale entropy
MSRMS	Multi-scale root-mean-square
MSVM	Multiclass support vector machine
NN	Neural network
NO	Brand new condition
NW	Worn but undamaged condition
OFS	Original feature set
OR	Outer race fault condition
PCA	Principal Component analysis

PCC	Pearson correlation coefficients
PCs	Principal components
PDF	Probability density function
PF	Polynomial function
PNN	Probabilistic neural network
PS	Power spectrum
PSO	Particle swarm optimisation
QPSO	Quantum behaved particle swarm
RBFN	Radial basis function network
RBMs	Restricted Boltzmann machines
RE	Rolling element fault condition
RF	Random forest
RIP	Restricted Isometry property
RL	Reinforcement learning
RMS	Root mean square
RMSE	Root mean squared error
RNNs	Recurrent neural networks
SAE	Sparse autoencoder
SampEn	Sample Entropy
SBFS	Sequential backward floating selection
SBS	Sequential backward selection
SCG	Scale conjugate gradient
SF	Sigmoid function
SFFS	Sequential forward floating selection
SFS	Sequential forward selection
SK	Spectral kurtosis
SM-LFDA	Support margin local fisher discriminant analysis
SMV-CS	Single measurement vector based compressive sampling
SOM	Self-organising map
STD	Standard deviation
STE	Standard error
STFT	Short time Fourier transform
SVM	Support vector machine

SVM-REF	Support vector machine – Recursive feature elimination
TBM	Time-based maintenance
TSA	Time synchronous averaging
UB	Upper bound of the histogram
VPMCD	Variable predictive model-based class discrimination
VR	Variance
WPT	Wavelet packet transform
WSNs	Wireless sensor networks
WT	Wavelet transform
WVD	Wigner-Ville distribution
ZC	Zero crossing

Acknowledgments

I would like to thank my parents, my wife Intesar, and my children Monia, Malaz, Mohamed, and Abubaker for all their patience and support without which I would not have been able to complete this thesis. Also, I would like to thank my first supervisor Prof Asoke Nandi for his patient supervision, great guidance, and advice for my work in this thesis. Additionally, I thank him for motivating me to focus on publishing my work. I also thank him for his confidence and mentorship over the years. Moreover, I would like to thank Prof Dennis Wong for his fruitful collaboration, useful discussions, and valuable suggestions. In addition, I thank my second supervisor Dr. Maysam Abbod for his kind support and help.

I would like to thank Brunel University London for all their support including the computing facilities and the training opportunities within Brunel Researcher Development Programme that have been available to me and all researchers in the university. I also extend my sincere thanks to all members of Brunel Graduate School and the Department of Electronic and Computer Engineering for all their support during my Ph.D. study. Additionally, I would like to thank Professor Nandi for making available some financial support.

I would like to thank Prof. Chris Hudson, Dr. Federico Colecchia, Dr. Basel Abujamous, Dr. Chao Liu, and Dr. Ismail Hburi for their invaluable advice at the early stage of my Ph.D. study. Last but not the least, I would like to thank all my colleagues and friends who have filled my Ph.D. study with happiness and fun.

Chapter 1
Introduction

1.1 Background

The need for an effective condition monitoring and machinery maintenance program exists wherever complex and expensive machinery is used to deliver critical functions of businesses. For example, manufacturing companies in today's global marketplace race, use their best endeavours to cut costs and improve product quality to maintain their competitiveness. Rotating machine is a central part of the manufacturing procedure, and its health and availability have their direct effects on production schedules, production quality, and production costs. Unforeseen machine failures may lead to unexpected machine downtime, accidents, and injuries. In reality, rotating machine components including motors, bearings, gearboxes, etc. are engaged to operate effectively to keep a stable and healthy condition of the rotating machine. For that reason, maintenance is performed in order to ensure that machines remain in a healthy condition by repairing, modifying, or replacing these components.

Maintenance can be accomplished using two main approaches, namely, corrective and preventive maintenance (Wang et al., 2007). Corrective maintenance is the most basic maintenance technique that performs after machine failure and can be very expensive particularly for large-scale applications of rotating machines. Preventive maintenance can be applied to prevent a failure using either time-based maintenance (TBM) or condition-based maintenance (CBM), which can be Localised CBM, or Remote CBM (Higgs et al., 2004; Ahmad et al., 2012). TBM uses a calendar schedule that sets in advance to perform the maintenance regardless of the health condition of the machine, which makes this approach also expensive in some large and complex machines. In addition, TBM may not prevent machines from failures.

It has been reported that 99% of rotating equipment failures are preceded by non-specific conditions indicating such a failure is going to happen (Bloch et al., 2012). CBM is an efficient maintenance approach that can help in avoiding unnecessary maintenance tasks of TBM approach. Numerous studies have shown the economic advantages of CBM in several applications of rotating machine (for example, McMillan et al., 2007; Verma et al., 2013; Van Dam and Bond, 2015; Kim et al., 2016). In CBM, the decision of maintenance is made based on the current machine health condition that can be identified through the Condition Monitoring (CM) system. Once a fault occurs, the accurate CM technique allows early detection of faults and correct identification of the type of faults. Thus, the more accurate and sensitive CM system, the more correct maintenance decision is made and more time available to plan and perform maintenance before machine breakdowns.

Condition monitoring of rotating machine components can minimize the risk of failure by identifying machine health condition via early fault detection. The aim of CM is to avoid catastrophic machine failure that may cause secondary damage, downtime, potentially safety incidents, production lost, and higher costs associated with repairs. CM techniques in rotating machinery encompass the practice of monitoring measurable data (e.g., vibration, acoustic, etc.), that can be used individually or in a mixture to identify changes in machine condition. This allows the CBM program to be arranged, or other actions to be taken to prevent machine breakdowns (Jardine et al., 2006). Based on types of sensor data acquired from rotating machines, bearing condition monitoring techniques can be grouped into the following: vibration monitoring, acoustic emission monitoring, electric current monitoring, temperature monitoring, chemical monitoring, and laser monitoring. Of these techniques, vibration-based bearing condition monitoring has been widely studied and has become a well-accepted technique of many planned maintenance managements (Lacey, 2008; Randall and Antoni, 2011). In reality, different fault conditions generate different patterns of vibration spectrums. Thus, vibration analysis in principle allows us to examine the inner parts of the machine and analyse the health conditions of the operating machine without physically opening it (Nandi et al., 2013). Moreover, various characteristic features can be observed from vibration signals that make it one of the best selections for machine CM.

In vibration-based condition monitoring by analysing the physical features of the acquired vibration signals, one is able to categorise the acquired vibration signal into the corresponding condition correctly, which is generally a multi-class classification problem. A simple vibration based CM system consists of three key steps as shown in Figure 1.1. First, data acquisition step; in which a sensor (e.g., velocity sensors and accelerometers) is mounted to the component of interest to collect input vibration measurements, i.e., raw data, that can be transmitted, stored, and processed. Second, the vibration data analysis step, include pre-processing, filtering, feature extraction and selection of the vibration data acquired in the first step. Finally, machine health diagnosis that involves detection and identification of a fault uses a classifier to discriminate the data signals into different classes utilising the extracted features, while prognosis aims to predict the residual life of a machine before breakdown takes place (Nandi and Jack, 2004; Worden et al., 2011).

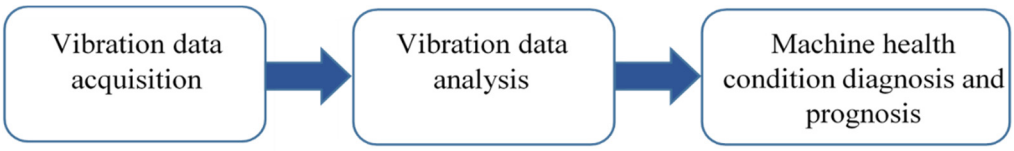


Figure 1.1. The overall framework of vibration based machine condition monitoring.

The sampling theorems including the Shannon-Nyquist theorem are in the core of the current sensing systems. However, Nyquist sampling rate which is at least twice the highest frequency contained in the signal is high for some modern developing applications, e.g., industrial rotating machine (Eldar et al., 2015). One aspect of much of the literature on using Nyquist sampling rate is that it may result in measuring a large amount of data that need to be transmitted, stored, and processed.

Moreover, some applications that include wideband it is often very costly to collect samples at the necessary rate. It is clear that acquiring a large amount of data requires large storage and time for signal processing and this also may limit the number of machines that can be monitored remotely across Wireless Sensor Network (WSNs) due to bandwidth and power constraints.

For this reason, it is currently becoming of essential importance to develop new CM methods that not only have the ability to achieve accurate detection and identification of machine health conditions but are also have the capability to address two main challenges: (1) costs of learning from a large amount of vibration data, i.e., transmission costs, computation costs, and power needed for computations, and (2) the demand for early detection of fault.

1.2 Rolling bearings condition monitoring

Considering their role in maintaining motion between static and moving parts in rotating machinery, rolling bearings are critical components of the whole system of rotating machines. In practice, rolling bearings failures may lead to more major failures in machines. It is stated that approximately 40 – 90% of rotating machine failures are related to bearing faults (Immovilli et al., 2010) based on the machine size. Thus, in most production processes in the industry, roller bearings need to be kept in healthy condition to guarantee continuity of production. Therefore, it is very important to monitor roller bearings to avoid machine breakdowns. There are two types of bearings: (1) plain (sliding) bearings that maintain motion through sliding contact, and (2) rolling element bearings that maintain motion through rolling contact (Collins et al., 2010). The latter is widely used in most applications of rotating machinery and can be categorised into two main groups, ball bearings (spherical rolling elements) and roller bearings (nominally cylindrical rolling elements). As shown in Figure 1.2, a roller bearings consists of several components: (1) the inner race in which the shaft drives, (2) the outer race that normally positioned into a hole or housing, (3) the rolling elements that is normally placed between the inner race and the outer race, and (4) the cage that also called retainer, can be made of plastic or metal, and is used to keep the rolling elements separated equally.

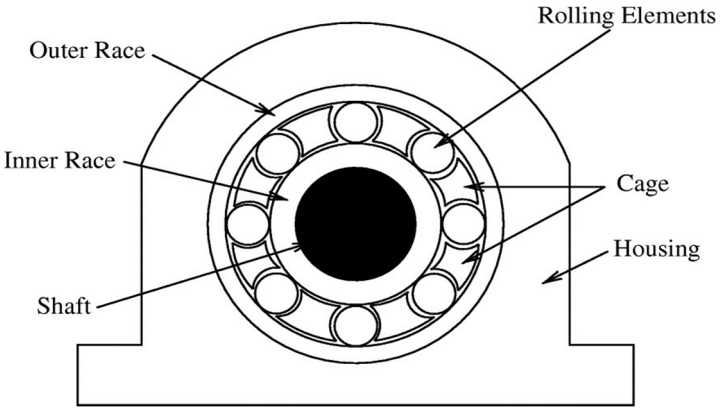


Figure 1.2. A typical roller bearing.

Bearings defects can occur for many reasons including (1) Fatigue, which happens when there is too much load on the bearing; (2) Incorrect lubrication, which can be over-lubrication or under-lubrication; (3) Contamination and corrosion; and (4) Wrong bearing installation (Harris, 2001; Nandi et al., 2005). The literature on vibration-based bearings CM identifies numerous computational methods for fault diagnosis that achieved many interesting results. However, the performance of these methods is limited by the large amounts of Nyquist rate-based sampled vibration data need to be acquired from rotating machines to achieve the anticipated accuracy of bearing fault detection and classification.

A reasonable approach to tackle the challenges involved in dealing with too many samples could be to compress the data. One of the most well-known advanced and recent techniques for signal compression is transform coding that depends on mapping the signal samples into bases that provide sparse or compressible representations of the signal of interest (Rao et al., 2000). Recent advances in techniques beyond bandlimited sampling offer lower sampling rates and a reduced amount of data (Eldar et al., 2015). New advances in transform coding techniques have facilitated investigation of Compressive sampling (CS) framework (Donoho, 2006; Candès and Wakin, 2008) that relies on linear dimensionality reduction. CS supports sampling below the Nyquist rate for signals that have a sparse or compressible description. Accordingly, if the signal has a sparse representation in a known basis then one is able to reduce the number of measurements that need to be stored, transmitted, and processed. CS is being considered in a large diversity of applications including medical imaging, seismic imaging, and radio detection and ranging, and communications and networks (Holland et al., 2010; Qaisar et al., 2013; Merlet et al. 2013; Rossi et al., 2014). The basic idea of CS is that a finite-dimensional signal having sparse or compressible representation can be reconstructed from fewer linear measurements far below than the Nyquist sampling rate. In the last few years, there has been a growing interest in the application of CS in machine fault diagnosis since machine vibration signals have a compressible representation in several domains, e.g., frequency domain. Previous studies investigating CS-based compressive signal processing and its application in machine fault diagnosis can be categorised into two types:

(1) CS-based vibration signal recovery for machine fault diagnosis

The possibility to diagnosis machine faults from reconstructed signals based on CS has been validated by (Li et al., 2012; Wong et al., 2015) where the CS-based compressively sampled signals are used and then followed by signal reconstruction. The literature on compressive sensing shows a variety of approaches to reconstruct original signals from compressive signals (Needell et al., 2009; Dai et al. 2009; Renna et al., 2014; Stanković et al., 2014). However, signal reconstruction techniques may not be practical in all applications and make no attempt to address the question of *whether or not it is possible to learn in the compressed domain rather than having to recover the original signals.* For instance, bearing vibration signal is always acquired for faults detection and estimation, and as long as

it is possible to detect faulty signals in the measurement domain, then it is not necessary to recover the original signal to identify faults.

(2) CS-based fault diagnosis using compressed measurements and incomplete signal reconstruction

Most research in compressive sensing based methods has emphasized the use of sparse representations, compressed measurements, and incomplete signal reconstruction for the bearing fault diagnosis. For example, Tang and colleagues (Tang, et al., 2015) developed a sparse classification strategy based on compressive sensing by extracting and classifying fault features through sparse representation combined with random dimensionality reduction. In this strategy, the original vibration data is sampled and preserved by using a small number of random projection and a redundant dictionary then constructed to sparse representation. Zhang and colleagues suggest a bearing fault diagnosis method based on the low-dimensional compressed vibration signal by training several over-complete dictionaries that can be effective in signal sparse decomposition for each vibration signal state (Zhang, et al. 2015). Another learning dictionary basis for extracting impulse components is described by Chen and colleagues (Chen, et al., 2014). An interesting approach is proposed in (Tang et al., 2015) where authors attempted to observe the characteristic harmonics from sparse measurements through a compressive matching pursuit strategy during the process of incomplete reconstruction. Recently, Shao and colleagues (Shao et al, 2018) proposed an improved convolutional deep belief network (CDBN) with CS for rolling bearing. In this method, CS is used for reducing the vibration data amount to improve analysis efficiency. Then, a CDBN with Gaussian visible units is constructed to enhance the feature learning ability for the compressed data. Finally, the exponential moving average (EMA) method is used to improve the generalization performance of CDBN. In general, a significant analysis and discussion on the subject of how to solve a range of signal detection and estimation problems given compressed measurements without reconstructing the original signal can be found in (Davenport et al., 2006).

Even though the efficiency of CS in machine fault diagnosis has been validated in these studies, there are two main problems with these studies: (1) CS-based sparse signal reconstruction is a complex computational problem that depends on the sparsity of the measured vibration signal. Therefore, CS-based signal recovery methods may not be useful in reducing computational complexity for condition monitoring of rolling bearings, and (2) most of the methods that are based on learning directly from the compressed measurements achieved good classification accuracy by only increasing the sampling rate, thereby requiring higher computational complexity. Consequently, two practical questions arise when dealing with the compressively-sampled signals. 1) Is it important to identify the ideal compressible representations of vibration signals that can be used within the CS framework to generate compressively-sampled signals that possess the quality of the original signals? and 2) How does one identify what type of features can be learned from the compressively-sampled signals that have the

ability to achieve high classification accuracy using low-rate of compressed measurements? The advantages of compressed measurements with low-rate, i.e., highly compressed measurements can be summarised as follows:

1. Reduced computations: CS is able to reduce a large amount of the acquired vibration data. The larger the reduction in the amount of vibration data results in much reduction in the computation.
2. Reduced transmission costs: In the cases of having to send vibration data from remote places by wireless (e.g. in the case of off-shore wind turbines) or wired transmission, the cost of transmission will be less as CS reduces the amount of vibration data.
3. Benefits to the environment: As the application of CS results in reduced computations, it helps to reduce the amount of power needed for both computations and transmission. In consequence, CS offers much benefit to the environment.
4. Increase the number of machines that can be monitored remotely across WSNs: It is clear that acquiring a large amount of data limit the number of machines that can be monitored remotely across WSNs due to bandwidth and power constraints. Therefore, the large reduction in the amount of data will increase the number of machines that can be monitored remotely.

The scope of this thesis is the development and application of vibration analysis methods of a large amount of acquired vibration dataset for condition monitoring of rolling bearing. The main aim of this analysis is to obtain compressively-sampled dataset which possesses the quality of the original vibration dataset and then learn the features of rolling bearing health condition from these compressively-sampled signals directly without reconstructing the original signal. The generation of the compressively-sampled signals serves as a first dimensionality reduction step. Regardless of its importance in addressing the challenge of learning from a large amount of vibration dataset, no existing method possesses the ability to learn directly from low-rate of compressed measurements.

This thesis presents new and tested vibration condition monitoring methods for roller bearings that can receive a large amount of vibration dataset as input and produces fewer features that can sufficiently represent the original input vibration signals. These methods use CS to reduce a large amount of the original vibration signal by obtaining compressively-sampled signals that possess the quality of the original signal. Then, they can further learn features from these compressively-sampled signals that have the ability to increase the identifying power of the compressed measurements in rolling bearing faults diagnosis.

1.3 Structure of the thesis

The remaining part of this chapter describes the contributions of this thesis and lists my publications. **Chapter 2** provides a literature review of vibration-based rolling bearings condition monitoring. Of the large volume of published studies in the area of vibration-based condition monitoring, this chapter discusses the time domain, frequency domain, and time-frequency domain vibration analysis methods. Various commonly used techniques and their application in roller bearing fault diagnosis are presented. This chapter is summarised in section 2.8 while Itemising existing issues and challenges in vibration signal analysis for bearing fault diagnosis that being effectively addressed in this thesis compared to the existing methods. **Chapter 3** discusses the sources, data acquisition, and compression of the different bearing vibration datasets used in the experiments conducted in this thesis. In addition, a brief description of some of the characteristics of the different datasets and their corresponding compressed data matrices sizes were also given. **Chapter 4** is concerned with the methods and techniques that are used in this thesis. These include the new methods that represent the original contribution of this thesis in designing new methods for vibration-based rolling bearing condition monitoring.

Chapter 5, Chapter 6, and Chapter 7 details numerous experiments that have been conducted to assess the newly proposed methods and demonstrate their validity in roller bearing condition monitoring using vibration signals. **Chapter 8**, concludes the thesis and provides insights into future work. Finally, **Appendix I** presents the list of references and **Appendix II** is an index.

1.4 Summary of contributions

The work presented in chapters 4, 5, 6, 7, and 8 discuss intelligent vibration-based methods and their validation in rolling bearings condition monitoring. These methods have been presented in five accepted and published conference papers and three accepted and published journal papers. The original contributions of this thesis can be summarized as follows:

1. **CSFR framework:** the compressive sampling and feature ranking framework combines CS based on multiple measurement vectors (MMV) and feature ranking and selection techniques to learn optimally fewer features from a large amount of vibration data. With these learned features, bearing health condition can be classified using a machine learning classifier. To apply the proposed framework to bearing vibration dataset, we investigate the combination of CS and several feature ranking techniques, namely, Fisher score (FS), Laplacian score (LS), Relief-F, Pearson correlation coefficients (PCC), and Chi-square (Chi-2) to reduce a large amount of bearing vibration signals and select fewer representative features for fault classification. Multinomial Logistic regression (LRC), artificial neural network (ANN), and support vector machine (SVM) classifiers are then

used to produce the final results of bearing health condition. In spite of that, the CSFR framework has the capability to make use of existing feature selection and classification methods in addition to the new developing methods. The CSFR framework for rolling bearings is described in Section 4.2 and its application to rolling bearings condition monitoring was published in (Ahmed et al., 2017a; Ahmed and Nandi, 2017; Ahmed and Nandi, 2018b).

2. **CSLSL based techniques:** the combination of compressive sampling and linear subspace learning techniques is introduced to learn fewer features from the raw vibration data. These learned features can be successfully used for rolling bearing fault diagnosis. Based on this combination we introduced three methods, compressive sampling with principal component analysis (CS-PCA), compressive sampling with linear discriminant analysis (CS-LDA), and compressive sampling with correlated principal and discriminant components (CS-CPDC), which have been tested for rolling bearing fault diagnosis. These methods are described in Section 4.3 and were published in (Ahmed et al., 2017b; Ahmed and Nandi, 2018c; Ahmed and Nandi, 2018a)
3. The design of an intelligent classification method for bearing faults from highly compressed measurements using sparse-over-complete features and training DNN through SAE (**CS-SAE-DNN**). This method includes the extraction of over-complete sparse representations from highly compressed measurements. It involves the unsupervised feature learning algorithm sparse autoencoder (SAE) for learning feature representations in multi-stages of non-linear feature transformation based on deep neural network (DNN). The accuracy of the proposed method is verified using highly compressed datasets of rolling element bearings signals obtained using different compressed sampling rates. These compressed datasets contain fewer samples for each bearing condition. This method is described in Section 4.4 and was published in (Ahmed et al., 2018).
4. Two compressible representations of bearing vibration signals that can be used within CS framework to generate compressively-sampled signals, namely, Fast Fourier Transform (FFT) based coefficients and thresholded Wavelet Transform (WT) based coefficients, were investigated.

1.5 List of publications

1.5.1 Journal publications

1. H. Ahmed and A. Nandi, "Three-stage Hybrid Fault Diagnosis for Rolling Bearings with Compressively-sampled data and Subspace Learning Techniques," in IEEE Transactions on Industrial Electronics. doi: 10.1109/TIE.2018.2868259. (Ahmed and Nandi, 2018a).

2. Ahmed, H. and Nandi, A.K., 2018. Compressive Sampling and Feature Ranking Framework for Bearing Fault Classification with Vibration Signals. *IEEE Access, 6, pp.44731-44746.* (Ahmed and Nandi, 2018b).

3. H. O. A. Ahmed, M. L. D. Wong, and Asoke K. Nandi. “Intelligent condition monitoring method for bearing faults from highly compressed measurements using sparse over-complete features”. *Mechanical systems and signal processing* 2018, 99: 459-477, doi: 10.1016/j.ymssp.2017.06.027. (Ahmed et al., 2018).

1.5.2 Full-length international conference publications: (Accepted/Published)

1. H. O. A. Ahmed and A. K. Nandi, “Three-stage Method for Rotating Machine Health Condition Monitoring Using Vibration Signals,". The *9th Prognostic and System Health Management conference PHM 2018,* Chongqing, China, October. 26th to 28th, 2018.
   (Accepted for publication)
2. Ahmed, H.O.A. and Nandi, A.K., 2018, September. Intelligent Condition Monitoring for Rotating Machinery Using Compressively-Sampled Data and Sub-space Learning Techniques. In *International Conference on Rotor Dynamics* (pp. 238-251). Springer, Cham. (Ahmed and Nandi, 2018c).
3. H. O. A. Ahmed, M. L. Dennis Wong, and A. K. Nandi, "Classification of bearing faults combining compressive sampling, Laplacian score, and support vector machine," *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society,* Beijing, China, 2017, pp. 8053-8058. (Ahmed et al., 2017a).
4. H. O. A. Ahmed and A. K. Nandi, "Multiple measurement vector compressive sampling and Fisher score feature selection for fault classification of roller bearings," *2017 22nd International Conference on Digital Signal Processing (DSP)*, London, 2017, pp. 1-5. doi: 10.1109/ICDSP.2017.8096076. (Ahmed and Nandi, 2017).
5. Z. Di, X. Gong, J. Shi, H. O. A. Ahmed and A. K. Nandi, "Detection of IAD based on personality questionnaires of Chinese college students and SVMs," 2017 10th International Congress on Image and Signal Processing, *Biomedical Engineering and Informatics (CISP-BMEI)*, Shanghai, 2017, pp. 1-6. doi: 10.1109/CISP-BMEI.2017.8302282.
6. H. O. A. Ahmed, M. L. D. Wong, and Asoke K. Nandi. “Compressive sensing strategy for classification of bearing faults”. *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2017, New Orleans, USA, pp. 2182 – 2186. (Ahmed et al., 2017b).
7. H. O. A. Ahmed, M. L. D. Wong, and Asoke K. Nandi. “Effects of deep neural network parameters on classification of bearing faults”. *Proceedings of the Annual Conference of IEEE Industrial Electronics Society (IECON)*, 2016, Piazza Adua, 1 - Firenze (Florence), Italy, pp. 6329 – 6334. (Ahmed et al., 2016).

1.5.3. Conference poster (without full-length papers)

1. H. O. A. Ahmed and Asoke K. Nandi. “Intelligent Condition Monitoring Technique for Bearing Faults from Compressively-sampled Vibration Signals”. *Brunel Annual Student Research Conference.* 2017, London, United Kingdom.

Chapter 2

Review of Vibration-Based Rolling Bearings Condition Monitoring

This chapter reviews the literature of condition monitoring using vibration signals and their applications in roller bearings condition monitoring. To start with, Section 2.1 is an introduction to vibration-based condition monitoring. Section 2.2 is concerned with time domain analysis of vibration signals. Section 2.3, is devoted to an explanation of different techniques that can be used to extract various frequency spectrum features. Section 2.4 gives a description of several techniques that can be used to examine time-frequency characteristics of the time indexed series signal. Section 2.5 presents commonly appropriate linear dimensionality reduction methods that can be used for vibration signal. Section 2.6 introduces generally applicable methods that can be used for features selections. Section 2.7 provides an explanation of three commonly used classification algorithms for fault diagnosis. Some of these sections include some demonstrative examples that not meant to be complete. Section 2.8 summarises the chapter.

2.1 Introduction

Machine condition monitoring (MCM) is an essential part of Condition-based maintenance (CBM). The key motive for applying MCM is to produce useful and accurate information on the current health condition of the machine. In consequence, with this type of dependable information, the correct decision of maintenance activities (if any) are required can be made to avoid machines breakdowns. Owing to the importance of rolling bearings in rotating machines, condition monitoring of rolling element bearings have been studied extensively over the past decades. Based on types of sensor data acquired from rotating machines, bearing condition monitoring techniques can be grouped into the following: vibration monitoring, acoustic emission monitoring, electric current monitoring, temperature monitoring, chemical monitoring, and laser monitoring. The advantages and limitations of these methods can be found in (Zhou et al., 2007). Of these methods, vibration-based bearing condition monitoring has been widely studied and has become a well-accepted technique of many planned maintenance management (Lacey, 2008; Randall et al., 2011). In vibration-based condition monitoring by analysing the physical features of the acquired vibration signals, one is able to categorise the acquired vibration signal into the corresponding condition correctly, which is generally a multi-class classification problem.

As was mentioned in the previous chapter, in vibration-based CM system, first, vibration data are acquired at specific time series using a sensor, e.g., velocity sensors and accelerometers that mounted to the component of interest. Then, the collected vibration data set is analysed for the purpose of machine health diagnosis and prognosis.

Faults in rolling bearing generate a series of impulses that repeat periodically at a rate called as bearing fundamental defect frequency (BFDF), which relies on the shaft speed, the geometry of the bearing (Figure 2.1), and the site of faults. Based on the damaged part, BFDFs can be categorized into four types, bearing pass frequency of outer race (BPFO), bearing pass frequency of inner race (BPFI), ball spin frequency (BSF), and fundamental train frequency (FTF), which relates to the fault at the outer race, the inner race, the rolling element, and the cage, respectively (McFadden and Smith, 1985; Rai et al., 2016). The BFDFs can be computed using the following formulas:

$$BPFO = \frac{N_b S_{sh}}{2}\left(1 - \frac{d_b}{D_p} cos\varphi\right) \tag{2.1}$$

$$BPFI = \frac{N_b S_{sh}}{2}\left(1 + \frac{d_b}{D_p} cos\varphi\right) \tag{2.2}$$

$$BSF = \frac{D_p}{2d_b}\left(1 - \left(\frac{d_b}{D_p} cos\varphi\right)^2\right) \tag{2.3}$$

$$FTF = \frac{S_{sh}}{2}\left(1 - \frac{d_b}{D_p} cos\varphi\right) \tag{2.4}$$

Here N_b is the number of rolling elements, S_{sh} is the shaft speed, d_b is the rolling element diameter, D_p is the pitch diameter, and φ is the angle of the load from the radial plane. Based on these BFDFs, the frequency of the collected bearing vibration signal indicates the source of the fault and the amplitude indicates the fault severity. However, the extraction of features from the collected bearing vibration signal that contains noise due to environmental conditions is quite difficult as it makes the interpretations of spectrum quite complex (Rai et al., 2016). For integrative fault diagnosis, several types of methods need to be adopted in a cascade of steps starting from raw vibration datasets and ending at final mature sets of results. These include vibration analysis techniques that have the ability to obtain useful information of machine condition from the raw vibration datasets, which can be successfully used for fault diagnosis. Vibration signal analysis can be performed in three main groups of waveform data analysis – time domain, frequency domain, and time-frequency domain.

2.2 Time domain analysis

The vibration signal collected from the rotating machine using vibration transducers is often in the time domain. It is a collection of time-indexed data points that collected over historical time, representing acceleration, velocity, or proximity based on the type of transducer used to collect the

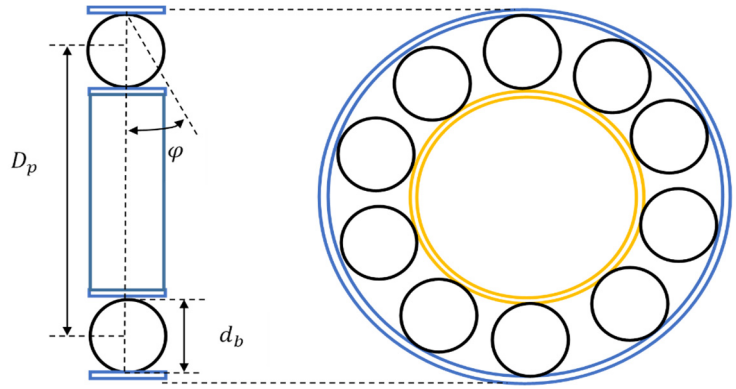


Figure 2.1. Rolling element bearing geometry.

signal. In practice, the collected vibration signal usually contains a large collection of responses from several sources in the rotating machine and some background noise. These make the direct usage of the acquired vibration signal in rotating machine fault diagnosis either manual inspection or automatic monitoring is challenging. As an alternative to processing the raw vibration signals, the common method is to compute certain features of the raw vibration signal that can describe the signal in essence (Nandi et al., 2013). In the machine learning community, these features are also called characteristics, signatures or attributes. The manual inspection of the vibration signal in the time domain fault diagnosis may be divided into two main types, namely, visual inspection and feature-based inspection.

In the visual inspection type, the machine condition can be assessed by using a measured vibration signal and compare it to a previously measured vibration signal of a machine in a normal condition, i.e., measured from a new or healthy machine. In this case, both signals should be measured on the same frequency range. A higher than normal level of vibration measurements indicates that the machine is in a fault condition, which causes the machine to produce more vibration. For instance, Figure 2.2 (a) shows a typical time domain vibration signal of a brand new condition of roller bearings and Figure 2.2 (b) represents the inner race fault condition of roller bearings. Obviously, in the case shown in this Figure, the time waveform in Figure 2.2 (b) shows spikes with high levels of amplitude in some locations of the vibration signal while other locations of the vibration remain in the lower amplitude level compared to the normal condition vibration signal in Figure 2.2 (a). This can exactly tell us that the machine is in an abnormal condition. This technique is a simple and cost-effective method of condition monitoring, which uses an oscilloscope to view the vibration signal or computer based-aids to collect data, record or display information. Readers who are interested in more details of visual inspection systems may be referred to (Davies, 1998). Nevertheless, this type of inspection is not dependable in the rotating machine condition monitoring field, because of the following four reasons: the first is that not all time waveform signals of rotating machines provide clear visual differences (Guo et al, 2005). For example, Figure 2.3 presents two typical vibration signals of roller bearing for a worn but undamaged condition (Figure 2.3 (a)) and an outer race fault condition (Figure 2.3 (b)), in this case, it is difficult to depend on visual inspection to analyse the waveforms and identify whether or not a

machine is in a fault condition; the second is that in practice we deal with a large collection of vibration signals that usually contain some background noise; the third is that we sometimes deal with low amplitude signals measured in noisy background, and the fourth is the demand for early detection of fault makes the manual inspection for all the collected signals impractical.

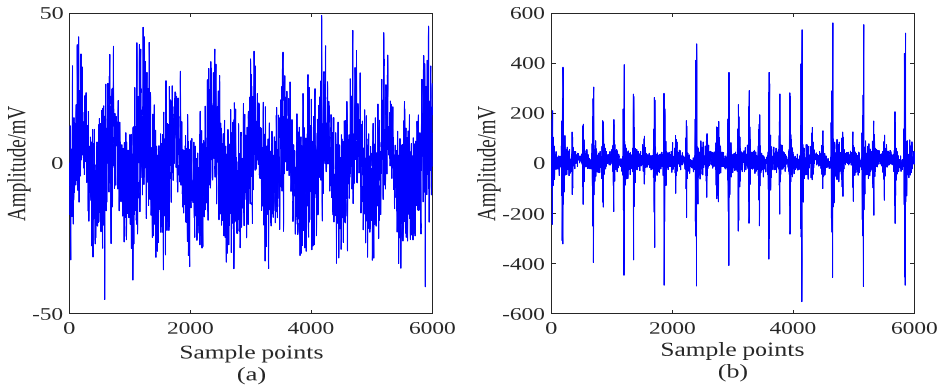


Figure 2.2 Time domain vibration signal of roller bearing (a) brand new condition and (b) inner race fault condition.

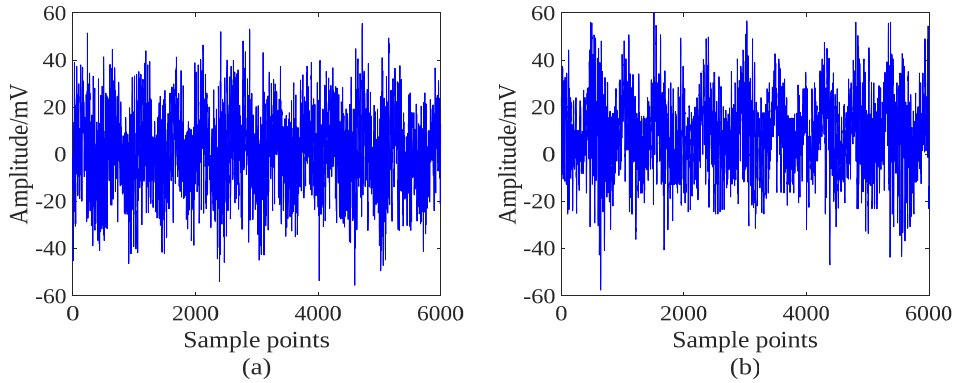


Figure 2.3 Time domain vibration signal of roller bearing (a) a worn but undamaged condition and (b) outer race fault condition.

In the feature-based type, the machine condition can be evaluated by computing certain features of the raw vibration signal. These features can be used to identify the difference between the two vibration signals.

In the case of the automatic monitoring of rotating machine, we use machine learning classifiers to classify the signal to its correct condition type using either the raw vibration signal or computed features of it in the time domain. For a complete view of the field, this section introduces vibration signal processing in the time domain, by giving an explanation of statistical functions and other advanced techniques that can be used to extract features, i.e., basic signal information, from time indexed raw vibration dataset, which can represent sufficiently machine health conditions. This will position time domain - based features in its place within the wider context of machine fault diagnosis. The other two types of vibration signal analysis, i.e., frequency domain and time-frequency domain analysis, will be covered in details in section 2.3 and section 2.4 respectively.

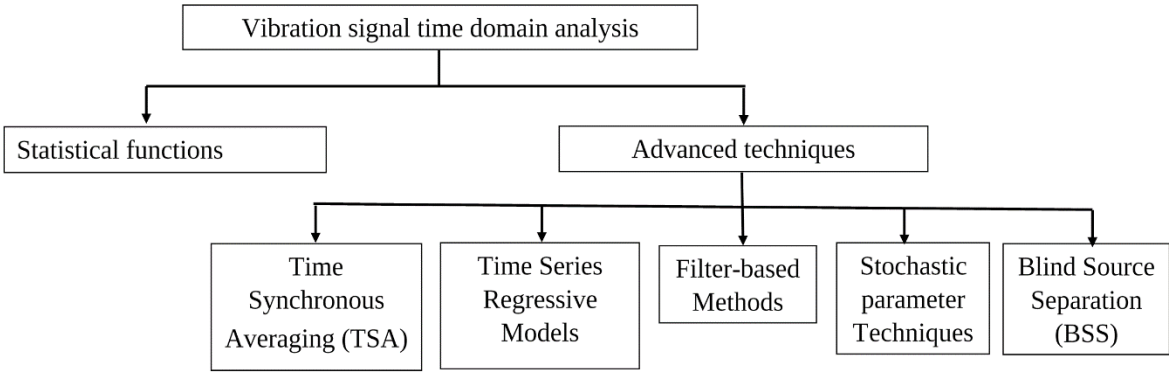


Figure 2.4 Vibration signal time domain analysis techniques

Various time domain based techniques are used for vibration signal analysis. They are summarised in Figure 2.4.

2.2.1. Statistical Functions

The acquired vibration signals are usually obtained from several sources in a rotating machine with random behaviour. With their randomness characteristics, these vibration signals cannot be defined by a direct mathematical formula and can be analysed only using statistical techniques with respect to time. Therefore, it is not surprising to find earlier works in this area focus on time domain descriptive statistics - based features that can be used either for manual inspection or automatic monitoring (Randall, 2011; Nandi et al, 2013). Numerous types of statistical functions have been heavily used to extract features from vibration signals in the time domain based on the signal amplitude. The subsequent subsections discuss these statistical functions in more detail.

2.2.1.1. Peak Amplitude

Peak amplitude, x_p is the maximum positive amplitude of the vibration signal $x(t)$ that also can be defined as half the difference between the maximum and minimum vibration amplitude, i.e., the maximum positive peak amplitude and the maximum negative peak amplitude. This can be mathematically given by Equation (2.5),

$$x_p = \frac{1}{2}[x_{max}(t) - x_{min}(t)] \tag{2.5}$$

2.2.1.2. Mean Amplitude

Mean amplitude, $\bar{x}$ is the average of the vibration signal $x(t)$ over a sampled interval, which can be computed using the following Equation (2.6),

$$\bar{x} = \frac{1}{T}\int x(t)\, dt \tag{2.6}$$

where T is the sampled signal duration. For the discrete sampled signal Equation (2.6) can be rewritten as Equation (2.7),

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2.7}$$

where N is the number of sampled points and x_i is an element of signal x.

2.2.1.3. Root Mean Square Amplitude

The root mean square (RMS) amplitude, x_{RMS} is the variance of the vibration signal magnitude. The mathematical expression of x_{RMS} is shown in Equation (2.8),

$$x_{RMS} = \sqrt{\frac{1}{T}\int |x(t)|^2 dt} \tag{2.8}$$

where T is the sampled signal duration and $x(t)$ is the vibration signal. RMS amplitude is resilient to spurious peaks in the steady state operating condition. If the vibration signal is in discrete form, Equation (2.8) can be rewritten as Equation (2.9.)

$$x_{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} |x_i|^2} \tag{2.9}$$

2.2.1.4. Peak to Peak Amplitude

The peak to peak amplitude that also called range, x_{p-p} is the range of the vibration signal, $x_{max}(t) - x_{min}(t)$ which denotes the difference between the maximum positive peak amplitude and the maximum negative peak amplitude.

2.2.1.5. Crest Factor (CF)

The crest factor, x_{CF} is defined as the ratio of peak amplitude value,x_p and the root mean square amplitude,x_{RMS} of the vibration signal. This can be computed using the following Equation (2.10),

$$x_{CF} = \frac{x_p}{x_{RMS}} \tag{2.10}$$

The crest factor is useful in detecting early stages of a fault condition and is used in on-line monitoring (Nandi et al, 2013). It is often utilised as a measure of the impulsive nature of a vibration signal that will give basic information of how much changing is occurring in a normal condition vibration waveform. For instance, in a fixed period of a pure sine wave (x) (Figure 2.5), with 100 samples, a maximum positive amplitude of 1, and a maximum negative amplitude of -1, the x_{RMS} value is equal to 0.707 and x_{CF} is 1.414. Hence, a signal with a value of x_{CF} higher than 1.414 indicates an abnormal state in the signal.

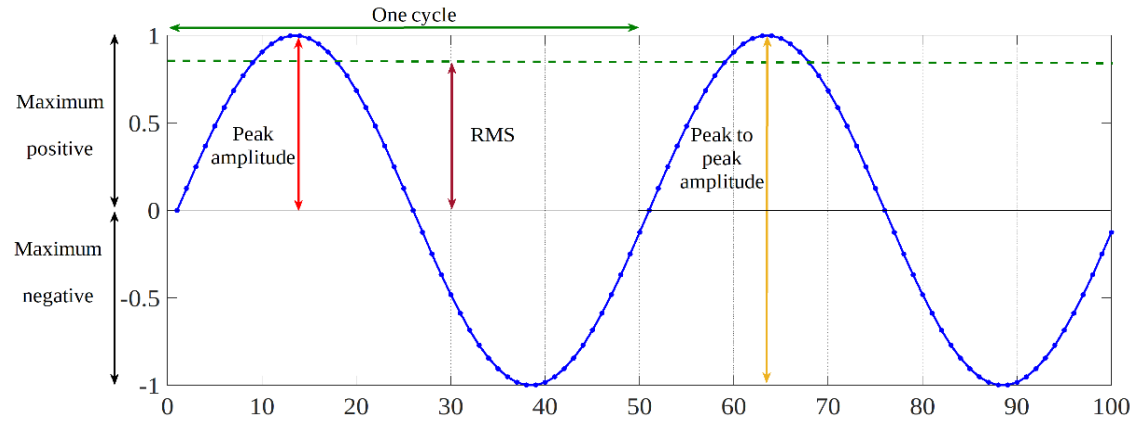


Figure 2.5. A pure sine wave with amplitude of 1 and 100 sample points.

2.2.1.6. Variance and Standard Deviation

The variance, σ_x^2 defines deviations of the vibration signal energy from the mean value, which can be mathematically given by Equation (2.11),

$$\sigma_x^2 = \frac{\sum(x_i - \bar{x})^2}{N-1} \tag{2.11}$$

The square root of the variance, i.e., σ_x is called the standard deviation of the signal x, and is expressed as shown in Equation (2.12),

$$\sigma_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{N-1}} \tag{2.12}$$

here x_i represents an element of x, $\bar{x}$ is the mean of x and N is the number of sampled points.

2.2.1.7. Standard Error

The standard error of a predicted y for an individual x in the regression, y_{STE} can be expressed in the following Equation (2.13),

$$y_{STE} = \sqrt{\frac{1}{n-1}\left[\sum(y - \bar{y})^2 - \frac{[\sum(x - \bar{x})(y - \bar{y})]^2}{\sum(x - \bar{x})^2}\right]} \tag{2.13}$$

here *n* is the sample size, $\bar{x}$ and $\bar{y}$ are the sample means.

2.2.1.8. Zero Crossing

The digitized vibration signal has a portion above the zero and a portion below the zero therefore when the signal crosses the x-axis the amplitude value is equal to zero. This location of the signal crosses the x-axis is called a zero crossing. Hence, the zero crossing,x_{ZC} can be identified as the number of times the signal crosses the x-axis if it satisfies the following criteria in Equation (2.14),

$$x_i > 0 \; and \; x_{i-1} < 0$$

$$x_i < 0 \; and \; x_{i-1} > 0 \tag{2.14}$$

where x_i is the current signal value and x_{i-1} is the previous signal value. To avoid background noise contained in the vibration signal, a threshold τ may be used instead of the amplitude value of zero, i.e., instead of counting the number of crossing when the amplitude value equal zero, the number of crossing may be counted for the amplitude of value τ (Figure 2.6) such that

$$|\, x_i - \, x_{i-1}| \geq \tau \tag{2.15}$$

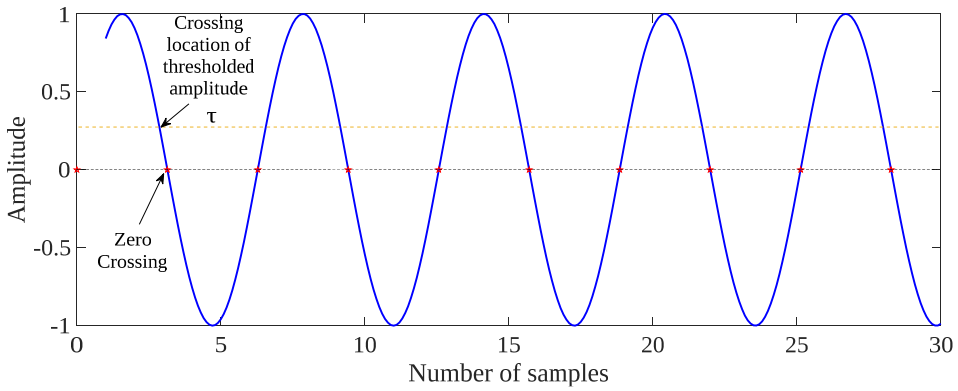


Figure 2.6 Crossing locations of amplitude equal to zero and a τ thresholded amplitude.

