UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

INTELLIGENT FAULT DIAGNOSIS OF RECIPROCATING COMPRESSOR USING DEEP LEARNING METHODS

by

Ying Zhang

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

Doctor of Philosophy

Sydney, Australia

2020

Certificate of Authorship/Originality

I, Ying Zhang certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This research is supported by the Australian Government Research Training Program.

Signature of Student:

Date: 15 November 2020

Production Note:
Signature removed prior to publication.

Abstract

Fault diagnosis of reciprocating compressors (RCs) based on vibration signals plays a vital role in guaranteeing a high operating reliability in RCs. Conventional maintenance schemes, which are carried out on a regular basis, can lead to unnecessary maintenance and shutdowns. Online health monitoring can monitor the working conditions of RCs continuously and provide more specific information, thus allowing the RC to be maintained as needed.

This PhD research focuses on the development of effective fault diagnosis methods using deep learning methods, thereby greatly advancing traditional health condition monitoring methods. Most traditional data-driven methods analyze the operating conditions using shallow models, which are incompetent at obtaining more confident results. To overcome this problem, a novel scheme based on deep learning models is proposed and applied to RC fault diagnosis. Traditional fault diagnosis methods select and extract features of raw signal with expertise and fuse them with shallow models. However, these methods cannot analyze the characteristics of signal in depth and thus degrade the performance of health monitoring. Deep learning methods are introduced in this research to calculate more representative features self-adaptive from the RC signals to improve fault diagnosis performance. As most fault diagnosis methods are based on vibration signals being the single information source, they cannot reflect the RC operating condition comprehensively. In this research, multi-source signals are collected and analysed for fault diagnosis. A scheme fusing multi-source information is proposed, as well as an auto-denoising network for denoising RC signals self-adaptively.

This PhD thesis consists of seven chapters. Chapter 1 provides research background. Chapter 2 presents a literature review. Chapter 3 proposes a method using intrinsic vibration feature fusion and a Grassmann manifold-based similarity. Chap-

ter 4 introduces the method of RC fault diagnosis using mode isolation-convolutional deep belief networks. Chapter 5 presents the intelligent fault diagnosis method using an optimized convolutional deep belief network. Chapter 6 proposes a novel ensemble empirical mode decomposition-convolutional deep belief network for RC fault diagnosis, and chapter 7 presents the conclusion and discusses future research in this area.

Acknowledgements

My research career as a Ph.D. candidate is coming to an end. For this duration, I would like to first express my sincere gratitude to my principal supervisor, A/Prof. Jinchen Ji. He provided me with great guidance and valuable suggestions on my research work. I am grateful for his patience and insightful ideas on revising my papers, and especially thank him for teaching me many research skills that I should equip with to be an eligible Ph.D. student.

I would like to thank my co-supervisor, A/Prof. Dongbin Wei, for his valuable suggestions on my research and immediate help as I needed. I feel great honor to have a supervisor who has rich research experience in fatigue analysis field to broaden my research horizon.

I would like to thank all my friends and colleagues for their help in my life. I am happy that we can work together and share this experience with each other. Much of our discussion has inspired me a lot.

I would like to thank all the staff members of UTS for their help in my research and thank UTS for providing us with elaborative learning resources. These really facilitate our research and help advance our research skills.

I would like to thank the China scholarship council (CSC) and University of Technology Sydney for the financial supports. Only with the great supports can I progress my research and pursue my PhD career.

Last but not least, I really appreciate my parents and my brother for their unconditioned love and powerful supports forever and ever.

Ying Zhang

Sydney, Australia, 2020.

List of Publications

Journal Papers

J-1. **Zhang, Y.**, Ji, J.C., 2020. Intelligent fault diagnosis of a reciprocating compressor using mode isolation-convolutional deep belief networks. *IEEE/ASME Transaction on Mechatronics*. [Published, DOI: 10.1109/TMECH.2020.3027912]

J-2. **Zhang Y.**, Ji, J.C. and Ma, B., 2020. Reciprocating compressor fault diagnosis using an optimized convolutional deep belief network. *Journal of Vibration and Control*, p.1077546319900115.

J-3. **Zhang, Y.**, Ji, J.C. and Ma, B., 2020. Fault diagnosis of reciprocating compressor using a novel ensemble empirical mode decomposition-convolutional deep belief network. *Measurement*, vol. 156, p.107619.

J-4. Ma, B., Zhao, Y., **Zhang, Y.**, Jiang, Q.L. and Hou, X.Q., 2019. Machinery Early Fault Detection Based on Dirichlet Process Mixture Model. *IEEE Access*, 7, pp.89226-89233.

J-5. **Zhang, Y.**, Ji, J.C., Use of intrinsic vibration feature fusion and a Grassmann manifold-based similarity for intelligent fault diagnosis of a reciprocating compressor, IEEE Transactions on Industrial Informatics. [Under review]

J-6. **Zhang, Y.**, Ji, J.C., Wei, D.B., Wear and fatigue analysis of support ring of reciprocating compressor using a deep belief network-conditional random field method. [Submitted]

Statement of Contribution of Authors

The chapters of this thesis contain materials from 4 papers published or submitted for publication in the journals.

Zhang, Y., Ji, J.C., 2020. Intelligent fault diagnosis of a reciprocating compressor using mode isolation-convolutional deep belief networks. *IEEE/ASME Transaction on Mechatronics*. [Published, DOI: 10.1109/TMECH.2020.3027912]

Contributor	Statement of Contribution	Thesis Chapter
Ying Zhang	Literature review, Conceptualization, Methodology, Validation, Investigation, Manuscript Writing- Original Draft. Overall contribution: 80%	Chapter 4
Jinchen Ji	]Conceptualization, Manuscript Writing-Review & Editing, Supervision. Overall contribution: 20%	

Zhang Y., Ji, J.C. and Ma, B., 2020. Reciprocating compressor fault diagnosis using an optimized convolutional deep belief network. *Journal of Vibration and Control*, p.1077546319900115.

Contributor	Statement of Contribution	Thesis Chapter
Ying Zhang	Literature review, Conceptualization, Methodology, Validation, Investigation, Manuscript Writing- Original Draft. Overall contribution: 80%	Chapter 4
Jinchen Ji	Conceptualization, Manuscript Writing- Review & Editing, Supervision. Overall contribution: 15%	
Bo Ma	Experimental Data Collection Overall contribution: 5%	

Zhang, Y., Ji, J.C. and Ma, B., 2020. Fault diagnosis of reciprocating compressor using a novel ensemble empirical mode decomposition-convolutional deep belief network. *Measurement*, vol. 156, p.107619.

Contributor	Statement of Contribution	Thesis Chapter
Ying Zhang	Literature review, Conceptualization, Methodology, Validation, Investigation, Manuscript Writing- Original Draft. Overall contribution: 80%	Chapter 4
Jinchen Ji	Conceptualization, Manuscript Writing- Review & Editing, Supervision. Overall contribution: 15%	
Bo Ma	Experimental Data Collection Overall contribution: 5%	

Zhang, Y., Ji, J.C., Use of intrinsic vibration feature fusion and a Grassmann manifold-based similarity for intelligent fault diagnosis of a reciprocating compressor, IEEE Transactions on Industrial Informatics. [Under review]

Contributor	Statement of Contribution	Thesis Chapter
Ying Zhang	Literature review, Conceptualization, Methodology, Validation, Investigation, Manuscript Writing- Original Draft. Overall contribution: 80%	Chapter 4
Jinchen Ji	Conceptualization, Manuscript Writing- Review & Editing, Supervision. Overall contribution: 20%	

Contents

List of Figures

List of Tables

Abbreviation

RC - Reciprocating Compressor

EMD - Empirical Mode Decomposition

IMF - Intrinsic mode function

PCA - Principal Component Analysis

KPCA - Kernel Principal Component Analysis

DBN - Deep Belief Network

GM - Grassmann manifolds

RBM - Restricted Boltzmann Machines

SVD - Singular Value Decomposition

SVM - Support Vector Machine

TD - Time-domain

FD - Frequency-domain

CNN - Convolutional Neural Network

CDBN - Convolutional Deep Belief Betwork

LDA - Latent Dirichlet allocation

AE - Auto Encode

CS - Compressed Sensing

SF - Sparse Filtering

CRBM - Convolutional Restricted Boltzmann Machine

ANN - Artificial Neural Network

EEMD - Ensemble Empirical Mode Decomposition

CC - Correlation Coefficient

WT - Wavelet Transforms

PCM - Probabilistic Committee Machine

GPC - Gaussian Process Classifier

EM - Expectation Maximization

MI - Mode isolation

GMM - Gaussian mixture model

p-V - Pressure-volume

Chapter 1

Introduction

1.1 Background

Reciprocating compressors (RC) are one of the most important machineries in the petroleum industry, such as in offshore oil and gas production applications [83]. Once faults or failures occur in the RC, it may cause operation disruption and serious economic loss for manufacturers. Therefore, Intelligent fault diagnosis of the RC plays an important role in early detection of potential failures to ensure a reliable operation [45]. With the rapid development of the Internet of Things and smart manufacturing, the amount of condition monitoring data available is growing at an explosive speed [106]. Firstly, this implies that fault diagnosis has entered an era of big data. Secondly, the data have a more complex structure and more useful hidden information [7]. Therefore, how to implement a more reliable fault diagnosis with this "big data" has become a challenging problem in this field.

Fault diagnosis has drawn significant attention in various research fields [96, 42, 115]. Fault is defined as the deviation of observations or parameters from the acceptable range in a certain process. The aim of fault diagnosis is to find the defective components operating out of the normal ranges on the machinery. Although fault diagnosis has been extensively studied for several decades, RC fault diagnosis is still a challenging problem due to various vibration excitation sources and complex transfer paths of vibration. Currently, fault diagnosis methods can be categorized into three main groups: model-based methods, knowledge-based methods and data-driven based methods. As the former two methods require extensive knowledge

about fault mechanism, they are highly limited when compared to the current availability of data [21]. Consequently, data-driven methods have been used extensively for fault diagnosis due to its potential in processing big data [33, 27].

In recent decades, considerable effort has been devoted to the development of condition monitoring methodologies [24, 100, 43, 89]. However, some aspects can be further improved to increase diagnosis accuracy. First, most existing methods of feature extraction are labour-intensive and time-consuming. For instance, prior knowledge is required to design a suitable scheme of feature extraction [56]. The absence of prior knowledge on the scheme of feature extraction could lead to inappropriate feature selection and further degrade the fault diagnosis performance [125]. Second, the sophisticated structure and operating conditions of RCs contribute to the complex structure of RC vibration signals [95]. Direct analysis of these complex RC raw signals would be unable to highlight useful information that can reflect the salient local characteristics of the operating conditions, as the mode mixture of signals and other information unrelated to fault may disturb the fault identification, such as noise [103]. Furthermore, the straight analysis could decrease the robustness of the fault diagnosis results and lead to poor performance of RC fault diagnosis.

To address these problems, deep learning methods were explored in this research. Recently, deep learning has attracted increasing attention from researchers of various research communities due to its state-of-the-art performance in refining deep information from data. Deep learning refers to machine learning techniques that use supervised and/or unsupervised strategies to automatically learn hierarchical representations from deep network architectures [28]. The most significant advantage of deep learning methods is that they can construct deeper networks and calculate more representative features even from big data [39]. This method has been successfully applied in many research areas [80]. Therefore, this research was launched on the basis of the deep learning method.

1.2 Research objectives

The aims of the project were to implement intelligent RC fault diagnosis, which included the following aspects.

i. Data acquisition.

   To monitor RC health condition, data indicating the RC operating condition were collected from sensors, including accelerometers monitoring the cylinder and crankcase vibration, temperature sensors measuring temperature variation of the cylinder and phase sensors measuring the rotating speed and piston location. The acquired data were used to validate the proposed method.

ii. Optimization of traditional fault diagnosis method using deep learning.

   Traditional methods extract features from raw signals and fuse features with conventional models. Features from conventional models could include more irrelevant information that disturb fault diagnosis. To overcome this problem, deep learning methods will be used to fuse features in depth and improve fault diagnosis performance.

iii. Self-adaptive feature extraction based on a deep learning method.

   Typically, traditional methods select features in the RC fault diagnosis with expertise. However, inadequate selection of features by insufficient experience would decrease the performance of fault diagnosis. To address this issue, a framework based deep learning was adopted in this research.

iv. Optimization of a deep learning method in the application of RC fault diagnosis.

   In terms of the RC vibration mechanism, optimization of the deep learning method was proposed to enhance its performance in the application of RC fault diagnosis.

v. Fault diagnosis based on multi-source information fusion.

Currently, most research on RC fault diagnosis adopts vibration signal as the single source of information for fault diagnosis. This does not reflect the operating conditions comprehensively. To encompass more information sources from RCs, a scheme fusing multi-source information was proposed to enhance the RC fault diagnosis performance.

1.3 Thesis organization

The rest of the thesis is organised as follows:

- *Chapter 2*: This chapter presents the literature review of the research. It introduces the development of RC fault diagnosis and other related techniques.

- *Chapter 3*: Traditional methods extract features and fuse features with conventional model with simple matrix manipulations.To fuse features in depth, this chapter proposes a method which fused features in depth to enhance the RC fault diagnosis performance. Empirical mode decomposition (EMD) was used to analyse the intrinsic vibration of RC. Deep belief network (DBN) was used to fuse features in depth and obtain high-representative features. A Grassmann manifold-based similarity was also proposed for RC fault diagnosis to preserve the non-linearity of singals.

- *Chapter 4*: Chapter 3 used a conventional method for faultdiagnosis. This required prior knowledge for feature extraction. Insufficient knowlwdge could degrade the performance of fault diagnosis. This chapter adopted a deep learning based framework. Considering the RC vibration mechanism, a mode isolation-convolutional deep belief network was proposed and validated by the RC condition monitoring data in this chapter. Mode isolation-convolutional

deep belief network (MI-CDBN) can isolate multi-modal data and extract features self-adaptively.

- *Chapter 5*: The CDBN in chapter 4 adopted a max pooling. Though it preserved some desirable features, maxmium value could degrade the generalization ability of CDBN. To improve the performance of CDBN, this chapter proposes an optimized convolutional deep belief network and presents the resulting application to RC fault diagnosis. An optimized probabilistic out was proposed in this chapter to enhance the generalization of the convolutional deep belief network (CDBN).

- *Chapter 6*: The prior research used vibration signals as the single source of fault diagnosis information, which could not relect the operating conditions comprehensively. In addition, those method did not consider the negative effect of background noise on fault diagnosis. To onvercome these problems, this chapter presents a method of denoising signals and a scheme of fusing multi-source signals. An auto-denoising network was proposed to eliminate the noise existing in the RC vibration signals. A probabilistic committee machine (PCM)-based method was proposed to fuse multiple sources of information with a more reasonable weight.

- *Chapter 7*: A brief summary of the thesis contents and its contributions are given in the final chapter. Recommendations for future works are given as well.

Chapter 2

Literature review

To conduct a suitable RC fault diagnosis, an extensive literature review, covering the general overview of multiple methods, is explored accordingly.

2.1 Fault diagnosis based on the traditional methods

Traditional data-driven methods of RC fault diagnosis mainly include three steps: data acquisition, feature extraction, and feature classification or fault diagnosis [84]. Among these three steps, feature extraction is the most crucial step in fault diagnosis [127]. The quality of the feature can significantly affect the diagnosis results. Typical vibration features include kurtosis, peak value, peak-peak value, mean square root, and impulsive index [66, 79]. Fault diagnosis based on the vibration feature extraction has received substantial research effort. For example, Feng et al. extracted frequency domain features from an indicator diagram to diagnose RC faults. A discrete 2D-Curvelet transform was adopted to extract the representative features from the indicator diagram. Next, nonlinear principal component analysis (PCA) was employed for multi-class recognition to reduce dimensionality, as well as for novelty detection. Finally, multi-class and one-class support vector machines (SVMs) were used as the classifier and novelty detector, respectively. Experimental results showed the effectiveness of the proposed approach [25]. Althobiani et al. used an Teager-Kaiser energy operator and deep brief networks (DBNs) to diagnose the faults of an RC valve. To reveal the fault patterns contained in this signal, the Teager–Kaiser energy operation was proposed to estimate the amplitude envelopes. In case of pressure and current, random noise was removed using a denoising method based

on a wavelet transform. Subsequently, statistical measures were extracted from all signals to represent the characteristics of the valve conditions. To classify the faults of compressor valves, a new type of learning architecture for deep generative models, called DBNs, was applied. The experimental results proved the effectiveness of the proposed method [4]. Ahmed et al. extracted several time-domain, frequency-domain and envelop-domain features to represent machinery operating conditions and used a relevance vector machine for fault classification [1]. Zhao et al. used local mean decomposition and multi-scale fuzzy entropy to calculate the features for bearing clearance fault in the RC [125]. Li et al. proposed a new method for non-destructive RC fault diagnosis using a strain-based pressure-volume (p-V) diagram. This method extracted the key feature points on the piston rod load curve that reflect the opening and closing events of the compressor valves. The algorithm was validated by comparing the p–V diagrams obtained from direct pressure measurement and strain-based derivation. The reconstructed p–V diagrams were further used for RC fault diagnosis. The results indicated that this method was able to monitor the operating conditions and identify fault type and location [52]. Cabrera et al. used a long short-term memory model to extract features for RC valve fault diagnosis and used a Bayesian model to adjust the parameter of the model. Experimental data confirmed the effectiveness of the proposed method of [10]. Yan et al. extracted sensitive features from roller bearing signal, and proposed an optimized support vector machine to conduct fault diagnosis. The result shows the effectiveness of the proposed method [111]. Zhang and Deng proposed a method integrating adaptive neuro fuzzy inference system (ANFIS) and Dempster–Shafer theory (DST) to operate the fault diagnosis of engine. The experimental data proved the practicability of the proposed method [121]. Jiang et al. proposed an initial center frequency-guided variational mode decomposition for fault diagnosis of rotating machines [41]. The conventional method were applied in various fields and achieved

good performance. These methods extract one, or multiple, time and/or frequency features from original signals to represent the operating conditions. These features can reflect the operating conditions to some extent. However, achieving appropriate feature selection and identification of faults using this method relies heavily on prior knowledge, which limits its application in the industry.

2.2 Feature dimension reduction

Once features have been extracted, conventional models are used to fuse features and reduce feature dimension. For example, Liu et al. proposed a method to calculate high-dimensional features of roller bearing using a local characteristic-scale decomposition-Teager energy operator.Intrinsic multifractality features were extracted from decomposed signals and constructed into a high-dimensional feature. Then the dimension was reduced by principal component analysis (PCA). The experimental results proved that the PCA can fuse features by linear transformation [57]. De et al. applied optimized kernel principal component analysis (KPCA) to reduce feature dimension and used an artificial neural network to perform fault diagnosis.The result illustrated the effectiveness of the proposed method in industrial system [22]. Sakthivel et al. extracted statisticial features and used feature dimension reduction techniques to reduce feature dimensions. This paper compared different dimension reduction techniques for the fault diagnosis of a monoblock centrifugal pump using vibration signals, including PCA, KPCA, isomap, and Laplacian eigenmaps.The result proved that the PCA plus decision tree obtained the best performance in the fault diagnosis of mono block centrifugal pump [79]. Although these methods can fuse features effectively and reduce feature dimension, they may not be adequate for obtaining high-representative features, particularly in the current big data field. This is due to their inability to reveal the deep relationship between complex data/signal structure and health conditions.

Deep learning methods have attracted increasing attention from academic communities due to their state-of-the-art performance in deep feature fusion and feature extraction without any assistance on feature extraction. These methods use supervised and/or unsupervised strategies to automatically learn hierarchical representations in deep architectures for feature extraction, transformation and classification [47]. Among all the deep learning methods, the DBN has demonstrated advantages in constructing deeper networks and calculating low-dimensional features with excellent representations from big data [18]. Liu et al. adopted DBN to extract high discriminative features for an electronics-rich analog system. Experimental results show the fault diagnosis based on DBN achieved superior diagnostic performance than the traditional feature extraction methods. [62]. Arsa et al. proposed a dimensionality reduction method using DBN for hyperspectral image classification, which also reduces the computational cost. In the proposed framework, the first DBN is used to reduce the dimension of spectral bands and the second DBN is used to extract spectral-spatial feature and as a classifier. An Indian Pines data set that consisted of 16 classes were used to validate the proposed method and to compare the proposed DBN with the PCA. The results indicated that using DBN as a dimensionality reduction method performed better than PCA in hyperspectral image classification [5]. PCA can fuse features and obtain a low dimensional, but cannot fuse features in depth due to the shallow structure. DBN, with its deep structures, is better equipped to solve this problem. Mutual information of neighbouring layers is applied in DBN feature dimension reduction, which, along with deep structures, allows features to be fused in-depth. The obtained high-representative features can reflect the data information more effectively.

2.3 Feature extraction by deep learning methods

To be unrestrained by prior knowledge, the deep learning method was also proposed to calculate signal features self-adaptively [126] and has achieved great success in multiple fields [15, 110, 117], including object detection [34] and fault detection in wind turbines [14]. DBN and convolutional neural network (CNN), as two deep learning methods, have achieved high recognition due to their excellent performance in unsupervised feature learning. Mohamed et al. used DBN to establish models of acoustic signals in language processing [70]. Hassan et al. identified features from images to implement face recognition [35]. Thus, it is evident that the DBN method can calculate features with high efficiency by generative models. Gu et al. conducted a review on the development of the CNN method, which provided a broad survey of the recent advances in CNNs, including layer design, activation function, loss function, regularization, optimization and fast computation. Additionally, various applications of CNNs in computer vision, speech and natural language processing were also introduced [29]. Xia et al. used CNN to implement the fault diagnosis of rotating machinery [108]. Janssens et al. conducted machine detection and oil prediction using CNN to learn unsupervised features [38]. Cheng et al. extracted features using CNN to estimate remaining bearing life [19]. Thus, the CNN method has demonstrated its ability to extract useful information from raw data of high dimension and preserve the property of shift-invariance. To further enhance the advantages of deep learning methods, CDBN was proposed and applied to multiple domains. Lee et al. used CDBN to extract features by unsupervised learning from audio signals and perform audio identification [48]. Ren and Wu obtained more effective information from signals using CDBN to extract features from electroencephalographic signals [75]. Li et al. combined DBN and CNN to extract features which are used to implement the fault diagnosis of rotating machinery [53]. Shao et al. proposed an optimized CDBN method to accomplish bearing fault diagnosis with

excellent application performance. This paper showed the convolution of CDBN and the operation of the proposed method, which were shown in figure 2.1 and figure 2.2 [82]. In application, the effectiveness of feature extracted is evaluated by the

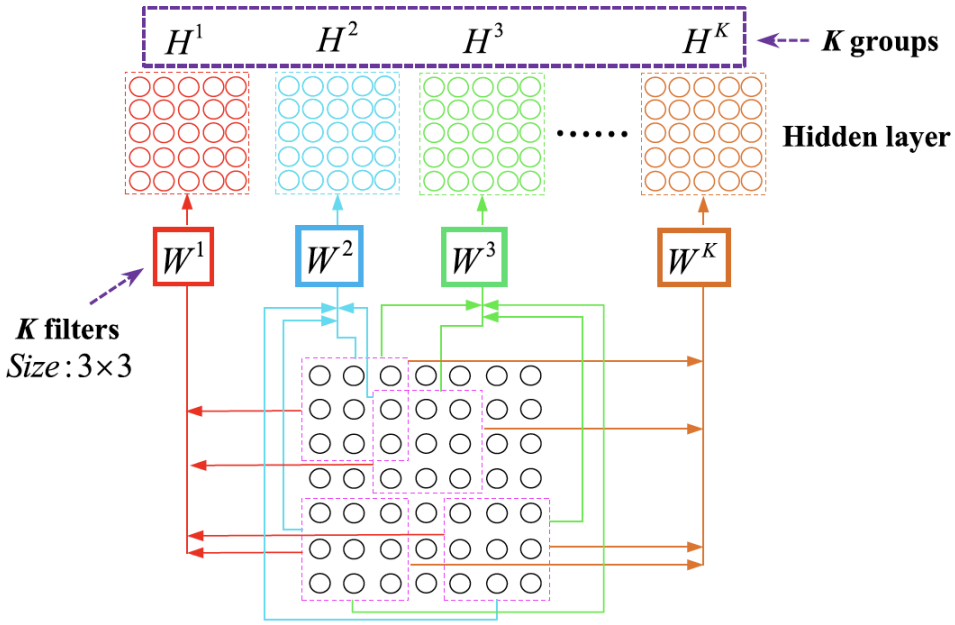


Figure 2.1 : Convolution of CDBN

fault diagnosis result, namely the accuracy. The experimental results show their effectiveness and great success in various applications. The CDBN has shown its superiority to the DBN and the CNN in terms of the accuracy metric. The CDBN preserves the property of shift-invariance of CNN and the high efficiency of DBN. Moreover, CDBN can be extended to handle multi-channel data.

