

Development of Deep Learning Based Methodology on Rotating Machine Fault Diagnosis

BY

MIAO HE

B.S., University of Science and Technology Beijing, 2012

M.S., Purdue University, 2014

THESIS

Submitted as partial fulfillment of the requirements
for the degree Doctor of Philosophy in Industrial Engineering and Operations Research
in the Graduate College of the
University of Illinois at Chicago, 2018

Defense Committee:

David He, Chair and Advisor

Michael A. Brown, Mechanical and Industrial Engineering

Lin Li, Mechanical and Industrial Engineering

Yayue Pan, Mechanical and Industrial Engineering

Eric Bechhoefer, Green Power Monitoring Systems, Inc.

This dissertation is dedicated to my parents, Xingzhao He and Chunfang Dang, for their eternal love and unswerving supports, and continually imparting the importance of education.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude and appreciation to my Ph.D. advisor, Dr. David He, who inspired and encouraged me constantly to this successful completion of my doctorate in the Industrial Engineering and Operation Research degree. His expertise, technical and editorial advises guided me through this journey and helped me matured academically, professionally and personally.

I would also like to express my appreciation to all my dissertation committee members, Dr. Michael A. Brown, Dr. Lin Li, Dr. Yayue Pan, and Dr. Eric Bechhoefer, for their time spent and valuable comments provided.

Furthermore, I would like to thank Dr. Eric Bechhoefer, Dr. Ruoyu Li, Dr. Jae Myung Yoon, for the precious career opportunities that had lasting effect. In addition, a thank you to all the past and current colleagues, Dr. Jae Myung Yoon, Dr. Brandon Van Hecke, Dr. Yongzhi Qu, Jason Deutsch and Khaled Akkad, for their valuable discussion and suggestions on my projects. It was my honor to be part of the ISM&D team and conduct researches with them.

CONTRIBUTION OF AUTHORS

Chapter 1 is an introduction to rotating machinery fault diagnosis, research background, acoustic emission and deep learning background. Chapter 2 is a literature review covering the state of the art techniques available for rotating machinery fault diagnosis, deep learning related application in condition-based maintenance. Chapter 3 explains the motivation for the research presented in this dissertation. Chapter 4 presents the detailed methodology for the validation results in Chapter 6. The majority of the content is composed of previously published work (He, M., He, D., 2018, “Wind turbine planetary gearbox feature extraction and fault diagnosis using a deep learning based approach”, *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, Copyright © 2018 (SAGE Publications). Reprinted by permission of SAGE Publications. DOI: 10.1177/1748006X18768701; He, M., He, D., 2018, “Simultaneous bearing fault diagnosis and severity detection using a LAMSTAR network-based approach”, *IET Science, Measurement & Technology*, DOI: 10.1049/iet-smt.2017.0528; and He, M and He, D., 2017, “A deep learning based approach for bearing fault diagnosis”, *IEEE Transactions on Industrial Application*, Vol. 53, No. 3, pp. 1-9, DOI: 10.1109/TIA.2017.2661250) for which I was the first author and perform the research. My advisor, Dr. David He supervised the research and edited the manuscript. Chapter 5 presents the experimental setup for the research herein. Chapter 6 presents the validation results of multiple case studies. Section 6.1 and Section 6.3 are results from the previously mentioned publications. Section 6.2, Section 6.4 and Section 6.5 present the unpublished investigations on integrating signal processing into deep learning algorithm explained in Section 4.6. Finally, Chapter 7 provides a synthesis of the research presented in this dissertation.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	ii
CONTRIBUTION OF AUTHORS	iii
TABLE OF CONTENTS	iv
SUMMARY	x
1. INTRODUCTION	1
1.1 Rotating Machinery – A Brief Introduction.....	1
1.2 Fault Diagnosis and Condition-Based Maintenance on Rotating Machinery.....	3
1.3 Deep Learning Based Machinery Fault Diagnosis Methodology	5
1.4 Objective	7
1.5 Outline	8
2. LITERATURE REVIEW.....	9
2.1 Rotating Machinery Fault Diagnosis.....	9
2.1.1 Rotating bearing fault diagnosis.....	9
2.2 Deep Learning Based Methodology and Application in Fault Diagnosis	15
2.2.2 Sparse Coding and Application in Fault Diagnosis	18
2.3 Signal Processing Techniques on Rotating Machinery Fault Diagnosis	21
3. Limitations of the Current Deep Learning Based Fault Diagnosis Methods	26
4. METHODOLOGY	27
4.1 Overview of the LAMSTAR Neural Network Based Rotating Machinery Fault Diagnosis Methodology	27
4.2 Basic LAMSTAR Network Structure.....	28
4.3 The Design of the LAMSTAR Network for Rotating Machinery Fault Diagnosis	32
4.3.1 Sub-pattern Generation	33
4.3.2 Input Normalization	35
4.3.3 Fundamental Principles Used in Dynamic SOM Module Design	35
4.3.4 Dynamic Formation of Neurons and Weights in SOM Module.....	37
4.3.5 Determination of the Link Weights	42
4.4 LAMSTAR Network Based Bearing Fault Diagnosis	42
4.5 Sparse Coding and Combination with LAMSTAR Network on Rotating Machinery Fault Diagnosis	42

4.5.1	<i>Sparse Coding Model</i>	43
4.5.2	<i>Sparse Coding and Dictionary Learning From Time Series Data</i>	45
4.5.3	<i>LAMSTAR-DL Based PGB Feature Extraction and Fault Diagnosis</i>	47
4.6	Hybrid Deep Signal Processing Approach for Bearing Fault Diagnosis Using Vibration Signals	50
5	EXPERIMENTAL SETUP	72
5.1	Introduction of the Bearing Test rig	72
5.2	Introduction of the PGB Test rig.....	76
6	VALIDATION RESULTS.....	79
6.1	The Validation Results of Bearing Fault Diagnosis Using AE signals.....	79
6.2	The Validation Results of Bearing Fault Diagnosis Using Vibration Signals	84
6.3	The Validation Results of PGB Fault Diagnosis Using Vibration Signals.....	98
6.3.1	<i>The Designed LAMSTAR Network and LAMSTAR-DL Network for the Validation Test</i>	98
6.3.2	<i>The Validation Results</i>	101
6.4	The Validation Results of Deep Hybrid Signal Processing Method on Bearing Fault Diagnosis using Vibration Signals	105
6.4.1	<i>Experimental Setup and Data Acquisition</i>	105
6.4.2	<i>The Validation Results</i>	106
6.5	The Validation Results of Deep Hybrid Signal Processing Method on Gear Prognostics Using Vibration Signals.....	113
6.5.1	<i>Description of the Validation Data Sets</i>	113
6.5.2	<i>The Validation Results</i>	115
7	CONCLUSIONS	118
8	REFERENCES	122
	APPENDIX	134
	VITA.....	138

LIST OF TABLES

Table 5.1. Shaft speed and number of samples collected for all bearing conditions.....	76
Table 5.2. Vibration DAQ settings.....	78
Table 6.1 Detailed description of samples and generated patterns.....	79
Table 6.2 Bearing conditions and LAMSTAR network output neuron firing sequence representations.....	81
Table 6.3 Overall accuracy at different shaft speeds.....	81
Table 6.4. Bearing details and fault frequencies (CWRU).....	85
Table 6.5. Detailed description of vibration data sets for validation.....	86
Table 6.6. Diagnostic results with constant loading of 0 Hp and sampling rate as 48 kHz.....	87
Table 6.7. Diagnostic results with mixed loading conditions and a sampling rate of 12 kHz.....	90
Table 6.8. Diagnostic results with mixed loading condition and a sampling rate of 48 kHz.....	90
Table 6.9. Accumulated variance of the first three principle components.....	93
Table 6.10. Detailed description of collected samples and generated patterns.....	99
Table 6.11. PGB health conditions and LAMSTAR network based approaches firing sequence representations.....	101
Table 6.12. PGB fault diagnosis results using LAMSTAR and LAMSTAR-DL.....	103
Table 6.13 Classification results with different shaft rotating speeds.....	107
Table 6.14 Classification results with various pre-defined segment length.....	108
Table 6.15 Classification results with different operation conditions for training and testing datasets.....	109
Table 6.16 Comparison Results with Other Deep Learning Methods.....	112
Table 6.17 Detailed information of carried out experiments.....	114

LIST OF FIGURES

Figure 1.1. Components of a rolling bearing.....	2
Figure 1.2. Components of a planetary gearbox.....	2
Figure 4.1. General structure of the LAMSTAR based method for bearing fault diagnosis and severity detection	28
Figure 4.2. General tructure of the LAMSTAR network.....	30
Figure 4.3. The procedure for obtaining each sub-pattern from spectrum matrix as input to LAMSTAR network.	34
Figure 4.4. Flow chart of training procedure using dynamic neurons in the input SOM modules starting from the second training input pattern.....	39
Figure 4.5. Flow chart of testing procedure using dynamic neurons in the input SOM modules.	41
Figure 4.6 The framework of the LAMSTAR-DL network based PGB feature extraction and fault diagnosis.....	46
Figure 4.7. The scheme to obtain the sub-patterns from raw signal as input to LAMSTAR-DL network for dictionary learning.	48
Figure. 4.8 Framework of the hybrid deep signal processing approach for bearing fault diagnosis using vibration signals	50
Figure. 4.9. Synthesized tachometer generation from speed profile	51
Figure. 4.10. Schematic of convolution layer in the CNN with adjustable filter size	54
Figure 4.11. Illustration definition of an adjustable filter size k	55
Figure. 4.12. Illustration of the expanding and shrinking process.....	58
Figure 4.13. The structure of DFT-IDFT autoencoder with dual inputs layers	61
Figure. 4.14 The DFT-IDFT autoencoder to compute the TSA.....	65
Figure. 4.15. Validation result using TSA embedded neural network on sample signal.....	66
Figure 4.16. The DFT-IDFT autoencoder to compute the TSR and reconstructed vibration signals	67

Figure. 4.17. Reconstructed signals using theoretic TSR result and DFT-IDFT autoencoder result with outer race fault bearing signal.....	70
Figure 5.1. The bearing test rig.....	73
Figure 5.2. The steel bearing with seeded faults	74
Figure 5.3. Data acquisition system: demodulation board, sampling board, and function generator	74
Figure 5.4. The PGB test rig for wind turbine simulator	77
Figure. 5.5. Seeded faults: (a) sun gear fault, (b) planet gear fault, (c) ring gear fault.....	78
Figure 6.1. Average output values of the winning neurons at different shaft speeds.....	83
Figure 6.2. The seeded bearing test rig from Case Western University	84
Figure 6.3. Plots of the principal components of the winning neuron outputs using data collected at sampling rate of:(a) 12 kHz and (b) 48 kHz.....	91
Figure 6.4. Plots of the principal components of the winning neuron outputs using the data collected at the sampling rate of:(a)12 kHz and (b) 48 kHz.	92
Figure 6.5. Average output value of the winning neurons using data collected at the sampling rate of:(a)12 kHz and (b) 48 kHz.	94
Figure 6.6. Diagnosis results comparison of different window sizes	96
Figure 6.7. Diagnosis results comparison with different percentages of samples for training.	96
Figure 6.8. Average value of winning neurons using LAMSTAR network.....	102
Figure 6.9. Average value of winning neurons using LAMSTAR-DL network.....	102
Figure 6.10. The bearing test rig.....	105
Figure 6.11. Visualization of randomly selected row vectors in W_{i_re} trained by signals with mixed shaft speeds	111
Figure 6.12. RMS by shaft speed using hybrid signal processing method	111
Figure 6.13. Peak by shaft speed using hybrid signal processing method.....	112
Figure 6.14 Spiral Bevel Gear Fatigue Rig.....	114

Figure 6.15 The comparison results from experiment 3. (a) RMS extracted from output of the neural network; (b)RMS extracted from TSA calculation; (c) Other CIs extracted from TSA calculation 116

Figure 6.16 The comparison results from experiment 4. (a) RMS extracted from output of the neural network; (b)RMS extracted from TSA calculation; (c) Other CIs extracted from TSA calculation 116

Figure 6.17 The comparison results from experiment 6. (a) RMS extracted from output of the neural network; (b)RMS extracted from TSA calculation; (c) Other CIs extracted from TSA calculation 117

Figure 6.18 The comparison results from experiment 7. (a) RMS extracted from output of the neural network; (b)RMS extracted from TSA calculation; (c) Other CIs extracted from TSA calculation 117

SUMMARY

Rotating machines are widely used in various industrial applications. It is necessary to implement the condition based maintenance for rotating machines to prevent failures, increase reliability and decrease maintenance cost. Fault detection and diagnosis are critical in rotating machinery condition-based maintenance. Traditionally, the most critical issue in developing rotating machine fault detection and diagnosis methods is to extract and quantify the complicated signal processing-based fault features. With the combination of data mining techniques, faults can be diagnosed accurately using previously extracted features. However, nowadays there are challenges in using existing methods for rotating machinery fault diagnosis. In the age of Internet of Things and Industrial 4.0, massive real-time data were collected from health monitoring systems for fault diagnosis. The traditional methods to extract features from monitoring data manually with expertise in signal processing and prior knowledge in fault diagnosis is rarely accomplishable on a machinery big data platform. Therefore, a novel methodology that can automatically extract the adaptive fault features from monitoring data and, diagnose the fault pattern intelligently, is expected to realize rotating machinery fault detection and diagnosis on machinery big data platform.

With its deep architecture, deep learning can automatically extract features from the data and hence eliminate the process of handcrafting features from the data. Though there is a growing interest in using deep learning for machinery fault detection and diagnosis, some challenges still exist. The most reported applications of deep learning on fault diagnosis of rotating machinery include convolutional neural network (CNN) (Janssens *et al.*, 2016), deep neural network (DNN) (Jia *et al.*, 2016), and deep belief network (Gan *et al.*, 2016). However, the raw monitoring data were processed with complicated signal processing algorithms such as wavelet-package transform (WPT) (Gan *et al.*, 2016), or pre-processed to obtain features such as RMS and kurtosis (Janssens *et al.*, 2016). The complicated signal processing is still required in many reported deep learning based fault diagnosis applications in literature. Besides, DNN based bearing fault diagnosis depends greatly on supervised fine-tuning process to ensure a satisfactory classification accuracy

(Jia *et al.*, 2016). As Jia *et al.* (2016) mentioned, the current selection of DNN architecture is trial and error based. The selection of deep learning architecture has not been well investigated yet. Until now, only the vibration condition monitoring data was studied with the application of deep learning based approaches. Other monitoring data such as acoustic emission (AE) data and piezoelectric (PE) data have yet to be processed with deep learning based approaches.

In this research, novel deep learning based methodologies that can automatically extract the adaptive fault features from monitoring data and intelligently diagnose the faults with machinery big data is developed to address the issues stated above. Specifically, the following new effective and efficient rotating machine fault diagnosis are presented: a deep learning based approach for bearing fault diagnosis using AE signals, a deep learning based approach for simultaneous bearing fault diagnosis and fault severity detection using vibration signals, a deep learning based approach for planetary gear box (PGB) fault diagnosis, a signal processing integrated deep learning approach for bearing fault diagnosis using vibration signals. Initially, a large memory storage and retrieval (LAMSTAR) neural network based approach for rotating machinery fault diagnosis and severity detection is investigated. The LAMSTAR neural network based method pre-processes the monitoring signals by using short time Fourier transform (STFT). With simply processed data, an optimized LAMSTAR neural network is used to simultaneously diagnose faults and detect severity. Inspired by sparse coding, a novel sparse coding and LAMSTAR combined neural network with deep architecture is proposed. The combination of sparse coding and LAMSTAR neural network results in sparse feature extraction with adaptive property. The realization of adaptive feature extraction and learning can reduce the ratio of training samples to testing samples. Furthermore, a novel signal processing integrated deep learning method is proposed to capture the hidden time and frequency features in the monitoring data. The introduction of signal processing into deep learning method provides a view of effective deep learning method on time series monitoring signals. To validate the proposed methodology on rotating machinery fault diagnosis, data collected from a bearing test rig and a planetary gear box (PGB) test rig were used. The data was collected

from the runs on bearings and gears with seeded typical faults. Vibration and AE data were collected at the bearing test rig, while vibration, AE, and PE data were collected on the PGB test rig.

The potential contributions of this research include:

- (1) Development of effective deep learning based methodology on bearing fault diagnosis using AE signals. This methodology is developed by using LAMSTAR network to obtain the self-learned fault features and it requires only simple pre-processing methods such as FFT and STFT.
- (2) Validation of the developed deep learning based bearing fault diagnosis method with AE signals using steel bearing tested at low (2-10 Hz) and high (greater than 10 Hz) applications. The diagnosis performance on all 4 bearing fault types in the low shaft rotating speed outstands the previously reported results.
- (3) Development of effective deep learning based methodology on simultaneous bearing fault diagnosis and fault severity detection. The LAMSTAR network is employed to capture the fault features from time-frequency spectra of raw signal through unsupervised learning. Then the self-learned features are used to locate the bearing faults and detect fault severity simultaneously.
- (4) Validation of the developed deep learning based simultaneous bearing fault diagnosis and severity detection methodology using seeded fault tests on a bearing test rig with vibration signals. The diagnosis performance outstands the previously reported results.
- (5) Development of effective deep learning based methodology for PGB fault diagnosis using vibration signals. This methodology is developed by integrating sparse coding and LAMSTAR network to obtain optimized deep architecture as LAMSTAR-DL. Then the fault features are extracted through LAMSTAR-DL and used for PGB fault diagnosis.
- (6) Validation of the developed deep learning based diagnostic method using seeded fault tests on a PGB test rig in laboratory.

- (7) Development of a hybrid deep signal processing approach for bearing fault diagnosis using vibration signals. This methodology is developed by mimicking the procedure of time synchronous resampling method in aspect of signal processing techniques.
- (8) Validation of the developed hybrid deep signal processing approach using seeded fault tests on a bearing test rig.
- (9) The developed hybrid signal processing approach is extended to extract fault features from simply processed monitoring signals for gearbox prognostics.
- (10) Validation of the hybrid signal processing approach on monitoring signals from run-to-failure experiments on a spear gearbox test rig.

1. INTRODUCTION

1.1 Rotating Machinery – A Brief Introduction

Rotating machinery plays a critical part in various industrial applications. The common rotating machinery includes multiple widely used components, such as rolling element bearings, gear box and planetary gear box (PGB). The rotating machinery operates in tough working conditions, which tends to cause faults and produce dysfunctional parts. The rotating machinery faults may decrease the machinery service performance in many ways such as reducing manufacturing quality, reducing manufacturing speed and operation safety, and even resulting in total breakdown. Such faults will significantly affect the manufacturing process efficiency and commercial benefits to the manufacturer. The rotating machinery in modern industry develops with more critical requirements on maintenance, machinery dimension, automation and precise performance. Therefore, there is a growing interest on increasing the reliability of rotating machinery. One common fault diagnosis strategy is using appropriate signal processing techniques to reveal the characteristics of faults from the monitoring signals (Fan *et al.*, 2008; Lei *et al.*, 2008). However, there is the limitation of such traditional methodology on automatic fault diagnosis and discovering critical fault information for adaptive fault diagnosis. Particularly, rolling element bearings and PGBs are studied for development of automatic and intelligent fault diagnosis methodology. Figure 1.1 and Figure 1.2 show the basic components of a rolling bearing and a PGB, respectively.

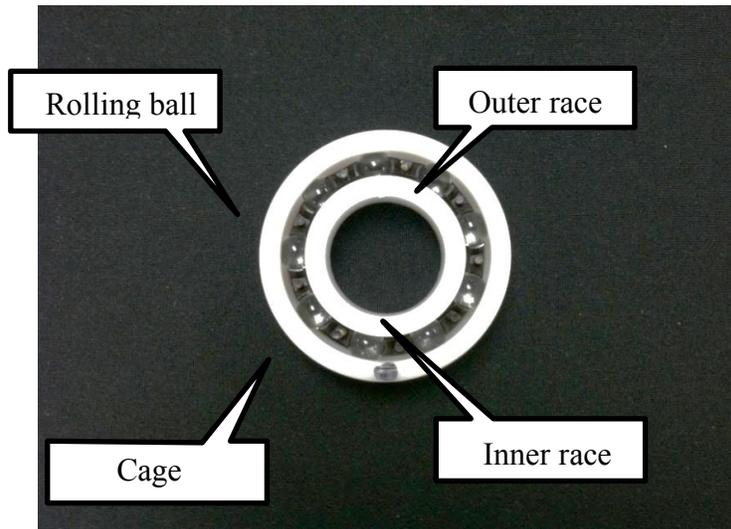


Figure 1.1. Components of a rolling bearing

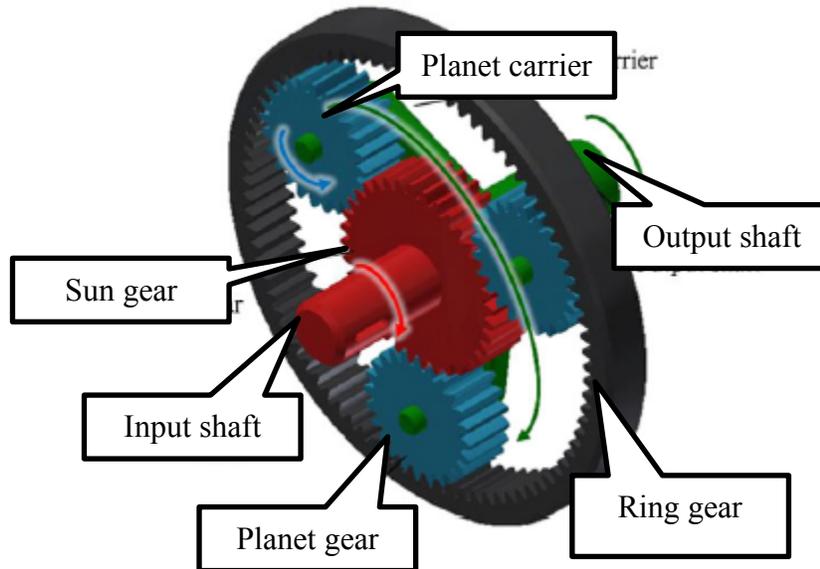


Figure 1.2. Components of a planetary gearbox

Rolling element bearings are designed in various dimensions, load capacity, weight, etc. They are used with an advantage of a balanced performance with considerations including cost, size, weight, friction, and carrying capacity. As shown in Figure 1.1, rolling element bearings are composed of rolling elements and two bearing rings, namely inner raceway (race) and outer raceway (race). The rolling element, regarded as balls or rollers, between two bearing rings, bears the external loading applied on bearing. Faults or damages commonly occur on both inner and outer raceways, rollers, and cage structure (Mathew and Alfredson, 1983).

PGBs are normally used in applications requiring high power density, such as wind turbine drivetrain systems and motor transmission systems. The rising deployment of commercial wind energy projects increases the demands of a more proactive maintenance strategy and a more intelligent methodology so that the wind energy can be more competitive to the traditional energy forms. The modern wind industry has experienced a higher wind turbine gearbox failure due to the various reasons such as: unexpected overloading caused by rare operating conditions, defective gearbox design, and ineffective maintenance. The gearbox failure is the most critical contributor to the total down time of wind turbines (Sheng *et al.*, 2011). The main faults of PGBs occur on sun gear and ring gear tooth cut. Due to the complicated structure of PGBs, the traditional vibration signals are incapable of revealing the characteristic of faults for diagnosis.

1.2 Fault Diagnosis and Condition-Based Maintenance on Rotating Machinery

A general fault diagnosis system should be functional in two tasks, fault detection and fault diagnosis (Simani *et al.*, 2000). Fault detection is to determine if fault occurred on the target machine. The health monitoring data of the target machine will be collected and processed to detect if there are any changes compared with the one obtained under normal machinery condition. With positive fault detection on the target machine, the specific fault can be located and determined. To an issue as rotating machinery, the fault could be originated from basic defective on critical component, such as rolling element bearing and gear box. To increase the operational reliability of rotating machines, it is necessary to keep monitoring critical components. Typically, accelerometers are widely used for collecting vibration data from the target machine. However, the theoretical amplitude modulation (AM) effect in vibration signal complicated the fault diagnosis problem on rotating machinery. Particularly in PGB fault diagnosis with vibration signals, AM effect caused by time variant vibration transfer paths of planet gear due to the unique dynamic structure of PGB makes it difficult to diagnosis faults on PGB. Such effect also existed in rolling element bearing

monitoring data. As bearing faults develop, the AM effect has an increasing impact on bearing monitoring data. Also, the fact that accelerometers can only be mounted on the housing or frame of rotating components directly brings extra noise into the collected raw data. Various signal processing techniques have been developed and implemented to extract fault information from raw signals. Widely studied signal processing techniques include Hilbert Huang transform (HHT), wavelet transform (WT), wavelet package transform (WPT), time synchronous average (TSA) and so on.

Apart from vibration signals being studied in this research, other types of machine monitoring data will also be studied using deep learning based approaches. Signals collected by acoustic emission (AE) is commonly studied alternative monitoring signal resources. AE signals have been proven to be effective in investigating material behavior under mechanical stress. AE is defined as transient elastic waves within material that caused by localized stress energy. These waves are transient burst of energy generated by sudden change in material state. As a physical property, all materials have their specific elastic properties. In comparison with vibration signals, AE signals have certain advantages, one of which is the capability of incipient fault detection (Van Hecke *et al.*, 2014). Since AE signals are sensitive to fault location, they are more suitable than vibration signals to locate faults. Moreover, the information in frequencies higher than the sampling frequencies of vibration signals can be retrieved by AE signals.

With rapid development of measurement and data storage techniques, an amount of data-driven based fault detection and diagnosis methodologies have been studied to determine fault condition using extracted fault features. Data-driven methods such as typical data mining techniques and artificial neural networks (ANN) have been proven effective on condition identification (He *et al.*, 2011). However, as one characteristic of data-driven methods, a significant amount of training data is required to achieve a good performance with a satisfactory accuracy. Though the data-driven methods show the promising results on fault diagnosis, it should be noticed that such results are dependent greatly on signal processing procedure to extract effective features carrying fault information. So far, most of rotating machinery fault diagnosis

systems are designed for using vibration signals. As explained previously, the theoretical AM effect of vibration signal will affect the fault diagnosis greatly. Besides, only vibration signals have been reported in the references as validation of rotating machinery fault diagnosis using deep learning. There is a necessity to study the adaptive deep learning based automatic rotating fault diagnosis methodologies using AE signals. The performance of appropriately designed deep learning based methodologies for AE and PE signals still needs to be validated and evaluated.

1.3 Deep Learning Based Machinery Fault Diagnosis Methodology

Most recently, deep learning as a new trending tool set has risen to attract interests in various research fields, including machinery fault diagnosis. With the complicated structure and ability to dig out hidden non-linear relationship between items of deep learning, it has great potential to diagnose fault without great dependency on feature extraction.

As an extension to the ANN, deep learning can be regarded as a branch of machine learning methods which are complicated ANNs with multiple hidden layers. Deep learning attempts to model high level abstractions in data by using the structure with multiple hidden layers. The multiple hidden layers can be composed of multiple linear or non-linear transformation. It has been reported that one promising outstanding contribution of deep learning is the elimination on manually feature extraction with unsupervised or semi-supervised automatically feature extraction using efficient algorithm (Song and Lee, 2013). Initially, deep architecture as the architecture with multiple hidden layer was developed for high level abstractions representation, such as in vision, language, etc. (Bengio, 2009). The advantage of automatically feature extraction in models with deep architecture has drawn attention from scholars in prognostic and health monitoring field. The fault diagnosis system can possibly be realized without solid support from complicated signal processing and feature extraction. Furthermore, feature extraction and learning without handcraft procedure enable the model to process large volume of data generated rapidly. Successful cases of

implementation deep learning based methods on big data set include Alpha Go player developed by google, intelligent translation system by google (Wu *et al.*, 2016), and the developing automatic driving technique.

Commonly used deep learning based methods include deep belief network (DBN), convolutional neural network (CNN), stacked autoencoder (SAE), and other optimized models based on these basic structures. It should be noticed that DBN is composed by stacking restricted Boltzmann machine (RBM). RBM and autoencoder share the similar structure that there are only three layers in each of them, namely one input layer, one hidden layer and one output layer. The output variables from output layer share the same value as input variables into input layer. Thus, the operating function of hidden layer is to adjust weight by training procedure, so that output variables from model can retrieve the input variables after processing input data through the model. In such case, the layer structure in both RBM and autoencoder can also be presented as composition of two layers, which is input layer and hidden layer with dual direction connection. RBM/autoencoder can be stacked by using hidden layer of previous RBM/autoencoder as input layer to the following RBM/autoencoder, to obtain the neural networks with deep architecture. Different from typical layer-by-layer greedy searching deep learning based methods as DBN or stacked autoencoder, large memory storage and retrieval (LAMSTAR) neural network is composed by multiple parallel self-organizing-map (SOM) layers. Taking the multiple SOM modules as multiple hidden layers, LAMSTAR network can also be regarded as the network with deep architecture. The connection between each SOM layer and output layer is constructed in LAMSTAR network, and will be used for feature extraction and classification along with input data. In the case of typical deep learning based models, the models are composed by neural networks with deep architecture for feature extraction and a simple classifier to identify the extracted features. In LAMSTAR network based models, the feature extraction is accomplished in SOM models. More detailed explanation is provided in Section 4.

In recent years, deep learning based signal processing methods for bearing fault diagnosis have been developed. Unlike vibration analysis based techniques such as time synchronous average algorithm which

requires much human expertise on signal processing, one major benefit of deep learning based approaches is the adaptive learning capability of deep learning for automatic signal processing and feature extraction. With the deep structures, deep learning related methods can dig the hidden non-linear complicated relationship in the original signal without using complicated signal processing techniques. When applied to the rotating machine fault diagnosis, the reported deep learning related studies have shown the outstanding performance in comparison with traditional data-driven fault diagnosis methods (Jia *et al.*, 2016; Gan *et al.*, 2016). A detailed literature review of deep learning based fault diagnosis is provided in Section 2.

1.4 Objective

The objectives of this research include: (1) developing optimized deep learning based rotating machinery fault diagnosis methodologies which are suitable for various types of sensors and, (2) validating the proposed methods with experiments. Moreover, with potential ability of fast processing data, the developed deep learning based methodology are developed with the goal of efficiently processing industrial big data platform for accurate monitoring and fault diagnosis.

To achieve these goals, the deep learning and related machinery fault diagnosis will be studied. Different deep learning algorithms will be compared and LAMSTAR network will be used mainly to construct the optimized deep learning based rotating machinery fault diagnosis methodology. Then, the LAMSTAR based bearing fault diagnosis method is extended to simultaneously diagnose bearing fault and fault severity. Next, LAMSTAR and sparse coding are combined for both bearing and PGB fault diagnosis. Furthermore, a hybrid deep signal processing method is developed for bearing fault diagnosis using vibration data. Additionally, the hybrid deep signal processing method is extended to extract features for PGB prognostics. Given that a bearing test rig and a PGB test rig have been constructed, validation and evaluation will be performed on monitoring data from both test rigs. Specifically, vibration and AE signals were collected from the bearing test rig and vibration, AE and PE signals were collected from PGB test rig. Two data acquisition systems were designed and set up for bearing and PGB test rig, respectively. Both tested

bearings and gears were seeded with multiple single-point faults. There is only one certain fault on every specific tested bearing and gear for individual fault data collection. The location and dimension of seeded fault have been observed and studied both in prior research and industrial frame. The bearings were seeded with inner race, outer race, ball, and cage faults separately for data collection. The outcome of this research is a effective and efficient deep learning based fault diagnosis methodology that can automatically extract the effective features from simple pre-processed signals with satisfied classification accuracy. Additionally, this research provides theoretical foundation to optimize the deep structure based on the characteristics of monitoring data rather than typical trial-and-error method. Particularly for the study of hybrid deep signal processing method on gear prognostics, the monitoring vibration data provided by NASA Glenn Spiral Bevel Gear Fatigue Rig is used for validation. The outcome of this extended research is effective methodology and tool that can extract features showing clear degradation trend of damages on gears that can be used for gear prognostics.

1.5 Outline

The remaining part of the proposal is outlined as follows. Section 2 provides a detailed literature review on the current state of rotating machinery fault detection and diagnosis. The techniques used in rotating machinery fault diagnosis are also introduced in Section 2, as well as the literature related with deep learning and implementations of deep learning in fault diagnosis and gear prognostics. Section 3 presents the limitation of current deep learning based rotating machinery fault diagnosis and motivation as to the necessity of this research. In Section 4, the proposed methodology is explained in detail. Section 5 presents the experimental setup and explanation of the collected data. Then, Section 6 presents the validation results. Lastly, Section 7 concludes the dissertation.

2. LITERATURE REVIEW

2.1 Rotating Machinery Fault Diagnosis

(Parts of the literature review in this chapter were previously published as (1) He, M., He, D., 2018, “Wind turbine planetary gearbox feature extraction and fault diagnosis using a deep learning based approach”, *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, © 2018 (SAGE Publications), Reprinted by permission of SAGE Publications. DOI: 10.1177/1748006X18768701; (2) He, M., He, D., 2018, “Simultaneous bearing fault diagnosis and severity detection using a LAMSTAR network-based approach”, *IET Science, Measurement & Technology*, DOI: 10.1049/iet-smt.2017.0528; (3) He, M and He, D., 2017, “A deep learning based approach for bearing fault diagnosis”, *IEEE Transactions on Industrial Application*, Vol. 53, No. 3, pp. 1-9, DOI: 10.1109/TIA.2017.2661250)

2.1.1 Rotating bearing fault diagnosis

As a typical mechanical component, rotating machinery is widely used as vital component in auto-manufacturing, heavy duty machinery, etc. Possible occurrence of faults on rotating machinery potentially leads to fatal breakdown of machines and further develops into significant economic loss as result. To completely inspect health conditions of rotating machinery, real-time data from machines is collected through condition monitoring system and ended with massive data collection after long time operation. Commonly used condition monitoring techniques include vibration, AE, temperature measurement, oil debris, electrostatic and ultrasound. Among all various studied condition monitoring techniques, vibration and AE are most commonly used techniques for mechanical fault diagnosis in time domain analysis, frequency domain analysis, time and frequency domain analysis (Shah and Patel, 2014). In the studies reported recently, many advanced signal processing techniques have been developed for rotating machinery fault diagnosis. Advanced signal processing techniques such as HHT, empirical mode decomposition (EMD), wavelet transform (WT), and fuzzy entropy were well studied in rolling bearing fault diagnosis on both vibration and AE signals. Pandya *et al.* (2013) processed the AE signals with HHT and used a k-nearest neighbor (KNN) based classifier for the rolling bearing fault pattern recognition. Chacon *et al.* (2015) used AE signals

collected from the bearing with outer race defective and processed by WT method to detect and identify the fault location at early stage. With development of data collection and storage technique, it is beyond the capability of existed traditional diagnosis methods to process large amount of data with accurate results (Wang and Peter, 2015).

A group of intelligent fault diagnosis methods that can effectively analyze the monitoring data with accurate diagnosis results were presented in the literatures for machinery health diagnosis, mainly including ANNs, SVM, fuzzy inferences, gene expression programming, etc. (Lei *et al.*, 2007; Wang *et al.*, 2010; Widodo and Yang., 2007; Ning *et al.*, 2012; Dong and Zhu, 2012). Combined with various vibration signal processing and feature extraction methods, the intelligent fault diagnosis methods were used widely in the field of fault diagnosis of rotating machinery (Yu *et al.*, 2006; Samantha and Al-Blushi, 2003; Al-Raheem *et al.*, 2007). Yu *et al.* (2006) extracted energy entropy features from EMD components as input vector into ANN to identify roller bearing fault patterns with accurate and effective result. Samantha *et al.* (2003) selected time-frequency domain features to recognize different bearing health pattern with application of ANN and SVM. Tran *et al.* (2009) used statistic features to identify the health conditions of induction motor and then imported the features into adaptive neuro-fuzzy inference systems (ANFIS) to distinguish. Lei *et al.* (2007) proposed two features from multiple sensors signal fusion result to identify health conditions of complex PGBs and ANFIS was applied to recognize various health patterns.