In ZC-based feature extraction, the density of the time intervals between successive ZC and the excess threshold measurement, are the two commonly used measurements for representing the information enclosed in the ZC features. William and Hoffman empirically demonstrated that ZC features extracted from time domain vibrations using the duration between successive ZC are useful in the early detection and identification of bearing faults (William and Hoffman, 2011).

2.2.1.9. Wavelength

The wavelength,x_{WL} is a measure of distance between two consecutive positive peaks or negative peaks of the vibration signal that can be computed using Equation (2.16),

$$x_{WL} = \sum_{i=1}^{N} |x_i - \, x_{i-1}| \tag{2.16}$$

The value of x_{WL} decreases as the frequency of the vibration signal increases.

2.2.1.10. Willison Amplitude

The Willison amplitude,x_{WA} is defined as the number of times that the difference between a vibration signal amplitude amongst two adjacent samples exceeds a predefined threshold τ. This can be computed using the following Equation (2.17)

$$x_{WA} = \sum_{i=1}^{N} f(|x_i - \, x_{i+1}|)$$

$$s.t \quad f(x) = \begin{cases} 1 & if\ x \geq \tau \\ 0 & otherwise \end{cases} \tag{2.17}$$

2.2.1.11. Slope Sign Change

The slope sign change, x_{ssc} is defined as the number of times that the slope of the vibration signal changes sign. Similarly to x_{ZC}, x_{ssc} needs to introduce a threshold τ to reduce the background noise made at slope sign changes. Given three continuous data points, x_{i-1}, x_i and x_{i+1}, x_{ssc} can be computed using Equation (2.18)

$$x_{ssc} = \sum_{i=1}^{N}[g((x_i - x_{i-1})(x_i - x_{i+1}))]$$

$$s.t \quad g(x) = \begin{cases} 1 & if\ x \geq \tau \\ 0 & otherwise \end{cases} \tag{2.18}$$

2.2.1.12. Impulse Factor

The impulse factor, x_{IF}, is defined as the ration of the peak value to the average of the absolute value of the vibration signal that can be expressed as in Equation (2.19),

$$x_{IF} = \frac{x_{peak}}{\frac{1}{N}\sum_{i=1}^{N}|x_i|} \tag{2.19}$$

The impulse factor is useful in measuring the impact of a fault generated in the vibration signal.

2.2.1.13. Margin Factor

The margin factor,x_{MF} can be calculated using the following Equation (2.20),

$$x_{MF} = \frac{x_{peak}}{(\frac{1}{N}\sum_{i=1}^{N}\sqrt{|x_i|})^2} \tag{2.20}$$

The margin factor value changes significantly with the change in the peak value, which makes it very sensitive to impulse fault especially.

2.2.1.14. Shape Factor

The shape factor, x_{SF}, is defined as the ration of the root mean square value to the average of the absolute value of the vibration signal that can be expressed as in Equation (2.21),

$$x_{SF} = \frac{x_{RMS}}{\frac{1}{N}\sum_{i=1}^{N}|x_i|} \tag{2.21}$$

The shape factor is useful in measuring the change resulted in the vibration signal under defects of unbalance and misalignment.

2.2.1.15. Clearance Factor

The clearance factor, x_{CLF}, is defined as the ration of the maximum value of the input vibration signal to the mean square root of the absolute value of the vibration signal that can be expressed as in Equation (2.12),

$$x_{CLF} = \frac{x_{max}}{(\frac{1}{N}\sum_{i=1}^{N}\sqrt{|x_i|})^2} \tag{2.22}$$

2.2.1.16. Skewness

The skewness that also called the third normalized central statistical moment, x_{SK}, is a measure of the asymmetry behaviour of the vibration signal through its probability density function (PDF), i.e., it measure whether the vibration signal is skewed to the left or to the right side of the distribution of the normal state of the vibration signal. For a signal with N sample points, x_{SK} can be presented by Equation (2.23),

$$x_{SK} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^3}{N\,\sigma_x^3} \tag{2.23}$$

The value of x_{SK} for the normal condition is often equal to zero.

2.2.1.17. Kurtosis

The kurtosis that also called the fourth normalized central statistical moment, x_{KURT}, is a measure of the peak value of the vibration signal through its probability density function (PDF), i.e., it measure whether the peak is higher or lower than the peak of the distribution corresponding to a normal condition of the vibration signal. For a signal with N sample points, x_{KURT} can be formulated as shown in Equation (2.24),

$$x_{KURT} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^4}{N\,\sigma_x^4} \tag{2.24}$$

Other higher moments from fifth (HOM5) to ninth central statistical moments (HOM9) can be calculated by raising the power expression in Equation (2.24) correspondingly. These can be represented by the following Equations (2.25), (2.26), (2.27), (2.28), and (2.29).

$$HOM5 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^5}{N\,\sigma_x^5} \tag{2.25}$$

$$HOM6 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^6}{N\,\sigma_x^6} \tag{2.26}$$

$$HOM7 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^7}{N\,\sigma_x^7} \tag{2.27}$$

$$HOM8 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^8}{N\,\sigma_x^8} \quad (2.28)$$

$$HOM9 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^9}{N\,\sigma_x^9} \quad (2.29)$$

2.2.1.18. Histogram

The histograms can be assumed as a discrete PDF of the vibration signal. Two types of features can be obtained from the histogram, namely, lower bound of the histogram (LB) and upper bound histogram, which can be expressed using the following Equation (2.30) and (2.31),

$$LB = x_{min} - 0.5\,(\frac{x_{max} - x_{min}}{N-1}) \quad (2.30)$$

$$UB = x_{max} - 0.5\,(\frac{x_{max} - x_{min}}{N-1}) \quad (2.31)$$

2.2.19 Normal/Weibull Negative log-likelihood value

The negative log-likelihood of the time domain vibration signal (x) can be expressed using the following Equation (2.32),

$$-logL = -\sum_{i=1}^{N} log[f(a, b\backslash x_i)] \quad (2.32)$$

where $f(a, b\backslash x_i)$ is the PDF of the vibration signal. The normal negative log-likelihood (Nnl) and the Weibull negative log-likelihood (Wnl) can be used as features of the time domain vibration signals. Nnl and Wnl can be computed using Equation (2.32) where their PDFs can be expressed using Equations (2.33) and (2.34),

$$Normal\ PDF = \frac{1}{\sigma\sqrt{2\pi}}\ exp^{-(x_i - \frac{\mu}{2\sigma^2})^2} \quad (2.33)$$

$$Weibull\ PDF = \frac{b}{a}\,(\frac{x_i}{a})^{b-1} exp^{-(\frac{x_i}{a})} \quad (2.34)$$

here μ is the signal mean and σ is the standard deviation.

A considerable amount of literature has been published on rolling bearing vibration monitoring using statistical time domain techniques to pre-process the vibration signals as input features individually or in a combination with other techniques. These studies are summarised in Table 2.1. As can be seen from this table, all the listed studies used more than one-time domain statistical techniques to extract features from the raw vibration data. Each study, at most, used six techniques and some used more than ten techniques to extract features from the raw vibration data. Of these techniques, kurtosis technique is used in all the mentioned studies. Moreover, skewness, shape factor, impulse factor, variance, crest factor, peak-to-peak, root mean square, and mean are among the most used techniques in these studies.

Table 3.1. Summary of the time domain statistical features that have been used in different studies of machine condition monitoring.

Studies	PEAK	MEAN	RMS	MAX	MIN	SUM	P-P	CF	VR	STD	ZC	WL	WA	SSC	IF	MF	SF	CLF	SK	KURT	HIST	Nnl	WnL	STE	LF	HO5 to HO9
McCormick and Nandi, 1996.		✓							✓																	
McCormick and Nandi, 1997.	✓							✓											✓	✓						
McCormick and Nandi, 1998.	✓							✓											✓	✓						
Jack and Nandi, 2000.	✓							✓											✓	✓						
Samanta *et al*, 2003.		✓	✓						✓										✓	✓						✓
Sun *et al*, 2004.	✓		✓					✓							✓			✓		✓						
Guo *et al*, 2005.		✓							✓										✓	✓						
Zhang, et al., 2005		✓							✓										✓	✓						
Rojas and Nandi, 2006.	✓							✓											✓	✓						
Saxena and Saad, 2006.	✓							✓											✓	✓						
Yang *et al*, 2007		✓	✓				✓	✓	✓						✓		✓		✓	✓						
Sugumaran *et al*, 2007.				✓	✓	✓	✓												✓	✓						
Sassi et al, 2007.	✓	✓	✓					✓							✓		✓			✓						
Sreejith *et al*, 2008.	✓		✓					✓		✓					✓		✓	✓		✓		✓	✓			
Chebil *et al*, 2011.	✓	✓	✓					✓							✓		✓	✓		✓						
Kankar *et al*, 2011.		✓					✓	✓		✓									✓	✓						
Saimurugan *et al*, 2011				✓	✓	✓	✓		✓	✓									✓	✓						
Yiakopoulos *et al*, 2011			✓					✓	✓						✓		✓	✓	✓	✓						
Sugumaran *et al*, 2011				✓	✓	✓	✓		✓	✓									✓	✓	✓			✓		
Prieto *et al*, 2013.		✓	✓	✓				✓	✓	✓					✓		✓		✓	✓					✓	✓
Lakshmi *et al*, 2014.		✓	✓						✓										✓	✓						✓
Ali *et al*, 2015.			✓				✓	✓							✓	✓	✓		✓	✓						
Rauber *et al*, 2015			✓				✓	✓							✓	✓	✓		✓	✓						
Nayana *et al*, 2017.		✓	✓						✓	✓	✓	✓	✓	✓					✓	✓						
Tahir *et al*, 2017		✓	✓				✓	✓	✓						✓		✓		✓	✓						

VR, Variance; STD, Standard deviation; HIST, Histogram; STE, Standard error; HO5 – HO9, Fifth higher order moment to ninth higher order moment.

2.2.2 Time Synchronous Averaging (TSA) Method

The time synchronous averaging, x_{TSA} can be defined as a periodicity feature of the vibration signal. It extracts periodic waveform from noisy vibration data, which was first introduced by Braun (Braun, 1975) and is still of interest to rotating machine condition monitoring research especially gearbox condition monitoring (Wegerich, 2004; Combet and Gelman, 2007; Bechhoefer and Kingsley, 2009; Ha et al, 2016). Also, it has been used for bearing fault diagnosis (McFadden and Toozhy, 2000; Christian et al, 2007; Ahamed et al, 2014). This technique can be performed by averaging the time domain vibration signal in synchronization with the sampling frequency or sampling time used to acquire vibration signals from the rotating machine of interest. The mathematical expression of x_{TSA} is shown in Equation (2.35),

$$x_{TSA} = \frac{1}{N} \sum_{n=0}^{N-1} x(t + nT) \tag{2.35}$$

where T is the period of averaging and N is the number of sample points.

This technique considered one of the most useful techniques for vibration signal analysis that remove any periodic events not certainly synchronous with specific sampling frequency or sampling time of a vibration signal. This also may allow time domain vibration signals concealed in noise to be viewed. For example, Figure 2.7 (a) presents a typical vibration signal of a roller bearing for inner race fault condition and Figure 2.7 (b) shows its 12 kHz corresponding synchronized signal in red and time-synchronous averaging signal in blue.

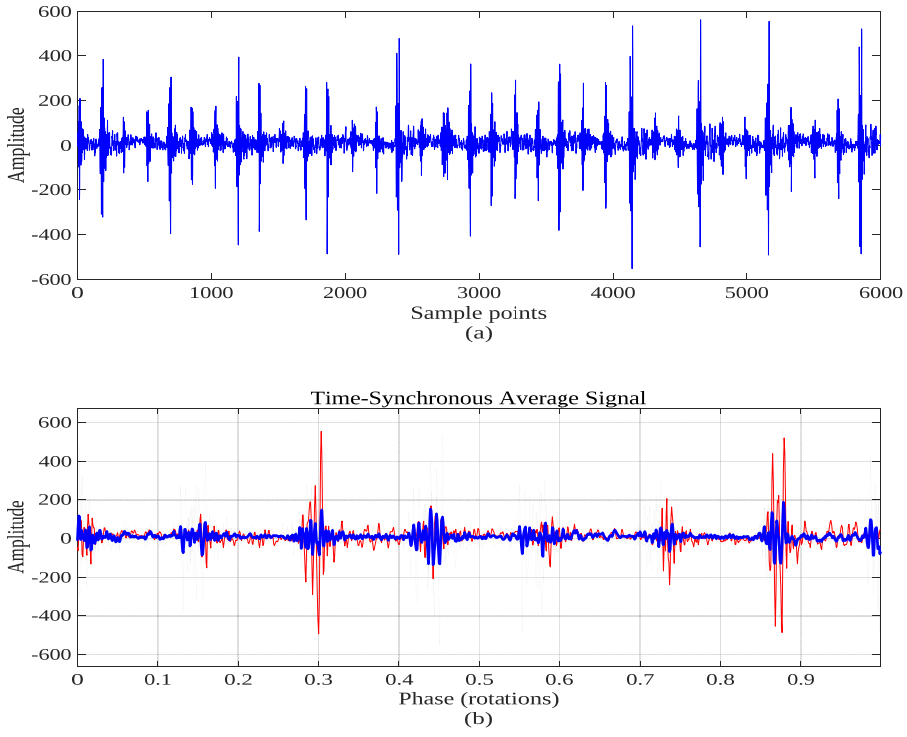


Figure 2.7. Example of (a) Time domain vibration signal of roller bearing with inner race fault condition, and (b) its time-synchronous average signal.

2.2.3 Time Series Regressive Models

Model-based techniques for vibration monitoring can provide a means of detection machine faults even if data are only available from machine in its normal condition (McCormick et al, 1998). In regressive model-based vibration monitoring, autoregressive (AR), autoregressive moving average (ARMA) which is also known as the mixture of AR and moving average (MA), and autoregressive integrated moving average (ARIMA) have been the most used techniques. The subsequent subsections discuss those types of models in more detail.

2.2.3.1 AR model

The autoregressive model (AR), AR (p), is basically a linear regression analysis of the current signal values, i.e., the estimated signal values, of the vibration time series against previous values of the time series, i.e., the values of the measured time series signal. This can be expressed mathematically as in Equation (2.36),

$$x_t = a_1 x_{t-1} + a_2 x_{t-2} + \cdots + a_p x_{t-p} + \mu_t = \mu_t + \sum_{i=1}^{p} a_i \, x_{t-i} \tag{2.36}$$

here a_1 to a_p are the model parameters, μ_t is the white noise that also called random shock or innovation, and p is the model order. Usually, the Yule-Walker equations can be used to estimate the parameters of AR model for a given vibration signal, and the model order can be chosen using Akaike's Information Criterion (AIC) (McCormick et al, 1998; Ayaz, 2012). Figures 2.8 (a) to 2.8 (d) show examples of the original Autogressive signal and the linear predictor-based estimated signal of a brand new condition bearing vibration signal (see Figure 2.2(a)) using different values of p.

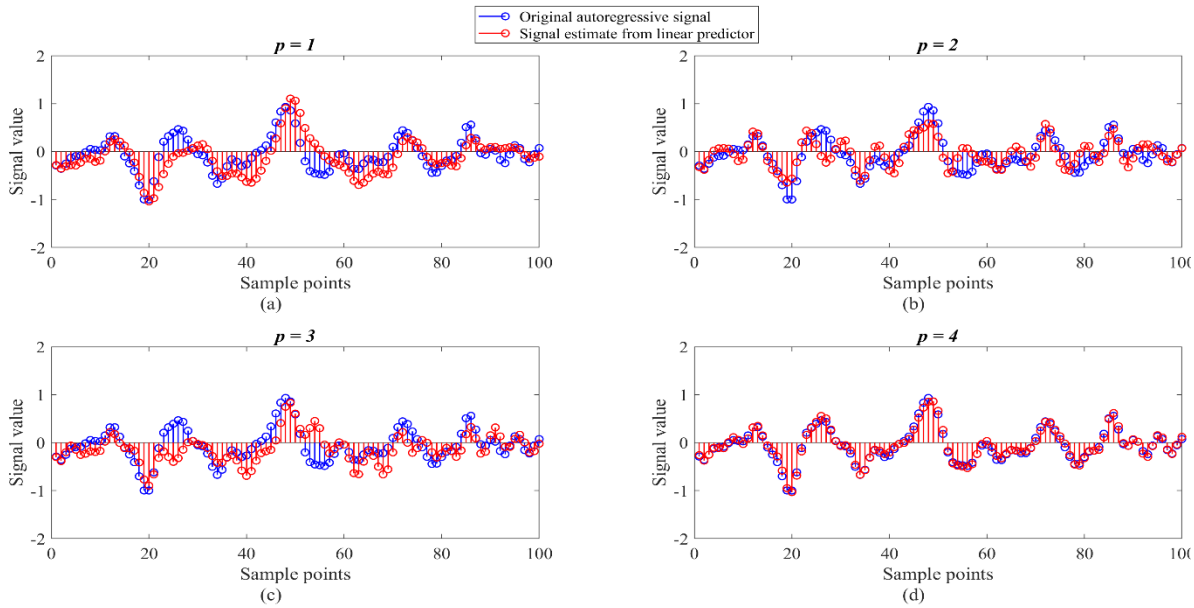


Figure 2.8. Examples of The original Autogressive signal and the linear predictor-based estimated signal of a brand new condition bearing vibration signal using different values of *p*.

2.2.3.2 MA model

The moving average (MA) model, MA (q), is basically a linear regression analysis where the current signal series is modelled based on the weighted sum of values of the time series. This can be expressed as in Equation (2.37),

$$x_t = b_1\mu_{t-1} + b_2\mu_{t-2} + \cdots + b_q\mu_{t-q} + \mu_t = \mu_t + \sum_{i=1}^{q} b_i\,\mu_{t-i} \tag{2.37}$$

here b_1 to b_q are the model parameters, μ is the white noise, and q is the model order

2.2.3.3 ARMA model

The autoregressive moving average (ARMA) model, ARMA (p, q), is a combination of AR (p) and MA (q) to achieve better flexibility in fitting of actual time series (Box et al., 2015). This can be expressed mathematically in Equation (2.38),

$$\begin{aligned} x_t &= a_1x_{t-1} + \ldots + a_px_{t-p} + \mu_t + b_1\mu_{t-1} + \ldots + b_q\mu_{t-q} \\ &= \mu_t + \textstyle\sum_{i=1}^{p} a_i\,x_{t-i} + \sum_{i=1}^{q} b_i\,\mu_{t-i} \end{aligned} \tag{2.38}$$

where a_1 to a_p and b_1 to b_q are the model parameters, μ is the white noise, and p and q are the model orders for AR and MA respectively.

2.2.3.4 ARIMA model

The AR and ARMA models, which we discussed above can be used for stationary time series vibration data. Box and colleagues (Box et al., 2015) brought the idea of using ARMA model for applications of non-stationary time series by applying the differencing technique, is done by computing the difference between consecutive observations, on the non-stationary time series to become stationary. This developed model is normally called autoregressive integrated moving average, ARIMA (p, D, q), which is a combination AR (P), integration (I), and MA (q), where p and q are the model orders for AR and MA respectively, and D is the number of differencing operators. This can be represented in Equation (2.39),

$$\Delta^D x_t = a_1\Delta^D x_{t-1} + \ldots + a_p\Delta^D x_{t-p} + \mu_t + b_1\mu_{t-1} + \ldots + b_q\mu_{t-q} \tag{2.39}$$

where Δ^D is the difference, a_1 to a_p and b_1 to b_q are the model parameters, μ is the white noise, and p and q are the model orders for AR and MA respectively.

There are also many other types of regressive models. Readers who are interested in more details of the above-introduced algorithms and other types of algorithms may be referred to (Palit and Popovic, 2006; Box et al., 2015).

Numerous studies have used autoregressive models for bearing fault diagnosis. For instance, Baillie and Mathew compared three techniques of AR modeling for their performance and reliability under conditions of several vibration signal lengths of induced faults in a rolling element bearing (Baillie and Mathew, 1996). McCormick and colleagues (McCormick et al., 1998) demonstrated the application of periodic time-variant AR models to the problems of fault detection and diagnosis in machinery. Poulimenos and Fassois (Poulimenos and Fassois, 2006) presented a survey and comparisons of time-dependent ARMA-based methods for non-stationary random vibration modeling and analysis. Furthermore, Li and colleagues (Li et al., 2007) proposed a higher-order-statistics based ARMA model for bearing fault detection that has the ability to eliminate the effects of noise and obtain clearer information of bearing fault. Recently, Ayaz (Ayaz, 2014) investigated the effectiveness of AR modeling with the order in the range of 1 to 200 for extracting bearing fault characteristics from the vibration signal.

2.2.4 Filter-based Methods

The acquired vibration signal often contains some background noise and interferences with unknown frequency contents. Many researchers have utilised filter-based methods to remove noise and isolate signals from raw signals. These include demodulation, Prony model, and adaptive noise cancellation (ANC). The subsequent subsections discuss those types of methods in more detail.

2.2.4.1 Demodulation

The demodulation process is the inverse of the modulation process, which can be amplitude demodulation or phase demodulation. That amplitude demodulation process that also goes by the name “high-frequency resonance”, “resonance demodulation”, or “envelop analysis” separates low-level frequency from high-frequency background noise (Singh and Vishwakarma, 2015). The demodulation process consists of three steps, first the raw vibration signal is band-pass filtered, then rectified or enveloped by folding the bottom part of the time waveform on the top part, this usually done using Hilbert-Huang Transform (HHT), and finally transformed utilising Fast Fourier Transform (FFT). Many researchers have utilised envelop analysis for bearing fault diagnosis (Toersen, 1998; Randall et al., 2000; Konstantin-Hansen, 2003; Patidar and Soni, 2013).

2.2.4.2 Prony model

Similar to AR, ARMA, and ARIMA models, Prony model attempts to fit a model to the sampled data. It computes the modal information such as amplitude, damping, frequency, and phase shift, which can be utilised for fault diagnosis or to recover the original signal. Given a time series signal x (t), its Prony model can be computed using the following Equation (2.40),

$$x(t) = \sum_{i=1}^{n} A_i e^{-\sigma_i t} \cos(w_i t + \phi_i) \qquad (2.40)$$

where A_i is the amplitude, w_i is the angular frequency, σ_i is the damping coefficients, and ϕ_i is the phase shift of the *i*-th sample (Tawfik and Morcos, 2001).

This method formed the central focus of a study by Chen and Mechefske in (Chen and Mechefske, 2002) in which the authors found that the Prony model-based method is able to be an efficient machine fault diagnosis technique using short duration transient vibration signals.

2.2.4.3 Adaptive Noise Cancellation (ANC)

The adaptive noise cancellation (ANC) is a technique to eliminate the background noise or interference contained in the time waveform. This technique utilises a primary input signal that contains the background noise and a reference input signal that contains noise related to the main input. To obtain the estimated signal, the reference signal is adaptively filtered and subtracted from the primary input signal (Widrow et al., 1975). This reference signal is often acquired from one or more sensors located at points in the noise field where the signal of interest is weak or undetectable (Benesty and Huang, 2013). ANC has been used for roller bearing fault diagnosis by many studies (Chaturvedi and Thomas, 1982; LI and FU, 1990; Lu et al., 2009; Elasha et al., 2016).

2.2.5 Stochastic parameter techniques

Stochastic parameters such as Chaos, correlation dimension, and thresholding methods, i.e., soft threshold and hard threshold, have been considered as effective techniques to analyse time series vibration signal. For instance, Jiang and Shao proposed a fault diagnosis method for rolling bearings using chaos theory (Jiang and Shao, 2014). Logan and Mathew introduced the correlation dimension technique for vibration fault diagnosis of rolling bearings (Logan and Mathew, 1996). Yang and colleagues applied the capacity dimension, information dimension, and correlation dimension to classify fault conditions of rolling bearings (Yang, et al., 2007). Moreover, the application of thresholding methods for rolling bearings is studied in (Lin and Qu, 2000).

2.2.6 Blind Source Separation (BSS)

The blind source separation (BSS) is a signal processing method that recovers the unobserved signals from a set of observations of numerous combinations of them (Gelle et al., 2003). This technique is useful in cases where there is a lack of knowledge about the different combinations of signals received by each sensor. The BSS model assumes the presence of L statistically independent signals $X(n) = [x_1(n), \dots, x_L(n)]$ and the observations of L signals $Y(n) = [y_1(n), \dots, y_L(n)]$ and this can be expressed by the following Equation (2.41),

$$Y(n) = f(X(n), X(n-1), \dots, X(0) + \aleph(n) \qquad (2.41)$$

where f is a nonlinear and time dependent function and $\aleph(n)$ is an additive noise. More details of the mathematical formulation of BSS can be found in (Yu et al., 2013).

Many researchers considered BSS-based techniques for machine fault diagnosis. For example, Gelle and colleagues in (Gelle et al., 2003) examined the recovery of the vibration information acquired from a single rotating machine working in a noisy environment by separating the sensor signal from the influence of another working machine. Thus, they demonstrated that BSS can be employed as a pre-processing step for fault diagnosis of rotating machine using vibration signals. A BSS-based technique that involves a signal separation in a context of spatially correlated noise is proposed in (Serviere and Fabry, 2005). Chen et al., presented a fault diagnosis method for rolling bearings based on BSS to separate signals collected from rolling bearings and gearbox (Chen, et al., 2012).

2.3 Frequency domain analysis

The frequency domain analysis techniques have the ability to divulge some information based on the frequency characteristics that are not easily observed in time-domain. In practice, the time-domain signal is transformed into frequency-domain representation by using Fourier Transform (FT). There are two main types of FT, Continuous Fourier Transform (CFT) and the Discrete Fourier Transform (DFT). Of these, DFT is an important tool in the frequency analysis of discrete time signals (Van Loan, 1992; Diniz et al., 2010).

2.3.1 Discrete Fourier transform (DFT)

The discrete Fourier transform (DFT) of a discrete signal $x(n)$ can be defined in Equation (2.42),

$$X_{DFT}(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi nk/N}, \qquad k = 0, 1, \dots, N-1 \tag{2.42}$$

Or more efficiently

$$X_{DFT}(k) = \sum_{n=0}^{N-1} x(n) {W_N}^{nk}, \qquad k = 0, 1, \dots, N-1 \tag{2.43}$$

where

$$W_N = e^{-\frac{j2\pi}{N}} = \cos\left(\frac{2\pi}{N}\right) - j\, sin\, \left(\frac{2\pi}{N}\right) \tag{2.44}$$

The inverse of DFT that transforms $X_{DFT}(k)$ back into $x(n)$ can be expressed using the following Equation (2.45),

$$x(n) = \frac{1}{N} \sum_{n=0}^{N-1} X_{DFT}(k) {W_N}^{-nk}, \qquad n = 0, 1, \dots, N-1 \tag{2.45}$$

To compute the DFT of a signal of length N, one needs N^2 complex multiplications that limits its practical use for signals with a large number of samples (Diniz et al., 2010).

2.3.2 Fast Fourier Transform (FFT)

The Fast Fourier Transform (FFT) is an efficient algorithm that computes the DFT and its inverse of stationary time series signal with a significant reduction in the complexity. In fact, FFT computes the DFT of a signal of length N using $N\,log_2\,N$ complex multiplications instead of N^2 complex multiplications. FFT technique was first proposed by Cooley and Tukey (Cooley and Tukey, 1965) based upon sparse factorization of the Discrete Fourier Transform (DFT) (Van Loan, 1992). Briefly, we present the simplified FFT form as follows:

Suppose a discrete time series signal *x* with length N sample points is separated into two parts, x_{even} that composed of even-numbered points and x_{odd} that composed of the odd-numbered points, each of which has half of the total sampled points (N/2). Based on x_{n-even} and x_{n-odd} the DFT of signal *x (n)* can be given by the following Equation (2.46),

$$X_{DFT}(k) = \sum_{n-even=0}^{(\frac{N}{2})-1} x_{n-even} {W_N}^{nk} + \sum_{n-odd=0}^{(\frac{N}{2})-1} x_{n-odd} {W_N}^{nk} \ , \quad k = 0, 1, \dots, N-1 \quad (2.46)$$

More details of the mathematical formulation of FFT for DFT calculation can be found in (Cochran et al., 1967; Van Loan, 1992; Diniz et al., 2010).

2.3.3 Frequency spectrum statistical features

Various statistical features can be computed from the transformed signal in the frequency domain. Of these features, spectral kurtosis has been investigated in many studies of rolling bearings fault diagnosis. For example, Vrabie et al., studied the application of spectral kurtosis to rolling bearings fault detection (Vrabie et al., 2004). Tian and colleagues (Tian et al., 2016) proposed a method for fault detection and degradation of bearings in electric motors. This method first extracts fault features using spectral kurtosis and cross-correlation and then combined these features using principal component analysis (PCA) and a semi-supervised k-nearest neighbor (KNN). Furthermore, a good review of the application of spectral kurtosis for fault detection, diagnosis, and prognostics of rotating machines can be found in (Wang, et al., 2016). Moreover, higher-order spectra, bispectral, and trispectral have been applied successfully for machine condition monitoring (McCormick et al., 1998; McCormick and Nandi, 1999).

2.4 Time-frequency domain analysis

The time-frequency domain has been used for non-stationary waveform signals which are very common when machinery fault occurs. Thus far, several time-frequency analysis techniques have been

developed and applied to machinery fault diagnosis, e.g., Wavelet Transform (WT), Short Time Fourier Transform (STFT), Hilbert-Huang Transform (HHT), Empirical Mode Decomposition (EMD), and Local Mean Decomposition (LMD) etc. The subsequent Subsections describe these techniques in more details.

2.4.1 Wavelet analysis

Wavelet analysis is a tool for the analysis of non-stationary waveform signal. There are several types of wavelet families that are widely utilised in signal analysis, e.g., Haar wavelets, Coiflets, Morlet wavelet, Biorthogonal wavelets, etc. The mother wavelet $\psi(t)$ can be expressed mathematically by the following Equation (2.47),

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\,\psi\left(\frac{t-\tau}{s}\right), \tag{2.47}$$

where $s > 0$ and represent the scaling parameter, τ is the transformation parameter, and t is the time. The three main transforms in wavelets analysis are Continuous Wavelet Transform (CWT), Discrete Wavelet Transform (DWT), and Wavelet Packet Transform (WPT) (Burrus et al., 1998).

2.4.1.1 Continuous Wavelet Transform (CWT)

The continuous wavelet transform (CWT) of the time domain vibration signal (x), $W_{x(t)}(s,\tau)$ can be expressed using the following Equation (2.48),

$$W_{x(t)}(s,\tau) = \frac{1}{\sqrt{s}}\int x(t)\,\psi^*\left(\frac{t-\tau}{s}\right)dt \tag{2.48}$$

here $\psi^*\left(\frac{t-\tau}{s}\right)$ represents the complex conjugate of $\psi(t)$ that scaled and shifted using s and τ parameters respectively.

2.4.1.1 Discrete Wavelet Transform (DWT)

The discrete wavelet transform (DWT) of the time domain vibration signal (x), $W_{x(t)}(s,\tau)$ can be expressed using the following Equation (2.49),

$$W_{x(t)}(s,\tau) = \frac{1}{\sqrt{2^j}}\int x(t)\,\psi^*\left(\frac{t-k2^j}{2^j}\right)dt, \tag{2.49}$$

Here DWT represents the discrete form of CWT where $\psi_{s,\tau}(t)$discretised using dyadic scales, i.e., is $s = 2^j$ and $\tau = k2^j$, where *j* and *k* are integers (Yan et al., 2014).

In practice, the DWT can be implemented by low-pass scaling filter $h(k)$ and high-pass wavelet filter$g\,(k) = (-1)^k h(1-k)$. These filters are created from $\psi(t)$ and scaling function $\phi(t)$ that can represented by the following Equations (2.50) and (2.51).

$$\phi(t) = \sqrt{2}\sum_k h(k)\phi(2t-k) \quad (2.50)$$

$$\psi(t) = \sqrt{2}\sum_k g(k)\,\psi(2t-k) \quad (2.51)$$

2.4.1.1 Wavelet Packet Transform (WPT)

The wavelet packet transform (WPT) is an improvement of DWT, in which every detail signal obtained by DWT is further decomposed into an approximation signal and a detail signal (Shen et al., 2013). Accordingly, the time domain vibration signal $x(t)$ can be decomposed using the following Equations (2.52) and (2.53),

$$d_{j+1,2n} = \sum_m h(m-2k)d_{j,n} \quad (2.52)$$

$$d_{j+1,2n+1} = \sum_m g(m-2k)d_{j,n} \quad (2.53)$$

here *m* is the number of coefficients, $d_{j,n}$ $d_{j+1,2n}$, and $d_{j+1,2n+1}$are the wavelet coefficients at sub-band n, 2n, and 2n+1 respectively.

2.4.2 Short Time Fourier Transform (STFT)

The short time Fourier transform (STFT) of a time domain vibration signal $x(t)$,$STFT_{x(t)}(t,f)$ can be expressed using the following Equation (2.54),

$$STFT_{x(t)}(t,f) = \int_{-\infty}^{+\infty} x(\tau)w(\tau-t)\exp(-j2\pi f\tau)\,d\tau \quad (2.54)$$

where $w(\tau-t)$ is a window function, τ is a time variable

2.4.3 Hilbert-Huang Transform (HHT)

The Hilbert-Huang transform (HHT) is defined as the convolution of signal $x(t)$ with 1/t and can emphasise the local properties of $x(t)$ (Peng et al., 2005) such that,

$$y(t) = \frac{P}{\pi}\int_{-\infty}^{+\infty}\frac{x(\tau)}{t-\tau}d\tau \quad (2.55)$$

here *P* is the Cauchy principal value. The analytic signal $z(t)$ of $x(t)$ can be expressed as follows,

$$z(t) = x(t) + iy(t) = a(t)e^{i\varphi(t)} \quad (2.56)$$

where $a(t)$ is the instantaneous amplitude of $x(t)$ and can be computed using Equation (2.57),

$$a(t) = [x^2(t) + y^2(t)]^{\frac{1}{2}} \quad (2.57)$$

$\varphi(t)$ is the instantaneous phase of $x(t)$ and can be calculated by Equation (2.54),

$$\varphi(t) = \arctan(\frac{\mathrm{y(t)}}{\mathrm{x(t)}}) \tag{2.58}$$

2.4.4 Empirical Mode Decomposition (EMD)

The empirical mode decomposition (EMD) is a nonlinear and adaptive data analysis technique that decomposes the time domain signal $x(t)$ into different scales or intrinsic mode functions (IMFs) (Huang et al., 1998; Wu and Huang, 2009) such that,

$$x(t) = \sum_{j=1}^{n} c_j + r_n \,, \tag{2.59}$$

where c_j is the *j*-th IMF and r_n is the residue of data $x(t)$ after the extraction of the *n* IMFs.

2.4.5 Wigner-Ville Distribution (WVD)

The Wigner distribution (WD) for a signal $x(t)$ can be expressed mathematically by the following Equation (2.60) (Staszewski et al., 1997),

$$W_x(t,f) = \int_{-\infty}^{+\infty} x\left(t+\frac{\tau}{2}\right) x^*\left(t+\frac{\tau}{2}\right) e^{-2\pi f t} d\tau \tag{2.60}$$

where $x^*(t)$ is the complex conjugate of $x(t)$.

The Wigner-Ville Distribution (WVD) is defined as the Wigner distribution (WD) for the analytic signal $z(t)$. Here $z(t)$ can be represented mathematically by the following Equation (2.61),

$$z(t) = x(t) + j\hat{x}(t) \tag{2.61}$$

where $\hat{x}(t)$ is the Hilbert transform of $x(t)$.

2.4.6 Local Mean Decomposition

The local mean decomposition (LMD) is an adaptive analysis technique that decomposes a complicated signal into a set of product functions (PFs) that composed of frequency modulated signal and an amplitude envelop signal. The LMD procedure for a time domain signal $x(t)$ can be described as follows (Smith, 2005):

- Compute the mean value and the envelop estimate of the maximum and minimum points of each half-wave oscillation of the signal. Thus the *i*-th mean value (m_i) is given by Equation (2.62),

  $$m_i = \frac{(n_i + n_{i+1})}{2} \tag{2.62}$$

  and the *i*-th envelop estimate (a_i) is given by Equation (2.63),

  $$a_i = \frac{|n_i - n_{i+1}|}{2} \tag{2.63}$$

- Obtain the smoothed varying continuous local mean function $m_{11}(t)$ and the smoothed varying continuous envelop function $a_{11}(t)$ using MA.
- Subtract $m_{11}(t)$ from $x(t)$ to obtain the residual signal $h_{11}(t)$ such that,

$$h_{11}(t) = x(t) - m_{11}(t) \tag{2.64}$$

- Divide $h_{11}(t)$ by $a_{11}(t)$ to obtain $s_{11}(t)$ such that,

$$s_{11}(t) = h_{11}(t)/a_{11}(t) \tag{2.65}$$

- Compute envelop $a_{12}(t)$ of $s_{11}(t)$. If $a_{12}(t) \neq 1$ the process needs to be repeated for $s_{11}(t)$.
- Compute a smoothed local mean $m_{12}(t)$ for $s_{11}(t)$, subtract it from $s_{11}(t)$ to obtain $h_{12}(t)$, and divide $h_{12}(t)$ by $a_{12}(t)$ to obtain $s_{12}(t)$. Repeat this process n times until a purely frequency modulated signal $s_{1n}(t)$ is obtained.

2.4.7 Spectral Kurtosis and Kurtogram

To compute the spectral kurtosis (SK) the signal is first decomposed into the time-frequency domain where the kurtosis values are defined for each frequency group. The SK of a signal $x(t)$, $K_x(f)$ is defined as the fourth-order normalised cumulant. This can be computed using the following Equation (2.62) (Antoni and Randall, 2006),

$$K_x(f) = \frac{\langle |H(t,f)|^4 \rangle}{\langle |H(t,f)|^2 \rangle^2} - 2 \tag{2.66}$$

where $H(t,f)$ is the complex envelop function of $x(t)$ at frequency f that obtained using STFT algorithm such that,

$$H(t,f) = \int_{-\infty}^{+\infty} x(\tau) w(\tau - t) e^{-j2\pi f \tau} d\tau \tag{2.67}$$

The Kurtogram algorithm (KUR) computes SK for several window size using a bandpass filter and the fast Kurtogram algorithm uses a series of digital filters rather than STFT (Antoni, 2007; Randall, 2011).

A considerable amount of literature has been published on the use of frequency domain and time-frequency domain techniques for bearing fault diagnosis. Table 2.2 presents a summary of the frequency and time-frequency domain vibration analysis techniques that have been used in different studies of rolling bearings condition monitoring. According to these studies, these techniques can be used individually or in a mixture of several techniques to extract features from the raw vibration data.

Table 2.2. Summary of the frequency and time-frequency domain analysis techniques that have been used in different studies of rolling bearing condition monitoring.

Studies	FFT	CWT	DWT	WPT	STFT	HHT	EMD	LMD	WVD	SK	KUR
Lin and Qu, 2000; Peng et al, 2005.				✓		✓	✓				
Peter et al, 2001; Paliwal et al., 2014.	✓	✓									
Sun and Tang, 2002; Luo et al., 2003; Hong and Liang, 2007; Li et al, 2007; Zhu et al., 2009; Su et al, 2010; Kankar et al, 2011; Li et al, 2011.		✓									
Nikolaou et al, 2002; Ocak et al, 2007; Wang et al, 2015.				✓							
Prabhakar et al, 2002; Lou and Loparo, 2004; Purushotham, et al., 2005; Tyagi, 2008; Djebala et al., 2008; Xian, 2010; Kumar and Singh, 2013.			✓								
Yu et al, 2005.						✓	✓				
Junsheng et al, 2006; Yu et al, 2006; Zhao et al, 2013; Dybala and Zimroz, 2014.							✓				
Li and Zhang, 2006; Li et al, 2006.							✓		✓		
Antoni and Randall, 2006.										✓	✓
Sawalhi et al, 2007.										✓	
Rai et al, 2007; Pang et al., 2018.	✓					✓					
Li et al., 2008.			✓	✓							
Zhang et al., 2009.											✓
Immovilli et al., 2009; Wang et al, 2011.										✓	
Wensheng et al, 2010.							✓			✓	
Lei et al, 2011; Wang et al, 2013.				✓							✓
Zhou et al., 2011.									✓		
Linsuo et al., 2011.					✓				✓	✓	
Cheng et al, 2012; Liu and Han, 2014; Tian et al, 2015; Li et al, 2016.								✓			
Cozorici et al., 2012.	✓	✓	✓								
Jiang et al, 2013.				✓			✓				
Singhal and Khandekar, 2013; Lin et al., 2016.	✓										
Liu et al, 2014.						✓					
Liu et al, 2016.					✓						
Jacop et al., 2017.	✓	✓			✓						

2.5 Linear subspace learning

Linear subspace learning techniques can be used to extract a new low dimensional feature space that is usually a linear combination of the original high dimensional feature space. Of these techniques, Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Independent Component Analysis (ICA) are amongst the most common techniques that have been used in machine fault diagnosis. The subsequent subsections discuss those types of techniques in more detail.

2.5.1 Principal Component Analysis (PCA)

PCA is an orthogonal linear feature projection algorithm which aims to find all the components (eigenvectors) in descending order of significance. The procedure of PCA involves the following steps.

- Calculate the mean vector of the data.
- Compute the covariance matrix of the data.
- Obtain the eigenvalues and eigenvectors of the covariance matrix.

PCA can be used to form a low-dimensional feature vector (Wold et al., 1987). To reduce the dimensionality of the data by means of PCA, one ignores the least significant components from the PCA. Suppose that the input dataset $X = [x_1 x_2 \dots x_L]$ has L observations and *n*-dimensional space. PCA transforms *X* to $\hat{X} = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_L]$ in a new $n1$-dimensional space such that

$$\hat{X} = W^T X \tag{2.68}$$

where W is the projection matrix in which each column vector is composed of the corresponding eigenvectors of $n1$ largest eigenvalues ($n1 \ll n$) of the covariance matrix *C* that can be computed as follows

$$C = \frac{1}{L} \sum_{i=1}^{L} (x_i - \bar{x})(x_i - \bar{x})^T \tag{2.69}$$

Here $\bar{x} = \frac{1}{L} \sum_{i=1}^{L} x_i$.

2.5.2 Independent Component Analysis (ICA)

The basic assumption in the standard ICA model is that the observed data *X* is a mix of sources or a vector of independent components (s_i) (Comon, 1994) such that,

$$X = AS \tag{2.70}$$

where $X = [x_1 x_2 \dots x_L] \in R^{n \times L}$, $S = [s_1 s_2 \dots s_L] \in R^{k \times L}$ is the independent components matrix ($k \leq n$), and $A \in R^{n \times k}$ is a mixing matrix that is invertible and square in the simplest case. The

independent components can be recovered from the data by using the inverse matrix *W* of the mixing matrix *A* such that

$$\hat{S} = A^{-1}X = WX \tag{2.71}$$

The estimation of *W* is based on cost functions, also called objective functions or contrast functions. Solutions *W* are found at the minima or maxima of these functions. The estimation of the generative model in (2.70) can be done using the non-normalized likelihood and this can be represented by the following equation,

$$log\, p(x) = \sum_{i=1}^{n} G\left(W_i^T X\right) + Z(W_1, \dots, W_n) \tag{2.72}$$

where Z is the normalization constant that is equal to $-log\,|detW|$, W_i is one row in *W* matrix, and the distribution of the sources *S* is selected to be logistic, we define,

$$G(S) = -2\, log\, cosh\left(\frac{\pi}{2\sqrt{3}}\, S\right) - log4 \tag{2.73}$$

2.5.3 Linear Discriminant Analysis (LDA)

Different from PCA, searching for the most important components of samples, LDA aims to find discriminant components that distinguish different class samples (Balakrishnama and Ganapathiraju, 1998). In fact, LDA collects the samples from the same class and expand the margin of samples from different classes. Fisher LDA analysis method (Sugiyama, 2006) considers maximizing the Fisher criterion function *J (W)*, i.e., the ratio of the between the class scatter (S_B) to the within class scatter (S_w) such that

$$J(W) = \frac{|W^T S_B\, W|}{|W^T S_w\, W|} \tag{2.74}$$

$$S_B = \frac{1}{L}\sum_{i=1}^{c} l_i \left(\mu^i - \mu\right)(\mu^i - \mu)^T \tag{2.75}$$

where

$$S_w = \frac{1}{L}\sum_{i=1}^{c}\sum_{j=1}^{l_i}(x_j^i - \mu^i)(x_j^i - \mu^i)^T \tag{2.76}$$

where *L* is the total number of observations, *c* is the number of classes, and μ^i is the mean vector of class *i*. $x \in R^{n\,x\,L}$ is the training dataset, x_1^i represents the dataset belong to the *c-th* class, n_i is the number of measurements of the *i-th* class, μ^i is the mean vector of class *i*, and μ is the mean vector of all training dataset. LDA projects the space of the original data onto a (c – 1) – dimension space by finding the optimal projection matrix *W* that maximizes the *J (W)* in equation (2.74) such that,

$$\widehat{W} = arg\, \max_{W} J(W) \tag{2.77}$$

Here $\widehat{W}$ is composed of the selected eigenvectors $(\widehat{w}_1, \dots, \widehat{w}_{m2})$ with the first *n2* largest eigenvalues (*n2 = c - 1*).