2.4 Signal denoising

Due to the sophisticated production environment of the RC, the acquired signals are usually contaminated by background noise [74]. Noise disturbance can negatively affect the fault diagnosis results and undermine the fault diagnosis performance [86]. Therefore, it is essential to construct an auto-denoising network to eliminate the

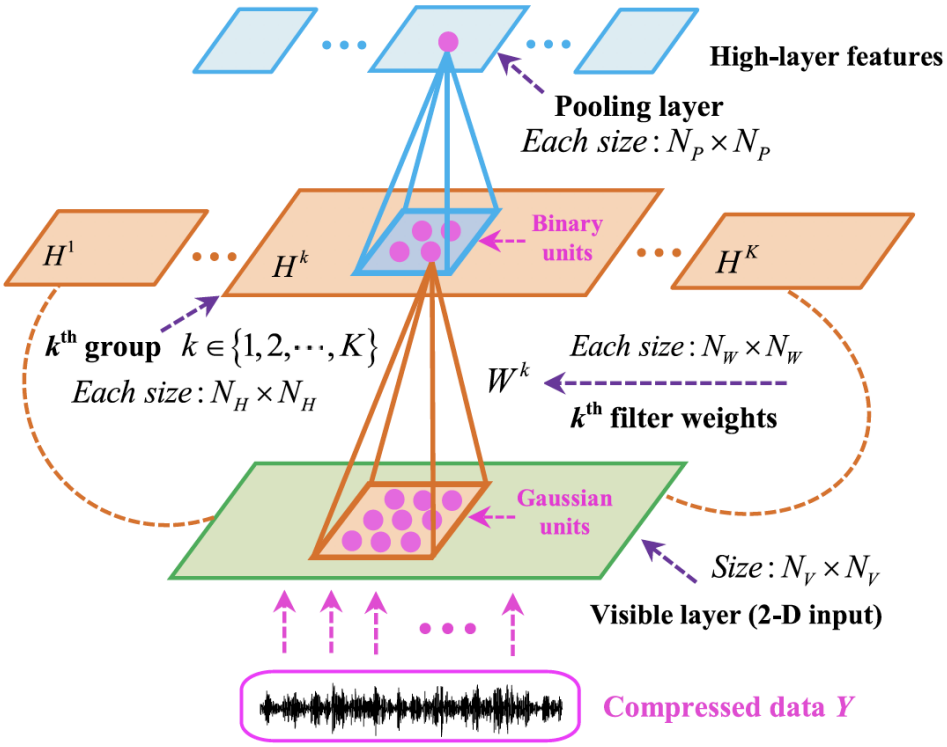


Figure 2.2 : Operation of CDBN

noise existing in signals. A variety of methods have been explored for denoising by filtering [24]. Wavelet transforms (WT) are widely used in a variety of research fields for signal denoising [59]. This method requires manual selection of base functions before denoising [69]. This means that an inappropriate base function could be selected for denoising, due to insufficient expertise [13]. As a result, this can generate poor denoising performance and decrease fault diagnosis accuracy. To overcome this, EMD was proposed to extract useful data from noisy and non-stationary signals self-adaptively [123]. Ali et al. used EMD to decompose signals into a series of intrinsic mode functions (IMFs) and employed artificial neural networks to identify faults of roller bearings [3]. Li et al. applied an EMD-based method to predict the load of power system operation and control [51]. Wang et al. proposed an EMD-assisted manifold for fault diagnosis of rotating machinery [101]. Li et al. used

optimized EMD to diagnose gear faults [54]. Wang et al. utilized integrating EMD manifold to extract features and accomplish machinery fault diagnosis. EMD was used to decompose signals and useful IMFs were selected by correlation analysis for machinery fault diagnosis [97]. An improved EMD method, based on the multi-objective optimization, was proposed in this paper and applied to extract the fault feature of rolling bearing with inner and outer race fault [31]. EMD, as a self-adaptive decomposing method of decomposing signals, can decompose signals into several IMFs in terms of the time-scale of the signal itself [77]. This reduces human interference with signal decompositions [103], which is well suited for analyzing non-linear and non-stationary signals (the signal properties of RCs) [73]. Compared to wavelet transforms, EMD does not required selection of basis functions in advance, thus avoiding reduced fault diagnosis accuracy due to unreasonable selections.

However, mode mixing and end effects are two disadvantages of the EMD [68]. To solve these problems, ensemble empirical mode decomposition (EEMD) was proposed, as it can separate each mode of data more precisely from raw signals [30]. Wang et al. proposed an EEMD denoising method with singular spectrum constraint for Ladar signals. The IMFs decomposed by EEMD are stacked together to obtain the denoised signals. Tests on synthetic and real data demonstrated that the proposed method, compared to the EMD denoising method, could suppress more noise but filter out less useful signals in the FMCW Ladar signal denoising [102]. Cheng et al. proposed a segmentation singular value decomposition (SVD)-lifting wavelet transform (LWT) denoising algorithm based on ensemble empirical mode decomposition (EEMD) to better suppress noise in an atmospheric lidar return signal. The EEMD method was used to distinguish IMFs of the noise and signal, and remove the IMF with noise as its main component[20]. The EEMD could distinguish the hidden mode of signals more accurately by averaging signals with white noise added, which allows signals to be denoised effectively and self-adaptively. This

method could be more suitable to handle the RC signal denoising problem.

2.5 Multi-source information fusion

The fusion of multi-source information is a challenging obstacle [119, 78]. Fusing information from multiple sensors can provide more comprehensive information on operating conditions and further enhance the performance of the fault diagnosis method [18].

Traditional methods tend to fuse information of each source with equal weights, which assumes that each source of information has the same sensitivity to faults and contributes equally to the fault diagnosis. This may not be a reasonable scheme, as different types of features can have unequal sensitivities to different faults and contribute differently to diagnosing various faults [127]. To overcome this problem, much research has been carried out. Cai et al. fused multi-source information for fault diagnosis of a ground-source heat pump using a Bayesian network. This method increased the diagnostic accuracy of a ground-source heat pump system, especially for multiple-simultaneous faults [11]. Zhong et al. conducted scene classification based on the multi-feature fusion probabilistic topic model for high spatial resolution remote sensing imagery. A geological survey data set and the UC Merced data set were utilized to evaluate the proposed method in comparison with conventional methods. The experimental results confirmed the superiority of the proposed method [128]. Xu and Yu proposed a novel approach to information fusion in multi-source datasets from a granular computing viewpoint. This method allowed valuable and reliable information sources to be chosen by transforming the original information of each object into a triangular fuzzy information granule. Experiments confirmed the effectiveness of the proposed method [109]. Liu et al. proposed an innovative information fusion method using adaptive Kalman filtering. This method can integrate information from INS/GPS navigation of autonomous vehicles [60]. Li

et al. proposed the physics of failure-based reliability prediction method using multi-source information fusion to predict the reliability of aero engine turbine blades. In the proposed method, the fuzzy theory was employed to represent uncertainties involved in prediction. Case studies of reliability prediction under fuzzy stress with and without fuzzy strength were conducted using a dynamic stress-strength interference model which considers types of aero-engine cycles. Results indicated that the proposed method was more in line with engineering practice and more flexible in decision making. Furthermore, the proposed method could predict the reliability of aero-engine turbine blades as an interval by utilizing the proposed linear fusion algorithm [50]. Che et al. designed a multi-granulation probabilistic rough set based on evidence theory, probability theory and information entropy to fuse uncertain data in a multi-source information system. This study is helpful for integrating the uncertain information of multiple sources and consequential for creating a route of granular computing [12]. Wang et al. presented a deep learning-based model named multi-resolution and multi-sensor fusion network for motor fault diagnosis, achieved through multi-scale analysis of motor vibration and stator current signals. This method can automatically learn discriminative features through the network training process without any assistance of prior knowledge [100]. From the literature review, it can be seen that fusing multiple sources of data can provide more comprehensive information for handling problems in different fields and enhance the performance of final results. It is an effective framework to fuse information with the probabilistic scheme.

2.6 Summary

Currently, RC fault diagnosis mainly are conducted by traditional methods. Features are extracted with prior knowledge. Insufficient knowledge could lead to inappropriate feature selections and further decrease the performance of fault diagnosis.

By addressing this problem, deep learning methods are developed due to the ability of unsupervised feature learning. In this research, deep learning method is introduced to conduct fault diagnosis of RC. Due to the advantage of CDBN over other deep learning methods, RC fault diagnosis is conducted based on the CDBN. In addition, in the RC fault diagnosis, vibration signals are the primary single source of information. To combine more information, a scheme fusing multi-source information is proposed to improve fault diagnosis performance.

Chapter 3

Intelligent fault diagnosis using intrinsic vibration feature fusion and a Grassmann manifold-based similarity

3.1 Introduction

Traditional monitoring methods extract only one to a few time and/or frequency features from original signals [24] and use a conventional model, such as a principal component analysis [2], to fuse features and reduce feature dimension or use trending analysis to track changes in the health condition[120] *. Although these methods achieve good performance in fault diagnosis and classification to a certain extent, there are three remaining limitations: (1) Traditional methods extract features from raw signals which ignores more local features. Local features means the feature extracted from segmentation of signals or decomposed components of signals, which reflect more local characteristics from various scaling of signals. Features extracted from raw signals could only reflect operating conditions globally. For example, features extracted from a cycle of signal can reflect characteristics of the whole cycle. This could decrease the reliability of the features. (2) Conventional models may not be adequate in feature dimension reduction of big data, as they cannot reveal deep relationships between hidden information in big data and the RC operating conditions. (3) As RC signals are characterized by a non-linearity and non-stationary nature [33], Euclidean distance-based methods or any assumption on data distribu-

*This chapter has been submitted as: Zhang, Y., Ji, J.C., Use of intrinsic vibration feature fusion and a Grassmann manifold-based similarity for intelligent fault diagnosis of a reciprocating compressor, IEEE Transactions on Industrial Informatics.[Under review]

tion would not be suitable for RC fault identification. Motivated to overcome these issues, this chapter launches the research from the following three aspects.

EMD is adopted in this chapter to decompose raw vibration signals of RCs into a collection of IMFs, which can reflect the intrinsic vibration of an RC. These intrinsic functions can reflect intrinsic vibrations of an RC from local perspectives, which could represent the operating condition more confidently.

To obtain useful information from intrinsic vibrations comprehensively, multiple features were extracted from each IMF and constructed into a high dimensional vector. This vector can adequately reflect the RC operating conditions comprehensively [26]. Due to the complex production environment of RCs, RC vibration signals could be contaminated by noise or other unhelpful disturbances. The extracted features could include redundant information, which can negatively affect fault diagnosis. Therefore, dimension reduction techniques have been adopted to remove redundant information in feature vectors and enhance the fault diagnosis performance. In view of the advantage of DBN in unsupervised feature learning, DBN has been used for feature dimension reduction of RC fault diagnosis.

After reducing the feature dimension, a pattern recognition method was applied to perform the labelling and identification of the fault categories intelligently. Most methods of fault identification are conducted with Euclidean distance [64] or under assumptions, such as that the data follows a certain distribution [9]. As RC vibration signals present nonlinear and non-stationary properties [116], this assumption may be considered unreasonable. It may neglect more information of nonlinearity and thus impair the performance of the fault diagnosis method. To overcome this problem, a similarity based on Grassmann manifolds (GM) was designed in this chapter for RC fault diagnosis. GM is composed of a series of subspaces that can represent the working conditions of RCs [85]. The similarity of subspaces on GM can

be measured by geodesic distance [90]. It can preserve more non-linear properties of original subspace which reflect the RC operating conditions.

Aiming at intelligent fault diagnosis in the context of big data, this chapter proposes a hybrid method incorporating EMD for intrinsic vibration analysis, DBN for feature dimension reduction and a Grassmann manifold-based similarity for RC fault identification. The contributions of this chapter are listed as follows.

i. EMD was utilized to decompose signals into a series of IMFs. The IMFs provided a physical interpretation in the RC diagnosis, which indicates the intrinsic vibrations generated from the RC operation.

ii. Multiple features were extracted from each IMF and then constructed into high-dimensional feature vectors to reflect health condition comprehensively. Then the feature vectors were fused deeply by DBNs to reduce feature dimensions. Due to the deep structure of DBNs, the fused features had a high representativeness of the operating conditions compared with conventional feature dimension reduction methods.

iii. The low-dimensional feature vectors were transformed into subspaces on the GM. The similarity of subspace on the GM was proposed to determine RC faults. Compared with Euclidean methods, this method can calculate the similarity while preserving more linear properties of RC signal and enhancing the performance of RC fault diagnosis.

This chapter is organized as follows. Section 3.2 introduces the proposed method and gives a brief introduction to the related theories. Section 3.3 analyzes the experimental data and validates the proposed method. Section 3.4 concludes the paper.

3.2 Proposed method

A novel method is proposed to enhance the performance of RC fault diagnosis. EMD is used to analyze the local information from the perspective of intrinsic vibrations. This method decomposes RC vibration signals into IMFs. Features are extracted from each IMF and constructed into high-dimensional vectors in order to reflect more detailed and comprehensive characteristics of operation. Then DBN is employed to reduce the dimension of feature vectors and remove the redundant information. As the deep structure of DBN, it can fuse RC features in depth by complex non-linear mapping. Meanwhile, these low-dimensional feature vectors are constructed into subspaces on the GM. The faults could then be determined by calculating the similarity between subspaces. As GM is a geodesic distance, it could preserve more nonlinear property of RC signals. The detailed procedure is illustrated in Fig. 3.1 and described as follows.

i. RC vibration signals were collected using accelerometers mounted on the cylinder.

ii. The measured signals were decomposed into a series of IMFs by EMD.

iii. Features were extracted from each IMF to reflect RC operating conditions with more local information.

iv. The extracted features were constructed into a high-dimensional feature vector to represent the working condition.

v. These feature vectors were split into training data and testing data.

vi. Feature vectors were input into DBN to reduce the feature vector dimension.

vii. Feature vectors of training data were transformed into subspaces on the Grassmann manifold for each fault as the base subspace.

viii. Subspaces of testing data were calculated by repeating Step 6 and Step 7.

ix. The similarities between subspaces of testing data and fault data were calculated to determine the fault type.

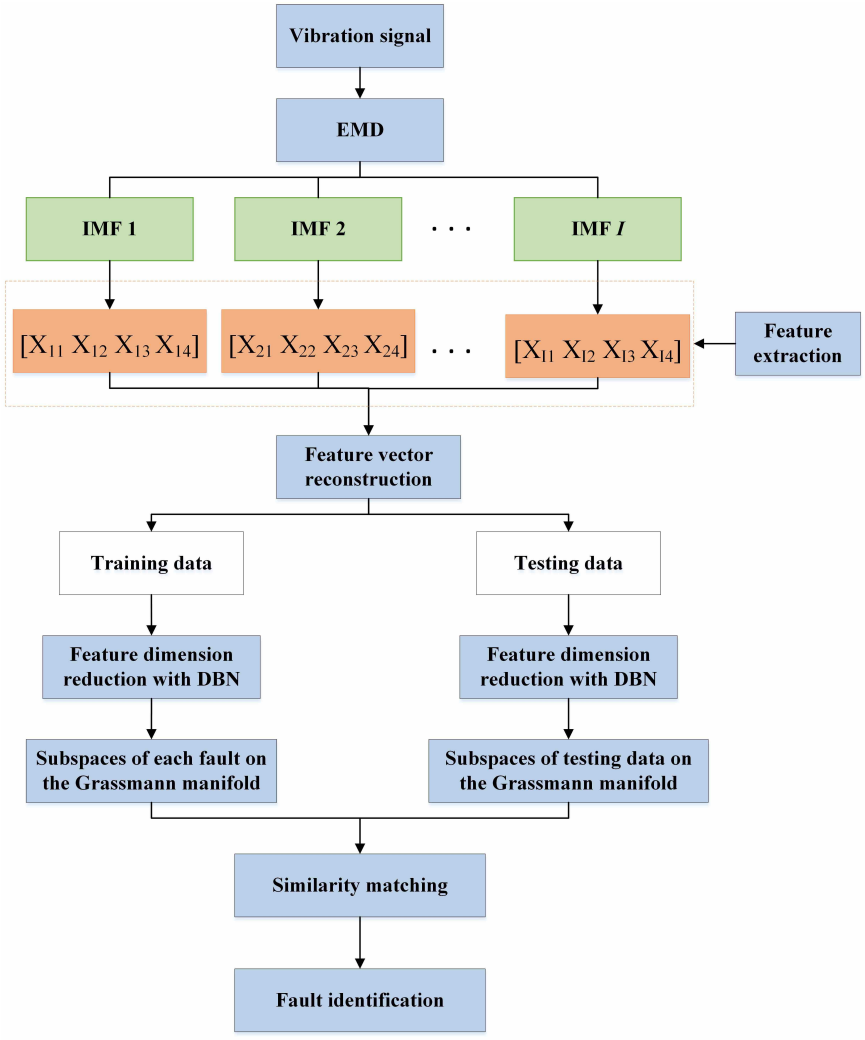


Figure 3.1 : Schematic of the proposed method

3.2.1 Empirical mode decomposition

EMD can decompose signals into a series of IMFs, and each IMF represents an intrinsic vibration mode generated by an RC. Therefore, IMF can present more local information of the operating conditions. EMD can decompose a signal into a

collection of IMFs. These IMFs are of different time scales, which means they can provide information from different time scales and can reflect operating conditions by more local characteristics, thereby improving the fault diagnosis performance. The general steps of EMD are listed below.

i. The upper envelops and lower envelops of original signals as well as their means m_1 are calculated.

ii. The difference between the raw signals $x(t)$ and m_1 is calculated and can be written as

$$x = x(t) - m_1 \tag{3.1}$$

iii. x is considered a raw signal to compute the first IMFs by repeating Steps 1 and 2 k times. The first IMF can be expressed as

$$\mathrm{IMF}_1 = x_1 = x_{1(k-1)} - m_{1k} \tag{3.2}$$

iv. The above three steps are repeated until satisfying the stop criterion [65] and all the IMFs are computed.

v. From the above steps, it can be induced that

$$x(t) = \sum_{i=1}^{I} \mathrm{IMF}_i + r \tag{3.3}$$

where I is the total number of IMFs and r denotes the residual function.

3.2.2 High-dimensional feature extraction and feature vector reconstruction

To reflect the RC operating conditions from more local perspectives, features were extracted from each IMF. The extracted features are shown in Table 3.1. These extracted features are typical features applied in health condition monitoring. After features were extracted from each IMF, they were constructed into a

Table 3.1 : Extracted features

Feature	Equation	Feature symbol	Meaning
Root mean square	$X_{RMS} = \sqrt{\frac{\sum_{i=1}^{n}(x(i))^2}{n}}$	X_1	The average energy of signals
Crest factor	$X_{Crest} = \frac{\max\{x(i)\}}{X_{RMS}}$	X_2	The peak amplitude divided by the RMS value
Kurtosis	$X_{Kurt} = \frac{\sum_{i=1}^{n}(x(i)-\bar{x})}{n(X_{RMS})}$	X_3	The shape of a probability, sensitive to impulsive faults
Peak	$X_{Peak} = \max\|x\|$	X_4	Indicating the intensity of vibration

high-dimensional feature vector. The vector can be expressed as

$$X = (X_1, X_2, \cdots, X_i, \cdots, X_I) \tag{3.4}$$

where $X_i = [X_{i1}, X_{i2}, X_{i3}, X_{i4}]$, indicating features extracted from the i-th IMF. These features were concatenated into a high-dimensional feature vector to represent the operating condition comprehensively.

3.2.3 Deep belief networks

A deep belief network consists of deep architectures that are capable of learning feature representations from unlabeled data and exploring complex data characteristics [124]. Figure 3.2 shows a typical DBN architecture. X $= (X_1, X_2, \cdots, X_m)$ represents the high-dimensional features of IMFs and $O = (o_1, o_2, \cdots, o_n)$ indicates the low-dimensional features calculated by the DBN with unsupervised learning. The DBN consists of i hidden layers with each layer having $m_1, m_2, \cdots, m_k$ nodes, respectively, and is comprised of a stack of restricted Boltzmann Machines (RBMs).

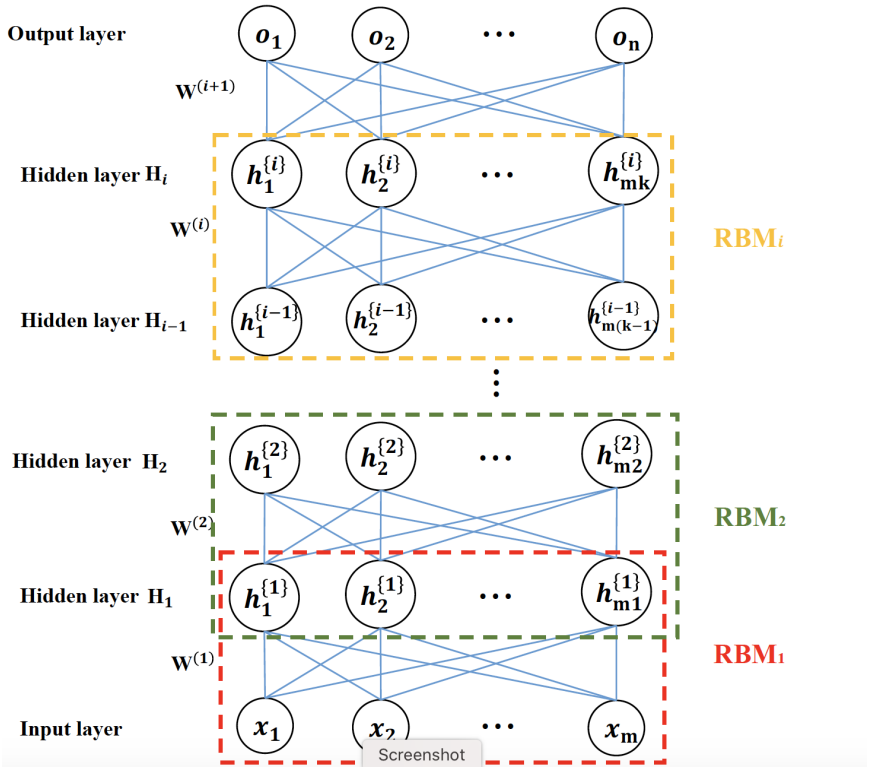


Figure 3.2 : The DBN structure

The RBMs are probabilistic generative models that learn a joint probability distribution from training data without data labels. A RBM consists of two layers, namely the input layer and output layer. There is no connection among nodes in the same layer, and the neighboring layers are connected by weights (and biases) matrices. The outputs of the current RBM are the inputs of the following RBM. The energy of the input can be computed efficiently by

$$E(v,h) = -a^T v - b^T h - v^T W h \tag{3.5}$$

In general Boltzmann machines, probability distribution over h and/or v is defined by the following energy function

$$P(v,h) = \frac{e^{-E(v,h)}}{Z} \tag{3.6}$$

where Z is a partition function representing the sum of all possible configurations.