There are two stages required for the application of intelligent fault diagnosis methods on rotating machinery, feature extraction from monitoring signals involving time series signal processing procedure and fault classification using extracted features with implementation of artificial intelligence methods. It can be observed from previously reported literatures on intelligent rotating machinery fault diagnosis that ANN is one of the most commonly used techniques to classify and recognize rotating machinery health conditions. Most reported ANNs implemented in intelligent rotating machinery fault diagnosis are back propagation (BP) networks. However, BP networks have the potential danger of being captured by local minima (Graupe,

2007). Also, in comparison to other ANNs, BP network takes more computation time to train the fitted model. In addition, the disadvantage of being a network with shallow structure limits BP networks to seek hidden relationship and pattern among input feature vector (Bin *et al.*, 2012, Kankar *et al.*, 2011). Moreover, the features imported into network are extracted and selected from the collected data ahead of being classified. The manually extracted features largely depend on diagnosis scenarios and unique to the specific issues. With development of artificial intelligence technique, the limitation of traditional fault diagnosis using neural networks with shallow structure can be recovered by introducing deep learning method.

2.1.2 Planetary Gearbox Fault Diagnosis

The rising deployment of commercial wind energy projects increases the demands of a more proactive maintenance strategy and a more intelligent methodology so that the wind energy can be more competitive to the traditional energy forms. The modern wind industry has experienced a higher wind turbine gearbox failure due to the various reasons such as: unexpected overloading caused by rare operating conditions, defective gearbox design, and ineffective maintenance (Feng *et al.*, 2012).

With benefit from their compressive dynamic structure and unique ability to distribute the input torque in parallel, the planetary gearboxes (PGBs) are widely used in the design of wind turbine drivetrain systems. As reported by Sheng *et al.* (2011), the gearbox failure is the first leading cause (22%) to total wind turbine downtime. Furthermore, the gear failure constitutes 26 and 25% of total gearbox failures recorded in year 2013 and 2014, respectively (Sheng, 2014). Condition monitoring for wind turbine has been studied over the past decade (Zaher *et al.*, 2009). Condition monitoring data of wind turbines is increasingly important as the dimension and remote locations of wind turbines used nowadays lead to crucial technical availability of wind turbines (Zhang and Wang, 2014). Most wind turbines are installed with supervisory control and data acquisition (SCADA) system monitoring critical components such as PGB. The effective formats of measurement are collected and stored via SCADA systems (Becker and Poste, 2006). Though SCADA has been reported as cost-efficient and beneficial for irregular component behavior, it is difficult to be applied due to the large volume of SCADA data generated in the monitoring process (Dao *et al.*, 2017).

The monitoring data can rapidly accumulate to create large and unmanageable volumes that can potentially lead critical faults to be undetectable. These massive real-time data are characterized with large volume, diversity, and high velocity. Typically, there are two categories of fault diagnosis methods, namely model-based methods and data-driven methods. The comprehensive physical and mathematical model of studied system is essential to model-based methods and normally is not available for the wind turbine PGB. With continuously monitoring and regularly stored monitoring data, data-driven methods are preferred (Laouti *et al.*, 2011). Different methods have been developed for gearbox condition monitoring and fault diagnosis. A general procedure of successful fault diagnosis normally constitutes by three stages: selecting a measurement to reflect machinery information, extracting the fault sensitive features, and categorizing the health conditions. Various formats of measurement have been proven effective in fault diagnosis, including vibration signals (Chen and Feng, 2016), oil debris (Li *et al.*, 2014), acoustic emission signals (Yoon and He, 2015; Zhang *et al.*, 2017), thermal signals (Wong *et al.*, 2010), and strain signals (Yoon *et al.*, 2015).

Most available condition monitoring relies on vibration signals analysis using accelerators, combined with complicated signal processing in both time and frequency domain (Lei *et al.*, 2014). As pointed out by McName (2002), continuous-time Fourier analysis explains the natural occurrence of asymmetry in the spectrum that should not alarm the fault. The estimation of local scaling exponent of monitoring vibration data has been successfully used to identify faulty PGB with cracks (Loutridis, 2008). Time synchronously averaged vibration data from helicopter gearbox triaxial vibration data has been proved carrying supplementary information in frequency domain and dynamics of the vibration (Tumer *et al.* 2003). The effects of unequal planet gear spacing on PGB vibration systems have been studied to enhance the understanding of vibration characteristics of PGBs (Patrick *et al.*, 2012). The results have shown the potential of using relative size of the nondominant sidebands magnitude to the dominant sidebands magnitude as the feature to detect carrier-plate crack fault in PGB. The different features sensitive to fault detection have been proposed and applied by researchers. The features extracted from vibration signals are generally located in the time domain, frequency domain, and time-frequency domain. A new time domain feature was introduced

to monitor PGBs in time-varying working conditions (Bartelmus and Zimroz, 2009). Cyclostationarity was applied as condition indicator to diagnose gear faults in the time domain. Information in frequency domain can be very sensitive to the fault patterns (Raad and Sidahmed, 2008). Morlet wavelet was introduced as a filter to effectively remove the noise in the vibration signals by Lin and Zuo (Lin and Zuo, 2004). The period impulse of the filtered signals was then obtained by the degree of cyclostationary as the feature to diagnose a gear fault. In the study reported by Lu *et al.* (2015), the detection of weak nonstationary signal in the monitoring vibration signal was enhanced by introducing time-delayed feedback into stochastic resonance. Such detection can be extensively used in gearbox fault diagnosis. As pointed out (Randall *et al.*, 2011), spectral kurtosis is one of the most used frequency domain fault features. Compared to the features extracted only from time domain or frequency domain, the features from time-frequency domain have attracted much attention in recent researches. It was shown in the research that the features extracted based on second order transient analysis with particle swarm optimized adaptive band filtering were more sensitive to the initial fault development (Hussain, 2017). Complicated signal processing techniques including continuous wavelet transform (Zuo *et al.*, 2015), discrete wavelet transform (Bairis *et al.*, 2016), wavelet packet transform (Shen *et al.*, 2013), Hilbert-Huang transform (Yoon *et al.*, 2016), adaptive optimal kernel time-frequency analysis (Feng & Liang, 2014), empirical wavelet transform (Chen *et al.*, 2016) and other time-frequency tools have been successfully applied to extract fault features from monitoring signals. The combination of various signal processing methods has been applied on wind turbine PGB fault diagnosis. Study has shown the effectiveness of ensemble intrinsic time-scale decomposition, wavelet packet transform, and correlation dimension combination on PGB fault diagnosis (Hu *et al.*, 2015). However, the application results can only identify the faults on the high-speed shaft gear. Furthermore, the researchers have studied the performance of combined features extracted from multiple domains in application of fault diagnosis. Two diagnosis parameters generated from the features extracted in both time and frequency domains were applied as condition indicator to diagnose PGB faults (Lei *et al.*, 2012).

After the features are extracted from the monitoring signals, a variety of classifiers can be used to categorize the health conditions. The most reported algorithms used as health condition classifiers include support vector machines (SVMs) and artificial neural networks (ANNs). The wind turbine blade damage classification abilities of SVMs and logistic regression as two typical supervised learning methods have been studied (Regan *et al.*, 2017). Both two methods can identify the wind turbine blade faults with minimal overfitting. An intelligent fault diagnosis method combined with envelope spectrum and SVM was proposed for fault diagnosis (Guo *et al.*, 2009). An intelligent model for gearbox fault diagnosis based on SVM and immune genetic algorithm (IGA) was generated (Chen *et al.*, 2013). The IGA was applied to select the optimal parameters for SVM used in the proposed model for better classification accuracy. It has been reported that a dynamic neural network was applied for intelligent fault diagnosis with success (Tayarani-Bathaie *et al.*, 2014). An ANN was applied with the features extracted from empirical mode decomposition of vibration signals to accomplish automatic fault diagnosis (Ali *et al.*, 2015). Manifold learning and Shannon wavelet SVM was combined to recognize the faults in wind turbine transmission system (Tang *et al.*, 2014). A strategy was proposed to detect and isolate faults for wind turbine benchmark (de Bessa *et al.*, 2016). The strategy detects the change in the monitoring time series data and applies fuzzy Bayesian network to generate the possibility of potential faults.

The above-mentioned condition monitoring and fault diagnosis methods face three challenges in the age of big data. Firstly, the features used for classification are manually extracted. This feature extracting process depends greatly on complicated signal processing techniques and fault diagnosis expertise. Such characteristic constrained the extensibility and capability of the existing methods on large volume of data, since human intervention is required to accomplish the intelligent process. Secondly, though the manually selected and extracted features have been proven effective to express fault information, the features are selected on specific operational condition or components. Thus, the feature extraction and selection process need to be adjusted regarding each fault diagnosis system with different components and environment, requiring the experienced professional to accomplish. The third challenge is that widely used ANNs in the

existing studies were constituted by shallow structure of single hidden layer. Such shallow structure limits the capacity of ANNs to adaptively self-learn and mine the hidden information regarding the fault diagnosis issues. In the age of Internet of Things and Industrial 4.0, massive data will be generated by condition monitoring systems. There is an urgent need to develop an intelligent condition monitoring and fault diagnosis being capable of automatic feature extraction and fault diagnosis.

2.2 Deep Learning Based Methodology and Application in Fault Diagnosis

2.2.1 Deep Learning – A Brief Introduction

Deep learning has been widely studied in various fields for its unique capability of adaptively capturing the representative information from raw data through multiple non-linear transformations and approximating complex non-linear functions with a small error (Schmidhuber,2015). Deep learning refers to a class of machine learning techniques where many layers of information processing stages in deep architectures are exploited for pattern classification and other tasks (Deng, 2014). Development of deep learning based approaches for mechanical fault diagnosis has been reported (Tamilselvam and Wang, 2013; Tran *et al.*, 2014; Ma *et al.*, 2014; Lv *et al.*, 2016; Jia *et al.*,2016; Gan *et al.*, 2016). Tamilselvan and Wang (2013) reported a multi-sensor method using DBN for chemical benchmark classification, aircraft engine health diagnosis, and electric power transformer health diagnosis. Tran *et al.* (2014) implemented DBN on two-stage reciprocating compressor fault diagnosis under different valve conditions. The validation results showed the efficiency of the reported method with high accuracy on diagnosis. Ma *et al.* (2014) implemented a deep learning based approach in a multi-sourced big data environment. The validation results obtained from power transformer and circuit breaker showed the capability of deep learning on accurate classification. Lv *et al.* (2016) applied a deep learning based approach on distributed control system fault detection. The validation results showed that the proposed method can capture the features that could not be obtained by statistic techniques.

Recently, several studies on rotating machinery fault diagnosis using deep learning based methods have been reported. Jia *et al.* (2016) proposed an intelligent diagnosis method by using deep neural networks (DNN) on vibration bearing monitoring data along with simple signal processing technique as fast Fourier transfer (FFT). DNN is composed by stacking multiple autoencoders. The validation results showed that the DNN can extract fault features from monitoring signals that led to a satisfactory diagnosis accuracy. The method reported by Jia *et al.* (2016) relies heavily on supervised fine tuning to determine the appropriate parameters of designed DNN, including the number of hidden layers and the number of neurons in each layer. Gan *et al.* (2016) constructed a hierarchical diagnosis network (HDN) by stacking DBN for bearing fault diagnosis. The first DBM was employed to identify the fault locations, the then second DBN was employed to identify the fault severity for the specific fault location. However, features that were manually extracted from the monitoring signals by using wavelet packet energy method were required. Besides, the reported HDN is trained independently and sequentially for fault diagnosis and severity detection, which will potentially reduce the reliability of the diagnosis and detection results. Though DNN and DBN based deep learning approaches have been implemented in rotating machine fault diagnosis successfully, determination of significant parameters of designed model depends greatly on trial-and-error in precisely supervised fine-tuning procedure. Both time and experience on fine tuning procedure is required to achieve the goal as satisfied accurate performance. A method that does not require supervised fine tuning to determine model structure is needed for automated intelligent fault diagnosis with big data.

A deep learning based intelligent rotating machinery fault diagnosis methodology is needed to simultaneously perform fault location and severity diagnosis without great dependency on complicated signal processing or supervised fine tuning. Large memory storage retrieval (LAMSTAR) neural network is selected as deep architecture in this research. The LAMSTAR neural network can be regarded as an ANN with deep architecture for the composition of multiple parallel SOM modules. With the similar structure to the typical deep learning, LAMSTAR neural network is potential to overcome the deficiencies in current intelligent diagnosis methods. LAMSTAR neural network has been implemented in multiple fields including image

recognition (Girado, 2013), biomedical diagnosis (Nigam and Graupe, 2004; Sivaramakrishna and Graupe, 2004; Waxman *et al.*, 2010; Isola *et al.*, 2012), finance analysis (Dong *et al.*, 2012), with solid results showing the capability of LAMSTAR network for rapidly processing large volume of data with less error than regular machine learning algorithms. By introducing the forgetting, rewarding and punishing features into a traditional neural network, LAMSTAR network works on a closer level of mimicking the working process of a natural brain. LAMSTAR network can grow and shrink in dimension without changing the original structure and maintain fast training speed due to the operation of parallel SOM modules. The standard perceptron-like neurons are employed in LAMSTAR network, arranging in SOM modules. The SOM modules in LAMSTAR network are governed by winner-taking-all (WTA) strategy and the memories of those neurons in SOM are stored in bidirectional associative memory function. The optimization of a LAMSTAR network is to determine the link weights that store relation message between various input SOM modules and between neurons in input SOM modules and decision SOM module. The link weights contribute along with information stored in winner neuron from each individual input SOM module to decision SOM module. With all the characteristics discussed above, a LAMSTAR network can learn and understand system information more systemically and intelligently. Yoon *et al.* (2013) have successfully applied a LAMSTAR network on vibration signals for full ceramic bearing fault diagnosis. In that study, vibration signals were pre-processed by using HHT to extract conditional indicators as inputs to BP neural network, LAMSTAR network, and KNN model, respectively. The classification results from that study showed faster learning speed and higher accuracy obtained by the LAMSTAR network.

The proposed LAMSTAR network based rotating machinery fault diagnosis and severity detection approach pre-processes monitoring signals using the simple short-term Fourier-transform (STFT) method rather than computationally complicated signal processing and feature extraction algorithms such as wavelet transform and HHT. After transforming the raw vibration signals to generate a time spectrum matrix using STFT, sub-patterns are then generated from the time spectrum matrix and used to obtain the optimized LAMSTAR network for bearing fault diagnosis and severity detection simultaneously.

2.2.2 Sparse Coding and Application in Fault Diagnosis

There is a growing interest in implementation of sparse coding in signal processing field. Initially, the idea of redundancy reduction was raised by Barlow (1961) as the fundamental principle underlying the primary processing in mammalian neural circuits for perception. Based on the extension of redundancy reduction to sparse coding by Foldiak (1990), Olshausen and Field (1996) declared that basic functions sharing similar properties with the neurons in the primary visual cortex of mammals were learned out from natural images. The basic model suggests that natural signals can be represented, or efficiently approximated, by a linear combination of pre-learned atom signals with sparse linear coefficients, meaning most of linear coefficients are zero. With the capability of high level abstract presentation, the sparse coding is becoming widely studied in many fundamental signal and image processing tasks (Long *et al.*, 2016; Hu *et al.*, 2016; Elad and Aharon, 2006; Potter and Elad, 2009; Zibulevsky and Pearlmutter, 2001; Chen *et al.*, 2014; Liao and Sapiro, 2008; Wrights, *et al.*, 2009). Generally, given $x \in R^N$ as a column signal, where N represents the number of data points in column signal, and arranging the atom signals as the columns of the dictionary $D \in R^{N \times L}$, the sparsity assumption is described by the following sparse approximation problem. Assume that a sparse solution exists such that

$$\hat{s} = \text{Arg} \min_s \|s\|_0^0 \quad \text{Subject to } \|x - Ds\|_2 < \epsilon \quad (2.1)$$

In Eq. (2.1), \hat{s} is the sparse representation of x and γ the error tolerance. The function $\|*\|_0^0$ is referred to as the l^0 - norm that counts the nonzero entries of a vector. Though known to be an NP-hard problem in general (Davis *et al.*, 1997), there are a variety of methods to provide approximation on this problem (Donoho *et al.*; 2012, Sardy *et al.*, 2000).

A critical issue in using sparse coding model for signal processing is the selection and generation of dictionary D . Two major methods were reported in the literature, namely the analytic method and the machine learning based method. For the analytic approach, a mathematical model of the data is formulated, and an

analytic construction is generated to efficiently represent the model. Dictionaries developed using the analytic approach are referred to as implicit dictionaries due to how they are described by algorithm. Typical analytical approach based dictionaries include curvelets (Gorodnitsky and Rao, 1997), shearlets (Labate *et al.*, 2005), wavelets (Mallat, 2009), complex wavelets (Selesnick, *et al.*, 2005) and so on. For the machine learning based approach, the dictionaries are generated using machine learning technique from a set of examples. Different from analytical approach based dictionaries, machine learning based dictionaries are represented as an explicit matrix, and a training algorithm is required to adapt the matrix coefficients to the examples. Typical machine learning algorithms have been combined into sparse coding models for dictionaries generation, including principal component analysis (Jolliffe, 2002), method of optimal directions (Engan *et al.*, 1999), and K-SVD (Aharon *et al.*, 2006). Though the fine-tuned dictionaries from machine learning based methods provide significantly better performance in sparse coding model application, the production of unstructured dictionaries leads to more cost for application. Also, the complexity of typical machine learning methods limits the size of dictionaries and the dimension of signals to be processed (Rubinsein *et al.*, 2010).

Though many sparse coding based signal processing studies have been reported, only a few are investigated for machinery fault diagnosis. Martin-del-Campo and Snadin (2015) studied the possibility to automate the condition monitoring process by continuously learning a dictionary of optimized shift-invariant feature vector. The features extracted from vibration signals carrying ball bearing condition information show the obvious difference in shape and frequency in different bearing condition signals under various operation conditions. Such features can be useful for identify the bearing conditions. Liu *et al.* (2011) extracted features from bearing vibration signals by using shift-invariant sparse coding method, and classified the bearing condition through a linear discriminant analysis classifier. The classification error rates from validation results ranges between 0% to 77.31%. Li *et al.* (2015) presented a rolling bearing fault diagnosis method based on locality-constrained sparse coding. The basic time domain features, including mean, median, variance, skewness, min value, max value, kurtosis value, and other frequently used time domain features, were extracted and used for dictionary generation and training. A support vector machine classifier is selected

to classify the bearing conditions with extracted sparse features obtained from sparse coding model. The vibration signals under 0 hp load was selected for validation. The validation results showed good classification performance with highest classification error as 2.54%. Li *et al.* (2014) presented the bearing fault diagnosis using Fisher discrimination sparse coding methods. The feature extraction method added the Fisher discriminant criterion into sparse coding framework to extract more discriminative and conductive features for classification. The validation results involving vibration signals carrying different bearing faults with mixed loading showed classification error ranging from 0% to 3.63%. It should be noted that analytical approach was selected for dictionary learning process in this reference to realize sparse coding based methodology on machinery fault diagnosis. Also, only seeded vibration bearing fault monitoring data was used for validation.

Given the limitation of current machine learning based dictionary learning approach and its significant computational effort, a LAMSTAR neural network can be introduced into sparse coding frame as well as classification stage. It is expected that such combination will increase the classification accuracy and reduce the computational effort.

It is also observed that only shallow classifiers were employed on the extracted sparse features for bearing fault diagnosis (Liu *et al.*, 2011; Li *et al.*, 2015; Li *et al.*, 2014). Classifiers including support vector machine (SVM) (Liu *et al.*, 2011; Li *et al.*, 2014) and linear discriminate analysis (LDA) (Li *et al.*, 2015) were reported with classification error rate in range of 0% to 6%, when training dataset and testing dataset were formed under same operation condition. The training datasets were designed as half of testing datasets in the above literatures. Such design requires large ratio of labeled data for satisfied classification results when applying the sparse coding based models. However, the reported sparse feature extracted using sparse coding based method did not give a good adaptive performance. The classification error in the model with training and testing datasets obtained under different operation conditions can be as high as 77%. The capability of deep learning based methods on obtaining connection between various input dataset and digging

hidden information in the pattern is a potential solution to the adaptive characteristics on sparse feature application. More specifically, the connection between faults and their severity will be generated and gathered in LAMSTAR neural network through extracted sparse features under various operation conditions, and lead to accurate classification results. It has been proven shown that the application of sparse coding as a deep learning method efficiently increased the robustness and adaptiveness of the models (Xavier, *et al.*, 2011; Ranzato, *et al.*, 2008; Lee, *et al.*, 2008). Similarly, it can be expected to introduce sparse coding method into LAMSTAR neural network with low risk of overfitting. The amount of labeled data required in supervised training procedure can be reduced. Thus, with small portion of labeled data, the accurate machinery fault diagnosis can be achieved automatically.

2.3 Signal Processing Techniques on Rotating Machinery Fault Diagnosis

2.3.1 Signal Processing Techniques on Bearing Fault Diagnosis

Signal processing is an important task for machine fault diagnosis. Over the years, many vibration signal processing methods have been developed for bearing fault diagnosis. In general, they can be classified into two major categories: vibration analysis based and deep learning based. vibration analysis based signal processing techniques includes: time-domain analysis, frequency-domain analysis, and time-frequency domain analysis (Jardine, *et al.*, 2016).

Since the vibration signals are collected as time series data, time-domain analysis is applied to the time waveform directly. Typical time-domain analysis calculates the statistical condition indicators from the time waveform signals that can describe the uniqueness of the signals. Commonly used descriptive statistics include mean, peak, peak-to-peak interval, standard deviation. Simple condition indicators such as root-mean-square (RMS) and crest factor have been applied for the detection of localized bearing defects (Miyachi and Seki, 1986; Tandon, 1994), with limited success. Some high-order statistical parameters such as probability density and kurtosis have been used for bearing defect detection (Cikkacittm, 1977, Dyer and Stewart, 1978). Apart from the condition indicators calculated directly from the time-domain, display of

waveform data on oscilloscope or on a chart recorder for periodic peaks observation can locate the bearing local defects (Igarashi and Hamada, 1982; Igarashi and Kato, 1985). The shock pulse method detecting the structural resonances existed in the high frequency zone has been widely studied for the bearing fault detection (Butler, 1973; Li *et al.*, 2007). Another popular studied time-domain analysis approach is time synchronous average (TSA). TSA can extract the periodic waveform information from the noisy data by the ensemble average of the raw signal over multiple evolutions. TSA has been successfully applied to bearing local faults diagnosis (Siegel *et al.*, 2012; McFadden and Toozy, 2000; Christian *et al.*, 2007). Other time-domain waveform vibration analysis methods for bearing fault diagnosis are briefly mentioned as the followings: time series model application (Polyhonen *et al.*, 2004), modified least mean square for non-stationary signals (Zhuge and Lu, 1991), principle component analysis (PCA) (Baydar *et al.*, 2001).

Frequency-domain analysis is the most widely used methods on machinery fault, including the bearing fault diagnosis (Tandon and Choudhury, 1999). As the frequency-domain analysis can easily isolate and identify certain interested frequency components it has certain advantages over regular time-domain analysis. The typical spectrum analysis is the fast Fourier transform (FFT). FFT can obtain narrowband spectra more efficiently. In case of bearing fault diagnosis, both the high and low ranges of frequency spectra components are useful. Particularly, the defect in rolling element bearing produces short duration pulse during operation, resulting in the high energy in a specific range of frequency band. TSA in frequency domain has been applied to extract bearing fault features (Van Hecke *et al.*, 2014; Van Hecke *et al.*, 2016). Power spectrum using the result of FFT is a commonly used tool for bearing fault diagnosis. Other efficient supplementary tools to analyze the spectrum include visualization of spectrum, frequency filters, envelope analysis, side band analysis (Ho and Randall, 2000; Randall *et al.*, 2001; Stack *et al.*, 2004; Blankenship and Singh, 1995). Cepstrum analysis related methods that have some advantages over the FFT based methods on certain cases include: power cepstrum (Van der Mervwe and Hoffman, 2002), bispectrum (Yang *et al.*, 2002), and holospectrum (Qu *et al.*, 1996). Even the application of cepstrum analysis related methods have been

successfully applied to diagnosing bearing faults, the estimation of power spectrum is complicated with certain degrees of inaccuracies (Haves, 1996; Slami *et al.*, 2002).

Time-frequency domain analysis is used to extend the capability of frequency domain analysis on non-stationary waveform signals. Time-frequency analysis methods such as short-term Fourier transform (STFT) (Wang and McFadden, 1993; Andrade *et al.*, 1999), Wigner-Ville distribution (Meng and Qu, 1991), and wavelet transform (WT) (Luo *et al.*, 2003) can process the non-stationary signals for machinery fault diagnosis. Other widely studied time-frequency analysis methods based on WT have shown their effectiveness on identifying bearing faults (Yan *et al.*, 2014). However, the non-reachable simultaneous best time and frequency resolution of STFT, the cross-term interface of Wigner-Ville distribution, the energy leakage issue and precise base function selection of wavelet transform (Yan *et al.*, 2014; Peng *et al.*, 2005) limit their application to bearing fault diagnosis with better results. As a powerful time-frequency analysis method applied to rotating machinery fault diagnosis, empirical mode decomposition (EMD) of Hilbert-Huang transform (HHT) decomposes the signal without base function like wavelet transform and can be applied to both stationary and nonstationary signals (Lei *et al.*, 2013). Though HHT related methods have been implemented with solid results, it suffers from the computational expense and complicated results with changing working conditions (Liu *et al.*, 2014; Tse *et al.*, 2001; Yam *et al.*, 2009; Peng *et al.*, 2005).

2.3.2 Signal Processing Techniques on Gear Fault Diagnosis

As a critical component in the modern manufacturing systems, prediction of remaining useful life (RUL) has been used to make condition-based maintenance decision to avoid disastrous events and extend machine life (Sikorska *et al.*, 2011; Wang and Shen, 2016). In comparison with traditional physical-model based top-down modeling, the data-driven machine health monitoring methods provide the bottom-up solution for both fault diagnosis and prognostics to be easily accessed without degradation model describing the nature of failure (Jardine *et al.*, 2006). The conventional data-driven prognostic methods include three main stages: design of effective condition indicators (CI), CIs extraction and selection, and prognostics using manually extracted features. The satisfactory prediction of machine future working condition relies heavily

on the effective CIs and appropriate CIs selection (Javd, 2014). Over the years, many vibration signal processing methods have been developed and applied on effective CIs extraction for machine health monitoring, mainly include time-domain analysis, frequency-domain analysis, and time-frequency-domain analysis (Jardine *et al.*, 2016). Since the vibration signals are collected as time series data, time-domain analysis is applied to the time waveform directly. Typical time-domain analysis calculates the statistical condition indicators from the time waveform signals that can describe the uniqueness of the signals. Commonly used descriptive statistics include mean, peak, peak-to-peak interval, standard deviation. Another popular studied time-domain analysis method is time synchronous average (TSA). By the ensemble average of the raw signal over multiple revolutions, TSA can extract the periodic waveform information from the noisy data. TSA has been proved effective on locating gear faults (Qu *et al.*, 2014; Yoon *et al.*, 2015; Dapliaz *et al.*, 2000). Frequency-domain analysis is the most widely used methods on machinery fault, including the bearing fault diagnosis (Tandon and Chouldhury, 1999). The typical spectrum analysis is the fast Fourier transform (FFT). As the frequency-domain analysis can easily isolate and identify certain interested frequency components it has certain advantages over regular time-domain analysis. Bispectrum as the Fourier transform of the third statistics of the time waveform has been studied in locating gear faults (Xiong *et al.*, 2002; Zhang *et al.*, 2003). Time-frequency domain analysis is used to extend the capability of frequency domain analysis on non-stationary waveform signals. Time-frequency analysis methods such as short-term Fourier transform (STFT) (Wang and McFadden, 1993; Andrade *et al.*, 1999), Wigner-Ville distribution (Meng and Qu, 1991), and wavelet transform (WT) (Luo *et al.*, 2003) can process the non-stationary signals for machinery fault diagnosis. The processes of designing, extracting and selecting appropriate CIs for the prediction of target machine working condition require human labor and expertise on both signal processing and prognostics. They have been proved effective on gear prognostics successfully but still with limitations on certain constrains and complicated designed signal processing procedures. Besides, the improvement of gear prognostics performance using the manually extracted features yields to a higher computational cost

(Heydarzadeh *et al.*, 2016; Sharma and Parey, 2015). Therefore, it is difficult to apply the conventional data-driven methods to automatically process machine big data.

3. Limitations of the Current Deep Learning Based Fault Diagnosis Methods

(Parts of the literature review in this chapter were previously published as (1) He, M., He, D., 2018, “Wind turbine planetary gearbox feature extraction and fault diagnosis using a deep learning based approach”, *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, © 2018 (SAGE Publications), Reprinted by permission of SAGE Publications. DOI: 10.1177/1748006X18768701; (2) He, M., He, D., 2018, “Simultaneous bearing fault diagnosis and severity detection using a LAMSTAR network-based approach”, *IET Science, Measurement & Technology*, DOI: 10.1049/iet-smt.2017.0528;)

Even with the success of deep learning for bearing fault diagnosis, deep learning based approaches suffer the following limitations.

(1). The structure of developed neural network cannot be explained with signal related physical meaning to the date. Most of deep learning based machinery fault diagnosis design the structure by trail-and-error test without explicit instruction. The commonly used constrain in neural network such as weight decay and sparsity coefficient cannot force the network to learn the unique properties of signal. The connection stated by Lei *et al.* (2016) is concluded by observing the trained filter and Gabor filter without detailed explanation. The orthonormality constraint introduced by Jia *et al.* (2018) lacks the explicit physical explanation. Thus, a deep structure guided with vibration analysis techniques will be useful.

(2). Most of the deep learning based approach cannot be used directly on the periodic input. As a typical rotating machinery component, bearing signals contain shift variant information. The features learnt by the reported deep learning based approach with shift variant properties could lead to the misclassification.

(3). For the deep learning based prognostics methods, the features are pre-extracted from raw data to be used in the following deep learning model for prognostics. Thus, the final performance of the deep learning based models still relies on the signal processing technique and features selection. The mechanism of using deep learning algorithms to express the monitoring data abstractly has not been explored yet.

4. METHODOLOGY

(The majority of the content in this chapter is composed of previously published work as (1) He, M., He, D., 2018, “Wind turbine planetary gearbox feature extraction and fault diagnosis using a deep learning based approach”, *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, © 2018 (SAGE Publications), Reprinted by permission of SAGE Publications. DOI: 10.1177/1748006X18768701; (2) He, M., He, D., 2018, “Simultaneous bearing fault diagnosis and severity detection using a LAMSTAR network-based approach”, *IET Science, Measurement & Technology*, DOI: 10.1049/iet-smt.2017.0528; (3) He, M and He, D., 2017, “A deep learning based approach for bearing fault diagnosis”, *IEEE Transactions on Industrial Application*, Vol. 53, No. 3, pp. 1-9, DOI: 10.1109/TIA.2017.2661250)

4.1 Overview of the LAMSTAR Neural Network Based Rotating Machinery Fault Diagnosis Methodology

The general procedure of the LAMSTAR neural network based approach for rotating machinery fault diagnosis and severity detection is shown in Figure 4.1. As shown in Figure 4.1, the presented method includes three major steps. The first step is to obtain a time-spectrum matrix S using STFT to pre-process the monitoring raw data. The motivation behind such pre-processing is that frequency spectrum of rotating machinery shows how their constitutive components are distributed with discrete frequencies and potentially provide valuable information about machinery health conditions (Jia *et al.*, 2016). The second step is to import the obtained time-spectrum matrix S into a LAMSTAR network structure to obtain optimized deep learning model for fault diagnosis. Specially, sub-patterns are generated from the time-spectrum matrix S and channeled into the optimal LAMSTAR network model for rotating machinery fault diagnosis. The final step is to employ the optimized LAMSTAR network model to diagnose the bearing faults and detect the fault severity. The basic LAMSTAR network structure and specific design of the LAMSTAR network model are explained next.

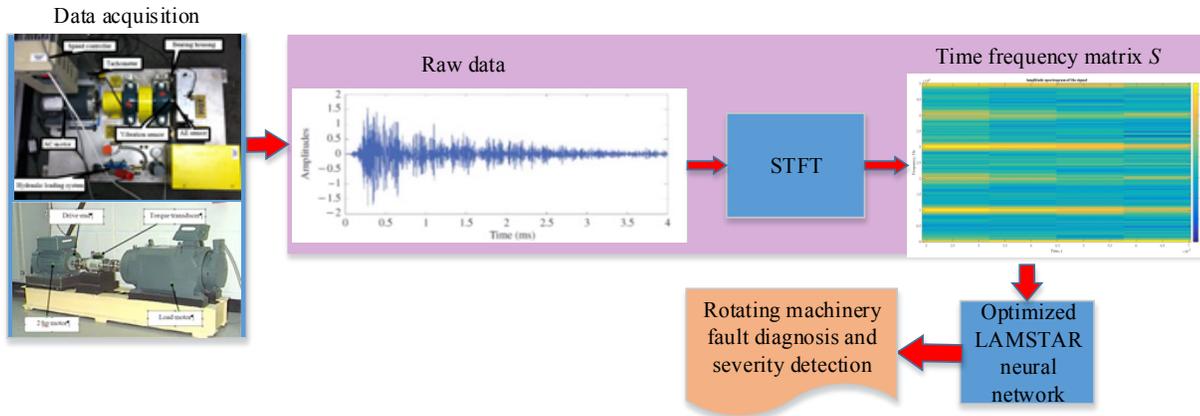


Figure 4.1. General structure of the LAMSTAR based method for bearing fault diagnosis and severity detection

4.2 Basic LAMSTAR Network Structure

LAMSTAR network is specifically designed for retrieval, classification, prediction and decision making problems, especially those involving large amount of data. The LAMSTAR network is built with one input layer, multiple input SOM modules and one decision SOM module. Two types of SOM modules are employed with different purposes, including multiple input SOM modules as hidden layers and one decision SOM module as output layer. With advantages of SOM-based network and statistical decision tools, LAMSTAR network is capable of pattern storage and retrieval. The information in the network is continuously being ranked for each data sample through learning and correlation, to make LAMSTAR network an intelligent system.

The decision-making principle in a LAMSTAR network is the same as the classical neural networks. If n denotes the number of inputs fed into j^{th} neuron as $\{v_{ij}, i = 1, 2, \dots, n\}$, then output y_j of the j^{th} neuron can be expressed as:

$$y_j = f_N \left(\sum_{i=1}^n w_{ij} v_{ij} \right) \quad (4.1)$$

where $f_N(\cdot)$ represents nonlinear activation function. Variable w_{ij} are the weights assigned to the i^{th} inputs of j^{th} neuron and whose setting is the learning action of the LAMSTAR network. The information in LAMSTAR network is saved and then processed through correlation links between individual neurons in separate SOM modules. Given a coded real matrix \mathbf{X} as input pattern:

$$\mathbf{X} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \mathbf{x}_3^T, \dots, \mathbf{x}_N^T] \quad (4.2)$$

where \mathbf{x}_i^T stands for transpose of sub-pattern \mathbf{x}_i . Each sub-pattern \mathbf{x}_i is transported to a corresponding i^{th} input SOM module storing the corresponding data of the i^{th} category in the input pattern. A general structure of LAMSTAR network is presented in Figure 4.2. In Figure 4.2, input pattern represents each signal to be diagnosed, containing all sub-patterns generated from the spectrum matrix S . LAMSTAR network does not create neurons for an input pattern. Instead, only individual sub-patterns are stored in input SOM modules, and correlations between sub-patterns are stored as link weights (Graupe, 2013). The sub-pattern construction is explained in detail in Section 4.3. The procedure for generating and adjusting the link weights in the LAMSTAR network is explained next.

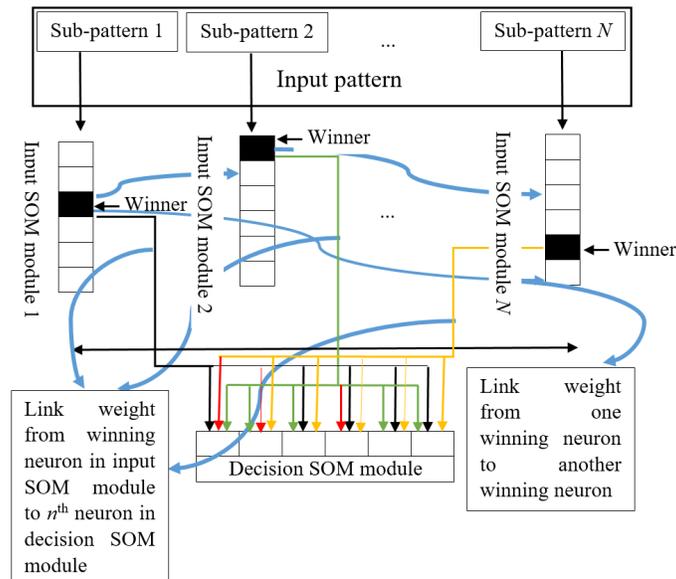


Figure 4.2. General structure of the LAMSTAR network

When a new input pattern is loaded into a LAMSTAR network, it checks all the storage-weight w_i in the i^{th} input SOM module corresponding to the i^{th} sub-pattern to be stored. A neuron will be claimed as the winning neuron for this particular input sub-pattern if it can match the input sub-pattern within a pre-defined error tolerance.