Many researchers have used linear subspace learning-based methods to extract features from a large amount of collected vibration data for the purpose of roller bearings fault diagnosis. For instance, a PCA-based approach to select the most representative features for the classification bearing faults is proposed by (Malhi and Gao, 2004) and showed improvements in the classification accuracy for both NN and Radial Basis Function Network (RBFN). Widodo and colleagues developed a method that combined ICA and SVM for fault diagnosis of induction motors (Widodo, et al., 2007). Shuang and Meng proposed a method based on PCA and SVM and showed its efficiency in bearing fault diagnosis (Shuang and Meng, 2007). In a similar way, the combination of Neural Network (NN) and ICA is proposed by Chang and Jiao and they found that it can achieve a considerable classification accuracy of rotating machinery fault diagnosis (Chang and Jiao, 2012). Jiang and colleagues proposed a method for condition monitoring for rolling element bearing based on PCA and Phase Space Reconstruction (PSR) that can effectively identify different conditions of rolling element bearings (Jiang et al., 2013). Likewise, a PCA-based technique on defined time-frequency statistical features is proposed in (Wang et al., 2015) and the effectiveness of the proposed method in rolling bearing faults diagnosis was evaluated using a fuzzy C-means (FCM) model. Also, Ciabattoni and colleagues set up a series of experiments using LDA-based method and show that the proposed algorithm improves the classification accuracy when the classes of the motor bearing are overlapped (Ciabattoni et al., 2015). Overall, these studies highlight the need for feature extraction to improve the efficiency of diagnosis methods and decrease the computational time.

2.6 Feature selection

Feature selection techniques are used to select a subset of features that can sufficiently represent the characteristic of the original features. In view of that, this will reduce the computational cost and may remove irrelevant and redundant features. Feature selection methods can be categorised into three groups, supervised, unsupervised, and semi-supervised feature selection techniques. Also, it can be further grouped into filter models, wrapper models, and embedded models. The subsequent subsections discuss those types of methods in more detail.

2.6.1 Filter model-based feature selection

The filter model-based feature selection is based on measures of various characteristics of the training data such as similarity, dependency, information, and correlation. Therefore, it is fast and requires low computational complexity compared to other methods. The filtering can be performed using univariate feature filters that rank every single feature or using multivariate feature filters which evaluate a feature subset (Tang et al., 2014; Chandrashekar et al., 2014). This section gives brief descriptions of several feature ranking (FR) methods that can be used to rank the features of a vibration signal.

2.6.1.1 Fisher Score (FS)

Fisher score (FS) is a filter-based feature selection method and one of the commonly used supervised feature selection methods (Duda et al., 2001; Gu et al., 2012). The main idea of FS is to compute a subset of features with a large distance between data points in different classes and small distance between data points in the same class. To describe briefly FS method, assume the input matrix $X \in R^{n\,x\,L}$ reduces to $Z \in R^{k\,x\,L}$ matrix. The FS feature of the *i-th* can be computed by the following equation:

$$FS(X^i) = \frac{\sum_{c=1}^{C} L_c\,(\mu_c^i - \mu^i)^2}{(\sigma^i)^2} \tag{2.78}$$

where $X^i \in R^{1\,x\,L}$, L_c is the size of the *c-th* class, $(\sigma^i)^2 = \sum_{c=1}^{C} L_c\,(\sigma_c^i)^2$, μ_c^i and σ_c^i are the mean and standard deviation of *c-th* class corresponding to the *i-th* feature; μ^i and σ^i are the mean and standard deviation of the entire dataset corresponding to the *i-th* feature. Usually, the FS of each feature is computed independently.

2.6.1.2 Laplacian Score (LS)

Laplacian Score (LS) is an unsupervised filter based technique that rank features depending on their locality preserving power. In fact, LS is mainly based on Laplacian Eigenmaps and Locality Preserving Projection, and can be briefly described as follows (He et al., 2006). Given a dataset $X = [x_1, x_2, \dots, x_L]$, where $X \in R^{n\,x\,L}$, suppose the Laplacian Score of the *r-th* feature is L_r and f_{ri} represent the *i*-th sample of the *r-th* feature where $i = 1, \dots, n$ and $r = 1, \dots, L$. First the LS algorithm constructs the nearest neighbour graph G with *n* nodes, where the *i-th* node corresponds to x_i. Next, an edge between nodes *i* and *j* is placed, if x_i is among k nearest neighbors of x_j or vice versa, then *i* and j are connected. The elements of the weight matrix of graph G is S_{ij} and can be defined as follows:

$$S_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{t}}, & i \text{ and } j \text{ are connected} \\ 0, & \text{otherwise} \end{cases} \tag{2.79}$$

where *t* is an appropriate constant. The Laplacian score L_r for each sample can be computed as follows:

$$L_r = \frac{\tilde{f}_r^{\,T} L \tilde{f}_r}{\tilde{f}_r^{\,T} D \tilde{f}_r} \tag{2.80}$$

Here $D = diag(S\mathbf{1})$ is the identity matrix, $\mathbf{1} = [1, \dots, 1]^T, L = D - S$ is the graph Laplacian matrix, and $\tilde{f}_r$ can be calculated using the following equation:

$$\tilde{f}_r = f_r - \frac{f_r^T D\mathbf{1}}{\mathbf{1}^T D\mathbf{1}} \tag{2.81}$$

where $f_r = [f_{r1}, f_{r2}, \dots, f_{rn}]^T$.

2.6.1.3 Relief-F

Relief-F is a supervised feature ranking algorithm that commonly used as a pre-processing technique for a feature subset selection. Relief-F is an extension of the traditional Relief algorithm (Liu et al., 2007) that has the ability to deal with noisy, incomplete, and multi-class datasets. It uses a statistical approach to select the important features based on their weight *W*. The main idea of Relief-F is to randomly compute examples from the training data and then calculate their nearest neighbours from the same class, also called the nearest hit, and the other nearest neighbours from different class, also called the nearest miss. The function diff (Attribute, Instance1, and Instance2) is used to compute the distance between instances to find the nearest neighbours. The good attribute should have the same value of weights for instances from the same class and discriminate between instances from different classes (Kononenko et al., 1997). The procedure of Relief-F algorithm is summarized below in algorithm 1(Liu et al., 2007).

Algorithm 1 Relief-F

Input: ℓ learning instances, n features and c classes;
Probabilities of classes $p_{y,}$ Sampling parameter a; Number of nearest instances from each class d;
Output: for each feature f_i a feature weight $-1 \leq W[i] \leq 1$;
1 for $i = 1$ to n do $W[i] = 0.0$; end for;
2 for $h = 1$ to a do
3 randomly compute an instance $\mathbf{x}_k$ with class y_k;
4 for y = 1 to c do
5 find d nearest instances x[j, y] from class y, $j = 1 \ldots d$;
6 for $i = 1$ to n do
7 for $j = 1$ to d do
8 if $y = y_k$ {nearest hit}
9 then $W[i] = W[i] - \text{diff}(i, \mathbf{x}_k, \text{x}[j, y])/ (a*d)$;
10 else $W[i] = W[i] + p_y / (1 - p_{yk}) * \text{diff}(i, \mathbf{x}_k, \text{x}[j, y])/ (a*d)$;
11 end if;
12 end for; {j} end for; {i}
13 end for; {y} end for ; {h}
14 return (W);

2.6.1.4 Pearson Correlation Coefficient (PCC)

Pearson Correlation Coefficient (PCC) (Liu et al., 2007) is a supervised filter-based ranking technique that examines the relationship between two variables according to their correlation coefficient (r), $-1 \leq r \leq 1$. Here the negative values indicate inverse relations, the positive values indicate a correlated relation, and the value 0 indicates no relation. PCC can be computed as follows:

$$r(i) = \frac{cov(x_i, y)}{\sqrt{var(x_i) * var(y)}} \tag{2.82}$$

Here x_i is the i_{th} variable, y is the class labels.

2.6.1.5 Chi-Squared (Chi-2)

Feature ranking and selection using chi-square (chi-2) is based on the χ^2 test statistics (Yang et al., 1997). Chi-2 evaluate the importance of a feature by calculating the χ^2 test with respect to the class labels. The χ^2 value for each feature *f* in a class labels group c can be computed using the following equation:

$$\chi^2(f,c) = \frac{L(E_{c,f}E - E_cE_f)^2}{(E_{c,f}+E_c)(E_f+E)(E_{c,f}+E_f)(E_c+E)} \tag{2.83}$$

where L is the total number of examples, $E_{c,f}$ is the number of times *f* and *c* co-occur, E_f is the number of time the feature *f* occurs without *c*, E_c is the number of times *c* occurs without *f*, and *E* is the number of times neither *f* nor *c* occurs. The bigger value of χ^2 indicates that the features are highly related.

There are many other filter-based feature selection methods available for vibration-based fault diagnosis. These include Mutual Information (MI), Information Gain (IG), the Gain ratio (GR), and Minimum Redundancy Maximum Relevance (mRmR) (Peng et al., 2005; Kappaganthu and Nataraj, 2011; Zhang et al., 2011; Liu et al., 2014).

The use of filter model-based techniques as feature selection for diagnosis bearing faults was investigated by several researchers. For example, Tian and colleagues (Tian et al., 2012) combines wavelet packet transform (WPT), PCC, and envelop analysis to extract the bearing signals from the masking background. A multi-scale analysis technique is employed to extract the possible fault-related features in different scales, such as the multi-scale entropy (MSE), multi-scale permutation entropy (MPE), multi-scale root-mean-square (MSRMS) and multi-band spectrum entropy (MBSE). Some of the features are then selected as the inputs of the support vector machine (SVM) classifier through the Fisher score (FS) as well as the Mahalanobis distance (MD) evaluations (Wu et al., 2013). Zheng and colleagues (Zheng et al., 2014) proposed a method for rolling bearing fault diagnosis based on multi-scale fuzzy entropy (MFE), Laplacian Score (LS) and variable predictive model-based class discrimination (VPMCD). The application of chi-2 feature ranking technique and random forest classifier for fault classification of bearing faults was presented in (Vinay et al., 2015). Vakharia and colleagues (Vakharia, et al., 2016) extracted various features from the time domain, frequency domain, and discrete wavelet transform (DWT). Then, Chi-2 and Relief-F methods are used to select the most informative feature to reduce the size of the feature vector for bearing fault diagnosis. Saucedo Dorantes and colleagues (Saucedo Dorantes et al., 2016) proposed a fault diagnosis method based on hybrid feature reduction using EMD, time-domain statistical features, genetic algorithm (GA), FS, and LDA is proposed to learn a reduced set of features. With these reduced features NN is used to classify multiple faults of the induction motor.

Recently, Haroun et al., utilised multiple features extraction techniques from time, frequency, and time-frequency domain to extract features from the bearing vibration signal. Then, Relief-F and

mRmR feature selection techniques are used to select the optimal features. Finally, the self-organising map (SOM) is used for the classification of bearing faults (Haroun, et al., 2017). Li and colleagues (Li, et al., 2017) proposed a strategy for rolling bearing fault diagnosis based on multiscale permutation entropy (MPE), LS, and least squares support vector machine-Quantum behaved particle swarm optimization (QPSO-LSSVM) classifier.

2.6.2 Wrapper model-based feature selection

The wrapper-based feature selection is based on the predictive performance of a predefined predictor. It is usually expensive compared to filter-based feature selection. Wrapper methods can be categorized into sequential selection algorithms and heuristic search algorithms (Chandrashekar et al., 2014).

2.6.2.1 Sequential selection algorithms

There are two types of sequential selection algorithms. The first one is the sequential forward selection (SFS), which start with an empty set and sequentially adding features, evaluate the selected features, and repeat until there is no improvement in the prediction. The second one is the sequential backward selection (SBS), which start with the complete set of features and sequentially removing features, evaluate the selected features, and repeat until there is no improvement in the prediction (Devijver and Kittler, 1982). Moreover, Pudil and colleagues (Pudil et al., 1994) suggested floating search methods in feature selection where a dynamically changing number of features added or removed at each step of sequential search methods. These are the sequential forward floating selection (SFF) and the sequential backward floating selection (SBFS).

Many researchers have utilised sequential selection algorithms to select features from the high dimensional feature space of vibration data for roller bearing fault diagnosis. For instance, Zhang and colleagues (Zhang et al., 2011) proposed a hybrid model for machinery fault diagnosis. The proposed model combines multiple feature selection models including eight filter models and two wrapper models to select the most significant features from all potentially relevant features. These are data variance, PCC, Relief, FS, class separability, Chi-2, IG, GR, a Binary Search (BS) model, and an SBS model respectively. Rauber and colleagues (Rauber et al., 2015) proposed different feature models that utilised in a single pool. These models rely on statistical time parameters, complex envelope spectrum, and wavelet packet analysis. Furthermore, the most significant features are selected using PCA, SFS, SBS, SFFS, and SBFS. Islam and colleagues (Islam et al., 2016) proposed a hybrid fault diagnosis model for bearing, which extracts features from the acoustic emission signal. They investigated the feature selection approaches including SFS, SFFS, and genetic algorithm (GA) for selecting optimal features that can be used for fault diagnosis.

2.6.2.2 Heuristic -based selection algorithms

In the wrapper approach, the size of the search space for n features is ($O(2^n)$ that makes it impractical for high dimensional feature space. The heuristic and meta-heuristic search-based feature selection algorithms was proposed to improve the searching performance (Kohavi and John, 1997; Bozorg-Haddad, et al., 2017). Of these algorithms, meta-heuristic-based algorithms are amongst the most common algorithms used for feature selection in bearing fault diagnosis. These include, Ant Colony Optimization (ACO), Genetic Algorithm (GA), and Particle Swarm Optimization (PSO). For example, ACO is used by many researchers to select the parameters of SVM to improve the classification performance of roller bearing faults (Li et al., 2013 and Zhang et al., 2015). Yuan and Chu proposed a method that jointly optimises the feature selection and the SVM parameters with a modified discrete PSO for fault diagnosis (Yuan and Chu, 2007). Kanović and colleagues presented a detailed theoretical and empirical analysis of PSO and generalised PSO with application in fault diagnosis (Kanović et al., 2011). Liu and colleagues (Liu, et al., 2013) proposed a multi-fault classification model based on the wavelet kernel method of SVM and PSO is used to seek the optimal parameters of the proposed method. In the same vein, Van and Kang (Van and Kang, 2015) proposed method based on a wavelet kernel function and linear local fisher discriminant analysis (LFDA) and PSO is used to seek the optimal parameters of the proposed method. Zhu and colleagues (Zhu et al., 2014) proposed a fault feature extraction method based on hierarchical entropy and SVM with PSO.

Furthermore, Jack and Nandi (Jack and Nandi, 2000) examined the use of GA to select the most important features from a large number of input features in machine fault classification using Artificial Neural Network (ANN). Similarly, Samanta and colleagues (Samanta, et al., 2006) compared the performance of MLP, RBF, and PNN for bearing fault detection. GA is applied to select the parameters and the input features of the classifiers. Genetic programming (GP) was also demonstrated effective in selecting the best features in machine fault classification (Zhang et al., 2005; Zhang and Nandi, 2007).

2.6.3 Embedded model-based feature selection

The embedded model-based feature selection methods are built in the classification algorithm to accomplish the feature selection. LASSO, Elastic Net, the Classification and Regression Tree (CART), C4.5, and SVM–Recursive Feature Elimination (SVM-RFE) are amongst common used embedded methods. Several researchers utilised embedded model-based feature selection for bearing fault diagnosis. For example, Rajeswari and colleagues (Rajeswari et al., 2015) examined the performance of Multiclass Support Vector Machines (MSVM) for bearing fault classification using different feature selection techniques. In the data pre-processing step, the wavelet transform is used to extract the features from bearing vibration signal. To reduce the dimensionality of the feature SVM-RFE, Wrapper subset method, Relief-F method, and PCA feature selection techniques are used. Peng and Chiang employed a C4.5 decision tree and random forest algorithm for bearing fault diagnosis

(Peng and Chiang, 2011). Seera and colleagues (Seera, et al., 2016) proposed a hybrid online learning model that combines the fuzzy min-max (FMM) neural network and the CART for motor fault diagnosis. Duque-Perez et al. used the LASSO technique to improve the performance of a logistic regression classifier (LRC) to diagnosis bearing health conditions.

2.7 Classification algorithms

2.7.1 Multinomial Logistic Regression Classifier (LRC)

Multinomial logistic regression (Hosmer et al., 2013), also called Softmax regression in ANN, is a linear supervised regression model that generalizes the logistic regression where labels are binary, i.e., $c^{(i)} \in \{0,1\}$ to multi-classification problems that have labels {1… c} where c is the number of classes. Briefly, we present the simplified multinomial logistic regression model as follows:

Let there be a training set $\{(x^{(1)}, c^{(1)}), \dots, (x^{(L)}, c^{(L)})\}$ of L labeled examples and input features $x^{(i)} \in R^k$. In logistic regression with binary labels,$c^{(i)} \in \{0,1\}$ our hypothesis can be written as follows:

$$h_0(x) = \frac{1}{1+exp(-\theta^T x)} \tag{2.84}$$

Here θ are model parameters that are trained to minimize the cost function $J(\theta)$ defined by the following equation

$$J(\theta) = -\left[\sum_{i=1}^{L} c^{(i)} \log h_\theta(x^i) + \left(1 - c^{(i)}\right) \log\left(1 - h_\theta(x^i)\right)\right] \tag{2.85}$$

In multinomial logistic regression with multi-labels $c^{(i)} \in \{1, \dots, c\}$ the aim is to estimate the probability $P(c = c^{(i)}|x)$ for each value of $c^{(i)}$ = 1 to c, such that

$$h_\theta(x) = \begin{bmatrix} P(c=1|x;\theta) \\ P(c=2|x;\theta) \\ \vdots \\ P(c=K|x;\theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{K} \exp(\theta^{(j)T} x)} \begin{bmatrix} \exp(\theta^{(1)T} x) \\ \exp(\theta^{(2)T} x) \\ \vdots \\ \exp(\theta^{(K)T} x) \end{bmatrix} \tag{2.86}$$

where $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(K)} \in R^n$ are the parameters of the multinomial logistic regression model.

LRC has been adopted for rolling bearings fault diagnosis in many studies, e.g., (Widodo et al., 2009; Caesarendra et al., 2010; Pandya et al., 2014).

2.7.2 Artificial Neural Network

Artificial neural network (ANN) is a supervised learning algorithm that has the ability to learn real, discrete, and vector-valued target function (Nandi et al, 2013). It has been used successfully in bearing fault diagnosis, e.g., (McCormick and Nandi, 1997; Jack et al., 1999; Jack and Nandi, 2002;

Guo et al., 2008; Chang and Jiao, 2012). There are different types of ANN, e.g., Radial Basis Function (RBF), Probabilistic Neural Network (PNN), Multi-Layer Perceptron (MLP), etc. The MLP ANN (Figure 2.9) is one of the most commonly used methods. As shown in Figure 2.9, it involves an input layer, one to several hidden layers, and an output layer. Each layer consists of a number of neurons. The neuron receive inputs, multiply it by the weights of each input and combine the results of the multiplication. Then, the combined multiplications of the signals and weights are then passed to a transfer function to generate the output of the neuron (Figure 2.10).

In this study, we have tested that pattern recognition networks that are feedforward networks with one hidden layer and 10 neurons that trained using Scaled Conjugate Gradient (SCG) backpropagation function (Møller, 1993) are sufficient to deliver high classification accuracy.

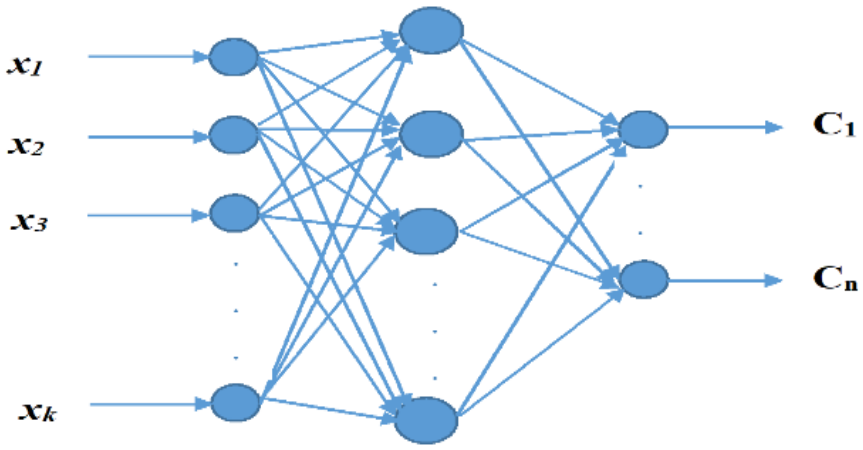


Figure 2.9. A Multilayer Perceptron Model for ANN.

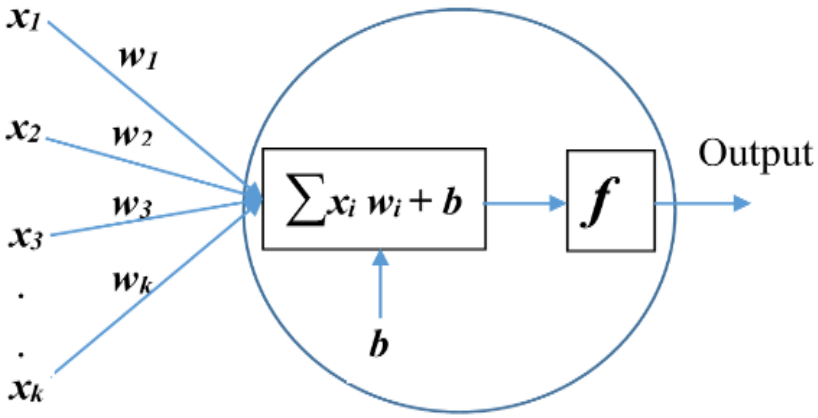


Figure 2.10 Model of an artificial neuron.

2.7.3 Deep Neural Network (DNN)

Although, supervised learning based artificial neural network (ANN) with many hidden layers are found to be difficult in practice, DNNs have been well developed as a research topic and have been made practically feasible with the assistance of unsupervised learning. Moreover, DNNs have attracted

extensive attention by outperforming other machine learning methods. Each layer of DNN performs a non-linear transformation of the input samples in the preceding layer to the following one. A good overview of DNNs can be found in (Schmidhuber, 2015). Different from ANN, DNNs can be trained in a supervised or unsupervised manner (Bengio, et al., 2007; Erhan, et al., 2010) and they are also appropriate in the general area of Reinforcement Learning (RL) (Lange and Riedmiller, 2010; Salimans and Kingma, 2016). The basic idea of training DNN is that we first train the network layer by layer using an unsupervised learning algorithm, e.g. autoencoder; this process is called DNN pre-training. In this process, the output from each layer will be the input to the succeeding layer. Then the DNN is retrained in a supervised way with backpropagation algorithm for classification.

2.7.3.1 Sparse Autoencoder

An autoencoder neural network provides a means of an unsupervised learning algorithm that sets the target values, i.e., the outputs to be equal to the inputs and applies backpropagation (Ng, 2011). As shown in Figure 2.11, like many unsupervised feature learning methods the design of an autoencoder relies on an encoder-decoder architecture, where the encoder produces a feature vector from the input samples and the decoder recovers the input from this feature vector. The encoder part is a feature extraction function f_θ that computes a feature vector h (xi) from an input x_i, we define

$$h(x_i) = f_\theta(x_i) \tag{2.87}$$

where $h(x_i)$ is the feature representation. The decoder part is a recovery function gθ that reconstructs the input space $\widetilde{x_i}$ from the feature space $h(x_i)$ such that

$$\widetilde{x_i} = g_\theta\,(h(x_i)) \tag{2.88}$$

The autoencoder is attempting to learn an approximation such that x_i is similar to $\widetilde{x_i}$, i.e., is trying to attain the lowest possible reconstruction error E (x_i, $\widetilde{x_i}$) that measure the discrepancy between x_i and $\widetilde{x_i}$. Hence the following equation is obtained

$$E(x_i, \widetilde{x_i}) \;=\; \|x_i - \widetilde{x_i}\|^2 \tag{2.89}$$

In fact, autoencoders were mainly developed as a multi-layer perceptron (MLP) and the most commonly used forms for the encoder and decoder are affine transformations that keep collinearity followed by a nonlinearity:

$$f_\theta(x_i) = s_f\,(b + x_i W) \tag{2.90}$$

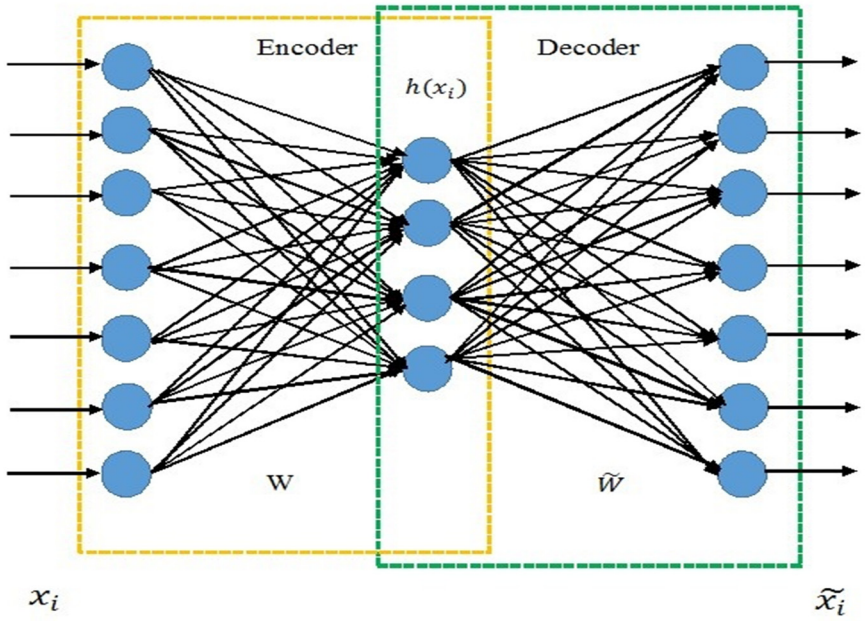


Figure 2.11. Autoencoder architecture.

$$g_\theta\big(h(x_i)\big) = s_g\,(c+x_i\widetilde{W}) \tag{2.91}$$

where, s_f and s_g are the encoder and decoder activation function, e.g. sigmoid and hyperbolic tangent, b and c are the encoder and decoder bias vectors, W and $\widetilde{W}$ are the encoder and decoder weight matrices.

The autoencoder is one of the more practical ways of reinforcement learning. For instance, by forcing some constraints on the autoencoder network such as limiting the number of hidden units and imposing some regularizers, Autoencoder may learn interesting feature structure about the data. Therefore, different constraints give different forms of autoencoders. Sparse Autoencoder (SAE) is an autoencoder that contains sparsity constraint on the hidden unit's activation that must typically be near 0. This may be accomplished by adding Kullback-Leibler (KL) divergence penalty term

$$\sum_{j=1}^{d} KL\,(\rho\|\hat{\rho}) \tag{2.92}$$

where,

$$KL(\rho\|\hat{\rho}) = \rho \log\frac{\rho}{\hat{\rho}_j} + (1-\rho)\log\frac{1-\rho}{1-\hat{\rho}_j} \tag{2.93}$$

and ρ is a sparsity parameter, normally its value can be small and close to zero, e.g., $\rho = 0.2$, while $\hat{\rho}$ is the average threshold activation of hidden units and can be calculated by the following equation:

$$\hat{\rho}_j = \frac{1}{n}\sum_{i=1}^{n}\left[a_j^2(x_i)\right] \tag{2.94}$$

where a_j^2 represents the activation of hidden unit j. By minimizing this penalty term $\hat{\rho}$ would be close to ρ, and the overall cost function *(CF)* can be calculated by the following equation

$$CF_{sparse}(W, b) = \frac{1}{2n}\sum_{i=1}^{n}\|\tilde{x}_i - x_i\|^2 + \lambda\|W\|^2 + \beta\sum_{j=1}^{d} KL(\rho\|\hat{\rho}) \quad (2.95)$$

where n is the input size, d is the hidden layer size, λ represents the weight decay parameter, and β is the weight of the sparsity penalty term.

Several studies have been undertaken to use deep neural network using an autoencoder algorithm for the purpose of machinery fault diagnosis. For example, Tao and colleagues (Tao et al., 2015) proposed a deep neural network algorithm framework for bearing fault diagnosis based on stacked Autoencoder and softmax regression. Jia et al. in (Jia, et al., 2016) shown the effectiveness of a proposed DNN-based intelligent method in the classification of different datasets from rolling element bearings and planetary gearboxes with massive samples using Autoencoder as a learning algorithm. In a recent paper by Sun et al., a sparse Autoencoder-based deep neural network approach with the help of the denoising coding and dropout method using one hidden layer was proposed for induction motor fault classification with 600 data samples and 2000 features from each induction motor working condition (Sun, et al., 2016.). The results of these investigations validate the effectiveness of DNN based on autoencoder learning algorithm in machinery fault classification.

Moving on now to consider other types of deep learning architectures, e.g., convolution deep neural networks (CNNs), deep belief networks (DBNs), and recurrent neural networks (RNNs), that have been used for machine fault diagnosis. For example, unlike the standard Neural Network (NN), the architecture of CNN is usually composed of a convolutional layer and a sub-sampling layer also called a pooling layer. CNN learns abstract features from alternating and stacking convolutional layers and pooling operation. The convolution layers convolve multiple local filters with raw input data and generate invariant local features and the pooling layers extract most significant features (Ahmed et al., 2018). Many studies have used CNN for bearing fault diagnosis (Guo et al., 2016; Fuan et al., 2017; Zhang et al., 2018; and Shao et al., 2018).

DBNs are generative neural networks that stack multiple restricted Boltzmann machines (RBM) that can be trained in a greedy layer-wise unsupervised way, then it can be further fine-tuned with respect to labels of training data by adding a softmax layer in the top layer. Many researchers have used DBN for bearing fault diagnosis (Shao et al., 2015; Ma et al., 2016; Tao et al, 2016). RNN builds connections between units from a direct cycle and map form the entire history of previous inputs to target vectors in principal and allows a memory of previous inputs to be kept in the network state. As is the case with DBNs, RNNs can be trained via backpropagation through time for supervised tasks with sequential input data and target outputs. Examples of studies that used RNN for bearing fault diagnosis can be found in (Lu et al., 2017; Jiang et al., 2018). A good overview of these deep learning architectures can be found in (Zhao et al, 2016; Khan et al., 2018).

2.7.4 Support vector machine classifier

Support vector machine (SVM) is another classification algorithm that can be employed to classify machine health condition using the selected feature space. It has been used for bearing fault classification in many different studies (Jack and Nandi, 2001; Rojas and Nandi, 2006; Yang et al., 2007; Zhang et al., 2013; Zhu et al., 2014; Soualhi et al., 2015; Zheng et al., 2017). It is a supervised machine learning method that was first proposed for binary classification problem (Cortes et al., 1995). The basic idea of SVM is that it can find the best hyperplane(s) to separate data from two different classes such that the distance between the two classes, i.e., the margin, is maximized. Based on the features of the data, SVM can make linear or non-linear classifications by using different kernel functions. The linear hyperplane can be designed using the following function.

$$g(x) = w^T x + w_0 \tag{2.96}$$

Here $x = [x_1, x_2, \dots x_k]$ is the input feature vector. Each hyperplane is described by: (1) its direction which is determined by w, and (2) its position that determined by w_0. To calculate the parameters w and w_0 of the hyperplane, the following optimization problem is performed

$$\text{maximize } J(w, w_0) = \frac{2}{\|w\|^2} \tag{2.97}$$

$$\text{subject to } y_i(w^T x_i + w_0) \geq 1, i = 1,2, \dots, k \tag{2.98}$$

Here y_i is the class indicator. Figure 2.12 shows an illustration of a linear SVM for a two-class classification problem.

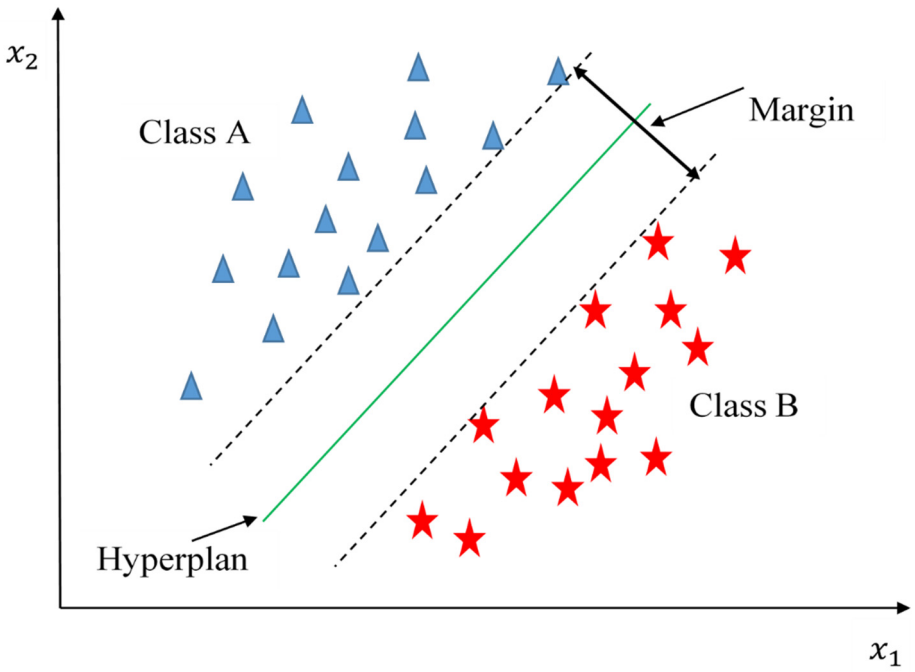


Figure. 2.12. An example of linear classifier for a two class problem.

The non-linear SVM classifier uses a non-linear kernel e.g., Radial basis function (RBF), Polynomial function (PF), and Sigmoid function (SF) (Theodoridis et al., 2010). For multiclass classification problems, several SVM classifiers can together deal with the multiclass problems. There are various methods based on the binary classification used in multiclass problems, e.g., one-against-all, one-against-one, error-correcting output codes (ECOC), etc. (Hsu et al., 2002; Abe, 2005). In this research, we have tested that SVMs with ECOC is sufficient to deliver high classification accuracy. The implementation of SVM with ECOC is from "fitcecoc" function in Statistics and Machine Learning Toolbox in MATLAB, which fit multiclass models for SVMs. It uses c(c-1)/2 binary SVM models using one-versus-one coding design, where c is the number of unique class labels. This will return a fully trained error-correcting output code (ECOC), multiclass model.

2.8 Discussion and Summary

This chapter has provided a review of vibration-based rolling bearings condition monitoring, which covered several topics of vibration analysis techniques that have been used for rolling bearings fault diagnosis. Numerous techniques have been proposed using features extracted in the time domain, frequency domain, time-frequency domain or a mixture of them. Time domain techniques extract features from the raw vibration signals using some statistical parameters, e.g., peak-to-peak value, root mean square, Crest Factor, skewness, and kurtosis, and other advanced techniques such as TSA, AR, ARIMA, and filter-based methods. The frequency domain analysis techniques have the ability to divulge some information based on the frequency characteristics that are not easily observed in time-domain. In practice, the time-domain signal is transformed into frequency domain by using FFT. The time-frequency domain has been used for non-stationary waveform signals which are very common when machinery fault occurs. Thus far, several time-frequency analysis techniques have been developed and applied to machinery fault diagnosis, e.g., WT, STFT, HHT, LMD, and EMD.

To reduce the high dimension of the extracted features one may use linear subspace learning techniques, e.g., PCA, LDA, and ICA, to extract a new low dimensional feature space that is usually a linear combination of the original high dimensional feature space. Furthermore, feature selection techniques can be used to select a subset of features that can sufficiently represent the characteristic of the original features. In addition, there are several types of fault classification methods, e.g., SVM, ANN, and LRC, to process the extracted features. The following conclusions were drawn from the literature review:

- The literature on vibration-based bearings CM identifies numerous computational methods for fault diagnosis that achieved many interesting results. However, the performance of these methods is limited by the large amounts of Nyquist rate-based sampled vibration data need to be

acquired from rotating machines to achieve the anticipated accuracy of bearing fault detection and classification.

- In order to learn more useful features for rolling bearings fault diagnosis, most of the proposed methods combine two or more of the various analysis techniques. For example, linear subspace learning techniques such as PCA and LDA have been used in different combinations with other methods of feature extraction or feature selection in the time domain, frequency domain or time-frequency domain. Hence, the challenge will always remain of producing possible approaches to machine condition monitoring capable of improving fault diagnosis accuracy and reduce computations.
- Several studies have been undertaken to use deep neural network using an autoencoder algorithm for the purpose of machinery fault diagnosis. However, their focus was mainly on using autoencoder as a dimensionality reduction technique, i.e., the number of hidden nodes in each hidden layer is less than the number of input samples for the purpose of fault diagnosis with a large amount of input data.

Chapter 3

Datasets, Acquisition and Compression

This chapter provides a summary of the datasets utilised in the experiments in this thesis. Two sets of bearing vibration data are used to evaluate the proposed methods in this thesis. In section 3.1, the description of the data acquisition of the first bearings vibration data is presented, together with an overview of different faults features. Section 3.2 describes the second bearings vibration data provided by Case Western Reserve University (CWRU). Section 3.3 outlines the compression of the acquired vibration data using the CS framework.

3.1 The first bearing vibration data

The first bearing vibration dataset is acquired from experiments on a test rig that simulates running roller bearings' environment. In these experiments, several interchangeable faulty roller bearings are inserted in the test rig to symbolise the type of faults that can normally happen in roller bearings. The test rig (Figure 3.1) used to acquire the first vibration dataset of bearings consists of a 12V DC electric motor driving the shaft through a flexible coupling. The shaft supported by two Plummer bearing blocks where a series of damaged bearing were inserted. Two accelerometers were used to measure the resultant vibrations in the horizontal and vertical planes. The output from the accelerometers was fed back via a charge amplifier to a Loughborough Sound Images DSP32 ADC card utilising a low-pass filter by means of a cut-off of 18 kHz. The sampling rate was 48 kHz, giving slight oversampling. Six health conditions of roller bearings have been recorded with two normal conditions, i.e., brand new condition (NO) and worn but undamaged condition (NW), and four faulty condition including, inner race fault (IR), an outer race fault (OR), rolling element fault (RE), and cage fault (CA). Table 3.1 presents an explanation of the corresponding characteristics of these bearing health conditions.

The data recorded using 16 different speeds in the range 25 – 75 rev/s. In each speed, ten-time series were recorded for each condition, i.e., 160 examples per condition. This resulted in a total of 960 examples with 6000 data points to work with. Figure 3.2 illustrates some typical time series plots for the six different aforementioned conditions.

As shown in Figure 3.2, each fault modulates the vibration signals with their own unique patterns. For instance, based on the level of damage to the rolling element and the loading of the bearing,

IR and OR fault conditions have a fairly periodic signal, RE fault condition may or may not be periodic, and CA fault condition generates a random distortion.

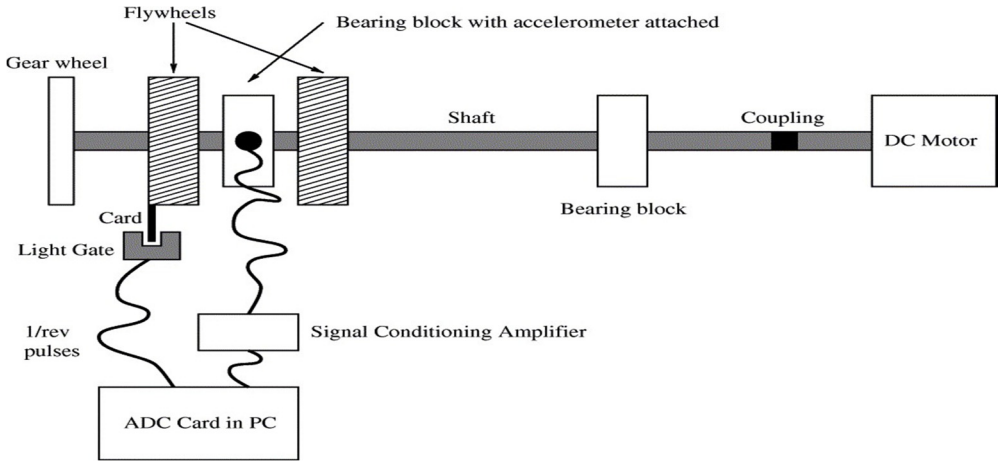


Figure 3.1. The test rig used to collect the first vibration data of bearings.

Table 3.1. The characteristics of bearings health conditions in the first bearing dataset.

Condition	Characteristic
NO	The bearing is a brand new and in perfect condition.
NW	The bearing is in service for some period of time but in good condition.
IR	Inner race fault. This fault is created by cutting a small groove in the raceway of the inner race.
OR	Outer race fault. This fault is created by cutting a small groove in the raceway of the outer race.
RE	Roller element fault. This fault created by using electrical etcher to mark the surface of the balls, simulating corrosion.
CA	Cage fault. This fault is created by removing the plastic cage from one of the bearings, cutting away a section of the cage so that two of the balls were not held at a regular space and free to move.

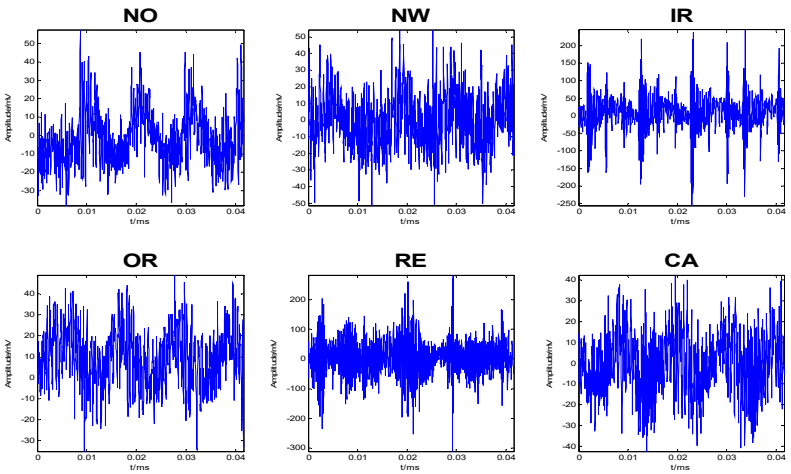


Figure 3.2. Typical time – domain vibration signals for the six different conditions

3.2 The second bearing vibration data

The second bearing vibration data is provided by the Case Western Reserve University (CWRU) Bearing Data Center (http://csegroups.case.edu/bearingdatacenter/home). This data is freely available and commonly used in roller bearings fault diagnosis field. Figure 3.3 shows the test rig that is used to acquire this vibration data. It is comprised of a 2 horsepower electric motor driving a shaft that contains a torque transducer and encoder. A dynamometer and electronic control system are used to apply torque to the shaft. A series of faults with width ranging from 0.18 to 0.71 mm (0.007 to 0.028 in) were seeded on the drive end bearing (in this case SKF deep-groove ball bearings 6205-2RS JEM were used) of the electric motor utilising electro-discharger machining.

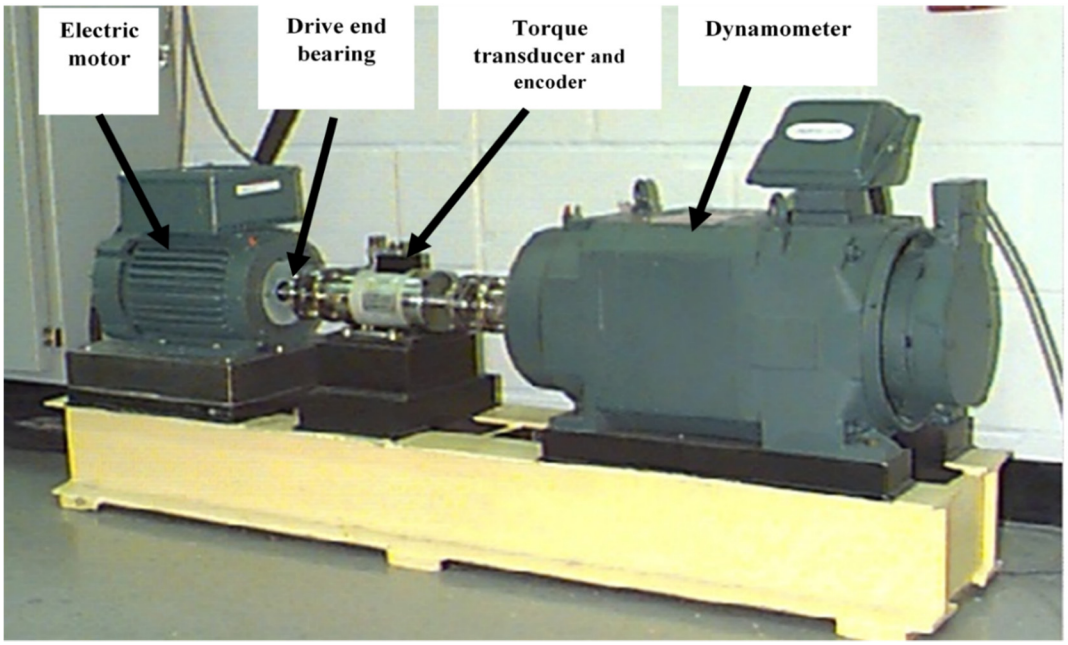


Figure 3.3. The test rig used to collect the first vibration data of bearings (CWRU Bearing Data Center).