The marginal probability of a Boolean visible vector is defined as the sum of all possible hidden layer configurations, which is expressed by

$$P(v,h) = \frac{1}{Z}\sum_{h} e^{-E(v,h)} \tag{3.7}$$

The conditional probability of the visible units under a given hidden unit can be written as

$$P(v\,|h) = \prod_{i=1}^{m} P(v_i\,|h) \tag{3.8}$$

Conversely, the conditional probability of h under the given v is

$$P(h\,|v) = \prod_{j=1}^{mk} P(h_j\,|v) \tag{3.9}$$

The individual activation probabilities can be calculated by

$$P(h_j = 1\,|v) = \sigma\left(b_j + \sum_{i=1}^{m} w_{ij}v_i\right) \tag{3.10}$$

$$P(v_i = 1\,|h) = \sigma\left(a_i + \sum_{j=1}^{mk} w_{ij}h_j\right) \tag{3.11}$$

where σ indicates sigmoid function. For more information on DBN refer to [94].

3.2.4 Grassmann manifold-based similarity

After the feature dimensions have been decreased by DBN, a similarity based on Grassmann manifold was designed to identify fault types. Grassmann manifold is applied in the research, as it can calculate feature difference by geodesic distance. It considers the nonlinear property of data [67], which is suitable to handle RC vibration signals. This method constructs a series of subspaces on the Grassmann manifold and then calculates the geodesic distance between subspaces to determine fault types.

Grassmann manifolds is a Riemannian manifold that is embedded in a high dimensional Hilbert space [91]. A GM $\mathcal{G}_{d,D}$ is composed of a set of d-dimensional

subspaces of $\mathbb{R}^D$. The subspaces are spanned by orthonormal matrices Y and represented by $span(Y)$. Two low-dimensional feature matrices O_1 and O_2 were assumed to be obtained from DBN and are spanned by Y_1,$Y_2 \in \mathbb{R}^{D\times d}$. Principal angles were calculated to measure the similarity of two subspaces $span(Y_1)$ and $span(Y_2)$, and the principal angles of two subspaces could be calculated by

$$\cos\theta_i = \max_{u_i \in span(Y_1)} \max_{v_i \in span(Y_2)} u_i v_i \tag{3.12}$$

with the following conditions

$$\begin{gathered} u_i' u_i = v_i' v_i = 1 \\ u_i' u_j = v_i' v_j = 0 \\ i = (1, 2, \cdots, d) \quad j = (1, 2, \cdots, i-1) \end{gathered} \tag{3.13}$$

where u and v are the principal vectors. In this chapter, singular value decomposition (SVD) was adopted to calculate principal angles [118]. The equation can be expressed as

$$Y_1' Y_2 = USV^* \tag{3.14}$$

where $U = [u_1, u_2, \cdots, u_d]$ is a unitary matrix, $S = [\cos\theta_1, \cos\theta_2, \cdots, \cos\theta_d]$ is a diagonal matrix, and V^* is the conjugate transpose $V = [v_1, v_2, \cdots, v_d]$. The similarity between two subspaces was calculated from geodesic distance and defined as

$$dist(Y_1, Y_2) = \sum_{i=1}^{d} \cos^2\theta_i \tag{3.15}$$

When $\theta_1 = \theta_2 = \cdots = \theta_d = 0$, two subspaces $span(Y_1)$ and $span(Y_1)$ are considered into one. This means a larger value of $dist(Y_1, Y_2)$ indicates a higher similarity between subspaces, and it is more likely that feature matrices O_1 and O_2 are of the same RC fault data. The subspaces of each fault were calculated as a reference and were matched to the subspace of testing data for identification to obtain the similarity. The fault type is determined by the value of similarity.

3.3 Experimental verification and analysis

3.3.1 Data description

Vibration data that reflects RC operating conditions were collected by vibration accelerometers (sensor type: PCB 608A11) mounted above the crosshead. The acquired data were used to validate the proposed method. An RC schematic is shown in figure 3.3, which illustrates the structure of the RC and the sensor locations. The

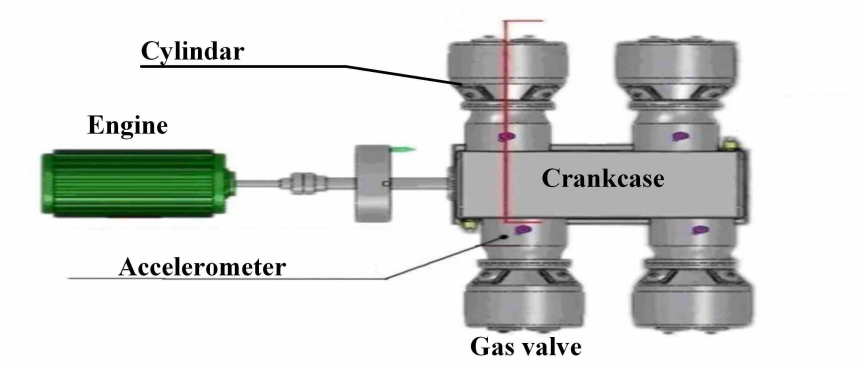


Figure 3.3 : Schematic of RC structure

engine lying on the left end provided the output power to drive the rotation of the shaft. Phase sensors were installed on the flywheel near the engine to monitor the rotating speed. Accelerometers were mounted on the crosshead to measure the cylinder vibrations, which can reflect the operating conditions of the RC. The piston rods in the cylinder can take in and push out gas via the gas valve. A crank-link mechanism was located inside the crankcase and transformed shaft rotation into compression motion of the piston rod in the cylinder to complete gas compression. Figure 3.4 shows four typical RC faults occurring in an oil refinery plant, gas valve leakage, piston rod breaking, cylinder scraping, and bearing shell wear. Figure 3.4(a) shows a broken gas valve which causes gas valve leakage. Figure 3.4(b) displays a broken piston rod. Extended runtime under component fatigue can lead to piston breaking. Figures 3.4(c) and (d) show cylinder scraping and the wear of bearing

shell faults, respectively. Due to worn-down components and delayed replacement of assembly, piston rods and bearings will scrape cylinders or bearing shells directly. This significant contact friction can lead to more serious faults, such as piston rod breaking. These faults or failures may cause serious issues in operation or even disastrous incidents without early detection.

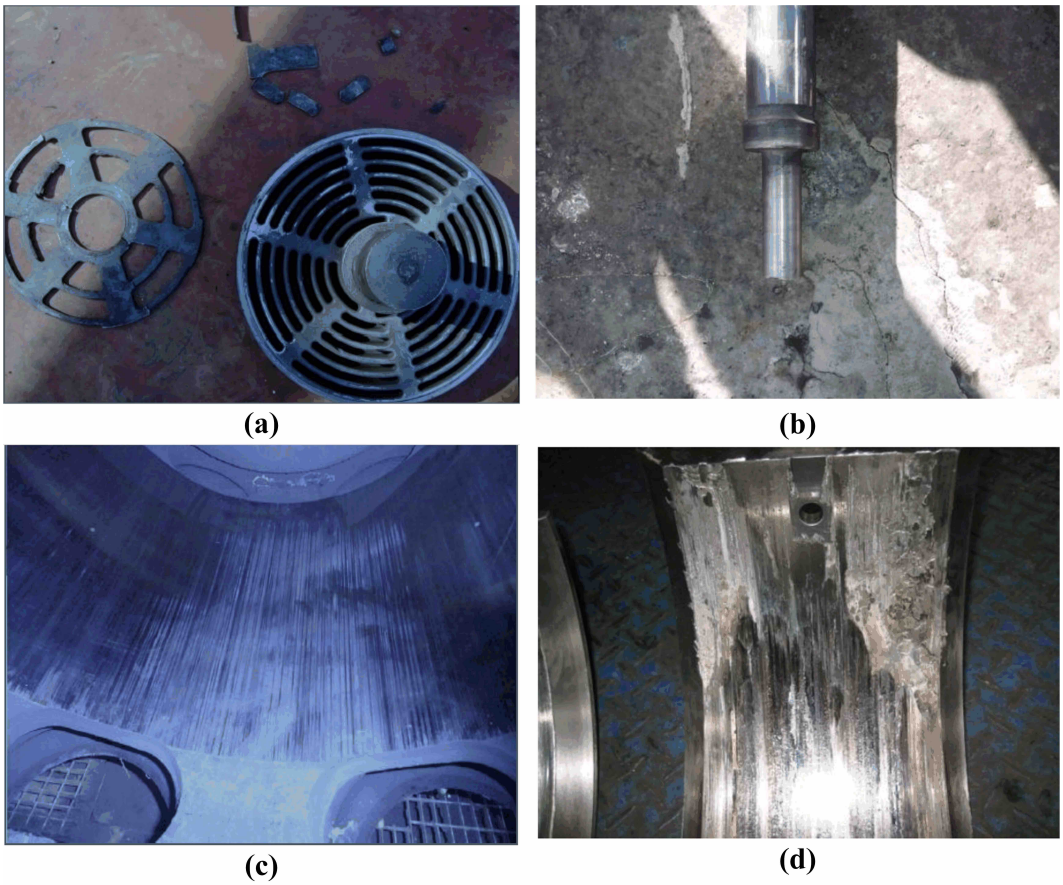


Figure 3.4 : Faults of RC

Vibration monitoring is an effective method in RC fault diagnosis. Vibrations change after faults occur in the RC. Therefore, vibration signals were collected for method validation in this research. Detailed information on the fault data are listed in table 3.2. Signals of five operating conditions were acquired under the rotating speed of 375rpm and the sampling frequency of 12.8kHz to validate the

Table 3.2 : Date description

Fault type	Description	Sampling frequency/kHz	Size of training/ testing data
F1	Wear of bearing shell	12.8	200/100
F2	Cylinder scraping	12.8	200/100
F3	Gas valve	12.8	200/100
F4	Piston rod breaking	12.8	200/100

developed method. As the rotating speed is 375rpm, the duration of each signal is 60/375=0.16s. Each type of data included 300 signals with 200 randomly selected samples as the training data and the rest as the testing data.

3.3.2 RC fault diagnosis using the proposed method

EMD was first used to decompose signals into a series of IMFs. The decomposed results are shown in figure 3.5. Raw signals were decomposed into 9 IMFs. Each IMF can reflect an intrinsic vibration mode of original signals. Four features in table 3.1 were extracted from each IMF and constructed into a high-dimensional vector. The dimension of the feature vector was equal to 36. To remove redundant information from the feature vector, high-dimensional features were input into the DBN for feature dimension reduction. The DBN was composed of three RBMs. The parameter setting of the DBN is listed in table 3.3. There is no criteria for the parameter setting [82]. They are typically tuned according to the performance in the fault diagnosis. The last layer of the DBN included 10 nodes, thus the output dimension of the feature vector was 10. The first three components of the lower-dimensional feature vector are visualized and displayed in Figure 3.6, where different colors represent different operating conditions. It can be seen from the figure that

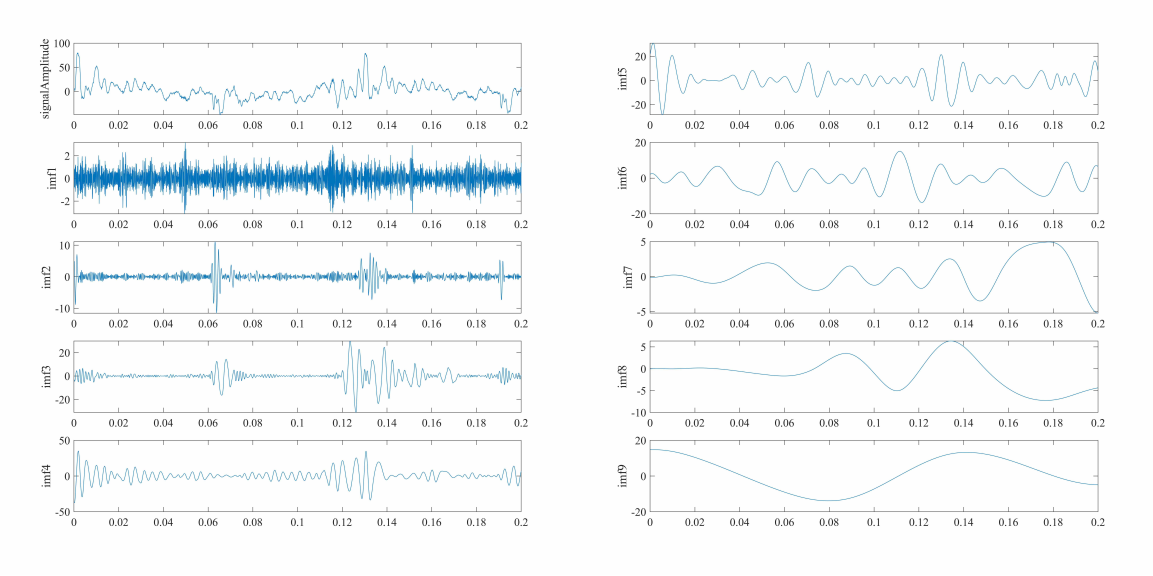


Figure 3.5 : EMD results

the features extracted from the same condition are clustered tightly, whereas features of different conditions are clearly distinguishable. This proves that DBN can refine the highly representative features from raw signals effectively.

Subspaces of the low-dimensional feature vector were calculated on the Grassmann manifold to determine the fault type. The performance of the proposed method was evaluated for application to RC fault diagnosis. The accuracies of fault diagnosis are displayed by the confusion matrix in Figure 6.11. The horizontal and vertical axes indicate the actual labels and predicted labels of data, respectively. The values in yellow are the fault diagnosis accuracy for the data of each fault type while the values in green are the detailed misclassified ratio of each fault. The accuracy of the proposed method for fault identification reached up to 86.6%, 84.2%, 85.5% and 87.6% for fault F1, F2, F3 and F4, respectively. The diagnosis accuracies in this chapter were averaged over ten trials to ensure stable final results.

Table 3.3 : Parameter setting of the DBN

Parameter	Setting
The number of RBMs	3
The number of input nodes of the 1st RBM	36
The number of output nodes of the 1st RBM	30
The number of output nodes of the 2nd RBM	20
The number of output nodes of the 3rd RBM	10
Learning rate	0.05
The number of epochs	30

3.3.3 Parameter analysis and method evaluation

The number of nodes in the DBN output layer was investigated to determine the most suitable feature vector dimension for RC fault diagnosis. The relationship between the number of nodes in the output layer and the diagnosis accuracy is exhibited in Figure 3.8. With increasing output dimension, accuracy rose to achieve the peak value of 85.98% at the dimension of 10. After this, accuracy begins to decrease. This reveals that the lower-dimensional features consist of the most useful information at the dimension of 10. When the dimension exceeds 10, more redundant information is included in the features, and it affects the performance of the proposed method negatively. Therefore, the dimension of 10 was adopted as a reasonable output dimension. The proposed method was compared with other methods in three aspects, feature extraction methods, dimension reduction techniques and feature identification methods. The comparison results are listed in Table 3.4. In method 1, features were extracted from original time-series signals and resulted in a poor performance in application. Method 2 used EMD to decompose vibration signals into multiple intrinsic vibrations and could reflect more local characteristics

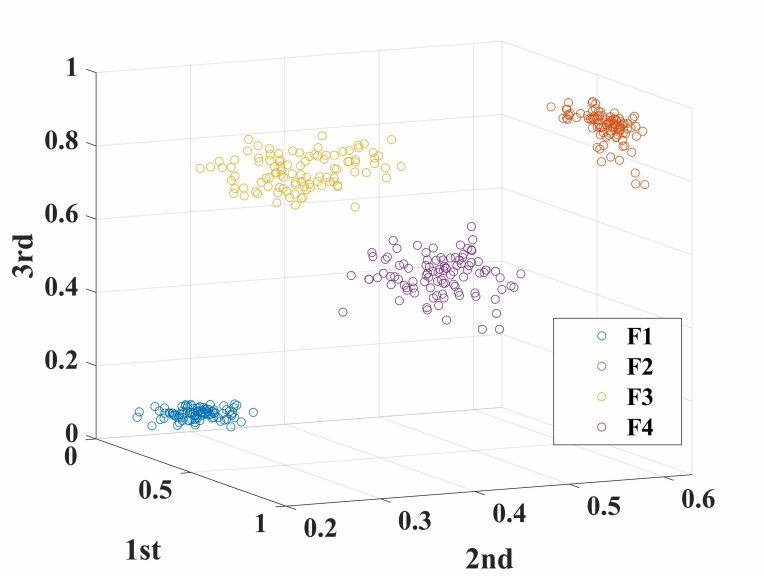


Figure 3.6 : DBN feature visualization

of the operating condition, thus the diagnosis performance improved. In methods 3, 4 and 5, feature dimension reduction techniques were applied to refine the high-representative information. This further enhanced the accuracy of fault diagnosis. However, method 5 achieved higher accuracy than method 3, which proves that DBN can fuse features more effectively and preserve more useful information. For example, non-linear information, which tends to be neglected by traditional methods. Additionally, method 5 surpassed method 4 in accuracy. This reveals that GM can preserve more reliable properties of RC operating conditions and provide a more reasonable similarity between testing data and fault data than a Euclidean distance-based method (e.g. SVM) in RC fault diagnosis.

3.4 Conclusion

To refine more reliable features in depth and implement intelligent fault diagnosis for reciprocating compressors, this chapter proposed a novel method fusing intrinsic vibration features by DBN and measuring the similarity of feature subspaces on

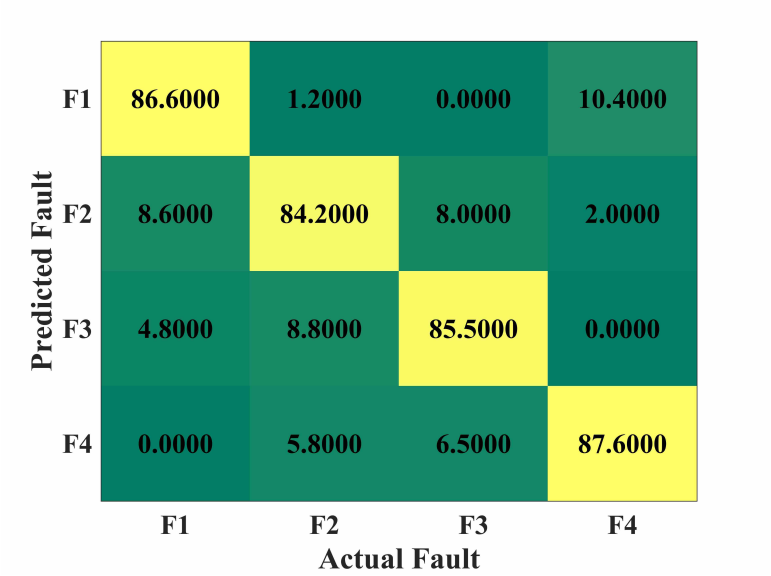


Figure 3.7 : Confusion matrix

Grassmann manifold.

First, signals were decomposed into IMFs by EMD. This decomposition could separate intrinsic vibration modes from raw vibration signals effectively and reveal more local and reliable information.

Secondly, multiple features were extracted from each IMF and constructed into a high-dimensional feature vector. These features could reflect more comprehensive information of the intrinsic vibration.

Thirdly, the high-dimensional feature vector were fused in depth by DBN to obtain a lower-dimensional feature vector. The deep network structure allowed the features to be fused in depth and further guaranteed more reliable information was preserved, for example non-linearity information.

Finally, feature vectors were transformed into subspaces on the Grassmann manifold for multiple faults. A subspace of testing data and subspaces of each fault were matched to determine fault types by their similarities. The GM-based similarity preserved more non-linearity information by adopting a geodesic metric.

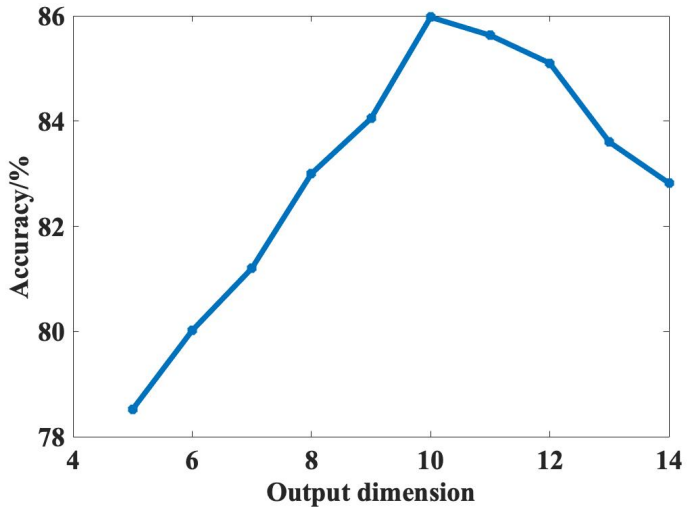


Figure 3.8 : The relationship between output dimension and accuracy

Table 3.4 : Method evaluation

Method	F1	F2	F3	F4
features + GM	72.3	70.8	71.5	69.3
EMD+features + GM	75.6	74.2	74.6	76.6
EMD+features + KPCA + GM	82.2	83.5	82.4	84.9
EMD+features + DBN + SVM	85.8	83.6	83.5	85.8
EMD+features+ DBN + GM	86.6	84.2	85.5	87.6

RC fault data were collected from an industrial oil refinery plant to validate the proposed method. The results confirmed that the proposed method was able to extract more detailed information and fuse features in depth. The Grassmann manifold-based similarity could identify RC faults more effectively.

Chapter 4

Intelligent fault diagnosis using mode isolation-convolutional deep belief networks

4.1 Introduction

CDBN shows superiority in feature extraction with unsupervised learning*. This chapter adopts the CDBN method to calculate features in the RC fault diagnosis. Although the raw signals can reflect the RC operating condition from a global perspective, more useful local information is unfortunately neglected. To take into account more useful local features and to improve the performance of CDBN in RC fault diagnosis, the MI-CDBN is proposed in this chapter. The transfer path theory is introduced to the RC fault diagnosis to analyze the potential structures of the RC signals from the perspective of multi-modal data isolation, which is inspired by the theory of transfer path analysis in [95]. Vibration signals were considered as the superposition of multi-modal data generated from different transfer paths. Each mode was calculated by a data-driven method, considering the absence of prior knowledge on vibration mechanism. To isolate multi-modal data with unknown distribution, a Gaussian mixture model (GMM) was adopted to establish the models for data following complex distributions, due to its flexibility in data modelling [61]. As the parameters of GMM can represent vibration conditions robustly, they were used to construct into state space for feature learning from CDBN. The state space is

*This chapter has been accepted as: Zhang, Y., Ji, J.C., 2020. Intelligent fault diagnosis of a reciprocating compressor using mode isolation-convolutional deep belief networks. *IEEE/ASME Transaction on Mechatronics*. [Published, DOI: 10.1109/TMECH.2020.3027912]

referred to as time-invariant [46]. Consequently, they can represent operating conditions confidently and facilitate the RC fault diagnosis. Subsequently, the calculated features from MI-CDBN were fed into classifiers to identify fault types of an RC. A multi-class logistic regression classifier was adopted to recognize multiple faults of the RC [6].