For most of the applications with storage of purely numerical input sub-patterns, the storage of such sub-patterns into an input SOM module can be simplified by directly mapping each SOM module into a pre-defined range of value. For example, the sub-pattern with a value of 0.6 will be stored in an input neuron representing range from 0.5 to 0.75. The searching procedure and decision of winning neuron in each module for SOM dynamic weights construction is explained in Section 4.3.

The correlation between sub-patterns is stored in link weights that connect neurons in different input SOM modules. Thus, the link weights become fundamental to allow interpolation and exploration of sub-patterns. The link weights are updated as for a given input pattern with the determined winning k^{th} neuron in the i^{th} input SOM module and the winning m^{th} neuron in the j^{th} input SOM module. A winning neuron is determined for each input sub-pattern according to how similar the input sub-pattern and a weight vector

w (stored information). For a given sub-pattern \mathbf{x}_i , the winning neuron is determined by minimizing the distance norm $\|*\|$ as below:

$$d(i, j) = \|\mathbf{x}_i - \mathbf{w}_j\| \leq \|\mathbf{x}_i - \mathbf{w}_k\|, \forall k \neq j \quad (4.3)$$

The vector \mathbf{x}_i will be stored in weights w_{ij} of vector \mathbf{w}_j relating to the j^{th} neuron when the distance satisfies Eq.(4.3). The representing link weight $L_{ij}^{k,m}$ is calculated by adding a pre-defined reward. Meanwhile, all other link weights will be decreased with a predefined punishment. The link weights update can be expressed as followings:

$$L_{i,j}^{k,m}(t+1) = L_{i,j}^{k,m}(t) + \Delta R \quad (4.4)$$

$$L_{i,j}^{k,m}(t+1) = L_{i,j}^{k,m}(t) - \Delta P \quad (4.5)$$

$$L(0) = 0 \quad (4.6)$$

where, in Eq.(4.4) and Eq.(4.5), $L_{i,j}^{k,m}$ donates link weights between the i^{th} in the k^{th} input SOM module and j^{th} winning neuron in the m^{th} input SOM module, ΔR and ΔP are pre-defined reward and punishment values, and t represents the number of the iterations that the link weights are updated. The initial link weight is set as 0. The output that matches with target input will be rewarded by a non-zero increment; otherwise it will be punished by a non-zero decrement.

The decision at the decision SOM module is made based on the collected link weights between neurons in the decision SOM module and neurons in all input SOM modules. To make such a decision, the LAMSTAR network produces a winning decision neuron n from the set of neurons J in the decision SOM module by searching for the neuron in the decision SOM module with the highest cumulative value of link weights connecting to the selected winning neurons in each input SOM modules. The equations to make such a decision for the decision SOM module are given as follows:

$$E(j) = \sum_{\kappa^*}^N L_{\kappa^*}^j, \forall j \in J \quad (4.7)$$

$$E(n) \geq E(j), \forall j \in J \quad (4.8)$$

where k^* donates the winning neuron in every k^{th} input SOM module; N is the total number of input SOM modules, $L_{k^*}^j$ stands for link weight between the winning neuron in k^{th} input SOM module and neuron j in decision SOM module, J is set of neurons in decision SOM module, and n donates the selected winning neuron in the decision SOM module. Thus, $E(j)$ represents the sum of link weights connecting to the j^{th} neuron in decision SOM module from each winning neuron k^* in k^{th} input SOM module, and $E(n)$ represents the sum of link weights connecting to the winning neuron n in the decision SOM from each winning neuron k^* in the k^{th} input SOM module.

In the LAMSTAR network based rotating machinery fault diagnosis method, the training data will be used firstly to update the link weights. More specifically, it will find the winning neuron in each input SOM module and use WTA principle to update the link weights. In testing procedure, the LAMSTAR network will calculate the winning neuron in the decision SOM module serving as the label information. In the end, the accuracy of the fault diagnosis is defined as ratio of the number of successful classifications over the total number of tested data points.

4.3 The Design of the LAMSTAR Network for Rotating Machinery Fault Diagnosis

The design of the LAMSTAR network for rotating machinery fault diagnosis involves the following tasks: sub-pattern generation, input data normalization, dynamic formation of neurons in SOM, determination of the link weights, and neural network test. They are explained detailed in the next sections.

4.3.1 Sub-pattern Generation

The basic storage modules of the LAMSTAR network are SOM modules as discussed previously in Section 4.2. Since the decisions were made based on the link weights existed in both input SOM modules and the ones between input SOM modules and decision module, LAMSTAR network focused on correlation between patterns rather than the patterns themselves..

The collected raw data are firstly pre-processed by STFT. The STFT pre-processing result is a spectrum 2-D matrix with a size of $l \times n$. Given a real-time monitoring signal $s(t)$, the spectrum matrix can be written as $S_{l \times n}$ with a size of $l \times n$, where l denotes row number and n denotes column number. Each signal sample is viewed as one input pattern in this case. By considering the spectrum matrix as a spectrum plot with known elements inside, the sub-sampling method used in the LAMSTAR network based image recognition application (Girado, 2013; Graupe, 2013) can be used to generate sub-patterns. Sub-patterns are generated by taking samples from previously generated spectrum matrix.

For each pattern, data subsets are sampled using a sliding box with a size of $d \times d$ by sliding the spectrum matrix from left to right, then top to bottom, sequentially. Every subset obtained from the sliding box will be transformed column by column into a 1-D vector, taken as one sub-pattern in an input SOM module. The procedure for obtaining sub-patterns from spectrum matrix is presented in Figure 4.3.

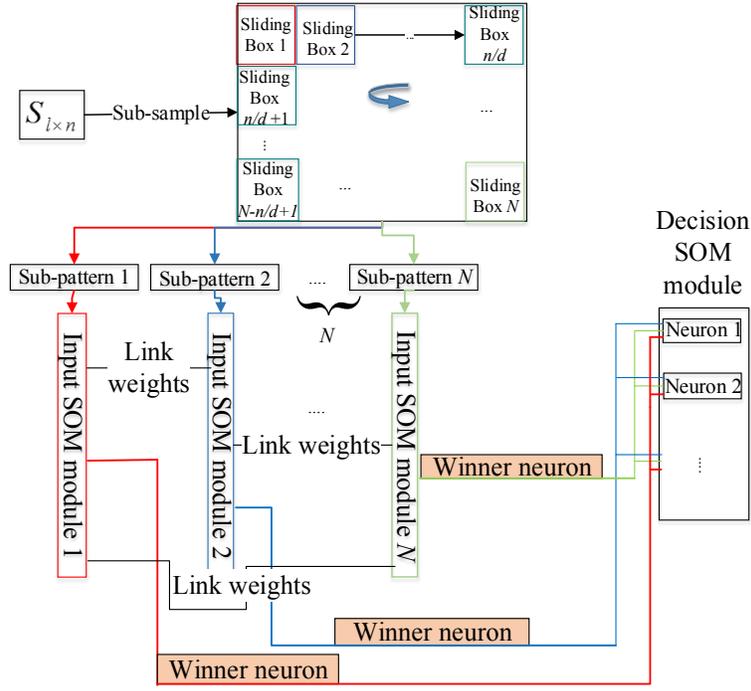


Figure 4.3. The procedure for obtaining each sub-pattern from spectrum matrix as input to LAMSTAR network.

With a defined size of sliding box as $B_{d \times d}$, the total number of generated sub-patterns N is decided as:

$$N = (l/d) \times (n/d) \quad (4.9)$$

For each sub-pattern, the number of elements is d^2 . Since each spectrum matrix $S_{l \times n}$ is considered as one pattern, then one can define a sub-pattern as Ω_v , $v = 1, 2, \dots, N$. Thus, a sub-pattern can be written as:

$$\Omega_v(f, g) = S_{l \times n} \left[\left[\frac{N}{n} \right] \times f, \text{mod} \left(\frac{l}{d} \right) \times g \right], \forall f, g < d \quad (4.10)$$

The data sampled from the spectrum matrix is stored in sub-patterns, and each sub-pattern will be modeled as one input SOM module. Thus, the number of input SOM modules is the same as the number of

generated sub-patterns. The goal of determining an appropriate sliding box dimension is to select a sub-pattern that contains as much variant information as possible and thus can be used as a feature. Therefore, the entropy of elements per sliding box can be used as one criterion when selecting the appropriate sliding box dimension. In rotating machinery fault diagnosis, each fault will be represented by an output neuron firing sequence in the decision SOM module. Hence, the decision SOM module should contain enough number of output neurons such that a complete permutation of the output neuron firing sequences can be used to represent the rotating machine conditions to be classified.

4.3.2 Input Normalization

For each sub-pattern $\mathbf{x}_i = [x_1, x_2, \dots, x_j, \dots]$, the normalization of \mathbf{x}_i is computed as:

$$\mathbf{x}_i^{norm} = \frac{\mathbf{x}_i}{\sqrt{\sum_j x_j^2}} = \left[\frac{x_1}{\sqrt{\sum_j x_j^2}}, \frac{x_2}{\sqrt{\sum_j x_j^2}}, \dots, \frac{x_j}{\sqrt{\sum_j x_j^2}}, \dots \right] \quad (4.11)$$

Thus, the input sub-pattern is normalized between 0 and 1.

4.3.3 Fundamental Principles Used in Dynamic SOM Module Design

A SOM module is a learning algorithm that was originally proposed by Kohonen (1982) and the SOM modules were designed to learn to cluster groups of similar input patterns in a high dimensional input space using a non-linear function (Kohonen, 2001). The physical arrangement of neurons in the output layer is taken into consideration in the SOM modules. The neurons physically located close to each other in the output layer of the SOM have similar input patterns. The SOM layer employs WTA strategy, in which the output neurons compete among themselves to be a winning neuron, with the result that only one winning neuron is generated at one time. Such competition ends up with the neurons being forced to organize themselves. The main advantages of SOM modules are that it is non-linear and it can preserve the topological structure of dataset (Corn *et al.*, 1999). In general, the SOM modules cluster the input patterns into classes with meaningful maps. The neurons in Kohonen SOM module are connected to every neuron in the input layer

through adjustable weights or network parameters. The weight vectors in the Kohonen SOM module give a representation of the distribution of the input vectors in an ordered fashion.

Let $\mathbf{x} = [x_1, x_2, \dots, x_m]^T$ be the input vector to the Kohonen SOM neuron and $\mathbf{w}_j = [w_{1j}, w_{2j}, \dots, w_{mj}]^T$ be the weight vector assigned to the Kohonen SOM neuron. Then the net output of a Kohonen SOM neuron k_j can be defined as:

$$k_j = \sum_{i=1}^m w_{ij} x_i = \mathbf{w}_j^T \mathbf{x} \quad (4.12)$$

where $j=1, 2, \dots, p$, with p being the number of different classes considered and m the dimension of the input vector.

Subsequently, for the h^{th} neuron ($j \neq h$), if $k_h > k_j$ then weight vector \mathbf{w}_j is updated such that:

$$k_h = \sum_{i=1}^m w_{ij} x_i = \mathbf{w}_h^T \mathbf{x} = 1 \quad (4.13)$$

With the normalized input vector, the winning neuron can be decided by searching for the highest value as: $k_h^* = (\mathbf{x}_i^{\text{norm}})^T \mathbf{w}_h$. After the winning neuron is decided, the weights of winning neuron can be adjusted to yield a unity output $k_h^* = 1$, that is $k_h^* = (\mathbf{x}_i^{\text{norm}})^T \mathbf{w}_h = 1$. And since the input vector is normalized, the following equation can be obtained:

$$(\mathbf{x}_i^{\text{norm}})^T \mathbf{x}_i^{\text{norm}} = 1 \quad (4.14)$$

Thus, the normalized input vector can be updated as:

$$\mathbf{x}_i^{\text{norm}} = \mathbf{w}_h \quad (4.15)$$

Eq.(4.15) shows the weight vector of the winning neuron in a SOM module equals to the normalized input vector. Starting with the weights of the winning neuron to be the ones that best approximates the input vector, the weights can be updated closer to input vector as:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \alpha [\mathbf{x} - \mathbf{w}(n)] \quad (4.16)$$

where learning constant α is set between 0 and 1 and normally as 0.8 and \mathbf{x} denotes sub-patterns, n the iteration times.

4.3.4 Dynamic Formation of Neurons and Weights in SOM Module

In this paper, the neurons in the SOM models are built dynamically instead of setting a fixed number of neurons arbitrarily. The network is built to have neurons depending on the class to which a given input to a sub-pattern might belong. Such designed network produces less number of neurons and the time required to fire a neuron at the classification stage is reduced considerably.

The first neurons in all the SOM modules are constructed as Kohonen neurons. One neuron is built with input sub-pattern and randomly initialized weights to start with, and the initial weights are normalized following the same equation for input sub-patterns. Then the weights are updated until the output of the neuron is made equal to 1 with pre-defined error tolerance. Let $w(n)$ and $w(n + 1)$ be the weight at iteration n and $n+1$, then the weight is updated as Eq. (4.16) stated above.

The output value of a neuron is computed as:

$$z = \mathbf{W} * \mathbf{X}^T \quad (4.17)$$

In the dynamic SOM weight construction process, all incoming sub-patterns are checked to see if they are zero sub-patterns (i.e. zero vectors). The trained output weight is set to be zero and the searching procedure for winning neuron step is skipped if a zero sub-pattern exists. Otherwise, any incoming sub-pattern searches among the previously constructed neurons and corresponding weights if any neuron generates output z that equals to 1 with the pre-defined error tolerance. The neuron satisfies with the condition that the output equals to 1 with pre-defined error tolerance is claimed as the winning neuron. If the searching for winning neuron fails, another neuron and corresponding weight set are constructed additionally with the pre-defined error tolerance. After all the sub-patterns are imported into the respective input SOM modules, the output at any of the previously built neuron is compared to 1 with the pre-defined error tolerance. The neuron that satisfies the condition will be rewarded with a non-zero increment otherwise punished with a small non-zero decrement.

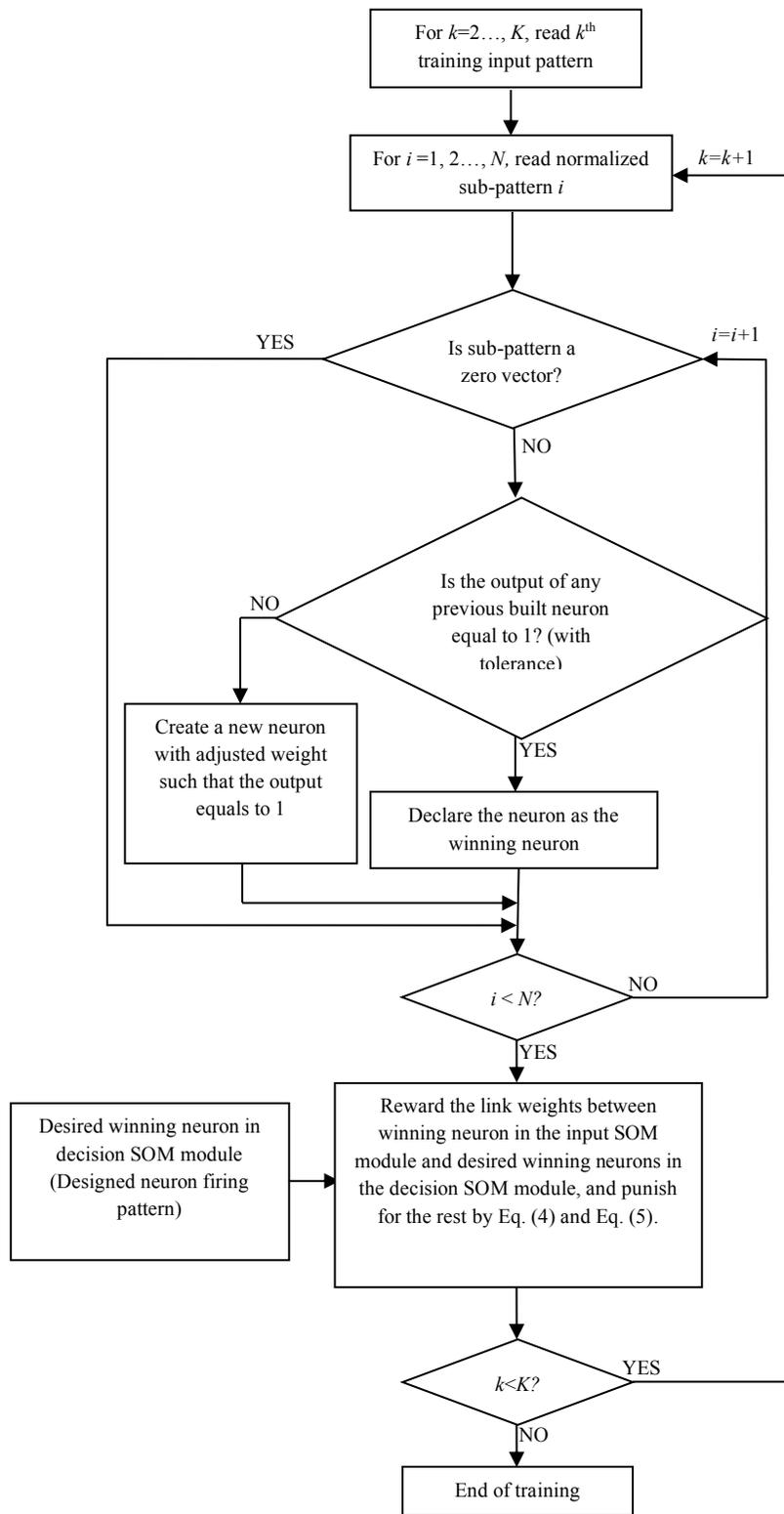


Figure 4.4. Flow chart of training procedure using dynamic neurons in the input SOM modules starting from the second training input pattern

As explained previously, the first neurons in all the input SOM modules are constructed as Kohonen neurons. When the first training input pattern is imported to the network, one neuron is built with inputs and randomly determined input weights. The initial input weights are normalized the same way as the input sub-patterns. Then the input weights are adjusted such that the output of the first neuron is equal to 1 (with pre-defined tolerance) using Eq.(4.16) and Eq.(4.17). The dynamic formation of neurons and their weights in the input SOM modules after the first training pattern in the training procedure is displayed in Figure 4.4. The testing procedure using the dynamic formation of neuron sin the input SOM modules is displayed in Figure 4.5.

As shown in Figure 4.4 and Figure 4.5, k represents index of input patterns, K the total number of training input patterns, M the total number of testing input patterns, i the index of sub-patterns of an input pattern, and N the total number of sub-patterns of an input pattern.

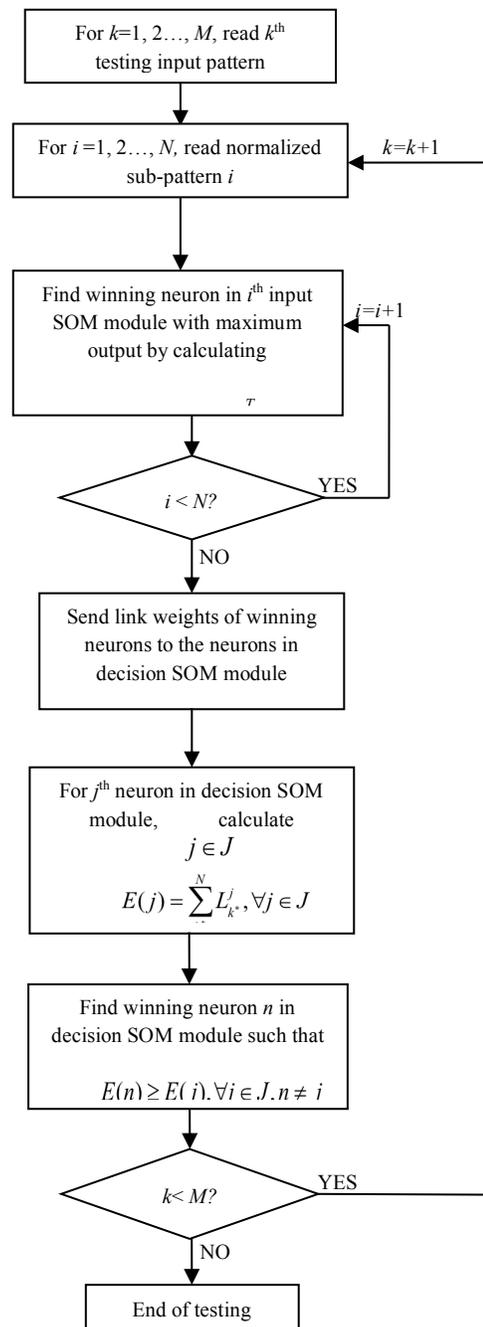


Figure 4.5. Flow chart of testing procedure using dynamic neurons in the input SOM modules

4.3.5 Determination of the Link Weights

Link weights are the weights of links connecting the winning neuron at every input SOM module to the decision SOM module. The value of link weight will be changed in every iteration based on reward/punish policy. In a modified version of the LAMSTAR network, the link weight $L_{i,j}(m, k)$ that represent correlation between m^{th} neuron in the k^{th} input SOM module to j^{th} neuron in the i^{th} decision SOM module is normalized as:

$$L_{i,j}^{\text{norm}}(m, k) = L_{i,j}^{k,m} / n(m, k) \quad (4.18)$$

where $n(m, k)$ denotes the number of times that neuron m is the winning neuron in the k^{th} input SOM module.

4.4 LAMSTAR Network Based Bearing Fault Diagnosis

The scheme of the LAMSTAR network based rotating machinery fault diagnosis scheme involves 5 steps as follows: (1) A LAMSTAR network with N input SOM modules is created as a start of the training procedure. The value of N is decided accordingly to the sub-pattern generation process. (2) Import the training dataset of patterns constructed by sub-patterns into the LAMSTAR network. Dynamic SOM weights and link weights are optimized using normalized sub-patterns. (3) Store the generated dynamic SOM weights and link weights. (4). Import the testing dataset into the LAMSTAR network. With storage of SOM weights and link weights, the winning neuron in each SOM module for testing data is calculated, and the winning neuron in decision SOM module is decided. (5) Achieve the machinery fault diagnosis with information of winning neuron in decision SOM module.

4.5 Sparse Coding and Combination with LAMSTAR Network on Rotating Machinery Fault Diagnosis

4.5.1 Sparse Coding Model

Sparse coding means to represent a signal as a linear combination of a few atoms of a given dictionary.

Given an input signal as $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$, the input signal can be expressed as a linear weighted combination of basic functions plus additive noise:

$$\mathbf{x} = \mathbf{D}\mathbf{s} + \varepsilon = \sum_{k=1}^K \mathbf{d}_k s_k + \varepsilon \quad (4.19)$$

where $\mathbf{D} \in \mathbb{R}^{n \times K}$ is the dictionary matrix, with each column as the basic function \mathbf{d}_k also known as atoms in dictionary learning model, $\mathbf{s} = [s_1, s_2, \dots, s_k]^T$ as the sparse representation of input signal \mathbf{x} , n as the number of data points in vector \mathbf{x} , and ε as assumed Gaussian white noise.

The goal of sparse coding is to find sparse representation coefficients s_k based on given input data and dictionary. Normally, the number of basic functions is greater than the dimension of input signal, which means $n < K$ for most of cases. Thus, the over-complete set of basic functions can be obtained. Sparse coding expects a succinct sparse representation of each input signal with most of its coefficients s_k are zero or nearly. In such case, there would not be infinite number of solutions for sparse coefficients vector with full-ranked matrix \mathbf{D} . The formal expression of sparse coding model can be written as:

$$\mathbf{x} \min_s \|s\|_0 \text{ subject to } \|\mathbf{x} - \mathbf{D}\mathbf{s}\|_2 \leq \gamma \quad (4.20)$$

where function $\|\cdot\|_0$ is referred to as the l^0 - norm that counts the nonzero entries of a vector, as a sparsity measurement; the approximation accuracy accessed by the l^2 - norm, $\|\cdot\|_2$; and γ as the approximation error tolerance.

Basic functions of dictionary matrix \mathbf{D} can be either manually extracted or automatically learned from the input data. As mentioned previously, the manually extracted basic functions are simple and will lead to fast algorithms, however with poor performance on matching the structure in the analyzed data. An adaptive

dictionary should be learned from input data through machine learning based methods, with basic functions capturing a maximal amount of structures of the data. Several potential options include RBM, autoencoder and LAMSTAR neural network due to the characteristics of automatically feature extraction. Particularly for the LAMSTAR neural network, the winning neurons in each self-organized map (SOM) module can be taken as basic function.

To get the exact determination of the sparse coefficients of a signal using a generic dictionary has been proven to be a NP-hard problem (Davis *et al.*, 1997). Instead, an approximate algorithm for inferring \mathbf{s} is to find a maximum a posterior (MAP) estimation (Lewiski and Sejnowski, 2000) as following:

$$\hat{\mathbf{s}} = \max_{\mathbf{s}} P(\mathbf{s}|\mathbf{x}, \mathbf{D}) = \max_{\mathbf{s}} P(\mathbf{x}|\mathbf{D}, \mathbf{s})P(\mathbf{s}) \quad (4.21)$$

where $P(\mathbf{s}|\mathbf{x}, \mathbf{D})$ represents the posterior distribution of \mathbf{s} , $P(\mathbf{x}|\mathbf{D}, \mathbf{s})$ as the likelihood of a signal for a given state of the coefficients, and $P(\mathbf{s})$ represents the prior probability distribution over the basic coefficients with assumption that \mathbf{s} is independent to \mathbf{D} . With Gaussian white noise assumed in Eq. (4.20), the likelihood can be obtained by Eq. (4.22) as following:

$$\log P(\mathbf{s}|\mathbf{x}, \mathbf{D}) \propto -\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{D}\mathbf{s}\|_2 \quad (4.22)$$

where σ^2 represents the noise variance. With consideration of forcing sparsity of sparse representations, $P(s_k)$ is defined as Laplace distribution as $P(s_k) \propto \exp(-\theta|s_k|)$, on the assumption of independent sparse coefficients $P(\mathbf{s}) = \prod_k P(s_k)$. Then the MAP estimation can be written as:

$$\hat{\mathbf{s}} = \max_{\mathbf{s}} P(\mathbf{s}|\mathbf{x}, \mathbf{D}) = \min_{\mathbf{s}} [-\log P(\mathbf{s}|\mathbf{x}, \mathbf{D}) - \log P(\mathbf{s})] = \min_{\mathbf{s}} \left(\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{D}\mathbf{s}\|_2 + \theta^T \|\mathbf{s}\|_1 \right) \quad (4.23)$$

It is common to replace l^0 - norm with l^1 - norm (Aharon *et al.*, 2006).

4.5.2 Sparse Coding and Dictionary Learning From Time Series Data

Machinery monitoring data is usually collected as time series data type. A time series data can always be partitioned into smaller blocks for signal processing. Thus, the characteristic patterns can be shifted in any time location due to various block length in partition. Moreover, the periodical defective impulse can be captured in monitoring data for bearing fault diagnosis. Shift-invariant sparse coding can effectively remove the periodical defective impulse effects, by representing the same defective impulse at different time location with only one basic function.

Thus, additional parameters need to be introduced for shift-invariant sparse coding as:

$$\mathbf{x} = \sum_{k=1}^k \mathbf{T}_l(\mathbf{d}_k) s_{k,l} + \varepsilon \quad (4.24)$$

where \mathbf{T}_l is a shift operator shifting an atom for $l \in [-L, L]$ offsets relatively; $s_{k,l}$ the coefficient associated with the dictionary \mathbf{d}_k shifted for l offsets; ε still the assumed Gaussian white noise.

Both the dictionary learning and sparse coefficients solutions in shift-invariant sparse coding are different from original sparse coding. The models for finding the sparse coefficients proposed by Lee *et al.* (2007) can be used in this research and can be briefly expressed as Eq. (4.25) below. Given a dataset containing N signals as $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$:

$$\min_{\mathbf{d}, \mathbf{s}} \sum_{i=1}^N \left(\left\| x_i - \sum_{k=1}^k \mathbf{d}_k * \mathbf{s}_{k,i} \right\|_2 + \beta \sum_{k=1}^k \|\mathbf{s}_{k,i}\|_1 \right) \quad (4.25)$$

$$\text{Subject to } \mathcal{C} \triangleq \{\|\mathbf{d}_k\|_2 \leq c, 1 \leq k \leq K\} \quad (4.26)$$

where $*$ represents the convolutional operator replacing the shift operator \mathbf{T}_l in Eq. (4.24); $\mathbf{s}_{k,i} \in \mathbb{R}^{n-K+1}$ the sparse activation of \mathbf{d}_k for signal \mathbf{x}_i ; and \mathcal{C} represents a convex set of matrices as a constraint for preventing the situation with either too large \mathbf{d}_k or too small $\mathbf{s}_{k,i}$. Given a fixed set of bases \mathbf{d} , the optimization problem

in Eq. (4.27) over sparse coefficients \mathbf{s} decomposes over $i = 1, \dots, N$ into N independent optimization problems. Thus, the coefficient \mathbf{s}_{ki}^l corresponding to input signal \mathbf{x}_i and dictionary \mathbf{d}_k with temporal offset l can be solved independently of the coefficients corresponding to other input signals. Therefore, the optimization problem becomes finding the appropriate coefficients \mathbf{s} corresponding to a single input \mathbf{x}_i :

$$\min_{\mathbf{s}} \sum_{k=1}^K \|\mathbf{s}_{ki}^l\|_0 \quad (4.27)$$

$$\text{Subject to: } \left\| \mathbf{x}_i - \sum_{k=1}^K \mathbf{d}_k * \mathbf{s}_{ki}^l \right\|_2 \leq \gamma, l \in L \quad (4.28)$$

Still, the variable γ represents the approximation error tolerance as before. In such way, the sparse coefficients can be solved independently for each input signal with the obtained dictionary. Similar to the standard sparse coding, Eq. (4.21)-Eq. (4.23) can be employed for solving sparse coefficients.

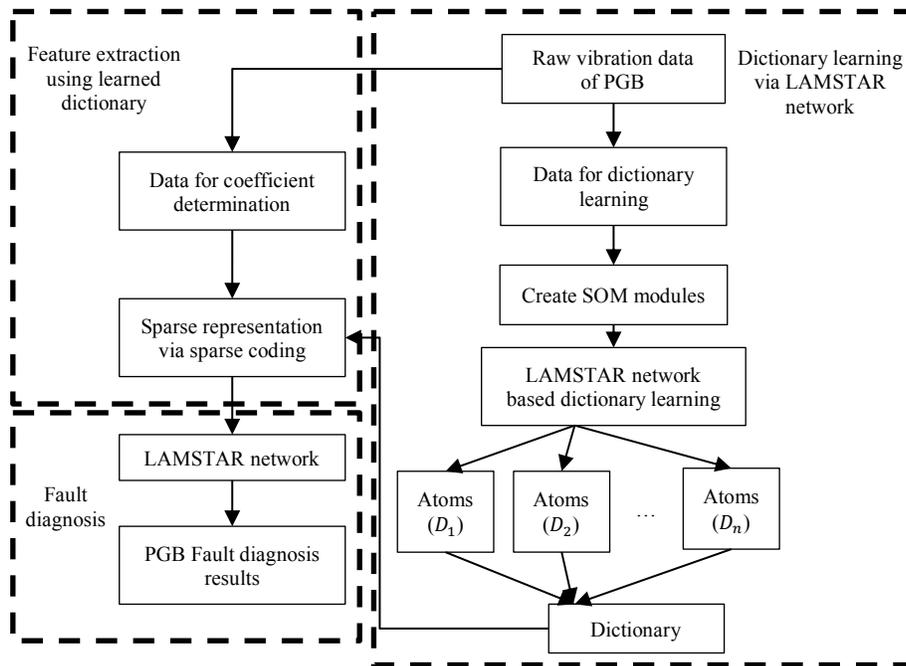


Figure 4.6 The framework of the LAMSTAR-DL network based PGB feature extraction and fault diagnosis

4.5.3 LAMSTAR-DL Based PGB Feature Extraction and Fault Diagnosis

The LAMSTAR-DL based PGB feature extraction and fault diagnosis framework is presented in Figure 4.6. As presented in Figure 4.6, the LAMSTAR-DL based approach is divided into two main stages: feature extraction, and fault diagnosis. The feature extraction stage involves two steps: dictionary learning and obtaining sparse representation. In the dictionary learning step, the randomly selected raw vibration signals are processed through LAMSTAR to obtain the basis functions as atoms. A dictionary is generated by arranging the obtained atoms together column by column into one matrix. Then, in the sparse representation generation step, the sparse representation the raw vibration testing data is obtained by shift-invariant sparse coding with the learnt dictionary. Finally, a LAMSTAR classifier network is trained with generated sparse representation of the vibration signals to diagnose the PGB faults. A significant advantage of the LAMSTAR-DL based approach is that it can extract effective fault features from raw vibration data without supervision, which can be further used for fault identification.

4.5.3.1 Dictionary learning using LAMSTAR

In using a LAMSTAR network for dictionary learning, the winning neuron in each input SOM module of the LAMSTAR network can be viewed as a self-learnt atom in the dictionary. Given a segment of monitoring data as \mathbf{x} , then a matrix \mathbf{X} containing n number of segments of data can be written as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{g}\mathbf{x}_n]$. Thus, the atom obtained from multiple segments of monitoring signals can be expressed as $\mathbf{a} = [\mathbf{\kappa}_1, \mathbf{\kappa}_2, \mathbf{\kappa}_3, \dots, \mathbf{\kappa}_m]$. For each individual segment of monitoring data, the sub-patterns are generated for dictionary learning. With the inspiration of LAMSTAR application for image recognition, a sliding window approach is applied to generate sub-patterns. For each pattern, the sub-patterns were generated by sampling the pattern using a sliding window with a length of d . Each sub-pattern sampled by the sliding window will be imported as a 1-D vector into an input SOM module. Thus, the number of input SOM modules and the number of the generated sub-patterns remain same. The procedure to obtain sub-patterns from one pattern is presented in Figure 4.7.

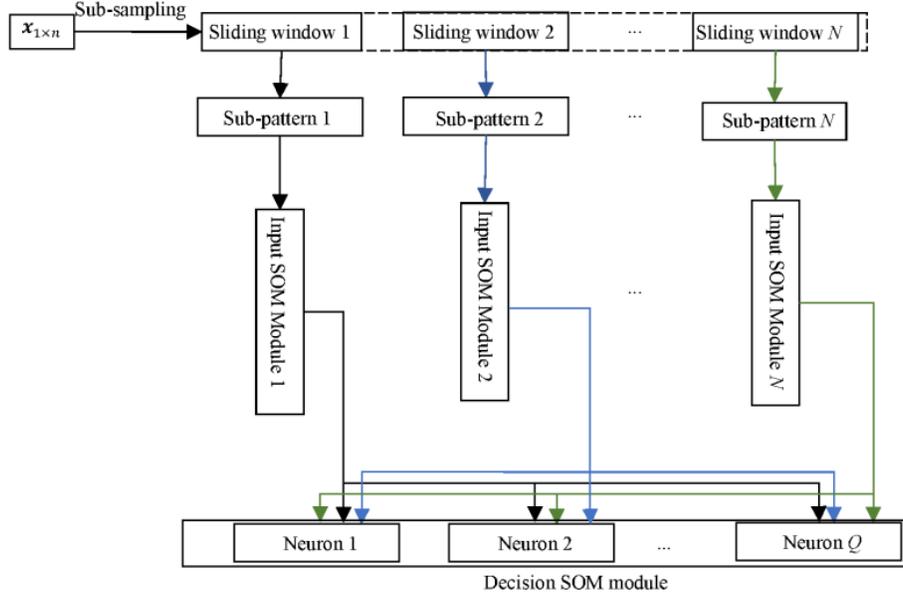


Figure 4.7. The scheme to obtain the sub-patterns from raw signal as input to LAMSTAR-DL network for dictionary learning.