The seeded faults include rolling elements, inner race, and outer race faults, and each faulty bearing was run for motor loads 0 – 3 horsepower at a constant speed in the range 1720 – 1797 rev/m. The sampling rates used were 12 kHz for some of the sampled data and 48 kHz for the rest of the sampled data. The bearing vibration signals were acquired under normal NO, IR, OR, and RE conditions for four different speeds. In each speed, 100 time-series were taken for each condition per load. For IR, OR, and RE condition, vibration signals for four different fault widths (0.18, 0.36, 0.53, and 0.71 mm) were separately recorded.

In this thesis, of these acquired vibration signals, three groups of datasets were organised to be used in the evaluation of the methods proposed. The first group of datasets is chosen from the data files of the vibration signals that sampled at 48 kHz with fault width (0.18, 0.36, and 0.53), fixed load including 1, 2, and 3 horsepower, and the number of examples chosen is 200 examples per condition. This gave three different datasets A, B, and C with 2000 total number of examples and 2400 data points

for each signal. The description of this three dataset is shown in Table 3.2. The second group of datasets is chosen from the vibration signals that sampled at 12 kHz with fault width (0.18, 0.36, and 0.53), fixed load including 0 and 3 horsepower, variable loads motor operating conditions including 0, 1, 2, and 3, and the number of examples chosen is 400 per condition. This gave three different datasets D, E, and F with 4000 total number of examples and 1200 data points for each signal. Table 3.3 shows a description of these datasets. A summary of the bearing health conditions included in A, B, and C is shown in Table 3.4, and for D, E, and F is presented in Table 3.5.

Table 3.2. Description of the first group of bearing datasets.

Datasets	Load	Number of samples in each condition
A	1	200
B	2	200
C	3	200

Table 3.3. Description of the second group of bearing datasets.

Datasets	Load	Number of samples in each condition
D	0	400
E	3	400
F	0, 1, 2, and 3	400

Table 3.4. Description of the bearing health conditions for A, B, and C datasets.

Fault type	Fault width in (mm)	Classification label
NO	0	1
IR1	0.18	2
IR2	0.36	3
IR3	0.53	4
RE1	0.18	5
RE2	0.36	6
RE3	0.53	7
OR1	0.18	8
OR2	0.36	9
OR3	0.53	10

Table 3.5. Description of the bearing health conditions for D, E and F datasets.

Fault type	Fault width in (mm)	Classification label
NO	0	1
OR1	0.18	2
OR2	0.36	3
OR3	0.53	4
IR1	0.18	5
IR2	0.36	6
IR3	0.53	7
RE1	0.18	8
RE2	0.36	9
RE3	0.53	10

The third type of bearing dataset is chosen from the data files of vibration signals that sampled at 12 kHz with fault size (0.18, 0.36, 0.53, and 0.71), load 2 horsepower, and the number of examples chosen is 60 examples per condition. This gave a dataset G with 720 total number of examples with 2000 data points for each signal. The description of this dataset is presented in Table 3.6. Some typical time series plots for the different health conditions of datasets A, B, and C are shown in Figure 3.4 – 3.6.

Table 3.6. Description of the bearing health conditions for dataset G.

Fault type	Fault width in (mm)	Classification label
NO	0	1
IR1	0.18	2
IR2	0.36	3
IR3	0.53	4
IR4	0.71	5
RE1	0.18	6
RE2	0.36	7
RE3	0.53	8
RE4	0.71	9
OR1	0.18	10
OR2	0.36	11
OR3	0.53	12

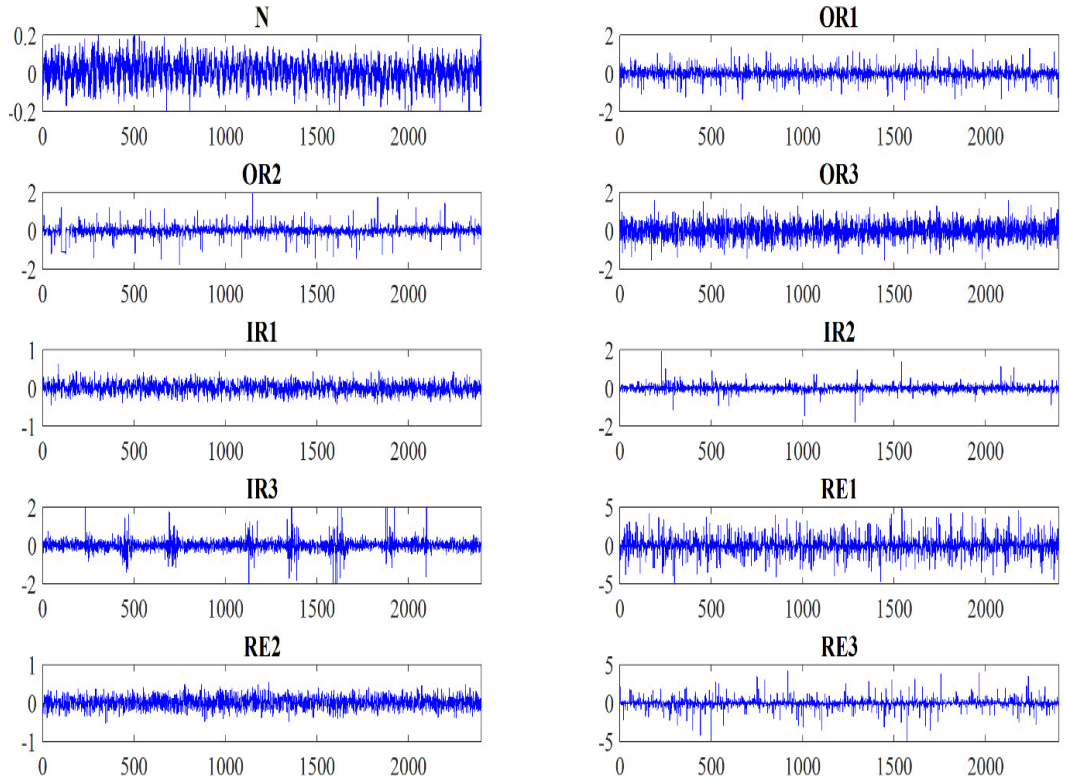


Fig. 3.4. Typical time – domain vibration signals for the ten different health conditions of dataset A

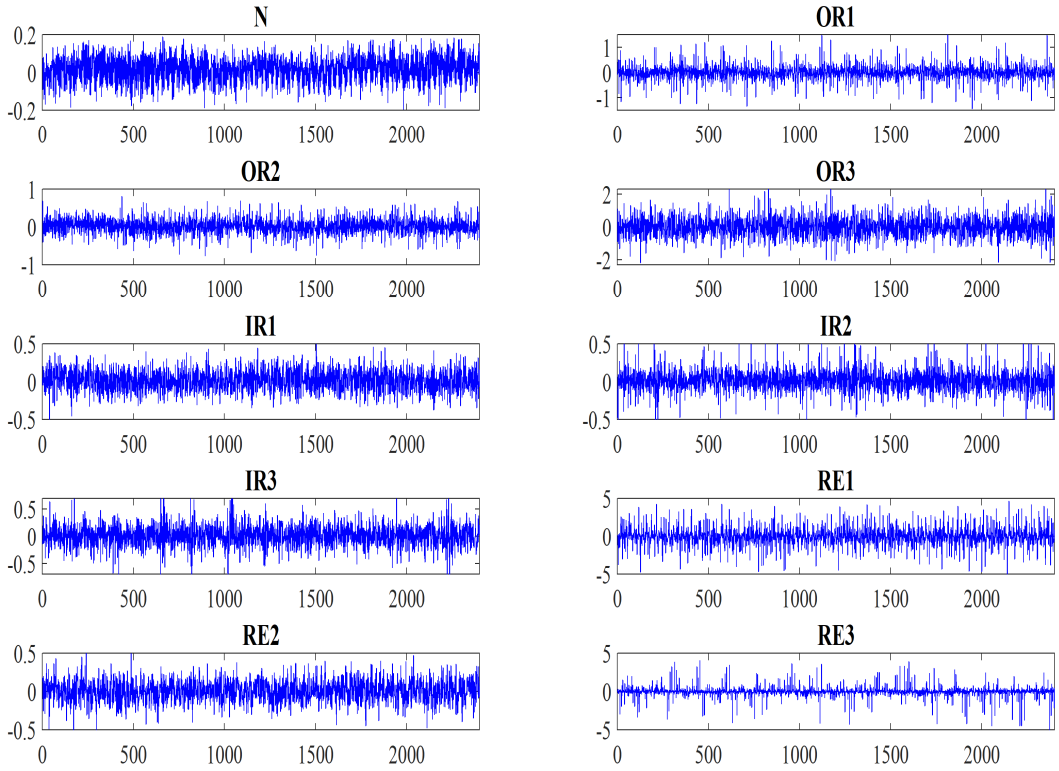


Fig. 3.5. Typical time – domain vibration signals for the ten different health conditions of dataset B

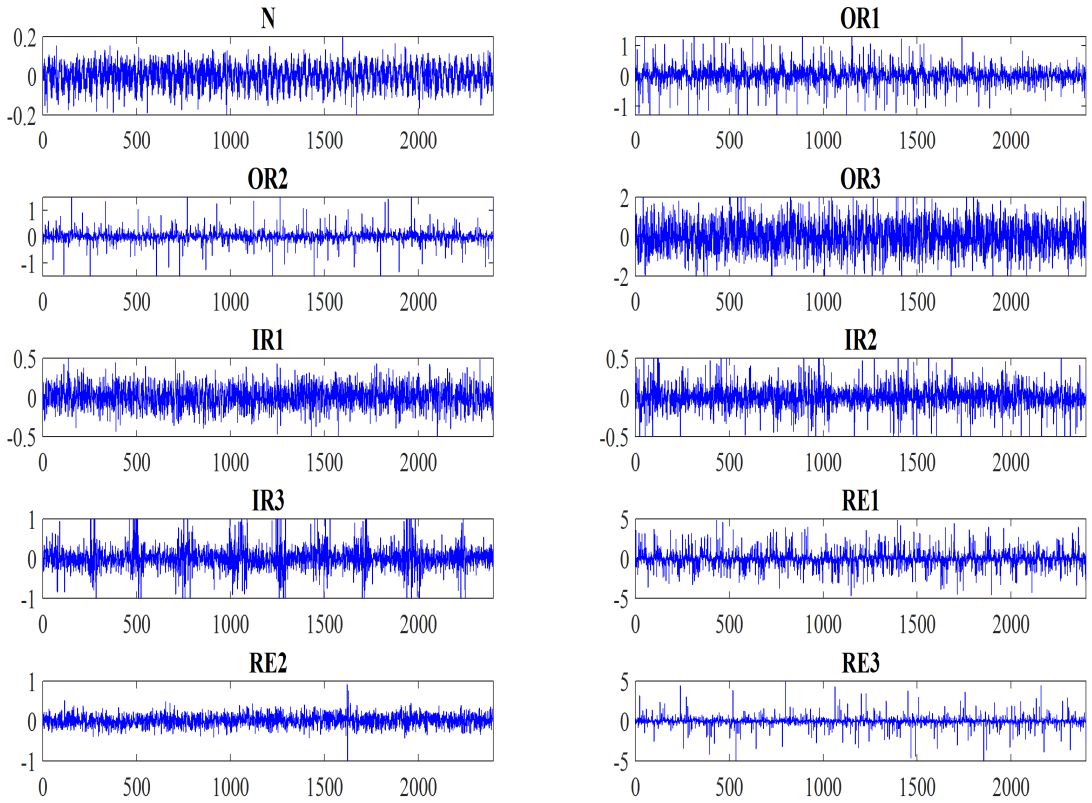


Fig. 3.6. Typical time – domain vibration signals for the ten different health conditions of dataset C

3.3 Vibration data compression

Processing a large amount of data requires a large training time and large storage and this also may limit the number of machines that can be monitored remotely across wireless sensor network (WSNs) due to bandwidth and power constraints. Instead of processing a large amount of the collected vibration data, we often depend on reducing the amount of the collected data without weakening the characteristics existing in the data. Transfer-coding that depends on finding a frame that delivers sparse or compressible representations with k nonzero coefficients of a signal of length n where $k << n$, is one of the most commonly used techniques. CS framework, which based on linear dimensionality reduction, allows a large reduction in sampling for signals that have sparse or compressible representation. Based on the CS framework, the data compression can be achieved using a sparse representation of the vibration data, measurement matrix, and compressive sampling rate. This will be described in more detail in chapter 4. Having the sampled vibration datasets as described in 3.1 and 3.2, CS framework with different values of compression sampling rate was used to compress these datasets and the resultant compressed data matrices of the first bearing vibration data and the second bearing vibration data are listed in Table 3.7 and Table 3.8 respectively.

Table 3.7. Description of the compressed data matrices of the first bearing vibration data using different values of α.

Data	Original data matrix size	Compressive sampling rate (α)	Compressed data matrix size
First bearing vibration data	6000 x 960	0.002	12 x 960
		0.003	18 x 960
		0.006	36 x 960
		0.01	60 x 960
		0.02	120 x 960
		0.03	180 x 960
		0.04	240 x 960
		0.05	300 x 960
		0.1	600 x 960
		0.2	1200 x 960
		0.3	1800 x 960
		0.4	2400 x 960

Table 3.8. Description of the compressed data matrices of the second bearing vibration datasets using different values of α.

Data	Original data matrix size	Compressive sampling rate (α)	Compressed data matrix size
A, B, and C	2400 x 2000	0.01	24 x 2000
		0.02	48 x 2000
		0.025	60 x 2000
		0.03	72 x 2000
		0.04	96 x 2000
		0.05	120 x 2000
		0.1	240 x 2000
		0.2	480 x 2000
		0.3	720 x 2000
		0.4	960 x 2000
D, E, and C	1200 x 4000	0.03	36 x 4000
		0.04	48 x 4000
		0.05	60 x 4000
		0.06	72 x 4000
		0.07	84 x 4000
		0.08	96 x 4000
		0.09	108 x 4000
		0.1	120 x 4000

G	2000 x 720	0.02	40 x 720
		0.04	80 x 720
		0.06	120 x 720
		0.08	160 x 720
		0.1	200 x 720
		0.2	400 x 720
		0.3	600 x 720
		0.4	800 x 720

3.4 Summary

This chapter has discussed the sources, data acquisition, and compression of the different bearing vibration datasets used in the experiments conducted in this thesis. In addition, a brief description of some of the characteristics of the different dataset and their corresponding compressed data matrices sizes were also given.

Chapter 4

Compressive Sampling Methodology

This chapter introduces the methods and techniques that have been used to produce the results presented and discussed in the following chapters. The sections 4.2, 4.3, and 4.4 present novel methods, namely the compressive sampling and feature learning (CSFR) framework based methods, the compressive sampling and subspace learning (CSLSL) based methods, and the compressive sampling and sparse autoencoder based deep neural network (CS-SAE-DNN) method. These methods represent the original contribution of this thesis in designing new methods for vibration-based rolling bearing condition monitoring. The majority of these methods have been published in (Ahmed et al., 2016; Ahmed et al., 2017; Ahmed and Nandi, 2017; Ahmed and Nandi, 2017b; Ahmed et al., 2018; Ahmed and Nandi, 2018a; Ahmed and Nandi, 2018b; Ahmed and Nandi, 2018c).

4.1 Compressive sampling (CS)

In place of processing the high dimensional of the collected vibration data, the common methodology is to identify a lower – dimensional features space that can represent the acquired large amount of vibration signals while retaining the important information of the machine conditions. Quite recently, considerable attention has been paid to CS for its ability to allow one to sample far below the Nyquist sampling rate and yet be able to reconstruct the original signal when needed. CS (Donoho, 2006; Candès and Wakin, 2008) is an extension of sparse representations and special case of it. The simple idea of CS is that many real-world signals have sparse or compressible representations in some domain, e.g., Fourier Transform (FT), can be recovered from fewer measurements under certain conditions. In fact, CS is based on two principles: (1) sparsity of the signal of interest, and (2) the measurements matrix that satisfies the data minimal information loss, i.e., fulfill Restricted Isometry Property (RIP) (Candes and Tao, 2006). Concisely, we describe the sparsity as follows:

Assume that x$\in R^{n\,x\,1}$ be an original time indexed signal. Given a sparsifying transform matrix $\psi \in R^{n\,x\,n}$ whose columns are the basis elements$\{\psi_i\}_{i=1}^{n}$. Based on this basis, x can be represented as follows:

$$x = \sum_{i=1}^{n} \psi_i \, s_i \tag{4.1}$$

Or more efficiently,

$$x = \psi s \tag{4.2}$$

Here s is an n*1 column vector of coefficients. If the basis ψ produces q-sparse representations, of x, i.e., x of length n can be represented with $q << n$ nonzero coefficients, then equation (1) can be rewritten as follows:

$$x = \sum_{i=1}^{q} \psi_{ni}\, s_{ni} \tag{4.3}$$

here ni is the index of the basis elements and the coefficients corresponding to the q non-zero elements. So, $s \in R^{n\,x\,1}$ is a vector column with only q non-zero elements and represents the sparse representation vector of x.

Based on the CS framework, m << n projections of the vector x with a group of measurement vectors $\{\emptyset_j\}_{j=1}^{m}$ and the sparse representations s of x can be produced from,

$$y = \emptyset\psi s = \theta s \tag{4.4}$$

where y is an m*1 column vector of the compressed measurements and $\theta = \emptyset\psi$ is the measurement matrix. To produce good compressed measurements that possess the quality of the original signal, i.e., the original signal can be reconstructed using these compressed measurements, the measurement matrix θ has to satisfy the data minimal information loss, i.e., satisfy the Restricted Isometry Property (RIP).

Definition 1.1: The measurement matrix θ satisfies the Restricted Isometry Property (RIP) if there exists a parameter $\delta \in (0,1)$ such that

$$(1 - \delta)\|s\|_{l_2}^2 \leq \|\theta s\|_{l_2}^2 \leq (1 + \delta)\|s\|_{l_2}^2 \tag{4.5}$$

Founded along the idea of compressive sensing, when Φ and Ψ are incoherent, the original signal can be recovered from m = O (q log (n)) Gaussian measurements or $q \leq C.m/\log(n/m)$ Bernoulli measurements (Baraniuk et al., 2008), here, C is constant and q is the sparsity level. Random matrix with i.i.d. Gaussian entries or Bernoulli (±1) matrix are both satisfy Restricted Isometry Property (RIP). The size of the measurement matrix (m*n) depends on the compressive sampling rate (α) (i.e., m = α*n). Fig. 1, shows an illustration of the CS framework. To estimate the vector s, the following optimization problem is solved:

$$\min_{s \in R^N} \frac{1}{2}\|\Theta\hat{s} - y\|_{l2}^2 + \gamma\|\hat{s}\|_{l1} \tag{4.6}$$

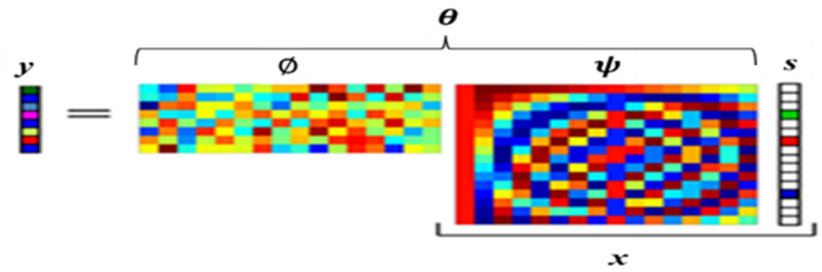


Figure 4.1. Compressive sampling framework.

Here $\|\Theta s - y\|_{\ell 2}^2 \leq \varepsilon$ for a chosen $\varepsilon > 0$, and a particular regularization parameter $\gamma > 0$ that controls the relative importance applied to the sparseness $\ell 1$ and the $\ell 2$ norms. Therefore, the original vector x can be reconstructed using the inverse of the sparsifying transform ψ and $\hat{s}$ such that

$$\hat{x} = \psi^{-1}\hat{s} \tag{4.7}$$

The model described above is meant to be Single Measurement Vector Compressive sampling (SMV-CS) that recovers one vector from its corresponding compressed measurement vector. But, Multiple Measurement Vectors Compressive Sampling (MMV-CS) is considered for signals that are represented as a matrix with a set of jointly sparse vectors such that

$$Y = \Theta S \tag{4.8}$$

where $Y \in R^{m\,x\,L}$, m is the number of compressed measurements and L is the number of observations, $\Theta \in R^{m\,x\,n}$ is a dictionary, and $S \in R^{n\,x\,L}$ is a sparse representation matrix. Several studies have been conducted to reconstruct jointly sparse signals (S) given multiple compressed measurement vector (Chen, et al., 2006; Sun, et al., 2009). Then, the original signal matrix X can be recovered using the inverse of the sparsifying transform and $\hat{S}$ such that

$$\hat{X} = \psi^{-1}\hat{S} \tag{4.9}$$

Here $\hat{X}$ and $\hat{S}$ is the estimation of X and S respectively. The better signal reconstruction indicates that the compressed samples possess the quality of the original signal. In this thesis, MMV-CS has been used to obtain compressively sampled signals since the dataset consists of a matrix of multiple measurements. Also, since it is possible to recover the original signal (X) from the compressed data (Y) this indicates that Y possesses the quality of the original signal (X). Hence, we use the compressed measurements directly without recovering the original data.

4.1.1 CS for compressible frequency representation

The CS framework requires that the signal of interest have a sparse or compressible representation in a known transform domain. A commonly utilised sparse basis for vibration signal is Fast Fourier Transform (FFT) matrix (Rudelson and Vershynin, 2008; Zhang et al., 2015; Wong et al., 2015; Yuan et al., 2017). We assume the time domain vibration signal of the rolling bearing is compressible in the frequency domain and that the FFT-based frequency representation of vibration signal preserves a compressible structure. Given an acquired vibration dataset $X = \{x_1, x_2, \ldots, x_L\} \in R^n$ the process to obtain compressively sampled dataset using MMV-CS framework in (4.8). First the compressible representations ($S \in R^{nxL}$) that consists of only a small number of $q \ll n$ of nonzero coefficients are obtained from the raw vibration signals ($X \in R^{nxL}$) using the FFT algorithm, which computes n-point complex discrete Fourier transform (DFT) of signal X. In our case we used the magnitude of the DFT, i.e., the absolute value of DFT of signal X, to obtain S. Then, the obtained (S) is projected into a suitable measurement matrix ($\Theta \in R^{mxn}$) that satisfy the RIP. In this thesis, a random

matrix with i.i.d. Gaussian entries matrix and a compressed sampling rate (α) is used to generate the compressively-sampled signals ($Y \in R^{mxL}$) where m is the number of compressed signal elements (i.e., m = α*n). These procedures are summarized below in Algorithm 1.

Algorithm 1: Compressive sampling with FFT

Input: $X \in R^{n\,x\,L}$; $\Theta \in R^{mxn}$; and α

Output: $Y \in R^{m\,x\,L}$

1: abs (FFT(X)) ⟶ $S \in R^{n\,x\,L}$

2: Project S into Θ with compressed sampling rate α to obtain compressively-sampled signal $Y \in R^{m\,x\,L}$

Figure 4.2 shows an example of an obtained compressively-sampled signal of bearing outer race fault signal x_{OR} using Algorithm 1with α = 0.1.

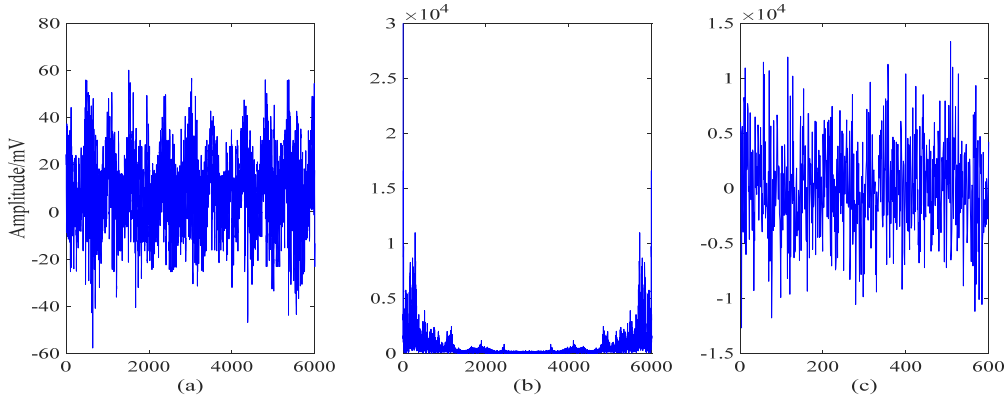


Figure. 4.2 An example of (a) an outer race fault time domain signal x_{OR}, (b) the corresponding absolute values of Fourier coefficients for x_{OR}, and (c) the obtained compressively-sampled signal of x_{OR}.

In this study, the implementation of FFT is from the *fft* function in Fourier analysis and Filtering Toolbox in MATLAB.

4.1.2 CS for sparse time-frequency representation

The Wavelet transform that decomposes the signal into low and high-frequency levels is used to obtain the sparse components of the vibration signal that are demanded by the compressive sensing framework. One of the choices is the Haar wavelet basis that has been used as sparse representations for vibration signals in (Bao, et al., 2011). Given an acquired vibration dataset $X = \{x_1, x_2, \dots, x_L\} \in R^n$ the process to obtain compressively sampled dataset using MMV-CS framework in (4.8). First the compressible representations ($S \in R^{nxL}$) that consists of only a small number of $q \ll n$ of nonzero coefficients are obtained from raw vibration signals ($X \in R^{nxL}$) using thresholded Haar wavelet basis with five decomposition levels as a sparsifying transform. For example, the wavelet coefficients of the bearing outer race vibration signal x_{OR} is displayed in Figure 4.3 (b). After applying the hard threshold, which preserves the input if it is bigger than the threshold τ; else, it is set to zero (Chang et al., 2000),

the wavelet coefficients is sparse in the Haar wavelet domain as shown in Figure 4.3 (c). Then, the obtained (S) is projected into a random matrix with i.i.d. Gaussian entries matrix, with the compressed sampling rate (α), is used to generate the compressively-sampled signals ($Y \in R^{mxL}$) where m is the number of compressed signal elements. These procedures are summarized below in Algorithm 2.

Algorithm 2: Compressive sampling with thresholded Haar WT

Input: $X \in R^{n\,x\,L}$; $\Theta \in R^{mxn}$; and α

Output: $Y \in R^{m\,x\,L}$

1: thresholded Haar WT(X)) ⟶ $S \in R^{n\,x\,L}$

2: Project S into Θ with compressed sampling rate α to obtain compressively-sampled signal $Y \in R^{m\,x\,L}$

Figure 4.3 (d) shows an example of an obtained compressively-sampled signal of bearing outer race fault signal x_{OR} using Algorithm 2 with α = 0.1.

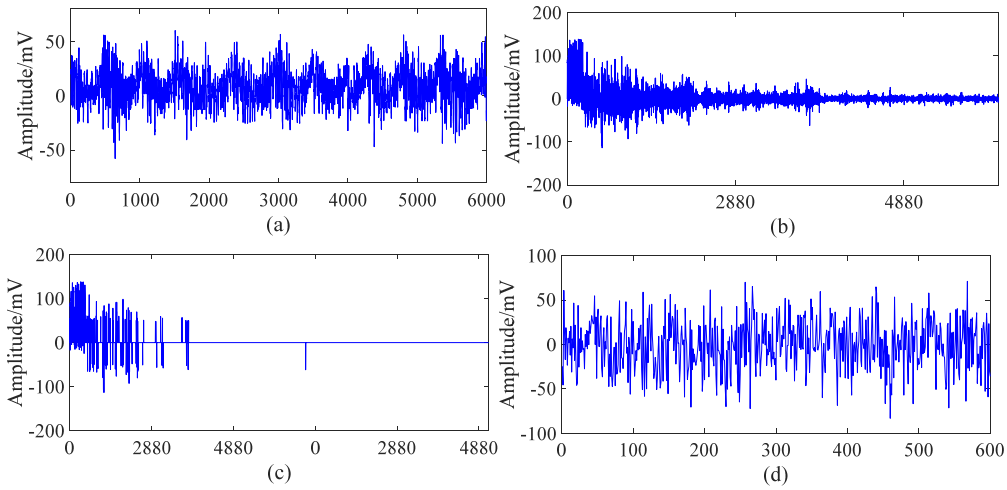


Figure. 4.3 An example of (a) an outer race fault time domain signal x_{OR}, (b) the corresponding Haar WT coefficients of x_{OR}, (c) the corresponding thresholded Haar WT coefficients of x_{OR}, and (d) the obtained compressively-sampled signal of x_{OR}.

In this study, the implementation of Haar WT is from the Wavelet Toolbox in MATLAB.

4.2 Compressive sampling and feature ranking (CSFR) framework

Despite the fact that the obtained CS-based compressed measurements are able to recover the original signal, they may not provide the best bearing fault classification. Moreover, these compressed measurements may still represent a large amount of data collected in real operating condition.The compressive sampling and feature ranking framework (CSFR) combines CS based on multiple measurement vectors (MMV) and feature ranking and selection techniques to learn optimally fewer features from a large amount of vibration data. With these learned features, bearing health condition can be classified using a machine learning classifier. CSFR receive a large amount of vibration data as

input and produce fewer features as output, which can be used for fault diagnosis. As shown in Figure 4.4, the proposed framework first compress the vibration data and then rank the features of the compressed data from which the most significant features can be selected to be used for classification. The details are as follows:

(1) Vibration data compression: With the aim to reduce computations, transmission costs, and reduce demands on the environment compared to other techniques. CSFR employs MMV-CS model to produce compressively-sampled signals, i.e., compressed data $Y = \{y_1, y_2, \dots, y_L\} \in R^m$ that have enough information of the original bearing raw data $X = \{x_1, x_2, \dots, x_L\} \in R^n$. Here m << n.

(2) Feature ranking and selection: as long as the compressively-sampled signals produced by CS model have enough information about the original vibration signals, we may further filter the compressively-sampled signals using feature ranking and selection techniques to rank and select fewer features from the compressively-sampled signals that can sufficiently represent characteristics of bearing health conditions.

(3) Fault classification: with these fewer selected features a classifier is used to classify bearing health condition.

Figure 4.5 shows an illustration of the data compression and feature selection process in the proposed framework.

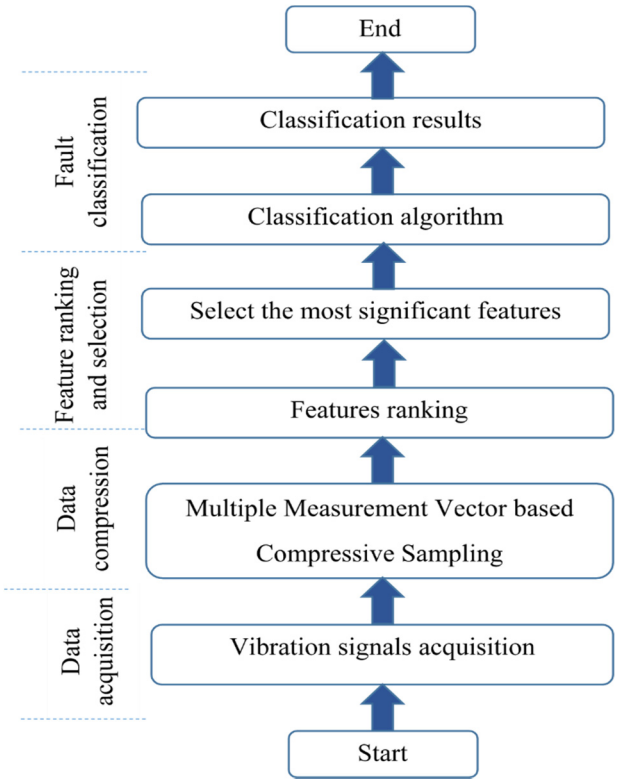


Figure. 4.4 Flowchart summarizing the steps of the CSFR framework.

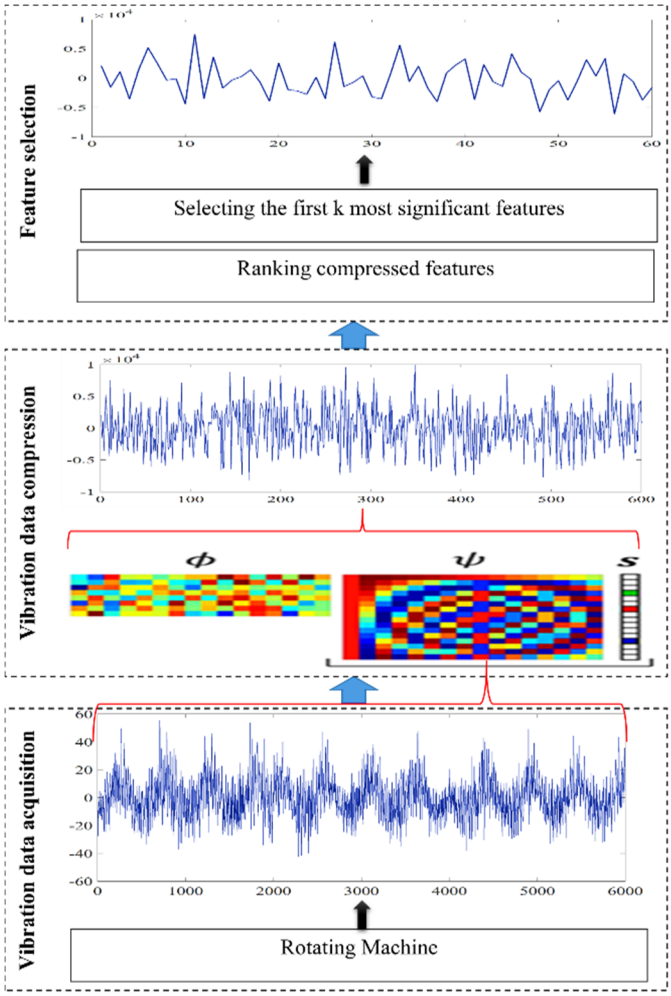


Figure. 4.5 Illustration of the data compression and feature selection.

Based on our proposed framework, we considered two techniques of feature selection to select fewer features of the compressively sampled signals. These are:

(1) Similarity-based methods: that assign similar values to the compressively sampled signals that are close to each other. Three algorithms (LS, FS, and Relief-F algorithms) were investigated to select fewer features based on similarity.

(2) Statistical-based methods: that measure the importance of feature of the compressively sampled signals using different statistical measures. Two algorithms, PCC and Chi-2 were investigated to select fewer features based on correlation and independence test respectively.

4.2.1 CS-LS

The compressive sampling and the Laplacian score feature selection (CS-LS) method receive a large amount of vibration data as input and produce fewer features with the smaller Laplacian scores as output, which can be used for fault classification of roller bearings (Ahmed, et al., 2017a). CS-LS first obtain the compressively-sampled data (Y) and then uses LS to rank the features of the compressively-sampled data. The LS rank the features depending on their locality preserving power. Given a

dataset $Y = [y_1, y_2, \dots, y_n]$, where $Y \in R^{m\,x\,L}$, suppose the Laplacian Score of the *r-th* feature is L_r and f_{ri} represent the *i-th* sample of the *r-th* feature where $i = 1, \dots, m$ and $r = 1, \dots, L$. First LS algorithm constructs the nearest neighbour graph G with *m* nodes, where the *i-th* node corresponds to y_i. Next, an edge between nodes *i* and *j* is placed, if y_i is among k nearest neighbors of y_j or vice versa, then *i* and j are connected. The elements of the weight matrix of graph G is S_{ij} and can be defined as follows:

$$S_{ij} = \begin{cases} e^{-\frac{\|y_i - y_j\|^2}{t}}, & i \text{ and } j \text{ are connected} \\ 0, & \text{otherwise} \end{cases} \tag{4.10}$$

Here *t* is a suitable constant. The Laplacian score L_r for each sample can be computed as follows:

$$L_r = \frac{\tilde{f}_r^T L \tilde{f}_r}{\tilde{f}_r^T D \tilde{f}_r} \tag{4.11}$$

where $D = diag(S\mathbf{1})$ is the identity matrix, $\mathbf{1} = [1, \dots, 1]^T$, $L = D - S$ is the graph Laplacian matrix, and $\tilde{f}_r$ can be calculated using the following equation:

$$\tilde{f}_r = f_r - \frac{f_r^T D\mathbf{1}}{\mathbf{1}^T D\mathbf{1}} \tag{4.12}$$

here $f_r = [f_{r1}, f_{r2}, \dots, f_{rm}]^T$.

CS-FR selects the k features with the smaller Laplacian scores (L_r) that can be used for fault classification of roller bearings (k < m).

4.2.2 CS-FS

The compressive sampling and Fisher score feature selection (CS-FS) method receive a large amount of vibration data as input and produce fewer features with the larger fisher scores as output, which can be used for fault classification of roller bearings (Ahmed and Nandi, 2017). Given a dataset of bearings vibration $X \in R^{nxL}$, CS-FS first uses CS based on MMV model to produce compressively-sampled signals, i.e., compressed data $Y = \{y_1, y_2, \dots, y_L\} \in R^m$ that have enough information of the original bearing raw data $X = \{x_1, x_2, \dots, x_L\} \in R^n$ where m << n. Then, it employs FS to rank the features of the compressively-sampled data (Y). The main idea of FS is to compute a subset of features with a large distance between the compressively-sampled data points in different classes and small distance between data points in the same class. The fisher score of the *i-th* feature can be computed by the following equation:

$$FS(Y^i) = \frac{\sum_{c=1}^{C} L_c\,(\mu_c^i - \mu^i)^2}{(\sigma^i)^2} \tag{4.13}$$

Here $Y^i \in R^{1\,x\,L}$, L_c is the size of the *c-th* class, $(\sigma^i)^2 = \sum_{c=1}^{C} L_c\,(\sigma_c^i)^2$, μ_c^i and σ_c^i are the mean and standard deviation of *c-th* class corresponding to the *i-th* feature; μ^i and σ^i are the mean and standard deviation of the entire dataset corresponding to the *i-th* feature. Finally, CS-FR selects the k features with the larger fisher scores that can be used for fault classification of roller bearings (k < m).

4.2.3 CS-Relief-F

The compressive sampling and Relief-F feature selection (CS-Relief-F) method receive a large amount of vibration data as input and produce fewer features as output, which can be used for fault classification of roller bearings. Similar to CS-FS and CS-LS methods described above, CS-Relief-F first obtain the compressively-sampled data ($Y \in R^{m\,x\,L}$) and then uses Relief-F to rank the features of the compressively-sampled data. The Relief-F technique uses a statistical approach to select the important features from the compressively-sampled data based on their weight *W*. The main idea of Relief-F is to randomly compute examples from Y and then calculate their nearest neighbours from the same class, also called the nearest hit, and the other nearest neighbours from different class, also called the nearest miss. The procedure of CS-Relief-F algorithm for feature ranking of the compressively-sampled signals is summarized below in algorithm 3.

Algorithm 3 Relief-F

Input: *l* learning instances, *m* features and *c* classes; Probabilities of classes $p_{y,}$ Sampling parameter *a*; Number of nearest instances from each class *d*;

Output: for each feature f_i a feature weight $-1 \le W[i] \le 1$;

1 for $i = 1$ to m do $W[i] = 0.0$; end for;

2 for $h = 1$ to a do

3 randomly compute an instance $\mathbf{y}_k$ with class y_k;

4 for y = 1 to c do

5 find d nearest instances y[j, c] from class c, $j = 1 \ldots d$;

6 for $i = 1$ to m do

7 for $j = 1$ to d do

8 if $y = y_k$ {nearest hit}

9 then $W[i] = W[i] - \text{diff}(i, \mathbf{y}_k, y[j, c])/(a*d)$;

10 else $W[i] = W[i] + p_y / (1 - p_{yk}) * \text{diff}(i, \mathbf{y}_k, y[j, y])/(a*d)$;

11 end if;

12 end for; {j} end for; {i}

13 end for; {y} end for ; {h}

14 return (W);

In this study, we applied *relieff* function in Statistics and Machine Learning Toolbox in MATLAB. This function returns the ranks and weights of features of the input matrix data and class labels vector c with *d* nearest neighbours. CS-Relief-F selects the k most important features, which can be used for fault classification of roller bearings (k < m).

4.2.4 CS-PCC

The compressive sampling and Pearson correlation coefficients feature selection (CS-PCC) method receive a large amount of vibration data as input and produce fewer features as output, which can be used for fault classification of roller bearings. Similar to CS-FS, CS-LS, CS-Relief-F methods described above, CS-PCC first obtains the compressively-sampled data (Y) and then uses PCC to rank the features of the compressively-sampled data. PCC examines the relationship between two variables according to their correlation coefficient (r), $-1 \leq r \leq 1$. Here the negative values indicate inverse relations, the positive values indicate a correlated relation, and the value 0 indicates no relation. PCC can be computed as follows:

$$r(i) = \frac{cov(y_i, c)}{\sqrt{var(y_i) * var(c)}} \tag{4.14}$$

Here $\boldsymbol{y_i}$ is the $\boldsymbol{i_{th}}$ variable, c is the class labels. CS-PCC selects the k features that are correlated with the class labels, which can be used for fault classification of roller bearings (k < m).

4.2.5 CS-Chi-2

The compressive sampling and Chi-square feature selection (CS-Chi-2) method receive a large amount of vibration data as input and produce fewer features as output, which can be used for fault classification of roller bearings. Similar to CS-FS and CS-LS methods described above, CS-Chi-2 first obtain the compressively-sampled data (Y) and then uses Chi-2 to rank the features of the compressively-sampled data. The χ^2 value for each feature *f* in a class labels group c of Y can be computed using the following equation:

$$\chi^2(f, c) = \frac{L(E_{c,f}E - E_c E_f)^2}{(E_{c,f} + E_c)(E_f + E)(E_{c,f} + E_f)(E_c + E)} \tag{4.15}$$

where *L* is the total number of examples in Y, $E_{c,f}$ is the number of times *f* and *c* co-occur, E_f is the number of time the feature *f* occurs without *c*, E_c is the number of times *c* occurs without *f*, and *E* is the number of times neither *f* nor *c* occurs. The bigger value of χ^2 indicates that the features are highly related.

In this study, we applied cross-tabulation function in Statistics and Machine Learning Toolbox in MATLAB. The cross-tabulation function returns the chi-square statistic and the obtained values of chi-2 are sorted in descending order to create a new feature vector with ranked features. CS-Chi-2 select

the k features with the bigger value of χ^2 that can be used for fault classification of roller bearings (k < m).

4.3 Compressive sampling and linear subspace learning (CSLSL) framework

The compressive sampling and linear subspace learning (CSLSL) based techniques receive a large amount of bearing vibration data and produce a fewer feature that can be used for fault classification of roller bearings. CSLSL based techniques first use MMV-CS model to produce compressively-sampled signals, i.e., compressed data $Y = \{y_1, y_2, \dots, y_L\} \in R^m$ that have enough information of the original bearing raw data $X = \{x_1, x_2, \dots, x_L\} \in R^n$. To extract feature representations of these signals, CSLSL based techniques perform a linear transformation to map the m-dimensional space of the compressively sampled vibration to a lower dimensional feature space, using the following equation:

$$\hat{y}_r = W^T y_r \tag{4.16}$$

Here r = 1, 2 … L, $\hat{y}_r$ is the transformed feature vector with reduced dimension, and W is a transformation matrix. There are three techniques based on CSLSL that have been proposed and tested in this thesis.

4.3.1 CS-PCA

The compressive sampling and Principal component analysis (CS-PCA) method receive a large amount of vibration data as input and produce fewer features as output, which can be used for fault classification of roller bearings (Ahmed, et al., 2017b). Given a dataset of bearings vibration $X \in R^{nxL}$, CS-PCA first uses CS based on MMV model to produce compressively-sampled signals, i.e., compressed data $Y = \{y_1, y_2, \dots, y_L\} \in R^m$ where $1 \leq l \leq L$, m << n and let each of these signals fit in with one of the *c* classes of roller bearings conditions. To find the larger attributes of the compressively-sampled vibration signals, CS-PCA uses PCA to compute W projection matrix using the scatter matrix, i.e., the covariance matrix *C* of the compressively-sampled data, which can be computed as follows:

$$C = \frac{1}{L} \sum_{i=1}^{L} (y_i - \bar{y})(y_i - \bar{y})^T \tag{4.17}$$

Here $\bar{y}$ is the mean of all samples. In the produced projection matrix W, successive column vectors from left to right correspond to decreasing eigenvalues. We select the m1 eigenvectors corresponding to the m1 largest eigenvalues. Hence, a new m1-dimensional space $\hat{Y}1 \in R^{m1\,x\,L}$ is produced from $Y \in R^{mxL}$, where m1 << m.