The contributions of this chapter are summarized in the following three aspects.

i. An MI-CDBN is proposed to enhance the performance of the standard CDBN in implementing the RC fault diagnosis. Direct analysis of raw signal could hide more local information from different modes and degrade the fault diagnosis performance. To extract more local information, GMM is used to isolate multi-modal data from raw signals, then the isolated multi-modal data is input into CDBN to learn more useful local information. To highlight more local information, multi-modal data are isolated from raw signals. GMM is used to isolate multi-modal data in this paper due to its flexibility in data modelling, then the network self-adaptively calculates features from isolated modal data. This method can highlight more information hidden in the raw signals. This method also provides a novel physical interpretation of the RC vibration signals from the perspective of the transfer path.

ii. The CDBN is used to extract features for the raw signals. Compared with the other deep learning methods, CDBN can extract features using a deep network structure with shift-invariance and high efficiency. By strengthening the advantages of the CNN and the DBN, this method can obtain more reliable information from the collected RC vibration signals and ensure improved performance in the RC fault diagnosis.

iii. A three-stage method is adopted in the RC intelligent fault diagnosis based on a deep learning-based framework. First, sparse filtering is employed to

compress signals into more compact series to reduce computing cost. Then, the MI-CDBN is used to calculate features by unsupervised learning. Finally, fault types are identified by a classifier. This method can remove the dependence on prior knowledge and enhance computational efficiency.

iv. The GMM is used to isolate multi-modal data generated from various transfer paths. The parameters of the GMM are calculated using latent Dirichlet allocation (LDA) instead of being set at a constant. This improves the robustness of the constructed state space in the RC fault diagnosis.

The rest of the chapter is organized as follows. Section 4.2 introduces the related preliminary of the proposed method, and Section 4.3 gives a detailed description of the proposed method. In Section 4.4, site measurement data from the petroleum industry are employed to validate the proposed method and Section 4.5 concludes the chapter.

4.2 Preliminary

4.2.1 Sparse filtering

Due to the technological advancements in data acquisition and storage, voluminous monitoring data of the RC can be collected by sensors, which makes it challenging to handle big data [110]. Therefore, it is essential to compress data and refine the highly representative information for fault diagnosis. Sparse filtering (SF) is one of the most promising tools due to its advantages of a basic hyperparameter setting and low computational cost [71]. Only the output dimension of the SF requires tuning for data compression. This can reduce computation cost compared with compressive sensing (CS) and allow the SF to obtain a good performance in intelligent fault diagnosis [49]. Moreover, SF can compress data with unknown distribution with little distortion of raw signals [49]. This would be more suitable for

compressing RC vibration data than the auto-encoder (AE) method in terms of the complex structure of vibration signals [71].

Sparse filtering was used to compress raw signals into compact time series with the fault-related information extracted. Sparse filtering captures three properties: sparse features per example, sparse features across examples, and uniform activity distribution [71]. This method represents raw signals with a few non-zero high representative coefficients (features). The compressed feature can be calculated by

$$f_j^{(i)} = \mathrm{w}_j^{\mathrm{T}} \mathrm{s}^{(i)} \tag{4.1}$$

where $f_j^{(i)}$ denotes the j-th feature of the i-th data sample in the feature matrix with each column representing a sample and w_j representing the weight matrix. ℓ_2-norm was adopted to normalize features within the same range and ensure each feature was activated equally.

$$\tilde{\mathrm{f}}_{\mathrm{j}} = \mathrm{f}_{\mathrm{j}} / \left\| \mathrm{f}_{\mathrm{j}} \right\| \tag{4.2}$$

Then each column was normalized within a unit ℓ_2-ball.

$$\hat{\mathrm{f}}^{(i)} = \tilde{\mathrm{f}}^{(i)} / \left\| \tilde{\mathrm{f}}^{(i)} \right\| \tag{4.3}$$

ℓ_1 penalty was applied to optimize the sparsity of the normalized features. The objective function of sparse filtering for a dataset including N samples is given by

$$\text{minimize} \quad \sum_{i=1}^{N} \left\| \hat{\mathrm{f}}^{(\mathrm{i})} \right\|_1 \tag{4.4}$$

The weight matrix satisfying Eq.(4.4) could be regarded as the optimal results of compression and the input $\mathbf{x}$ of the MI-CDBN.

4.2.2 Traditional CDBN

Convolutional deep belief network has been widely used in many fields. It inherits the advantages of DBN and CNN by integrating them. The traditional CDBN was constructed by stacking convolutional restricted Boltzmann machine (CRBM), whose energy function is denoted as [82]

$$\begin{aligned} -\log P(\mathbf{v},\mathbf{h}) &\propto E(\mathbf{v},\mathbf{h}) \\ &= -\sum_{k=1}^{K_c}\sum_{i,j=1}^{N_H}\sum_{r,s=1}^{N_W} h_{i,j}^{k} W_{r,s}^{k} v_{i+r-1,j+s-1} \\ &\quad -\sum_{k=1}^{K_c} b_k \sum_{i,j=1}^{N_H} h_{i,j}^{k} - c\sum_{r,s=1}^{N_V} v_{i,j} \end{aligned} \tag{4.5}$$

where $\mathbf{W}$, $\mathbf{v}$ and $\mathbf{h}$ represent the filter, visible vector and hidden vector, respectively. $v_{i,j}$ is an element located in the i-th row and the j-th column of the k-th filter. b_k is the bias of each hidden group and all visible units share the same bias c. The conditional probabilities of the standard CRBM can be expressed as:

$$P\left(h_{i,j}^{k}=1 \mid \mathbf{v}\right)=\sigma\left(\left(\tilde{W^{k}} * v\right)_{i,j}+b_k\right) \tag{4.6}$$

$$P\left(v_{i,j}=1 \mid \mathbf{h}\right)=\sigma\left(\sum_{k,i,j}\left(W^{k} * h^{k}\right)_{i,j}+c\right) \tag{4.7}$$

where $\sigma(x)=1/(1+e^{-x})$ is the sigmoid function, $*$ is the convolutional operation, and $\tilde{W_{i,j}^{k}}=W_{N_W-j+1}^{k}$. Once CDBN identifies the features, they could be categorized by multi-class logistic regression [36].

4.3 The Proposed method

4.3.1 General procedure of the proposed method

A novel method was proposed to conduct intelligent fault diagnosis of the RC by making the best use of transfer path analysis and multi-modal data isolation. The proposed method primarily involved three stages. First, the raw data was

compressed using sparse filtering to obtain high representative features and reduce computational cost. Then, the MI-CDBN was employed to isolate multi-modal data and to learn unsupervised features from the compressed signals. Finally, the representative features were input into a multi-class logistic regression to identify fault types. The theoretical framework is displayed in figure 4.1. The detailed procedure consists of the following six steps.

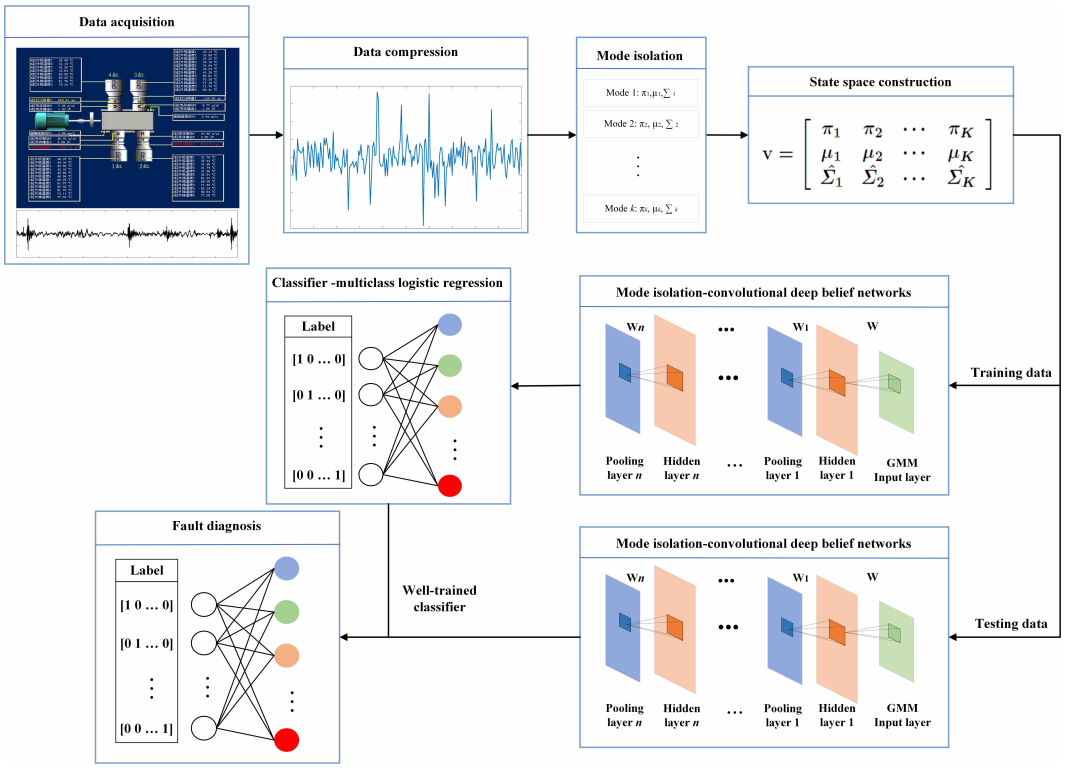


Figure 4.1 : Framework of the proposed method

Step 1: Vibration data were measured by accelerometers installed on the crosshead of the RC.

Step 2: Data were compressed using sparse filtering to refine and achieve highly representative information and decrease the computational cost.

Step 3: The GMM was used to isolate multi-modal data and the state space was constructed from the parameters of each mode to represent the operating conditions

of the RC robustly.

Step 4: The MI-CDBN was adopted to calculate the features from the constructed state-space by unsupervised learning.

Step 5: Multi-class logistic regression classifier was trained to implement the fault diagnosis of the RC.

Step 6: Testing data were input into the well-trained classifier to evaluate the performance of the proposed method.

4.3.2 Mode isolation-convolutional deep belief network

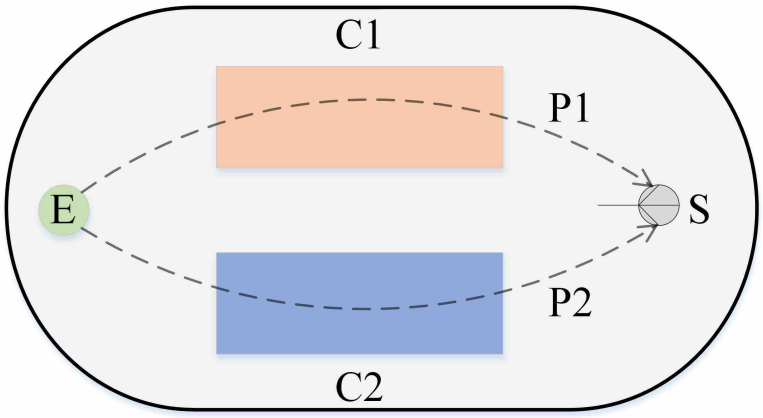


Figure 4.2 : Diagram of transfer path on the RC

Transfer path analysis is an effective approach for analyzing complex vibration mechanism when it is difficult to measure vibration excitation directly [95]. This method considers a mode of data generated from an excitation via a specific path. The collected signals are a combination of multiple modes of data. However, different modes of vibration data cannot be measured by sensors directly. Therefore, these modal data are isolated using the data-driven method by assuming that these modal data follow Gaussian distribution. GMM can establish models for data with complex distributions. This is suitable for handling RC signals which are generated

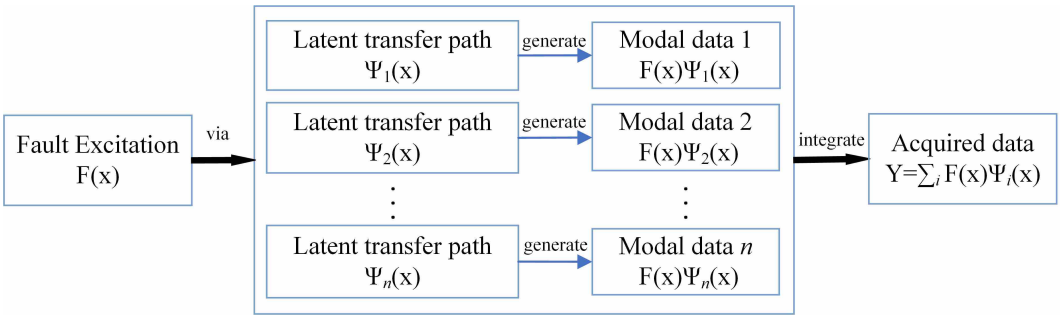


Figure 4.3 : The relationship among fault excitation, transfer paths and acquired data

from complex vibrations and follow complex distributions. Thus, GMM was adopted to resolve the problem and projects raw data into state-space. The parameters of GMM are used to construct the state space, and currently, they are mostly estimated by expectation maximization (EM) directly. It is highly likely to obtain an overfitting value of GMM parameter using the EM method [76], which decreases the generalization ability of GMM [104]. Considering the Bayesian method advantages of learning parameters from prior knowledge [122], latent Dirichlet allocation (LDA) was employed to establish the GMM model parameters. LDA has an excellent performance in constructing models for discrete data [128, 99, 114], which can represent various potential operating conditions of the RC [17]. From the LDA model, the running condition of the highest possibility will be selected for fault identification.

In the RC fault diagnosis, a diagram of one source of excitations is shown as an example in figure 4.2. Each source of excitations was assumed to be independent of each other. When a fault occurs in the RC, a vibration excitation E will be generated. The vibration induced by E cannot be measured directly by sensors. It will transfer along a certain latent path P in the RC body. For example, path $P1$ (dotted curve) via component $C1$ (red square), path $P2$ (dotted curve) via component $C2$ (blue square). Finally, the vibrations could be measured by a sensor

S. The modal data generated from the two paths can be expressed as $\mathrm{F}(\mathrm{x})\,\Psi_1(\mathrm{x})$ and $\mathrm{F}(\mathrm{x})\,\Psi_2(\mathrm{x})$, as the vibration of $C1$ and $C2$ are modulated by the vibration of E. Here only two modal data are taken for explanation. In the following introduction to the theoretical framework, the transfer path was extended to a random number n. The collected data presents the non-linear and non-stationary nature, with n types of modal data generated from various transfer paths combined. The relationship among vibration excitations, transfer paths, and acquired data were formulated and are shown in figure 4.3. In figure 4.3, the vibration data for the RC fault diagnosis incorporated multi-modal data and each modal data was generated by the vibration of one fault excitation transferring along a certain transfer path. Different mode combinations of multi-modal data indicate different operating conditions. These combinations can represent transfer paths and their weights, as the fault excitation dominates the variation of a path and the corresponding weight. In this research, the modes were isolated by projecting the raw signals to the state-space to reflect local features. Straight analysis of the raw signals with multiple-mode data mixed could degrade the performance of the fault diagnosis method and hide or neglect more useful information. In contrast, the local information obtained from mode-isolated data could characterize and reflect the operating condition more robustly. In practice, it is difficult to measure the fault excitation F(x) and transfer paths $\Psi(\mathrm{x})$ directly [95]. Thus a data-driven method was proposed to establish the models for multi-modal data generated from various transfer paths. Theoretically, the raw signals can be expressed as

$$\mathrm{Y} = \sum_i F(\mathrm{x})\,\Psi_i(\mathrm{x}) \tag{4.8}$$

As the vibration mode generated from each transfer path cannot be measured by sensors directly, a data-driven method was adopted to isolate multi-modal data from vibration data and establish the model for each modal data. A GMM was used to establish models for the RC data. The GMM consisted of multiple Gaussian

components, and different components represented different modal data generated via different transfer paths under specific vibration excitations. Multi-modal data can be regarded as linear combinations of Gaussian components.

The Gaussian mixture model can be expressed as

$$p\left(\mathbf{x}\right) = \sum_{k=1}^{K} \pi_k p\left(\mathbf{x} \mid \mu_k, \Sigma_k\right) \tag{4.9}$$

where $\mathbf{x}$ is the compressed signals consisting of M points. $p\left(\mathbf{x} \mid \mu_k, \Sigma_k\right)$ is one component of the Gaussian mixture model parameterized by (μ_k, Σ_k), π_k is the prior input of $\mathbf{x}$, and K is the total number of Gaussian components. In this chapter, a novel physical interpretation is proposed for GMM. Each $F(x)\Psi_i(x)$ assumed to follow Gaussian distribution. The $\mathrm{p}(\mathbf{x} \mid \mu_k, \Sigma_k)$ and π_k denote a transfer path characterizing the fault excitation and its corresponding weight, respectively [95].

The parameters of the GMM can represent the vibration conditions powerfully and thus were used to construct the state-space. The dimensions of the parameters (π, μ, Σ) are $K \times 1$, $M \times 1$ and $M \times M$, respectively. As Σ is a symmetric matrix, the lower or the upper diagonal of the matrix can be converted into a vector. The conversion of the covariance matrix $\hat{\Sigma}$ can be denoted as

$$\begin{aligned} \hat{\Sigma} = &\left[\Sigma\left(1,1\right), \Sigma\left(2,1\right), \Sigma\left(2,2\right), \cdots,\right. \\ &\left.\Sigma\left(M,1\right), \Sigma\left(M,2\right), \cdots \Sigma\left(M,M\right)\right] \end{aligned} \tag{4.10}$$

Each component's parameters can be concatenated into a vector and the vectors from all GMM components were constructed into a matrix v. In other words, a concatenated vector was considered a channel of v, and each channel can represent a type of modal data generated via a transfer path. The input can be expressed as

$$\mathrm{v} = \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \pi_K \\ \mu_1 & \mu_2 & \cdots & \mu_K \\ \hat{\Sigma}_1 & \hat{\Sigma}_2 & \cdots & \hat{\Sigma}_K \end{bmatrix} \tag{4.11}$$

where $\hat{\Sigma}_K$ is the conversion of the covariance matrix of the K-th GMM component.

LDA was adopted to calculate the GMM parameters used to construct the state-space. A K-dimensional binary vector was introduced with a particular element equal to 1 and the rest 0, namely the vector was subjected to the following condition:

$$\begin{cases} z_k & \in \{0,1\} \\ \sum_k z_k & = 1 \end{cases} \tag{4.12}$$

such that

$$p\left(z_k = 1\right) = \pi_k$$

$$s.t. \quad 0 \leq \pi_k \leq 1 \text{ and } \sum_{k=1}^{\mathrm{K}} \pi_{\mathrm{k}} = 1 \tag{4.13}$$

Traditionally, the parameters of GMM are estimated using an EM algorithm [112]. This method tends to cause overfitting, so LDA was used to calculate the GMM parameters. The label of z_k were assumed to follow multinomial distribution with Q realization.

$$p\left(z_i \mid \xi_i\right) = \frac{\mathrm{Q!}}{\prod_{j=1}^{\mathrm{K}} \left(z_{ij}\right)!} \prod_{j=1}^{\mathrm{K}} \left(\xi_{ij}\right)^{z_{ij}} \tag{4.14}$$

where $\xi_i = (\xi_{i1}, \xi_{i2}, \ldots, \xi_{ij})$ denotes the parameter of the distribution, which is subjected to the constraints that $\xi_{ij} \geq 0$ and $\sum_{j=1}^{k} \xi_{ij} = 1$. To calculate the probability of each possible ξ_{ij} , Dirichlet adopted to establish the model for ξ_{ij}. The probabilities are defined as

$$p\left(\xi_i \mid \alpha_i\right) = \frac{\Gamma\left(\sum_{\mathrm{j=1}}^{\mathrm{K}} \alpha_{ij}\right)}{\prod_{j=1}^{K} \Gamma\left(\alpha_{ij}\right)} \prod_{j=1}^{\mathrm{K}} \left(\xi_{ij}\right)^{(\alpha_{ij}-1)} \tag{4.15}$$

where $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{iK})$ is the parameter vector of Dirichlet distribution for ξ_i. α_{ij} is a non-negative value, and $\Gamma(\cdot)$ represents the Gamma function. The

probability of the label is computed by

$$p\left(z_i \mid \alpha_i\right) = \int_0^1 p\left(z_i \mid \xi_i\right) p\left(\xi_i \mid \alpha_i\right) \mathrm{d}\xi_i \tag{4.16}$$

By substituting Eq.(4.14) and Eq.(4.15) into Eq.(4.16), the distribution of the label parameterized by α_i is obtained by

$$p\left(z_i \mid \alpha_i\right) = \frac{\mathrm{Q!}}{\prod_{j=1}^{\mathrm{K}} \left(z_{ij}\right)!} \frac{\Gamma\left(\sum_{j=1}^{\mathrm{K}} \alpha_{ij}\right)}{\Gamma\left(\sum_{j=1}^{\mathrm{K}} \left(\alpha_{ij} + \mathrm{z}_{ij}\right)\right)} \prod_{j=1}^{\mathrm{K}} \frac{\Gamma\left(\alpha_{ij} + z_{ij}\right)}{\Gamma\left(\alpha_{ij}\right)} \tag{4.17}$$

Substituting the label in Eq.(9) into Eq.(14) with one realization (Q=1) yields the prior distribution π_{ij} as

$$\pi_{ij} = p\left(z_{ij} = 1 \mid \alpha_i\right) = \frac{\alpha_{ij}}{\sum_{k=1}^{K} \alpha_{ik}} \tag{4.18}$$

The log-likelihood function can be expressed as

$$L\left(\mathbf{x} \mid \alpha, \mu, \Sigma\right) = \sum_{i=1}^{N} \log \left\{ \sum_{j=1}^{\mathrm{K}} \frac{\alpha_{ij}}{\sum_{k=1}^{K} \alpha_{ik}} p\left(\mathbf{x} \mid \mu_j, \Sigma_j\right) \right\} \tag{4.19}$$

where N indicates the number of samples in $\mathbf{x}$. The parameter (α, μ, Σ) of the model could then be calculated using EM. This calculation process can avoid overfitting and enhance the generalization ability of the method, as the parameters of GMM are selected from LDA rather than setting a fixed value.

4.4 Experimental verification and analysis

4.4.1 Data description

Data used to validate the proposed method were collected from industrial plants. When a fault occurs in the RC, the vibration is transferred along a particular transfer

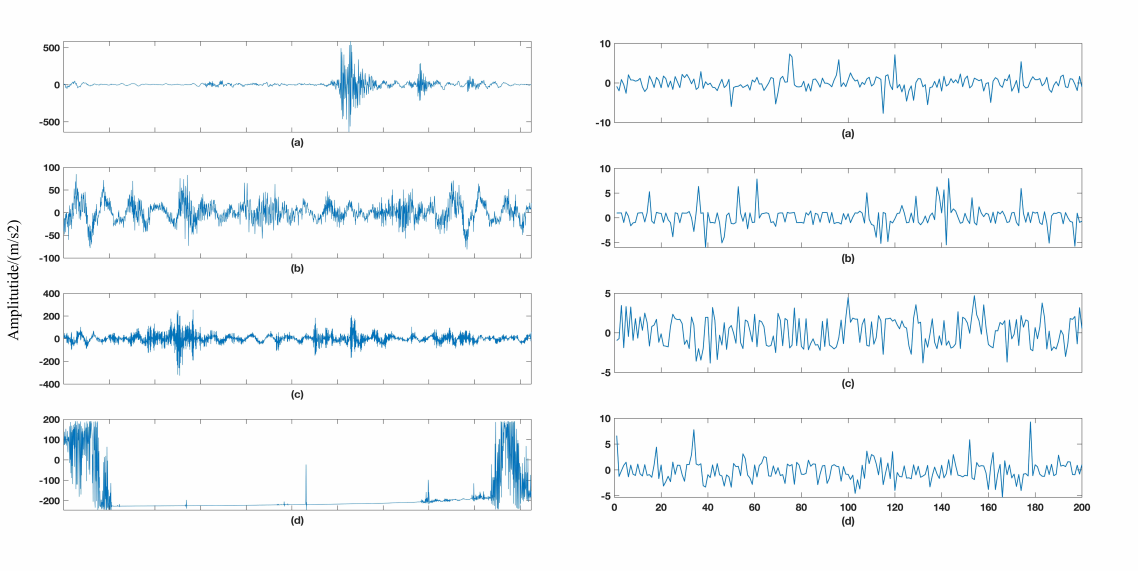


Figure 4.4 : Raw signals and compressed signals via sparse filtering

path in the body of the RC and measured by the cylinder accelerometers. Four types of data were collected from industrial plants. Random selection of 800 data sets of each fault was used to train the proposed method and 400 were used to test the method. The sampling frequency was 12.8kHz. The rotating speed under the four operating conditions were 333, 375, 300 and 333 respectively. A sample contains points of one circle of shaft rotation, ensuring signals covered information of a complete movement.