With a sliding window $W_{1 \times l}$ whose size has been pre-defined, the total number of generated sub-patterns N can be calculated as $N = n/d$. Since each segment of the raw signal is regarded as one pattern, the sub-pattern can be defined as $\Omega_v, v = 1, 2, \dots, N$. Then, Ω_v is written as:

$$\Omega_v(i) = \mathbf{x}_{1 \times n}[(v-1) \times i] \quad (4.29)$$

To determine an appropriate sliding window dimension, a sampled sub-pattern should contain as much variant information as possible for features extraction. Therefore, a selection criterion can be designed as to choose a sliding window of size d such that the sum of the Von Neumann entropy of all the sliding windows is maximized as:

$$\arg \max_d \sum_{v=1}^N \frac{H(\Omega_v)}{d^2} \quad (4.30)$$

where $H(\Phi_v)$ represents the Von Neumann entropy of each sliding window $\Phi_v, v = 1, 2, \dots, N$. The detailed theory of Von Neumann entropy can be found in (Bengtsson and Zyczkowski, 2006). Since the number of input SOM modules equals to the number of generated sub-patterns, the LAMSTAR-DL network can be established with dynamic neurons in each input SOM module.

To generate the sparse representation of the vibration signals using shift-invariant sparse coding, Eq. (4.27) and Eq. (4.28) will be solved with the dictionary \mathbf{d}_k learnt by the LAMSTAR-DL network. Since the LAMSTAR-DL network learns the dictionary with the resampled sub-patterns, the temporal offset l can be defined as $l \in N$. Thus, the number of possible offsets is decided by the number of sub-patterns, which is selected based on the criterion in Eq.(4.31).

4.5.3.2 Designed LAMSTAR network for PGB fault diagnosis on sparse representation

With the dictionary learnt through LAMSTAR-DL network, the sparse representation of a segment of the raw data can be obtained by shift-invariant sparse coding explained in Section 4.5.3.1. The obtained sparse representations of the original vibration signals will be used for PGB fault diagnosis in the next stage. A LAMSTAR network will be used as the classifier in this study for the advantage of fast processing massive data. Besides, the dimension of LAMSTAR classifier network can be adjusted easily by arranging the number of SOM modules matching with the dimension of sparse representations. The sparse representations of original vibration signals can be imported into the LAMSTAR classifier network for PGB fault diagnosis without other feature extraction process. The LAMSTAR network was applied to acoustic emission signals with much higher diagnosis accuracy on PGB fault diagnosis in comparison with other common classifiers including k -nearest neighbors (KNN) and backpropagation (BP) network (Yoon *et al.*, 2015). The obtained sparse representation of the original raw vibration signal will be channeled row by row into the LAMSTAR classifier network for PGB fault diagnosis. For each row of the sparse representation, an input SOM module will be constructed. The row number of the sparse representation remains the same as number of basis functions in the learnt dictionary, which is equal to the number of the input SOM modules in the LAMSTAR-

DL network. Therefore, the number of input SOM modules in the LAMSTAR classifier network is equal to the number of SOM modules in LAMSTAR-DL network. PGB health conditions will be represented by the neurons firing sequence in the decision SOM module. Thus, the PGB faults can be diagnosed. Therefore, the LAMSTAR-DL network based PGB feature extraction and fault diagnosis are performed by two stacked LAMSTAR networks. The first LAMSTAR network was used for dictionary learning and the second one for PGB fault diagnosis by using the sparse representations of the original vibration signals.

4.6 Hybrid Deep Signal Processing Approach for Bearing Fault Diagnosis Using Vibration Signals

The framework of the hybrid deep signal processing approach for bearing fault diagnosis is shown in Figure. 4.8.

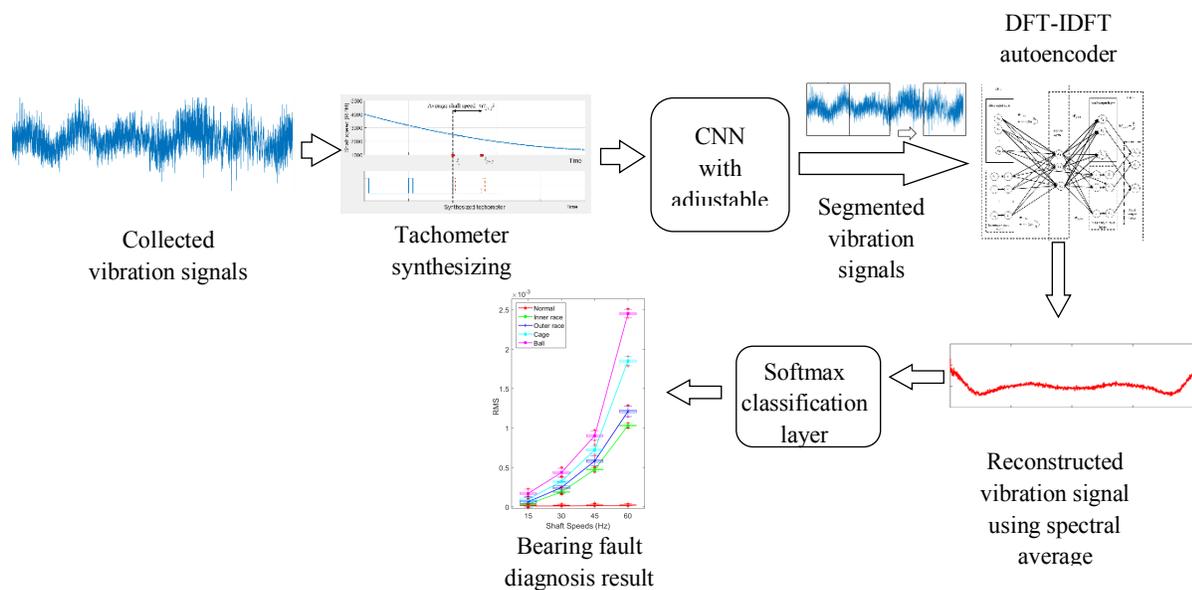


Figure. 4.8 Framework of the hybrid deep signal processing approach for bearing fault

4.6.1. Signal Segment

To integrate vibration analysis techniques such TSA/TSR into deep learning to process vibration signals, it is important for bearing fault diagnosis to select the correct number of data points in each revolution of the shaft rotation, i.e., signal segmentation. The number of data point in each revolution can be extracted by calculating zero crossing times in tachometer signal. However, in many machine health monitoring and fault diagnosis applications, obtaining tachometer signal is almost impossible. For example, in some aircraft

engine, the high-speed turbine shaft is not accessible for a tachometer. Therefore, an alternative way to estimate the appropriate length of segment is needed. According to the study by Luo *et al.* [49], a tachometer signal can be synthesized from shaft speed profile. The relation between the shaft speed profile and the synthesized tachometer signal is provided in Figure. 4.9.

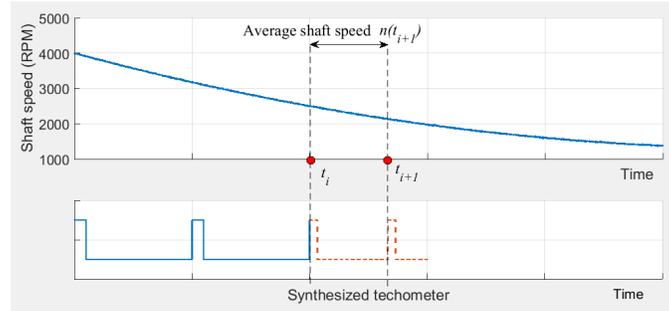


Figure. 4.9. Synthesized tachometer generation from speed profile

The shaft speed profile can be generated using the instantaneous frequency of the vibration signals, which is obtained from the derivative of the unwrapped phase signal. Given the original vibration signal as $x(t)$, then the phase information $\psi(t)$ of the vibration signal can be expressed as:

$$\psi(t) = \arctan (z(t)/x(t)) \quad (4.31)$$

$$z(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau \quad (4.32)$$

where $z(t)$ represents the Hilbert transform of the original vibration signal, as the time domain convolution result between $1/(\pi t)$ and $x(t)$. The tachometer synthesizing procedure is provided next.

Step 1. Given the shaft speed profile as shown in Figure. 4.9, initially assume a synchrophaser pulse at time zero.

Step 2. Once the i^{th} synchrophaser pulse is located, at t_i , assume the $(i + 1)^{th}$ pulse is located at t_{i+1} .

Step 3. Calculate the average shaft speed $n(t_{i+1})$ from the given shaft speed profile $S(t)$ as a function of t_{i+1} :

$$n(t_{i+1}) = \frac{1}{t_{i+1} - t_i} \int_{t_i}^{t_{i+1}} S(t) dt \quad (4.33)$$

Step 4. The time difference between t_i and t_{i+1} is written as:

$$\Delta t_1 = t_{i+1} - t_i \quad (4.34)$$

The time elapsed by one rotation can be written based on Eq. (4.31):

$$\Delta t_2 = \frac{60}{n(t_{i+1})} \quad (4.35)$$

Step 5. The target t_{i+1} can be solved as:

$$\min_{t_{i+1}} |\Delta t_1 - \Delta t_2| \quad (4.36)$$

As one can see from the tachometer synthesizing procedure, the location of possible impulse can only be determined when the closest previous impulse is certain. By considering the tachometer in each revolution, the synthesized tachometer is proved to be accurate with small error. However, for automatic vibration signal processing and fault diagnosis, an automated tachometer synthesizer that can generate synthesized tachometer signals from the vibration signals is needed. Assuming the equal length of the signal segments, then the length of a signal segment will be the number of data points between two continuous impulses in shaft speed profile, Δt_1 in Eq. (4.34) and Eq. (4.36) stay stable. Hence, the objective function of searching for the impulse times can be set as:

$$\min_{t_1, t_2, \dots, t_M} \left| M\Delta t_1 - \sum_{i=1}^{M-1} \frac{60}{n(t_{i+1})} \right| \quad (4.37)$$

where M represents the total number of segments and t_i the occurrence time of the i^{th} impulse. The form of segments in a continuous time series signal is similar to the sliding window or filter, but with varying size. The objective of the designed neural network is to find the optimized filter size that can meet Eq. (4.37). Inspired by the adjustable filter size in a convolutional neural network (Han *et al.*, 2016), a simple linear 1-D convolutional layer with adjustable filter size is employed to find the optimal length of segment. Given an input 1-D signal denoted by \mathbf{x} with length as N , then the activation of \mathbf{x} denoted as \mathbf{y} , calculated as inner product of \mathbf{x} and convolutional filter can be written as:

$$\mathbf{y}_i(k) = \mathbf{w}(k)\mathbf{x}_j(k) + \mathbf{b}_i, \quad i = 1, 2, \dots, f; j = (i - 1)k + 1 \quad (4.38)$$

where $\mathbf{w}(k)$ is the convolutional filter with size as $1 \times k$, $\mathbf{y}_i(k)$ the output of the i^{th} filter with filter size as k , f the total number of required filters, $\mathbf{x}_j(k)$ the input with a receptive field of length k starting at the j^{th} data point in vector \mathbf{x} , \mathbf{b}_i the bias to be added. Normally, the filter size k is a predefined integer and constant throughout the training and testing procedure. Here $k \in \mathbb{R}^+$ is defined as a continuous variable that can be adjusted throughout the training procedure. Based the concept of using filter as sliding window to cover the two neighboring impulses, the filter slides from the beginning of speed profile with stride as the length of the filter. Thus, the number of filters required in the convolutional layer can be decided as:

$$f = \frac{N - k + 2P}{k} + 1 \quad (4.39)$$

where f represents the number of required filters, P the amount of zero padding on the border of the original input. It is possible that Eq. (4.39) yields to a non-integer result for the number of filters, indicating that the neurons in the input layer can fit into filter size properly. In the designed network, the extra zeros are padded

on the border of the original input vector to make it fit. The structure of the convolutional layer is shown in Figure. 4.10.

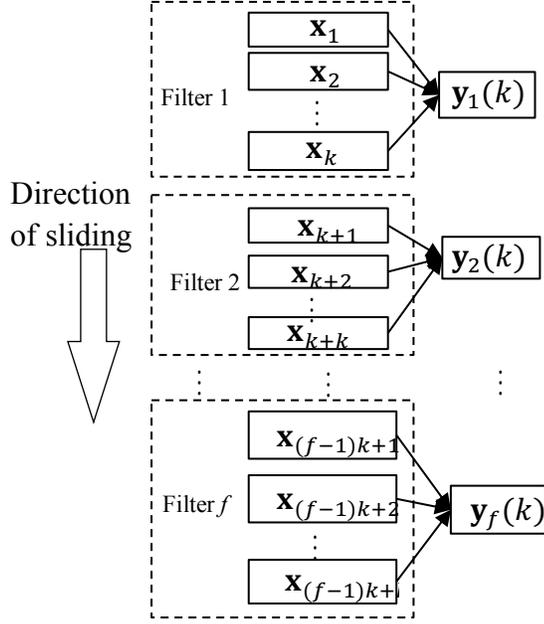


Figure. 4.10. Schematic of convolution layer in the CNN with adjustable filter size

To establish the relationship between activation and the adjustable filter size, an upper-bounded filter $w(k_u)$ and a lower-bounded filter $w(k_l)$ are defined, respectively. Specifically, the size of the upper-bounded filter k_u is the smallest odd number bigger than k and the size of the lower-bounded filter k_l is the largest odd number smaller or equal to k . The definition of these filter sizes can be expressed as:

$$k_u = \lfloor (k + 1)/2 \rfloor * 2 + 1, k_l = \lfloor (k + 1)/2 \rfloor * 2 - 1 \quad (4.40)$$

$$\alpha = (k - k_l)/2 \quad (4.41)$$

Therefore, the activation y_i can be calculated as a linear combination of the upper-bounded filter and lower-bounded filter as:

$$y_i(k) = \alpha y_i(k_u) + (1 - \alpha) y_i(k_l) \quad (4.42)$$

where $\mathbf{y}_i(k_u)$ and $\mathbf{y}_i(k_l)$ represent the activation of the upper-bounded filter and the lower-bounded filter, respectively and α the linear combination weight. The activations are calculated as in Eq. (4.38) with the same bias. Eq.(4.41) can be easily rewritten as:

$$k = k_l + 2\alpha \quad (4.43)$$

Thus, the optimal filter size k is a weight related filter size in the defined boundary. Additionally, the relationship between filters with different sizes can be expressed as:

$$\mathbf{w}(k) = \alpha\Delta\mathbf{w}(k_u) + \mathbf{w}(k_l) \quad (4.44)$$

where $\Delta\mathbf{w}(k_u)$ represents the unique part of upper-bounded filter subtract the shared part with lower-bounded filter. The relationship described in Eq. (4.44) can be illustrated in Figure 4.11. In Figure 4.11, the yellow area represents the elements shared by upper and lower-bounded filters. The unique part of upper-bounded filter $\Delta\mathbf{w}(k)$ is shown as the blue area with zeros inside in Figure 4.11. The weight of upper-bounded filter α is defined in Eq. (4.41).

$$\begin{array}{c}
 \boxed{w_1 \mid w_2 \mid \dots \mid w_{n-1} \mid w_n} \\
 \mathbf{w}(k_u)
 \end{array}
 -
 \begin{array}{c}
 \boxed{w_2 \mid \dots \mid w_{n-1}} \\
 \mathbf{w}(k_l)
 \end{array}
 =
 \begin{array}{c}
 \boxed{w_1 \mid 0 \mid \dots \mid 0 \mid w_n} \\
 \Delta\mathbf{w}(k_u)
 \end{array}$$

$$\begin{array}{c}
 \boxed{w_1 \mid 0 \mid \dots \mid 0 \mid w_n} \\
 \Delta\mathbf{w}(k_u)
 \end{array}
 * \alpha +
 \begin{array}{c}
 \boxed{w_2 \mid \dots \mid w_{n-1}} \\
 \mathbf{w}(k_l)
 \end{array}
 =
 \begin{array}{c}
 \boxed{\alpha w_1 \mid w_2 \mid \dots \mid w_{n-1} \mid \alpha w_n} \\
 \mathbf{w}(k)
 \end{array}$$

Figure 4.11. Illustration definition of an adjustable filter size k

With Eq. (4.38), Eq.(4.40) can be deduced as:

$$\mathbf{y}_i(k) = \alpha \mathbf{w}(k_u) \mathbf{x}_j(k_u) + (1 - \alpha) \mathbf{w}(k_l) \mathbf{x}_j(k_l) + \mathbf{b}_i \quad (4.45)$$

Note that after padding zeros into $\mathbf{w}(k_l)$, $\mathbf{w}(k_l) \mathbf{x}_j(k_l)$ equals to $\mathbf{w}(k_l) \mathbf{x}_j(k_u)$. Thus, Eq. (4.44) can be derived as:

$$\begin{aligned} \mathbf{y}_i(k) &= \alpha \mathbf{w}(k_u) \mathbf{x}_j(k_u) + (1 - \alpha) \mathbf{w}(k_l) \mathbf{x}_j(k_l) + \mathbf{b}_i \\ &= [\alpha \mathbf{w}(k_u) + (1 - \alpha) \mathbf{w}(k_l)] \mathbf{x}_j(k_u) + \mathbf{b}_i \\ &= [\alpha \Delta \mathbf{w}(k_u) + \mathbf{w}(k_l)] \mathbf{x}_j(k_u) + \mathbf{b}_i \end{aligned} \quad (4.46)$$

Substitute Eq. (4.44) into Eq. (4.46) to obtain the simplified result as:

$$\mathbf{y}_i(k) = \mathbf{w}(k) \mathbf{x}_j(k_u) + \mathbf{b}_i \quad (4.47)$$

Thus, the partial derivative of the activation with respect to the filter size can be calculated as:

$$\frac{\partial \mathbf{y}_i(k)}{\partial k} = \frac{\mathbf{y}_i(k_u) - \mathbf{y}_i(k_l)}{k_u - k_l} \quad (4.48)$$

With relationship stated in Eq. (4.36) and padding zeros into k_l , Eq. (4.48) can be simplified as:

$$\frac{\partial \mathbf{y}_i(k)}{\partial k} = \frac{\Delta \mathbf{w}(k_u) \mathbf{x}_j(k_u)}{k_u - k_l} \quad (4.49)$$

Based on Eq. (4.49), the partial derivative of loss L with respect to filter size k can be calculated as:

$$\frac{\partial L}{\partial k} = \sum_i \frac{\partial L}{\partial \mathbf{y}_i} \frac{\partial \mathbf{y}_i}{\partial k} \quad (4.50)$$

Therefore, the filter size k can be updated as:

$$k^{t+1} = k^t - \gamma \frac{\partial L}{\partial k} \quad (4.51)$$

where t represents the iteration number.

Similarly, the partial derivative of the activation with respect to the upper-bounded filter can be expressed as:

$$\frac{\partial \mathbf{y}_i^t}{\partial \mathbf{w}^t(k_u^t)} = \mathbf{x}_j(k_u^t) + \alpha^t \Delta \mathbf{x}_j(k_u^t) \quad (4.52)$$

The derivative of loss L with respect to upper-bounded filter can be calculated as:

$$\frac{\partial L}{\partial \mathbf{w}^t(k_u^t)} = \sum_i \frac{\partial L}{\partial \mathbf{y}_i^t} \frac{\partial \mathbf{y}_i^t}{\partial \mathbf{w}^t(k_u^t)} \quad (4.53)$$

After each iteration, the filter size k^{t+1} is possible to be greater than the upper-bounded filter size or smaller than the lower bounded filter size. Thus, both the sizes of the upper-bounded and lower-bounded filters should be updated. Next, two processes are defined to manage the boundary filter size update process, namely expanding and shrinking procedures.

Expanding: When the updated filter size is greater than the previous upper-bounded filter size, i.e., $k^{t+1} > k_u^t$, both the upper-bounded filter and lower-bounded filter should be expanded as:

$$\begin{aligned} w^{t+1}(k_u^{t+1}) &= \text{expand}[w^{t+1}(k_u^{t+1})] \\ w^{t+1}(k_l^{t+1}) &= w^{t+1}(k_u^{t+1}) \end{aligned} \quad (4.54)$$

where $\text{expand}(\cdot)$ denotes a function to increase the filter size by padding values from the nearest neighbors of the original filter.

Shrinking: When the updated filter size is smaller than the previous lower-bounded filter size, i.e., $k^{t+1} < k_l^t$, both the upper-bounded filter and lower-bounded filter will be shrunk as:

$$\begin{aligned} w^{t+1}(k_u^{t+1}) &= w^{t+1}(k_l^{t+1}) \\ w^{t+1}(k_l^{t+1}) &= \text{shrink}[w^{t+1}(k_l^{t+1})] \end{aligned} \quad (4.55)$$

where $\text{shrink}(\cdot)$ denotes a function to decrease the filter size by filling the first and the end unit as zero. The process of expanding and shrinking is illustrated in Figure. 4.12 .

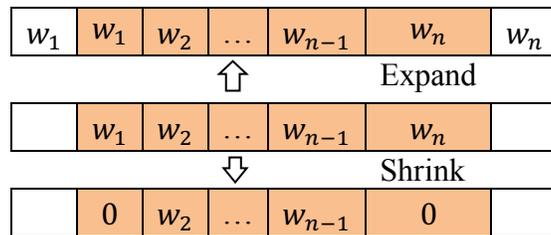


Figure. 4.12. Illustration of the expanding and shrinking process

4.6.2 The DFT-IDFT autoencoder

Inspired by the fact that the extraordinary impulses are easier to identify in frequency domain, a neural network that can automatically learn the characteristics of vibration signals in the frequency domain is developed. Let the general mapping function $f()$ be the relationship between a D -dimensional vector and a scalar, then a conventional L -layer feed-forward neural network can be written as the function as:

$$f(\mathbf{x}) = g_L \left\{ \sum_k \mathbf{w}_{L-1k} [\dots g_1 \left(\sum_i \mathbf{w}_{1i} \mathbf{x}_i \right)] \right\} \quad (4.56)$$

where $\mathbf{w}_l (l = 1, 2, \dots, L)$ denotes the weights matrix of the l^{th} layer, $g_l (l = 1, 2, \dots, L)$ the activation function of the l^{th} layer, and $f(\mathbf{x})$ the output scalar value. Given the special condition where $L=1$, then Eq. (4.56) can be written as:

$$f(\mathbf{x}) = g \left(\sum_i \mathbf{w}_i \mathbf{x}_i \right) \quad (4.57)$$

Let \mathbf{x}_n be a time series signal and consider the discrete Fourier transform (DFT) of \mathbf{x}_n as:

$$\mathbf{X}_k = \sum_{n=0}^{N-1} \mathbf{x}_n \cdot e^{-jkn\frac{2\pi}{N}}, k = 0, 1, \dots, N-1 \quad (4.58)$$

where \mathbf{X}_k represents the k^{th} spectral line of \mathbf{x}_n . It can be observed that by taking \mathbf{x} as the input vector, $e^{-jkn\frac{2\pi}{N}}$ the connecting weights between input layer and the output layer, activation

function as linear function, the k^{th} spectral line of a discrete time series signal can be calculated by implementing a single layered neural network.

As described by Velik (2008), with the weight matrix initialized as the discrete Fourier transform (DFT) coefficients $e^{-jkn\frac{2\pi}{N}}$, the DFT of a time series vibration signal can be performed in a single layered neural network without training process. Furthermore, the DFT coefficient can be learned with the calculated DFT as the target in a single layered neural network. In such case, the weights are randomly initialized and are learned through back propagation process with an approximation solution. The frequency component of the DFT can be used to reconstruct the original signal by taking the inverse discrete Fourier transform (IDFT). The IDFT can be expressed as:

$$\hat{x}_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k \cdot e^{-jkn\frac{2\pi}{N}}, k = 0, 1, \dots, N - 1 \quad (4.59)$$

where \hat{x}_n represents the n^{th} data point of the reconstructed signal.

By viewing the DFT and IDFT of the segmented vibration signals as an encoding and decoding process, respectively, then the vibration signal processing can be performed by an autoencoder structure embedded with vibration analysis techniques. The DFT-IDFT autoencoder is shown in Figure 4.13.

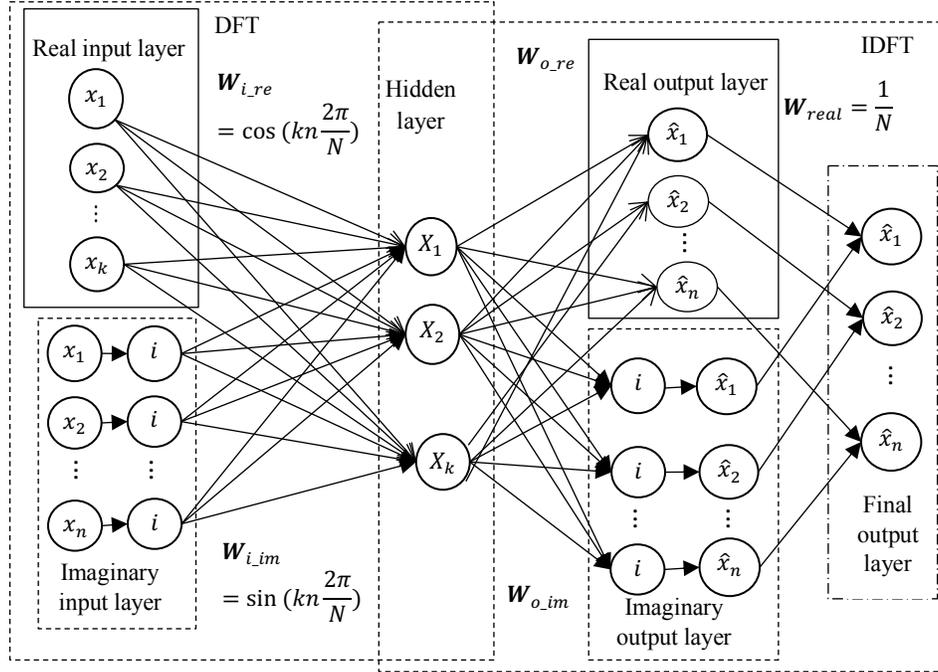


Figure 4.13. The structure of DFT-IDFT autoencoder with dual inputs layers

As shown in Figure 4.13, the DFT-IDFT autoencoder can learn the DFT and IDFT coefficients by mimicking encoding and decoding procedures of an autoencoder through unsupervised learning. The DFT and IDFT coefficients are complex values as Eq. (28) and Eq. (29) indicate. To handle the complex values in an autoencoder, the DFT-IDFT autoencoder with dual input layers is constructed. In Figure 4.13, the solid line block and dashed line block represent the structure for learning the real part and imaginary part of the transformation coefficients, respectively. Thus, the dual input layers represent the real part and imaginary part of the DFT

coefficients so that both the real and imaginary parts can be learnt through the network. According to the Euler's formula, Eq. (4.58) and Eq.(4.59) can be rewritten as:

$$\mathbf{X}_k = \sum_{n=0}^{N-1} \mathbf{x}_n \cdot \{\cos[kn(2\pi/N)] + j\sin[kn(2\pi/N)]\}, k = 0,1, \dots, N - 1 \quad (4.60)$$

$$\hat{\mathbf{x}}_n = \frac{1}{N} \sum_{k=0}^{N-1} \mathbf{X}_k \cdot \{\cos[kn(2\pi/N)] + j\sin[kn(2\pi/N)]\}, k = 0,1, \dots, N - 1 \quad (4.61)$$

where $j = \sqrt{-1}$ represents the imaginary unit. Thus, the real part and imaginary part of DFT coefficients can be calculated. A full connection is constructed between the real input layer and the hidden layer, and the connection weights represent the real parts of the DFT coefficients. In the imaginary input layer, double-connected neurons will be used. In each double-connected neuron, one neuron takes the real value from the input vibration signal. This neuron is connected to an imaginary neuron with a constant weight of 1 in order to convert the real value into an imaginary value. Thus, Eq. (4.58) is calculated as the output of hidden layer. Similarly, the real part and the imaginary part will calculate the output layer, through the real output layer and the imaginary output layer, respectively. In the end, the outputs from the real output layer and the imaginary output layer are simply added up to form the final output layer. The result from the final output layer is the construction of the original vibration signals, expressed by Eq. (4.61).

According to Eq. (4.60), the weights between real input layer and the hidden layer can be expressed as:

$$\mathbf{W}_{i_{re}} = \cos[kn(2\pi/N)] \quad (4.62)$$

Note that the imaginary part of the DFT coefficients is updated by adjusting the weight matrix with constant imaginary unit. Therefore, the weight matrix between the imaginary input layer and the hidden layer can be simply expressed as:

$$\mathbf{W}_{i_{im}} = \sin[kn(2\pi/N)] \quad (4.63)$$

Furthermore, $\mathbf{W}_{i_{im}}$ can be rewritten according to the trigonometric functions as:

$$\mathbf{W}_{i_{im}} = \sqrt{1 - \mathbf{W}_{i_{re}}^2} \quad (4.64)$$

By observing Eq. (4.58) – Eq.(4.64), the IDFT coefficients are the transpose of the DFT coefficients for the same input signal. Thus, the weights between the hidden layer and the real output layer, and the weights between the hidden layer and the imaginary output layer can be written as:

$$\mathbf{W}_{o_{im}} = \mathbf{W}_{i_{im}}^T \quad (4.65)$$

$$\mathbf{W}_{o_{re}} = \mathbf{W}_{i_{re}}^T \quad (4.66)$$

The connections between real output layer and final output layer, and between imaginary output layer and the final output layer are kept constant as $1/N$ to obtain the reconstruction result of original signals. Therefore, the complete cost function of the DFT-IDFT autoencoder can be expressed as:

$$\begin{aligned}
J(W) = & \frac{1}{N} \sum_{n=1}^N \|x_n - \text{real}(\hat{x}_n)\|^2 \\
& + \frac{\lambda}{2} \sum_{n=1}^N \sum_{k=0}^{N-1} \left[(\mathbf{W}_{i_re}^{nk})^2 + (\mathbf{W}_{i_im}^{nk})^2 + (\mathbf{W}_{o_re}^{nk})^2 + (\mathbf{W}_{o_im}^{nk})^2 \right]
\end{aligned} \tag{4.67}$$

Thus, DFT and IDFT coefficients of the input vibration signals can be learnt through the unsupervised training method. The constraints between four different weight sets expressed as Eq. (4.62) – Eq.(4.66) enforce the autoencoder to accomplish the DFT and IDFT procedures through one iteration. The training procedure is completed by using backpropagation algorithm to minimize the cost function. The weights are updated after one iteration through the gradient descent method. Once the weights \mathbf{W}_{i_re} are determined, the other weights \mathbf{W}_{i_im} , \mathbf{W}_{o_im} , and \mathbf{W}_{o_re} can be determined by Eq. (4.63) and Eq. (4.65). Therefore, only \mathbf{W}_{i_re} needs to be updated after one iteration. The updating equation can be written as:

$$\mathbf{W}_{i_re} = \mathbf{W}_{i_re} - \lambda \frac{\partial J(W)}{\partial \mathbf{W}} \tag{4.68}$$

where λ represents the predefined learning rate. Once the DFT autoencoder is trained with sufficient weight sets to accomplish the DFT and IDFT procedures, a maxpooling layer can be added after the hidden layer. The maxpooling layer will keep the maximum value in the pooling window and replace the rest with zero. Hence, by keeping the maximum value in the spectrum of one revolution, the optimized neural network can compute the time synchronous average (TSA) of the original vibration signal. The designed TSA embedded neural network can be illustrated in Figure. 4.14.

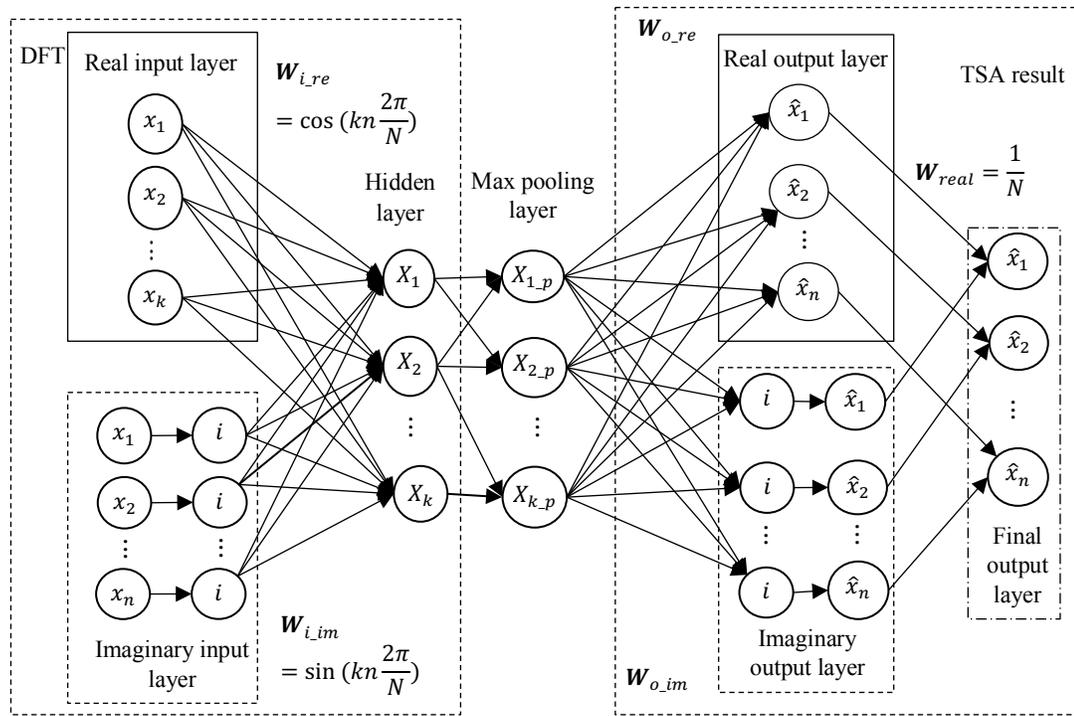


Figure. 4.14 The DFT-IDFT autoencoder to compute the TSA

To validate the DFT-IDFT autoencoder, a sample signal with random noise was used to train the autoencoder and the TSA result was obtained at the final output layer. The original signal with sample was expressed as $s = \sin\left(2\pi 12t + \frac{\pi}{8}\right) + \omega(0,1)$. As shown in Figure. 4.15, the signal to noise ratio of sample signal was -2.1544. As displayed in Figure. 4.15, the plot of result from final output layer is close to the plot of theoretical TSA calculation result. The root mean

squared error (RMSE) between the results of DFT-IDFT autoencoder and the theoretical TSA is as low as 0.012 for the plots shown in Figure. 4.15.

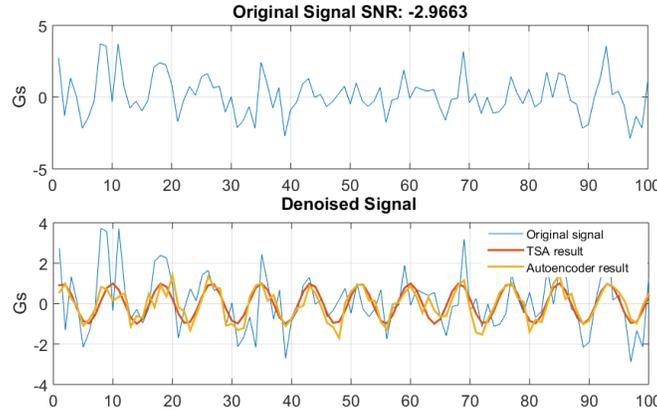


Figure. 4.15. Validation result using TSA embedded neural network on sample signal

4.6.3 TSR and reconstructed vibration signals

With the TSA embedded neural network, signal processing techniques based on TSA methods can be carried out in the neural network like structure. Inspired by the successful applications of TSA based gear pitting diagnosis, Van Hecke *et al.* (2014) used time synchronous resampling (TSR) on acoustic emission (AE) signals for bearing fault diagnosis. The diagnosis results have shown the effectiveness of the TSR on noise remove. However, the TSA/TSR based method requires shaft zero crossing times (ZCT) for calculation of a TSA. Usually, the ZCTs are defined by a tachometer signal, or an angular reference to estimate the one per revolution. Besides, the data points in each revolution need to be interpolated so that the length of each revolution segment stays constant for TSA calculation. To address these critical challenges during implementation of the TSA related method, the neural network embedded with TSA is developed. The objective of using TSA embedded neural network for signal processing is to eliminate the complexity of TSA related methods and resampling process. Generally, the raw data will be resampled by a convolutional kernel to obtain the periodograms. Then, the average of the

periodograms for all the resampled segments are calculated. Finally, time domain expression of the averaged periodograms are obtained through the IDFT layer. The designed structure of TSA embedded neural network for TSR can be illustrated in Figure 4.16.