4.3.2 CS-LDA

The compressive sampling and Linear discriminant analysis (CS-LDA) method receive a large amount of roller bearings vibration data as input and produce fewer features as output, which can be used for fault classification of roller bearings (Ahmed et al., 2017b). CS-LDA produce a set of compressively-sampled signals $Y \in R^{mxL}$, here Y can be presented as $Y = [y_1, y_2, \dots, y_l]$ where$1 \leq l \leq L$, and let each of these signals fit in with one of the c classes of roller bearings conditions. To compute discriminant attributes from the compressively-sampled signals, CS-LDA method adopts LDA that considers maximizing the Fisher criterion function J (W), i.e., the ratio of the between the class scatter (S_B) to the within class scatter (S_w) such that

$$J(W) = \frac{|W^T S_B W|}{|W^T S_w W|} \tag{4.18}$$

where

$$S_B = \frac{1}{L}\sum_{i=1}^{c} l_i \left(\mu^i - \mu\right)(\mu^i - \mu)^T \tag{4.19}$$

$$S_w = \frac{1}{L}\sum_{i=1}^{c}\sum_{j=1}^{l_i}(y_j^i - \mu^i)(y_j^i - \mu^i)^T \tag{4.20}$$

Here μ^i is the mean vector of class i, y ϵ R of size L*m is the training dataset, y_1^i represents the dataset belong to the c-th class, n_i is the number of measurements of the i-th class, μ^i is the mean vector of class i, and μ is the mean vector of all training dataset. LDA projects the space of the compressively-sampled data onto a (c – 1)–dimension space by finding the optimal projection matrix W by maximizing J (W). Now W is composed of the selected eigenvectors ($\widehat{w}_1, \dots, \widehat{w}_{m2}$) with the first m2 largest eigenvalues (m2 = c - 1). Consequently, a new m2-dimensional space of discriminant attributes $\widehat{Y}2 \in R^{\,m2\,x\,L}$ is produced from $Y \in R^{mxL}$, where m2 << m.

With the fewer learned features, both CS-PCA and CS-LDA utilise Multinomial Logistic Regression classifier (LRC) described in section 2.7.1 to classify roller bearings health conditions.

4.3.3 CS with correlated principal and discriminant components (CS-CPDC)

The compressive sampling with Correlated Principal and Discriminant Components (CS-CPDC) is a three-stage hybrid method for classification of bearing faults (Ahmed and Nandi, 2018a, 2018c). In the first stage (Figure 4.6(a)), CS-CPDC uses MMV-CS to obtain compressively sampled raw vibration signals. In the second stage (Figure 4.6 (b)), it employs a multi-step approach of PCA, LDA, and Canonical correlation analysis (CCA) to extract features from the obtained compressively sampled signals. In the third stage (Figure 4.6 (c)), it applies SVM to classify bearing health condition using the learned features from the previous stage.

In the first stage, with the intention of reducing the amount of data and improving analysis effectiveness, CS-CPDC acquires compressively sampled signals using MMV-CS framework described in Algorithm 1. While the CS projections obtained in the first stage help to recover the original signal from low dimensional features, they may not be the best from a discriminant point of view. Furthermore, the size of the CS projections may still represent a large amount of data collected in real operating condition. Consequently, techniques to extract fewer features of the CS projections are required. Accordingly, PCA and LDA are commonly used. However, while an individual set of features (e.g., either PCA or LDA) can be good for representations, it may not be good for classifications. Thus, the aim of the second stage is to generate features for superior classification accuracy. The second stage consists of three steps as shown in Figure 4.6 (b).

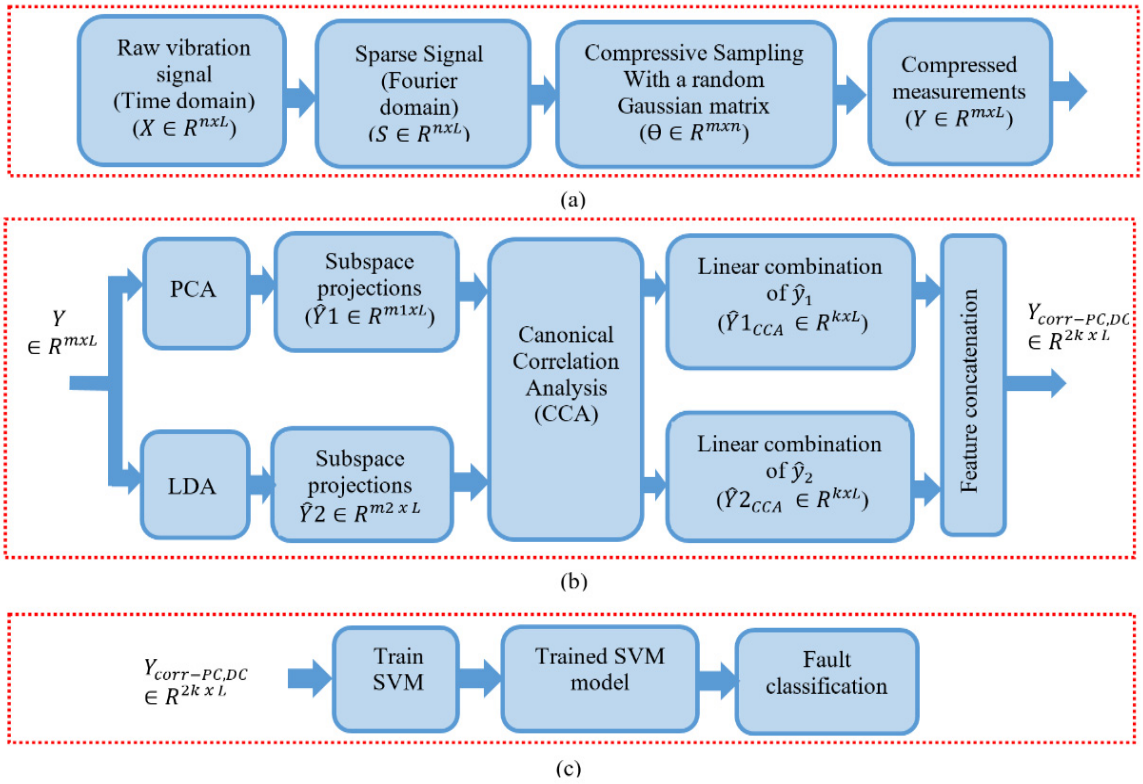


Figure. 4.6 Training of CS-CPDC method: (a) The first stage, (b) The second stage (c) The third stage

In the first step, CS-CPDC finds two feature representations from the compressively sampled signals using PCA and LDA respectively. Hence, we transform the characteristic space of the compressively sampled signal into a low dimensional space defined by those basis vectors corresponding to larger eigenvalue components (PCA). Furthermore, we augment these basis vectors with discriminant attributes learned through supervised learning (LDA). Let us consider a set of compressively-sampled signals $Y \in R^{mxL}$, here Y can be presented as $Y = [y_1, y_2, \dots, y_l]$ where $1 \leq l \leq L$, and let each of these signals fit in with one of the c classes of machine conditions. To extract feature representations of these signals, CS-CPDC performs a linear transformation to map the m-

dimensional space of the compressively sampled vibration to a lower dimensional feature space, using Equation (4.16) above.

To find the larger attributes of the compressively-sampled vibration signals, CS-CPDC uses PCA to compute W projection matrix as described above in subsection 4.3.1. Hence, a new m1-dimensional space $\hat{Y}1 \in R^{m1\,x\,L}$ is produced from $Y \in R^{mxL}$, where m1 << m. Furthermore, we employed LDA to compute discriminant attributes from the compressively-sampled signals as described above in subsection 4.3.2. Consequently, a new m2-dimensional space of discriminant attributes $\hat{Y}2 \in R^{m2\,x\,L}$ is produced from $Y \in R^{mxL}$, where m2 << m.

These different feature representations extracted from the same dataset always reflect different characteristics of the original signals. The best combination of them retains the multiple features of the integration that can be used effectively for classification. We propose CCA (Hardoon, et al., 2004) to combine PCA and LDA features to obtain superior classification.

The second step of the multi-step procedure of CS-CPDC utilises CCA to combine the different feature representations $\hat{Y}1$ and $\hat{Y}2$ by forming the relationship between them, i.e., maximising the overlapping variance between $\hat{Y}1$ and $\hat{Y}2$. The main idea is to find linear combinations of $\hat{Y}1$ and $\hat{Y}2$ that can maximize the correlation between them based on the following objective function

$$(W_1, W_2) = arg \max_{W_1, W_2} \acute{W}_1 C_{\hat{Y}1\hat{Y}2} \tag{4.21}$$

$$\text{s.t} \quad \acute{W}_1 C_{\hat{Y}1\hat{Y}1} W_1 = 1, \acute{W}_2 C_{\hat{Y}2\hat{Y}2} W_2 = 1.$$

where $C_{\hat{Y}1\hat{Y}2}$ is the cross-covariance matrix of $\hat{Y}1$ and $\hat{Y}2$ that can be computed using the following Equation (4.22),

$$C(\hat{Y}1, \hat{Y}2) = \hat{E}\left[\begin{pmatrix}\hat{Y}1\\ \hat{Y}2\end{pmatrix}\begin{pmatrix}\hat{Y}1\\ \hat{Y}2\end{pmatrix}'\right] = \begin{bmatrix} C_{\hat{Y}1\hat{Y}1} & C_{\hat{Y}1\hat{Y}2} \\ C_{\hat{Y}2\hat{Y}1} & C_{\hat{Y}2\hat{Y}2}\end{bmatrix} \tag{4.22}$$

The resulting linear combinations of $\hat{Y}1$ ($\hat{Y}1_{CCA} = W_1 * \hat{Y}1$) and the $\hat{Y}2$ ($\hat{Y}2_{CCA} = W_2 * \hat{Y}2$) will maximize their correlation. Finally, in the third step, the learned features $\hat{Y}1_{CCA}$ and $\hat{Y}2_{CCA}$ are concatenated to obtain a vector ($Y_{corr-PC,DC} \in R^{L\,x\,2k}$) that comprises highly correlated representations of principal and discriminative components where k is equal to the minimal dimension size of $m1$ and $m2$. These procedures are summarised in Algorithm 4. Figure 4.7 shows an illustration of the training process of the first and second stage of our proposed method.

In the third stage, CS-CPDC utilises Multi-class Support Vector Machine (SVM) classifier described in section 2.7.4 to classify roller bearings health conditions.

Algorithm 4: Feature learning stage

Input: $Y \in R^{m\,x\,L}$, $y \in R^{1\,x\,L}$: label information vector for each data points, c: number of classes, $m1$: selected number of principal components

Output: $Y_{corr-PC,DC} \in R^{L\,x\,2k}$

1: PCA(Y) $\longrightarrow$ $E1 \in R^{m\,xm1}$

2: $\hat{Y}1 = Y^T * E1$

3: LDA(Y, y) $\longrightarrow$ $E2 \in R^{m\,x\,m2}$; $m2 = c - 1$.

4: $\hat{Y}2 = Y^T * E2$

5: CCA ($\hat{Y}1, \hat{Y}2$) $\longrightarrow$ $w_1, w_2 \in R^{L\,x\,k}$; k =*min* ($m1$,$m2$).

6: $\hat{Y}1_{CCA} = w_1 * \hat{Y}1, \hat{Y}2_{CCA} = w_2 * \hat{Y}2$

7: $Y_{corr-PC,LD} = [\hat{Y}1_{CCA}\ \hat{Y}2_{CCA}]$

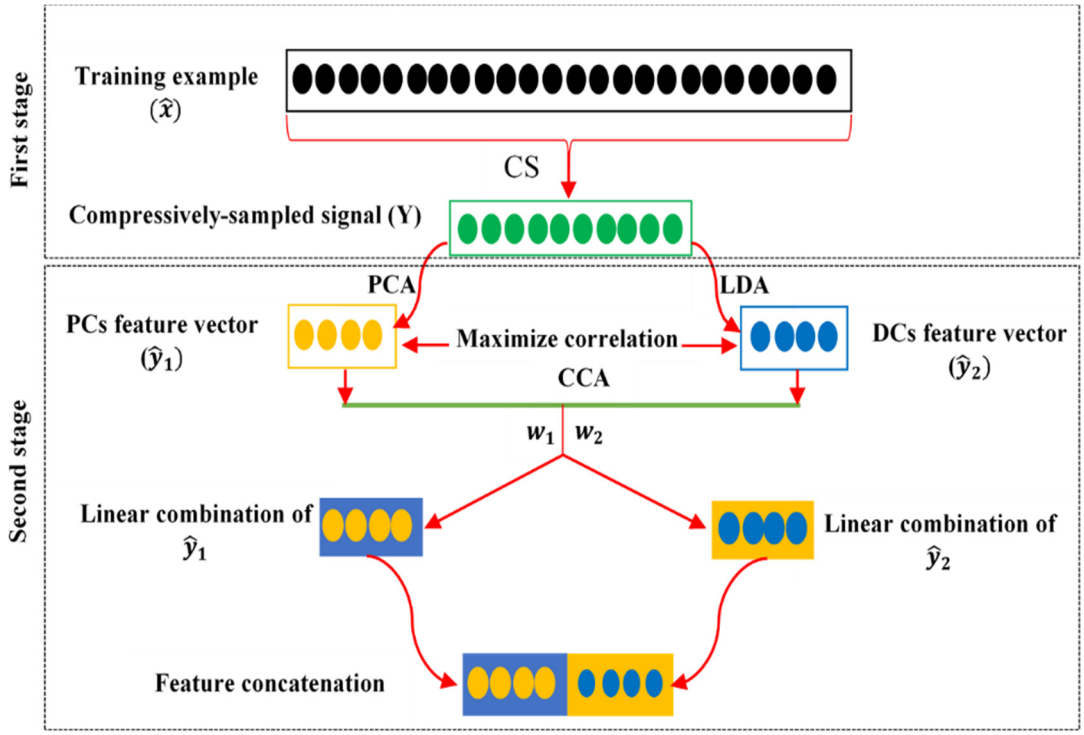


Figure. 4.7 Illustration of the training process of the first and second stage of the proposed method.

4.4 CS and sparse autoencoder based deep neural network (CS-SAE-DNN) method

This section presents a novel intelligent classification method for bearing faults from highly compressed measurements using sparse-over-complete features and training DNN through SAE (Ahmed, et al., 2018). In this method, we impose some flexible constraints for regularization on the hidden units of the sparse-Autoencoder. These include sparsity constraint that can be controlled by

different parameters such as, sparsity parameter, weight decay parameter and the weight of the sparsity penalty parameter. For the intent of learning sparse over-complete representations of our highly compressed measurements, the number of hidden units in each hidden layer is set to be greater than the number of input samples, and we used the encoder part of our unsupervised learning algorithm (i.e. The SAE). One important aspect in this proposed method is to pre-train the DNNs in an unsupervised manner using SAE that described in Subsection 2.7.3.1and then to fine-tune it with back-propagation (BP) algorithm for classification. The difficulty of multilayer training can be overcome with an appropriate set of features. One of the advantages of pre-training and fine-tuning in this approach is the power to mine flexibly fault features from highly compressed signals. Thus, the proposed approach is expected to achieve better classification accuracy compared with methods based on under-complete feature representations. Consequently, the efficiency of compressive sensing in machine faults classification is expected to be improved.

Previous research has shown that sparse representations of signals are able to signify the diagnosis features for machinery fault (Liu et al., 2011; Tang et al., 2014; Zhu et al., 2015; Fan et al., 2015; Ren et al., 2016). The advantages of sparse over-complete representations i.e. the number of obtained features is greater than the number of input samples have been studied Lewicki and Sejnowski, who concluded that over-complete bases can produce a better approximation of the underlying statistical distribution of the data (Lewicki and Sejnowski, 2000). Olshausen and Field (Olshausen and Field, 1997) and Doi et al. (Doi, et al., 2006) identify several advantages of over-complete basis set, for example, their robustness to noise and their ability to improve classification performance. Sparse feature learning methods normally contain two stages: (1) produce a dictionary W that represents the data $\{x_i\}_{i=1}^{N}$ sparsely using a learning algorithm, e.g. training artificial neural network (ANN) with sparsity penalties; and (2) obtain a feature vector from a new input vector using encoding algorithm.

Various recent studies investigating sparse feature representation have been carried out. These include Sparse Autoencoder (SAE), Sparse Coding (Liu et al., 2011), and Restricted Boltzmann machines (RBMs) (Lee et al., 2008). SAE approach has a number of attractive features: (1) simple to train, (2) the encoding stage is very fast, and (3) the ability to learn features when the number of hidden units is greater than the number of input samples. Therefore it was decided that SAE is an appropriate method to adopt for our investigation. In an analysis of Autoencoder (AE), Bengio and colleagues in (Bengio et al., 2007) found that AE can be used as a building block of a deep neural network (DNN), using greedy layer-wise pre-training.

CS-SAE-DNN applies a learning algorithm in multi-stages of non-linear feature transformation, each stage is a kind of feature transformation. One way to do this is by using DNN with multiple hidden layers and each layer is connected to the layers below it in a non-linear combination.

In the pre-training stage, sparse-autoencoder is used to train the DNN, the encoder part of the sparse-autoencoder with sigmoid activation function is used to learn the over-complete feature representations.

As expressed in Figure 4.8, the proposed method produces over-complete representations from the input compressed measurements (obtained using the CS model described in Algorithm 2 above) by setting the number of hidden units (d_i) to be greater than the number of input samples(m), i.e., $d_i > m$ in each hidden layer (i), where $d_{i+1} > d_i$ for i =1, 2, 3, …, n. and Input (n) represent the output of Encoder (n-1), and d_n is the number of hidden layers in Encoder (n). As drawn in section 3, DNN training includes two levels of training, namely, pre-training using unsupervised learning algorithm and re-training using backpropagation algorithm. In the pre-training stage, the unlabelled bearing compressed measurements (y) are first used to train DNN by setting the parameters in each hidden layer and compute the sparse-over-complete feature representations. In fact, in the DNN based on sparse autoencoder, we are making use of the SAE algorithm applied multiple times through the network. Therefore, the output over-complete feature vector from the first encoder is the input of the second encoder.

Finally, the fault classification is achieved using two stages, namely, (1) pre-training classification based on stacked autoencoder and softmax regression layer which is the deep net stage (the first stage), and (2) re-training classification based on backpropagation (BP) algorithm and that is the fine-tuning stage (the second stage). Figure 4.9 shows an illustration of the pre-training process using two hidden layers. With enforcement of sparsity constraints and by setting the number of units in each hidden layer to be greater than input samples, each autoencoder learns useful features of the compressed unlabelled training samples. The training process is performed by optimizing the cost function $CF_{sparse}(W, b)$ in the following equation.

$$\mathrm{CF_{sparse}(W,b)} = \frac{1}{2m}\sum_{i=1}^{m}\left\|\tilde{y}_i - y_i\right\|^2 + \lambda\|W\|^2 + \beta\sum_{j=1}^{d} KL\,(\rho\|\hat{\rho}) \qquad (4.23)$$

where m is the input size, i.e., the highly compressed measurements size, d is the hidden layer size, λ is the weight decay parameter, and β is the weight of the sparsity penalty term, ρ is a sparsity parameter, and $\hat{\rho}$ is the average threshold activation of hidden units.

The optimization is performed using the Scaled Conjugate Gradient (SCG) which is a member of the Conjugate Gradient (CG) methods (Møller, 1993). In the first learning stage, the encoder part of the first SAE with sigmoid functions of the range of unit activation function values [0, 1] is used to learn features from compressed vibration signals of length m, where the number of hidden units $d_1 > m$ and the extracted over-complete features (v_1) are used as the input signals for the second learning stage. Then Encoder 2 of the second SAE with a number of hidden units d_2 is used to extract over-

complete features (v_2) from (v_1). Finally, softmax regression is trained using (v_2) to classify bearings health conditions.

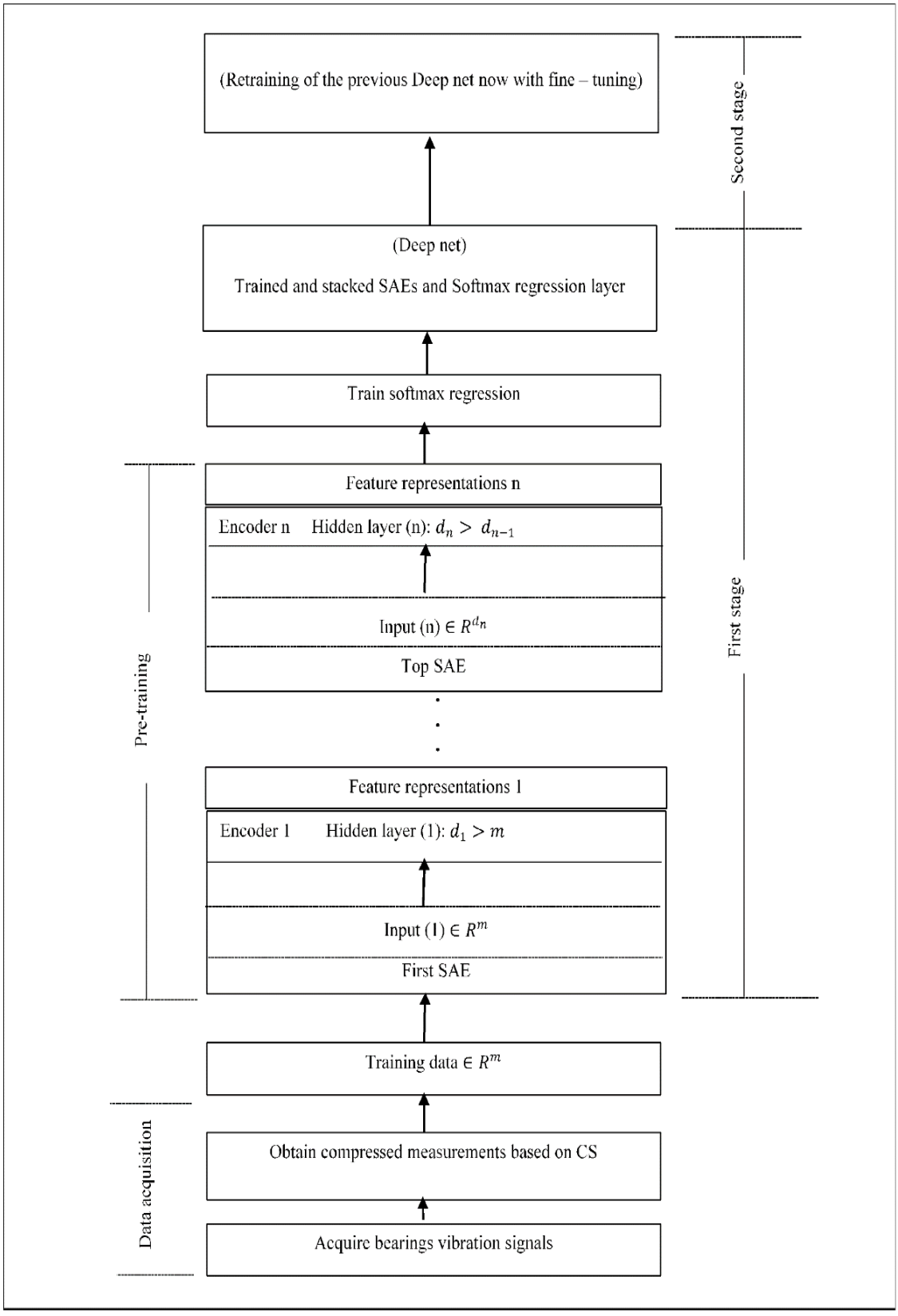


Figure. 4.8 Training of our proposed method.

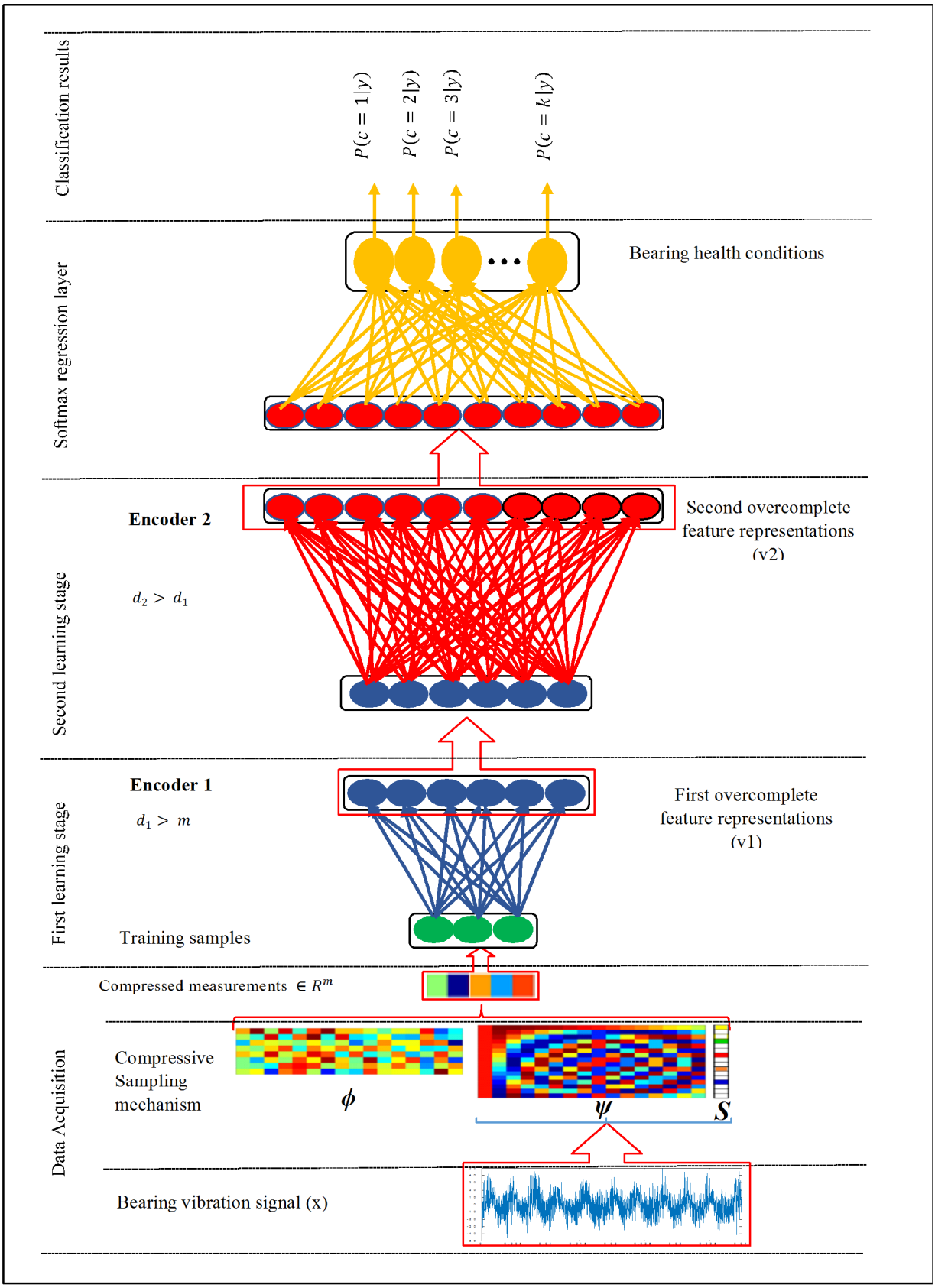


Figure. 4.9 Illustration of the proposed method using two hidden layers. Data flow from the bottom to the top.

The pre-training process can be described in the following steps:

Given a DNN of n hidden layers, the pre-training process using sparse autoencoder to learn over-complete features will be conducted on each layer.

(1) Initialization, m is the number of input samples, n is the number of hidden layers, $d_0 = m, i = 1, 2, \dots, n$

(2) For $i = 1\ to\ n$,

(3) Set up the number of hidden units (d_i) to be greater than the input samples (m), i.e. $d_i > d_{i-1}$.

(4) Set up sparsity parameter, weight decay parameter, weight of the sparsity penalty term and the maximum training epoch that achieve the lowest possible reconstruction error E ($x_i, \tilde{x_i}$).

(5) Use scale conjugate gradient training (SCG) algorithm for network training. By utilizing SCG, the learning rate is automatically adopted at each epoch and the average threshold activation of hidden units ($\hat{\rho}$) can be computed using Eq. (2.94).

(6) Based on Eq. (4.23), compute the cost function.

(7) Using the encoder part of the sparse autoencoder calculate the output over-complete feature vector v_i and use it as the input of the following hidden layer.

(8) $i = i + 1$.

(9) Repeat steps 2, to 8 until $i = n$.

(10) Use the over-complete feature vector of the last hidden layer v_n to be the input of the softmax regression layer.

Chapter 5
CSFR framework Assessment and Validation

In this chapter, we present several experiments that we have conducted to verify the capability of the CSFR framework, which described in Section 4.2, for rolling bearing health conditions diagnosis. This achieved by investigating the combination of MMV-CS and several feature ranking techniques – i.e., FS, LS, Relief-F, PCC, and Chi-2 – to learn fewer features from a large amount of vibration data from which bearing health conditions are classified . Three classification algorithms, LRC, ANN, and SVM are used to classify bearing health conditions. Two case studies of bearing health conditions diagnosis are used in this investigation with the aims to (1) validate the proposed framework for bearing health conditions diagnosis, and (2) observe the best combinations of MMV-CS, feature selection techniques, and classifiers with reduced complexity and improved classification accuracy. This work has been published in (Ahmed et al., 2017a; Ahmed and Nandi, 2017; Ahmed and Nandi, 2018b).

In an attempt to validate the proposed framework, various experiments were conducted using two of the bearing vibration datasets that were discussed in chapter 3. These are the first bearing vibration data and dataset G of the second bearing vibration data. Each of these datasets was applied to our proposed framework using different combinations of techniques detailed in Section 4.2.

5.1 The first bearing vibration data

5.1.1 Experimental setup

To apply our proposed framework in this dataset, 50% of the vibration data are randomly selected for training and the other 50% are used for testing. To obtain compressively sampled signals from the original vibration signals of roller bearings, MMV-CS model with two different sparse representations techniques, i.e., thresholded Haar WT and FFT, which described in Subsection 4.1, are used. First, we used the Haar wavelet basis with five decomposition levels as sparsifying transform where the wavelet coefficients are thresholded using the penalized hard threshold to obtain sparse representations of the original vibration signals. Second, we used FFT to obtain the sparse components. Then we utilised compressive sampling framework with different sampling rates (α) (0.1, 0.2, 0.3 and 0.4) with 600, 1200, 1800, and 2400 compressed measurements from our original vibration signals using a random matrix with i.i.d. Gaussian entries which satisfy the RIP.

To ensure that our CS model generates enough samples for fault classification, we used the generated compressively sampled signals in each of the sparse representation methods to reconstruct

the original signal X using the Compressive Sampling Matching Pursuit (CoSaMP) algorithm (Needell, et al., 2010). The reconstruction errors are measured by Root Mean Squared Error (RMSE). By using thresholded WT based CS with α = 0.1, the RMSE for the six conditions of bearings are 8.5% (NO), 24.6% (NW), 15.23% (IR), 12.71% (OR), 11.87 % (RE), and 5.29% (CA). This has been studied in details in (Ahmed *et al.*, 2018a). While for FFT based CS using the same sampling rate α = 0.1, the RMSE values are 4.8% (NO), 8.9% (NW), 6.3% (IR), 5.6% (OR), 4.7 % (RE), and 3.6% (CA), which indicate good signal reconstruction.

Based on the theory of CS, these compressively sampled signals possess the quality of the original signals. For further filtering, we applied FS, LS, Relief-F, PCC, and Chi-2 to select fewer features (*k*) from these compressively-sampled signals. Finally, with these selected features, we applied multinomial LRC, ANN, and SVM with ten-fold cross-validation to deal with the classification problem. Classification accuracies are obtained by averaging the results of twenty trials for each classifier and for each experiment.

5.1.2 Results

Table 5.1, Table 5.2, and Table 5.3 present testing classification results for LRC, ANN, and SVM respectively using two different sparsifying transforms, i.e., FFT and thresholded Haar WT, to obtain the compressively-sampled signals using the aforementioned compressive sampling rates. The results in Table 5.1, Table 5.2, and Table 5.3 show that among the various proposed combinations of CS with FFT, feature selection techniques, and classifiers, most of the combinations with LRC and ANN achieved better results than with SVM with "fitcecoc" function. In particular, results from CS-Chi-2 and CS-Relief-F for all values of the sampling rate (α) and the number of selected features (k) with both LRC and ANN are better than with SVM. Also, all the combinations of CS with FFT and the considered feature selection techniques with LRC and ANN achieved very high classification accuracies (all above 99%) for all values of α with k = 120.

Also, CS-FS, CS-LS, and CS-PCC with LRC and ANN achieved better results than with SVM for all values of α with k = 60 and 120. Moreover, all classification accuracies are above 99% for all the classifiers considered with CS-FS, CS-Relief-F, CS-PCC, and CS-Chi-2 with both WT and FFT sparse representations techniques using α = 0.4 and k =180. For CS-LS all considered classifiers achieved accuracy results above 99% using α = 0.4 and k =180 with FFT only. SVM achieved good results in several scenarios with the larger number of selected features, i.e., k =180: (1) using CS-FS with FFT, α = 0.2, 0.3, and 0.4, and with thresholded WT and α = 0.4, (2) using CS-LS with FFT and all values of α, (3) using CS-Relief-F with WT and α = 0.3 and 0.4, (4) CS-PCC with FFT, and with WT for α = 0.4, (5) using CS-Chi-2 with WT and all values of α.

Table 5.1. Classification accuracy of roller bearings health conditions for LRC with different combinations of MMV-CS and feature ranking and selection techniques (all classification accuracies ≥ 99% in bold).

Sparsifying method		FFT				WT			
Sampling rate (α)		0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
Classifier: LRC									
Method	k								
CS-FS	60	94.3 ± 5.6	98.7 ± 0.7	98.8 ± 0.5	98.1 ± 0.9	92.2 ± 7.5	96.4 ± 3.9	97.2 ± 2.8	98.0 ± 2.2
	120	**99.7 ± 0.4**	**99.8 ± 0.3**	**99.8 ± 0.2**	**99.9 ± 0.4**	96.3 ± 3.8	96.2 ± 4.0	**99.2 ± 0.8**	**99.9 ± 0.1**
	180	**99.9 ± 0.1**	**99.9 ± 0.2**	**99.9 ± 0.1**	**100.0 ± 0.0**	98.9 ± 1.2	97.8 ± 2.2	**99.7 ± 0.3**	**99.9 ± 0.2**
CS-LS	60	95.8 ± 0.9	93.4 ± 1.2	95.2 ± 3.7	97.4 ± 2.8	91.9 ± 8.2	93.3 ± 6.9	95.1 ± 4.8	96.6 ± 3.5
	120	**99.5 ± 0.3**	**99.8 ± 0.6**	**99.4 ± 0.6**	**99.9 ± 0.1**	92.8 ± 7.2	94.2 ± 5.7	96.6 ± 3.4	98.0 ± 2.1
	180	**99.8 ± 0.2**	**99.9 ± 0.1**	**99.9 ± 0.1**	**99.9 ± 0.1**	95.4 ± 4.6	95.5 ± 4.4	96.5 ± 3.7	98.9 ± 1.1
CS-Relief-F	60	**99.5 ± 0.3**	**99.3 ± 0.6**	**99.5 ± 0.2**	**99.4 ± 0.3**	94.1 ± 5.1	93.7 ± 6.4	95.8 ± 3.4	96.6 ± 3.5
	120	**99.8 ± 0.3**	**99.9 ± 0.2**	**99.9 ± 0.1**	**99.9 ± 0.1**	96.4 ± 3.2	95.9 ± 4.2	98.3 ± 1.8	**99.9 ± 0.1**
	180	**99.9 ± 0.1**	**99.9 ± 0.1**	**100.0 ± 0.0**	**100.0 ± 0.0**	98.4 ± 1.6	**99.3 ± 0.7**	**100.0 ± 0.0**	**100 ± 0.0**
CS-PCC	60	98.5 ± 1.3	**99.3 ± 0.5**	98.8 ± 0.4	98.4 ± 0.9	94.1 ± 5.7	94.3 ± 4.6	97.4 ± 2.3	98.6 ± 1.5
	120	**99.8 ± 0.3**	**99.5 ± 0.2**	**99.8 ± 0.2**	**99.9 ± 0.1**	95.0 ± 4.9	95.0 ± 4.9	98.1 ± 1.7	98.9 ± 1.3
	180	**99.9 ± 0.1**	**99.9 ± 0.2**	**99.9 ± 0.2**	**99.9 ± 0.1**	98.8 ± 1.2	**99.0 ± 1.1**	98.8 ± 1.2	**99.4 ± 0.6**
CS-Chi-2	60	98.0 ± 2.2	98.7 ± 1.4	98.8 ± 1.2	98.8 ± 1.2	95.1 ± 4.5	96.7 ± 2.7	98.3 ± 2.8	98.9 ± 1.2
	120	**99.5 ± 0.5**	**99.5 ± 0.4**	**99.5 ± 0.3**	**99.5 ± 0.5**	96.8 ± 2.9	97.9 ± 3.2	96.6 ± 3.4	**99.9 ± 0.1**
	180	**99.9 ± 0.1**	**99.9 ± 0.1**	**100.0 ± 0.0**	**99.9 ± 0.1**	**99.5 ± 0.4**	**99.3 ± 0.7**	**100.0 ± 0.0**	**100.0 ± 0.0**

Table 5.2. Classification accuracy of roller bearings health conditions for ANN with different combinations of MMV-CS and feature ranking and selection techniques (all classification accuracies ≥ 99% in bold).

Sparsifying method		FFT				WT			
Sampling rate (α)		0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
Classifier: ANN									
Method	k								
CS-FS	60	93.8 ± 6.3	98.3 ± 0.8	**99.0 ± 1.1**	98.9 ± 0.4	91.6 ± 8.3	97.1 ± 2.8	98.9 ± 1.2	97.7 ± 2.3
	120	**99.2 ± 1.3**	**99.7 ± 0.7**	**99.7 ± 0.6**	**99.6 ± 0.9**	97.2 ± 2.8	98.3 ± 1.7	**99.0 ± 0.9**	**99.3 ± 0.7**
	180	**99.8 ± 0.5**	**99.8 ± 0.3**	**99.7 ± 0.3**	**100.0 ± 0.0**	98.2 ± 1.9	**99.1 ± 0.9**	**99.3 ± 0.8**	**99.7 ± 0.3**
CS-LS	60	98.6 ± 0.9	97.3 ± 2.1	97.9 ± 2.2	95.8 ± 3.3	92.9 ± 7.0	93.7 ± 6.3	94.4 ± 5.5	97.1 ± 2.7
	120	**99.9± 0.2**	**99.4 ± 0.7**	**99.9 ± 0.1**	**99.9 ± 0.1**	95.3 + 4.8	95.2 + 4.8	96.9 + 3.2	98.3 + 1.7
	180	**99.9 ± 0.1**	**99.9 ± 0.2**	**99.9 ± 0.1**	**99.9 ± 0.1**	96.8 ± 2.9	95.9 ± 4.1	97.8 ± 2.1	98.9 ± 1.0
CS-Relief-F	60	**99.7 ± 0.6**	**99.7 ± 0.3**	**99.2 ± 0.9**	**99.6± 0.3**	94.6 ± 5.4	94.4 ± 5.7	93.2 ± 6.7	97.1 ± 2.7
	120	**99.9 ± 0.2**	**99.7 ± 0.4**	**99.6 ± 0.9**	**99.7 ± 0.4**	97.2 ± 2.8	97.0 ± 3.1	**99.5 ± 0.5**	**99.8 ± 0.2**
	180	**99.9± 0.3**	**100 ± 0.0**	**100.0 ± 0.0**	**99.9 ± 0.3**	**99.1 ± 0.9**	**99.9 ± 0.1**	**100.0 ± 0.0**	**100.0 ± 0.0**
CS-PCC	60	**99.4 ± 0.3**	**99.4 ± 0.6**	**99.2 ± 0.9**	**99.5 ± 0.5**	94.9 ± 5.3	95.7 ± 4.3	98.5 ± 1.6	94.4 ± 5.5
	120	**99.8 ± 0.3**	**99.7 ± 0.4**	**99.9 ± 0.3**	**99.9 ± 0.1**	96.4 ± 3.6	96.3 ± 3.8	**99.0 ± 0.9**	98.5 ± 1.4
	180	**99.7± 0.5**	**100 ± 0.0**	**100 ± 0.0**	**99.9 ± 0.2**	**99.3 ± 0.7**	**99.6 ± 0.3**	**99.3 ± 0.7**	**99.9 ± 0.1**
CS-Chi-2	60	96.9± 2.0	95.2± 3.9	97.8 ± 0.9	98.3 ± 0.9	95.8 ± 3.9	96.3 ± 3.7	98.5 ± 2.7	98.9 ± 1.2
	120	**99.2 ± 0.9**	**99.4 ± 0.8**	**99.0 ± 1.1**	**99.3± 0.8**	98.8 ± 0.4	98.1 ± 2.9	**99.5 ± 0.5**	**99.9 ± 0.1**
	180	**99.9± 0.3**	**99.9 ± 0.3**	**99.6 ± 0.6**	**99.9 ± 0.2**	**99.2 ± 0.7**	**99.9 ± 0.1**	**100.0 ± 0.0**	**100.0 ± 0.0**

Table 5.3. Classification accuracy of roller bearings health conditions for SVM with different combinations of MMV-CS and feature ranking and selection techniques (all classification accuracies ≥ 99% in bold).

Sparsifying method		FFT				WT			
Sampling rate (α)		0.1	0.2	0.3	0.4	0.1	0.2	0.3	0.4
Classifier: SVM Method	k								
CS-FS	60	85.8 ± 8.4	89.5 ± 5.1	74.8 ± 5.6	69.9 ± 2.6	91.5 ± 8.4	95.2 ± 4.7	95.9 ± 3.7	95.7 ± 3.9
	120	89.6 ± 11.4	92.3 ± 6.2	98.4 ± 1.6	97.9 ± 3.9	94.9 ± 4.9	94.8 ± 5.2	97.5 ± 2.3	97.9 ± 2.4
	180	97.4 ± 5.2	**99.1 ± 0.9**	**99.3 ±0.7**	**99.8 ± 1.2**	95.8 ± 3.9	96.3 ± 3.8	98.8 ± 1.4	**99.4 ± 0.7**
CS-LS	60	92.7 ± 7.0	92.1 ± 5.1	92.0 ± 0.7	95.3 ± 4.7	92.6 ± 7.4	92.2 ± 7.7	91.9 ± 8.3	92.7 ± 6.9
	120	98.7 ± 1.4	98.9 ± 1.1	**99.3 ± 0.8**	**99.7 ± 0.3**	91.4 ± 6.9	93.5 ± 6.4	94.7 ± 5.2	95.7 ± 4.2
	180	**99.2 ± 0.8**	**99.2 ± 0.7**	**99.2 ± 0.9**	**99.9 ± 0.1**	94.4 ± 5.4	93.2 ± 6.9	95.3 ± 4.6	97.8 ± 2.3
CS-Relief-F	60	85.8 ± 13.7	77.9 ± 1.5	68.4 ± 16.1	67.8 ± 16.4	94.7 ± 5.3	94.2 ± 5.9	92.7 ± 7.4	94.2 ± 4.8
	120	89.3 ± 8.4	90.9 ± 7.8	83.9 ± 12.2	78.5 ± 9.7	95.3 ± 4.7	95.5 ± 4.4	97.5 ± 2.7	98.8 ± 1.3
	180	95.6 ± 5.8	96.2 ± 4.8	88.0 ± 9.4	96.8 ± 4.5	97.7 ± 2.1	98.8 ± 1.2	**99.3 ± 0.6**	**99.0 ± 1.2**
CS-PCC	60	78.3 ± 14.0	73.2 ± 7.4	69.4 ± 13.1	65.1 ± 15.4	71.5 ± 17.8	79.9 ± 17.8	85.8 ± 13.2	94.7 ± 5.3
	120	93.2 ± 7.5	93.8 ± 7.9	93.3 ± 8.5	79.9 ± 2.2	78.2 ± 16.2	82.2 ± 15.8	88.2 ± 10.9	92.4 ± 7.5
	180	97.7 ± 5.1	96.7 ± 5.7	97.8 ± 2.8	**99.3 ± 0.7**	83.5 ± 13.7	87.4 ± 11.6	92.7 ± 7.6	**99.3 ± 0.7**
CS-Chi-2	60	64.2 ± 7.1	68.3± 7.6	63.7± 9.3	69.1 ± 18.4	94.3 ± 3.8	95.2 ± 3.2	97.7 ± 3.4	98.2 ± 2.8
	120	83.2 ± 9.0	76.9 ± 9.3	74.8 ± 11.9	71.4 ± 7.1	96.2 ± 3.7	97.1 ± 3.4	98.5 ± 1.7	98.8 ± 1.2
	180	95.2 ± 5.2	94.9 ± 6.3	90.7 ± 7.0	82.3 ± 14.4	**99.6 ± 0.3**	**99.6 ± 0.2**	**99.8 ± 0.2**	**99.9 ± 0.2**

Generally, the classification accuracies of all the proposed methods, i.e., CS-FS, CS-LS, CS-Relief-F, CS-PCC, and CS-Chi-2 are based on the compressed sampling rate (α), the feature selection method, and the number of the selected features (k). However, the features are selected from random compressed projections of length (m) that do not include representations of all the attributes in the original data of length (n), i.e., $m \neq n$. Therefore, the assumption that when α gets larger the accuracy gets better may not apply in every set of selected features. For example, it can be clearly seen from the results in Table 5.1 that the classification accuracy of all the proposed methods with FFT sparsifying method becomes better when α becomes larger with k = 180. While with k = 120, only one variation was found using CS-PCC with accuracy 99.8% and 99.5% for α equal to 0.1 and 0.2 respectively. With k = 60, five variations were found; one variation using CS-FS with accuracy 98.8% and 98.1% for α equal to 0.3 and 0.4 respectively; one variation using CS-LS with accuracy 95.8% and 93.4% for α equal to 0.1 and 0.2 respectively; two variations using CS-Relief-F, first with accuracy 99.5%, 99.3% for α equal to 0.1 and 0.2, and second with 99.5% and 99.4% accuracy for 0.3, and 0.4 respectively; one variation using CS-PCC with 98.8% and 98.4% for 0.3, and 0.4 respectively.