4.4.2 Fault diagnosis by the proposed method

The raw signals were compressed using sparse filtering to decrease the computational cost of fault diagnosis. The input dimension of the raw data is listed in table 3.2 and the output dimension was 200. The compression results of each operating condition are displayed on the right in figure 4.4. They represent fault F1, F2, F3 and F4, respectively.

Next, the compression results were input into MI-CDBN to learn the features.

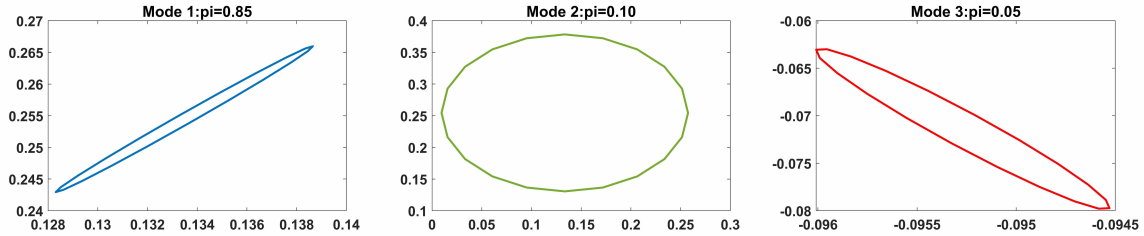


Figure 4.5 : Isolated mode data

The mode isolation results of LDA-GMM are shown in figure 4.5, where each sub-figure denotes a component of LDA-GMM. The pi or π, the ellipse centre and the direction of the ellipse along the axis indicate the prior, means, and direction of covariance of the compressed signal, respectively. Then, the state-space was constructed by the parameters of LDA-GMM as equation 4.10 and fed into CDBN to calculate the unsupervised features. The MI-CDBN included three CRBMs and one fully connected layer. The input consisted of three channels with the input size of 20301 for each channel. The epochs of each CRBM were 30 and the size of weight matrix in each CRBM was 6. Currently, there exists no systematic method that calculates the parameters of CDBN or any other networks of deep learning [82]. This chapter achieves an optimal result by trial and error, and the parameters of MI-CDBN are listed in table 4.1.

Principal component analysis was adopted to visualize the features in figure 4.6, where blue, red, yellow and green points represent the features of the four faults, F1, F2, F3 and F4, respectively. The first three principal components (PC) were displayed. It can be seen from figure 4.6 that different classes of unsupervised features learned by the proposed method were separated clearly and the features of the same class were distributed tightly. The MI-CDBN can effectively extract representative features from the constructed state-space. Finally, the features were input into a multi-class logistic regression to identify faults. The accuracies are

Table 4.1 : Parameter setting of the MI-CDBN

Parameter	setting
input dimension	20301
the number of channel	3
α of LDA	13
the number of GMM components	3
the number of CRBMs	3
learning rate	0.01/0.005/0.0005
size of weight matrix	6/6/6
the number of weight matrix in CDBN	80/50/15
epoch	30
pooling scale	2

shown in table 4.2. The confusion matrix in figure 4.7 displays the detail of fault diagnosis accuracy. The yellow squares display the rate of correct classification, and the green squares show the rate of misclassification. The horizontal axis is the actual fault, and the vertical axis is the predicted fault.

4.4.3 Parameter analysis

Two parameters are explored in this subsection, the output length of data compression and the value α of LDA. It should be noted that the experimental results are an average of 20 trials to reduce the influence of randomness in the following calculations.

The parameters of sparse filtering were adjusted to ensure the majority of useful information was preserved with less distortion than the original signal [49]. The fluctuation of accuracy and computing time with various compressed data lengths

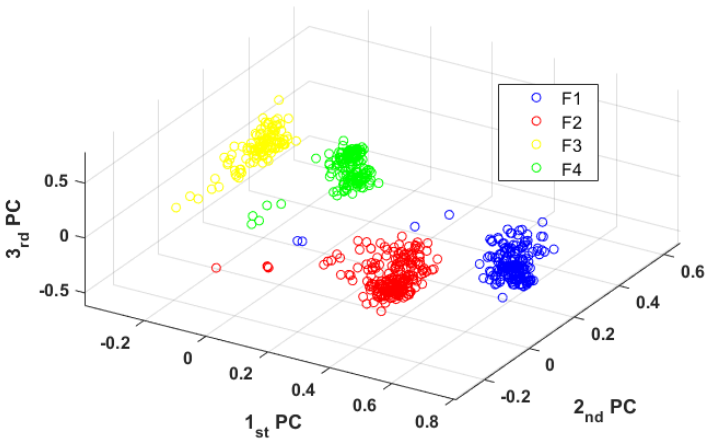


Figure 4.6 : Features calculated by MI-CDBN

are illustrated in figure 4.8. The figure suggests that with an increase of data length, the classification accuracies increase and peak at a data length of 200. Meanwhile, the computing cost increased consistently. Thus, a data length at 200 is a suitable selection.

Additionally, the value selection of α is explored in the application of RC fault diagnosis. The accuracies obtained on different values of α are shown in figure 4.9. The figure shows that when α (alpha) equals to 13, the model achieves the best diagnostic performance with an overall accuracy reaching 89.3%.

4.4.4 Performance evaluation and comparison

The proposed method was compared with several other competitive methods in the field of RC fault diagnosis in the following aspects.

Comparison of data compression methods

Currently, the most commonly used methods listed in table 4.2 are compared with the SF, which was used in the proposed method.

The input data without compression achieved a poor performance compared

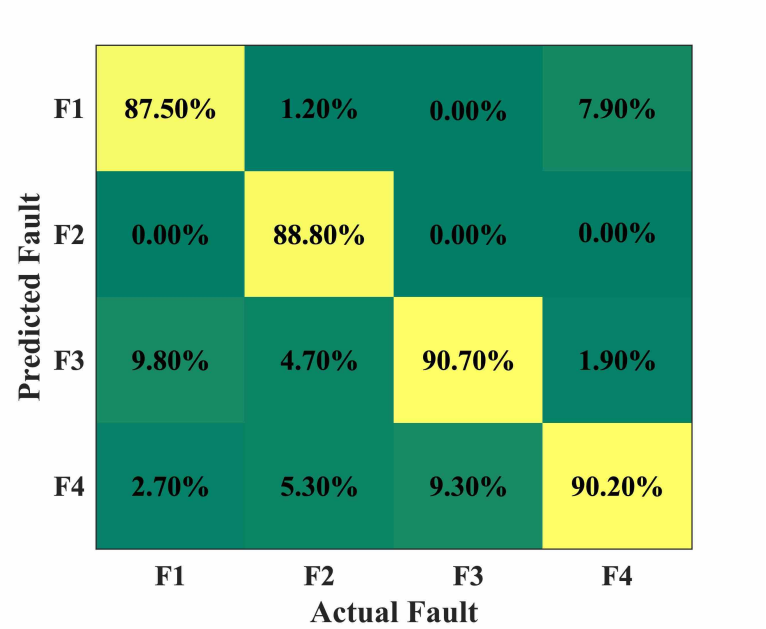


Figure 4.7 : The accuracy of fault diagnosis

Table 4.2 : Comparison of data compression methods

Method	Output dimension	Average Accuracy/%	Time/s
Raw data+MI-CDBN	-	85.4	2782
AE[82]+MI-CDBN	180	87.3	376
CS[81]+MI-CDBN	230	88.6	565
SF+MI-CDBN	**200**	**89.3**	**362**

with the others, as it contained more redundant information that disturbs the RC fault diagnosis. With data compression techniques applied, the accuracies increased overall. Compared with AE and CS, SF performed better. Only one parameter of SF required tuning, which can save substantial computing resource. The results reveal that the data compression method can extract highly-representative information from raw data effectively.

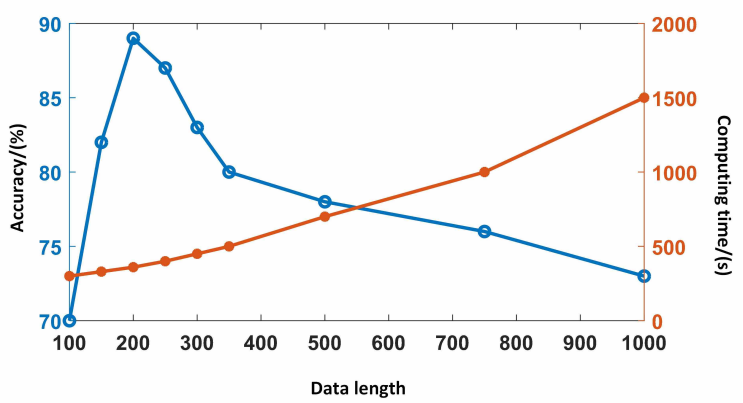


Figure 4.8 : Fault diagnosis performance with various compressed data lengths

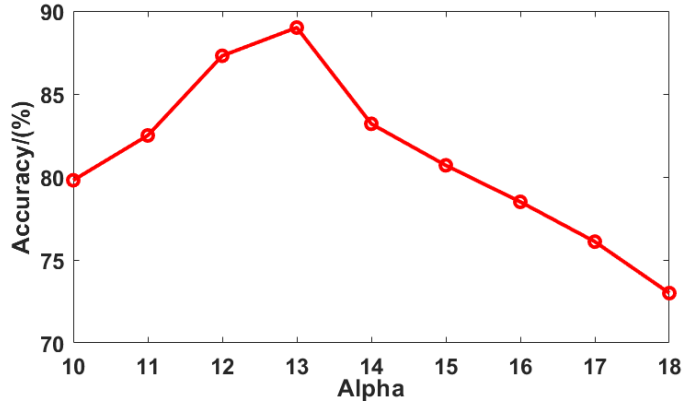


Figure 4.9 : Accuracy of fault diagnosis with various α

Comparison of deep learning methods

The performance of deep learning methods were compared and are listed in table 4.3. The raw signals were compressed by sparse filtering and then input into different deep learning methods for comparison. The parameters of different methods are listed as follows.

DBN: The DBN included four layers with the number of nodes in each layer set to 200-100-80-10. The epoch was 30 and the learning rate was set as 0.1.

CNN: The CNN consisted of 3 convolutional layers, 3 pooling layers and 1 fully

Table 4.3 : Comparison of deep learning methods

Method	Accuracy/(%)				Standard
	F1	F2	F3	F4	deviation
SF+DBN[23]	79.3	81.9	80.6	85.6	0.9769
SF+CNN[37]	80.6	83.8	84.8	84.8	0.8635
SF+CDBN[48]	82.7	84.0	86.6	85.9	0.6637
SF+GMM-CDBN	85.1	86.4	87.8	88.3	0.5046
SF+MI-CDBN	**87.5**	**88.8**	**90.7**	**90.2**	**0.4288**

connected layer. The dimension of the input layer was 200*50 with a weight matrix size of 6*6, a pooling ratio of 2, and an epoch number of 30.

Convolutional deep belief networks (CDBN): The CDBN included three CRBMs and the input dimension was 200. The size of weight matrix in each layer was 6, the pooling ratio was 2, and the epoch number was 30.

GMM-CDBN: The number of GMM components was 3 and the remaining parameters were the same as MI-CDBN. The GMM function is available on Matlab.

The proposed method achieved the best performance among all the methods with an average accuracy of 89.3%. The CDBN method reached accuracy rates of up to 82.7%, 84.0%, 86.6% and 85.9% on the four types of data, which were higher than DBN and CNN as it combined the advantages of the two methods. To improve the standard CDBN's disadvantages of transfer path analysis and mode isolation, the GMM-CDBN was compared and found to obtain a higher accuracy than CDBN. This is because GMM-CDBN can refine more robust local information from raw signals by isolating multi-modal data. GMM-CDBN and MI-CDBN achieved an average accuracy of 86.9% and 89.3%, respectively. This indicates that LDA-GMM performs better at RC fault diagnosis. LDA established a model for all

Table 4.4 : Comparison of the state-of-the-art methods

Method	Accuracy/(%)				Standard
	F1	F2	F3	F4	deviation
TD+SVM	58.8	60.1	55.2	65.1	1.7122
FD+SVM	62.1	66.4	59.5	65.9	1.4903
TD + ANN	62.8	63.4	58.9	67.4	1.6017
FD + ANN	65.7	66.4	60.9	69.2	1.2202
The proposed method	**87.5**	**88.8**	**90.7**	**90.2**	**0.4288**

possible GMM parameters and selected GMM parameters in terms of the calculated distribution of LDA rather than settling with a single selection obtained by EM. This could enhance the generalization of the method by alleviating the overfitting and further improve the fault diagnosis performance. It should be noted from table 4.3 that the proposed method sacrificed some computational efficiency to obtain a higher accuracy, as it takes more time to do the calculation. An optimized design of the input of the network would reduce the computational burden, which will be a future research topic.

Comparison of the state-of-the-art methods

The comparison results between the proposed method and other state-of-the-art methods are listed in table 4.4.

In table 4.4, the extracted TD-based methods (time-domain features) and FD-based methods (frequency-domain features) refer to [2, 116, 125]. Seven statistic features were extracted from the raw signals and frequency-domain signals. These methods were compared with the proposed method, and the results suggest that traditional methods have lower accuracy in fault diagnosis. Despite completing

the fault diagnosis in less time, they had difficulty reflecting the complex running conditions of RCs comprehensively. Presumably, this is due to traditional methods using only several statistic features. Moreover, SVM and ANN, as conventional models, are ineffectual in extracting more useful information from signals and thus cannot guarantee good classification performance. These results confirm that the MI-CDBN can extract more confident features from the input for RC fault diagnosis.

4.5 Conclusion

A novel MI-CDBN has been proposed for fault diagnosis of a reciprocating compressor from the perspective of transfer path analysis and mode isolation.

First, sparse filtering was applied to compress data into lower dimensions. This method could effectively refine highly-representative information from raw signals and increase the computational efficiency.

Secondly, MI-CDBN was used to isolate multi-modal data generated via various transfer paths of fault excitation, as well as to calculate unsupervised learning features. In the network, the parameters of the GMM model were calculated using LDA and transformed into state-space to represent RC operating conditions. The enhanced accuracy confirms that the MI-CDBN can isolate the modal data effectively, and the LDA-calculated parameters can construct robust state-space for RC fault diagnosis.

Thirdly, CDBN was used to extract features by unsupervised learning without the assistance of prior knowledge. Thus, this method can eliminate the limitation of experience and guarantee a desirable performance of fault diagnosis.

Finally, the three-stage method was validated using RC data acquired from industrial plants. The results demonstrated that the proposed method could effectively diagnose RC faults with better performance.

Chapter 5

Intelligent fault diagnosis using an optimized convolutional deep belief network

5.1 Introduction

The CDBN is advantageous in RC fault diagnosis*. The traditional CDBN consists of an input layer, an output layer, and a stack of hidden and maxout pooling layers [44]. Specifically, a variety of filters make a convolution with one unit, and the maximum of all the convolution results obtained via various filters is regarded as the pooling output. While maintaining the desirable properties of data, the max pooling results could make relatively larger generalization errors [29]. To mitigate this problem, probabilistic pooling was proposed as a tradeoff between the generalization ability and the desirable properties [88]. In this method, multiple filters were applied to a single input and then pooling results were calculated by sampling from convolution results that follow multinomial distribution instead of using the maximum. While this method enhanced the performance of the standard CDBN, a fixedly-set parameter of multinomial distribution may not be suitable to pool the data acquired under various operating conditions. Furthermore, once an inappropriate value is selected, it can reduce the generalization ability of the CDBN and deteriorate the performance of fault diagnosis.

To overcome the problem and enhance the generalization ability of the CDBN

*This chapter has been published as: Zhang Y., Ji, J.C. and Ma, B., 2020. Reciprocating compressor fault diagnosis using an optimized convolutional deep belief network. *Journal of Vibration and Control*, p.1077546319900115.

with standard probabilistic output, an optimized CDBN was proposed. The parameter of multinomial distribution was calculated using LDA instead of being set a constant. LDA can select different reasonable parameter values for various operating conditions and thus reduce the generalization error. Currently, LDA has been used in many fields to establish models for discrete data. Wang et al. applied LDA to text classification [98]. Terenin et al. employed LDA in natural language processing and validated the methods by several public corporals [93]. LDA can establish models for discrete data without any assumption of data distribution and can adjust its distribution shape self-adaptively with parameter variation [32]. In view of these advantages, the LDA was incorporated in the approach to optimize the probabilistic output of the pooling layer in the CDBN.

In the context of big data, the computing efficiency is of great significance for intelligent fault diagnosis. Typically, the acquired condition monitoring data is longer in length at a high sampling frequency, which will increase the computational burden and require more time for data processing. This could delay the response time of fault diagnosis and fail to conduct early detection of severe faults. Therefore, SF was used to compress signals in this chapter.

Once unsupervised features have been learned by the proposed CDBN, a softmax regression classifier was connected with the CDBN as the last layer to implement intelligent fault identification. The softmax regression classifier makes the best use of its excellent classification capabilities evidenced in various fields, such as text classification [40], bearing fault identification [92], image classification [55], and disease diagnosis [113, 58].

Motivated by the above discussions, this chapter proposes an optimized method for intelligent RC fault diagnosis based on a CDBN with an optimized probabilistic output. A novel framework based on the CDBN was adopted for the RC fault

diagnosis. First, sparse filtering was used to compress raw signals into compact time series and obtain the most representative information. Then, the unsupervised features were learned by the CDBN with an optimized probabilistic output for pooling. Finally, a softmax regression classifier was connected as the last layer of the CDBN to identify fault types of the RC. Data from industrial plants were collected to validate the method. The contribution of this part can be summarized as follows.

i. This framework adopted the technique of data compression to enhance computational efficiency, which is crucial for a highly-efficient RC fault diagnosis, especially concerning big data.

ii. Adopting a deep learning method allowed the feature extraction of RC fault diagnosis to be independent of expertise, which can reduce errors made by human experience.

iii. The optimized probabilistic output was proposed for the pooling layer to improve the generalization ability of the CDBN. Instead of being set a fixed value, the parameter of multinomial distribution was obtained from the established LDA models.

The rest of the chapter is arranged as follows. Section 5.2 gives a brief preliminary of the proposed method. Section 5.3 describes the proposed method in detail. Then industrial data were analyzed to validate the proposed method in Section 5.4, and conclusions are drawn in Section 5.5.

5.2 Preliminary

5.2.1 Probabilistic out

Given an input unit $\mathbf{v}$ which can be the compressed signal or the output of previous CRBM, the activation of a probabilistic output unit was obtained by convolution

with k filters which is given by

$$z_i = \mathbf{w}_i \mathbf{v} + b_i, i \in [1, k] \tag{5.1}$$

where k is the total number of filters of the weight matrix. Then all the possible outputs of probabilistic pooling h_{pbout} would be

$$h_{pbout}(\mathbf{v}) : [z_1, ..., z_k] \tag{5.2}$$

A Boltzmann distribution was introduced over k linear mappings and then unit $h_{pbout}(\mathbf{v})$ was sampled in terms of probabilities calculated by the distribution. The Boltzmann distribution is defined as

$$p_i = \frac{e^{\lambda z_i}}{\sum_{j=1}^{k} e^{\lambda z_i}} \tag{5.3}$$

where λ is a hyperparameter dominating the variance of the distribution. The activation $h_{pbout}(\mathbf{x})$ was then sampled as

$$h_{pbout}(\mathbf{v}) = z_i,\ p(i|\xi_{ij}) \sim Multinomial\{p_1, ..., p_k\} \tag{5.4}$$

where ξ_{ij} denotes the parameter of multinomial distribution. It can be observed from equation 5.4 that the activation output was obtained by sampling from all possible unit outputs. The probabilistic output unit could preserve most properties of the maxout unit and improved the overall generalization ability. More details about the CDBN can be found in Ref [48].

5.3 The proposed method

5.3.1 Optimized convolutional deep belief network

An optimized convolutional deep belief network was proposed by optimizing conventional probabilistic pooling to improve the model generalization ability. $\mathbf{x}$ is defined as the input of the CDBN with the corresponding actual label y and the

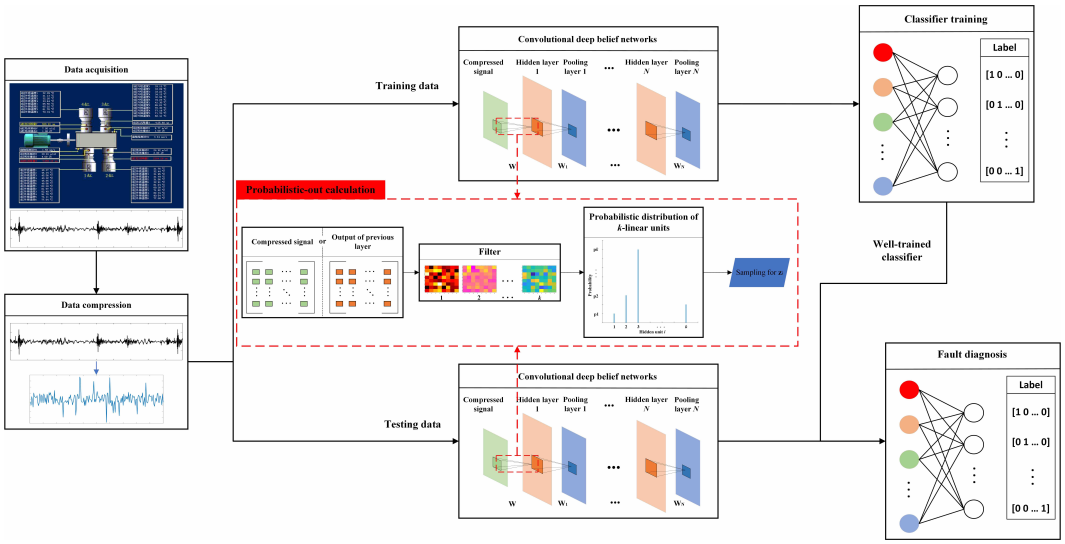


Figure 5.1 : Framework of the proposed method

predicted label $\hat{y}$, which are both C-dimensional vectors denoting the C type of the RC faults. The CDBN was constructed by stacking N CRBMs which incorporates a visible layer $\mathbf{v}$ and a hidden layer $\mathbf{h}$. $h_i^{(l)}, l \in [1, N]$ indicates the unit in the hidden layers $\mathrm{h}^{(1)}$,...,$\mathrm{h}^{(N)}$. The input was mapped sequentially into the N hidden layers to calculate the predicted result of fault types of the RC. Each unit in a layer was computed by a function $h_i^{(l)}\left(\mathbf{v}; \mathbf{w}_i^{(l)}, b_i^{(l)}\right)$, which maps its input $\mathbf{v}$ to an output $h_i^{(l)}$ using weight $\mathbf{w}_i^{(l)} and bias b_i^{(l)}$. Then the CDBN was connected to a softmax function as the final layer to identify RC faults which is expressed as [49]

$$\hat{\mathbf{y}} = softmax\left(\mathbf{W}^{N+1}\mathbf{h}^{N} + \mathbf{b}^{N+1}\right) \tag{5.5}$$

with weight $\mathbf{W}$ and bias $\mathbf{b}$. All parameters $\theta = \left\{W^{(1)}, b^{(1)}, ..., W^{(N+1)}, b^{(N+1)}, \right\}$ were then learned by minimizing the cross entropy loss between the predicted label $\hat{\mathbf{y}}$ and the actual label y .

$$\mathcal{L}\left(\hat{\mathrm{y}}, \mathrm{y}; \mathbf{x}\right) = -\sum_{i=1}^{C} y_i \log\left(\hat{y}_i\right) + \left(1 - y_i\right) \log\left(1 - \hat{y}_i\right) \tag{5.6}$$

To improve the generalization capability of the CDBN, a dropout technique was adopted to regularize the trained network[28]. The operation was achieved by

combining with the probabilistic sampling, and predefining the probabilities as $\hat{p}_0$. Then,

$$\hat{p}_i = \frac{e^{\lambda z_i}}{\hat{p}_0 . \sum_{j=1}^{k} e^{\lambda z_i}} \tag{5.7}$$

the value of $\hat{p}_0$ will be determined in the subsequent section. The probabilistic output activation function involving dropout $\hat{h}_{pbout}(\mathbf{v})$ was sampled as

$$\hat{h}_{pbout}(\mathbf{v}) = \begin{cases} 0 & \text{if } i = 0 \\ z_i & else \end{cases}$$

$$p(i|\xi_{ij}) \sim Multinomial\{\hat{p}_0, \hat{p}_1, ..., \hat{p}_k\} \tag{5.8}$$

To further increase the generalization ability of the probabilistic output in the pooling layer, the LDA was employed to establish the model of the parameter ξ_{ij} of multinomial distribution. LDA is defined as[72]

$$p(\xi_{ij} \mid \alpha_{ij}) = \frac{\Gamma\left(\sum_{\mathrm{m=1}}^{\mathrm{K}} \alpha_{ijm}\right)}{\prod_{m=1}^{K} \Gamma(\alpha_{ijm})} \prod_{m=1}^{\mathrm{K}} (\xi_{ij})^{(\alpha_{ijm}-1)} \tag{5.9}$$

where $\alpha_{ij} = (\alpha_{ij1}, \alpha_{ij2}, \ldots, \alpha_{ijK})$ is the parameter vector of Dirichlet distribution for ξ_{ij} . α_{ij} is a non-negative value. $\Gamma(\cdot)$ represents the Gamma function.