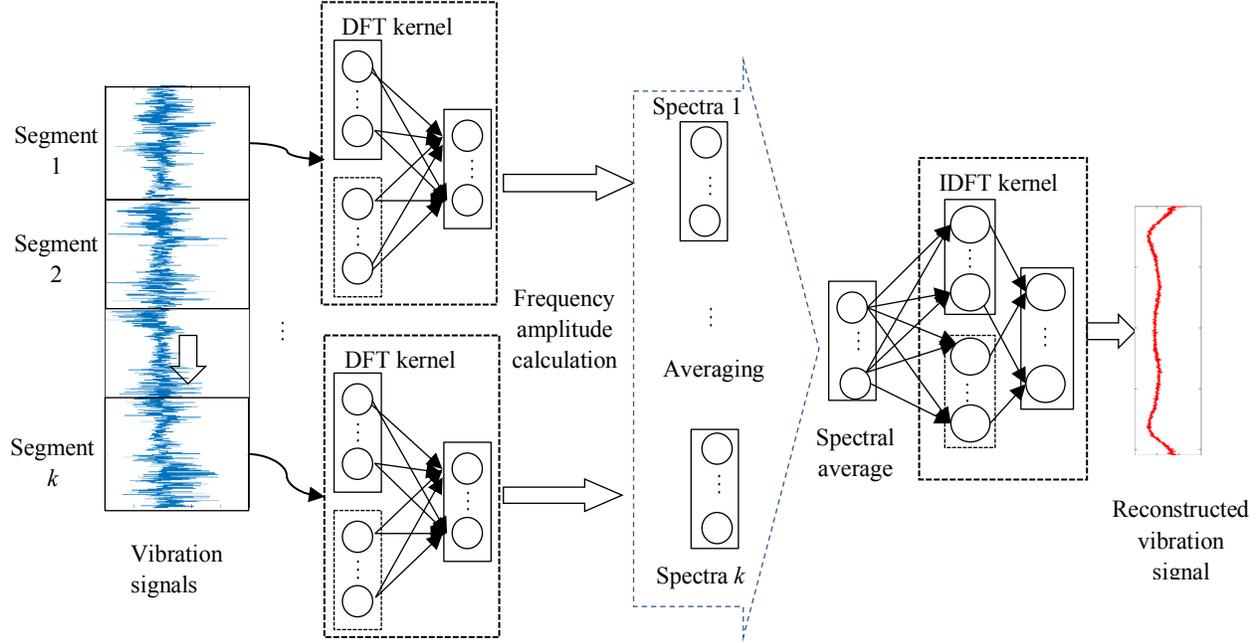


Figure 4.16. The DFT-IDFT autoencoder to compute the TSR and reconstructed vibration signals

As shown in Figure 4.16, the original data series are resampled by the sliding window into k segments. Each segment is processed through the pre-trained DFT dual layers (real layer and imaginary layer) to obtain the expression in frequency domain. Let $x(j)$, for $j = 1, 2, \dots, N$ be a collected signal sample, then the segments can be expressed as:

$$x_k(i) = x[i + (h - 1)M], \quad \text{for } i = 1, 2, \dots, k; h = 1, 2, \dots, H \quad (4.69)$$

where h represents the h^{th} segment sampled from the original signal, M the number of data points in the overlapping area between two continuous segments, k the length of each segment obtained

from the optimized filter size, H the total number of sampled segments. Particularly, M can be selected as 0.

Here in Figure 4.16, the DFT dual layers and the frequency domain output layer are displayed as DFT kernel to filter the original signal. The magnitude of frequency components is calculated following the output layer of the DFT kernel. By plugging Eq. (4.60) and Eq. (4.61) into Eq. (4.58)., the result after the DFT kernel as frequency domain expression can be written as:

$$X_h(i) = \sum_{i=0}^L \mathbf{W}_{i_re} x_h(i) + j\mathbf{W}_{i_im} x_h(i) \quad (4.70)$$

where $X_h(i)$ represents the i^{th} frequency component of the h^{th} segment. Then the spectral average of H number of segments frequency domain expressions can be obtained as:

$$SA(i) = \frac{1}{H} \sum_h^H (X_h(i))^2, \text{ for } i = 1, 2, \dots, k \quad (4.71)$$

Through the averaging layer, the spectral average can be obtained from the frequency magnitude of all segments. With the spectral average, the expression in time domain can be reconstructed through the pre-trained IDFT layer. Note that the weight sets in both DFT kernel and the IDFT kernel are obtained through unsupervised learning using the TSA embedded neural network. Finally, the TSR result in time domain from neural network can be expressed as:

$$\hat{x}(i) = \frac{1}{k} \sum_{i=0}^k \mathbf{W}_{o_re} SA(i) + j\mathbf{W}_{o_im} SA(i) \quad (4.72)$$

Furthermore, the supervised learning can be added in the TSR embedded neural network to reduce the influence of segment length selection. In the supervised learning procedure, the theoretic calculation results using the method by Van Hecke *et al.* (2014) can be regarded as the target. Let x_{tsa} be the theoretical calculation result of TSR, and x_{nn} be the result obtained from neural network, then the cost function of TSR embedded neural network supervised learning can be written as:

$$J(W) = \frac{1}{N} \sum_{n=1}^N \|x_{nn} - x_{tsa}\|^2 + \frac{\lambda}{2} \sum_{n=1}^N \sum_{k=0}^{N-1} \left[(w_{i_re}^{nk})^2 + (w_{i_im}^{nk})^2 + (w_{o_re}^{nk})^2 + (w_{o_im}^{nk})^2 \right] \quad (4.73)$$

The supervised learning procedure forces the neural network to learn a more generous solution to remove the noise with less influences of rotating speed. Therefore, length of segment can be estimated without neither accurate calculation nor tachometer signal.

To validate the developed TSR embedded neural network, the AE signals acquired from the bearing test rig were used. Initially, the length of segment was determined according to the equation:

$$k = 2^{\text{ceiling}[\log_2(r)]} \quad (4.74)$$

where k is length of segment and r the average number of points between shaft crossings. Once the length of segment was determined, 50 segments were randomly selected from the bearing signal with outer race fault, with k continuous data points in each segment. Then these segments

were used to train the DFT-IDFT autoencoder to obtain the TSR, spectral average, and the reconstructed signal. After the completion of TSA embedded neural network training, the complete bearing signal with outer race fault was introduced into the DFT-IDFT autoencoder. The reconstructed time domain expression of spectral average was calculated as the final output of the neural network. The real parts of the reconstructed signals obtained by the theoretical TSR and the DFT-IDFT autoencoder are provided in Figure. 4.17.

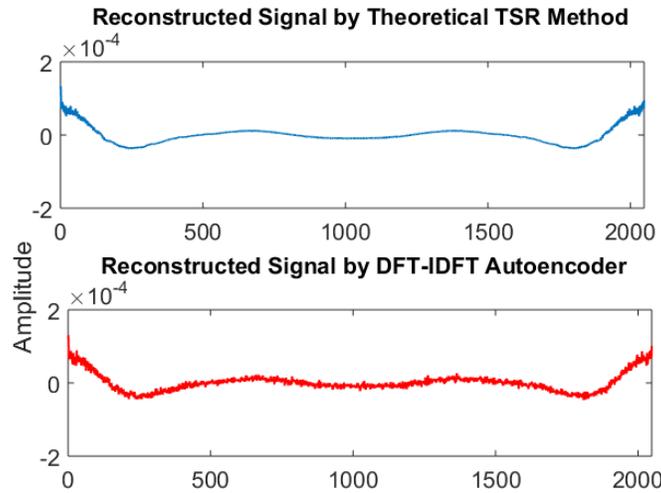


Figure. 4.17. Reconstructed signals using theoretic TSR result and DFT-IDFT autoencoder result with outer race fault bearing signal

The comparison results in Figure. 4.17 show the effectiveness of the DFT-IDFT autoencoder in computing the TSR and reconstructing the vibration signals. The RMSE between the results of the DFT-IDFT autoencoder and theoretical calculation is as low as 0.0004 for the plots shown in Figure. 4.17.

To diagnose the bearing health conditions using the reconstructed vibration signals, a classifier layer using softmax function is constructed right after the DFT-IDFT autoencoder. The output of DFT-IDFT autoencoder as the reconstructed vibration signals are fed into the softmax

classifier as the input, and the representative labels of bearing health conditions as the output. Therefore, the number of neurons in the softmax classifier remains the same as number of bearing health conditions.

5 EXPERIMENTAL SETUP

(The majority of the content in this chapter is composed of previously published work as (1) He, M., He, D., 2018, “Wind turbine planetary gearbox feature extraction and fault diagnosis using a deep learning based approach”, *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, © 2018 (SAGE Publications), Reprinted by permission of SAGE Publications. DOI: 10.1177/1748006X18768701; (2) He, M., He, D., 2018, “Simultaneous bearing fault diagnosis and severity detection using a LAMSTAR network-based approach”, *IET Science, Measurement & Technology*, DOI: 10.1049/iet-smt.2017.0528; (3) He, M and He, D., 2017, “A deep learning based approach for bearing fault diagnosis”, *IEEE Transactions on Industrial Application*, Vol. 53, No. 3, pp. 1-9, DOI: 10.1109/TIA.2017.2661250)

This section presents the experimental setup used to validate the proposed deep learning base rotating machinery fault diagnosis methodology. Specifically, there are two separate test rigs setup for validation, the bearing test rig for bearing fault diagnosis validation and the PGB test rig for gear fault diagnosis, respectively.

5.1 Introduction of the Bearing Test rig

This section presents the experimental setup used to validate the proposed AE based bearing fault diagnostic methodology. Figure 5.1 shows the experimental bearing test rig used for data collection along with indication of its main components. The motor controls the shaft which rotates the bearing located in the bearing housing. The AE sensor provides the input signal that allows the calculation of the AE envelope signal and the tachometer provides the trigger signal required for the synchronous average of the AE envelope signal. The rig also contains a hydraulic loading mechanism that affords the ability to apply a lateral load to the bearing housing if needed.

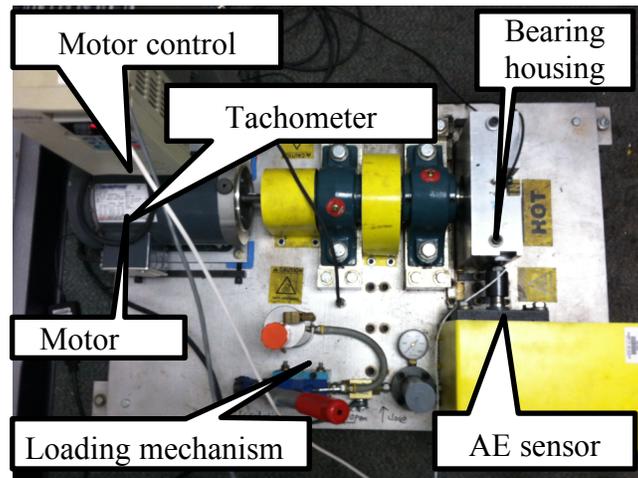


Figure 5.1. The bearing test rig

Type 6205-2RS steel ball bearing was selected for the experiment to collect validation AE data. Four single-point faults were seeded on the previously mentioned ball steel bearing, namely inner race fault, outer race fault, ball (rolling element) fault and cage fault (Van Hecke *et al.*, 2016). Location and shape of all four different faults are presented in Figure 5.2 as below. The test rig stays the same as reported by Van Hecke *et al.*(2016). According to Van Hecke *et al.* (2016): “Bearing inner race fault and outer race fault were generated by scratching the inner and outer steel race surfaces with a diamond tip grinding wheel bit to cover the ball contact surface. The dimension of seeded faults on both inner and outer race is approximately 1/16 inch in width and 1/250 inch in depth. The damage on rolling element was generated by cutting the steel cage in one of the ball positions, and then creating a dent with around 20% of volume of one steel ball using the diamond tip grinding machine.” The steel cage was cut in between two rolling ball still position to generate the cage fault, with size of about 50% of volume of one rolling ball. In the procedure of all seeded faults generation, the bearing seals and grease were removed and replaced following the implementation of the fault.



Figure 5.2. The steel bearing with seeded faults

It should be notified that the defects were intentionally seeded with sizes that were as small as possible to simulate the growing fault in propagation state, which is more frequently observed in practical industrial applications. In the propagation state, the defective fault size is likely be small and grows significantly over running time. Though the defective size in the experiment seeded faults is not that small to the incipient fault, it has been widely reported and studied in bearing fault diagnosis literature that one significant advantage of using an AE sensor is the capability of detect incipient faults in the early damage progression state.

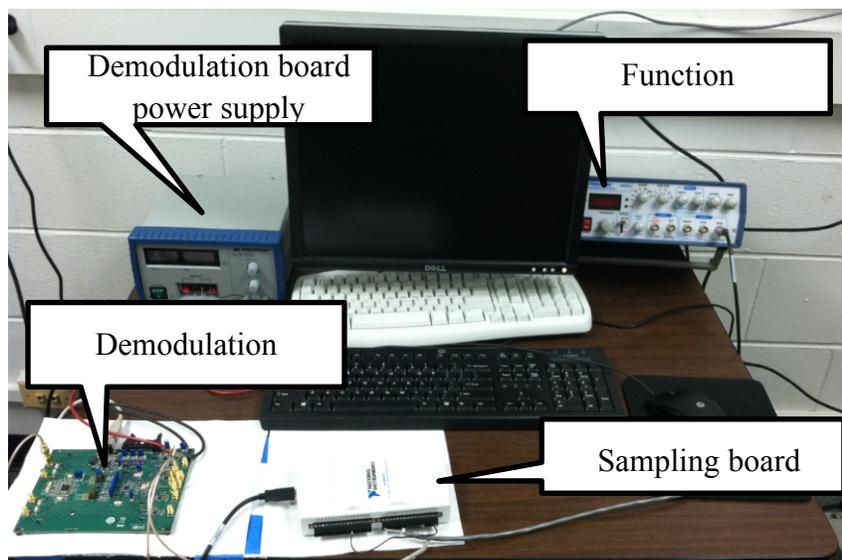


Figure 5.3. Data acquisition system: demodulation board, sampling board, and function generator

Data acquisition was accomplished via NI LabVIEW SignalExpress. All signals were sampled at 100kHz at varying durations throughout the experiment. Figure 5.3 shows the AE data

acquisition system consisting of a demodulation board, power supply, along with the function generator and sampling device. According to Van Hecke *et al.* (2016), “The demodulation board performs the multiplication of the AE sensor signal with the reference signal output from the function generator which allows a sampling frequency reduction technique to be implemented.” Two signals are imported into the demodulation board with one signal as output result. The output signal in lower frequency range as the result of down shifting the original signal is then imported into the sampling board while filtering out the high frequency component. The carrier frequency of the original AE signal needs to be determined for down shifting procedure. Thus, the goal was to determine the central AE carrier frequency and set the reference signal frequency as close to it as possible. A sweep function, created by the function generator, was used to record the output of the system. After examining the energy envelope at different frequencies, an estimate of the frequency range of the output was found. The central carrier frequency of original AE signal was determined as 400 kHz as well as the demodulation reference frequency. The sampling device is can handle a sampling frequency up to 250kS/s. For data acquisition, NI Labview signal express was employed to collect continuous AE signals with a sampling rate as 100 kHz. AE signals of the healthy bearing and the seeded faulty bearings were collected at various shaft speeds, including 2 Hz, 4 Hz, 6 Hz, 8 Hz, 10Hz, 30 Hz, and 45 Hz. At each rotating speed, there are 5 samples collected for each bearing type, leading to 35 samples in total for each bearing type. For consistency, the AE sensor was placed in the same axial location for all data acquisitions.

For the experiment with each studied shat speed, 5 samples were collected yielding to a total of 35 samples for each bearing condition. Table 5.1 presents the shaft speed (Hz) and number of samples for each of bearing conditions. The lowest shaft speed tested was 2 Hz and no loading

was applied during the entire experiment procedure. Additionally, for each fault, all data sets were recorded continuously as the shaft frequency was increased.

Table 5.1. Shaft speed and number of samples collected for all bearing conditions

Shaft Speeds (Hz)	Number of Samples/bearing condition
2	5
4	5
6	5
8	5
10	5
30	5
45	5
Total	35

5.2 Introduction of the PGB Test rig

A PGB test rig and a corresponding data acquisition were designed and set up for validation in the laboratory. Figure 5.4 shows the designed PGB test rig composed by 4 main parts, namely (1) a data acquisition (DAQ) system, (2) an AC motor, (3) the PGB, (4) and the load generator.

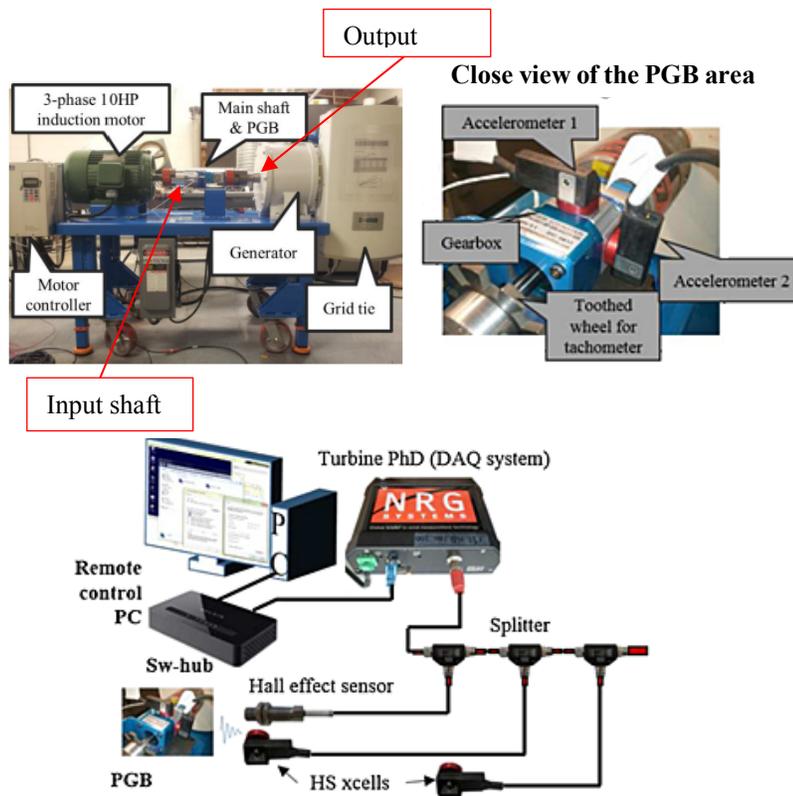


Figure 5.4. The PGB test rig for wind turbine simulator

The DAQ system contains a local data collector (LDC) (Turbine PhD produced by renewable NRG systems), two high speed accelerometers, and a tachometer. A Hall effect sensor and a toothed wheel mounted on the motor shaft were paired to record the real-time shaft rotating remarks. The detailed information of DAQ system is provided in Table 5.2. A 3-phase induction motor with 10 HP maximum loading is selected as driving motor with a motor controller. The output shaft of the gearbox is connected to a generator which serves as a load generator. The structure of the PGB test rig is like those used in a wind turbine. During the test, commercially available single stage PGB with a 5:1 speed reduction ratio was used.

Table 5.2. Vibration DAQ settings

	Vibration sensor 1	Vibration sensor 2	Tachometer
Sensor	High speed accelerometer	High speed accelerometer	Hall effect sensor
Manufacturer	NRG systems	NRG systems	Sensoronix
Sampling rate	6104 (Hz)	24414 (Hz)	1000 (Hz)

For the seeded gear fault experiment, three typical PGB gear faults were seeded on the gears. Figure. 5.5 below displays the details of seeded faults, including partial tooth cut on the sun gear, partial tooth cut on the planet gear and ring gear, separately. Both healthy and faulty gearboxes were tested under 20 various working conditions combining of 4 varying loading conditions, including 0% loading, 25% loading, 50% loading and 75% loading out of the rated torque of the PGB, and 5 varying shaft speeds including 10, 20, 30, 40 and 50 r/sec. Vibration signals were collected from each gearbox with different seeded faults sequentially. After switching one gearbox to another, the vibration sensors were mounted in the same locations on the PGB to ensure the experimental consistency.

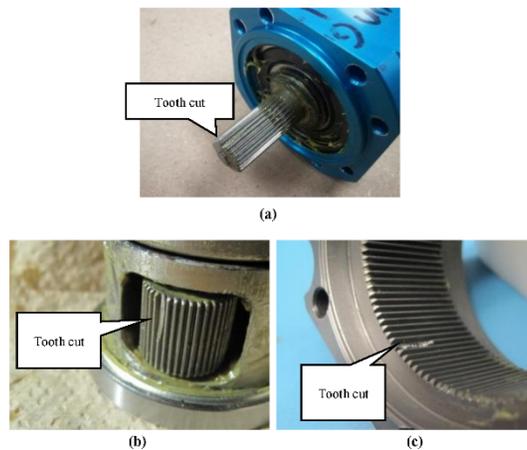


Figure. 5.5. Seeded faults: (a) sun gear fault, (b) planet gear fault, (c) ring gear fault

6 VALIDATION RESULTS

(The majority of the content in this chapter is composed of previously published work as (1) He, M., He, D., 2018, “Wind turbine planetary gearbox feature extraction and fault diagnosis using a deep learning based approach”, *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, © 2018 (SAGE Publications), Reprinted by permission of SAGE Publications. DOI: 10.1177/1748006X18768701; (2) He, M., He, D., 2018, “Simultaneous bearing fault diagnosis and severity detection using a LAMSTAR network-based approach”, *IET Science, Measurement & Technology*, DOI: 10.1049/iet-smt.2017.0528; (3) He, M and He, D., 2017, “A deep learning based approach for bearing fault diagnosis”, *IEEE Transactions on Industrial Application*, Vol. 53, No. 3, pp. 1-9, DOI: 10.1109/TIA.2017.2661250)

6.1 The Validation Results of Bearing Fault Diagnosis Using AE signals

6.1.1 The Designed LAMSTAR Neural Network for the Validation Test

Each of the AE signal dataset was transformed into a 2-D spectrum matrix as explained in Section 4.1. 10 patterns from each of the 5 bearing conditions: inner race fault, outer race fault, cage fault, ball fault and healthy condition were generated. Therefore, there were a total of 450 patterns generated for 5 bearing conditions at 9 different shaft speeds. Table 6.1 shows the detailed description of samples and patterns obtained at different shaft speeds.

Table 6.1 Detailed description of samples and generated patterns

Shaft speed (r/sec)	Number of samples	Number of generated patterns
60	5	50
45	5	50
30	5	50
15	5	50
10	5	50

8	5	50
6	5	50
4	5	50
2	5	50
Over all	45	450
Training patterns (60% of 450)		270
Testing patterns (40% of 450)		180

In the implementation of the LAMSTAR network, each spectrum matrix representing one pattern has a dimension of 250×4000 . A sliding box with size of 50×50 was selected. Thus, 40 input SOM modules were generated for 40 sub-patterns sampled from each spectrum matrix. In each input SOM module, the dynamic neurons were constructed for storing representative value of cell in matrix. Thus, the number of neurons in each input SOM module varied from 0 to 2500. Considering there are 5 bearing conditions in this study: inner race fault, outer race fault, cage fault, ball fault and healthy condition, three LAMSTAR network output neurons were used to give a complete permutation of 6 firing sequences with each sequence representing a condition. Recall Eq. (4.1), output value of neurons in the decision SOM module can be obtained by Eq. (4.1). Let y_i be the output value of each neuron in the decision SOM module, then one can assign a label of either 1 or 0 to each of neurons using the following equation:

$$L_i = \begin{cases} 1 & \text{if } y_i \geq 0 \\ 0 & \text{if } y_i < 0 \end{cases} \quad (6.1)$$

where L_i represents the label of the i^{th} neuron in the decision SOM module. Therefore, the firing sequence can be represented using the labels shown in Table 6.2 below. Table 6.2 shows the bearing conditions and their LAMSTAR network output neuron firing sequence representations. The error tolerance was set to be 10^{-9} ; and learning rate alpha was set to be 0.8 as a constant, and error tolerance for winning neuron decision in dynamic SOM weights construction was set to be

10⁻⁷. Based on the principle of neural network application, the collected dataset was divided into training and validation groups for training and validating generated network. In this paper, the ratio of training to testing dataset was 60% to 40% against total.

Table 6.2 Bearing conditions and LAMSTAR network output neuron firing sequence representations

Bearing condition	Output neuron 1	Output neuron 2	Output neuron 3
Inner race fault	0	0	1
Outer race fault	0	1	0
Cage fault	1	0	0
Ball fault	1	0	1
Healthy	1	1	0

6.1.2 The Validation Results

Using the AE signals collected during the bearing seeded fault tests, 270 patterns were used to train the LAMSTAR network, 10 for each bearing condition. Table 6.1 has shown the number of patterns used for training and testing procedure, separately. The dynamically built neurons in SOM modules enable a large reduction on training time as the search time to find the winning neuron was reduced to a small number of neurons in many cases. The neural network learns as it goes even if untrained. In addition to LAMSTAR network, another deep learning algorithm, CNN was used to perform the bearing fault diagnosis using the same datasets for comparison. Table 6.3 shows the bearing fault diagnosis results obtained by LAMSTAR and CNN at different shaft speeds.

Table 6.3 Overall accuracy at different shaft speeds

Shaft speed (r/sec)	LAMSTAR overall accuracy (%)	CNN overall accuracy (%)
60	100 (100/100/100/100/100)	92(100/90/90/90/90)
45	98(100/100/100/100/90)	92(100/90/90/90/90)
30	98(100/100/100/100/90)	90(100/80/90/90/90)

15	98(100/100/100/100/90)	88(100/80/90/90/80)
10	98(100/100/100/100/90)	88(100/90/90/80/80)
8	98(100/100/100/100/90)	88(100/90/90/80/80)
6	96(100/90/100/100/90)	84(100/80/80/80/80)
4	96(100/90/100/100/90)	84(100/80/80/80/80)
2	96(100/100/90/100/90)	80(80/80/80/80/80)

Note: in (a/b/c/d/e), a, b, c, d, and e represent diagnostic accuracy for healthy condition, inner race fault, outer race fault, cage fault, and ball fault, respectively.

From overall accuracy at different shaft speeds in Table 6.3, it can be observed that the performance of the LAMSTAR network remains almost the same when the shaft speed decreases, with a slight decrease from 100% to 96%. However, accuracy obtained by CNN drops from 92% to 80% when the shaft speeds decreases.

From columns of shaft speeds as 60, 45, 30 and 15 r/sec in Table 6.3, it can be observed that under the normal speeds, LAMSTAR network gives more accurate diagnosis performance than CNN. In addition, the time used to train a LAMSTAR network was 10 times less than CNN. As the shaft speed reduces from 60 r/sec to 15 r/sec, the diagnostic performance of the LAMSTAR network remains the same while the diagnostic performance of CNN gets worse.

As pointed out by Van Hecke *et al.* (2016), it is normally difficult to diagnose the bearing faults at low speeds in the range of 0.5 r/sec and 10 r/sec. Using a spectral averaging based approach, they only showed significant fault diagnosis results for rotating speeds over 30 r/sec. Table III also shows the diagnosis results obtained by the LAMSTAR network and CNN at relatively low speeds. As shown in the columns of shaft speeds as 10, 8, 6, 4, and 2 r/sec in Table 6.3, the LAMSTAR network shows more steady performance than CNN on fault diagnosis with decreasing rotating speeds. The classification accuracy from LAMSTAR network application decreases from 98% to 96 %, while the one from CNN drops from 88% to 80%. The results

presented in Table 6.3 show the powerful diagnostic performance of deep learning based approach for even relatively low speeds.

To show the effectiveness of using LAMSTAR network to extract AE features for bearing fault diagnosis, the AE features extracted using the LAMSTAR network were compared with the specialized AE signal processing and feature extraction techniques reported in the literature. Note that the outputs of each winning neuro of the decision SOM module in the LAMSTAR network represent the AE features similar to those condition indicators extracted using specialized signal processing techniques such as spectral averaging reported (Van Hecke *et al.*, 2014). Therefore, these outputs were used as the features extracted by the LAMSTAR network. Figure 6.1. represents average output values of winning neuron at the decision SOM module. The results in Figure 6.1 show the clear separation of the bearing conditions at all relatively low speeds from 2 r/sec to 10 r/sec.

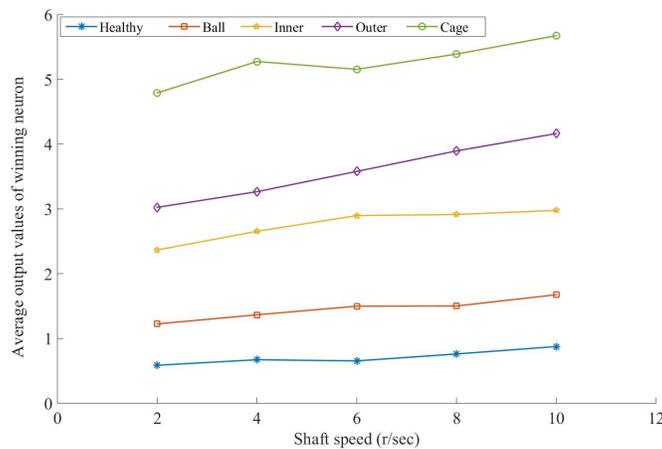


Figure 6.1. Average output values of the winning neurons at different shaft speeds.

In comparison with the fault diagnosis result on low rotating shaft speeds by using traditional signal processing and feature extraction method by Van Hecke *et al.* (2014) the deep learning based method has the advantage of requiring no specialized signal and feature extraction techniques. Due to the simplified signal processing, the large amount of monitoring data can be processed intelligently and rapidly with accurate diagnosis result.

6.2 The Validation Results of Bearing Fault Diagnosis Using Vibration Signals

6.2.1 Experimental Data Sets Used for Validation of the Methodology

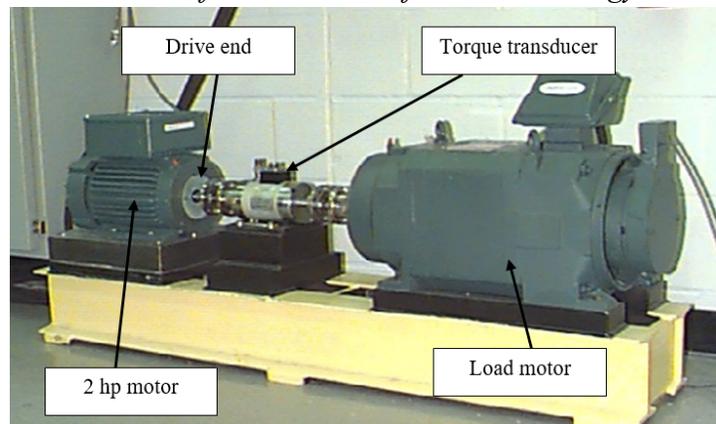


Figure 6.2. The seeded bearing test rig from Case Western University

Other than the AE signals collected on the seeded fault bearing test rig were used, the vibration signals provided by Case Western Reserve University Bearing Data Center (CWRU) were also used for validation and comparison with reported researches from other scholars. Figure 6.2. shows the seeded bearing test rig in CWRU. The main components of the experimental apparatus employed in conducting the experiments include a 2 hp motor, a torque transducer, a dynamometer, and a load motor. Three different single point faults, including outer raceway fault

(OR), inner raceway fault (IR) and ball fault (BF), with various dimension of 0.007 in, 0.014 in and 0.021 in are separately seeded on the testing drive-end bearings, respectively. SKF deep-groove ball bearing type 6205-2RS was used on test rig. Accelerometer was used to collect vibration data from test rig. During each test, accelerometer measures the acceleration on bearing housing at drive-end in the vertical direction. Relevant bearing details and fault frequencies are displayed in Table 6.4 as below.

Table 6.4. Bearing details and fault frequencies (CWRU)

Position on the rig	Model number	Fault frequencies (multiple of shaft speed)			
		BPMI	BPMO	FTF	BSF
Drive-end	SKF 6205-2RS	5.415	3.585	0.3983	2.357

In Table 6.4, bearing details and respective fault frequencies for each fault is displayed, where BPMI as ball pass frequency of inner race, BPMO as ball pass frequency of outer race, FT as fundamental train frequency (cage speed), and BSF as ball (roller) spin frequency.

Two different sampling frequencies were selected for vibration data collection separately, namely 12 kHz and 48 kHz. Also, bearing data in health state was collected under different loading of 0, 1,2, and 3 hp. The sampling frequency of health state bearing data was not stated. In sum, one health state and three defective conditions, with three defective dimensions for each of defective condition, compose the validation data set for the proposed deep learning based method.

The bearing data sets collected using vibration signal is from CWRU bearing test rig. Three defective, including outer race fault (OF), inner race fault (IF), ball fault (BF) were seeded on the tested drive-end bearing with three different sizes of 0.007, 0.014, and 0.021 inches in diameter,

respectively. A normal condition (N) dataset was also provided as the base line. Vibration data were recorded with two different sampling rates: 12 kHz and 48 kHz. For each sampling rate, vibration signals were collected under 4 different loadings: 0, 1, 2, and 3 hp. Correspondingly, the shaft rotating speed ranged from 1797 to 1720 rpm. Datasets with different sampling rates were tested separately. For each sampling rate, 100 samples were randomly selected for each bearing condition under one specific load. Thus, with various loads from 1 to 3 hp, 300 samples were selected for each bearing condition, and each sample contained 2500 data points. Detailed description of dataset is presented in Table 6.5

Table 6.5. Detailed description of vibration data sets for validation

Sampling rate	Load (hp)	Number of samples	Fault	Fault severity (in)	Index of neuron triggered in decision SOM module
12 kHz	1-3	300	N	0	1
		300	BF	0.007	2
		300	BF	0.014	3
		300	BF	0.021	4
		300	IF	0.007	5
		300	IF	0.014	6
		300	IF	0.021	7
		300	OF	0.007	8
		300	OF	0.014	9
		300	OF	0.021	10
48 kHz	1-3	300	N	0	1
		300	BF	0.007	2
		300	BF	0.014	3
		300	BF	0.021	4
		300	IF	0.007	5
		300	IF	0.014	6
		300	IF	0.021	7
		300	OF	0.007	8
		300	OF	0.014	9
		300	OF	0.021	10
48 kHz	0	100	N	0	1
		100	BF	0.007	2
		100	BF	0.014	3
		100	BF	0.021	4
		100	IF	0.007	5
		100	IF	0.014	6

100	IF	0.021	7
100	OF	0.007	8
100	OF	0.014	9
100	OF	0.021	10

6.2.2 The Validation Results

The dimension of the time-frequency matrix processed with STFT was 1025 by 100. A sliding box with a size of 25 by 25 was computed using Eq. (4.30). Thus, 164 sub-patterns were generated for each time-frequency matrix. Accordingly, 164 input SOM modules were generated to construct the LAMSTAR network. 10 neurons were constructed in the decision SOM module represent 10 different bearing conditions to be diagnosed. 60% of data were used for training and 40% for testing. Initially, the data collected with the sampling rate of 48 kHz under 0 hp were tested. The LAMSTAR network based mode was run with MATLAB 2015B on a PC: Intel Core i7-3770 CPU, 1T hard drive, 8G memory. The test results are provided in Table 6.6.

Table 6.6. Diagnostic results with constant loading of 0 Hp and sampling rate as 48 kHz

Fault type	Fault severity (in)	Classification accuracy (%)
N	0	98
BF	0.007	98
BF	0.014	100
BF	0.021	99
IF	0.007	99
IF	0.014	100
IF	0.021	98
OF	0.007	100
OF	0.014	100
OF	0.021	100
Over all		99.2

The classification accuracy φ is computed as:

$$\varphi = \frac{I}{J} \times 100\% \quad (6.2)$$

where I represents the number of successfully classified data points and J the total number of data points.

The test results in Table 6.6 show the capability of the LAMSTAR network on classifying various bearing faults under different severity situations. In comparison with the fault classification results by Gan *et al.* (2016), the classification accuracy obtained using the LAMSTAR network based method is at the same level as the one obtained using combined wavelet transformation and HDN. However, the HDN was implemented with manually extracted features using wavelet transform. Furthermore, the fault diagnosis and fault severity detection were trained in two separate stages. Since such hierarchical structure needs human intervention to train the network by stage, fault diagnosis and severity detection cannot be completed automatically.