Taken together, these results show that the proposed framework with various methods studied here has the ability to classify bearing health conditions with a high classification accuracy with the following comments:

1) FFT as a sparsifying transform method for our proposed MMV based CS model can achieve better results than thresholded WT.

2) LRC and ANN have the ability to achieve high classification accuracy with different values of the sampling rate (α) and a number of selected features (k) for all the considered CS and feature selection techniques combinations.

3) SVM has the ability to achieve good classification accuracy with the larger number of selected features, i.e., k =180, and larger values of α, e.g., α = 0.4, for certain combinations.

4) With the larger number of selected features, all the proposed methods achieved high classification accuracy. Thus, for the application of the proposed methods in fault diagnosis, we recommend selecting a larger number of features from the compressively-sampled signals.

5.1.3 Comparisons of results

For further verification of the efficiency of the proposed framework, complete comparison results of the classification accuracy using the different combinations based on the proposed framework compared with some recently published results using the same vibration dataset, for instance in (Wong *et al.*, 2015) results reported for three methods, one method uses all the original vibration data from which entropic features are extracted, and the other two uses compressed measurements to recover the original vibration signals and from the recovered signals entropic features are extracted. With the extracted features SVM used to classify bearing health conditions. Moreover, a hybrid model consisting of the Fuzzy Min-Max (FMM) neural network and Random Forest (RF) with Sample Entropy (SampEn) and Power Spectrum (PS) features is used to classify bearing health conditions and the results reported in (Seera et al., 2017). In (Guo et al., 2005), a Genetic Programming (GP) based approach is proposed for feature extraction from raw vibration data and with extracted features SVM and ANN are used to classify bearing health conditions. Table 5.4, presents classification results of bearing health conditions using our proposed methods with α = 0.4, 0.3, and 0.1 and the reported results in the above-mentioned studies using the same dataset used in this case study.

It can be clearly seen that classification results of our proposed methods are better than those reported in (Wong et al., 2015) and (Guo et al., 2005). Also, our results are as good as, if not better than results reported in (Seera et al., 2017) although we are using only 10% (α = 0.1) of the original vibration data that is not matched by the method in (Seera et al., 2017) using all the raw vibration data.

Table 5.4. A comparison with the classification results from literature on the first bearing dataset.

		Classification accuracy (%)
Raw Vibration with entropic features (Wong et al., 2015)		98.9 ± 1.2
Compressed sampled with α = 0.5 followed by signals reconstruction (Wong et al., 2015)		92.4 ± 0.5
Compressed sampled with α = 0.25 followed by signals reconstruction (Wong et al., 2015)		84.6 ± 3.4
FMM-RF (SampEn + PS) (Seera et al., 2017)		99.81± 0.41
GP generated feature sets (un-normalised) (Guo et al., 2005)		
	ANN	96. 5
	SVM	97.1
Our proposed framework with FFT, α = 0.1, k = 120		
CS-FS	LRC	99.7 ± 0.4
	ANN	99.2 ± 1.3
CS-LS	LRC	99.5 ± 0.3
	ANN	99.9± 0.2
CS-Relief-F	LRC	99.8 ± 0.3
	ANN	99.9 ± 0.2
CS-PCC	LRC	99.8 ± 0.3
	ANN	99.8 ± 0.3
CS-Chi-2	LRC	99.5 ± 0.5
	ANN	99.2 ± 0.9

This section has validated the proposed framework and has shown that the many combinations of CS and feature ranking methods achieved high classification accuracies of bearing faults. The next section of this paper will validate the usage of our proposed framework using publicly available bearing vibration dataset. The advantage of the shared dataset is that we can compare the results of other researchers easily.

5.2 Dataset G: The second bearing vibration data

5.2.1 Experimental setup

To apply our proposed framework in this dataset, 240 examples are randomly selected for training and 480 examples are used for testing. We applied the MMV-CS model with FFT to obtain compressively-sampled signals from the raw vibration signals using α equal to 0.1 and the same feature

selections methods as in the first bearing vibration data to select fewer features of these compressively-sampled signals. With these fewer selected features, we employed LRC, ANN, and SVM to deal with the classification problem. The classification accuracies are achieved by averaging the results of twenty trials for each classifier and for each experiment.

5.2.2 Results

Table 5.5 shows the accuracy of all experiments with a different number of selected features (k = 25 and 50). Figure 5.1 and Figure 5.2 show a column chart representations of the classification results presented in Table 5.5.

Table 5.5. Classification accuracy of roller bearings health conditions of bearing dataset G of the second machine with different Combinations of MMV-CS and Feature Ranking and Selection Techniques (all classification accuracies ≥ 99% in bold)

Method	LRC	ANN	SVM
CS-FS			
(k = 25)	98.4 ± 1.6	99.6 ± 0.5	97.4 ± 2.7
(k = 50)	99.9 ± 0.1	100 ± 0.0	99.9 ± 0.1
CS-LS			
(k = 25)	99.1 ± 0.8	99.2 ± 0.8	98.5 ± 1.6
(k = 50)	97.5 ± 2.6	99.4 ± 0.7	98.3 ± 1.7
CS-Relief-F			
(k = 25)	99.3 ± 0.6	99.2 ± 0.8	97.8 ± 2.4
(k = 50)	99.4 ± 0.5	100 ± 0.0	99.5 ± 0.5
CS-PCC			
(k = 25)	99.2 ± 0.8	99.5 ± 0.6	97.9 ± 1.9
(k = 50)	99.3 ± 0.7	99.9 ± 0.1	97.5 ± 2.4
CS-Chi-2			
(k = 25)	97.5 ± 2.6	99.9 ± 0.1	94.7 ± 5.4
(k = 50)	99.1 ± 0.9	100 ± 0.0	99.9 ± 0.1

As follows from the Table 5.5, all combinations of CS and feature ranking and selection techniques with ANN and a different number of selected features, i.e., k = 25 and 50 achieved high classification accuracies (all over 99%). As well as achieving high classification accuracy with ANN using 25 features, all combinations with ANN and 50 selected features are able to achieve even higher

classification accuracy. In particular, results of CS-FS, CS-Relief-F, and CS-Chi-2 with ANN and 50 selected features achieved 100% classification accuracy for every single run in our investigations.

The average classification accuracies of LRC obtained using 50 selected features using CS-FS, CS-Relief-F, CS-PCC, and CS-Chi-2 are above 99%. Also, with 25 selected features based on CS-LS, CS-Relief-F, and CS-PCC, LRC is able to achieve over 99% classification accuracy. In addition, with SVM and 50 selected features based on CS-FS, CS-Relief-F, and CS-Chi-2, the average classification accuracy rates are generally above 99%. These observations can be clearly seen in Fig. 5.1 and 5.2 below. However, from Table VII it can be clearly seen that for k = 25, the classification accuracies of CS-LS-LRC, CS-LS-SVM, and CS-PCC-SVM methods are 99.9%, 98.5%, and 97.9% respectively; while for k=50, the classification accuracies for the same methods are 97.5%, 98.3%, and 97.5% respectively. Therefore, for a fixed compressed signal size *m*, there is an optimal number of features *k* that makes the classification accuracy higher than other classification accuracies achieved using a different number of features that may be bigger or smaller than *k*.

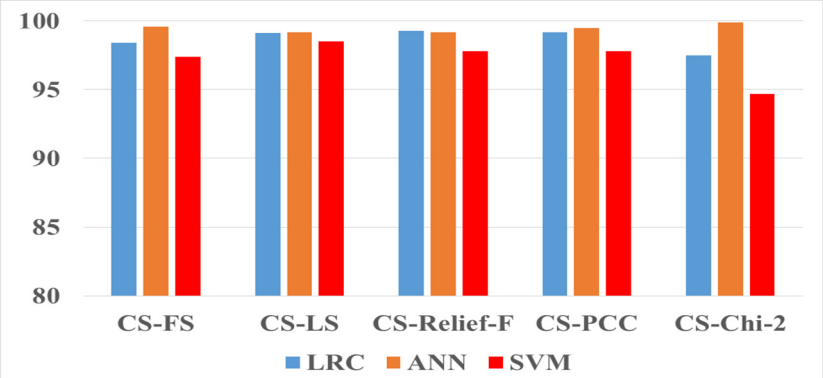


Figure. 5.1. Classification accuracy rates of 25 selected features

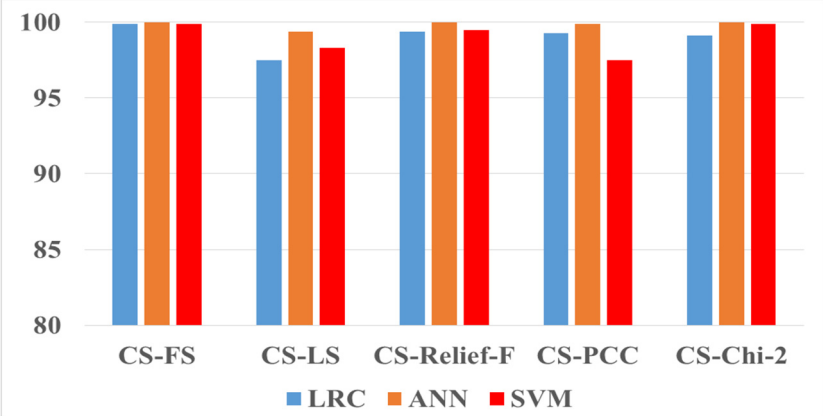


Figure. 5.2. Classification accuracy rates of 50 selected features

5.2.3 Comparisons of results

For further evaluation of the efficiency of the proposed MMV-CS and feature ranking analysis-based framework. Table 5.6 presents the comparisons with some recently published results in (Yu et al., 2018) with the same bearing dataset used in this case study. One method uses Feature Selection by

Adjunct Rand Index and Standard Deviation Ratio (FSAR) to select features from the original feature set (OFS). Other methods, use PCA, LDA, Local Fisher Discriminant Analysis (LFDA), and Support Margin LFDA (SM-LFDA) to reduce the dimension of selected features using FSAR. With the selected features, SVM is used for the purpose of classification. It is clear that all the results from our proposed framework outperforms results reported in (Yu et al., 2018).

Table 5.6. A comparison with the classification results from literature on bearing dataset G collected from the second machine.

Methods		Classification accuracy (%)
OFS-FSAR-SVM (Yu et al., 2018)		
	(25 selected features)	91.46
	(50 selected features)	69.58
OFS-FSAR-PCA-SVM (Yu et al., 2018)		
	(25 selected features)	91.67
	(50 selected features)	69.79
OFS-FSAR-LDA-SVM (Yu et al., 2018)		
	(25 selected features)	86.25
	(50 selected features)	92.7
OFS-FSAR-LFDA-SVM (Yu et al., 2018)		
	(25 selected features)	93.75
	(50 selected features)	94.38
FS-FSAR-(SM-LFDA)-SVM (Yu et al., 2018)		
	(25 selected features)	94.58
	(50 selected features)	95.63
Our proposed framework with FFT, $\alpha = 0.1$, $k = 25$		
CS-FS	LRC	98.4 ± 1.6
	ANN	99.6 ± 0.5
	SVM	97.4 ± 2.7
CS-LS	LRC	99.1 ± 0.8
	ANN	99.2 ± 0.8
	SVM	98.5 ± 1.6
CS-Relief-F	LRC	99.3 ± 0.6
	ANN	99.2 ± 0.8
	SVM	97.8 ± 2.4
CS-PCC	LRC	99.2 ± 0.8
	ANN	99.5 ± 0.6
	SVM	97.9 ± 1.9
CS-Chi-2	LRC	97.5 ± 2.6
	ANN	99.9 ± 0.1
	SVM	94.7 ± 5.4

5.3 Summary and conclusions

Two rolling bearings vibration datasets were analysed by five different techniques based on the CSFR framework to classify rolling bearings health condition. The CS-LS, CS-FS, CS-Relief-F, CS-PCC, and CS-Chi-2 methods were used to learn fewer features from a large amount of vibration data from which bearing health conditions were classified. Two different sparse representations techniques, i.e., thresholded Haar WT and FFT were used to obtain the compressible representations of the two rolling bearings vibration datasets. Moreover, three classification algorithms, LRC, ANN, and SVM were used to classify bearing health conditions. Interestingly, our validation experiments have demonstrated that the various combinations of CS and feature ranking techniques investigated in this study offer higher classification accuracies in all of the faults studied in this chapter. Having said that, CS-FR has the capability of answering two research questions with its various combinations of MMV-CS and feature ranking techniques with the following comments:

1) FFT as a sparsifying transform method for our proposed MMV based CS model can achieve better results than thresholded WT.

2) LRC and ANN have the ability to achieve high classification accuracy with different values of the sampling rate (α) and a number of selected features (k) for all the considered CS and feature selection techniques combinations.

3) SVM has the ability to achieve good classification accuracy with the larger number of selected features, i.e., k =180, and larger values of α, e.g., α = 0.4, for certain combinations.

4) With the larger number of selected features, all the proposed methods achieved high classification accuracy. Thus, for the application of the proposed methods in fault diagnosis, we recommend selecting a larger number of features from the compressively-sampled signals.

Chapter 6

Assessment and Validation of CSLSL based techniques for bearing health condition diagnosis

In this chapter, we describe various experiments that we have conducted to verify the ability of CSLSL framework, which described in Subsection 4.3, for rolling bearing health conditions diagnosis. This achieved by investigating the combination of MMV-CS and two linear feature extraction methods, namely, unsupervised PCA and supervised LDA, to extract fewer features from a large amount of vibration data. With this fewer features, LRC is used to classify bearing health conditions. Furthermore, the CS-CPDC method that described in Subsection 4.3.3, is assessed and validated for bearing health condition diagnosis. This work has been published in (Ahmed et al., 2017b; Ahmed and Nandi, 2018c; Ahmed and Nandi, 2018a)

6.1 Experimental setup

CS-PCA and CS-LDA have been applied to the first bearing vibration dataset that was described in Subsection 3.1. Half of the data is used for training and the other half is used to access the generalization of the classification performance. The classification performance evaluated using LRC with regularization parameter value ($\lambda = 0.1$) and by averaging the results of classification accuracy from ten experiments. CS-CPDC has been applied to four datasets of bearing vibration that described in chapter 3. These are, the first bearing vibration data, dataset D, dataset E, and dataset F. For the first bearing vibration data, fifty percent of the total observations were randomly selected for training and the other 50% for testing. While various training size were used for datasets D, E, and F. The classification performance evaluated using Multi-class SVM with “fitcecoc” function.

6.2 Validation of CS-PCA and CS-LDA

First, we obtain the bearing compressed vibration signals using different compressed sampling rate (α) ranging from 0.05 up to 0.4 with 256, 512, 1024 and 2048 compressed measurements of the bearings original vibration signal. These different sets of compressed measurements used for the fault diagnosis purpose. Half of this data is used for training and the other half is used to assess the generalization of classification performance. To apply our proposed framework in the vibration dataset

of the first machine, 50% of the vibration data are randomly selected for training and the other 50% are used for testing. To obtain compressively sampled signals from the original vibration signals of roller bearings, MMV-CS model with thresholded Haar WT that described in section 4.1.2. Here, we used the Haar wavelet basis with five decomposition levels as sparsifying transform where the wavelet coefficients are thresholded using the hard threshold to obtain sparse representations of the original vibration signals. Then we employed the CS framework with different sampling rates (α) (0.05, 0.1, 0.2 and 0.4) and a random matrix with i.i.d. Gaussian entries which satisfy the RIP to obtain the compressively sampled data.

For further filtering, PCA with 64 principal components and LDA with 5 discriminant components, i.e., (c-1) where c = 6 is the number of classes, were used to extract low dimensional spaces of features from the obtained CS-based compressed measurements. With these extracted features LRC with the regularization parameter ($\lambda = 0.1$) were employed to deal the classification problem. Table 6.1 summarizes the classification results for all experiments, where the classification accuracy is the average of ten trials for each experiment.

Table 6.1. Classification accuracies (%) and related standard deviations (in brackets) for compressed sensed datasets.

	Compressed Sensed data sets			
	Sampling rates (α) / Number of measurements (M)			
	0.05	**0.1**	**0.2**	**0.4**
	M=256	**M=512**	**M=1024**	**M=2048**
CS	98.2 (0.3)	98.6 (0.3)	98.9 (0.2)	99.0 (0.3)
CS-PCA	98.8 (0.7)	98.5 (0.4)	98.7 (0.6)	98.8 (0.7)
CS-LDA	72.5 (1.5)	89.8 (3.5)	100 (0.0)	100 (0.0)

It can be clearly seen from Table 6.1 that choices of small values α (0.05 and 0.1) can lead to high classification accuracies in both CS and CS-PCA, unlike CS-LDA that achieved low classification results although it obtained 100% classification accuracy for $\alpha = 0.2$ and 0.4. For further evaluation of the proposed method, we conducted several experiments using two well-known dimensionality reduction techniques, namely PCA and LDA to extract features from the raw vibration data. For the classification purpose, we used the same classifier LRC that used to obtain the results in Table 6.1. The classification results of these methods are shown in Table 6.2 and can be compared to the classification results of our proposed approach in Table 6.1 which shows the possibility to achieve high classification

performance with only a few compressed measurements comparable to the classification performance obtained using the high dimensional vibration signal.

Table 6.2. Summary of classification accuracies (%) and their related standard deviations for the raw vibration using PCA and LDA feature extraction methods.

	Feature Extraction method	
	PCA	**LDA**
Raw Vibration	99.6 ± 0.4	100 ± 0.0

Taken together, these results show the possibility to reduce the bandwidth consumption by up to 95% for remote machine condition monitoring while achieving good fault diagnosis performance comparable to fault classification performance from high dimensional vibration signal.

6.3 Validation of CS-CPDC using the first bearing vibration data

First, fifty percent of the total observations were randomly selected for training and the other 50% for testing. Then, we examined the selection of the compressed sampling rate (α) using different values (0.01, 0.02, 0.03, 0.04, 0.5, 0.1 and 0.2) to generate compressively-sampled vibration signals. To ensure that our CS model generates enough samples for the purpose of bearing fault classification, we used the generated compressively-sampled signals in the first stage to reconstruct the original signal X by applying the CoSaMP algorithm (Needell et al., 2010). For example, with $\alpha = 0.1$ the average percentage reconstruction errors for the six conditions of bearings are 1.8% (NO), 0.9% (NW), 3.3% (IR), 1.6% (OR), 0.7 % (RE), and 2.6% (CA), which indicate good signal reconstruction.

To learn features from the training set with compressed measurements, we proposed a multi-step approach, i.e., the second stage of our proposed method described in Figure 4.6 (b) using (c – 1) components for LDA and 40 principal components for PCA for each of the α values described above. These learned features from the second stage were used to train the multi-class SVM. To achieve better evaluations of the trained error-correcting output codes (ECOC) multiclass model, we applied 10-fold cross-validation in all our experiments. The training dataset is randomly subdivided into ten subsets. Each sub-set is validated on the classifier that is trained using the other nine subsets. The process is repeated 20 times and the training classification accuracy is the average taken from these 20 trials.

To evaluate the performance of the proposed method, we first compressively sampled each testing signal using the same values of α used to sample training set, and then the trained multi-step algorithm is used to obtain the learned features of the testing set. Once the features were learned, the trained SVM is used to classify the testing signals. The overall results are shown in Table 6.3, where the classification accuracy is the average of 20 trials for each experiment, and the time is obtained by

Table 6.3 Classification Results with Their Corresponding Root Mean Square Error (RMSE) and Computational time Using Various Compressed Sampling Rates

Compressed Sampling rate (α)	The cross-validated accuracy (%)	Testing Classification Accuracy (%)	Testing time (s)
α = 0.2 (1200 samples)	100.0 ± 0.0	99.9 ± 0.1	7.8± 0.01
α = 0.1 (600 samples)	100.0 ± 0.0	99.8 ± 0.2	6.7± 0.07
α = 0.05 (300 samples)	99.8 ± 0.2	99.3 ± 0.6	4.9 ± 0.03
α = 0.04 (240 samples)	99.4 ± 0.6	98.8 ± 1.2	4.3 ± 0.11
α = 0.03 (180 samples)	99.1 ± 0.8	98.4 ± 1.4	3.8 ± 0.03
α = 0.02 (120 samples)	99.1 ± 0.9	97.8 ± 1.3	3.2 ± 0.09
α = 0.01 (60 samples)	98.3 ± 1.6	96.4 ± 0.6	3.1 ± 0.16

averaging the testing time of these 20 trials. Table 6.3 shows that the value of α affects not only the classification accuracy results but also the time required by the CS-CPDC method to complete the classification task. It can be clearly seen that the larger the value of α is, the better is the classification accuracy and the more time the method requires. However, high levels of classification accuracy achieved with less than 25% of the original data samples. In particular, accuracies from our proposed method are 99.9%, 99.8%, and 99.3 for only 20%, 10%, and 5% of the whole data respectively.

6.3.1 Effect of Numbers of Principal Components on Classification Accuracy

To determine the effect of the number of PCs on classification accuracy of CS-CPDC, we tested it with α = 0.2, 0.1, 0.05, 0.04, 0.03, 0.02 and 0.01 using a different number of PCs in the range 10–50. Figure 6.1 shows the classification accuracies versus the number of PCs for each value of α. It is clear that most of the compressively-sampled signals require no more than 40 PCs to achieve high classification accuracy.

6.3.2 Comparison of Classification performance using Individual and Combined Features

In the second stage of the proposed multi-step features learning approach, four groups of features were extracted individually before the features concatenation step. These include PCA based

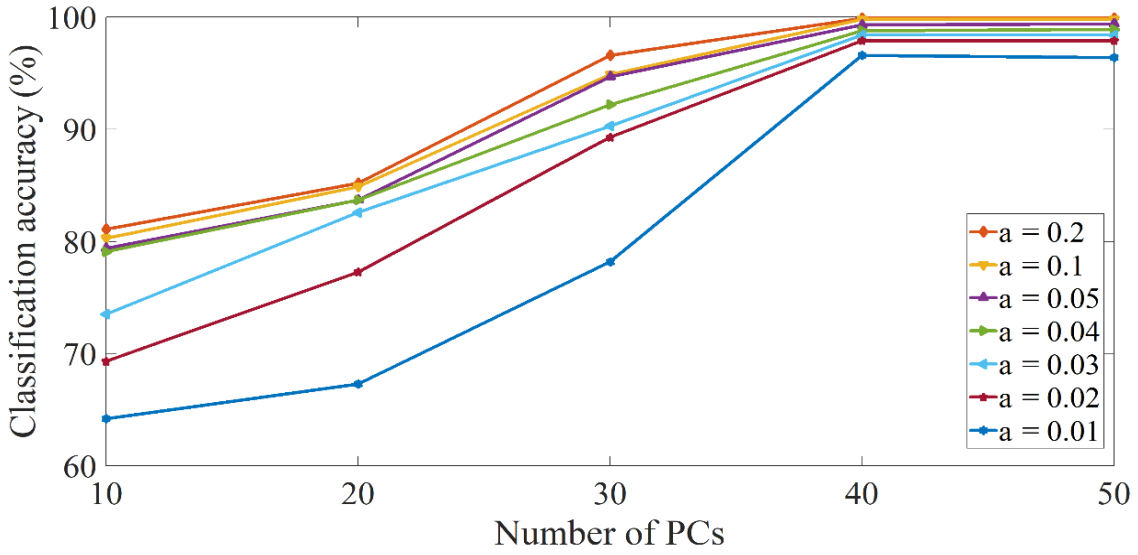


Fig. 6.1 Classification accuracies for different compressively-sampled signals versus the number of PCs

features($\hat{Y}1$), LDA based features($\hat{Y}2$), and the linear combinations features $\hat{Y}1_{CCA}$ and $\hat{Y}2_{CCA}$ of $\hat{Y}1$ and $\hat{Y}2$ these features will be referred as PCA' and LDA' respectively. Experiments are conducted using these features based on PCA, LDA, PCA', LDA' and concatenated features of PCA and LDA (PCA+LDA) with CS-CPDC to classify bearing faults. The test classification results are displayed in Fig. 5 and achieved by averaging the results of 20 trials for each experiment. Closer inspection of Figure 6.2 shows significant improvements in the classification accuracy achieved by PCA' and LDA' compared to PCA and LDA respectively. However, classification results from CS-CPDC method achieved best classification results for each value of α.

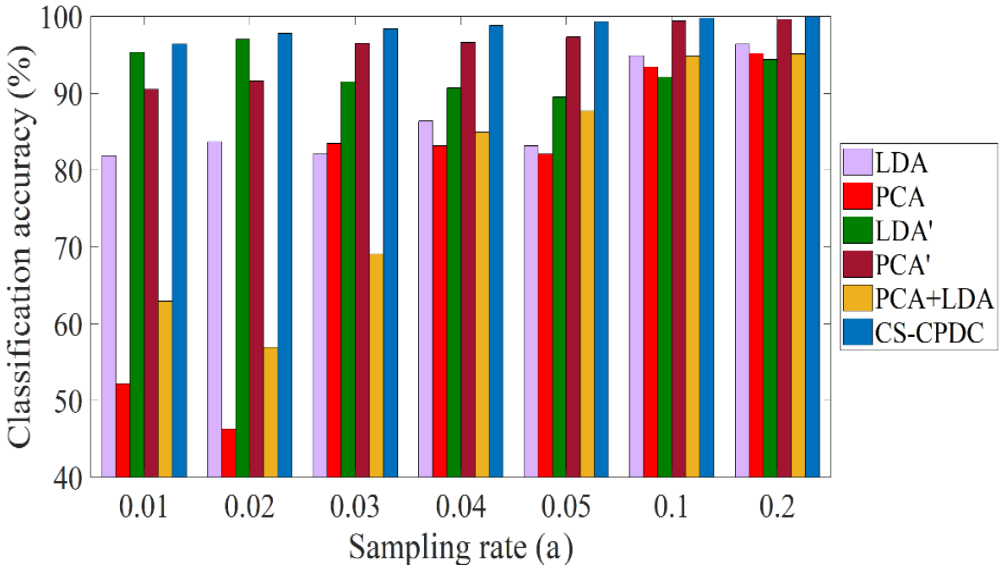


Figure 6.2 Comparison of Classification performance using Individual and Combined Features

6.3.3 Comparison of results

Different from the methods in (Wong, et al., 2015) that recovered the original signals from the compressively-sampled signals, CS-CPDC learns features directly from the compressed measurements. To show the superiority of CS-CPDC over the methods in (Wong, et al., 2015), Table 6.4 shows classification results of bearing faults using the same dataset used in the first case study of this work. It is clear that all results from CS-CPDC are better than those in (Wong, et al., 2015). For further verification of the efficiency of the CS-CPDC method, Table 6.5 presents the comparisons with the results of CS, CS-PCA, and CS-LDA presented in Table 6.1. It is clear that results from CS-CPDC are better than the results reported in (Ahmed, et al., 2017d) and the improvement is statistically significant. In particular, results from (Ahmed, et al., 2017d) are 98.6, 98.5, and 89.8 for CS, CS-PCA, and CS-LDA, while CS-CPDC method achieved 99.8%. Additionally, the RMSE of this method (0.2%) is smaller than the reported RMSEs in (Ahmed, et al., 2017d) (e.g., 0.3%, 0.4%, and 3.5%).

Table 6.4 Classification Comparison of Bearing Faults.

Methods	Sampling rate (α)	Testing Accuracy (%)
Compressed Sensed followed by signal recovery (Wong, et al., 2015)	0.5	92.4 ± 0.5
	0.25	84.6 ± 3.4
CS-CPDC	0.2	99.9 ± 0.1
	0.1	99.8 ± 0.2

Table 6.5 Classification Comparison of Bearing Faults.

Methods	Testing Accuracy (%)
	α = 0.1
CS (Ahmed, et al., 2017b)	98.6± 0.3
CS-PCA (Ahmed, et al., 2017b)	98.5± 0.4
CS-LDA (Ahmed, et al., 2017b)	89.8± 3.5
CS-CPDC	99.8± 0.2

6.3.4 Need for Compressive Sampling

In CS-CPDC, CS is employed to obtain compressively sampled signals in the first stage motivated by the following. 1) Reduced computations: We used CS to reduce a large amount of the

acquired vibration data. This reduction in the amount of vibration data resulted in much-reduced computation, i.e., less than 15% of the computation load of not using CS. 2) Reduced transmission costs: In the cases of having to send vibration data from remote places by wireless (e.g., in the case of off-shore wind turbines) or wired transmissions, the cost of transmission will be less as CS reduces the amount of vibration data. 3) Benefits to the environment: As the application of CS results in reduced computations, it helps to reduce the amount of power needed for both computations and transmission. In consequence, CS offers much benefit to the environment.

Several experiments were conducted to classify bearing fault using CS-CPDC without the compression in the first stage, i.e., with all the 6000 original samples from their sparse (Fourier Transform) domain. Table 6.6 contains the results where the classification accuracy is the average of 20 trials of testing accuracy. Two things are clear from the results presented in Table 6.6. Firstly, CS-CPDC with compressive sampling is significantly faster than (or requires less than 15% of the time of) the proposed method without compressive sampling. Secondly, CS-CPDC achieved better classification results with small RMSE. Of the remaining 0.6% (100% - 99.4%) accuracy, our method can make up 2/3 (= (99.8 – 99.4)/ (100 – 99.4)) of the missed accuracy, and it does so with significantly lower RMSE, i.e., 0.2 compared to 0.5. Thus the increase in accuracy of 0.4% with RMSE of 0.2% is practically significant.

Table 6.6. A comparison results to examine the Speed and accuracy performance of our proposed method with CS and without CS.

Methods	Classification accuracy (%)	Time (s)
CS-CPDC without CS (with 6000 inputs from FT)	99.4 ± 0.5	64.9± 0.3
CS-CPDC with CS $\alpha = 0.1$ (600 samples)	99.8± 0.2	6.7± 0.1

6.4 Validation of CS-CPDC using datasets D, E, F of the second bearing vibration data

To classify the motor bearing health condition in dataset F, the same steps section 6.1.2 were followed to apply CS-CPDC method. Half of the signal examples are selected randomly for training, and the rest of the signal examples are utilised for testing performance. Different compressed samples with 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09 and 0.1 sampling rates (α) of the original signals and 10 selected PCs are used for the overall classification results and their related root means square errors

(RMSEs) of 20 trials as shown in Table 6.7. In Table 6.7 CS-CPDC delivers high classification accuracies with small RMSEs. In particular, the classification accuracy for α = 0.1 is 99.9%, and the testing time required by CS-CPDC is only 1.62 seconds. In general, the computation time increased slightly with the increase in α value. For example, the total time for training and testing with the smallest value of α = 0.03 (5.35 s) increased by less than 20% compared with the total time required by the largest value of α = 0.1 (6.45 s).

Table 6.8 (a) and (b) shows some sample confusion matrices of ten types of health condition in the classification results for values of α = 0.1 and 0.07 respectively. It can be clearly seen that with α = 0.1 in Table 3 (a), only two signal examples of IR1 is likely to be estimated as IR3, i.e., the proposed method misclassified only 1% of testing examples of IR1 as IR3. With α = 0.07 in Table 6.8 (b), only one of IR1 (0.5 of IR1 testing examples) is likely to be confused with IR3, two of IR2 (1% of IR2 testing examples) to be classified as IR1, and five of IR3 (2.5% of testing examples) is expected to be classified as IR1.

Table 6.7. Classification results with their corresponding root Mean Square Error (RMSE) and computational time using various compressed sampling rates.

Sampling rate (α)	Training accuracy (%)	Training time (s)	Testing accuracy (%)	Testing time (s)
α = 0.1 120 samples	100	4.83 ± 0.01	99.9 ± 0.1	1.62 ± 0.02
α = 0.09 108 samples	100	4.72 ± 0.01	99.8 ± 0.2	1.55 ± 0.04
α = 0.08 96 samples	99.9 ± 0.1	4.62 ± 0.12	99.8 ± 0.1	1.47 ± 0.08
α = 0.07 84 samples	99.9 ± 0.1	4.61 ± 0.01	99.6 ± 0.2	1.41 ± 0.02
α = 0.06 72 samples	99.8 ± 0.2	4.50 ± 0.01	99.6 ± 0.2	1.32 ± 0.07
α = 0.05 60 samples	99.6 ± 0.2	4.36 ± 0.09	99.2 ± 0.3	1.24 ± 0.07
α = 0.04 48 samples	99.3 ± 0.3	4.24 ± 0.07	99.2± 0.5	1.19 ± 0.14
α = 0.03 36 samples	99.1 ± 0.4	4.20 ± 0.10	98.5± 0.8	1.15 ± 0.02

Table 6.8. Sample confusion matrix.

(a) Compressively-sampled data α = 0.1

True classes	Predicted classes										Class Prediction (%)
	NO	OR1	OR2	OR3	IR1	IR2	IR3	RE1	RE2	RE3	
NO	200	0	0	0	0	0	0	0	0	0	100
OR1	0	200	0	0	0	0	0	0	0	0	100
OR2	0	0	200	0	0	0	0	0	0	0	100
OR3	0	0	0	200	0	0	0	0	0	0	100
IR1	0	0	0	0	198	0	2	0	0	0	99
IR2	0	0	0	0	0	200	0	0	0	0	100
IR3	0	0	0	0	0	0	200	0	0	0	100
RE1	0	0	0	0	0	0	0	200	0	0	100
RE2	0	0	0	0	0	0	0	0	200	0	100
RE3	0	0	0	0	0	0	0	0	0	200	100

(b) Compressively-sampled data α = 0.07

True classes	Predicted classes										Class Prediction (%)
	NO	OR1	OR2	OR3	IR1	IR2	IR3	RE1	RE2	RE3	
NO	200	0	0	0	0	0	0	0	0	0	100
OR1	0	200	0	0	0	0	0	0	0	0	100
OR2	0	0	200	0	0	0	0	0	0	0	100
OR3	0	0	0	200	0	0	0	0	0	0	100
IR1	0	0	0	0	199	0	1	0	0	0	99.5
IR2	0	0	0	0	2	198	0	0	0	0	99
IR3	0	0	0	0	5	0	195	0	0	0	97.5
RE1	0	0	0	0	0	0	0	200	0	0	100
RE2	0	0	0	0	0	0	0	0	200	0	100
RE3	0	0	0	0	0	0	0	0	0	200	100

For further evaluation of the CS-CPDC method, experiments in datasets D, E, and F were conducted for α = 0.1 with a training size of 10% and 40%, and 20 trials for each experiment. The results are compared to some recently published results (Li et al., 2013; Jin et al., 2014; Du et al., 2014; Zhang et al., 2015; Van et al., 2015; Van et al., 2016; Xia et al., 2017; Xia et al., 2018) in Table 6.9. The first column refers to the scenarios of the motor operation and load conditions (fixed load and variable loads) in which the bearing samples collected. The second column defines the methods used for classification. The third column records the percentage of samples used to train these methods. The fourth column defines the related load of the data, and the fifth column records testing accuracies obtained using these methods.

Compared with the methods presented in Table 6.9, CS-CPDC with the smallest percentage (10%) of samples of the original data achieved the highest accuracy in both motor operation condition, i.e., fixed load and variable load.

For additional comparison, several experiments were conducted for α = 0.1, 0.09, and 0.08, using variable loads bearing dataset (0, 1, 2, 3 loads) with 10 classes to examine the speed and accuracy

performances of CS-CPDC compared to the method in (Lei et al., 2016) that used the same data. To match the experimental setup in (Lei et al., 2016), only 10% of signal examples are used for training; the testing classification accuracies and computation times were obtained by averaging 20 trials in each experiment. The results, as shown in Table 6.10, indicate that CS-CPDC method is significantly faster than the method in (Lei et al., 2016) and yet our classification accuracies, for all values of α, are better than the results reported in (Lei et al., 2016).

Table 6.9. A comparison with the results from literature on vibration datasets of roller bearings.

	Methods	**Training size (%)**	**No of classes**	**Load (hp)**	**Training Accuracy (%)**
Fixed load	(Jin et al., 2014)	10	10	3	92.5
	(Van et al., 2015)	35	10	3	92.65
	(Van et al., 2016)	N/A	10	0	97.89
	(Du et al., 2014)	75	10	0	88.9
	CS-CPDC	10	10	3	99.8 ± 0.2
				0	99.8 ± 0.1
Variable loads	(Li et al., 2013)	40	4	0,1,2,3	95.8
	(Xia et al., 2017)	75	9	0,1,2,3	97.59
	(Zhang et al., 2015)	40	11	0,1,2,3	97.91±0.09
	(Xia et al., 2018)	70	10	0,1,2,3	99.44
	CS-CPDC	40	10	0,1,2,3	99.9 ± 0.1

Table 6.10. A comparison results to examine the speed and accuracy performances of CS-CPDC.

Methods	**Testing accuracy (%)**	**Time (s)**
(Lei et al., 2016)	99.66 ± 0.19	12
CS-CPDC		
α = 0.1	99.8 ± 0.1	5.28
α = 0.09	99.8 ± 0.1	4.92
α = 0.08	99.7 ± 0.2	4.74

6.5 Summary and conclusions

Our validation experiments have demonstrated that choices of small values α (0.05 and 0.1) can lead to high classification accuracies in CS-PCA, unlike CS-LDA that achieved low classification results although it obtained 100% classification accuracy for α = 0.2 and 0.4. However, the classification accuracy obtained using the CS-CPDC method with 10% of the original data is better than the results obtained by CS-PCA and CS-LDA and the improvement is statistically significant.

The comparison of non-CS-based techniques such as in (Li et al., 2013; Jin et al., 2014; Du et al., 2014; Van et al., 2015; Zhang et al., 2015; Van et al., 2016; Xia et al., 2017&2018) with the proposed CS-CPDC technique show that the learned features of CS-CPDC achieved better classification results, even though we are using only 10% ((α = 0.1) of the original vibration data.

Chapter 7

CS-SAE based DNN technique Assessment and Validation

This chapter presents a number of experiments that have been conducted to validate the efficiency of CS-SAE-DNN technique, which described in Section 4.4, for rolling bearing health conditions diagnosis. This achieved by investigating the application of CS-SAE-DNN in four vibration datasets collected for the purpose of bearing health condition monitoring. These are the first bearing vibration data and datasets A, B, and C of the second bearing vibration data described in chapter 3. In these experiments, we explored computationally, the effects of SAE based over-complete sparse representations on the classification performance of CS-based highly compressed measurements of bearing vibration signals. For this study, the CS framework was used to produce highly compressed measurements of the original bearing dataset. Then, an effective DNN with unsupervised feature learning based on SAE is used for learning over-complete sparse representations of the compressed measurements. Finally, the fault classification is achieved using two stages, namely, pre-training classification based on stacked SAE and softmax regression layer, and re-training classification based on the BP algorithm. The experiments results show that the proposed method is able to achieve high levels of accuracy even with extremely compressed measurements. This work has been published in (Ahmed, et al., 2018).

7.1 Experimental setup

CS-SAE-DNN has been applied to four datasets of bearing vibration that described in chapter 3. These are, the first bearing vibration data, dataset A, dataset B, and dataset C. To obtain the compressively-sampled signals, first, we used Haar wavelet basis with five decomposition levels as a sparsifying transform. Then we applied compressive sampling with different sampling rates (α) (0.0016, 0.003, 0.006, 0.013, 0.025, 0.05, 0.1 and 0.2) with 8, 16, 32, 64, 128, 256, 512 and 1024 compressed measurements of our original vibration signal using random Gaussian matrix. The size of the Gaussian matrix is m by N, where N is the length of the original vibration signal measurements and m is the number of compressed signal elements (i.e., $m = \alpha * N$). Fifty percent of these compressed samples are randomly selected for the pre-training stage of DNN, then these samples are used to re-train the deep net, the other 50% of samples are used for testing the performance.

7.2 The first bearing vibration data

In order to verify the validity of the proposed method, we carried out several experiments to learn over-complete features of various highly compressed bearing data sets obtained using CS framework with different compressed sampling rate. Fifty percent of these compressed samples are randomly selected for the pre-training stage of DNN, then these samples are used to re-train the deep net, the other 50% of samples are used for testing the performance. Then, the obtained over-complete features used for classification using different settings of DNN. Finally, we compare our proposed method, using our highly compressed datasets, with several existing methods.

We began by obtaining the compressed vibration signal from the big data of rolling elements bearing vibration signal using thresholded Haar WT, which described in Subsection 4.1.2, and random Gaussian matrix. First, we used the Haar wavelet basis with five decomposition levels as sparsifying transform where the wavelet coefficients are thresholded using the penalized hard threshold to obtain sparse representations of the original vibration signals. As shown in Figure 7.1(a), the wavelet coefficients of the vibration. After applying the penalized hard threshold the wavelet coefficients are sparse in the Haar wavelet domain as shown in Figure 7.1(b) where only 216 are non-zeros (nnz) in NO wavelet coefficients that is 95.8% of the 5120 coefficients are zeros. Other conditions NW, IR, OR, RE and CA have 276, 209, 298, 199, and 299 non-zero elements respectively; and 94.6%, 95.9, 94.2%, 96.1 and 94.2% of 5120 coefficients are zeros.

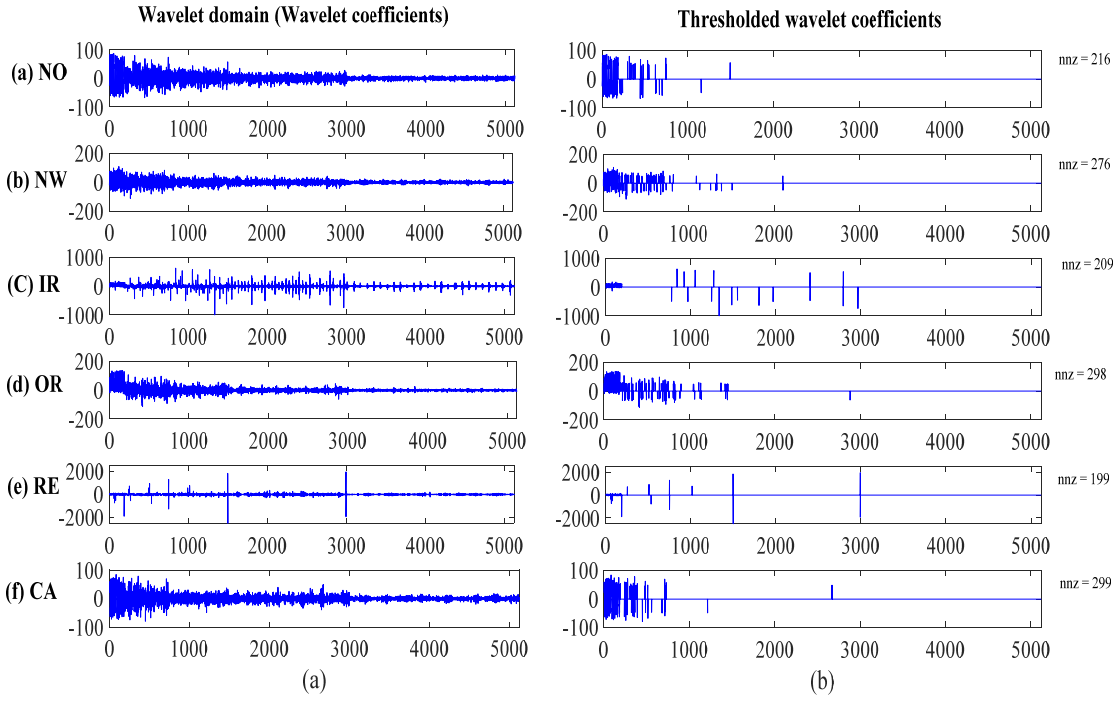


Figure 7.1 (a) Wavelet coefficients and (b) corresponding thresholded wavelet coefficients for each condition signal (nnz refers to number of non-zero elements).