5.3.2 General procedure of the proposed method

A CDBN with an optimized pooling method was proposed to enhance the generalization ability. The framework of the proposed method is illustrated in figure 5.1, and the detailed procedures are explained as follows.

Step 1: Vibration signals were collected under various operating conditions from an industrial petroleum plant.

Step 2: Sparse filtering compressed raw signals into lower dimensions to increase the computational efficiency of big data.

Step 3: The compressed signals were divided into training data and testing data with each group of data selected randomly.

Step 4: Without the need for manual feature extraction, the compressed training signals were fed into the CDBN, with the optimized probabilistic output, to obtain unsupervised features of each fault type. The subfigure inside the dash-line frame illustrates the improvements on the probabilistic output. In the CDBN, the LDA was used to calculate the parameters of the multinomial distribution. Then, the multinomial distribution was adopted to establish the model for multiple units computed by filters. Finally, the output of the pooling result was obtained by sampling from the multinomial distribution.

Step 5: The learned features from the training data were utilized to train the softmax classifier, and the testing data were used to verify the performance of the proposed method.

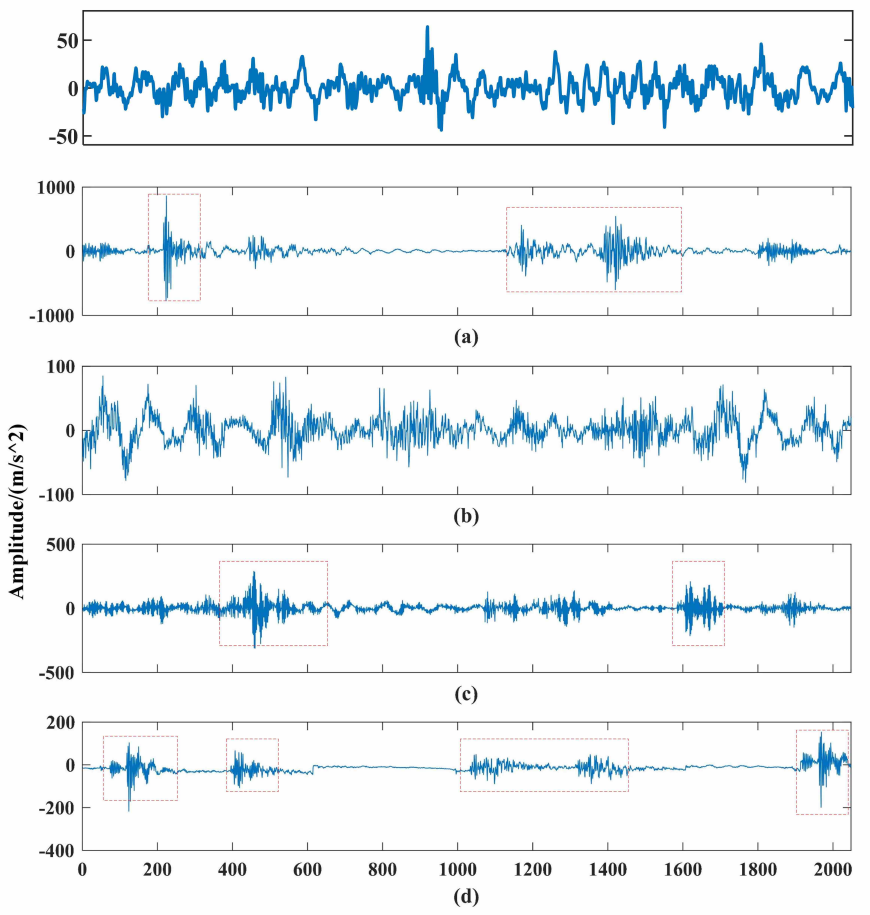
5.4 Experimental verification and analysis

5.4.1 Data description

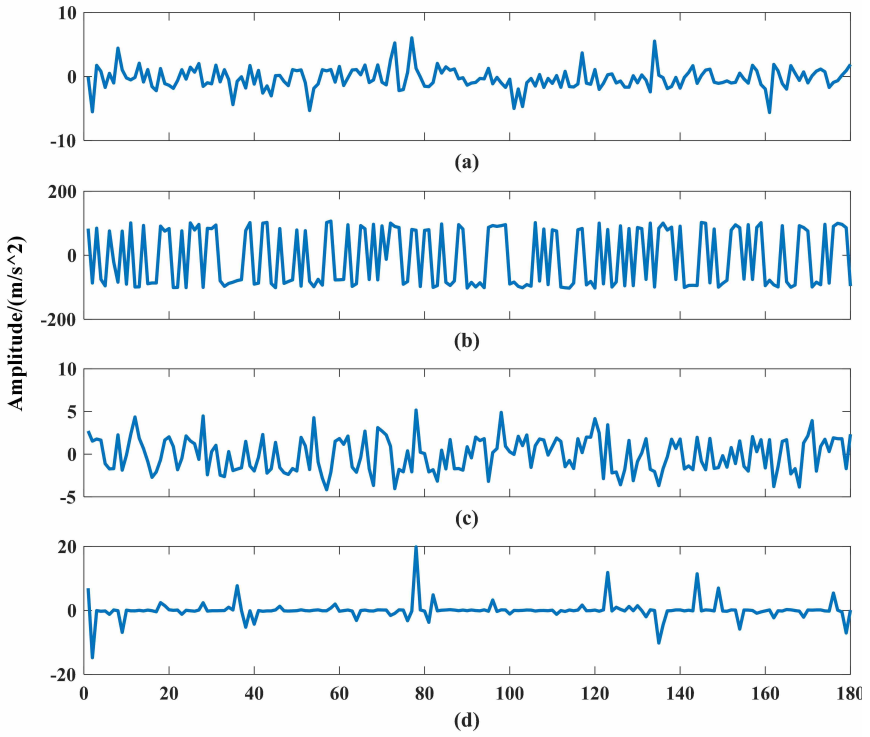
Details on the fault types are the same as in table 3.2. The four fault types are named as F1, F2, F3 and F4, respectively. The data were collected under the rotating speed of 375rpm and the sampling frequency of 12.8kHz. A set of 800 data for each group was randomly selected as training data and 400 as testing data.

5.4.2 Fault diagnosis by the proposed method

Vibration data monitoring operating conditions of the RC were collected from an industrial application to validate the proposed method. One cycle of a signal was fed into sparse filtering to obtain a higher-representative and lower-dimensional result. Figure5.2(b) displays the compression results of raw signals in figure 5.2(a). In the figure, raw signals were compressed into a dimension of 180 with redundant



(a)



(b)

Figure 5.2 : (a)Raw signals, and (b)compressed signals

information removed. The four compression results can represent each running condition with compact time series by refining the most useful information.

After raw signals were compressed with SF, the compressed results were input into the proposed CDBN to calculate features by unsupervised learning. PCA calculated the principal components of the learned features for visualization, and their distribution is shown in figure 5.3 where purple, red, green and blue points indicate

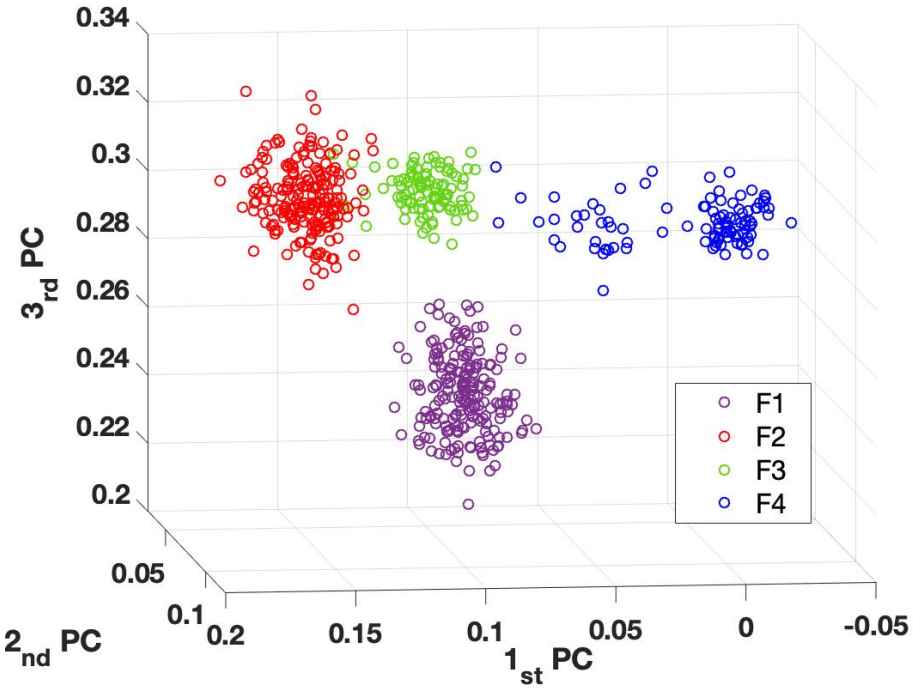


Figure 5.3 : Principal components of unsupervised feature

the learned features of F1, F2, F3, and F4, respectively. Different fault features are clearly distinguished, whereas, features of the same fault are clustered tightly. The results suggest the effectiveness of the features on characterizing various operating conditions of the RC.

After the proposed CDBN calculated unsupervised features, they were fed into a softmax classifier to accomplish intelligent fault diagnosis. The confusion matrix in figure 6.11 shows the accuracy of fault diagnosis. The vertical axis represents the

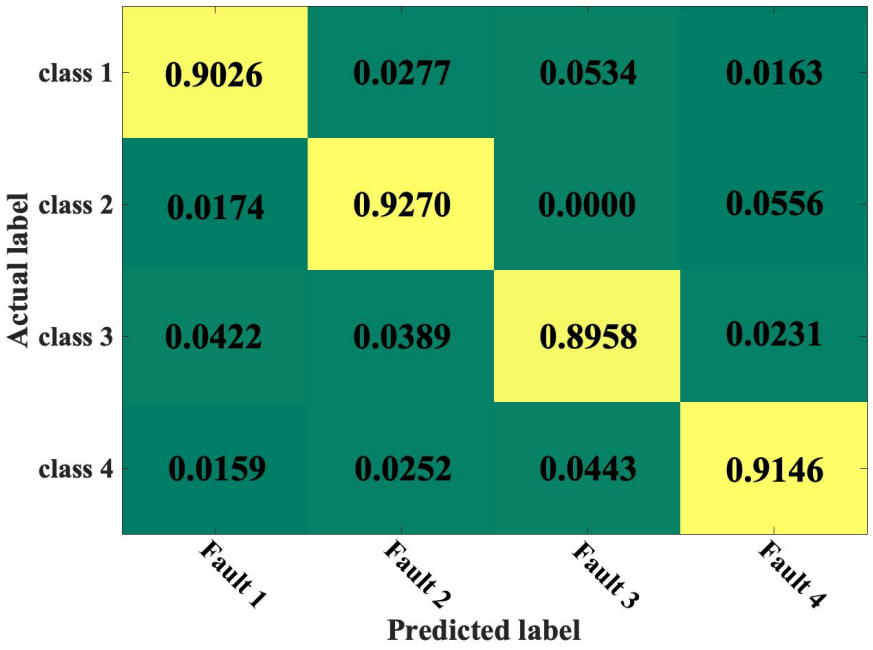


Figure 5.4 : Confusion matrix of diagnosis accuracy

actual fault labels of raw signals, and the horizontal axis denotes the misclassified fault type labels. The green square shows the detailed misclassification result of each data type, and the yellow square indicates the accuracies of the proposed method. The diagnosis accuracies of the four typical faults reach up to 90.26%, 92.70%, 89.58% and 91.46%, respectively, confirming that the proposed method can conduct RC fault diagnosis more effectively.

5.4.3 Parameter analysis

Several parameters exist in the proposed method, and the selection of these parameters are discussed in this section.

Firstly, the generalization ability of the proposed method was investigated.

The comparison of generalization error is shown in figure 5.5where the blue and red curves represent the errors of the CDBN with the standard and optimized prob-

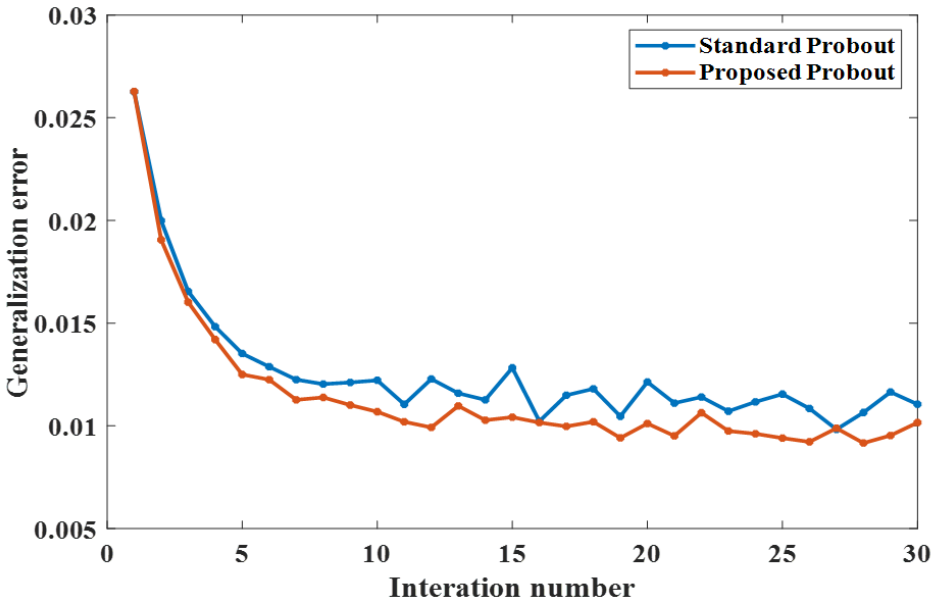


Figure 5.5 : Comparison of generalization error

abilistic output, respectively. The results show the average values over 20 trials. Both errors converged rapidly with an increase of iteration number. However, the optimized probabilistic output achieved an overall lower generalization error, suggesting that it could enhance the generalization capacity of the CDBN effectively by obtaining the parameter of the multinomial distribution from LDA.

Secondly, the relationship between accuracy and hyper-parameter λ was investigated and is demonstrated in figure 5.6. The accuracy continuously increased when λ was within the range of (0, 0.1), and then begun to decrease when λ exceeded 0.1. Therefore, λ was set to 0.1 to ensure the lowest generalization error and the best performance in the application of RC fault diagnosis.

Finally, the relationship between the sampling probability of dropout and the generalization error was explored. The results are shown in figure 5.7. It can be seen from the figure that with increased sampling probability, the accuracy increased and reached the peak value at a sampling probability of 0.7, after which accuracy decreased. The reason for this variation could be that a low sampling probability for

dropout could lose substantial useful information, and a higher probability could lead to overfitting, which could degrade CDBN performance. Therefore, the sampling probability was set to 0.7 in the RC fault diagnosis.

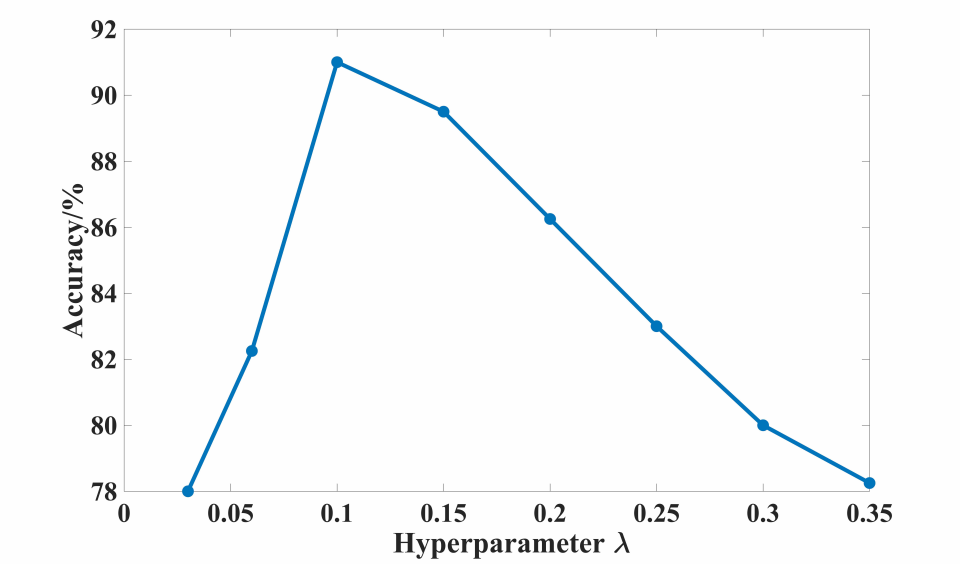


Figure 5.6 : The relationship between generalization error and λ

5.4.4 Method comparison and performance evaluation

This subsection makes a comparison of the method and evaluates its performance in RC fault diagnosis.

Table 5.1 : Comparison of pooling methods

Method	Averaging accuracy/(%)	Standard deviation
Standard Prob-out	90.75	0.8928
Proposed method	91.00	0.8523

Comparison of pooling methods

The performance of pooling methods in the RC fault diagnosis are listed in table 5.1.

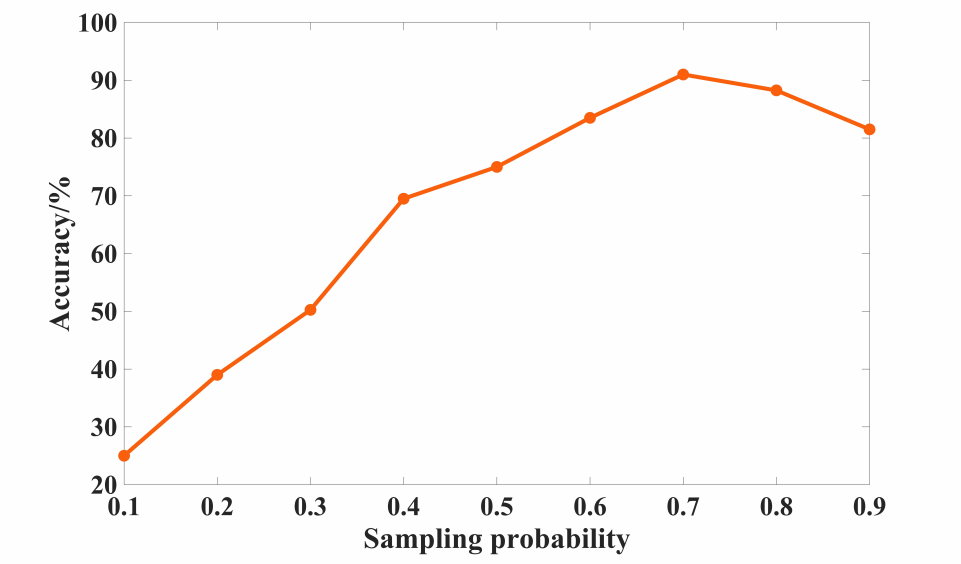


Figure 5.7 : The relationship between the accuracy and $\hat{p}_0$

Compared to standard probabilistic pooling, the proposed method yielded a higher accuracy rate up to 91.00% and a lower standard deviation. By adopting an optimized probabilistic output as the pooling result, the proposed method could enhance the generalization capacity of the CDBN effectively and ensure diagnosis results yield a more significant improvement.

5.5 Conclusions

An optimized CDBN was employed for the RC fault diagnosis to eliminate the dependence on prior knowledge and guarantee the extraction of desirable features. An optimized probabilistic output was proposed for the pooling layer to improve the generalization ability of the CDBN. A sparse filtering technique was first adopted to compress raw signals into low-dimensional ones and obtain the most representative information. Then compressed signals were fed as the input of the proposed CDBN to learn the features representing different operating conditions. Finally, a softmax regression classifier was connected as the last layer of CDBN to identify fault categories.

Vibration data characterizing various RC operating conditions were collected from a petroleum industry to validate the proposed method. The diagnosis and comparison results demonstrated the effectiveness of the proposed method on fault diagnosis.

Chapter 6

Intelligent fault diagnosis using a novel ensemble empirical mode decomposition-convolutional deep belief network

6.1 Introduction

Due to the sophisticated production environment of the RC, the acquired signals are usually contaminated by background noise*. The noise disturbances can negatively affect the fault diagnosis results and undermine the fault diagnosis performance. Therefore, it is essential to construct an auto-denoising network to eliminate the noise existing in signals. The ensemble empirical mode decomposition (EEMD) was adopted to decompose signals into various IMFs, which can be considered as intrinsic vibrations of machinery that indicate the operating conditions and noise that is generated by fault-irrelevant vibrations. Then the correlation coefficient (CC) was employed to select intrinsic functions associated with fault-related vibrations and remove noise for signal reconstruction and denoising [63].

To obtain more comprehensive information on the operating conditions, multiple types of sensors were attached to the RC to collect the displacement and vibration data of various parts. After signals were denoised and the unsupervised features were learned by the auto-denoising network, multi-source of information was fused based on the probabilistic committee machine (PCM) scheme. Conventional committee

*This chapter has been published as: Zhang, Y., Ji, J.C. and Ma, B., 2020. Fault diagnosis of reciprocating compressor using a novel ensemble empirical mode decomposition-convolutional deep belief network. *Measurement*, vol.156, p.107619.

members use nonlinear logistic regression, such as artificial neural network, support vector regression [87] and autoregression [16], as the probabilistic models to calculate the probability of each committee member. As these methods make a prediction based on point estimation, this may result in low robustness of fault diagnosis results. Therefore, the Gaussian process classifier (GPC) was adopted as the probabilistic model in the committee machine [8]. This method makes a probabilistic prediction based on a group of points instead of a single point, and thus enhances the robustness of fault diagnosis results.

In this chapter, a novel auto-denoising network called EEMD-CDBN and a framework fusing multi-source information was proposed to implement the RC fault diagnosis. The main contributions of the chapter can be summarized as follows.

i. The EEMD-CDBN was proposed to denoise the RC vibration signals and to enhance the robustness of learned features. The proposed network can eliminate the noise in signals self-adaptively and can learn the unsupervised features efficiently without prior knowledge.

ii. A novel framework fusing multi-source information was proposed for the RC fault diagnosis. This framework can fuse the features of signals from multiple types of sensors, which can demonstrate the RC working condition more comprehensively.

iii. The PCM was adopted to fuse features with more specific weights allocated to each committee component instead of equally distributed weights. A feature fusion method was proposed to calculate the weight distribution in terms of historical data of the RC. This can enhance the accuracy of the RC fault diagnosis.