Constant loading rarely happens in the practical industrial applications. Thus, mixed loading datasets were tested. Table 6.7 and Table 6.8 present the diagnostic results with sampling rate as 12 kHz and 48 kHz, respectively. The test results in Table 6.7 and Table 6.7 show the capability of the LAMSTAR network based method on classifying various bearing faults at different levels of severity. The overall classification accuracy reached as high as 99.23% with the sampling rate of 48 kHz, similar to the results obtained from the same dataset by using DNN (Jia *et al.*, 2016). The classification accuracy obtained with the sampling rate of 12 kHz is slightly lower than the one with the sampling rate of 48 kHz. This is reasonable as the data collected with a higher sampling rate normally contain more information. Comparing with the results reported by Jia *et al.* (2016), the diagnostic and detection accuracy presented in Table 6.8 is compatible. Furthermore, the DNN method (Jia *et al.*, 2016) relies greatly on the supervised fine-tuning process using classical backpropagation algorithm to get satisfactory results. The proposed LAMSTAR based approach can achieve the same level of accuracy without extra fine-tuning requirement and

time. In the age of big data, LAMSTAR based approach can provide more efficient solutions for bearing fault diagnosis and severity detection.

As stated by Smith and Randall (2015), the vibration signals collected at CWRU have degrees of diagnosability varying from easily diagnosable to impossibly diagnosable with different applied methods. The studied diagnosis methods included classical vibration signal analysis methods such as envelope analysis, cepstrum analysis, and benchmark model. The validation results showed that these classical signal analysis methods cannot identify fault location and severity effectively. Furthermore, their study only focused on fault diagnosis while this study conducted simultaneous fault diagnosis and severity detection without manual features extraction using deep learning based approaches. In comparison, the results reported by Smith and Randall (2015) showed that all the BF signals and small portion of IF and OF signals cannot be diagnosed by using traditional signal processing method. The presented LAMSTAR based approach shows more competitive performance on fault diagnosis and severity detection.

In comparison of the results obtained with constant loading of 0 hp in Table 6.6 with those obtained with the mixed loadings ranging from 1 to 3 hp in Table 6.8, one can observe that the diagnostic and detection accuracy obtained by the LAMSTAR network remains stable regardless of the working conditions. This comparison result shows the robustness of the LAMSTAR network based approach for simultaneous bearing fault diagnosis and severity detection.

In the data used in this paper, each fault severity is defined with a specific size, say 0.007, 0.014, and 0.021 inches in diameter. In real industrial applications, the fault severity is normally specified with different levels, for example, light damage with size in the interval $[0, 0.007]$, medium damage $[0.007, 0.014]$, heavy damage $[0.014, 0.021]$, and broken (> 0.021). In the case

when the training is composed just with light damage level and heavy damage level, and the medium damage level is presented during the test, then the diagnostic result will depend on how much the testing data is close to the training data. If more testing data is closer to the light damage level, then the testing data will be diagnosed as light damage level. Otherwise, the testing data will be diagnosed as heavy damage level. To make accurate diagnosis, the LAMSTAR network needs to be designed to process the temporal information in the data to keep tracking the degradation level of the fault over time. In other words, prognostic capability should be embedded into the LAMSTAR network. This will be a research topic for further investigation.

Table 6.7. Diagnostic results with mixed loading conditions and a sampling rate of 12 kHz

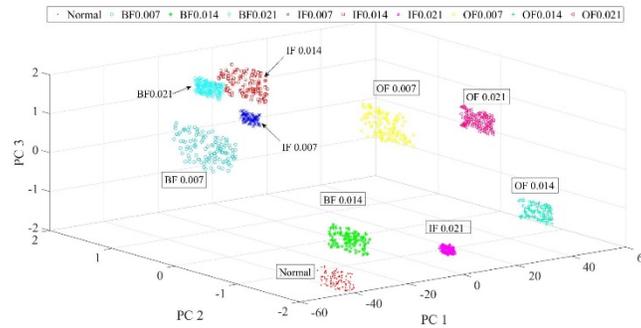
Fault type	Fault severity (in)	Classification accuracy (%)
N	0	100
BF	0.007	97
BF	0.014	98
BF	0.021	97
IF	0.007	98
IF	0.014	96
IF	0.021	98
OF	0.007	96
OF	0.014	97
OF	0.021	95
Over all		97.2

Table 6.8. Diagnostic results with mixed loading condition and a sampling rate of 48 kHz

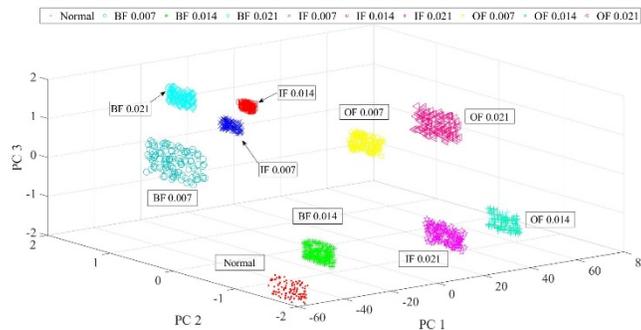
Fault type	Fault severity (in)	Classification accuracy (%)
N	100	100
BF	99	97
BF	98.67	98
BF	98.67	97
IF	99.67	98
IF	99.33	96
IF	99.33	98
OF	99.67	96

OF	98.67	97
OF	99.33	95
Over all		99.23

Note that the outputs of the winning neurons in the input SOM modules were fed into the output layer to diagnose the bearing faults and detect fault severity simultaneously. Thus, the outputs of the winning neurons in the input SOM models represent the bearing health condition features from the vibration signals extracted by the LAMSTAR network. Principle components obtained from the outputs of the winning neurons were extracted using the principle component analysis (PCA) method and are plotted in Figure 6.3 and Figure 6.4. Note that since there are many principle components can be obtained by PCA, here only the first three principle components that account for more than 97% of the total variation were used to make the plots.



(a)



(b)

Figure 6.3. Plots of the principal components of the winning neuron outputs using data collected at sampling rate of:(a) 12 kHz and (b) 48 kHz

Figure 6.3 (a) and Figure 6.3 (b) show the plots of the first three principal components of the winning neuron outputs in the input SOM modules in a 3-D space using the data collected with the sampling rate of 12 kHz and 48 kHz, respectively. From Figure 6.3 (a) and Figure 6.3 (b), one can see that clusters are formed according to the bearing conditions regardless the sampling rate used in collecting the data. However, with a higher sampling rate, clusters of the bearing conditions are more distinct.

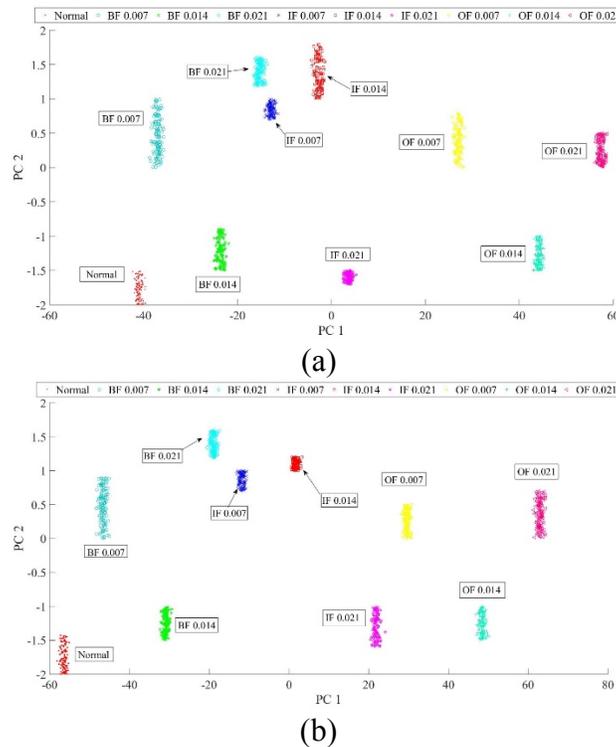


Figure 6.4. Plots of the principal components of the winning neuron outputs using the data collected at the sampling rate of:(a)12 kHz and (b) 48 kHz.

Figure 6.4 (a) and Figure 6.4 (b) show the plots of the first two principal components of the winning neuron outputs in the input SOM modules in a 2-D space using the data collected with the sampling rate of 12 kHz and 48 kHz, respectively. The results in Figure 6.4 (a) and Figure 6.4 (b) show that clusters are formed according to the bearing conditions regardless the sampling rate

used in collecting the data. Moreover, it can be observed that the same bearing fault with increasing fault size distributed increasingly along PC 1 axis. It can be concluded that the first principle component represents the feature of the bearing fault severity. To show the percentage of the variances in the original feature sets captured by the principle components, the accumulated variance of the i^{th} principle component can be calculated as:

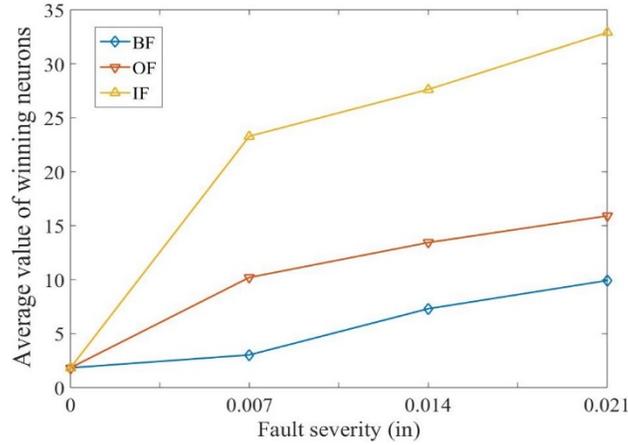
$$c_i = \frac{\sum_1^i \text{var}(\mathbf{p}_i)}{\sum_1^N \text{var}(\mathbf{p}_i)} \quad (6.3)$$

where c_i represents the accumulated variance of the i^{th} principle component \mathbf{p}_i , $\text{var}(\mathbf{p}_i)$ the variance of the i^{th} principle component \mathbf{p}_i , N the total number of principle components. The accumulated variance of the first three principle components is shown in Table 6.9. It can be observed that the first three components account for more than 97% of the variance of the original dataset.

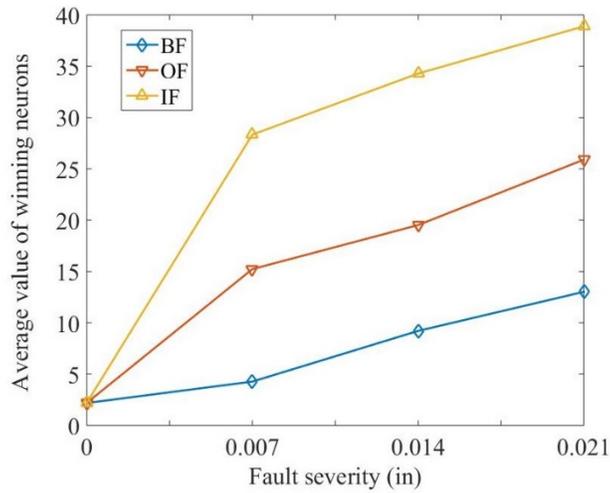
Table 6.9. Accumulated variance of the first three principle components

Sampling rate	Accumulated variance		
	1 st PC	2 nd PC	3 rd PC
12 kHz	0.64	0.89	0.97
48 kHz	0.79	0.94	0.99

One key aspect in determining whether a fault severity detection method is effective is to check to see if the extracted features can represent the monoatomic trend of the degradation process. To show that, the average output value of the wining neurons belonging to different bearing fault severity levels were calculated and are presented in Figure 6.5 (a) and Figure 6.5 (b) with the sampling rate of 12 kHz and 48 kHz, respectively. Note that the value of 0 fault severity represents the normal bearing condition.



(a)



(b)

Figure 6.5. Average output value of the winning neurons using data collected at the sampling rate of:(a)12 kHz and (b) 48 kHz.

As shown in Figure 6.5 (a) and Figure 6.5 (b), as the bearing fault severity increases, the average value of the output neurons increases. The results shown in Figure 6.5 (a) and Figure 6.5 (b) indicate that the features extracted using the LAMSTAR neural network can be used as a good indicator of bearing fault severity. Also, the magnitudes of the average output values have shown a distinct difference for different type of faults. It is interesting to see that the average output value of the normal bearing condition is close to that of BF 0.007, using both 12 kHz and 48 kHz sampling rates. This could lead to the misclassification of the two bearing conditions in the

LAMSTAR neural network. Note that no ability to capture the trending characteristic of the bearing degradation by the features extracted with deep learning based approaches was reported in references (Gan *et al.*, 2016; Jia *et al.*, 2016) and in the literature. The results presented in Figure 6.5 (a) and Figure 6.5 (b) have shown the significant advantage of the presented method in representing the characteristic trend of the bearing degradation process over the existing deep learning based approaches reported in the literature.

To demonstrate the robustness of the proposed method on simultaneous bearing fault diagnosis and severity detection, more experiments have been conducted with various hyperparameters of the LAMSTAR network. Firstly, the experiments to investigate the effects of window size on diagnosis results have been conducted with mixed loading conditions and at a sampling rate of 48 kHz. The investigated window size was selected in the range from 5 to 50, with a step length of 5. For each selected window size, 5 experiments were performed. The diagnosis accuracies using the testing dataset are plotted in Figure 6.6. Note that the diagnosis accuracy Figure 6.6 was computed as the average of the accuracies obtained from 5 experiments. It can be observed from Figure 6.6 that the highest diagnosis accuracy was obtained at the window size of 25. This window size of 25 is the optimized window size determined by Eq. (4.30). Starting with window size of 5, diagnosis accuracy climbs up from 95.47% to 99.91% as the window size approaches to 25. The diagnosis accuracy drops slightly with a fluctuation as the window size continues to increase beyond the optimized window size. Figure 6.6 shows that LAMSTAR network can achieve the best diagnosis results using the optimized window size. The diagnosis performance of the LAMSTAR network becomes relatively stable with a window size greater than the optimized one.

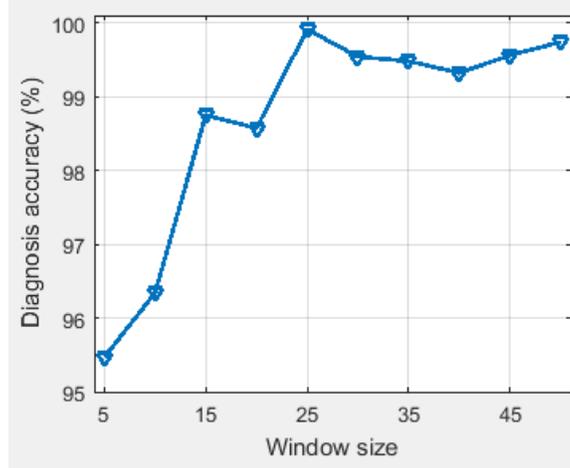


Figure 6.6. Diagnosis results comparison of different window sizes

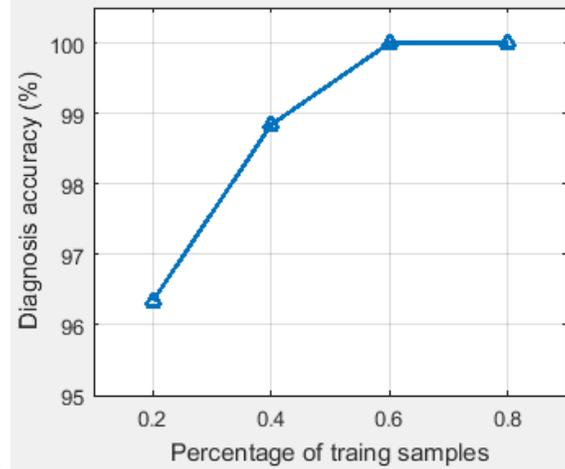


Figure 6.7. Diagnosis results comparison with different percentages of samples for training.

Furthermore, considering that insufficient labelled training data is a critical issue on data-driven fault diagnosis methods, the experiments with different percentages of training samples were also conducted. Similarly, for each training data percentage ranging from 20% to 80%, 5 experiments were performed. The diagnosis accuracies with the testing dataset are plotted in Figure 6.7. An obvious increase can be observed with increasing percentage of training samples. The diagnosis accuracies remain 100% as the training percentage reaches over 60%. Note that even with 20% training samples, the diagnosis accuracy can still reach at satisfactory level of 96.34%. It shows that the LAMSTR network can automatically learn the fault features from

relatively small amount of training data and gives accurate diagnosis results. In general, the proposed LAMSTAR network shows good performance robustness with slight changes on both the hyperparameter and percentage of the training samples.

6.2.3 Discussion

Based on the results presented in the previous section, a few points can be made as follows:

- (1) Different from model based bearing fault diagnosis methods, the presented method does not require model-based signal processing techniques to reveal the characteristics of the bearing faults. As reported by Smith and Randall (2015) the traditional vibration signal analysis methods including envelope analysis, cepstrum analysis and benchmark model cannot diagnose all the signals with BF and small portion of signals with IF and OF using the same dataset from CWRU. In a comparative study by Boudiaf *et al.* (2016), multiple vibration signal processing techniques, including FFT, cepstrum analysis, envelope analysis and wavelet transform, were implemented on the same data set for bearing fault diagnosis. It has drawn a similar conclusion that these traditional vibration signal processing techniques cannot diagnose the faults accurately. Besides, envelope analysis and wavelet transform require expertise knowledge of suitable selection of critical parameters such as filtering band and mother wavelet. Furthermore, their study only focused on fault diagnosis while this paper emphasized on simultaneous fault diagnosis and severity detection using deep learning based approaches.
- (2) The results presented in Figure 6.5 (a) and Figure 6.5 (b) have shown the ability to capture the trending characteristic of the bearing degradation with the features extracted using the method presented in this paper. As the fault severity increases, the average value of the winning neuron outputs increases. It is the first time that the monotonic features are extracted and presented to characterize the bearing degradation trending using deep learning based approach.

- (3) The results in Figure 6.6 and Figure 6.7 have shown that the proposed method can reach satisfactory diagnosis accuracies with even slight changes on both the hyperparameter and the percentage of the training samples. The maximum change in diagnosis accuracy is within 4% with the changes on window size and percentage of training samples.
- (4) In addition, in comparison with the deep learning based approaches applied to the same data set (Gan *et al.*, 2016; Jia *et al.*, 2016), the method presented in this paper can provide the same level of accuracy for simultaneous fault diagnosis and severity detection but requires no supervised fine tuning and trial and error experiment on setting the network structure parameters. This feature of the presented method adds significant advantages over the other method as it can provide more efficient solutions for bearing fault diagnosis and severity detection in the age of big data

6.3 The Validation Results of PGB Fault Diagnosis Using Vibration Signals

6.3.1 The Designed LAMSTAR Network and LAMSTAR-DL Network for the Validation Test

To compare the proposed LAMSTAR-DL based approach with the original LAMSTAR network based approach for PGB feature extraction and fault diagnosis, the vibration data collected from PGB seeded fault tests were used. The vibration signals were collected at sampling frequency of 24 kHz, under 5 various shaft speeds including 10, 20, 30, 40, and 50 r/sec. In addition to the shaft speeds variation, varying loading conditions were applied at the output shaft of the gearbox, namely 0%, 25%, 50%, and 75% of the maximum torque of the PGB. The faults were seeded artificially on different locations, including planet gear, sun gear, and ring gear, separately. Thus, 4 PGB health conditions were studied in total for classification, namely healthy, planet gear fault, sun gear fault and ring gear fault. For each health condition, 5 samples were collected under each specific shaft speed and specific loading. Given 5 different shaft speeds and 4 different loadings,

100 samples were collected in total for each specific PGB health condition. With length of 1024 data points in each pattern, 5 patterns were generated from each sample by random data segments. Thus, for 4 health conditions at 5 different shaft speeds and 4 different loading conditions, 2000 patterns were generated for validation. Table 6.10 shows the detailed description of samples and patterns obtained for validation.

Table 6.10. Detailed description of collected samples and generated patterns

		Number of samples at various loadings				Number of generated patterns
		0% max torque	25% max torque	50% max torque	75% max torque	
Shaft speeds (r/sec)	10	20	20	20	20	400
	20	20	20	20	20	400
	30	20	20	20	20	400
	40	20	20	20	20	400
	50	20	20	20	20	400
Total		100	100	100	100	2000
Number of training patterns (55% of total patterns)						1100
Number of testing patterns (45% of total patterns)						900

In the implementation of the original LAMSTAR network based approach for PGB fault diagnosis, the collected raw vibration signals were pre-processed firstly to obtain the frequency spectrum matrixes by using STFT method. The dimension of the generated spectrum matrix representing original raw vibration signals is 500×1000 . The dimension of sliding window for sub-pattern generation was decided as 100×100 , leading to that 50 sub-patterns were sampled from one spectrum matrix in total. Correspondingly, 50 input SOM modules were designed in the LAMSTAR network in total. The neurons were dynamically constructed in each of the input SOM module to store and process the representative sub-pattern sampled from matrix. Due to the employment of dynamic neurons, the number of neurons in each input SOM module as well as the

dimension of the input SOM module was varied in range of 0 to 10000. The other learning parameters of LAMSTAR network were set as following: error tolerance of decision SOM module as 10^{-9} , error tolerance for making decision of winning neuron in dynamically constructed SOM weights as 10^{-7} , learning rate α as 0.8.

To implement the LAMSTAR-DL network based PGB fault diagnosis method on the same data set, the raw vibration signals of 500 generated patterns were applied directly without signal pre-processing procedure in the LAMSTAR-DL network. Thus, each pattern with length of 1024 data points was used for both dictionary learning and PGB fault location diagnosis. Based on the selection criterion expressed in Eq. (4.30), the sliding window was selected with length as 100. Thus, 10 SOM input modules were constructed for 10 sub-patterns sampled from each raw data, leading to the 10 atoms in the dictionary for each PGB health condition class. Similarly, the neurons in each input SOM module were generated dynamically ranging from 1 to 100. Particularly, the input SOM module with only 1 neuron can be redundant. The sparse representations of original vibration signals were channeled into the classifier using LAMSTAR network row by row. Therefore, the number of SOM modules in the classifier LAMSTAR network were constructed the same as number of rows of original vibration signals sparse representations. Considering there are 4 different PGB health conditions in this paper, 2 neurons were designed in the output SOM module for both original LAMSTAR network and LAMSTAR-DL network classifier for a complete 4 firing sequences. Each firing sequence represents one health condition. The output value of neurons in the output SOM module can be calculated by using Eq. (4.1). Let y_i be the output value of the i^{th} neuron in the output SOM module, then firing condition of the i^{th} neuron can be obtained from Eq. (6.1). Thus, firing sequence and the corresponding PGB health condition can be defined in Table 6.11.

Table 6.11. PGB health conditions and LAMSTAR network based approaches firing sequence representations

PGB health condition	Output neuron 1	Output neuron 2
Healthy	0	0
Planet gear fault	1	0
Sun gear fault	0	1
Ring gear fault	1	1

6.3.2 The Validation Results

The vibration signals collected from PGB test rig were processed using two methods: LAMSTAR based method and LAMSTAR-DL based method. To implement the LAMSTAR based PGB fault diagnosis method on the selected signals, the dimension of spectrum matrix processed with STFT was 500×1000 . The sliding window was decided with size as 100×100 , resulted in 50 sub-patterns sampled from each spectrum matrix. Therefore, 50 input SOM modules were designed in the LAMSTAR network. In the implementation of the proposed LAMSTAR-DL based PGB fault diagnosis method on the collected signals, the raw vibration signals were directly used for dictionary learning and sparse coding without other signal process procedures. The dimension of each pattern applied in the LAMSTAR-DL network is 1024. A sliding window with length of 100 was selected. Therefore, 10 input SOM modules were constructed corresponding to the 100 sub-patterns sampled from each pattern. There were 2 neurons designed in the output SOM module for both LAMSTAR and LAMSTAR-DL based approaches validation, representing 4 different PGB health conditions.

To show the effectiveness of the features extracted using LAMSTAR-DL network, the average value of winning neurons in each SOM module of LAMSTAR-DL network were obtained for each health condition. Figure 6.8 and Figure 6.9 show the plots of winning neurons' average

values in two different LAMSTAR based networks. In Figure 6.8 and Figure 6.9, each point and the vertical bar represent the averaged winning neuron values and the 95% confidence interval.

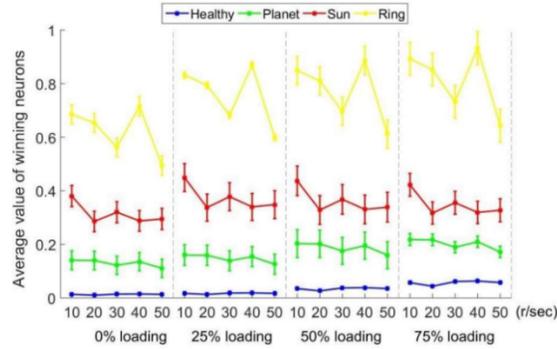


Figure 6.8. Average value of winning neurons using LAMSTAR network

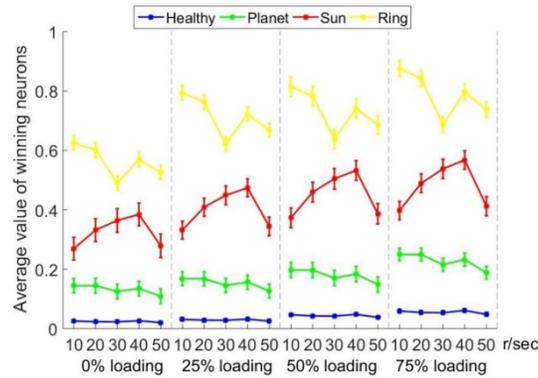


Figure 6.9. Average value of winning neurons using LAMSTAR-DL network

As one can observe in both Figure 6.8 and Figure 6.9, by calculating the average value of winning neurons in each SOM module of LAMSTAR based network, the 4 PGB health conditions can be separated with clear boundary. With the increasing loading, the division of gear health conditions is enhancing. The detectability of PGB faults gets better as loading increases. Furthermore, the average value of winning neurons stays in similar pattern within each loading, regardless changing shaft speeds. It indicates that the detection of PGB health conditions using the presented deep learning based methodologies is affected heavily by loading level of PGB. The

introduction of sparse coding into LAMSTAR shows more clear classification on PGB health conditions.

The collected vibration signals were partitioned into training and testing datasets for training and testing, respectively. 55% of the collected vibration signals were used for training and the rest for testing. Classification experiments were executed 5 times to get average diagnosis accuracy results. The fault diagnosis results obtained by using both LAMSTAR network and LAMSTAR-DL network are shown in Table 6.12.

Table 6.12. PGB fault diagnosis results using LAMSTAR and LAMSTAR-DL

PGB Faults	LAMSTAR	LAMSTAR-DL
Healthy	98.32	99.81
Planet gear fault	96.81	99.75
Ring gear fault	97.25	99.62
Sun gear fault	97.82	99.24
Over all	97.55	99.61

It can be observed from Table 6.12 that fault diagnosis accuracy achieves 97.55% overall by using LAMSTAR network alone. Higher accuracy was achieved by using the proposed LAMSTAR-DL, which is high as 99.61%. It shows the capacity of LAMSTAR and sparse coding on fault information extraction without complicated signal processing techniques. Regarding the LAMSTAR network as a model closer to the vanilla SOM model, the better classification performance obtained by LAMSTAR-DL network shows the advantage of LAMSTAR-DL over the simple SOM model. With limited amount of studies on SOM based fault diagnosis, the simple SOM based methods suffer from the weaknesses including: (1) time consuming (Katunin *et al.*, 2015), (2) weak identification ability (Zhong *et al.*, 2005), (3) dependence on faulty feature extraction and selection (German *et al.*, 2014).

In comparison with diagnosis results obtained by using the reported tradition signal processing technique (Bajric *et al.*, 2016), the presented method achieved higher accuracy without features extraction process. The presented method accomplishes simultaneous feature extraction and fault diagnosis process. Even in the comparison with other reported deep learning based PGB fault diagnosis (Chen *et al.*, 2017; Jia *et al.*, 2016), the presented method achieves higher diagnosis accuracy. Furthermore, both methods require extra supervised fine-tuning process (Chen *et al.*, 2017; Jia *et al.*, 2016). In the environment of wind turbine monitoring big data, such extensive processes can greatly reduce the computational efficiency. The proposed LAMSTAR-DL method for PGB fault diagnosis does not require the extra supervised fine-tuning to achieve the satisfactory diagnosis accuracy. Therefore, the LAMSTAR-DL method is more applicable in practical. Additionally, the structure of LAMSTAR-DL network is designed according to the characteristics of signals as explained in Eq. (4.30). In comparison with the trial-and-error test reported in the deep learning based fault diagnosis methods (Chen *et al.*, 2017; Jia *et al.*, 2016), LAMSTAR-DL reduces the computational complexity with more steady performance. Also, the method presented by Chen *et al.* (2017) implements the deep learning network simply as a classifier using features extracted with complicated signal processing techniques. The performance of such method that combines deep learning based methods and complicated feature extraction relies greatly on the complicated signal processing techniques and expertise.

In comparison with performance of features extracted from same vibration data sets using time synchronizing average (TSA) methodology (Yoon *et al.*, 2016), the average value of winning neurons in LAMSTAR shows the similar detectability on PGB faults as TSA root mean square (RMS) condition indicator (CI). Among the effective CIs, i.e. TSA RMS, TSA peak to peak (P2P), residual RMS, residual P2P, TSA RMS show the best separation results. Other CIs show the good

diagnosis results with 25% and higher loading condition, but not as good results as TSA RMS with 0% loading condition. Thus, it can be concluded that features learnt and extracted by using deep learning based method perform better on PGB fault diagnosis than CIs extracted by traditionally signal processing techniques.

6.4 The Validation Results of Deep Hybrid Signal Processing Method on Bearing Fault Diagnosis using Vibration Signals

6.4.1 Experimental Setup and Data Acquisition

This section covers the experimental setup used to validate the hybrid deep signal processing approach for bearing fault diagnosis using vibration signals. Figure 6.10 shows the bearing test rig used to conduct the steel bearing seeded fault tests. A model 608A11 Industrial accelerometer was mounted on the bearing housing for vibration data acquisition (Van Hecke *et al.*, 2014).

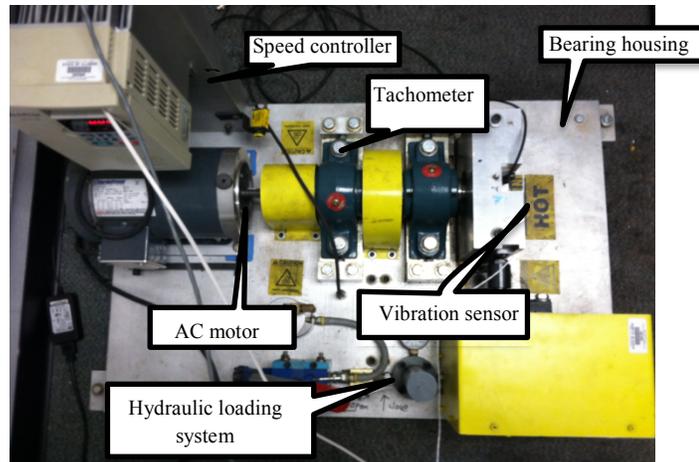


Figure 6.10. The bearing test rig

Type 6205-2RS steel FAG ball bearings were used for the testing. Four fault types were simulated on steel bearings: inner and outer race faults, rolling element fault, and cage fault. The inner and outer race faults were generated by scratching the steel race surfaces with a diamond tip grinding wheel bit to cover the ball contact surface. The scratches on both races were about 1/16

inch wide and 1/250 inch deep. The ball fault damage was created by cutting the steel cage in one of the ball locations and then using the diamond tip grinding wheel bit to create a small dent in one of the steel balls. The dent was about 20% of the ball volume. For the cage fault, the steel cage was cut in between two ball locations. The cut was about 50% of the ball diameter. For all seeded fault tests, the bearing seal and grease was removed and replaced following the creation of the fault.

For vibration acquisition, VibraQuest Pro by SpectraQuest, Inc. was used to digitize the continuous vibration signals at a sampling rate of 102.4 kHz (Van Hecke *et al.*, 2014). Vibration signals for the healthy bearing and the seeded fault bearings were recorded during the tests at four shaft speeds: 15 Hz, 30 Hz, 45 Hz, and 60 Hz. No load was applied for the duration of the seeded fault tests.

6.4.2. The Validation Results

To validate the proposed hybrid deep signal processing method on bearing fault diagnosis, AE data signals were acquired from the bearing test rig and used in different experiments. The tested bearings were operated at 4 different shaft rotating speeds: 15 Hz (900 rpm), 30 Hz (1800 rpm), 45 Hz (2700 rpm), and 60 Hz (3600 rpm). The loading condition can be regarded as nonstationary during the experiment. There were 5 bearing health conditions involved in the experiment: normal bearing, bearing with ball fault, bearing with cage fault, bearing with inner race fault, and bearing with outer race fault. The AE sensor mounted on the bearing housing collected the data with a sampling frequency of 100 kHz. For each operation condition, 400 samples were collected with a length of 10000 for each sample. Therefore, the bearing dataset contains 8000 samples under 4 different shaft rotating speeds for 5 bearing health conditions. Two different experiments were conducted. The deep structure used in the experiments was designed as: one input layer (10000 neurons), one convolutional layer with the optimal segment length, one

average pooling layer with optimal segment length, one IDFT layer optimal segment length, and one output layer (5 neurons). Note that layers including the convolutional layer, the average pooling layer and the IDFT kernel were built dynamically according to the signal with adjustable size. Initially, the samples collected under the same operation condition were used to carry the experiment. Then, the experiments were carried on the samples under the mixed operation conditions. For both experiments, 20% of the samples were randomly selected to obtain the segment length in the optimized length unit (OLU) and train the DFT encoder, the rest of samples were used for testing. Note that the classification accuracy of each experiment was obtained as average from 5 times conductions to reduce the effect of randomness. The experiments were carried out with MATLAB 2015B on a PC: Intel Core i7-3770 CPU, 1T hard drive, 8G memory. The classification results are presented in Table 6.13.

Table 6.13 Classification results with different shaft rotating speeds

Shaft speeds	Theoretical segment length	Optimized segment length	Training accuracy (%)	Testing accuracy (%)	Computational time (min)
15 Hz	7056	6700	99.91	99.82	85.23
30 Hz	3467	3200	99.97	99.92	44.16
45 Hz	2301	2100	100	100	33.13
60 Hz	1729	1800	100	100	20.67
Mixed	NA	3125	99.94	99.92	39.21

It can be observed from Table 6.13 that optimized segment length from OLU is close to the theoretical calculation result. It shows the capability of OLU on selecting the properly optimized segment length. The training and testing accuracies shown in the Table 6.13 are high overall, without obvious observation of overfitting result. The classification accuracies have shown that the proposed hybrid deep signal processing method can robustly obtain the bearing health

conditions classification result. However, also as shown in the Table 1, the computational time for the proposed method is extremely high. This is caused by the large architecture in both OLU and DFT encoder, leading to the huge number of parameters to be updated during the training procedure. To evaluate the influence of segment length on the classification results, experiments with pre-defined segment length were conducted. During the experiments with pre-defined length of segment, training procedure starts with training the DFT encoder unit. The classification results are shown in

Table 6.14.

Table 6.14 Classification results with various pre-defined segment length

Shaft speed	Length of pre-defined segment (Testing accuracy (%) /computational time (min))			
	500	1000	1500	2000
15 Hz	97.14/ 6.13	98.97 / 12.63	99.34 / 16.34	99.42 / 21.21
30 Hz	96.82 / 6.17	99.23 / 12.63	99.41 / 16.92	99.48 / 22.27
45 Hz	98.96 / 6.67	99.15 / 12.62	99.37 / 16.85	99.73/ 21.92
60 Hz	99.45 / 6.12	99.89 / 12.61	99.68 / 16.23	99.98 / 21.93
Mixed	98.91 / 6.78	99.22 / 12.78	99.59 / 16.54	99.88 / 22.06

It can be observed that the computational time drops significantly when the segment length decreases. The testing accuracy of samples from 30 Hz decreases mostly with the minimum segment length. Segment length has relatively less impact on the samples with the mixed shaft rotating speeds. In comparison with the classic TSA related method, the proposed method can obtain the good classification results with little impact of selection on segment length.