Then we applied compressive sampling with different sampling rates (α) (0.0016, 0.003, 0.006, 0.013, 0.025, 0.05 and 0.1) with 8, 16, 32, 64, 128, 256, and 512 compressed measurements of our original vibration signal using random Gaussian matrix. The size of the Gaussian matrix is *m* by *N*, where *N* is the length of the original vibration signal measurements and m is the number of compressed signal elements (i.e., $m = \alpha * N$). Based on compressive sampling framework, multiplying this matrix with our signal sparse representations generates different sets of compressed measurements of the vibration signal. The obtained compressed measurements must possess the quality of the original signal, i.e., have sufficient information of the original signal. Thus, we need to test that our CS model generates enough samples for the purpose of bearings fault classification. Roman et al. proposed a generalized flip test (Roman, et al., 2014) for CS model that has the ability to test the efficiency of any sparsity model, any signal class, any sampling operator, and any recovery algorithm towards accurate CS model. The basic idea of this test is to flip sparsity basis coefficients which represent the sparse representations and then perform a reconstruction with measurements matrix using the sampling operator and the recovery algorithm. If the sparse vector (s) is not recovered within a low tolerance, then decrease the thresholding level to obtain a sparse signal and repeat until s recovered exactly. More details of the original flip test can be found in (Roman, et al., 2014; Adcock, et al., 2017).

Following the idea of the generalized flip test, we tried different sampling rates (0.05, 0.1, 0.15, and 0.2) and tested the efficiency of our CS model by thresholding the wavelet coefficients to obtain a sparse signal (s) and then by reconstructing s from the obtained compressed measurements using random Gaussian matrix. The CoSaMP algorithm is used to reconstruct the sparse signal. The reconstruction errors measured by Root Mean Squared Error (RMSE) for the six conditions of bearings are presented in the following Table 7.1. The second column depicts the reconstruction errors compared to the original thresholded coefficients using 5% of the original signal for the six conditions, the third column shows the reconstructions errors using 10% of the original signal, the fourth is for 15% of the original signal and the fifth column is for 20% of the original signal.

Table 7.1. Results of root-mean-square-error (RMSE) for various sampling rates.

	α = 0.05	**α = 0.10**	**α = 0.15**	**α = 0.20**
NO	9.37	8.45	0.04	0.03
NW	11.29	24.64	5.8	0.06
IR	20.82	15.23	0.16	0.09
OR	14.38	12,71	8.18	0.07
RE	17.02	11.87	0.35	0.12
CA	8.3	5.29	0.14	0.04

It is clear that as α increases RMSE decreases, indicating better signal reconstruction. The better signal reconstruction indicates that the compressed measurements possess the quality of the original signal.

In order to verify the validity of the proposed method, we carried out several experiments to learn over-complete features of various highly compressed bearing datasets obtained by using different compressed sampling rate. Fifty percent of these compressed samples are randomly selected for the pre-training stage of DNN, then these samples are used to re-train the deep net, the other 50% of samples are used for testing the performance. Then, the obtained over-complete features used for classification using different settings of DNN. Finally, we compare our proposed method, using our highly compressed datasets, with several existing methods.

To learn features from these compressed measurements, by using one layer in the encoder, we achieved poor classification results. Then, by using multilayer as in our proposed method we have obtained better results. Hence, we used a sparse autoencoder neural network with a limited number of hidden layers (2, 3, and 4 hidden layers). The structures of these different hidden layers are chosen to be in the form of over-complete feature learning (expansion), where the number of neurons in different hidden layers network structure is twice the number of neurons in preceding layer, for example, if the number of input samples in the input layer is z then the number of nodes in the first hidden layer is 2z and 4z in the second hidden layer and so on. The number of nodes of the output layer is limited by the number of bearing conditions (6 conditions). A Bi-directional deep architecture of stacked autoencoders has been used for the purpose of deep learning, these include feedforward and backpropagation (BP). The parameters that control the effects of using regularizers by sparse autoencoder were set as follows: the weight decay (λ) was set to very small 0.002, the weight of the sparsity penalty term (β) was set to be 4, and the sparsity parameter (ρ) to 0.1. The maximum training epoch is 200.

7.2.1 Results

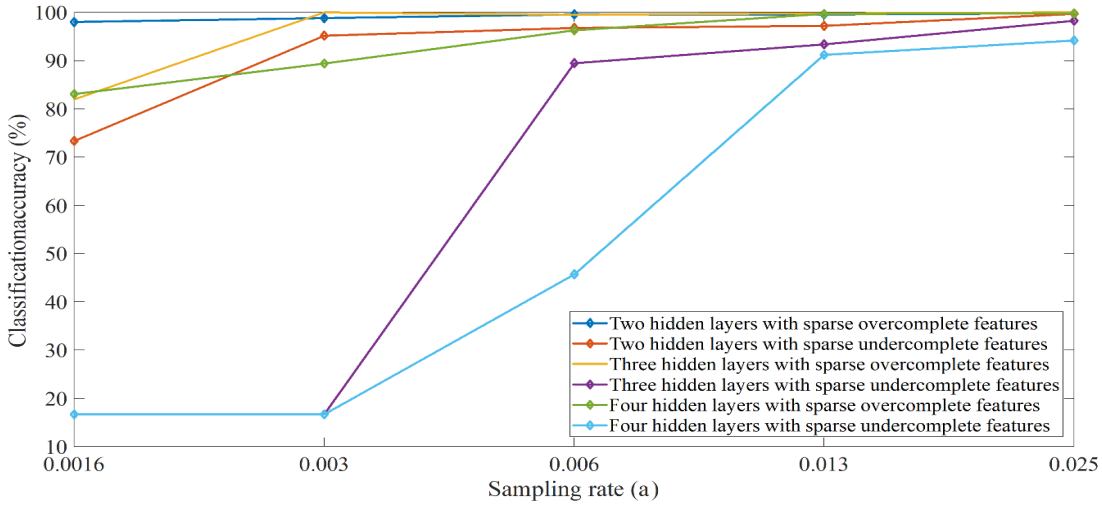
The previous sections have described the principle structure of our proposed method and the experimental setup. Various experiments were conducted to use our method for classifying bearing faults from different highly compressed vibration measurements. The overall classification results obtained from these experiments are shown in Table 7.2. The first major comment is that the classification accuracy after the second stage is better than that after the first stage for every dataset at these values of α. The classification in the deep net stage (the first stage) achieved good results for larger numbers of measurements, i.e., for values of *m* equal to 512, 256 and 128, and high accuracy was achieved by the two hidden layers DNN using only 64 samples of our signal. Most of the classification accuracies for the two, three and four hidden layers DNNs using fine-tuning stage (the second stage) are 99% or above and some are 100% for even less than 1% compressed measurements of the original vibration signal, i.e., when α = 0.006. The two hidden layers of DNN achieved high classification

accuracy (98%) for α equal to 0.003 and 0.0016 with 16 and 8 compressed measurements. Moreover, the three hidden layers of DNN achieved 100% with only 16 measurements, i.e., α=0.003. Taken together, these results show that the proposed method has the ability to classify the bearing conditions with a high accuracy from highly compressed bearing vibration measurements.

Table 7.2. Overall classification accuracies and their related standard deviations.

		Compressive Sampling rate						
Number of hidden layers	**Classification stage**	α = 0.0016 (m = 8)	α = 0.003 (m = 16)	α = 0.006 (m = 32)	α = 0.013 (m = 64)	α = 0.025 (m = 128)	α = 0.05 (m = 256)	α = 0.1 (m = 512)
2	Deep net (the first stage)	95.3 ± 1.0	95.9 ± 0.3	96.0 ± 1.3	99.1 ± 0.3	99.8 ± 0.1	100	99.8 ± 0.1
	Fine – tuning (the second stage)	98.0 ± 0.3	98.8 ± 0.3	99.6 ± 0.1	99.6 ± 0.1	99.8 ± 0.1	100	99.8 ± 0.1
3	Deep net (the first stage)	64.8 ± 4.2	94.9 ± 2.4	85.6 ± 4.4	93.6 ± 2.2	99.2 ± 2.6	99.6 ± 0.1	100
	Fine – tuning (the second stage)	82.0 ± 5.1	100	99.5 ± 0.3	99.8 ± 0.1	100	100	100
4	Deep net (the first stage)	36.6 ± 8.3	55.2 ± 6.1	90.5 ± 2.9	95.6 ± 1.2	99.8 ± 0.1	98.6 ± 3.9	100
	Fine – tuning (the second stage)	83.1 ± 1.6	89.4 ± 2.5	96.3 ± 0.8	99.7 ± 0.3	99.8 ± 0.1	99.7 ± 0.3	100

In comparison with DNN based on sparse autoencoder using under-complete representations, i.e., when the number of nodes in each hidden layer is less than the number of input samples. Classification results from several experiments using under-complete feature representations that also deal with the same highly compressed data sets are compared to the results acquired using over-complete representations as shown in Figure 7.2. From this figure, it can be clearly seen that all the scenarios of DNNs that used over-complete feature representations outperform all those utilizing under-complete sparse features when the input samples are extremely compressed datasets (i.e., with only 8 and 16 compressed measurements). Evidently, the two hidden layers DNNs in all scenarios achieved better results than other network structures, i.e., three hidden layers DNN and four hidden layers DNN.



representations with two, three and four hidden layers DNN.

In order to measure the training performance of the proposed over-complete feature based DNN compared to under-complete representations based DNN, a typical value of α = 0.025 and two hidden layers DNN has been used in this comparison. Figure 7.3 shows that minimum Mean Squared Error (MSE) value which is 0.003 was achieved with the training of three hidden layers DNN using over-complete features in epoch 200, compared to the same DNN structure using under-complete feature representations where MSE was 0.021.

For further verification of the performance of the proposed method, complete comparison results of the classification accuracy using the proposed technique compared with three classifiers, namely, logistic regression classifier (LRC), support vector machine (SVM) and neural network (NN) were used to classify faults from the same highly compressed measurements sets and the complete results are shown in Table 7.3. It is clear that the results from our proposed method with a smaller sampling rate as in α = 0.0016, 0.003, and 0.006 are better than those achieved by other classifiers.

Table 7.3. Complete classification results and their related standard deviations using LRC, SVM, NN and the proposed method.

	Compressive Sampling rate						
Classification method	α = 0.0016 (m = 8)	α =0.003 (m = 16)	α = 0.006 (m = 32)	α = 0.013 (m = 64)	α = 0.025 (m = 128)	α = 0.05 (m = 256)	α = 0.1 (m = 512)
LRC	64.2 ± 7.9	92.2 ± 5.5	96.2 ± 0.6	98.2 ± 0.5	99.0 ± 0.5	99.0 ± 0.3	99.3 ± 0.3
SVM	77.7 ± 19.3	94.3 ± 2.3	97.8 ± 0.4	98.4 ± 0.4	99.5 ± 0.3	99.7 ± 0.2	99.7 ± 0.2
ANN	60.4 ± 10.5	76.7 ± 6.2	98.0 ± 1.9	98.8 ± 0.3	99.6 ± 0.2	99.7 ± 0.2	99.7 ± 0.2
***proposed method**	98.0 ± 0.3	98.8 ± 0.3	99.6 ± 0.1	99.6 ± 0.1	99.8 ± 0.1	100	99.8 ± 0.1

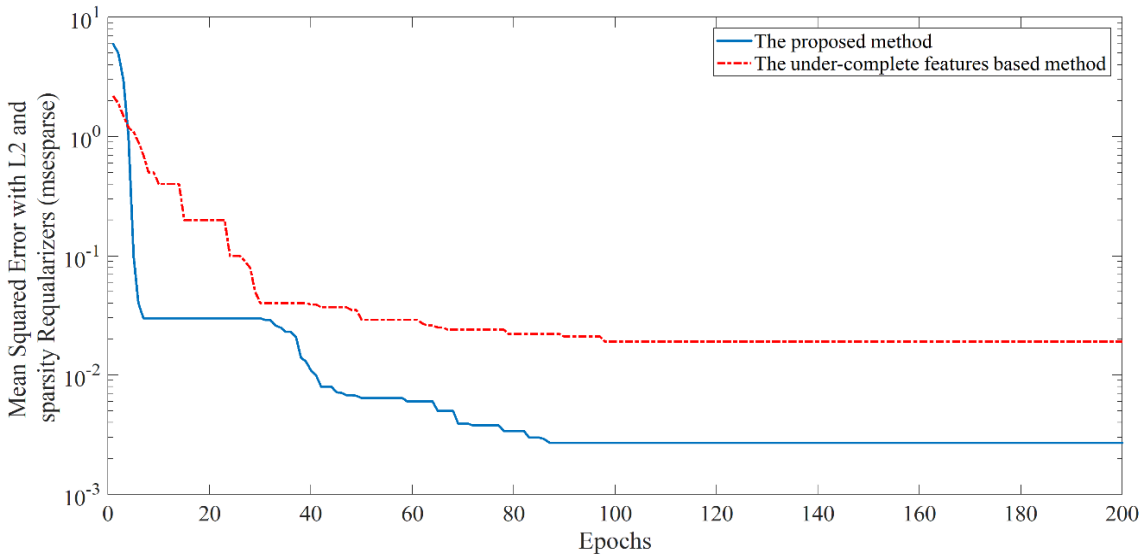


Figure 7.3 Training performance of over-complete feature based two hidden layers DNN and under complete feature based two hidden layers DNN (α = 0.025).

7.2.2 Effects of parameterization on the classification accuracy

To control the effects of parameterization on SAEs, some parameters need to be set, these include, sparsity parameter (ρ), weight decay (λ) and the weight of the sparsity penalty term (β). To test the influence of these parameters values on bearing fault classification performance, several experiments have been carried out using our proposed method with two hidden layers, and different values of SAE parameterization. The sampling rate α was set to 0.05 where two hidden layers achieved 100% classification accuracy for both classification stages, i.e., deep net (the first stage) and fine-tuning (the second stage) with the parameters values described in section 7.2. As can be seen from Figure 7.4, while the classification accuracy is sensitive to the value of λ, there is a very broad range of values for ρ and β for which classification accuracies are very high and stable.

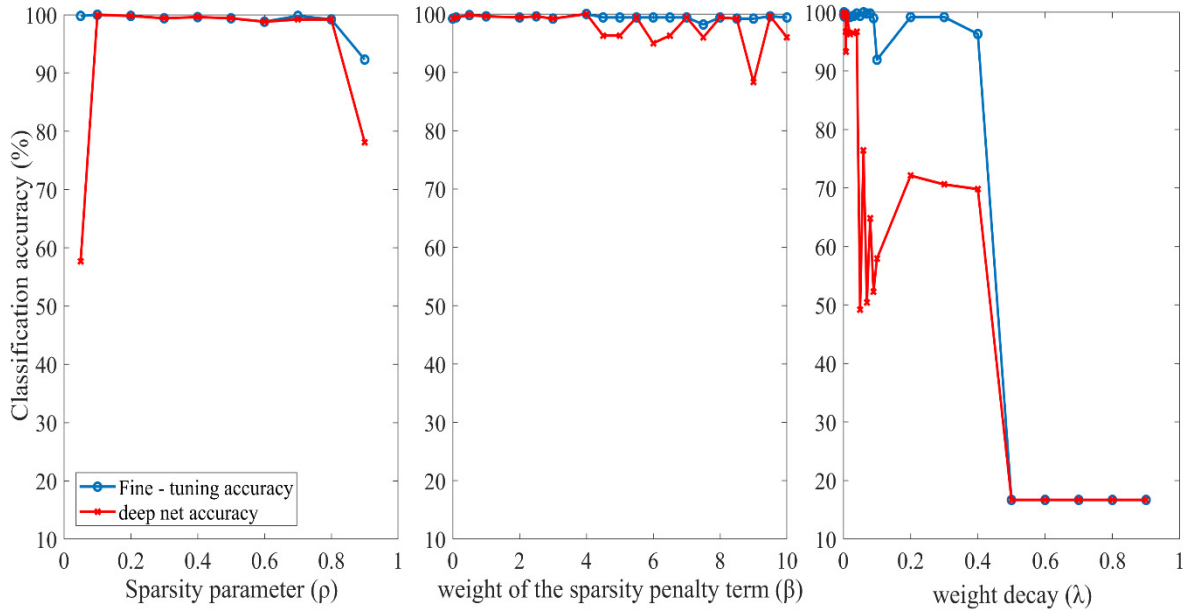


Figure. 7.4. Effects of parameterization on the classification accuracy.

7.2.3 Comparison of results

In this subsection, a comparison of several methods using the same vibration dataset as in (Wong, et al., 2015). One method uses all the original samples. Each of the two other methods uses compressed measurements (for values of α of 0.5 and 0.25) and then reconstruct the original signals. These three have been reported in (Wong, et al., 2015). The remaining three are our proposed methods to demonstrate the possibility to sample the vibration data of roller element bearings at less than Nyquist rate using CS and to perform fault classification without reconstructing the original signal. Table 7.4 shows classification results of bearing faults using our proposed method with two hidden layers and a sampling rate α of 0.5, 0.25 and 0.1 and the reported results in (Wong, et al., 2015) using the same dataset. It is clear that all our results are better than those in (Wong, et al., 2015).

Table 7.4. A comparison with the classification results from literature on bearing dataset

		Accuracy (%)
Raw Vibration (Wong et al., 2015)		98.9 ± 1.2
Compressed Sensed (α = 0.5) (Wong et al., 2015) followed by reconstruction		92.4 ± 0.5
Compressed Sensed (α = 0.25) (Wong et al., 2015)followed by reconstruction		84.6 ± 3.4
This work (α = 0.5) (two hidden layers)	Deep net (the first stage)	99.1 ± 1.7
	Fine – tuning (the second stage)	100 ± 0.0
This work (α = 0.25) (two hidden layers)	Deep net (the first stage)	99.6 ± 1.2
	Fine – tuning (the second stage)	100 ± 0.0
This work (α = 0.1) (two hidden layers)	Deep net (the first stage)	99.8 ± 0.1
	Fine – tuning (the second stage)	99.8 ± 0.1

7.3 Datasets A, B, and C: The second bearing vibration data

We applied the same data processing steps that described in Section 7.2 to each dataset (i.e., A, B and C) to obtain compressed vibration signals with different sampling rates (α) (as 0.025, 0.05 and 0.1, and 0.2) with 60, 120, 240 and 480 compressed measurements of A, B and C original vibration signals. Fifty percent of these compressed samples are randomly selected for the pre-training stage of DNN, then these samples are used to re-train the deep net, the other 50% of samples are used for testing the performance.

7.3.1. Results

The proposed method with two hidden layers is used to process the compressed measurements of each dataset. The classification accuracy rates are obtained by averaging the results of ten experiments for each compressed datasets obtained using the different sampling rates. The average accuracies and their corresponding standard deviations of ten experiments for each dataset are shown in Table 7.5. One of the more significant findings to emerge from the results in Table 7.5 is that classification results after the fine-tuning stage (the second stage) is better than that after the first stage for all datasets A, B, and C with different values of α. Also, it shows that the deep net stage (the first stage) achieved good results with 99.6% and 99.5% fora= 0.2 with datasets B and C respectively. Most of the classification accuracy results after the second stage are above 99% for values of α in the range of 0.05 to 0.2. In particular, results after the second stage of our proposed method for datasets B and C with α equal to 0.2 achieved 100% accuracy for every single run in our investigations, and also the

100% accuracy achieved for data C for α equal to 0.1. Overall, these results indicate that the proposed method has the ability to classify the bearing conditions with high accuracy from highly compressed vibration measurements.

Table 7.5. Classification results for bearing datasets A, B, and C of the second machine.

		Compressive Sampling rate			
Datasets	**Classification stage**	α = 0.025 (m = 60)	α = 0.05 (m = 120)	α = 0.1 (m = 240)	α = 0.2 (m = 480)
A	Deep net (the first stage)	89.7 ± 5.4	93.7 ± 2.9	92.5 ± 2.2	98.7 ± 0.8
	Fine – tuning (the second stage)	97.6 ± 0.7	98..4 ± 1.3	99.3 ± 0.6	99.8 ± 0.2
B	Deep net (the first stage)	94.5 ± 1.1	96.2 ± 1.6	98.2 ± 0.6	99.6 ± 0.6
	Fine – tuning (the second stage)	98.9 ± 0.2	99.3 ± 0.4	99.7 ± 0.5	100 ± 0.0
C	Deep net (the first stage)	97.3 ± 0.8	98.5 ± 1.4	98.2 ± 0.4	99.5 ± 0.2
	Fine – tuning (the second stage)	99.4 ± 0.6	99.7 ± 0.3	100 ± 0.0	100 ± 0.0

7.3.2 Comparison of results

To evaluate the effectiveness of our proposed method, Table 7.6 presents the comparisons with some recently published results in (De et al., 2015, and Jia et al., 2016) with the same roller bearing datasets A, B and C. The second left-hand column presents the classification results of DNN based method in (Jia et al., 2016) while the third column shows the classification results of back propagation neural network (BPNN) based method in (Jia et al., 2016). In (De et al., 2015) a generic multi-layer perceptron (MLP) was used for the classification purpose.

Table 7.6. A comparison with the results from literature on A, B, and C bearing vibration datasets of the second bearing vibration data.

Method / Dataset	(Jia et al., 2016) DNN	(Jia et al., 2016) BPNN	(De et al., 2016) MLP	Our proposed method with fine-tuning (the second stage) with 2 hidden layers α = 0.1
A	99.95 ± 0.06	65.20 ± 18.09	95.7	99.3 ± 0.6
B	99.61 ± 0.21	61.95 ± 22.09	99.6	99.7 ± 0.5
C	99.74 ± 0.16	69.82 ± 17.67	99.4	100 ± 0.0

It is clear that the results from our proposed methods with fine-tuning (second stage) are very competitive. In particular, results from our fine-tuning method with dataset C achieved 100% accuracy for every single run in our investigations, even though we are using a limited amount (only 10%) of the original data, which are not matched by any of the other methods using 100% of the data.

For further verification of the efficiency of the proposed method, we conducted three experiments (all with 2 hidden layers and fine-tuning) to examine the speed and accuracy performances in several scenarios. The results are presented in Table 7.7. The first column refers to the three datasets. The second and third columns describe accuracies and execution times of using a "traditional" autoencoder based DNN of (Jia et al., 2016) with 2400 inputs from Haar Wavelet (with no CS), while the fourth and fifth columns describe accuracies and execution times of using our sparse autoencoder based DNN with 2400 inputs from Haar Wavelet (with no CS). Two things are clear for each of the three datasets – 1) our autoencoder based DNN (even without CS) is much faster than (or requires only 80% of the time of) the "traditional" autoencoder based DNN, and 2) yet our classification results are very competitive.

Table 7.7. A comparison results to examine the speed and accuracy performances in several scenarios

Dataset	WT-DNN with Fine-tuning				Our proposed method with Fine-tuning (the second stage)	
	Using DNN implementation of (Jia et al., 2016) with 2 hidden layers (with no CS)		Using DNN implementation in this paper with 2 hidden layers (with no CS)		WT - CS- DNN (2 hidden layers)	
	Accuracy (%)	Time (mins)	Accuracy (%)	Time (mins)	Accuracy (%)	Time (mins)
A	99.0 ± 0.1	41.5 ± 0.3	99.4 ± 0.5	34.1 ± 0.7	99.3 ± 0.6	5.7 ± 0.1
B	99.1 ± 0.6	43.3 ± 0.7	99.5 ± 0.8	32.9 ± 0.3	99.7 ± 0.5	5.9 ± 0.4
C	99.6 ± 0.3	43.1 ± 0.2	99.8 ± 1.1	34.2 ± 0.9	100	5.7 ± 0.2

The sixth and seventh columns describe accuracies and execution times of using our proposed sparse autoencoder based DNN with 240 inputs from Haar Wavelet (with CS). Two things are clear for each of the three datasets – (1) our autoencoder based DNN (even with CS) is significantly faster than (or requires only 15% of the time of) the "traditional" autoencoder based DNN, and (2) yet our classification results are as good as, if not better than, the other two scenarios.

In summary, the significant reduction in computation time comes from two sources – (1) using our proposed sparse autoencoder and (2) using CS. Finally, our complete proposal (using sparse autoencoder and CS) achieves classification results for all three datasets that are as good as, if not better than, the other two scenarios.

7.4 Summary and conclusions

In this investigation, the aim was to assess the classification of bearing faults from highly compressed measurements based on CS. The proposed method includes the extraction of over-complete sparse representations from highly compressed measurements. It involves the unsupervised feature learning algorithm SAE for learning feature representations in multi-stages of non-linear feature

transformation based on DNN. The accuracy of the proposed method is verified using highly compressed datasets of rolling element bearings signals obtained using different compressed sampling rates. These compressed datasets contain fewer samples for each bearing condition. The most obvious finding to emerge from this study is that, despite achieving fairly high classification accuracy in the first stage, the proposed method is able to achieve higher classification accuracy in the second stage even from highly compressed measurements compared to the existing methods. Moreover, classification results from our proposed method outperform those achieved by reconstructing the original signals. Additionally, a significant reduction in computation time is achieved using our proposed method compared to another autoencoder based DNN method in (Jia et al., 2016), with better classification accuracies. The implication of this is the possibility that the proposed method of compressive sensing in machine faults classification will require fewer measurements thus it would reduce the computational complexity, storage requirement and the bandwidth for transmitting reduced data.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

There is a growing body of literature that identifies various methods for bearing fault detection and classification. However, the performance of these methods is limited by the massive amounts of data need to be collected for Machine Condition Monitoring (MCM). It is clear that acquiring a large amount of data requires large storage and time for signal processing and this also may limit the number of machines that can be monitored remotely across wireless sensor networks (WSNs) due to bandwidth and power constraints.

This thesis includes developments in the vibration-based computational analysis of a large amount of acquired vibration dataset for condition monitoring of rolling bearings. Aiming to address the challenges of learning from high dimensional data and embrace the efficiency of CS in improving rolling bearings condition monitoring, **chapter 4 (Methodology)** presents novel methods, namely the compressive sampling and feature ranking (CSFR) framework based methods, the compressive sampling and subspace learning (CSLSL) based methods, and the compressive sampling and sparse autoencoder based deep neural network (CS-SAE-DNN) method. These methods represent the original contribution of this thesis in designing new methods for vibration-based rolling bearing condition monitoring.

Chapter 5 (CSFR framework Assessment and Validation) presents several experiments that we have conducted to verify the capability of the CSFR framework for rolling bearing health condition diagnosis. This achieved by investigating the combination of MMV-CS and several feature ranking techniques – i.e., FS, LS, Relief-F, PCC, and Chi-2 – to learn fewer features from a large amount of vibration data from which bearing health conditions are classified . Three classification algorithms, LRC, ANN, and SVM are used to classify bearing health conditions. Two case studies of bearing health conditions diagnosis are used in this investigation with the aims to (1) validate the proposed framework for bearing health conditions diagnosis, and (2) observe the best combinations of MMV-CS, feature selection techniques, and classifiers with reduced complexity and improved classification accuracy.

Our validation experiments have demonstrated that our proposed framework, CSFR, has the ability to achieve high classification accuracy in all of the faults studied in this chapter.

Comparisons of CS-based vibration signal recovery for machine fault diagnosis method in (Wong et al., 2015) and non-CS-based techniques in (Guo et al., 2005), (Seera et al., 2017) and (Yu et al., 2018) show that the learned features of the proposed CSFR-based techniques achieved better classification results although we are using only 10% ($\alpha = 0.1$) of the original vibration data that not matched by the other method mentioned above.

Chapter 6 (Assessment and Validation of CSLSL based techniques for bearing health condition diagnosis) describes various experiments that we have conducted to verify the ability of CSLSL framework for rolling bearing health conditions diagnosis. This achieved by investigating the combination of MMV-CS and two linear feature extraction methods, namely, unsupervised PCA and supervised LDA, to extract fewer features from a large amount of vibration data. With this fewer features, LRC is used to classify bearing health conditions. Furthermore, the CS-CPDC method is assessed and validated for bearing health condition diagnosis.

Our validation experiments have demonstrated that choices of small values α (0.05 and 0.1) can lead to high classification accuracies in CS-PCA, unlike CS-LDA that achieved low classification results although it obtained 100% classification accuracy for $\alpha = 0.2$ and 0.4. However, the classification accuracy obtained using the CS-CPDC method with 10% of the original data is better than the results obtained by CS-PCA and CS-LDA and the improvement is statistically significant.

The comparison of non-CS-based techniques such as in (Li et al., 2013; Jin et al., 2014; Du et al., 2014; Van et al., 2015; Zhang et al., 2015; Van et al., 2016; Xia et al., 2017&2018) with the proposed CS-CPDC technique show that the learned features of CS-CPDC achieved better classification results, even though we are using only 10% ($\alpha = 0.1$) of the original vibration data.

Chapter 7 (CS-SAE based DNN technique Assessment and Validation) presents a number of experiments that have been conducted to validate the efficiency of CS-SAE-DNN technique, which is described in Section 4.4, for rolling bearing health condition diagnosis. This was achieved by the application of CS-SAE-DNN in four vibration datasets collected for the purpose of bearing health condition monitoring. These are the first bearing vibration data and datasets A, B, and C of the second bearing vibration data described in chapter 3. In these experiments, we explored computationally, the effects of SAE based over-complete sparse representations on the classification performance of CS-based highly compressed measurements of bearing vibration signals.

The most obvious finding to emerge from our validation experiments is that, despite achieving fairly high classification accuracy in the first stage, the proposed method is able to achieve even higher classification accuracy in the second stage from highly compressed measurements compared to the existing methods. In particular, most of the classification accuracies for the two, three and four hidden layers DNNs using fine-tuning stage (the second stage) are 99% or above and some are 100% for even less than 1% compressed measurements of the original vibration signal, i.e., when $\alpha = 0.006$. The two

hidden layers of DNN achieved high classification accuracy (98%) for α equal to 0.003 and 0.0016 with 16 and 8 compressed measurements. This provides strong evidence for the advantage of CS-SAE-DNN technique in cases where high levels of data compression are required to address the challenges of learning from high dimensional data.

Moreover, classification results from our proposed method outperform those achieved by reconstructing the original signals. Additionally, a significant reduction in computation time is achieved using our proposed method compared to another autoencoder based DNN method (Jia et al., 2016), with better classification accuracies.

Taken together, these results show that the proposed methods have the ability to classify bearing health conditions with a high classification accuracy using fewer measurements. The implication of this is the possibility that the proposed methods of compressive sampling in machine faults classification will require fewer measurements thus it would reduce the computational complexity, storage requirement and the bandwidth for transmitting reduced data.

8.2 Future work

These research findings provide the following insights for future work:

- If the research is to be moved forward, a better understanding of CS in machine fault diagnosis needs to be developed. This would be a fruitful area for further work of developing CS with respect to classification.
- With respect to the CSFR framework to bearing vibration dataset, we investigate the combination of CS and several feature ranking techniques to reduce a large amount of bearing vibration signals and select fewer representative features for fault classification. Three classifiers are then used to produce the final results of bearing health condition. In spite of that, the CSFR framework has the capability to make use of existing feature selection and classification methods in addition to the new developing methods. Hence, other alternative sparse representations, feature selection, and classification algorithms need to be investigated.
- In 2009, Baraniuk and Wakin demonstrated that random linear projections can efficiently preserve the structure of manifold (Baraniuk and Wakin, 2009). Hence, further exploration of the use of the other nonlinear variants of PCA and LDA like kernel-PCA (KPCA) and other manifold learning techniques in combination with CS may be beneficial to be examined.
- Investigation and experimentation into other alternatives methods for over-complete representations in combination with CS are strongly recommended.
- The future work may also include the application of the proposed method to fault diagnosis for other rotating machineries like induction motors, turbines, and gearboxes.

- This research provides a successful study of automatic fault diagnosis of rolling bearings using CS and learning algorithms. The proposed methods in this thesis and other methods that will be investigated in the future can be integrated into a complete toolbox for fault diagnosis, which will benefit the practitioners and researchers involved in the field of vibration-based machine condition monitoring.

Appendix I

Bibliography

Abe, S. 2005. Support vector machines for pattern classification (Vol. 2). London: Springer.

Adcock, B., Hansen, A.C., Poon, C. and Roman, B., 2017. Breaking the coherence barrier: A new theory for compressed sensing. In *Forum of Mathematics, Sigma* (Vol. 5). Cambridge University Press. arXiv: 1302.0561, 2014.

Ahamed, N., Pandya, Y. and Parey, A., 2014. Spur gear tooth root crack detection using time synchronous averaging under fluctuating speed. *Measurement*, *52*, pp.1-11.

Ahmad, R., & Kamaruddin, S. 2012. An overview of time-based and condition-based maintenance in industrial application. Computers & Industrial Engineering, 63(1), pp. 135-149.

Ali, J.B., Fnaiech, N., Saidi, L., Chebel-Morello, B. and Fnaiech, F., 2015. Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Applied Acoustics*, *89*, pp.16-27.

Antoni, J. and Randall, R.B., 2006. The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. *Mechanical systems and signal processing*, *20*(2), pp.308-331.

Antoni, J., 2007. Fast computation of the kurtogram for the detection of transient faults. *Mechanical Systems and Signal Processing*, *21*(1), pp.108-124.

Ayaz, E., 2012. Autoregressive modeling based bearing fault detection in induction motors. In *IX Symposium Industrial Electronics INDEL, Banja Luka*.

Ayaz, E., 2014. 1315. Autoregressive modeling approach of vibration data for bearing fault diagnosis in electric motors. *Journal of Vibroengineering*, *16*(5), pp. 2130 – 2138.

Baillie, D.C. and Mathew, J., 1996. A comparison of autoregressive modeling techniques for fault diagnosis of rolling element bearings. *Mechanical Systems and Signal Processing*, *10*(1), pp.1-17.

Balakrishnama, S. and Ganapathiraju, A., 1998. Linear discriminant analysis-a brief tutorial. *Institute for Signal and Information Processing*, *18*, pp.1-8.

Bao, Y., Beck, J.L. and Li, H., 2011. Compressive sampling for accelerometer signals in structural health monitoring. *Structural Health Monitoring*, *10*(3), pp.235-246.

Baraniuk, R., Davenport, M., DeVore, R., & Wakin, M. 2008. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, *28*(3), pp. 253-263.

Baraniuk, R.G., and Wakin, M.B., 2009. Random projections of smooth manifolds. *Foundations of computational mathematics*, *9*(1), pp.51-77.

Bechhoefer, E. and Kingsley, M., 2009. A review of time synchronous average algorithms. In *Annual Conference of the Prognostics and Health Management Society, San Diego, CA, Sept* (pp. 24-33).

Benesty, J. and Huang, Y. eds., 2013. *Adaptive signal processing: applications to real-world problems*. Springer Science & Business Media.

Bengio, Y., Lamblin, P., Popovici, D. and Larochelle, H., 2007. Greedy layer-wise training of deep networks. *Advances in neural information processing systems* (pp. 153-160).

Bengio, Y., Lamblin, P., Popovici, D. and Larochelle, H., 2007. Greedy layer-wise training of deep networks. In *Advances in neural information processing systems* (pp. 153-160).

Bloch, H. P., & Geitner, F. K. 2012. Machinery failure analysis and troubleshooting: practical machinery management for process plants. Gulf Professional Publishing Butterworth-Heinemann.

Box, G.E., Jenkins, G.M., Reinsel, G.C., and Ljung, G.M., 2015. *Time series analysis: forecasting and control*. John Wiley & Sons.

Bozorg-Haddad, O., Solgi, M. and LoÃ, H.A., 2017. *Meta-heuristic and evolutionary algorithms for engineering optimization* (Vol. 294). John Wiley & Sons.

Braun, S., 1975. The extraction of periodic waveforms by time domain averaging. *Acta Acustica United with Acustica*, *32*(2), pp.69-77.

Brereton, R. G., & Lloyd, G. R. 2010. Support vector machines for classification and regression. Analyst, 135(2), 230-267.

Burrus, C.S., Gopinath, R.A., Guo, H., Odegard, J.E. and Selesnick, I.W., 1998. *Introduction to wavelets and wavelet transforms a primer* (Vol. 1). New Jersey: Prentice Hall.

Caesarendra, W., Widodo, A., and Yang, B.S., 2010. Application of relevance vector machine and logistic regression for machine degradation assessment. *Mechanical Systems and Signal Processing*, *24*(4), pp.1161-1171.

Candes, E. J., & Tao, T. 2006. Near-optimal signal recovery from random projections: Universal encoding strategies. *IEEE transactions on information theory*, *52*(12), pp. 5406-5425.

Candès, E. J., & Wakin, M. B. 2008. An introduction to compressive sampling. *IEEE signal processing magazine*, *25*(2), pp. 21-30.

Chandrashekar, G., & Sahin, F. 2014. A survey on feature selection methods. Computers & Electrical Engineering, 40(1), pp. 16-28.

Chang, S.G., Yu, B. and Vetterli, M., 2000. Adaptive wavelet thresholding for image denoising and compression. *IEEE transactions on image processing*, *9*(9), pp.1532-1546.

Chang, Y. and Jiao, W., 2012. ICA-ANN method in fault diagnosis of rotating machinery. In *Computer Science and Automation Engineering (CSAE), 2012 IEEE International Conference on* (Vol. 3, pp. 236-240). IEEE.

Chaturvedi, G.K. and Thomas, D.W., 1982. Bearing fault detection using adaptive noise cancelling. *Journal of Mechanical Design*, *104*(2), pp.280-289.

Chebil, J., Hrairi, M. and Abushikhah, N., 2011. Signal analysis of vibration measurements for condition monitoring of bearings. *Australian Journal of Basic and Applied Sciences*, *5*(1), pp.70-78.

Chen, C.Z., Meng, Q., Zhou, H. and Zhang, Y., 2012. Rolling bearing fault diagnosis based on blind source separation. In *Applied Mechanics and Materials* (Vol. 217, pp. 2546-2549). Trans Tech Publications.

Chen, J., & Huo, X. 2006. Theoretical results on sparse representations of multiple-measurement vectors. *IEEE Transactions on Signal Processing*, *54*(12), pp. 4634-4643.

Chen, X., Du, Z., Li, J., Li, X., & Zhang, H. 2014. Compressed sensing based on dictionary learning for extracting impulse components. *Signal Processing*, *96*, pp. 94-109.

Chen, Z. and Mechefske, C.K., 2002. Diagnosis of machinery fault status using transient vibration signal parameters. *Modal Analysis*, *8*(3), pp.321-335.

Cheng, J., Yang, Y. and Yang, Y., 2012. A rotating machinery fault diagnosis method based on local mean decomposition. *Digital Signal Processing*, *22*(2), pp.356-366.

Christian, K., Mureithi, N., Lakis, A. and Thomas, M., 2007.On the use of time synchronous averaging, independent component analysis and support vector machines for bearing fault diagnosis, *Proceedings of the 1st International Conference on Industrial Risk Engineering*, Montreal, December 2007, pp. 610–624.

Ciabattoni, L., Cimini, G., Ferracuti, F., Freddi, A., Ippoliti, G., and Monteriu, A., 2015. A novel LDA-based approach for motor bearing fault detection. In *Industrial Informatics (INDIN), 2015 IEEE 13th International Conference on* (pp. 771-776). IEEE.

Cochran, W.T., Cooley, J.W., Favin, D.L., Helms, H.D., Kaenel, R.A., Lang, W.W., Maling, G.C., Nelson, D.E., Rader, C.M. and Welch, P.D., 1967. What is the fast Fourier transform? *Proceedings of the IEEE*, *55*(10), pp.1664-1674.

Collins, J. A., Busby, H. R., & Staab, G. H. 2010. Mechanical design of machine elements and machines: a failure prevention perspective. John Wiley & Sons.

Combet, F. and Gelman, L., 2007. An automated methodology for performing time synchronous averaging of a gearbox signal without speed sensor. *Mechanical systems and signal processing*, *21*(6), pp.2590-2606.

Comon, P., 1994. Independent component analysis, a new concept. *Signal processing*, *36*(3), pp.287-314.

Cooley, J. W., & Tukey, J. W. 1965. An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation*, *19*(90), 297-301.

Cortes, C., & Vapnik, V. 1995. Support-vector networks. Machine learning, 20(3), pp. 273-297.

Cozorici, I., Vadan, I. and Balan, H., 2012. Condition based monitoring and diagnosis of rotating electrical machines bearings using FFT and wavelet analysis. *Acta Electrotehnica*, *53*(4), pp.350-354.

Dai, W., & Milenkovic, O. 2009. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Transactions on Information Theory*, 55(5), pp. 2230-2249.

Davenport, M. A., Wakin, M. B., & Baraniuk, R. G. 2006. Detection and estimation with compressive measurements. *Dept. of ECE, Rice University, Tech. Rep*.

Davies, A. 1998. Visual inspection systems. In Handbook of Condition Monitoring. Springer*, Dordrecht*, pp. 57-77.

De Almeida, L. F., Bizarria, J. W., Bizarria, F. C., & Mathias, M. H. 201). Condition-based monitoring system for rolling element bearing using a generic multi-layer perceptron. Journal of Vibration and Control, 21(16), 3456-3464.

Devijver, P.A. and Kittler, J., 1982. *Pattern recognition: A statistical approach*. Prentice Hall.

Diniz, P.S., Da Silva, E.A. and Netto, S.L., 2010. *Digital signal processing: system analysis and design*. Cambridge University Press.

Djebala, A., Ouelaa, N. and Hamzaoui, N., 2008. Detection of rolling bearing defects using discrete wavelet analysis. *Meccanica*, *43*(3), pp.339-348.

Doi, E., Balcan, D.C. and Lewicki, M.S., 2006. A theoretical analysis of robust coding over noisy overcomplete channels. *Advances in neural information processing systems* (pp. 307-314).

Donoho, D. L. 2006. Compressed sensing. IEEE Transactions on information theory, 52(4), pp. 1289-1306.

Du, W., Tao, J., Li, Y. and Liu, C., 2014. Wavelet leaders multifractal features based fault diagnosis of rotating mechanism. *Mechanical Systems and Signal Processing*, *43*(1-2), pp.57-75.

Duda, R.O., Hart, P.E. and Stork, D.G., 2001. Pattern classification second edition John Wiley & Sons. *New York*, *58*, p.16.

Duque-Perez, O., Del Pozo-Gallego, C., Morinigo-Sotelo, D. and Godoy, W.F., 2017, August. Bearing fault diagnosis based on Lasso regularization method. In *Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED), 2017 IEEE 11th International Symposium on* (pp. 331-337). IEEE.

Dybała, J. and Zimroz, R., 2014. Rolling bearing diagnosing method based on empirical mode decomposition of machine vibration signal. *Applied Acoustics*, *77*, pp.195-203.

Elasha, F., Mba, D. and Ruiz-Carcel, C., 2016. A comparative study of adaptive filters in detecting a naturally degraded bearing within a gearbox. *Case Studies in Mechanical Systems and Signal Processing*, *3*, pp.1-8.

Eldar, Y. C. 2015. Sampling theory: Beyond bandlimited systems. Cambridge University Press.

Erhan, D., Bengio, Y., Courville, A., Manzagol, P.A., Vincent, P. and Bengio, S., 2010. Why does unsupervised pre-training help deep learning? *Journal of Machine Learning Research*, *11*(Feb), pp.625-660.

Fan, W., Cai, G., Zhu, Z.K., Shen, C., Huang, W. and Shang, L., 2015. Sparse representation of transients in wavelet basis and its application in gearbox fault feature extraction. *Mechanical Systems and Signal Processing*, *56*, pp.230-245.

Fuan, W., Hongkai, J., Haidong, S., Wenjing, D. and Shuaipeng, W., 2017. An adaptive deep convolutional neural network for rolling bearing fault diagnosis. *Measurement Science and Technology*, *28*(9), p.095005.

Gelle, G., Colas, M. and Servière, C., 2003. Blind source separation: A new pre-processing tool for rotating machines monitoring? *IEEE Transactions on Instrumentation and Measurement*, *52*(3), pp.790-795.

GU, Q., Li, Z. and Han, J., 2012. Generalized Fisher score for feature selection. *arXiv preprint arXiv: 1202.3725*.

Gu, Q., Li, Z., & Han, J. 2011. Generalized Fisher score for feature selection. In Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence, pp. 266-273.

Guo, H., Jack, L.B. and Nandi, A.K., 2005. Feature generation using genetic programming with application to fault classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *35*(1), pp.89-99.

Guo, H., Zhang, Q. and Nandi, A.K., 2008. Feature extraction and dimensionality reduction by genetic programming based on the Fisher criterion. *Expert Systems*, *25*(5), pp.444-459.

Guo, X., Chen, L. and Shen, C., 2016. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Measurement*, *93*, pp.490-502.

Ha, J.M., Youn, B.D., Oh, H., Han, B., Jung, Y. and Park, J., 2016. Autocorrelation-based time synchronous averaging for condition monitoring of planetary gearboxes in wind turbines. *Mechanical Systems and Signal Processing*, *70*, pp.161-175.

Hardoon, D.R., Szedmak, S. and Shawe-Taylor, J., 2004. Canonical correlation analysis: An overview with application to learning methods. *Neural Computation*, *16*(12), pp.2639-2664.

Haroun, S., Seghir, A.N. and Touati, S., 2017. Feature selection for enhancement of bearing fault detection and diagnosis based on self-organizing map. *Recent advances in electrical engineering and control applications* (pp. 233-246). Springer, Cham.

Harris, T. A. 2001. Rolling bearing analysis. John Wiley and sons.