The rest of the chapter is arranged as follows. Section 6.2 presents the proposed

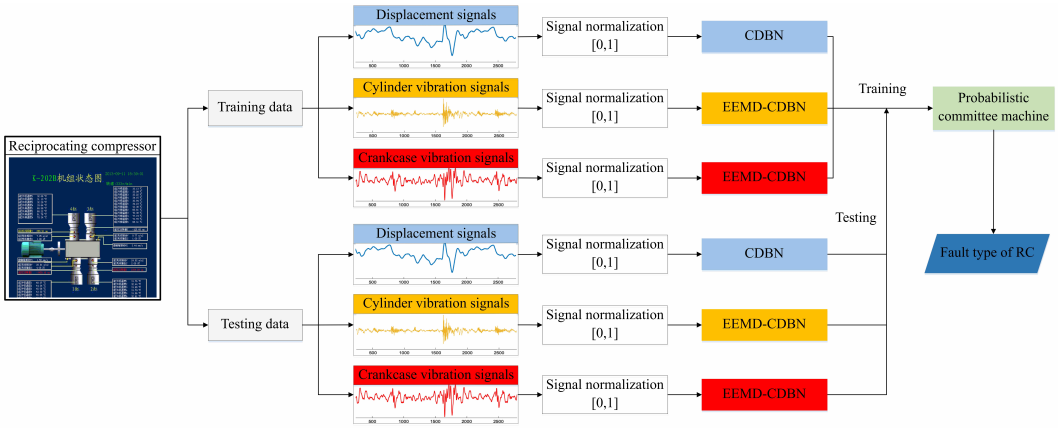


Figure 6.1 : Framework of the proposed method

method and related techniques. Section 6.3 shows the case study and experimental information. Section 6.4 evaluates the proposed method. Section 6.5 gives a discussion of the proposed method and concludes the chapter.

6.2 Proposed method

6.2.1 General framework

Aiming to combine more comprehensive information from the RC sensors and remove the background noise in raw signals, a novel ensemble empirical mode decomposition-convolutional deep belief network was proposed, as illustrated in figure 6.1. The detailed procedure of the general framework is described as follows.

i. Signals were collected from the sensors of the RC, including displacement signal, cylinder vibration signal and crankcase vibration signal.

ii. These signals were divided into training data and testing data randomly.

iii. Signals were normalized into the range of [0,1].

iv. Displacement signals were input into the CDBN, and the vibration signals of the cylinder and crankcase were input into the EEMD-CDBN for denoising and unsupervised feature learning.

v. Training data were applied to train the probabilistic committee machine.

vi. Testing data were input into the well-trained probabilistic committee machine for the fault diagnosis and method evaluations.

6.2.2 Ensemble empirical mode decomposition-convolutional deep belief network

To eliminate the noise contaminated in the original signals and enhance the robustness of fault diagnosis results, this chapter proposes a novel auto-denoising network, namely the EEMD-CDBN, as demonstrated in figure 6.2. The detailed procedure is listed below.

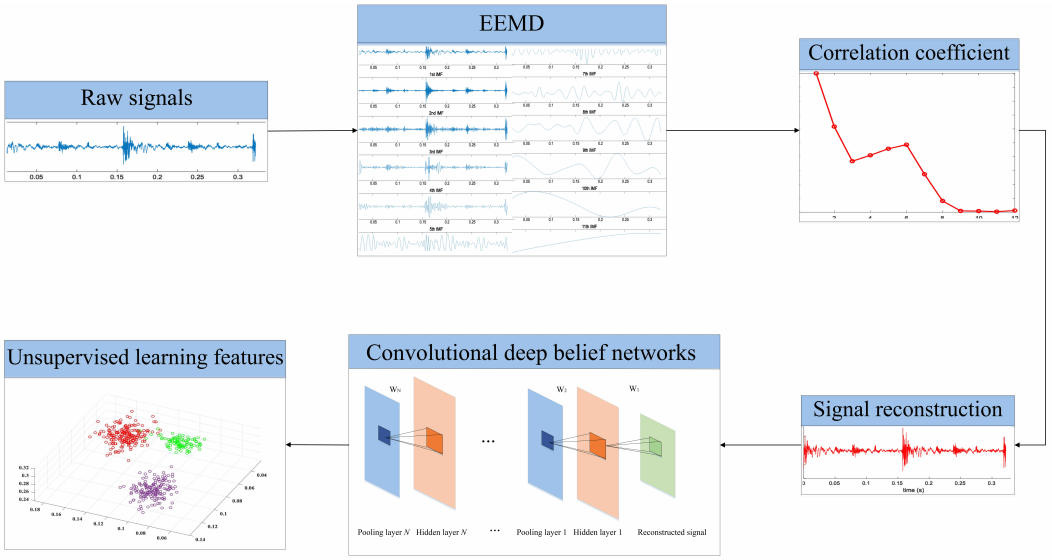


Figure 6.2 : Schematic of the EEMD-CDBN method

i. Raw signals were decomposed into a series of IMFs by the EEMD.

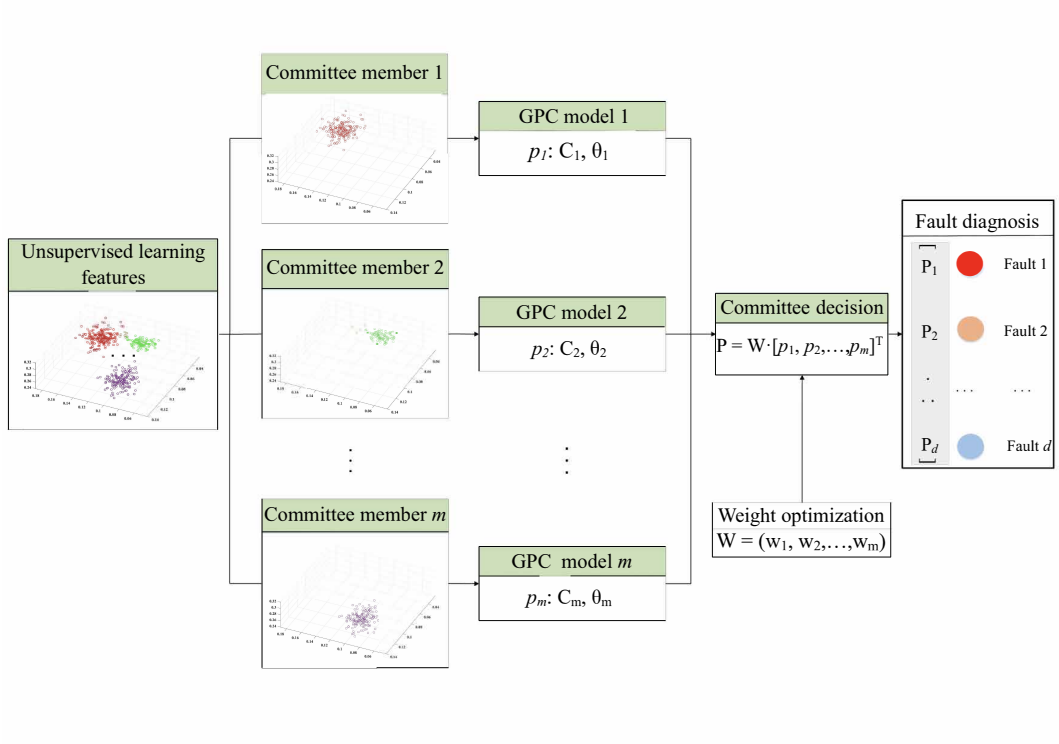


Figure 6.3 : Probabilistic committee machine

ii. CC was used to calculate the correlation between raw signals and each IMF.

iii. The IMFs of a large CC were selected to reconstruct signals.

Ensemble empirical mode decomposition

The raw signal of the RC x_r can be considered as the superposition of vibration from the fault component x and environment noise n_r

$$x_r = x + n_r \tag{6.1}$$

where x contains most of the information that can reflect the working condition. The EEMD was adopted to isolate the signal x from n_r, and the correlation coefficient was used to determine the IMFs that include the most powerful information.

The EEMD assumes that the decomposed results of each IMF were the ensemble means of multiple trials. With the properties of the EMD, the EEMD was conducted

as follows.

i. The ensemble number of the IMF was set as E

ii. A series of noise signals $(n_1, n_2, \cdots, n_E)$ and $e = 1$ were initialized.

iii. A white noise sequence n_e $(e = 1, 2, \cdots, E)$ was added to the original signal x. Then the new signal with the added noise could be expressed as

$$x_e = x + n_e \tag{6.2}$$

where x_e, x and n_e represent the generated signal, the original signal and the added noise in the e-th trial, respectively.

iv. The EMD decomposed the signal x_e into $imf_{i,e}$ $(i = 1, \cdots, I)$, where I is the total number of IMF.

v. If $e \leq E$, Step 2 and Step 3 were repeated.

vi. The ensemble means of the i-th IMF from E trials as the final decomposition results of the i-th IMF of EEMD were computed, namely,

$$\begin{gathered} \overline{imf_i} = \frac{1}{E} \sum_{e=1}^{E} imf_{i.e} \\ i = 1, 2, \cdots, I, \ e = 1, 2, \cdots, E \end{gathered} \tag{6.3}$$

where $imf_{i.e}$ is the i-th IMF decomposed from the e-th noise sequence x_e.

More details about the EEMD can be found in Ref [107].

When conducting the EEMD, the signals were decomposed into IMFs by the EEMD. These IMFs include both high-representative information associated with the RC operating condition and redundant information that is not necessary for the fault diagnosis. The IMFs with the most useful information were used for signal reconstruction. The reconstructed signal can be considered to have high-representative information for fault diagnosis.

Correlation coefficient and signal reconstruction

The correlation coefficient (CC) was adopted to select the most representative of the IMFs for signal reconstruction. It gives a correlation degree for each IMF with raw signals. A high CC value corresponds to an IMF with the most representative information, while a small CC value indicates that the IMF consists of more redundant information. The equation for calculating the correlation coefficient can be expressed as

$$CC = \frac{\sum_{i=1}^{m} \left(x_i\left(t\right) - \overline{x}\right)\left(imf_i\left(t\right) - \overline{imf_i}\right)}{\sqrt{\sum_{i=1}^{m} \left(x_i\left(t\right) - \overline{x}\right)^2}\sqrt{\sum_{i=1}^{m} \left(imf_i\left(t\right) - \overline{imf}\right)^2}} \tag{6.4}$$

where $x_i\left(t\right)$ and $imf_i\left(t\right)$ represent the i-th point in the raw signal $x\left(t\right)$ and the IMF of the i-th signal, respectively. m is the total number of points in the signal $x\left(t\right)$ and $imf\left(t\right)$. $\overline{x}$ and $\overline{imf_i}$ are the mean values of $x\left(t\right)$ and $imf\left(t\right)$, respectively. To eliminate noise and improve the signal to noise ratio, the IMFs of significantly low CC were neglected and the rest used for signal reconstruction. The reconstruction process can be written as

$$\hat{x} = imf_a + imf_b + \cdots + imf_n \tag{6.5}$$

where $\hat{x}$ denotes the reconstructed signal. imf_a, imf_b and imf_n are the selected IMFs with high CC values.

6.2.3 Probabilistic committee machine

In the process of RC fault diagnosis, low robust results are likely to occur when using single-source information. To fuse more information from various sensors and obtain more reliable fault diagnosis results, a novel scheme based on the PCM was proposed. The detailed procedure is discussed below.

i. The unsupervised features of signals from m sources were input into the

corresponding GPC of each committee machine to calculate the probability p_i $(i = 1, ..., m)$ that each fault occurs under the source of information.

ii. The proposed weight calculation method was adopted to determine the weight W of each committee member.

iii. The W was used to make a combined decision from each committee member.

iv. The fault corresponding to the largest probability in $\mathrm{P} = [\mathrm{P}_1, \mathrm{P}_2, ..., \mathrm{P}_d]$ was considered as the final diagnosis result.

In the PCM, the GPC was used to establish models for features extracted from different sources of signals and to calculate the probability that each fault occurs. The unsupervised learning data and their corresponding labels are expressed as

$$\begin{gathered} C : \{(\mathbf{x}_i, y_i) \,|i = 1, 2, \cdots, m\} \\ \mathbf{x}_i \in R^n,\ y_i \in [0, 1] \end{gathered} \tag{6.6}$$

where $\mathbf{x}_i$ are the unsupervised learning features from the CDBN and EEMD-CDBN, and y_i are the corresponding labels. The GPC result under the given data C can be found by

$$p\left(y \,|C, \theta, \mathbf{x}\right) = \int p\left(y \,|g(\mathbf{x})\right) p\left(g(\mathbf{x}) \,|C, \theta, \mathbf{x}\right) dg \tag{6.7}$$

where g indicates the latent function for mapping. θ represents the hyper-parameter of distribution p.

As the posterior distribution $p\left(y \,|C, \theta, \mathbf{x}\right)$ cannot be computed analytically, the Guassisan approximation was applied to simplify the calculation. The simplification with y belonging to Class 1 can be expressed as

$$\tilde{p}\left(y = 1 \,|C, \theta, \mathbf{x}\right) = \int \sigma\left(g(\mathbf{x})\right) \mathcal{N}\left(g(\mathbf{x}) \,|\mu, \Sigma^2\right) dg \tag{6.8}$$

where $\mathcal{N}\left(g(\mathbf{x}) \,|\mu, \Sigma^2\right)$ denotes the Gaussian process function with mean μ and covariance Σ^2. σ represents the sigmoid function. More information on parameters can be found in Ref [105].

An optimized scheme was proposed to fuse the committee decision and obtain the probability that each fault occurs under multi-source features [127]. This can be written as

$$P_j = \frac{\sum_{i=1}^{m} \mathrm{w}_i p_{ij}}{\sum_{i=1}^{m} \mathrm{w}_i} \quad i = 1, 2, \cdots, m \; j = 1, 2, ..., d \tag{6.9}$$

where w_i is the weight corresponding to the i-th committee machine. m is the number of committee machine. d is the number of fault type.

w_i was determined by historical fault diagnosis results. It is defined as

$$\mathrm{w}_i = \frac{n_{i-c}}{n_i} \tag{6.10}$$

where n_{i-c} indicates the number of features from the i-th sensor categorized into the correct fault class, and n_i is the total number of samples. Then the predicted label of fault diagnosis was determined as L_p. L_p is one label of $(1, 2, \cdots, d)$ with

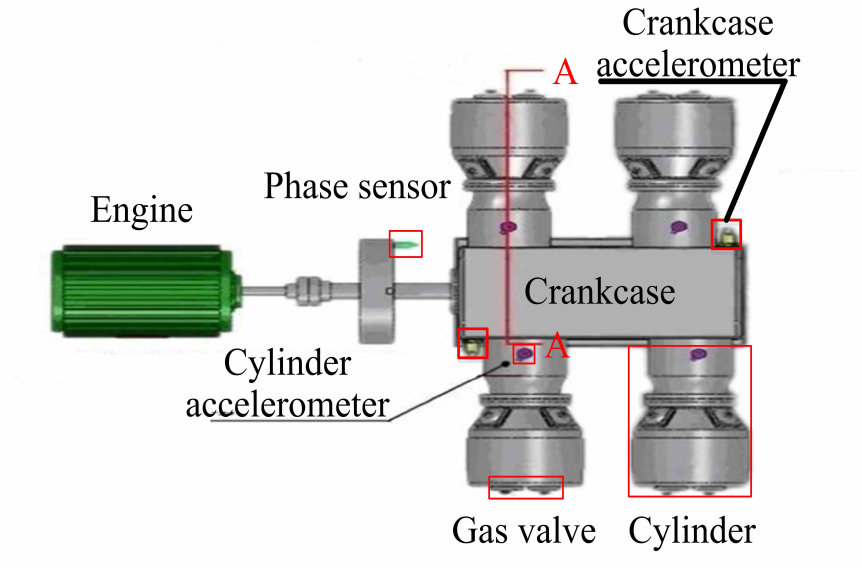
$$P_{L_p} = max\left(P_j\right) \; (j = 1, 2, ..., d) \tag{6.11}$$

6.3 Experimental verification and analysis

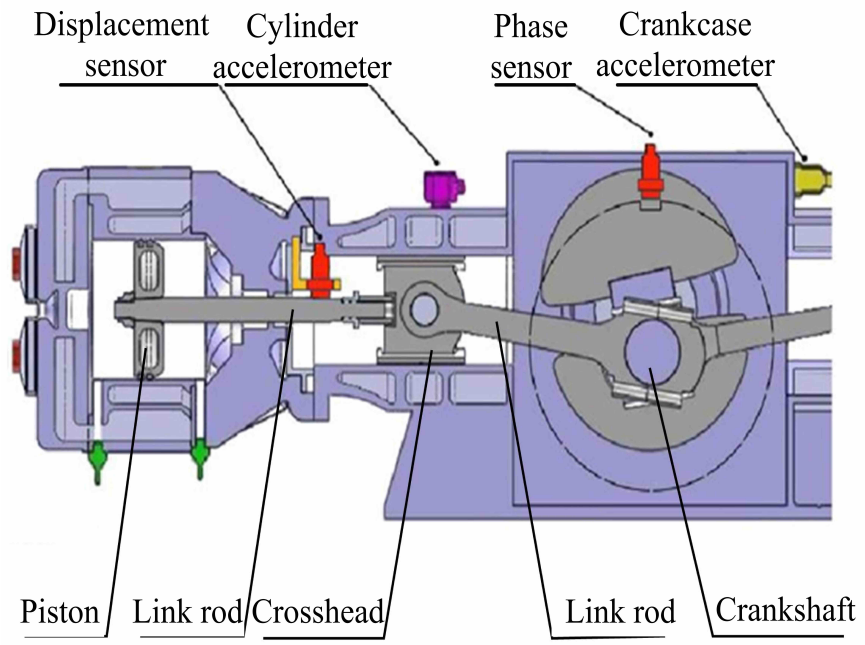
6.3.1 Data description

The RC fault data were collected from an industrial plant to validate the proposed method.

Figure 6.4 shows the typical structure of an RC with four cylinders and the sensor layout. An RC consists of three parts, an engine, cylinders and a crankcase. The cross-section schematic of the A-A in figure 6.4b illustrates the sensor layout and the inner structure of the cylinder and crankcase. The engine drives the crankshaft rotation, which allows the right link rod to swing around the crosshead and make further reciprocating movements of the crosshead and piston. The sensors installed in the RC are displayed in Figure 6.5. Figure 6.5a exhibits the phase sensor located in the flywheel, used to monitor the rotating speed of the RC. Figure 6.5b shows

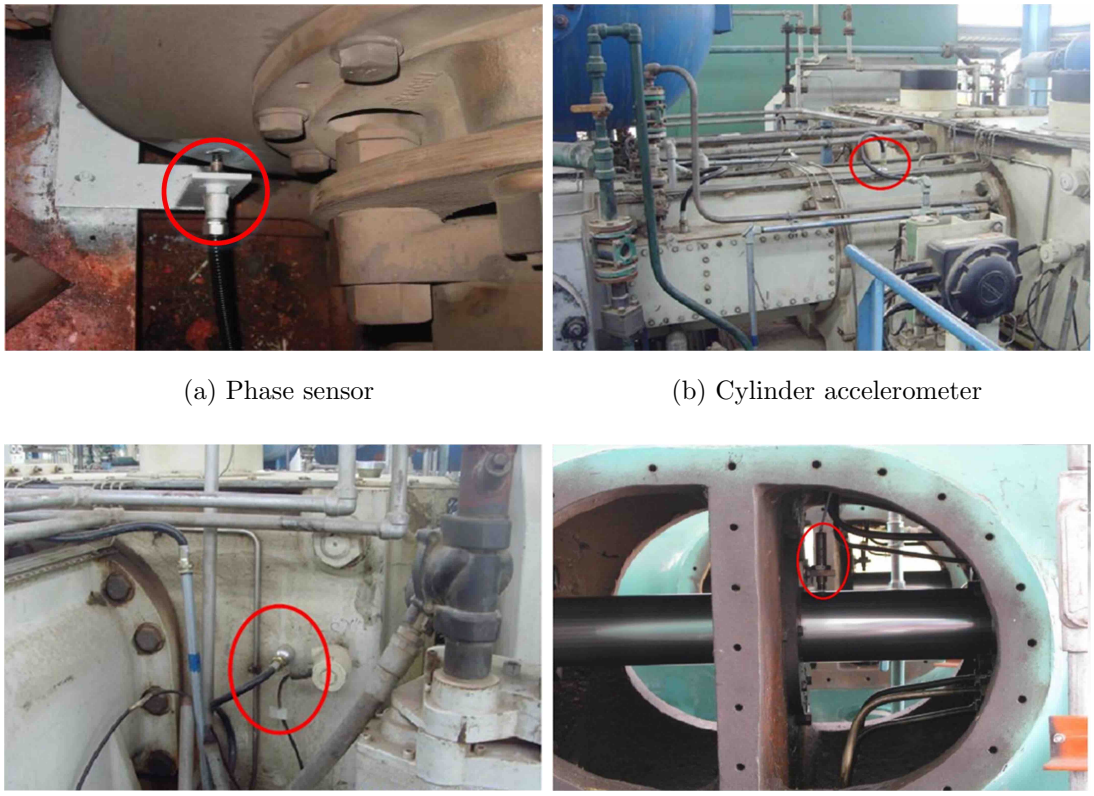


(a) Schematic diagram



(b) A-A cross-section

Figure 6.4 : Schematic diagram of the RC and the sensor layout on a cross-section



(a) Phase sensor

(b) Cylinder accelerometer

(c) Crankcase accelerometer

(d) Displacement sensor

Figure 6.5 : Locations of Sensors

a cylinder accelerometer mounted on the crosshead, which is used to measure the cylinder vibration. The sensor collects acceleration signals in the unit of m/s^2 and has a sensitivity of 98m/s^2/V. Figure 6.5c displays the crankcase accelerometer attached to the crankcase to monitor the operating condition of the crankcase. Data from a cylinder accelerometer were transformed into velocity data by the embedded integral module. The sensor collects velocity signals in m/s units and has a sensitivity of 98m/s/V. The displacement sensor on the link rod in figure 6.5d was used to measure the movement of the piston by calculating the distance to the link rod. It is an eddy-current sensor with a sensitivity of 254μm/V. The displacement signal is in the unit μm. Figure 6.6 illustrates the signals acquired from a displacement sensor, an accelerometer on the cylinder and an accelerometer on the crankcase, re-

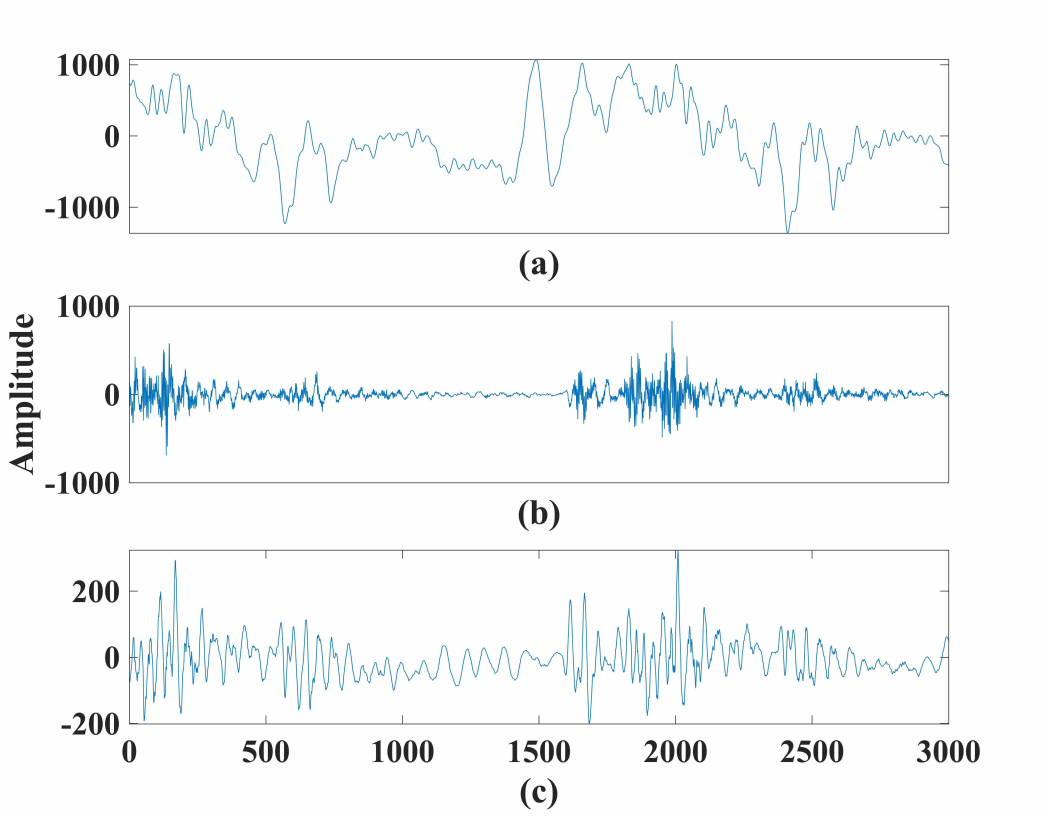


Figure 6.6 : Examples of raw signals: (a) displacement of piston rod, (b) vibration of cylinder, (c) vibration of crankcase.

spectively. These time-series data demonstrate the displacement of the piston rod, the vibration condition of the cylinder and the vibration condition of the crankcase.