To further evaluate the adaptive learning capability of the hybrid deep signal processing method, the experiments with signals under different operation conditions for training and testing procedures were also conducted. Since the network is constructed and applied on signals with different shaft speed, the optimal filter results of signals with mixed shaft speeds in Table 6.13 as 3125 was selected. Experiments with two scenarios were conducted: (1) training with high shaft speed signal and testing with low shaft speed; (2) training with low shaft speed signal and testing with high shaft speed. With the acquiesced bearing data, 15 Hz was regarded as low shaft speed and 60 Hz as high shaft speed. The validation results obtained from 5 times of executions are presented in Table 6.15.

Table 6.15 Classification results with different operation conditions for training and testing datasets

Training condition	Testing condition	Training accuracy	Testing accuracy
Low shaft speed	High shaft speed	99.89%	99.56%
High shaft speed	Low shaft speed	100%	99.98

The classification results displayed in Table 6.15 show that the hybrid deep signal processing method can extract the faulty information from signals under different shaft speed other than the one used for training adaptively. The extracted information can be used to accurately classify the bearing health conditions.

As introduced previously, the feature learning process in deep learning related machinery fault diagnosis topic remains ambiguous with attempted discussions. Recall Eq. (4.57), the value of the i^{th} output neuron is obtained by adding the inner product of the signal segment x_h and the i^{th} row vector $\mathbf{W}_{i_{re}}^i$, and signal segment x_h and the i^{th} row vector $\mathbf{W}_{i_{im}}^i$. Therefore, both $\mathbf{W}_{i_{re}}$ and $\mathbf{W}_{i_{im}}$ can be viewed as a set of normal orthogonal bases. Thus, the encoding procedure can

be regarded as decomposing signal segment in the normal orthogonal systems. 9 rows of the W_{i_re} trained with the mixed shaft speed signals are plotted in Figure 6.11 (a), along with the corresponding frequency spectra in Figure 6.11 (b). The local spikes can be observed in the time domain and the spectral spike in the frequency domain, showing the time-frequency properties of these weight vectors. The local spikes in the time domain have the specific patterns that can capture the various local characters of mechanical signals. Correspondingly, the narrow spectral band width observed in the frequency domain enables these weight vector to serve as good bandpass bases for the mechanical signals.

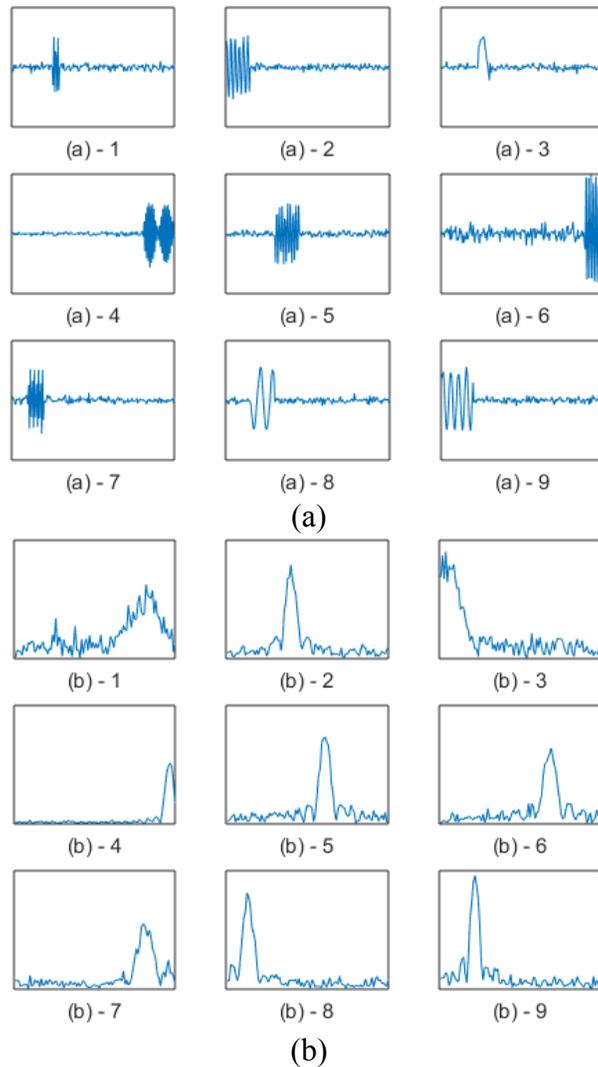


Figure 6.11. Visualization of randomly selected row vectors in $W_{i,re}$ trained by signals with mixed shaft speeds

Also, the data used for classification in the softmax layer were investigated. By calculating the root mean square (RMS) of the features from each sample, the boxplots of RMS are plotted and grouped by shaft speed in Figure 6.12.

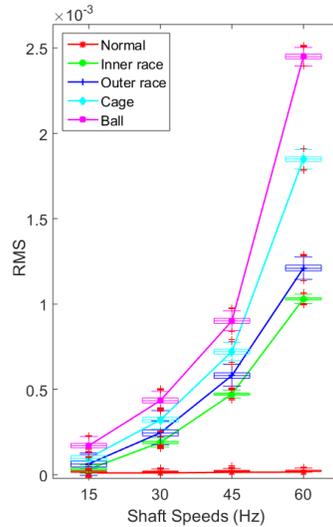


Figure 6.12. RMS by shaft speed using hybrid signal processing method

The RMS obtained after the hybrid deep signal processing method can distinguish the normal, inner race fault, outer race fault, cage fault and ball fault clearly. Furthermore, with increasing shaft speed, RMS of all bearing health conditions show the rising monotonic trend. Another frequently used condition indication (CI) peak value of the was also calculated on the signals being processed by the hybrid deep signal processing method. The peak CI of each bearing condition under different shaft speeds are plotted in Figure 6.13. The cluster of each bearing health condition can be observed clearly in Figure 6.13, with the rising monotonic trend along the increasing shaft speed. In comparison with the peak values extracted from reconstructed vibration signals using TSR method (Van Hecke *et al.*, 2014), the peak value extracted by using hybrid deep

signal processing method distribute more clearly, showing better performance on identifying bearing conditions.

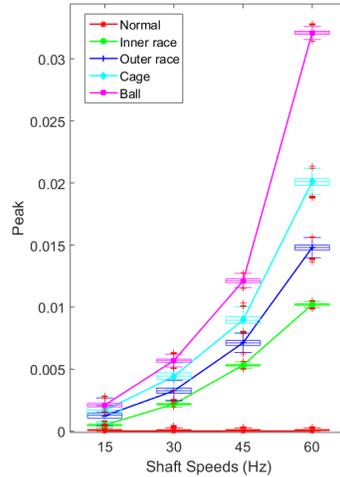


Figure 6.13. Peak by shaft speed using hybrid signal processing method

Other popularly used deep learning methods on machinery fault diagnosis have been reported with good performance, including deep neural network (DNN) (Jia *et al.*, 2016), convolutional neural network (CNN) were also adapted in this paper for comparison. To reduce the computational complexity, the frequency spectra of original raw vibration data was used as input fed into DNN and CNN. The comparison results of testing accuracies from the proposed hybrid deep signal processing method, DNN, and CNN are shown in Table 6.16.

Table 6.16 Comparison Results with Other Deep Learning Methods

Shaft speeds	Testing accuracy (%)		
	Hybrid deep signal processing	DNN	CNN
15	99.82	90.21	82.12
30	99.92	93.28	78.91
45	100	93.19	83.18
60	100	95.34	84.23
Mixed	99.92	87.34	80.45

Comparison results presented in Table 6.16 show that the hybrid deep signal processing can diagnose bearing faults more accurately than both DNN and CNN, even with raw vibration signals. The outstanding performance is contributed to the physical embedded artificial intelligence that can capture the time-frequency characteristics of signals.

6.5 The Validation Results of Deep Hybrid Signal Processing Method on Gear Prognostics Using Vibration Signals

6.5.1 Description of the Validation Data Sets

In preparation of the NASA Glenn Spiral Bevel Gear Fatigue Rig for upcoming tests of 46 newly designed spiral bevel gear sets, several existing gears (36 tooth gear/12 tooth pinion) were ran on the test rig. During these 8 checkout tests, vibration, oil debris, torque and speed data were collected once every minute with the NASA Glenn Labview based data acquisition system, MDSS. Figure 6.14 illustrates the setup of the sensors in the test rig. The left gear set (pinion/gear) was referenced as left and the right gear set (pinion/gear) was referenced as right in the MDSS system. For the MDSS tests, the optical once per rev sensor is located on the left gear shaft and the TSA (Time Synchronous Average) is only calculated for the 36 tooth gear. A second once per rev has been installed on the pinion shaft and will be used for pinion TSA calculations for future tests. The MDSS accelerometers are located on the right and left gearbox housing near the pinion. The

accelerometer is mounted axially in a vertical position. Photos of the accelerometer locations are also shown in Figure 6.14.

Eight run-to-failure experiments were conducted. No damage was observed on the gear teeth during tests 1, 2, and 5. During tests 3, 4, 6, 7 and 8, gears were inspected periodically and

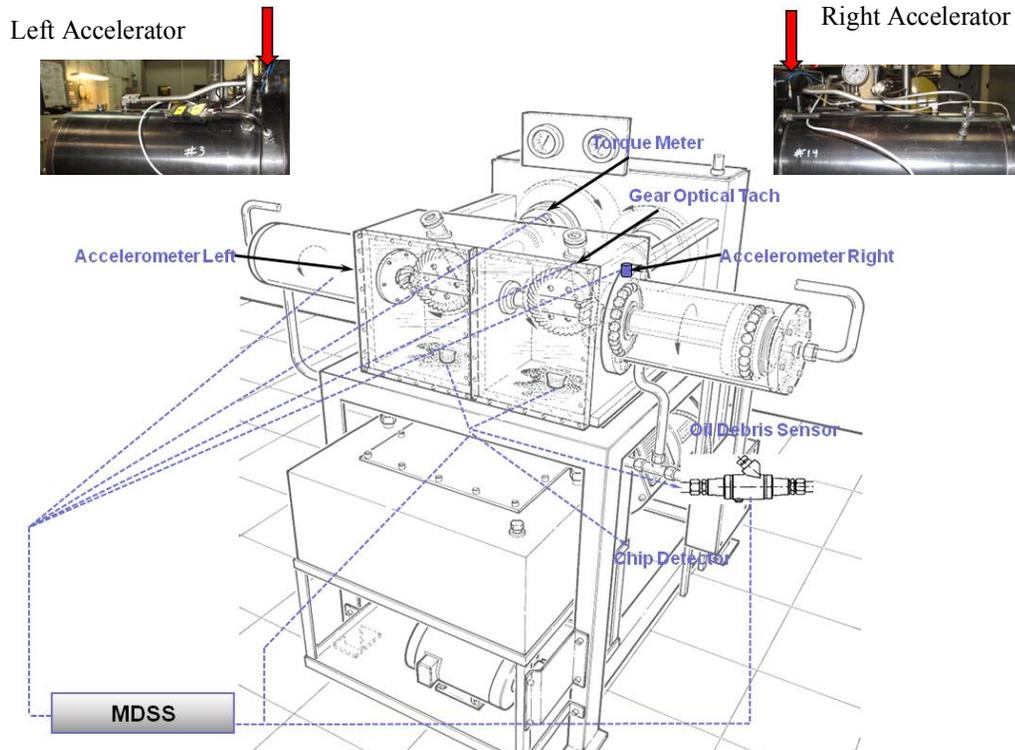


Figure 6.14 Spiral Bevel Gear Fatigue Rig

photos were taken of the damage to the gear teeth during testing. Table 6.17 is a list of detailed information of experiments to be discussed.

Table 6.17 Detailed information of carried out experiments

Experiment`	DAQMAX	Gear Set	Gear Set	RDGS	Occurred
Filename	Task	Left	Right	Hours	Failure
NGB_CHK1	NGB_Bevel1	Set 49	Set 47	3517	
	Archive 100%	s/n pin: 37	s/n pin: 30	59	No damage
NGB_CHK2	NGB_Bevel1	Set 69	Set 47	3069	No damage

	Archive 100%		s/n pin: 30	51	
NGB_CHK3	NGB_Bevel1	Set 69	Set 1	979	Right pinion and gear teeth damage
	Archive 100%		s/n pin: 33	16	
NGB_CHK4	NGB_Bevel1	Set 69	Set 35	2578	Right pinion and gear teeth damage
	Archive 30%			43	
NGB_CHK5	NGB_Bevel1	Set 43	Set 47	5366	No damage
	Archive 30%			89	
NGB_CHK6	NGB_Bevel1	Set 43	Set 47	3302	Left pinion scoring on three teeth
	Archive 50%			55	
NGB_CHK7	NGB_Bevel2	Set 24	Set 47	1057	Left pinion pitting on three teeth. Right pinion pitting on four teeth
	Archive 100%			18	
NGB_CHK8	NGB_Bevel2	Set 03	Set 14	1057	Left pinion pitting on two teeth
	Archive 100%			18	

6.5.2 The Validation Results

To demonstrate the effectiveness of the proposed method to extract gear fault features for gear prognostics and compare its performance with the traditional methods, vibration data sets collected from experiments 3, 4, 6 and 7 were used. For the collected data at every minute, a sample with length of 20000 data points was generated. In the data set obtained from each experiment, 100 samples were randomly selected to train the DFT-IDFT autoencoder with 1000 iterations set as constant. The optimal filter size was decided by the shrinkable CNN as 5000. The number of samples used for deciding the optimal filter size was selected as 200 samples. The trained DFT kernel was then used to process each sample as convolutional kernel to obtain the expression of the original input in frequency domain. Then the average pooling layer sharing the same dimension of filter size was employed to obtain the averaged frequency domain expression. Next, the averaged frequency spectra were transformed into time domain through the trained IDFT kernel. Finally, the RMS was calculated from the reconstructed time domain expression as feature

indicating gear status. In comparison with the proposed method, commonly used features extracted by TSA method were also calculated and plotted in the following validation results. The comparison results obtained from data sets of experiments 3,4,6 and 7 are shown in Figures 6.16-19.

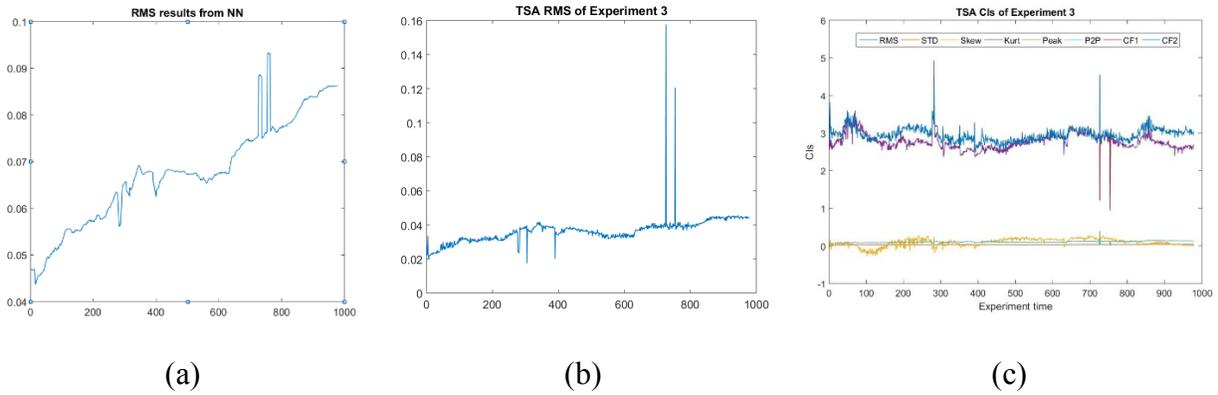


Figure 6.15 The comparison results from experiment 3. (a) RMS extracted from output of the neural network; (b)RMS extracted from TSA calculation; (c) Other CIs extracted from TSA calculation

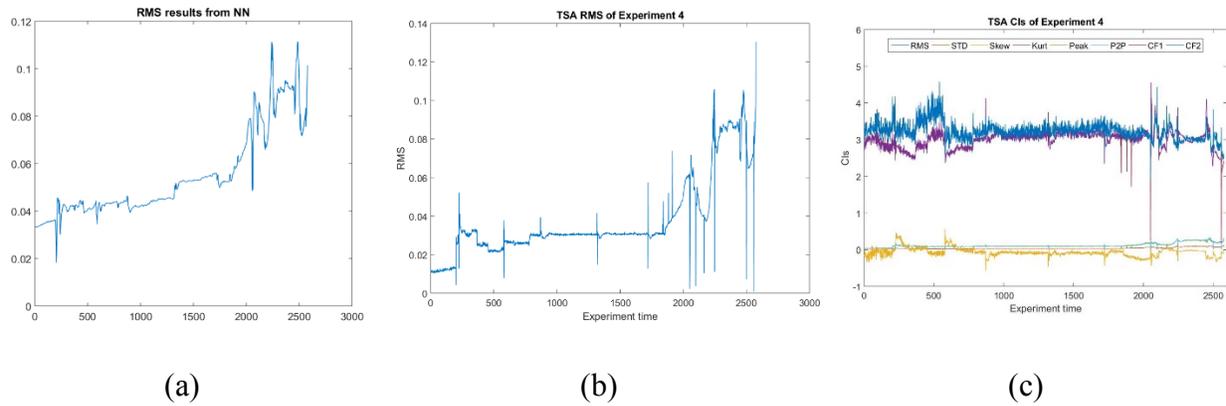


Figure 6.16 The comparison results from experiment 4. (a) RMS extracted from output of the neural network; (b)RMS extracted from TSA calculation; (c) Other CIs extracted from TSA calculation

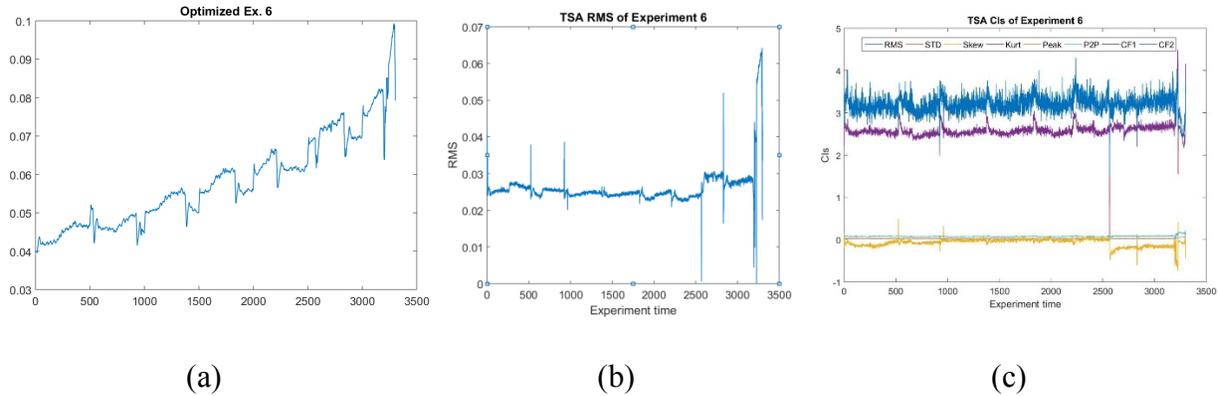


Figure 6.17 The comparison results from experiment 6. (a) RMS extracted from output of the neural network; (b)RMS extracted from TSA calculation; (c) Other CIs extracted from TSA calculation

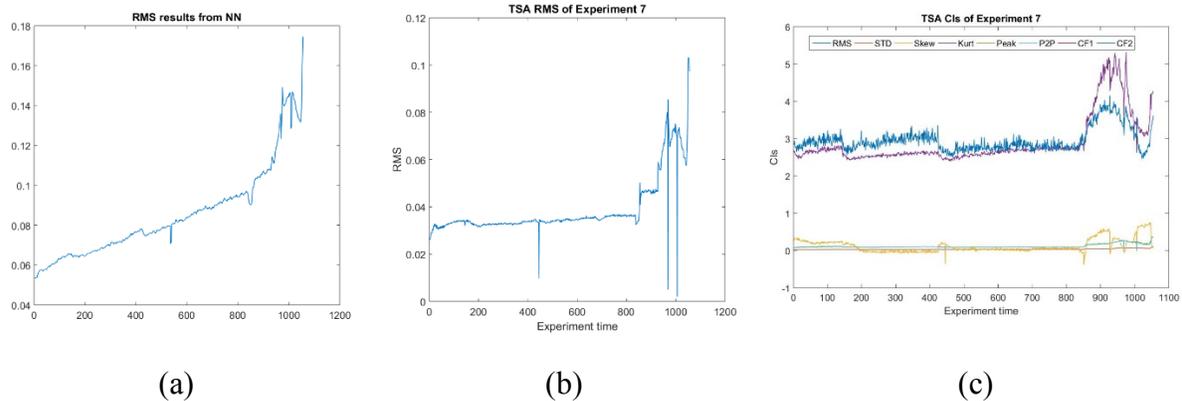


Figure 6.18 The comparison results from experiment 7. (a) RMS extracted from output of the neural network; (b)RMS extracted from TSA calculation; (c) Other CIs extracted from TSA calculation

It can be observed from Figure 5-8 that the RMS calculated from the output of the neural network shows obvious degradation trend in all the experiments with damage shown in the end. The features extracted from TSA results fail to reveal the degradation trends with damage developing on gears. The RMS of TSA results calculated from experiment 7 in Figure 6.18 shows the similar trend as in the RMS of outputs from neural network. It can be caused by that one tooth on the left pinion had pitting damage prior to the installation of the gears onto the rig.

7 CONCLUSIONS

In this dissertation, effective and efficient deep learning based rotating machinery fault diagnostics methods and tools are developed. The proposed methods and tools are validated on both a set of seeded localized faults on rotating bearings and a set of seeded localized faults on all gears in one PGB, respectively. Specifically, the all bearing faults locations include rolling ball, inner race, outer race and cage. The localized PGB faults are seeded on all gears including sun gear, planetary gear and ring gear. The integration of deep learning methods enables the proposed tools to extract fault features from collected raw data or simply processed data through unsupervised learning. Thus, the critical challenges of feature extraction through appropriate signal processing algorithms and effective feature selection for satisfactory fault diagnosis performance. Furthermore, to overcome the known issue of the AM effect caused by rotating vibration transfer path in vibration signals, AE signals were used to validate the proposed methods and to enhance the bearing fault diagnosis performance. This was the first reported application of AE data in combination with deep learning algorithm that was validated to diagnose all bearing fault types at both high and low shaft speed ranges.

This research has several significant contributions. First, a deep learning based bearing fault diagnosis method using AE sensors with big data has been proposed. This allows the network to extract effective fault features from time-frequency spectra of AE signals after short term Fourier transfer without human interaction involved. Such automatic feature extraction enables the intelligent bearing fault diagnosis with big data sets. Additionally, even though the methodology was originally proposed for AE signals, it can be easily adopted to bearing fault diagnosis using vibration signals. Secondly, the fault features carrying both fault location and fault severity level information can be extracted from vibration signals through the proposed deep learning based

method. With needs of fault severity diagnosis in modern industrial application, the LAMSTAR network can be adopted in fault severity diagnosis. The LAMSTAR network outperforms the other typical deep learning network on extracting fault location and severity level features simultaneously. Thirdly, a hybrid deep signal processing method is proposed. Integrating physical fundamental of signal processing technique into deep learning related network transforms the typical deep learning network into a more applicable tool for fault diagnosis. Thus, the mechanism of deep learning related method digging time and frequency characteristics can be revealed. Note that the methods proposed in this research do not rely on the trial-and-error test to select hyperparameters of network. Moreover, the hybrid signal processing technique has been extended on effective feature extraction for gear fault prognostics.

Through the seeded fault tests and runt-to-failure tests, the research contributions have been validated. The bearing fault diagnosis performance of the LAMSTAR network based approach was also compared with another deep learning method CNN. The results have shown that the LAMSTAR network based method gives better performance at both the normal and relative low input shaft speeds. In addition, the AE diagnostic features obtained using the LAMSTAR network were compared with those obtained by the specialized AE signal processing and feature extraction techniques reported in the literature. The results have shown that the AE diagnostic features obtained using the LAMSTAR network are as effective as those obtained by the specialized AE signal processing techniques reported in the literature. Additionally, integrating sparse coding into LAMSTAR network is extended easily on vibration signals for PGB fault diagnosis. The effectiveness of the presented methods was validated with vibration signals collected from seeded gear fault tests performed on a PGB test rig. The self-learnt features were extracted automatically from the raw vibration data by the presented approach. In comparison with the classical CIs for

PGB fault diagnosis (Yoon *et al.*, 2016), the self-learnt features by the presented method have shown a better separation performance on PGB feature extraction and fault diagnosis. The presented method resulted in a better feature extraction and fault diagnosis performance on the same dataset than those reported. Fault diagnosis accuracy reaches as high as 99.61% by the presented LAMSTAR-DL based method.

The hybrid deep signal processing method was validated with vibration signals collected from seeded bearing fault tests conducted in a bearing test rig. The validation results have shown that the proposed method can diagnose the bearing fault using vibration signals through unsupervised learning with high accuracy. Also, the presented method can eliminate the impact of signal segmentation and operation conditions. Furthermore, the RMS and peak values extracted from the reconstructed vibration signals show the monotonic trend along with the increasing shaft speed. The behavior of such statistical features agrees with the results reported by Van Hecke *et al.* [17]. Moreover, the peak value extracted by the presented method show better performance on distinguishing bearing health conditions. The introduction of TSR enables the deep learning structure to adaptively learn the time-frequency features from raw monitoring signals. The physical interpretation can be discovered to the weight vectors of both real and imaginary weights of the DFT-IDFT input layer. The weight vectors show the observable time-frequency properties as the bandpass bases for signal decomposing. Since the presented method uses a deep learning structure to achieve the function of TSR signal processing technique, the structure of deep learning network is designed with the principle of TSR signal processing. Thus, the optimal structure can be determined according to the characteristics of signals rather than trail-and-error test as typically adopted in the deep learning based fault diagnosis methods. The presented method fills in the gap between the traditional signal processing along with manual feature extraction and adaptive

features learning using deep learning techniques. The effectiveness of the proposed method was demonstrated with four gear run-to-failure teste datasets with different damages occurred. The RMS extracted from the output of neural network show the clear degradation trend of damages on gears. Moreover, multiple commonly used CIs were also calculated from the TSA results of the raw vibration signal for comparison. It can be concluded that for this case, the selected CIs calculated from TSA results fail to reveal the gear degradation trend.

In summary, this research shows that deep learning related algorithms can be successfully applied in condition based maintenance, including fault diagnosis and prognostics aspects. The designed LAMSTAR based network can extract fault features automatically for fault diagnosis, even simultaneous fault diagnosis and fault severity identification. Also, the novel deep learning structure integrated signal processing introduces the physical fundamental of application deep learning algorithm into rotating machinery fault diagnosis and prognostics.

8 REFERENCES

- Aharon, M., Elad, M., and Bruckstein, A. M., 2006, “The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation”, *IEEE Transactions on Signal Processing*, Vol. 54, No. 11, pp. 4311-4322.
- Andrade, F. A., East, I., and Badi, M. N. M., 1999, “Gearbox fault detection using statistical methods, time-frequency methods (STFT and Wigner-Ville distribution) and harmonic wavelet- a comparative study”, *Proceedings of COMADEM, Sunderland, United Kingdom*, 6-9 July, pp.77-85.
- Babu, G. S., Zhao, P., and Li, X.L., 2016, “Deep convolutional neural network based regression approach for estimation of remaining useful life”, *International Conference on Database Systems for Advanced Applications*. 1-3 August, pp. 214–228.
- Banaszak, D., 2001, “Comparison of piezoelectric strain sensors with strain gages”, *Proceedings of the Annual Meeting of the American Statistical Association*, Atlanta, GA, 5-9 August.
- Baraldi, P., Compare, M., Saucio, S., and Zio, E., 2013m “Ensemble neural network-based particle filtering for prognostics”, *Mechanical Systems and Signal Processing*, Vol. 41, pp. 288–300.
- Baraldi, P., Mangili, F., and Zio, E., 2012, “A Kalman filter-based ensemble approach with application to turbine creep prognostics”, *IEEE Transactions of Reliability*, Vol. 61, pp. 966–977.
- Barlow, H. B., 1961, “Possible principles underlying the transformations of sensory messages”, *Sensory Communication*, Ed W. Rosenblith Chapter 13, pp. 217-234, M.I.T. Press, Cambridge, Mass.
- Bechhoefer, E., Clark, S., and He, D., 2010, “A state space model for vibration based prognostics”, in *Proceedings of Annual Conference. Prognostics Health Management Society*, 13-16 October, Portland, OR, pp. 1–8.
- Bengio, Y., 2009, “Learning deep architectures for AI”, *Foundation and Trends in Machine Learning*, Vol. 2, No. 1, pp. 1-127.
- Bengtsson, I. and Zyczkowski, K., 2006, “Geometry of quantum states: an introduction to quantum entanglement”, *Cambridge University Press*, Cambridge, May.
- Bin, G. F., Gao, J. J., Li, X. J., and Dhillon, B. S., 2012, “Early fault diagnosis of rotating machinery based on wavelet packets-Empirical mode decomposition feature extraction and neural network”, *Mechanical Systems and Signal Processing*, Vol. 27, pp. 696-711.
- Boudiaf, A., Moussaoui, A., and Dahane, A. 2016, “A comparative study of various methods of bearing faults diagnosis using the Case Western Reserve University data”, *Journal of Failure Analysis and Prevention*, Vol. 16, pp. 271-284.

- Chacon, J. L. F., Kappatos, V., Balachandran, W., and Gan, T. H., 2015, “A novel approach for incipient defect detection in rolling bearings using acoustic emission technique”, *Applied Acoustics*, Vol.89, pp. 88-100.
- Chen, C., Zhang, B., Vachtsevanos, G., and Orchard, M., 2011, “Machine condition prediction based on adaptive neuro-fuzzy and high-order particle filtering”, *IEEE Transactions on Industrial Electronics*, Vol. 58, pp. 4353–4364.
- Chen, R., Chen, S., He, M., He, D., and Tang, B., 2017, “Rolling bearing fault severity identification using deep sparse auto-encoder network with noise added sample expansion”. *Proceedings of the institution of Mechanical Engineers Part O: Journal of Risk and Reliability*, Vol. 231, pp.666-679.
- Chen, X., Du, Z., Li, J., Li, X., and Zhang, H., 2014. “Compressed sensing based on dictionary learning for extracting impulse components”, *Signal Processing*, Vol. 96, pp. 94-109.
- Codetta-Raiteri, D. and Portinale, L., 2015, “Dynamic Bayesian networks for fault detection, identification, and recovery in autonomous spacecraft”, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 45, pp. 13–24.
- Collobert, R. and Weston, J., 2008, “A unified architecture for natural language processing: Deep neural networks with multitask learning”, in *Proceedings of the 25th International Conference on Machine Learning ACM*, 5-9 July, New York, NY, pp. 160–167.
- Corne, S., Murray, T., Openshaw, S., See, L., and Turton, I., 1999, “Using computational intelligence technique to model subglacial water systems”, *Journal of Geographical Systems*, Vol. 1, No. 1, pp. 37-60.
- Daigle, M. J. and Goebel, K., 2013, “Model-based prognostics with concurrent damage progression processes”, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 43, pp. 535–546.
- Dalpiatz, G., Rivola, A., and Rubini, R., 2000, “Effectiveness and sensitivity of vibration processing techniques for local fault detection in gears”, *Signal Processing*, Vol. 14, pp. 387–412.
- Darogheh, N., Baniamerian, A., Meskin, N., and Khorasani, K., 2017, “Prognosis and health monitoring of nonlinear systems using a hybrid scheme through integration of PFs and neural networks”, *IEEE Transactions on Systems, Man and Cybernetics*, 2017, Vol. 47, pp. 1990-2004.
- Davis, G., Mallat, S., and Avellaneda, M., 1997. “Adaptive greedy approximations”, *Constructive Approximation*, Vol. 13, No. 1, pp. 57-98.
- Deng, L., 2014, “A tutorial survey of architectures, algorithms, and applications for deep learning”, *APSIPA Transaction of Signal Information Processing*, Vol. 3, No. 2, pp.1-29.
- Dong, F., Shatz, S. M., Xu, H., and Majumbar, D., 2012, “Price comparison: A reliable approach to identifying shill bidding in online auctions?”, *Electronic Commerce Research and Applications*, Vol. 11, No. 2 pp. 171-179.

- Donoho, D. L., Tsaig, Y., Drori, I., and Starck, J. L., 2012, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit", *IEEE Transactions on Information Theory*, Vol. 58, No. 2, pp. 1094-1121.
- Elad, M. and Aharon, M., 2006, "Image denoising via sparse and redundant representations over learned dictionaries", *IEEE Transactions on Image Process*, Vol. 15, No. 12, pp. 3736-3745.
- Engan, K., Aase, S. O., and Husoy, J. H., 1999, "Method of optimal directions for frame design", in *Proceedings IEEE International Conference of Acoustic, Speed, Signal Processing (ICASSP)*, Vol. 5, pp. 2443-2446.
- Fan, X.F. and Zuo, M. J., 2008, "Machine fault feature extraction based on intrinsic mode functions", *Measurement Science Technology*, Vol. 19, pp. 1-12.
- Feng, Z. and Zuo, M. J., 2013, "Fault diagnosis of planetary gearboes via torsinal vibration signal analysis", *Mechanical Systems and Signal Processing*, Vol. 36, No. 22, pp. 401-421.
- Foldiak, P., 1990, "Forming sparse representations by local anti-hebbian learning", *Biological Cybernetics*, Vol. 64, No. 2, pp. 165-170.
- Gan, M., Wang, C., and Zhu, C., 2016, "Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings", *Mechanical Systems and Signal Processing*, Vol. 72-73, pp. 92-104.
- German, E., Basaran, M., and Fidan, M., 2014, "Sound based induction motor fault diagnosis using Kohonen self-organizing map", *Mechanical Systems and Signal Processing*, Vol. 46, pp. 45-58.
- Girado, I. J., Sandin, J. D., DeFanti, A. F., and Wolf, K. L., 2013, "Real-time camera-based face detection using a modified LAMSTAR neural network system," Processing of SPIE 5015. *Application of Artificial Neural Networks in Image Processing*, Santa Clara, CA, Mar 27, Vol. 8, No. 36, pp. 36-46.
- Gorodnitsky, I. F. and Rao, B. D., 1997, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm", *IEEE Transactions of Signal Processing*, Vol. 45, No. 3, pp. 600-616.
- Graupe, D., 2013, "Large Scale Memory Storage and Retrieval (LAMSTAR) network," in *Principles of Artificial Neural Networks*, 3rd edition, Chap. 9, pp.204-224, World Scientific, Hackensack, NJ.
- Han, S., Meng, Z., O'Reilly, J., Cai, J., Wang, X., and Tong, Y., 2016, "Optimizing filter size in convolutional neural network for facial action unit recognition", *Arxiv:1707.08603*.
- He, D., Bechhoefer, E., Ma, J., and Li, R., 2011, "Particle filtering based gear prognostics using one-dimensional health index", in *Proceedings of Annual Conference. Prognostics Health Management Society*, 25-29 September, Montreal, QC, Canada, pp. 1-9.
- He, M. and He, D., 2017, "A deep learning based approach for bearing fault diagnosis", *IEEE Transactions on Industrial Applications*, Vol. 53, No. 3, pp. 1-9.