He, X., Cai, D., & Niyogi, P. 2006. Laplacian score for feature selection. Advances in neural information processing systems, pp. 507-514.

Higgs, P. A., Parkin, R., Jackson, M., Al-Habaibeh, A., Zorriassatine, F., & Coy, J. 2004. A survey on condition monitoring systems in industry. In *ASME 7th Biennial Conference on Engineering Systems Design and Analysis*, pp. 163-178.

Holland, D. J., Malioutov, D. M., Blake, A., Sederman, A. J., & Gladden, L. F. 2010. Reducing data acquisition times in phase-encoded velocity imaging using compressed sensing. *Journal of magnetic resonance*, 203(2), pp. 236-246.

Hong, H. and Liang, M., 2007. Separation of fault features from a single-channel mechanical signal mixture using wavelet decomposition. *Mechanical Systems and Signal Processing*, *21*(5), pp.2025-2040.

Hosmer Jr, D.W., Lemeshow, S. and Sturdivant, R.X., 2013. *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Hsu, C. W., & Lin, C. J. 2002. A comparison of methods for multiclass support vector machines. IEEE Transactions on Neural Networks, 13(2), pp. 415-425.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C. and Liu, H.H., 1998, March. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. In *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences* (Vol. 454, No. 1971, pp. 903-995). The Royal Society.

Immovilli, F., Bellini, A., Rubini, R., & Tassoni, C. 2010. Diagnosis of bearing faults in induction machines by vibration or current signals: A critical comparison. *IEEE Transactions on Industry Applications*, *46*(4), 1350-1359.

Immovilli, F., Cocconcelli, M., Bellini, A. and Rubini, R., 2009. Detection of generalized-roughness bearing fault by spectral-kurtosis energy of vibration or current signals. *IEEE Transactions on Industrial Electronics*, *56*(11), pp.4710-4717.

Islam, M.R., Islam, M.M. and Kim, J.M., 2016. Feature selection techniques for increasing reliability of fault diagnosis of bearings. In *Electrical and Computer Engineering (ICECE), 2016 9th International Conference on* (pp. 396-399). IEEE.

Jack, L.B. and Nandi, A.K., 2000. Genetic algorithms for feature selection in machine condition monitoring with vibration signals. *IEE Proceedings-Vision, Image and Signal Processing*, *147*(3), pp.205-212.

Jack, L.B. and Nandi, A.K., 2000. Genetic algorithms for feature selection in machine condition monitoring with vibration signals. *IEE Proceedings-Vision, Image and Signal Processing*, *147*(3), pp.205-212.

Jack, L.B. and Nandi, A.K., 2001. Support vector machines for detection and characterization of rolling element bearing faults. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *215*(9), pp.1065-1074.

Jack, L.B. and Nandi, A.K., 2002. Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms. *Mechanical systems and signal processing*, *16*(2-3), pp.373-390.

Jack, L.B., Nandi, A.K. and McCormick, A.C., 1999. Diagnosis of rolling element bearing faults using radial basis function networks. *Applied signal processing*, *6*(1), pp.25-32.

Jacop, A., Khang, H.V., Robbersmyr, K.G. and Cardoso, A.J.M., 2017, August. Bearing fault detection for drivetrains using adaptive filters based wavelet transform. In *Electrical Machines and Systems (ICEMS), 2017 20th International Conference on* (pp. 1-6). IEEE.

Jardine, A. K., Lin, D., & Banjevic, D. 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, *20*(7), pp. 1483-1510.

Jia, F., Lei, Y., Lin, J., Zhou, X. and Lu, N., 2016. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing*, *72*, pp.303-315.

Jiang, H., Li, C. and Li, H., 2013. An improved EEMD with multiwavelet packet for rotating machinery multi-fault diagnosis. *Mechanical Systems and Signal Processing*, *36*(2), pp.225-239.

Jiang, H., Li, X., Shao, H. and Zhao, K., 2018. Intelligent fault diagnosis of rolling bearings using an improved deep recurrent neural network. *Measurement Science and Technology*, *29*(6), p.065107.

Jiang, L., Liu, Z., Cui, J. and Li, Z., 2013, Condition monitoring of rolling element bearing based on Phase-PCA. In *Control and Decision Conference (CCDC), 2013 25th Chinese* (pp. 1082-1086). IEEE.

Jiang, Y.L. and Shao, Y.X., 2014. Fault diagnosis of rolling bearing based on fuzzy neural network and chaos theory. *Vibroengineering PROCEDIA*, *4*, pp.211-216.

Jin, X., Zhao, M., Chow, T. W., & Pecht, M. 2014. Motor bearing fault diagnosis using trace ratio linear discriminant analysis. IEEE Transactions on Industrial Electronics, 61(5), pp. 2441-2451.

Junsheng, C., Dejie, Y. and Yu, Y., 2006. A fault diagnosis approach for roller bearings based on EMD method and AR model. *Mechanical Systems and Signal Processing*, *20*(2), pp.350-362.

Kankar, P.K., Sharma, S.C. and Harsha, S.P., 2011. Fault diagnosis of ball bearings using machine learning methods. *Expert Systems with Applications*, *38*(3), pp.1876-1886.

Kankar, P.K., Sharma, S.C. and Harsha, S.P., 2011. Fault diagnosis of ball bearings using continuous wavelet transform. *Applied Soft Computing*, *11*(2), pp.2300-2312.

Kanović, Ž, Rapaić, M.R. and Jeličić, Z.D., 2011. Generalized particle swarm optimization algorithm-Theoretical and empirical analysis with application in fault detection. *Applied Mathematics and Computation*, *217*(24), pp.10175-10186.

Kappaganthu, K. and Nataraj, C., 2011. Feature selection for fault detection in rolling element bearings using mutual information. *Journal of vibration and acoustics*, *133*(6), p.061001.

Khan, S. and Yairi, T., 2018. A review on the application of deep learning in system health management. *Mechanical Systems and Signal Processing*, *107*, pp.241-265.

Kim, J., Ahn, Y., & Yeo, H. 2016. A comparative study of time-based maintenance and condition-based maintenance for optimal choice of maintenance policy. *Structure and Infrastructure Engineering*, *12*(12), pp. 1525-1536.

Kohavi, R. and John, G.H., 1997. Wrappers for feature subset selection. *Artificial intelligence*, *97*(1-2), pp.273-324.

Kononenko, I., Šimec, E. and Robnik-Šikonja, M., 1997. Overcoming the myopia of inductive learning algorithms with RELIEFF. *Applied Intelligence*, *7*(1), pp.39-55.

Konstantin-Hansen, H., 2003. Envelope analysis for diagnostics of local faults in rolling element bearings. *Bruel & Kjaer*, pp.1-8.

Kumar, R. and Singh, M., 2013. Outer race defect width measurement in taper roller bearing using discrete wavelet transform of vibration signal. *Measurement*, *46*(1), pp.537-545.

Lacey, S. J. 2008. An overview of bearing vibration analysis. *Maintenance & asset management*, *23*(6), pp. 32-42.

Lakshmi Pratyusha, P., Shanmukha Priya, V. and Naidu, V.P.S., 2014. Bearing health condition monitoring: time domain analysis. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, pp.75-82.

Lange, S. and Riedmiller, M.A., 2010, July. Deep auto-encoder neural networks in reinforcement learning. In *IJCNN* (pp. 1-8).

Lee, H., Ekanadham, C., and Ng, A.Y., 2008. Sparse deep belief net model for visual area V2. In *Advances in neural information processing systems* (pp. 873-880).

Lei, Y., Jia, F., Lin, J., Xing, S., & Ding, S. X. 2016. An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. IEEE Transactions on Industrial Electronics, 63(5), pp. 3137-3147.

Lei, Y., Lin, J., He, Z. and Zi, Y., 2011. Application of an improved kurtogram method for fault diagnosis of rolling element bearings. *Mechanical Systems and Signal Processing*, *25*(5), pp.1738-1749.

Lewicki, M.S., and Sejnowski, T.J., 2000. Learning overcomplete representations. *Neural Computation*, *12*(2), pp.337-365.

Li, F., Meng, G., Ye, L. and Chen, P., 2008. Wavelet transform-based higher-order statistics for fault diagnosis in rolling element bearings. *Journal of Vibration and Control*, *14*(11), pp.1691-1709.

Li, F.C., Ye, L., Zhang, G.C. and Meng, G., 2007. Bearing fault detection using higher-order statistics based ARMA model. *Key engineering materials* (Vol. 347, pp. 271-276). Trans Tech Publications.

Li, H. and Zhang, Y., 2006, June. Bearing faults diagnosis based on EMD and Wigner-Ville distribution. In *Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on* (Vol. 2, pp. 5447-5451). IEEE.

Li, H., Fu, L., and Zheng, H., 2011. Bearing fault diagnosis based on amplitude and phase map of Hermitian wavelet transform. *Journal of Mechanical Science and Technology*, *25*(11), pp.2731-2740.

Li, H., Zheng, H. and Tang, L., 2006, September. Wigner-Ville distribution based on EMD for faults diagnosis of bearing. In *International Conference on Fuzzy Systems and Knowledge Discovery* (pp. 803-812). Springer, Berlin, Heidelberg.

Li, L., Qu, L. and Liao, X., 2007. Haar wavelet for machine fault diagnosis. *Mechanical Systems and Signal Processing*, *21*(4), pp.1773-1786.

Li, W., Qiu, M., Zhu, Z., Wu, B. and Zhou, G., 2016. Bearing fault diagnosis based on spectrum images of vibration signals. *Measurement Science and Technology*, *27*(3), p.035005.

Li, W., Zhang, S. and He, G., 2013. Semisupervised distance-preserving self-organizing map for machine-defect detection and classification. *IEEE Transactions on Instrumentation and Measurement*, *62*(5), pp.869-879.

Li, X. F., Fan, X. C., & Jia, L. M. 2012. Compressed sensing technology applied to fault diagnosis of train rolling bearing. *Applied Mechanics and Materials*, Vol. 226, pp. 2056-2061.

Li, X., Zhang, X., Li, C. and Zhang, L., 2013. Rolling element bearing fault detection using support vector machine with improved ant colony optimization. *Measurement*, *46*(8), pp.2726-2734.

Li, Y., Xu, M., Wang, R. and Huang, W., 2016. A fault diagnosis scheme for rolling bearing based on local mean decomposition and improved multiscale fuzzy entropy. *Journal of Sound and Vibration*, *360*, pp.277-299.

Li, Y., Zhang, W., Xiong, Q., Luo, D., Mei, G. and Zhang, T., 2017. A rolling bearing fault diagnosis strategy based on improved multiscale permutation entropy and least squares SVM. *Journal of Mechanical Science and Technology*, *31*(6), pp.2711-2722.

LI, Z. and FU, Y., 1990. Adaptive noise cancelling technique and bearing fault diagnosis. *Journal of Aerospace Power*, *5*, pp.199-203.

Lin, H.C., Ye, Y.C., Huang, B.J. and Su, J.L., 2016. Bearing vibration detection and analysis using an enhanced fast Fourier transform algorithm. *Advances in Mechanical Engineering*, *8*(10), p.1687814016675080.

Lin, J. and Qu, L., 2000. Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis. *Journal of sound and vibration*, *234*(1), pp.135-148.

Linsuo, S., Yazhou, Z. and Wenpeng, M., 2011. Application of Wigner-Ville-Distribution-Based Spectral Kurtosis Algorithm to Fault Diagnosis of Rolling Bearing [J]. *Journal of Vibration, Measurement & Diagnosis*, *1*, p.010.

Liu, C., Wang, G., Xie, Q. and Zhang, Y., 2014. Vibration sensor-based bearing fault diagnosis using ellipsoid-ARTMAP and differential evolution algorithms. *Sensors*, *14*(6), pp.10598-10618.

Liu, H. and Han, M., 2014. A fault diagnosis method based on local mean decomposition and multi-scale entropy for roller bearings. *Mechanism and Machine Theory*, *75*, pp.67-78.

Liu, H., & Motoda, H. (Eds.). 2007. Computational methods of feature selection. CRC Press.

Liu, H., Li, L., and Ma, J., 2016. Rolling bearing fault diagnosis based on STFT-deep learning and sound signals. *Shock and Vibration*, *2016*.

Liu, H., Liu, C. and Huang, Y., 2011. Adaptive feature extraction using sparse coding for machinery fault diagnosis. *Mechanical Systems and Signal Processing*, *25*(2), pp.558-574.

Liu, H., Wang, X. and Lu, C., 2014. Rolling bearing fault diagnosis under variable conditions using Hilbert-Huang transform and singular value decomposition. *Mathematical Problems in Engineering*, *2014*.

Liu, Z., Cao, H., Chen, X., He, Z. and Shen, Z., 2013. Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings. *Neurocomputing*, *99*, pp.399-410.

Logan, D. and Mathew, J., 1996. Using the correlation dimension for vibration fault diagnosis of rolling element bearings—I. Basic concepts. *Mechanical Systems and Signal Processing*, *10*(3), pp.241-250.

Lou, X. and Loparo, K.A., 2004. Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mechanical systems and signal processing*, *18*(5), pp.1077-1095.

Lu, B., Nowak, M., Grubic, S. and Habetler, T.G., 2009, September. An adaptive noise-cancellation method for detecting generalized roughness bearing faults under dynamic load conditions. In *Energy Conversion Congress and Exposition, 2009. ECCE 2009. IEEE* (pp. 1091-1097). IEEE.

Lu, C., Wang, Z. and Zhou, B., 2017. Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network-based health state classification. *Advanced Engineering Informatics*, *32*, pp.139-151.

Luo, G.Y., Osypiw, D. and Irle, M., 2003. On-line vibration analysis with fast continuous wavelet algorithm for condition monitoring of bearing. *Modal Analysis*, *9*(8), pp.931-947.

Ma, M., Chen, X., Wang, S., Liu, Y. and Li, W., 2016, August. Bearing degradation assessment based on Weibull distribution and deep belief network. In *Flexible Automation (ISFA), International Symposium on* (pp. 382-385). IEEE.

Malhi, A. and Gao, R.X., 2004. PCA-based feature selection scheme for machine defect classification. *IEEE Transactions on Instrumentation and Measurement*, *53*(6), pp.1517-1525.

McClintic, K., Lebold, M., Maynard, K., Byington, C., & Campbell, R. 2000. Residual and difference feature analysis with transitional gearbox data. In *54th Meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, VA,* pp. 1-4.

McClintic, K., Lebold, M., Maynard, K., Byington, C., & Campbell, R. 2000. Residual and difference feature analysis with transitional gearbox data. In *54th Meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, VA,* pp. 1-4.

McCormick, A.C. and Nandi, A.K., 1996, September. A comparison of artificial neural networks and other statistical methods for rotating machine condition classification. In *COLLOQUIUM DIGEST-IEE* (pp. 2-2). IEE INSTITUTION OF ELECTRICAL ENGINEERS.

McCormick, A.C. and Nandi, A.K., 1997, June. Neural network autoregressive modeling of vibrations for condition monitoring of rotating shafts. In *Neural Networks, 1997. International Conference on* (Vol. 4, pp. 2214-2218). IEEE.

McCormick, A.C. and Nandi, A.K., 1997. Classification of the rotating machine condition using artificial neural networks. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *211*(6), pp.439-450.

McCormick, A.C. and Nandi, A.K., 1999. Bispectral and trispectral features for machine condition diagnosis. *IEE Proceedings-Vision, Image and Signal Processing*, *146*(5), pp.229-234.

McCormick, A.C., Nandi, A.K. and Jack, L., 1998. Digital signal processing algorithms in condition monitoring. *International Journal of COMADEM*, *1*, pp.5-14.

McCormick, A.C., Nandi, A.K. and Jack, L.B., 1998. Application of periodic time-varying autoregressive models to the detection of bearing faults. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *212*(6), pp.417-428.

McCormick, A.C., Nandi, A.K. and Jack, L.B., 1998. Application of periodic time-varying autoregressive models to the detection of bearing faults. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *212*(6), pp.417-428.

McFadden, P. D., & Smith, J. D. 1985. The vibration produced by multiple point defects in a rolling element bearing. Journal of sound and vibration, 98(2), pp. 263-273.

McFadden, P.D. and Toozhy, M.M., 2000. Application of synchronous averaging to vibration monitoring of rolling element bearings. *Mechanical Systems and Signal Processing*, *14*(6), pp.891-906.

McMillan, D., & Ault, G. W. 2007. Quantification of condition monitoring benefit for offshore wind turbines. *Wind Engineering*, *31*(4), pp. 267-285.

Merlet, S. L., & Deriche, R. 2013. Continuous diffusion signal, EAP and ODF estimation via compressive sensing in diffusion MRI. *Medical image analysis*, *17*(5), pp. 556-572.

Møller, M.F., 1993. A scaled conjugate gradient algorithm for fast supervised learning. *Neural networks*, *6*(4), pp.525-533.

Nandi, A. K., Liu, C., & Wong, M. D. 2013. Intelligent vibration signal processing for condition monitoring. *In Proceedings of the International Conference Surveillance*, Vol. 7, pp. 1-15.

Nandi, A.K. and Jack, L.B., 2004. Advanced digital vibration signal processing for condition monitoring. *International Journal of COMADEM*, *7*(1), pp.3-12.

Nandi, S., Toliyat, H. A., & Li, X. 2005. Condition monitoring and fault diagnosis of electrical motors—A review. IEEE transactions on energy conversion, 20(4), pp. 719-729.

Nayana, B.R. and Geethanjali, P., 2017. Analysis of Statistical Time-Domain Features Effectiveness in Identification of Bearing Faults From Vibration Signal. *IEEE Sensors Journal*, *17*(17), pp.5618-5625.

Needell, D., & Tropp, J. A. 2009. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and computational harmonic analysis*, 26(3), pp. 301-321.

Ng, A., CS294A Lecture Notes Sparse Autoencoder, Cs294a. (2011) 1–19.

Nikolaou, N.G. and Antoniadis, I.A., 2002. Rolling element bearing fault diagnosis using wavelet packets. *Ndt & E International*, *35*(3), pp.197-205.

Ocak, H., Loparo, K.A., and Discenzo, F.M., 2007. Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: A method for bearing prognostics. *Journal of sound and vibration*, *302*(4-5), pp.951-961.

Olshausen, B.A. and Field, D.J., 1997. Sparse coding with an overcomplete basis set: A strategy employed by V1. *Vision Research*, *37*(23), pp.3311-3325.

Palit, A.K. and Popovic, D., 2006. *Computational intelligence in time series forecasting: theory and engineering applications*. Springer Science & Business Media.

Paliwal, D., Choudhur, A. and Govandhan, T., 2014. Identification of faults through wavelet transform vis-à-vis fast Fourier transform of noisy vibration signals emanated from defective rolling element bearings. *Frontiers of Mechanical Engineering*, *9*(2), pp.130-141.

Pandya, D.H., Upadhyay, S.H. and Harsha, S.P., 2014. Fault diagnosis of rolling element bearing by using multinomial logistic regression and wavelet packet transform. *Soft Computing*, *18*(2), pp.255-266.

Pang, B., Tang, G., Tian, T., and Zhou, C., 2018. Rolling Bearing Fault Diagnosis Based on an Improved HTT Transform. *Sensors*, *18*(4), p.1203.

Patidar, S. and Soni, P.K., 2013. An overview on vibration analysis techniques for the diagnosis of rolling element bearing faults. *International Journal of Engineering Trends and Technology (IJETT)*, *4*(5), pp.1804-1809.

Peng, H., Long, F., and Ding, C., 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, *27*(8), pp.1226-1238.

Peng, H.W. and Chiang, P.J., 2011, May. Control of mechatronics systems: Ball bearing fault diagnosis using machine learning techniques. In *Control Conference (ASCC), 2011 8th Asian* (pp. 175-180). IEEE.

Peng, Z.K., Peter, W.T. and Chu, F.L., 2005. A comparison study of improved Hilbert–Huang transform and wavelet transform: application to fault diagnosis for rolling bearing. *Mechanical systems and signal processing*, *19*(5), pp.974-988.

Peter, W.T., Peng, Y.A. and Yam, R., 2001. Wavelet analysis and envelope detection for rolling element bearing fault diagnosis—their effectiveness and flexibilities. *Journal of vibration and acoustics*, *123*(3), pp.303-310.

Poulimenos, A.G. and Fassois, S.D., 2006. Parametric time-domain methods for non-stationary random vibration modelling and analysis—a critical survey and comparison. *Mechanical Systems and Signal Processing*, *20*(4), pp.763-816.

Prabhakar, S., Mohanty, A.R. and Sekhar, A.S., 2002. Application of discrete wavelet transform for detection of ball bearing race faults. *Tribology International*, *35*(12), pp.793-800.

Prieto, M.D., Cirrincione, G., Espinosa, A.G., Ortega, J.A. and Henao, H., 2013. Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks. *IEEE Transactions on Industrial Electronics*, *60*(8), pp.3398-3407.

Pudil, P., Novovičová, J. and Kittler, J., 1994. Floating search methods in feature selection. *Pattern recognition letters*, *15*(11), pp.1119-1125.

Purushotham, V., Narayanan, S. and Prasad, S.A., 2005. Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition. *Ndt & E International*, *38*(8), pp.654-664.

Qaisar, S., Bilal, R. M., Iqbal, W., Naureen, M., & Lee, S. 2013. Compressive sensing: From theory to applications, a survey. *Journal of Communications and networks*, *15*(5), pp. 443-456.

Rai, A., & Upadhyay, S. H. 2016. A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. Tribology International, 96, pp. 289-306.

Rai, V.K. and Mohanty, A.R., 2007. Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform. *Mechanical Systems and Signal Processing*, *21*(6), pp.2607-2615.

Rajeswari, C., Sathiyabhama, B., Devendiran, S. and Manivannan, K., 2015. Bearing fault diagnosis using multiclass support vector machine with efficient feature selection methods. *International Journal of Mechanical and Mechatronics Engineering*, *15*(1), pp.1-12.

Randall, R. B., & Antoni, J. 2011. Rolling element bearing diagnostics—A tutorial. *Mechanical systems and signal processing*, *25*(2), pp. 485-520.

Randall, R.B., 2011. *Vibration-based condition monitoring: industrial, aerospace and automotive applications*. John Wiley & Sons.

Randall, R.B., Antoni, J. and Chobsaard, S., 2000. A comparison of cyclostationary and envelope analysis in the diagnostics of rolling element bearings. In *Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on* (Vol. 6, pp. 3882-3885). IEEE.

Rao, K. R., & Yip, P. C. (Eds.). 2000. *The transform and data compression handbook* (Vol. 1). CRC press.

Rauber, T.W., de Assis Boldt, F. and Varejão, F.M., 2015. Heterogeneous feature models and feature selection applied to bearing fault diagnosis. *IEEE Transactions on Industrial Electronics*, *62*(1), pp.637-646.

Ren, L., Lv, W., Jiang, S. and Xiao, Y., 2016. Fault diagnosis using a joint model based on sparse representation and SVM. *IEEE Transactions on Instrumentation and Measurement*, *65*(10), pp.2313-2320.

Renna, F., Calderbank, R., Carin, L., & Rodrigues, M. R. 2014. Reconstruction of signals drawn from a Gaussian mixture via noisy compressive measurements. *IEEE Transactions on Signal Processing*, 62(9), pp. 2265-2277.

Rojas, A. and Nandi, A.K., 2006. Practical scheme for fast detection and classification of rolling-element bearing faults using support vector machines. *Mechanical Systems and Signal Processing*, *20*(7), pp.1523-1536.

Rojas, A. and Nandi, A.K., 2006. Practical scheme for fast detection and classification of rolling-element bearing faults using support vector machines. *Mechanical Systems and Signal Processing*, *20*(7), pp.1523-1536.

Roman, B., Hansen, A. and Adcock, B., 2014. On asymptotic structure in compressed sensing. *arXiv preprint arXiv:1406.4178*.

Rossi, M., Haimovich, A. M., & Eldar, Y. C. 2014. Spatial compressive sensing for MIMO radar. *IEEE Transactions on Signal Processing*, 62(2), pp. 419-430.

Rudelson, M., & Vershynin, R. 2008. On sparse reconstruction from Fourier and Gaussian measurements. *Communications on Pure and Applied Mathematics*, *61*(8), 1025-1045.

Saimurugan, M., Ramachandran, K.I., Sugumaran, V., and Sakthivel, N.R., 2011. Multi component fault diagnosis of rotational mechanical system based on decision tree and support vector machine. *Expert Systems with Applications*, *38*(4), pp.3819-3826.

Sait, A. S., & Sharaf-Eldeen, Y. I. 2011. A review of gearbox condition monitoring based on vibration analysis techniques diagnostics and prognostics. In *Rotating Machinery, Structural Health Monitoring, Shock and Vibration, Volume 5*, pp. 307-324. Springer, New York, NY.

Salimans, T. and Kingma, D.P., 2016. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. In *Advances in Neural Information Processing Systems* (pp. 901-909).

Samanta, B., Al-Balushi, K.R. and Al-Araimi, S.A., 2003. Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. *Engineering applications of artificial intelligence*, *16*(7-8), pp.657-665.

Samanta, B., Al-Balushi, K.R. and Al-Araimi, S.A., 2006. Artificial neural networks and genetic algorithm for bearing fault detection. *Soft Computing*, *10*(3), pp.264-271.

Sassi, S., Badri, B. and Thomas, M., 2007. A numerical model to predict damaged bearing vibrations. *Journal of Vibration and Control*, *13*(11), pp.1603-1628.

Saucedo Dorantes, J.J., Delgado Prieto, M., Osornio Rios, R.A. and Romero Troncoso, R.D.J., 2016. Multifault diagnosis method applied to an electric machine based on high-dimensional feature reduction. *IEEE transactions on industry applications*, *53*(3), pp.3086-3097.

Sawalhi, N., Randall, R.B. and Endo, H., 2007. The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis. *Mechanical Systems and Signal Processing*, *21*(6), pp.2616-2633.

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. *Neural networks*, *61*, pp.85-117.

Seera, M., Lim, C.P. and Loo, C.K., 2016. Motor fault detection and diagnosis using a hybrid FMM-CART model with online learning. *Journal of intelligent manufacturing*, *27*(6), pp.1273-1285.

Seera, M., Wong, M. D., & Nandi, A. K. 2017. Classification of ball bearing faults using a hybrid intelligent model. Applied Soft Computing, 57, pp. 427-435.

Serviere, C. and Fabry, P., 2005. Principal component analysis and blind source separation of modulated sources for electro-mechanical systems diagnostic. *Mechanical systems and signal processing*, *19*(6), pp.1293-1311.

Shao, H., Jiang, H., Zhang, H., Duan, W., Liang, T. and Wu, S., 2018. Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing. *Mechanical Systems and Signal Processing*, *100*, pp.743-765.

Shao, H., Jiang, H., Zhang, X. and Niu, M., 2015. Rolling bearing fault diagnosis using an optimization deep belief network. *Measurement Science and Technology*, *26*(11), p.115002.

Shen, C., Wang, D., Kong, F. and Peter, W.T., 2013. Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier. *Measurement*, *46*(4), pp.1551-1564.

Shuang, L. and Meng, L., 2007, August. Bearing fault diagnosis based on PCA and SVM. In *Mechatronics and Automation, 2007. ICMA 2007. International Conference on* (pp. 3503-3507). IEEE.

Siddiqui, K. M., Sahay, K., & Giri, V. K. 2014. Health monitoring and fault diagnosis in induction motor-a review. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, *3*(1), pp. 6549-6565.

Singh, S. and Vishwakarma, D.M., 2015. A Review of Vibration Analysis Techniques for Rotating Machines. *International Journal of Engineering Research & Technology (IJERT), ISSN*, pp.2278-0181.

Singhal, A. and Khandekar, M.A., 2013, February. Bearing fault detection in induction motor using fast Fourier transform. In *IEEE Int. Conf. on Advanced Research in Engineering & Technology*.

Smith, J.S., 2005. The local mean decomposition and its application to EEG perception data. *Journal of the Royal Society Interface*, *2*(5), pp.443-454.

Soualhi, A., Medjaher, K., & Zerhouni, N. 2015. Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression. IEEE Transactions on Instrumentation and Measurement, 64(1), pp. 52-62.

Sreejith, B., Verma, A.K., and Srividya, A., 2008, December. Fault diagnosis of rolling element bearing using time-domain features and neural networks. In *Industrial and Information Systems, 2008. ICIIS 2008. IEEE Region 10 and the Third international Conference on* (pp. 1-6). IEEE.

Stanković, S., Orović, I., & Stanković, L. 2014. An automated signal reconstruction method based on analysis of compressive sensed signals in noisy environment. *Signal Processing*, 104, pp. 43-50.

Staszewski, W.J., Worden, K. and Tomlinson, G.R., 1997. Time-frequency analysis in gearbox fault detection using the Wigner–Ville distribution and pattern recognition. *Mechanical systems and signal processing*, *11*(5), pp.673-692.

Su, W., Wang, F., Zhu, H., Zhang, Z. and Guo, Z., 2010. Rolling element bearing faults diagnosis based on optimal Morlet wavelet filter and autocorrelation enhancement. *Mechanical systems and signal processing*, *24*(5), pp.1458-1472.

Sugiyama, M., 2006, June. Local fisher discriminant analysis for supervised dimensionality reduction. In *Proceedings of the 23rd international conference on Machine learning* (pp. 905-912). ACM.

Sugumaran, V. and Ramachandran, K.I., 2007. Automatic rule learning using decision tree for fuzzy classifier in fault diagnosis of roller bearing. *Mechanical Systems and Signal Processing*, *21*(5), pp.2237-2247.

Sugumaran, V. and Ramachandran, K.I., 2011. Effect of number of features on classification of roller bearing faults using SVM and PSVM. *Expert Systems with Applications*, *38*(4), pp.4088-4096.

Suh, W.J., Park, C.S. and Kim, D.W., 2011, November. Heuristic vs. meta-heuristic optimization for energy performance of a post office building. In *Proceedings of the 12th conference of international building performance simulation association. Sydney: IBPSA* (pp. 704-711).

Sun, L., Liu, J., Chen, J., & Ye, J. 2009. Efficient recovery of jointly sparse vectors. In *Advances in Neural Information Processing Systems*, pp. 1812-1820.

Sun, Q. and Tang, Y., 2002. Singularity analysis using continuous wavelet transform for bearing fault diagnosis. *Mechanical systems and signal processing*, *16*(6), pp.1025-1041.

Sun, Q., Chen, P., Zhang, D. and Xi, F., 2004. Pattern recognition for automatic machinery fault diagnosis. *Journal of vibration and acoustics*, *126*(2), pp.307-316.

Sun, W., Shao, S., Zhao, R., Yan, R., Zhang, X. and Chen, X., 2016. A sparse auto-encoder-based deep neural network approach for induction motor faults classification. *Measurement*, *89*, pp.171-178.

Tahir, M.M., Khan, A.Q., Iqbal, N., Hussain, A. and Badshah, S., 2017. Enhancing fault classification accuracy of ball bearing using central tendency based time domain features. *IEEE Access*, *5*, pp.72-83.

Tang, G., Hou, W., Wang, H., Luo, G., & Ma, J. 2015. Compressive sensing of roller bearing faults via harmonic detection from under-sampled vibration signals. *Sensors*, *15*(10), pp. 25648-25662.

Tang, G., Yang, Q., Wang, H. Q., Luo, G. G., & Ma, J. W. 2015. Sparse classification of rotating machinery faults based on compressive sensing strategy. *Mechatronics*, *31*, pp. 60-67.

Tang, H., Chen, J. and Dong, G., 2014. Sparse representation based latent components analysis for machinery weak fault detection. *Mechanical Systems and Signal Processing*, *46*(2), pp.373-388.

Tang, J., Alelyani, S. and Liu, H., 2014. Feature selection for classification: A review. *Data classification: Algorithms and applications*, p.37.

Tao, J., Liu, Y. and Yang, D., 2016. Bearing fault diagnosis based on deep belief network and multisensor information fusion. *Shock and Vibration*, *2016*.

Tao, S., Zhang, T., Yang, J., Wang, X. and Lu, W., 2015. Bearing fault diagnosis method based on stacked autoencoder and softmax regression. In *Control Conference (CCC), 2015 34th Chinese* (pp. 6331-6335). IEEE.

Tawfik, M.M. and Morcos, M.M., 2001. ANN-based techniques for estimating fault location on transmission lines using Prony method. *IEEE Transactions on Power Delivery*, *16*(2), pp.219-224.

Theodoridis, S., Pikrakis, A., Koutroumbas, K., & Cavouras, D. 2010. Introduction to pattern recognition: a Matlab approach. Academic Press.

Tian, J., Morillo, C., Azarian, M.H. and Pecht, M., 2016. Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with K-nearest neighbor distance analysis. *IEEE Transactions on Industrial Electronics*, *63*(3), pp.1793-1803.

Tian, J., Pecht, M. and Li, C., 2012. Diagnosis of rolling element bearing fault in bearing-gearbox union system using wavelet packet correlation analysis'. *Dayton, OH*, pp.24-26.

Tian, Y., Ma, J., Lu, C. and Wang, Z., 2015. Rolling bearing fault diagnosis under variable conditions using LMD-SVD and extreme learning machine. *Mechanism and Machine Theory*, *90*, pp.175-186.

Toersen, H., 1998. Application of an envelope technique in the detection of ball bearing defects in a laboratory experiment. *Tribotest*, *4*(3), pp.297-308.

Tyagi, C.S., 2008. A comparative study of SVM classifiers and artificial neural networks application for rolling element bearing fault diagnosis using wavelet transform preprocessing. *Neuron*, *1*, pp.309-317.

Vakharia, V., Gupta, V.K. and Kankar, P.K., 2016. Bearing Fault Diagnosis Using Feature Ranking Methods and Fault Identification Algorithms. *Procedia Engineering*, *144*, pp.343-350.

Van Dam, J., & Bond, L. J. 2015. Economics of online structural health monitoring of wind turbines: Cost benefit analysis. In *AIP Conference Proceedings* Vol. 1650, No. 1, pp. 899-908.

Van Loan, C. 1992. Computational frameworks for the fast Fourier transform (Vol. 10). Siam.

Van, M. and Kang, H.J., 2015. Wavelet kernel local fisher discriminant analysis with particle swarm optimization algorithm for bearing defect classification. *IEEE Transactions on Instrumentation and Measurement*, *64*(12), pp.3588-3600.

Van, M. and Kang, H.J., 2016. Two-stage feature selection for bearing fault diagnosis based on dual-tree complex wavelet transform and empirical mode decomposition. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *230*(2), pp.291-302.

Verma, N. K., Khatravath, S., & Salour, A. 2013. Cost benefit analysis for condition-based maintenance. In *Prognostics and Health Management (PHM), 2013 IEEE Conference,* pp. 1-6.

Vinay, V., Kumar, G.V. and Kumar, K.P., 2015. Application of chi-square feature ranking technique and random forest classifier for fault classification of bearing faults. In *Proceedings of the 22th International Congress on Sound and Vibration, Florence, Italy* (pp. 12-16).

Vrabie, V., Granjon, P., Maroni, C.S. and Leprettre, B., 2004. Application of spectral kurtosis to bearing fault detection in induction motors. In *5th International Conference on acoustical and vibratory surveillance methods and diagnostic techniques (Surveillance5).*

Wang, D., Peter, W.T. and Tsui, K.L., 2013. An enhanced Kurtogram method for fault diagnosis of rolling element bearings. *Mechanical Systems and Signal Processing*, *35*(1-2), pp.176-199.

Wang, F., Sun, J., Yan, D., Zhang, S., Cui, L. and Xu, Y., 2015. A feature extraction method for fault classification of rolling bearing based on PCA. In *Journal of Physics: Conference Series* (Vol. 628, No. 1, p. 012079). IOP Publishing.

Wang, L., Chu, J., & Wu, J. 2007. Selection of optimum maintenance strategies based on a fuzzy analytic hierarchy process. *International journal of production economics*, 107(1), pp. 151-163.

Wang, Y. and Liang, M., 2011. An adaptive SK technique and its application for fault detection of rolling element bearings. *Mechanical Systems and Signal Processing*, *25*(5), pp.1750-1764.

Wang, Y., Xiang, J., Markert, R. and Liang, M., 2016. Spectral kurtosis for fault detection, diagnosis, and prognostics of rotating machines: A review with applications. *Mechanical Systems and Signal Processing*, *66*, pp.679-698.

Wang, Y., Xu, G., Liang, L. and Jiang, K., 2015. Detection of weak transient signals based on wavelet packet transform and manifold learning for rolling element bearing fault diagnosis. *Mechanical Systems and Signal Processing*, *54*, pp.259-276.

Wegerich, S.W., 2004, March. Similarity based modeling of time synchronous averaged vibration signals for machinery health monitoring. In *Aerospace Conference, 2004. Proceedings. 2004 IEEE* (Vol. 6, pp. 3654-3662). IEEE.

Wensheng, S., Fengtao, W., Zhixin, Z., Guo, Z.G. and Li, H.K., 2010. Application of EMD denoising and spectral kurtosis in early fault diagnosis of rolling element bearings. *Journal of Vibration and Shock*, *29*(3), pp.18-21.

Widodo, A., Kim, E.Y., Son, J.D., Yang, B.S., Tan, A.C., Gu, D.S., Choi, B.K. and Mathew, J., 2009. Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine. *Expert systems with applications*, *36*(3), pp.7252-7261.

Widodo, A., Yang, B.S. and Han, T., 2007. Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors. *Expert systems with applications*, *32*(2), pp.299-312.

Widrow, B., Glover, J.R., McCool, J.M., Kaunitz, J., Williams, C.S., Hearn, R.H., Zeidler, J.R., Dong, J.E. and Goodlin, R.C., 1975. Adaptive noise cancelling: Principles and applications. *Proceedings of the IEEE*, *63*(12), pp.1692-1716.

William, P.E., and Hoffman, M.W., 2011. Identification of bearing faults using time domain zero-crossings. *Mechanical systems and signal processing*, *25*(8), pp.3078-3088.

Wold, S., Esbensen, K. and Geladi, P., 1987. Principal component analysis, Chemometr. Intell. Lab., 2, 37–52.

Wong, M. L. D., Zhang, M., & Nandi, A. K. 2015. Effects of compressed sensing on classification of bearing faults with entropic features. IEEE *Signal Processing Conference (EUSIPCO),* pp. 2256-2260.

Worden, K., Staszewski, W. J., & Hensman, J. J. 2011. Natural computing for mechanical systems research: A tutorial overview. *Mechanical Systems and Signal Processing,* 25(1), pp. 4-111.

Wu, S.D., Wu, C.W., Wu, T.Y. and Wang, C.C., 2013. Multi-scale analysis based ball bearing defect diagnostics using Mahalanobis distance and support vector machine. *Entropy*, *15*(2), pp.416-433.

Wu, Z. and Huang, N.E., 2009. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in adaptive data analysis*, *1*(01), pp.1-41.

Xia, M., Li, T., Liu, L., Xu, L. and de Silva, C.W., 2017. Intelligent fault diagnosis approach with unsupervised feature learning by stacked denoising autoencoder. *IET Science, Measurement & Technology*, *11*(6), pp.687-695.

Xia, M., Li, T., Xu, L., Liu, L. and de Silva, C.W., 2018. Fault Diagnosis for Rotating Machinery Using Multiple Sensors and Convolutional Neural Networks. *IEEE/ASME Transactions on Mechatronics*, *23*(1), pp.101-110.

Xian, G.M., 2010. Mechanical failure classification for spherical roller bearing of hydraulic injection molding machine using DWT–SVM. *Expert Systems with Applications*, *37*(10), pp.6742-6747.

Yan, R., Gao, R.X. and Chen, X., 2014. Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal processing*, *96*, pp.1-15.

Yang, J., Zhang, Y. and Zhu, Y., 2007. Intelligent fault diagnosis of rolling element bearing based on SVMs and fractal dimension. *Mechanical Systems and Signal Processing*, *21*(5), pp.2012-2024.

Yang, Y., & Pedersen, J. O. 1997. A comparative study on feature selection in text categorization. In Icml, Vol. 97, pp. 412-420.

Yang, Y., Yu, D., & Cheng, J. 2007. A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM. Measurement, 40(9-10), pp. 943-950.

Yiakopoulos, C.T., Gryllias, K.C. and Antoniadis, I.A., 2011. Rolling element bearing fault detection in industrial environments based on a K-means clustering approach. *Expert Systems with Applications*, *38*(3), pp.2888-2911.

Yu, D., Cheng, J. and Yang, Y., 2005. Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings. *Mechanical systems and signal processing*, *19*(2), pp.259-270.

Yu, X., Dong, F., Ding, E., Wu, S., & Fan, C. 2018. Rolling Bearing Fault Diagnosis Using Modified LFDA and EMD With Sensitive Feature Selection. IEEE Access, 6, 3715-3730.

Yu, X., Hu, D. and Xu, J., 2013. *Blind source separation: theory and applications*. John Wiley & Sons.

Yu, Y. and Junsheng, C., 2006. A roller bearing fault diagnosis method based on EMD energy entropy and ANN. *Journal of sound and vibration*, *294*(1-2), pp.269-277.

Yuan, H., & Lu, C. 2017. Rolling bearing fault diagnosis under fluctuant conditions based on compressed sensing. *Structural Control and Health Monitoring*, *24*(5).

Yuan, S.F. and Chu, F.L., 2007. Fault diagnostics based on particle swarm optimisation and support vector machines. *Mechanical Systems and Signal Processing*, *21*(4), pp.1787-1798.

Zakrajsek, J. J., Townsend, D. P., & Decker, H. J. 1993. *An analysis of gear fault detection methods as applied to pitting fatigue failure data* (No. NASA-E-7470). NATIONAL AERONAUTICS AND SPACE ADMINISTRATION CLEVELAND OH LEWIS RESEARCH CENTER.

Zhang, K., Li, Y., Scarf, P. and Ball, A., 2011. Feature selection for high-dimensional machinery fault diagnosis data using multiple models and radial basis function networks. *Neurocomputing*, *74*(17), pp.2941-2952.

Zhang, L. and Nandi, A.K., 2007. Fault classification using genetic programming. *Mechanical systems and signal processing*, *21*(3), pp.1273-1284.

Zhang, L., Jack, L.B. and Nandi, A.K., 2005. Fault detection using genetic programming. *Mechanical Systems and Signal Processing*, *19*(2), pp.271-289.

Zhang, W., Li, C., Peng, G., Chen, Y. and Zhang, Z., 2018. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mechanical Systems and Signal Processing*, *100*, pp.439-453.

Zhang, X., & Zhou, J. 2013. Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines. Mechanical Systems and Signal Processing, 41(1-2), pp. 127-140.

Zhang, X., Chen, W., Wang, B. and Chen, X., 2015. Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization. *Neurocomputing*, *167*, pp.260-279.

Zhang, X., Hu, N., Hu, L., Chen, L., & Cheng, Z. 2015. A bearing fault diagnosis method based on the low-dimensional compressed vibration signal. *Advances in Mechanical Engineering*, *7*(7), pp. 1 - 12.

Zhang, X., Liang, Y. and Zhou, J., 2015. A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM. *Measurement*, *69*, pp.164-179.

Zhang, Y. and Randall, R.B., 2009. Rolling element bearing fault diagnosis based on the combination of genetic algorithms and fast kurtogram. *Mechanical Systems and Signal Processing*, *23*(5), pp.1509-1517.

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P. and Gao, R.X., 2016. Deep learning and its applications to machine health monitoring: A survey. *arXiv preprint arXiv: 1612.07640*.

Zhao, S., Liang, L., Xu, G., Wang, J. and Zhang, W., 2013. Quantitative diagnosis of a spall-like fault of a rolling element bearing by empirical mode decomposition and the approximate entropy method. *Mechanical Systems and Signal Processing*, *40*(1), pp.154-177.

Zheng, J., Cheng, J., Yang, Y. and Luo, S., 2014. A rolling bearing fault diagnosis method based on multi-scale fuzzy entropy and variable predictive model-based class discrimination. *Mechanism and Machine Theory*, *78*, pp.187-200.

Zheng, J., Pan, H., & Cheng, J. 2017. Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines. Mechanical Systems and Signal Processing, 85, pp. 746-759.

Zhou, W., Habetler, T. G., & Harley, R. G. 2007. Bearing condition monitoring methods for electric machines: A general review. *Diagnostics for Electric Machines, Power Electronics, and Drives, SDEMPED 2007. IEEE International Symposium on*, pp. 3-6.

Zhou, Y., Chen, J., Dong, G.M., Xiao, W.B. and Wang, Z.Y., 2011. Wigner–Ville distribution based on cyclic spectral density and the application in rolling element bearings diagnosis. *Proceedings*

of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, *225*(12), pp.2831-2847.

Zhu, H., Wang, X., Zhao, Y., Li, Y., Wang, W. and Li, L., 2015. Sparse representation based on adaptive multiscale features for robust machinery fault diagnosis. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *229*(12), pp.2303-2313.

Zhu, K., Song, X. and Xue, D., 2014. A roller bearing fault diagnosis method based on hierarchical entropy and support vector machine with particle swarm optimization algorithm. *Measurement*, *47*, pp.669-675.

Zhu, Z.K., He, Z., Wang, A. and Wang, S., 2009. Synchronous enhancement of periodic transients on polar diagram for machine fault diagnosis. *International Journal of Wavelets, Multiresolution and Information Processing*, *7*(04), pp.427-442.

Appendix II

Index