- He, W., Zi, Y., Chen, B., Wu, F., and He, Z., 2015, “Automatic fault feature extraction of mechanical anomaly on induction motor bearing using ensemble super-wavelet transform”, *Mechanical Systems and Signal Processing*, Vol. 54-55, pp. 457-480.
- Henríquez, P., Alonso, J. B., Ferrer, M. A., and Travieso, C. M., 2014, “Review of automatic fault diagnosis systems using audio and vibration signals”, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 44, No. 5, pp. 642 – 652.
- Heydarzadeh, M., Kia, S. H., Nourani, M., Henao, H., and Capolino, G., 2016, “Gear fault diagnosis using discrete wavelet transform and deep neural networks”, *Industrial Electronics Society, IECON 2016*, 23-26 October, Florence, Italy.
- Hinton, G. E. and Salakhutdinov, R. R., 2006, “Reducing the dimensionality of data with neural networks”, *Science*, Vol. 313, No. 5786, pp. 504-507.
- Hinton, G., Deng, L., Yu, D., Dahl, G. E. Mohamed, A. R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., and Sainath, T.N., 2012, “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups”, *IEEE Signal Processing Magazine*, Vol. 29, pp. 82–97.
- Hu, W., Ding, X., Li, B., Wang, J., Gao, Y., Wang, F., and Maybank, S., 2016, “Multi-perspective cost-sensitive context-aware multi-instance sparse coding and its application to sensitive video recognition”, *IEEE Transactions on Multimedia*, Vol. 18, No. 1, 99.76-89.
- Isola, R., Crvallo, R. and Tripathy, A. K., “Knowledge discovery in medical systems using differential diagnosis, LAMSTAR, and k-NN,” *IEEE Transactions of Information in Technology Biomedicine*, Vol. 16, No. 6, pp. 1287-1294.
- Janssens, O., Slavkovikj, V., Vervisch, B., Stockham, K., Loccufier, M., Verstockt, S., Van de Walle, R., and Van Hoecke, S., 2016, “Convolutional neural network based fault detection for rotating machinery”, *Journal of Sound and Vibration*, Vol. 377, pp. 331-345.
- Jardine, A. K., Lin, D., and Banjevic, D., 2006, “A review on machinery diagnostics and prognostics implementing condition-based maintenance”, *Mechanical Systems and Signal Processing*, Vol. 20, No. 7, pp. 1483–1510.
- Javed, K. 2014, “A robust and reliable data-driven prognostics approach based on extreme learning machine and fuzzy clustering”, Doctoral thesis, University of Franche-Comte, April 9.
- Jeschke, S., Brecher, C., Song, H., and Rawat, D. B., 2017, *Industrial Internet of Things*. Springer Series in Wireless Technology, Switzerland, 2017.
- 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*, Vol. 72-73, pp. 303-315\.
- Jiang, X., Kim, K., Zhang, S., Johnson, J., and Salazar, G., 2014, “High-temperature piezoelectric sensing”, *Sensors*, Vol. 14, No. 1, pp. 144-169.
- Jolliffe, I. T., 2002, *Principle Component Analysis*, 2nd edition, New York: Springer.

- Kankar, P. K., Stish C. S., and Harsha, S. P., 2011, "Fault diagnosis of ball bearings using machine learning methods", *Expert Systems with Applications*, Vol. 38, pp. 1876-1886.
- Katunin, A., Amarowicz, M., and Chrzanowski, P., 2015, "Faults diagnosis using self-organizing maps: A case study on the DAMADICS Benchmark problem", *Proceedings of the Federal Conference on Computer Science and Information Systems*, Lodz, Poland, Vol. 5, pp. 1673-1681.
- Kohonen, T., 1982, "Analysis of a simple self-organizing process", *Biology Cybernetics*, Vol. 44, No. 2, pp. 135-140.
- Kohonen, T., 1982, "Self-organized formation of topologically correct feature maps", *Biology Cybernetics*, Vol. 43, No. 1, pp. 59-69`.
- Kohonen, T., 2001, "The basic SOM", *Self-Organizing Maps*, Springer-Verlag, Berlin Heidelberg.
- Labate, D., Lim, W., Kutyniok, G., and Weiss, G., 2005. "Sparse multidimensional representation using shearlets", in *Processing SPIE: Wavelets XI*, Vol. 5914, pp. 254-262.
- LeCun, Y., Bengio, Y., and Hinton, G., 2015, "Deep learning", *Nature*, Vol. 521, pp. 436-444.
- Lee, C. K. and O'Sullivan, T., 1991, "Piezoelectric strain rate gages", *Journal of the Acoustical Society of America*, Vol. 90, No. 2, pp. 945-953.
- Lee, H., Battle, A., Raina, R., and Ng, A. Y., 2007. "Efficient sparse coding algorithms", *Advances in Neural Information Processing Systems*, Vol. 19, pp. 801-808.
- Lee, H., Ekanadham, C., and Ng, A. Y., 2008. "Sparse deep belief net model for visual area V2", *Advances in Neural Information Processing Systems*, pp. 873-880.
- Lei, Y. G., He, Z. J., Li, Y. Y., and Hu, Q., 2007, "Fault diagnosis of rotating machinery based on multiple ANFIS combination with Gas", *Mechanical Systems and Signal Processing*, Vol. 21, No.5, pp. 2280-2294.
- Lei, Y. G., He, Z. J., Zi, Y. Y., and Chen, X., 2008, "New clustering algorithm-based fault diagnosis using compensation distance evaluation technique", *Mechanical System and Signal Processing*, Vol. 22, pp. 419-435.
- Lei, Y. G., Lin, J., He, Z. J., and Kong, D., 2012, "A method based on multi-sensor data fusion for fault detection of planetary gearboxes", *Sensors*, Vol. 12, No. 2, pp.2005-2017.
- Lei, Y., Jia, F., Lin, J., Xing, S., and Ding, S., 2016, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data", *IEEE Transactions on Industrial Electronics*, Vol. 63, pp. 3137-3147.
- Lei, Y.G., Lin, J., He, Z. J., and Zi, Y. Y., 2011, "Application of an improved kurtogram method for fault diagnosis of rolling element bearings", *Mechanical System and Signal Processing*, Vol. 25, pp. 1738-1749.
- Leung, M. K., Xiong, H. Y., Lee, L. J., and Frey, B. J., 2012, "Deep learning of the tissue-regulated splicing code", *IEEE Signal Processing Magazine*, Vol. 29, pp. 82-97.

- Lewiski, M.S. and Sejnowski, T. J., 2000, “Learning overcomplete representations”, *Neural Computation*, Vol. 12, No. 2, pp. 337-365.
- Li, C., Sanchez, R., Zurita, G., Cerrada, M., Cabrera, D., and Vasquez, R. E., 2015, “Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis”, *Neurocomputing*, Vol. 168, pp. 119-127.
- Li, C., Sanchez, R., Zurita, G., Cerrada, M., Cabrera, D., and Vasquez, R. E., 2016, “Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signal”, *Mechanical Systems and Signal Processing*, Vol. 77-76, pp. 283-293.
- Li, H., Zhang, Y. and Zheng, H., 2010, “Hilbert-Huang transform and marginal spectrum for detection and diagnosis of localized defects in roller bearings”, *Journal of Mechanical Science and Technology*, Vol. 29, No. 2, pp. 291-301.
- Li, R., Ma, J., Panyala, A., and He, D., 2010, “Hybrid ceramic bearing prognostics using particle filtering”. in *Technical Program for the MFPT 2010 Conference - Transition: From R and D to Produc.*, 13-15 April, Huntsville, AL, USA, pp. 57–69.
- Li, Y., Bu, S., Liu, Z., and Zhang, C., 2015. “Mechanical fault diagnosis of rolling bearing based on locality-constrained sparse coding”, *Prognostic and Health Monitoring (PHM)*, IEEE Conference on, June 22-25, Austin, TX, pp. 1-7.
- Li. C., Wang, Z. and Ding, C., 2014, “Fault diagnosis of rolling bearing based on fisher discrimination sparse coding”, *Proceedings of the First Symposium on Aviation Maintenance and Management*, Vol. 11, pp. 387-394.
- Liao, H. Y. and Sapiro, G., 2008, “Sparse representations for limited data tomography”, in *Proceedings of the 5th IEEE International Symposium Biomedical Image: Nano Macro (ISBI)*, pp. 1375-1378, Paris, France.
- Liao, L., Jin, W., and Pavel, R., 2016, “Enhanced restricted Boltzmann machine with prognosability regularization for prognostics and health assessment”, *IEEE Transactions on Industrial Electronics*, Vol. 63, pp.7076-7083.
- Lim, P., Goh, C. K., Tan, K. C., and Dutta, P., 2014, “Estimation of remaining useful life based on switching Kalman filter neural network ensemble”, in *Proceedings of Annual Conference. Prognostics Health Management Society*, 29 September – 2 October, Fort Worth, TX, pp. 2–9.
- Liu, H., Liu, C., and Huang, Y., 2011, “Adaptive feature extraction using sparse coding for machinery fault diagnosis”, *Mechanical Systems and Signal Processing*, Vol. 25, No. 1, pp. 559-574.
- Long, F. and Bartlett, M. S., 2016, “Video-based facial expression recognition using learned spatiotemporal pyramid sparse coding features”, *Neurocomputing*, Vol. 173, No. 3, pp. 2049-2054.
- Lou, X. S. and Loparo, K. A., 2004, “Bearing fault diagnosis based on wavelet transform and fuzzy inference”, *Mechanical Systems and Signal Processing*, Vol. 8, No. 5, pp. 1077-1095.

- Lu, C., Wang, Z. Y., Qin, W. L., and Ma, J., 2107, “Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification”, *Signal Processing*, 2017, Vol. 130, pp. 377-388.
- Lu, D., He, Q., Yuan, T., and Kong, F., 2016, “Online fault diagnosis of motor bearing via stochastic-resonance-based adaptive filter in an embedded system”, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 47, No. 7, pp. 1111-1122.
- Lund, D. MacGillivray, C., Turner, V., and Morales, M., 2014, “Worldwide and regional internet of things (iot) 2014–2020 forecast: A virtuous circle of proven value and demand”, International Data Corporation (IDC) Technology Report, 2014.
- Luo, G. Y., Osypiw, D., and Irle, M., 2003, “On-line vibration analysis with fast continuous wavelet algorithm for condition monitoring of bearing”, *Journal of Vibration Control*, Vol. 9, pp. 931-947.
- Luo, H., Qiu, H., Ghanime, G., Hirz, M., and Van der Merwe, G., 2010, “Synthesize synchronous sampling technique for differential bearing damage detection”, *Journal of Engineering for Gas Turbine and Power*, Vol. 132, pp. 1-8.
- Lv, F., Wen, C., Bao, Z., and Liu, M., 2016, “Fault diagnosis based on deep learning”, *American Control Conference*, Boston, MA, Jul. 6-8.
- Ma, M., Chen, X., Wang, S., Liu, Y., and Li, W., 2016, “Bearing degradation assessment based on weibull distribution and deep belief network”, in *Proceedings of 2016 International Symposium of Flexible Automation (ISFA)*, 1-3 August 2016, Cleveland, OH, pp. 1–4.
- Ma, Y., Guo, Z., Su, J., Chen, U., Du, X., Yang, Y., Li, C., Lin, Y., and Geng, Y., 2014, “Deep learning for fault diagnosis based on multi-sourced heterogeneous data”, *International Conference of Power Systems Technology*, October 20-22, Chengdu, China.
- Malhotra, P., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P., and Shroff, G., 2016, “Multi-sensor prognostics using an unsupervised health index based on LSTM encoder-decoder”, *1st ACM SIGKDD Workshop on Machine Learning for Prognostics and Health Management*, San Francisco, CA.
- Mallat, S., 2009, *A Wavelet Tour of Signal Processing*, 3rd ed., Academic Press., New York, New York.
- Martin-del-Campo, S. and Sandin, F., 2015, “Towards zero-configuration condition monitoring based on dictionary learning”, *Signal Processing Conference (EUSIPCO)*, 2015 23RD European, August 31 – September 4, Nice, France.
- Mathew, J. and Alfredson, R. J., 1983, “The condition monitoring of rolling element bearings using vibration analysis”, *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, Vol. 106, No. 3, pp. 447-453.
- Meng, Q. and Qu, L., 1991, “Rotating machinery fault diagnosis using Wigner distribution”, *Mechanical Systems and Signal Processing*, Vol. 5, pp. 77-85.

- Nigam, V. P. and Graupe, D., 2004, "A neural-network-based detection of epilepsy," *Neurological Research*, Vol. 26, No. 1, pp. 55-60.
- Olshausen, B. A. and Field, D. J., 1996. "Emergence of simple-cell receptive field properties by learning a sparse code for natural images", *Nature*, Vol. 381, No., 6583, pp. 607-609.
- Pandya, D. H., Upadhyay, S. H., and Harsha, S. P., 2013, "Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN", *Expert Systems with Application*, Vol. 40, pp. 4137-4145.
- Potter, M. and Elad, M., 2009., "Image sequence denoising via sparse and redundant representations", *IEEE Transactions on Image Process*, Vol. 18, No. 1, pp. 27-35.
- Qu, Y., He, D., Yoon, J., Van Hecke, B., Bechhoefer, E., and Zhu, J., 2014, "Gearbox tooth cut fault diagnostics using acoustic emission and vibration sensors-A comparative study", *Sensors*, Vol. 14, pp. 1372-1393.
- Raina, R., Battle, A., Lee, H., Packer, B., and Ng, A. Y., 2007., "Self-taught learning: transfer learning from unlabeled data", *ACM International Conference Proceeding Series*, Vol. 227, pp. 759-766.
- Ranzato, M., Y-Ian, B., L. Cun, Y., 2008., "Sparse feature learning for deep belief network", *Advances in Neural Information Processing Systems*, pp. 1185-1192.
- Rubinsein, Ron, Zibulevsky, M., and Elad, M., 2010., "Double sparsity: learning sparse dictionaries for sparse signal approximation", *IEEE Transactions on Signal Processing*, Vol. 58, No. 3, pp. 1553-1564.
- Samantha, B., Al-Balushi, K. R., and Al-Araimi, S. A., 2005, "Artificial neural networks and genetic algorithms for bearing fault detection", *Soft Computers*, Vol. 10, No. 3, pp.264-271.
- Samantha, 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*, Vol. 16, No. 7-8, pp. 657-665.
- Samuel, P. D., Conroy, J. K., Pines, and D. J., 2004, "Planetary transmission diagnosis", NASA Technical Report: NASA/CR-2004-213068, NASA Glenn Research Center, Cleveland, OH.
- Sardy, S., Bruce, A. G., and Tseng, P., 2000, "Block coordinate relaxation methods for nonparametric wavelet denoising", *Journal of Computational Graph Statistics*, Vol. 9, No. 2, pp. 361-379.
- Schmidhuber, J., 2015, "Deep learning in neural networks: an overview", *Neural Network*, Vol. 61, pp. 85-117.
- Selesnick, I. W., Baraniuk, R. G., and Kingsbury, N. C., 2005, "The dual-tree complex wavelet transform", *IEEE Signal Processing Magazine*, Vol. 22, No. 6, pp. 123-151.
- Shah, D. S. and Patel, V. N., 2014, "A review of dynamic modeling and fault identifications methods for rolling element bearing", *Procedia Technology*, Vol. 14, pp. 447-456

- Shao, H., Jiang, H., Zhao, and H., Wang, F., 2017, “A novel deep autoencoder feature learning method for rotating machinery fault diagnosis”, *Mechanical Systems and Signal Processing*, Vol. 95, pp. 187-204.
- Sharma, V. and Parey, A., 2015, “A review of gear fault diagnosis using various condition indicators”, *12TH International Conference on Vibration Problems, ICOVP 2015*, December 14-17, Guwahati, India.
- Sheng, S., Link, H., LaCava, W., Van Dam, J., McNiff, B., Veers, P., Keller, J., Butterfield, S., and Oyague, F., 2011, “Wind turbine drivetrain condition monitoring during GRC phase 1 and phase 2 testing”, NREL Technical Report: NREL/TP-5000-52748, National Renewable Energy Laboratory, Golden, CO, USA.
- Sikorska, J., Hodkiewicz, M., and Ma , L., 2011, “Prognostic modelling options for remaining useful life estimation by industry”, *Mechanical Systems and Signal Processing*, Vol. 25, pp.1803–1836.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., and Sifre, L., 2016, “Mastering the game of Go with deep neural networks and tree search”, *Nature*, Vol. 529, pp. 484-502.
- Simani, S. and Fantuzzi, C., 2000, “Fault diagnosis in power plant using neural network”, *Information Sciences*, Vol. 127, pp. 126-136.
- Sivaramakrishna, A. and Graupe, D., 2004, “Brain tumor demarcation by applying a LAMSTAR neural network to spectroscopy data,” *Neurological Research*, Vol. 26, No. 6, pp. 613-621.
- Smith, W. A. and Randall, R. B., “Rolling element bearing diagnostics using the Case Western Reserve University Data: A benchmark study”, *Mechanical Systems and Signal Processing*, Vol. 64-65, pp. 100-131.
- Song, H. A. and Lee, S. Y., 2013, “Hierarchical representation using NMF”, *Neural Information Processing. Lecture Notes in Computer Sciences*, Vol. 8226. Springer Berlin Heidelberg, pp. 466-473.
- Sun, W. J., Shao, S. Y., and Zhao, R., 2016, “A sparse auto-encoder-based deep neural network approach for induction motor faults classification”, *Measurement*, Vol. 89, pp. 171–178.
- Tamilselvan, P. and Wang, P., 2013, “Failure diagnosis using deep belief learning based health state classification”, *Reliability Engineering and System Safety*, Vol. 115, pp.124-135.
- Tandon, N. and Chouldhury, A., 1999, “A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings”, *Tribology International*, Vol. 32, pp. 469-480.
- Tran, V. T., AlThobiani, F., and Ball, A., 2014, “An approach to fault diagnosis of reciprocating compressor valves using Teager-Kaiser energy operator and deep belief networks”, *Expert Systems with Applications*, Vol. 41, No. 9, pp. 4113-4122.

- Tran, V. T., Yang, B. S., Oh, M. S., and Tan, A. C. C., 2009, "Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference", *Expert Applications with Artificial Intelligence*, Vol. 36, No. 2, pp. 1840-1849.
- Tran, V. T., AlThobiani, F., and Ball, A.D., 2014, "An approach to fault diagnosis of reciprocating compressor valves using Teager-Kaiser energy operator and deep belief networks", *Expert Systems with Applications*, Vol. 41, No. 9, pp. 4113-4122.
- Van Hecke, B., He, D., and Qu, Y., 2014, "On the use of spectral averaging of acoustic emission signals for bearing fault diagnostics", *Journal of Vibration and Acoustic*, Vol. 136, No. 6, pp. 1-13.
- Van Hecke, B., Qu, Y., He, D., Bechhoefer, E., 2014, "A new spectral average -based bearing fault diagnosis approach", *Journal of Failure Analysis and Prevention*, Vol. 14, No. 3, pp. 354-362.
- Van Hecke, B., Yoon, J., He, D., 2016, "Low speed bearing fault diagnosis using acoustic emission sensors", *Applied Acoustics*, Vol. 105, pp. 35-44.
- Velik, R., 2008, "Discrete Fourier transform computation using neural networks", *In Computational Intelligence Security, 2008. CIS'08. International Conference*, Vol. 1 2008, pp. 120-123.
- Wang , D. and Shen , C., 2016, "An equivalent cyclic energy indicator for bearing performance degradation assessment", *Journal of Vibration Control*, Vol. 22, pp. 2380-2388.
- Wang, D., Peter, W.T., Guo,W. T., and Qiang, M.,2010, "Support vector data description for fusion of multiple health indicators for enhancing gearbox fault diagnosis and prognosis", *Measurement Science and Technology*, Vol. 22, No. 2, pp.1-13.
- Wang, W. J. and McFadden, P. D., 1993, "Early detection of gear failure by vibration analysis I: Calculation of the time-frequency distribution", *Mechanical Systems and Signal Processing*, Vol. 7, pp. 193-203.
- Waxman, A., Graupe, D., and Carley, D. W., "Automated prediction of apnea and hypopnea, using a LAMSTAR artificial neural network," *American Journal of Respiratory and Critical Care Medicine*, Vol. 181, No. 7, pp. 727-733.
- Waxman, A., Graupe, D., and Carley, D. W., "Automated prediction of apnea and hypopnea, using a LAMSTAR artificial neural network," *American Journal of Respiratory and Critical Care Medicine*, Vol. 181, No. 7, pp. 727-733.
- Widodo, A. and Yang, B.S., 2007, "Support vector machine in machine condition monitoring and fault diagnosis", *Mechanical Systems and Signal Processing*, Vol. 21, No. 6, pp. 2560-2574.
- Worde, K., Staszewski, W.J., and Hensmen, J. J., 2011, "Natural computing for mechanical systems research: a tutorial overview", *Mechanical Systems with Signal Processing*, Vol. 25, No. 1, pp. 4-111.

- Wrights, J., Yang, A. Y., Gnesch, A., Sastry, S. S., and Ma, Y., 2009. "Robust face recognition via sparse representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 2, pp. 210-227.
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., 2016, "Google's neural machine translation system: bridging the gap between human and machine translation", Technical Report, arXiv:1609.08144
- Xavier, G., Bordes, A., and Bengio, Y., 2011. "Deep sparse rectifier neural networks", *14th International Conference on Artificial Intelligence and Statistics*, Vol. 15, No. 106, pp. 315-323. Fort Lauderdale, FL.
- Xiong, L., Shi, T., Yang, S., and Rao, R.B.K.N.. "A novel application of wavelet-based bispectrum analysis to diagnose faults in gears", *International Journal of COMADEM*, Vol. 5, No. 3, pp. 31-38.
- Yin, S., Li, X., Gao, H., and Kaynak, O., 2015, "Data-based techniques focused on modern industry: an overview.", *IEEE Transactions on Industrial Electronics*, Vol. 62, pp. 657-667.
- Yin, X. and Li, Z., 2016, "Reliable decentralized fault prognosis of discrete event systems", *IEEE Transactions of Systems, Man and Cybernetics, Systems*, Vol. 46, pp. 1598-1603.
- Yoon, J., He, D., and Van Hecke, B., 2014, "A PHM approach to additive manufacturing equipment health monitoring, fault diagnosis, and quality control", in *Annual Conference of the PHM*, Fort Worth, TX, Sep.27-Oct. 3, 2014.
- Yoon, J. and He, D., 2015, "Development of an efficient prognostic estimator", *Journal of Failure Analysis and Prevention*, Vol. 15, pp. 129-138.
- Yoon, J., He, D., and Van Hecke, B., 2015, "On the use of a single piezoelectric strain sensor for wind turbine planetary gearbox fault diagnosis", *IEEE Transactions of Industrial Electronics*, Vol. 62, pp. 6585-6593.
- Yoon, Y. M., He, D., and Qiu, B., 2013, "Full ceramic bearing fault diagnosis using LAMSTAR neural network," in *Prognostics and Health Management (PHM), IEEE Conference on*, pp. 1-9. Gaithersburg, MD, Jun 24-27, 2013.
- Zhang, G., Du, R., and Shi, T., 2003, "Extracting gear fault features using integrated bispectrum", IEEE International. Conference on Robotics, *Intelligent Systems and Signal Processing*, Changsha, Hunan, China, 8-3 October,2003, pp.548-533.
- Zhang, L., Xiong, G., Liu, H., Zou, H., and Guo, W., 2010, "Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference", *Expert Systems with Applications*, Vol. 37, No. 8, pp. 6077-6085.
- Zhao, R., Wang, J., Yan, R., and Mao, K., 2016, "Machine health monitoring with LSTM networks", *10th International Conference on Sensing Technology (ICST). IEEE*, November 11-13, Nanjing, P. R. China, pp. 1-6.

Zhong, F., Shi, T., and He, T., 2005, "Fault diagnosis of motor bearing using self-organizing maps", *Electrical Machines and Systems, ICEMS 2005*. Proceedings of the Eighth International Conference on, pp. 2411-2414, Nanjing, China.

Zibulevsky, M. and Pearlmutter, B. A., 2001. "Blind source separation by sparse decomposition in a signal dictionary", *Neural Computation*, Vol. 13, No. 4, pp. 863-882.

APPENDIX



Title: Deep Learning Based Approach for Bearing Fault Diagnosis
Author: Miao He; David He
Publication: Industry Applications, IEEE Transactions on
Publisher: IEEE
Date: May-June 2017

Copyright © 2017, IEEE

LOGIN
If you're a [copyright.com user](#), you can login to RightsLink using your copyright.com credentials. Already a [RightsLink user](#) or want to [learn more?](#)

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

- 1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
- 2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
- 3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

- 1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
- 2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
- 3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.





RightsLink®

Home

Create Account

Help



Title: Wind turbine planetary gearbox feature extraction and fault diagnosis using a deep-learning-based approach

Author: Miao He, David He, Jae Yoon, et al

Publication: Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability

Publisher: SAGE Publications

Date: 04/23/2018

Copyright © 2018, © SAGE Publications

LOGIN

If you're a [copyright.com](#) user, you can login to RightsLink using your [copyright.com](#) credentials. Already a [RightsLink](#) user or want to [learn more?](#)

If you are a SAGE journal author requesting permission to reuse material from your journal article, please note you may be able to reuse your content without requiring permission from SAGE. Please review SAGE's author re-use and archiving policies at <https://us.sagepub.com/en-us/nam/journal-author-archiving-policies-and-re-use> for more information.

If your request does not fall within SAGE's reuse guidelines, please proceed with submitting your request by selecting one of the other reuse categories that describes your use. Please note, a fee may be charged for reuse of content requiring permission. Please contact permissions@sagepub.co.uk if you have questions.

BACK

CLOSE WINDOW

Copyright © 2018 [Copyright Clearance Center, Inc.](#) All Rights Reserved. [Privacy statement](#). [Terms and Conditions](#). Comments? We would like to hear from you. E-mail us at customercare@copyright.com



[Home](#) > [Journals & magazines](#) > [IET Science, Measurement & Technology](#) > [E-First](#) > [Article](#)

Simultaneous bearing fault diagnosis and severity detection using a LAMSTAR network-based approach

Author(s): [Miao He](#)¹ and [David He](#)^{1, 2}

[View affiliations](#)

Source: [IET Science, Measurement & Technology](#), 0 pp.

DOI: [10.1049/iet-smt.2017.0528](#), Online ISSN: 1751-8830

Available online: 06 June 2018

[Access Full Text](#)

[Recommend to library](#)

© The Institution of Engineering and Technology

Received [21/12/2017](#), Accepted [02/05/2018](#), Revised [27/03/2018](#), Published [06/06/2018](#)

Article

Bearings are one of the most important components in many industrial machines. Effective bearing fault diagnosis and severity detection are critical for keeping the machines operate normally and safe. In this study, the problem of simultaneous bearing fault diagnosis and severity detection with deep learning is addressed. Existing solutions developed using deep learning rely on fault feature extraction using complicated signal processing techniques. They perform bearing fault diagnosis and severity detection separately and normally require extensive supervised fine tuning. This study presents an effective deep learning-based solution using a large memory storage and retrieval (LAMSTAR) neural network. The developed approach can automatically extract self-learned fault features and perform bearing fault diagnosis and severity detection simultaneously. The structure of the LAMSTAR network is determined by optimally selecting the sliding box size of the input time–frequency matrix. The effectiveness of the proposed approach is validated using data collected from rolling element bearing tests.

Inspec keywords: [mechanical engineering computing](#); [rolling bearings](#); [signal processing](#); [fault diagnosis](#); [neural nets](#); [feature extraction](#); [condition monitoring](#); [learning \(artificial intelligence\)](#); [machine bearings](#)

9. Green open access

Many funders now mandate that their sponsored research must be deposited in an institutional repository. For a list of funders and their open access requirements, see <http://www.sherpa.ac.uk/julie/>.

In order to allow authors to comply with these mandates, the IET offers a comprehensive green open access policy. With no embargo period, authors are permitted to deposit their accepted manuscript (reviewed revised and typeset, but not the published PDF), on their institutional repository or in repositories designated by their funding body, provided they refer to the published IET version. For full details of the policy and any relevant wordings, see section 9.1 below.

These options will allow our authors to comply with the mandates of a number of funders/research bodies, including, but not limited to EPSRC, RCUK, NIH, Wellcome Trust, Horizon 2020, HEFCE, and NERC.

If you are unsure if the IET will help you to comply with your funder's mandates, please contact us at journals@theiet.org.

9.1 Green open access policy

The following outlines the full IET green open access policy, making it straightforward for authors to comply with their funder mandates.

9.1.1 On submission of their manuscript, authors may deposit the submitted version on their personal, institutional, or online pre-print repository (for example, arXiv).

9.1.2 The first page of the manuscript must clearly display the following wording:

"This paper is a preprint of a paper submitted to [journal] and is subject to Institution of Engineering and Technology Copyright. If accepted, the copy of record will be available at IET Digital Library"

On acceptance, this may be changed to:

"This paper is a preprint of a paper accepted by [journal] and is subject to Institution of Engineering and Technology Copyright. When the final version is published, the copy of record will be available at IET Digital Library"

9.1.3 If the paper is rejected, authors must remove all mention of the journal.

9.1.4 On acceptance of their paper, authors may immediately deposit their accepted manuscript (reviewed revised and typeset, but not the published PDF), and must remove the original submitted pre-print.

9.1.5 The following wording clearly appears on the front page of the paper:

"This paper is a postprint of a paper submitted to and accepted for publication in [journal] and is subject to Institution of Engineering and Technology Copyright. The copy of record is available at IET Digital Library"

9.1.6 ResearchGate: The IET classes ResearchGate as a social media site and as such authors are able to upload their accepted manuscript (reviewed, revised and typeset) but not the published PDF.

Any questions should be addressed to the editorial officer (journals@theiet.org).

[Back to the top](#)

10. Copyright and Open Access License

10.1. Right to publish

An author submitting a paper should ensure that he or she has the right to publish the paper and that it contains nothing defamatory. The IET will assume that all co-authors have agreed to the submission of any paper received.

10.2. Copyright Form for Subscription Access

Authors who choose to publish through the subscription access model will be required to fill in the online copyright form in the ScholarOne Manuscripts Author Centre. The submitting author should sign on behalf of all co-authors.

Unless the paper is either 'United Kingdom Crown copyright' or 'a work of the US Government and in the Public Domain', the Bylaws of the IET require that copyright should belong to the IET.

10.3. Open Access: Creative Commons Licences

Authors who choose to publish their papers open access in the IET Research Journals will be asked to sign a Creative Commons Licence as opposed to a copyright form. All Creative Commons Licence types will incur an Article Processing Charge of \$1,450 USD. These licences allow the author to retain the copyright to their work. As we offer a number of different licence options for authors it is advised that authors take some time to consider which licence they would like to use when they submit their paper so as not to delay the publication of the paper on acceptance. Authors whose work is funded by the Wellcome Trust or Research Councils UK must sign the Creative Commons Attribution (CC BY). Other licences available are: Attribution-NoDerivs (CC BY-ND), Attribution-NonCommercial (CC BY-NC) and Attribution-NonCommercial-NoDerivs (CC BY-NC-ND). For more information on these options please see below.

Creative Commons Attribution (CC BY) This licence lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licences offered. Recommended for maximum dissemination and use of licensed materials. (If your work is funded by The Wellcome Trust or Research Councils UK you must use this licence.)

[Preview licence](#)

Creative Commons Attribution-NoDerivs (CC BY-ND) This licence allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you.

[Preview licence](#)

Creative Commons Attribution-NonCommercial (CC BY-NC) This licence lets others remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to licence their derivative works on the same terms.

[Preview licence](#)

Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) This licence is the most restrictive licence we offer, only allowing others to download your works and share them with others as long as they credit you, but they can't change them in any way or use them commercially.

[Preview licence](#)

10.4 Permissions to reproduce

10.4.1 If you wish to make use of previously published illustrations, diagrams or photographs in your paper, whether for commercial or non-commercial purposes, you must obtain the written permission of the [editorial office](#).

10.4.2 The IET will not make a charge where the work is being reproduced for non-commercial purposes, e.g. teaching, research presentations, etc.

10.4.3 Charges apply where IET copyrighted material is being reproduced for commercial purposes, e.g. book chapters, review articles.

10.4.4 The source of the material must also be acknowledged in full.

10.4.5 Please note, all authors have to obtain their own permissions for re-used text, figures and tables which do not belong to the IET in their submitted papers.

[Back to the top](#)

VITA

NAME: MIAO HE

EDUCATION:

B.S. in Mechanical Engineering, University of Science and Technology Beijing, Beijing, China, 2012

M.S. in Mechanical Engineering, Purdue University, Hammond, IN, 2014

PROFESSIONAL EXPERIENCE:

Job Title: Graduate Intern

*Date:*05/2017-08/2017

Employer: Siemens Corporate Technology

Location: Princeton, NJ

Job Description: Collaborated in the development of an online fault diagnosis system for train suspensions. Integrated signal processing, multi-level SVM, feature selection and Bayesian network on an embedded sensor system. Designed a Bayesian network on top of multi-level SVM, that helped to identify faulty components in suspension systems of train cart on dual vertical and horizontal directions, by improving 10% of accuracy in diagnosis (90% vs 80%). Implemented a stacked-sparse-autoencoder deep structure, that addressed the challenge on train cart bearing fault diagnosis with vibration signals by identifying occurrence of fault 4 months earlier than using signal processing techniques.

Job Title: Research Assistant

Date: 01/2015-05/2018

Employer: University of Illinois at Chicago

Location: Chicago, IL

Job Description: Developed a new signal processing and feature extraction module using manifold learning method, that helped to quickly extract efficient faulty features from bearing AE signals by showing clearer clusters of different bearing conditions. Developed a LAMSTAR network based module that solved the problem of intelligent bearing fault diagnosis using AE signals by improving 10% in diagnosis accuracy, accuracy as high as 96% even for low rotating speeds. Designed a module integrated with sparse coding and stacked autoencoder, which is implemented to diagnose gearbox fault by improving 15% detection accuracy using the self-learned features in comparison with standard stacked autoencoder. Proposed a deep structure using LAMSTAR that helps to detect planetary gear box faults used in wind turbines. Designed a hybrid deep signal processing method integrated discrete Fourier transform into autoencoder that helps to diagnose

bearing fault by improving 10% in diagnosis accuracies with self-learned features showing clear cluster and monotonic trend. Designed a module integrated with deep belief network and particle filter that solved remaining useful life prediction of hybrid ceramic bearing by improving 10% prediction accuracy.

Job Title: Teaching Assistant

Date: 01/2015-05/2018

Employer: University of Illinois at Chicago

Location: Chicago, IL

Job Description: Assisted the instructors with course management, e.g. homework, exams, and the coursework online platform. Lectured lab sessions with self-designed lectures including: big data analysis on Hadoop, Spark and RapidMiner, senior design with data mining using Python, computer-aided design and analysis using SolidWorks. Led the students' data mining related projects by 1) offering tutoring to the students on appropriate use of various data mining methods, 2) helping them to setup appropriate model of target issue and framework of seeking solutions.

Job Title: Laboratory Assistant

Date: 08/2013-12/2014

Employer: Purdue University Calumet

Location: Hammond, IN

Job Description: Developed LabView modules for heat transfer laboratories to aid students operating experiments to better understand the theory of heat transfer processes. Set up the material safety database for right-to-know information. Made a 5-year forecast for water withdrawals using an artificial neural network based on historical water withdrawals and industrial development data in Northwestern Indiana.

Job Title: Developer Intern

Date: 10/2011-04/2012

Employer: Intel China Center of Parallel Computing

Location: Beijing, China

Job Description: Assisted in design of Beijing Software Online Trading Platform with deep analysis of clients' needs. Cluster clients based on their user experience feedback by using logistic regression (LR) and random forest (RF) models. Optimize the recommendation system of e-trade platform based on clients' preferences from cluster results with 20% improvement in clients' trade records.

JOURNAL PUBLICATION:

Papers Published:

He, M., He, D., Wind turbine planetary gearbox feature extraction and fault diagnosis using a deep learning based approach, *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 2018.

He, M., He, D. and Qu, Y., A new signal processing and feature extraction approach for bearing fault diagnosis using AE sensors. *Journal of Failure Analysis and Prevention*, 2016, vol. 16, no. 5, pp. 821-827.

He, M and He, D., A deep learning based approach for bearing fault diagnosis. *IEEE Transactions on Industrial Application* 2017, vol. 53, no. 3, pp. 1-9.

He, M., He, D., A LAMSTAR network based approach for simultaneous bearing fault diagnosis and severity detection, *IET Science, Measurement & Technology*.

Zhou, Y., Lin, L., Wang, D., He, M., and He, D., A new method to classify railway vehicle axle fatigue crack AE signal. *Applied Acoustics*, vol. 131, pp. 173-185.

Qu, Y., He, M., Deutsch, J., and He, D., Detection of pitting in gears using a deep sparse autoencoder. *Applied Sciences*, 2017, vol. 7, no. 5, pp. 515.

Deutsch J., He, M., and He, D., Remaining useful life prediction of hybrid ceramic bearings using an integrated deep learning and particle filter approach. *Applied Sciences*, 2017, vol. 7, no., 7, pp. 649.

Chen, R., Chen, S., He, M., He, D., and Tang, B., Rolling bearing fault severity identification using deep sparse auto-encoder network with noise added sample expansion, *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*.

Papers Submitted:

He, M., He, D., A hybrid deep signal processing approach for bearing fault diagnosis using vibration signals, *Neurocomputing*

Conference Paper Published:

He, M., He, D., and Bechhoefer, E., Using deep learning based approaches for bearing fault diagnosis with AE sensors. *Annual Conference of PHM Society 2016*, Denver, CO.