Figure 3.4 displays four typical RC faults; these are gas valve failure, piston breaking, cylinder scraping and bearing shell wear, respectively. Details on the fault types are listed in table 3.2. The data were collected under the rotating speed of 375rpm and the sampling frequency of 12.8kHz. A set of 800 data for each group was selected randomly as training data and 400 as testing data.

6.3.2 Feature learning by the EEMD-CDBN

In the case study, signals were first denoised by the EEMD. The input dimension of each signal was set to 2560 to include sufficient information from one cycle

of crankshaft rotation. Seventy sets of noise signals were generated to calculate the ensemble results of the IMFs. The decomposed results of the raw signals are displayed in figure 6.7 and figure 6.8. In these two figures, the original signals were decomposed into multiple IMFs, and each IMF indicates a type of vibration modes in the raw signals. In other words, the raw signals can be regarded as a collection of RC vibration modes, which consist of noise components and fault-related vibrations. Next, the CC was employed to determine the correlations between the raw signals and IMFs. The IMFs of high CCs were then used to reconstruct to a denoised signal. The denoising results of the cylinder signal and crankcase signal are illustrated in figure 6.9 and figure 6.10, respectively. Figures 6.9a and 6.10a show the CCs of each IMF with raw signals. The IMF0 represents the raw signal. A high CC means the IMF is closely related to the raw signals and includes more representative information that is associated with the operating conditions. IMF1-IMF5 in figure 6.9a and IMF2-IMF6 in figure 6.10a have significantly larger CC values than the remaining IMFs; thus, they were selected to reconstruct the cylinder vibration signal and the crankcase vibration signal, respectively. The reconstructed signals are displayed in figures 6.9b and 6.10b. The blue lines are the raw signals, and the red lines are the reconstructed signals. It can be seen from the two signals in each figure that the EEMD has removed considerable noise and redundant information.

After signals were denoised and reconstructed, they were input into the CDBN to learn the unsupervised features. The operation of the CDBN refers to Ref [48]. The parameter setting of the EEMD-CDBN is listed in table 6.1. The EEMD-CDBN consists of three CRBMs, and each CRBM was configured with different parameters.

After the unsupervised features of different sources of signals were obtained via the EEMD-CDBN, they were input into the GPC to calculate the probability that each fault occurs. Then the results were fused by the PCM for fault identifications. Twenty trials were conducted to reduce the randomness of the data and enhance

Table 6.1 : Parameters of the EEMD-CDBN

Parameter		Value
The length of the raw data		2560
Ensemble number of raw signals		70
First CRBM	The number of weight matrix	24
	The dimension of weight matrix	6
Second CRBM	The number of weight matrix	16
	The dimension of weight matrix	6
Third CRBM	The number of weight matrix	9
	The dimension of weight matrix	6
Learning rate		0.1
The number of iteration		30
Pooling ratio		2

the robustness of the results. The diagnosis results are displayed in the confusion matrix in figure 6.11.

Table 6.2 : Comparison of denoising methods

Method	Accuracy/%	Standard deviation
WT-CDBN	90.93	0.8846
EMD-CDBN	91.29	0.8729
Proposed method	91.89	0.8689

Table 6.3 : Comparison of deep learning methods

Method	Accuracy/%	Standard deviation
EEMD-DBN	91.33	0.8765
EEMD-CNN	91.58	0.8712
Proposed method	91.89	0.8689

Table 6.4 : Comparison with the PCM-based and conventional methods

Method	Weight			Accuracy/%
	w_1	w_2	w_3	
Equal weight+ Gauss	1	1	1	89.23
PCM+SVM	-	-	-	91.33
PCM+ANN	-	-	-	91.66
Proposed method	0.9100	0.7023	0.2521	91.89

6.4 Evaluation of the proposed method

To illustrate its superiority, the proposed method was compared with several state-of-the-art methods in the following aspects. Fault diagnosis accuracy and standard deviation were selected for the performance evaluation.

6.4.1 Comparison of denoising methods

Multiple methods are used for signal denoising. The performance of several denoising methods are listed and compared in table 6.2, which shows that the proposed EEMD-CDBN method achieved the best performance with the highest accuracy rate of up to 91.89% and the lowest standard deviation of 0.8689. Compared with the WT-CDBN, the EMD-CDBN and the EEMD-CDBN achieved higher accura-

cies. This can be attributed to the advantage of EMD and EEMD to decompose signals in terms of the signal itself instead of the base functions selected using prior experience. In other words, these methods are independent of prior knowledge and can decompose signals self-adaptively. However, the EEMD-CDBN performs better than the EMD-CDBN, as EEMD alleviates mode mixing and end effect problems by averaging ensemble noise-added signals. These above-mentioned reasons account for the improved performance of the EEMD-CDBN in signal denoising and thus enhance the robustness of the traditional CDBN.

6.4.2 Comparison of deep learning methods

The CNN and DBN, as two widely used deep learning methods, were chosen to make a comparison with the CDBN. Raw signals were input into the DBN, the CNN and the proposed EEMD-CDBN to obtain the fault diagnosis results. The configuration of the CNN and the DBN are listed as follows.

DBN: The DBN includes one input layer and four hidden layers. The number of nodes in each layer was set as 2560-1000-600-100-30. The number of nodes was adjusted in terms of fault diagnosis performance. The number of iteration and the learning rate was set to 50 and 0.1, respectively.

CNN: The CNN includes one input layer, three convolutional layers and three pooling layers. The size of the input was 51*51. The number of weight matrix in the three convolutional layers were 10, 10 and 6, respectively. The pooling ratio, learning rate and iteration number were 2, 0.1 and 50, respectively.

Table 6.3 demonstrates that the proposed method achieved the best performance. Compared with the DBN-based and CNN-based method, the CDBN could accomplish fault diagnosis with higher accuracy, as the CDBN combines the advantages of the CNN and the DBN. It can extract more reliable information using its spatial-temporal invariance to assist fault diagnosis.

6.4.3 Effectiveness of the PCM

The effectiveness of the PCM is discussed in this subsection. Signals from three types of sensors were fused by the PCM in different weights according to their sensitivities to various faults. The fusion weight can be obtained from table 6.4 and according to equation 6.10. The fault diagnosis results with the PCM framework and the conventional fusion strategy are compared in table 6.4, from which it can be seen that the fusion method with equal weights performed poorly compared with other results. The main reason is that different sources of information can contribute differently to fault diagnosis. In other words, different sensors have different sensitivities to different faults. Additionally, the proposed method, namely the PCM framework plus Gaussian classifier, offers better performance in RC fault diagnosis, as the parameters of the Gaussian classifier were estimated based on a group of points rather than a single point estimation, as applied in the ANN and the SVM. This enhances the robustness of the estimated results and is more suitable for RC fault diagnosis.

6.5 Conclusion

This chapter proposed an auto-denoising convolutional deep belief network-based method, which uses the EEMD to denoise signals and uses a CDBN to extract features self-adaptively. The proposed method enhanced the fault diagnosis performance of the RC in two aspects.

Firstly, the proposed EEMD-CDBN can denoise the RC signal self-adaptively. The EEMD decomposed signals into the IMFs of different scales. Then, the correlation coefficient was adopted to calculate the correlation between the raw signal and each IMF. The IMFs of high CC values were used for signal reconstruction. This method can denoise signals effectively without any help of prior knowledge and thus enhanced the robustness of the fault diagnosis results.

Secondly, multi-source information was utilized to improve the performance of the RC fault diagnosis. A novel PCM-based strategy was applied to fuse information according to their various sensitivities to the RC faults. As multi-source information included more fault-related information, and the PCM can fuse features with a more reasonable weight, this proposed method improved the RC fault diagnosis performance.

Four types of fault data from an industrial plant were collected to validate the proposed method. The obtained results demonstrated that the proposed method could diagnose the RC faults with an accuracy rate of up to 91.89%. Additionally, the comparison with other methods illustrated the superiority of the proposed method in signal denoising, unsupervised feature learning and multi-source information fusion for the diagnosis of the RC faults.

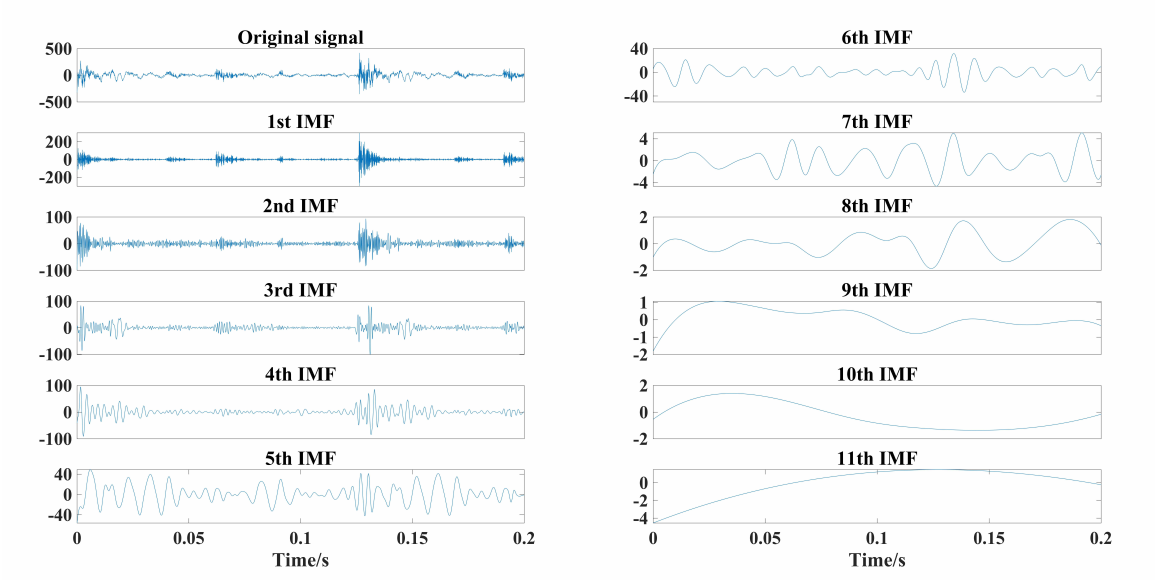


Figure 6.7 : The EEMD of cylinder signal

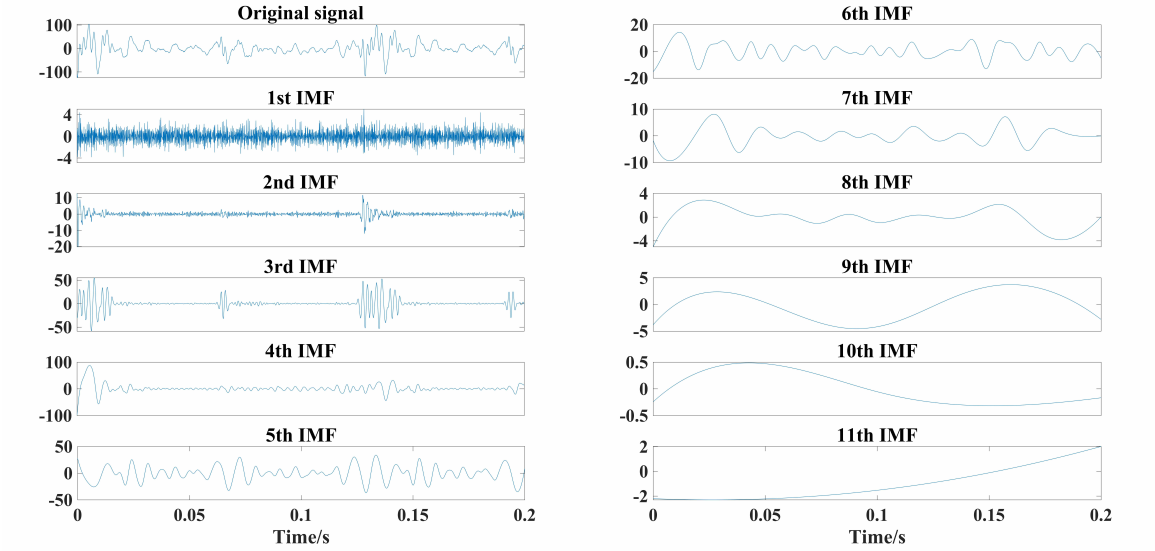
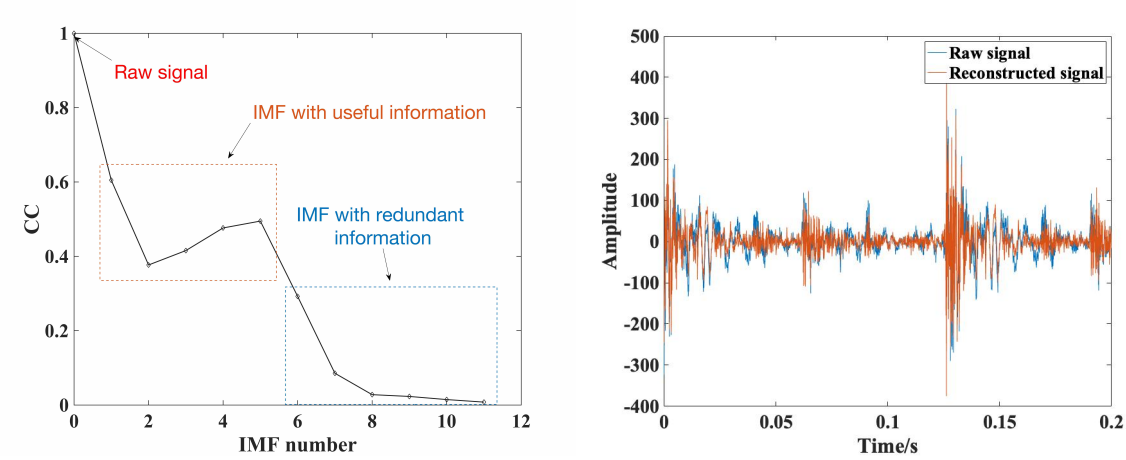
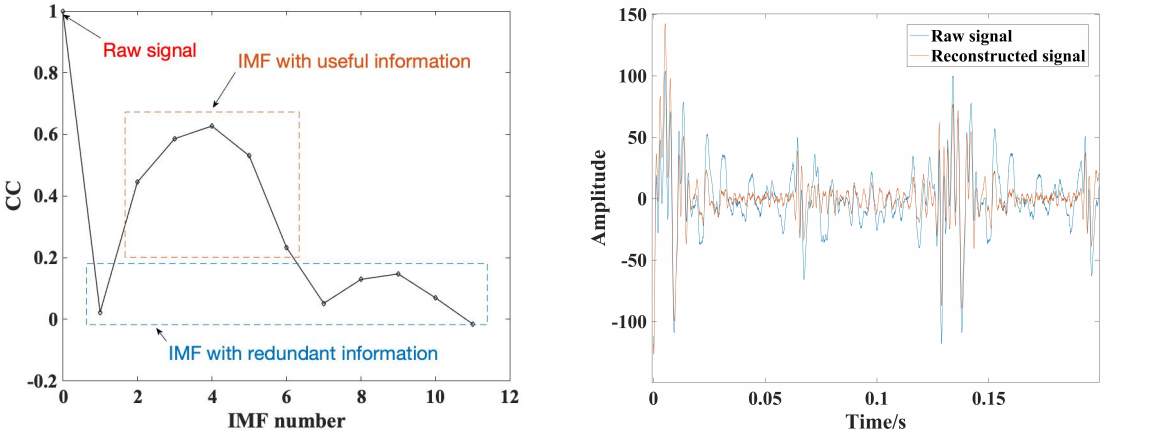


Figure 6.8 : The EEMD of crankcase signal



(a) CC of each IMF (b) Raw and reconstructed signals of cylinder

Figure 6.9 : Cylinder signal denoising



(a) CC of each IMF (b) Raw and reconstructed signals of crankcase

Figure 6.10 : Crankcase signal denoising

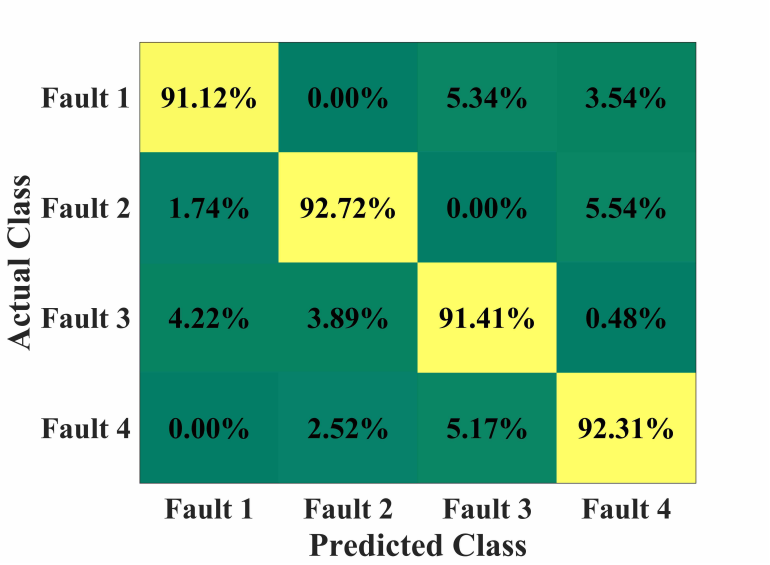


Figure 6.11 : Confusion matrix

Chapter 7

Conclusion and future work

This chapter presents the conclusions of the entire research and discusses future research opportunities.

7.1 Conclusion

Fault diagnosis of reciprocating compressors is an essential topic in the field of health condition monitoring. It can help provide a reliable environment for manufacturing. Motivated by this aim, the research was launched on RC fault diagnosis. The research progressed over four stages.

The first stage began based on traditional framework. EMD was adopted to analyze the intrinsic vibration of the RC. Then high-dimensional features were extracted from the intrinsic vibration. These features could reflect the operating conditions comprehensively. DBN was adopted to fuse features in-depth and obtain features of high-representativeness. A GM-based similarity was proposed for fault identification. The proposed method achieved better performance by retaining more nonlinear information in signals.

The second stage proposed a mode isolation-convolutional deep belief network method. To absorb the information of vibration mechanism, the proposed method isolated different modes of vibration, generated via different transfer paths, and calculated features from isolated modal data for fault diagnosis. With multi-modal data isolated, the proposed method could obtain features of high robustness, which assisted with RC fault diagnosis.

In the third stage, with the optimized convolutional deep belief network, a probabilistic pooling scheme was proposed. A value of high probability calculated from multinomial distribution was adopted as the pooling result. Next, latent Dirichlet allocation was used to calculate the parameters of multinomial distribution to enhance the generalization ability of the CDBN. Experimental data verified the effectiveness of the optimization of CDBN for RC fault diagnosis.

Finally, the fourth stage proposed an auto-denoising network. EEMD was added to CDBN for auto-denoising of the RC signal. This method could remove the noise of RC vibration signals and extract features self-adaptively. Furthermore, a scheme fusing multi-source information was developed for RC fault diagnosis. Multi-source signals were collected from different sensors installed on different parts of the RC to absorb more information from the RC for fault diagnosis. EEMD-CDBN and CDBN were used to extract features from signals of multiple sources after denoising. Then, a probabilistic committee machine was used to determine the fault according to the features of different sensors. The experimental results showed the effectiveness and superiority of the proposed method.

This research advanced the development of RC fault diagnosis by using deep learning methods. The framework of RC fault diagnosis based on deep learning can conduct feature extraction self-adaptively, which allows fault diagnosis to rely on prior knowledge minimally. In addition, this research provided more perspectives on how to extract local information to conduct more reliable information.

7.2 Future work

The research implemented RC fault diagnosis based on a deep learning method and proposed a set of optimizations to enhance their performance in the application. Although these methods enhanced the performance of RC fault diagnosis, future research could be broadened in the following aspects.

Firstly, fault detection should be conducted prior to fault diagnosis. This could assist fault diagnosis by distinguishing abnormal data from normal data, and thus further enhance the efficiency of health condition monitoring. Mathematical models can be established for normal condition. Once the model of real-time data deviates the established model of normal conditions and the deviation exceeds the threshold, the operating condition of machinery can be determined as being abnormal.

Secondly, the fault severity assessment could be conducted to assess the severity of a fault once it has been identified. Deep learning methods can be used to establish models for fault severity assessment by regression analysis. When a fault occurs in an RC, sometimes it is not necessary to shut down the machinery immediately, e.g. the wearing fault. Too frequent shutdowns of machinery will decrease the productivity. Fault severity assessment can assess the severity of current fault and select an appropriate moment to shut down. The fault severity assessment results can avoid unnecessary shutdowns and provide more specific guidance for maintenance.

Thirdly, the support ring is one of the most vulnerable components of an RC. It is essential to assess its operating condition and avoid any faults or failure related to the support ring. The model of assessing wearing state of support ring can be established. This problem can be designed as a classification problem or a regression problem.

Fourthly, only four types of fault data were collected in this research. In the future, we could collect more types of fault data for method validation from these industrial plants.

Bibliography

[1] M. Ahmed, A. Smith, F. Gu, and A. D. Ball, "Fault diagnosis of reciprocating compressors using revelance vector machines with a genetic algorithm based on vibration data," in *2014 20th International Conference on Automation and Computing*. IEEE, 2014, pp. 164–169.

[2] M. Ahmed, M. Baqqar, F. Gu, and A. Ball, "Fault detection and diagnosis using principal component analysis of vibration data from a reciprocating compressor," in *Proceedings of 2012 UKACC international conference on control*. IEEE, 2012, pp. 461–466.

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

[4] F. AlThobiani, A. Ball *et al.*, "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, 2014.

[5] D. M. S. Arsa, G. Jati, A. J. Mantau, and I. Wasito, "Dimensionality reduction using deep belief network in big data case study: Hyperspectral image classification," in *2016 International Workshop on Big Data and Information Security (IWBIS)*. IEEE, 2016, pp. 71–76.

[6] F. Attal, Y. Amirat, A. Chibani, and S. Mohammed, "Automatic recognition

of gait phases using a multiple regression hidden markov model," *IEEE/ASME Transactions on Mechatronics*, 2018.

[7] C. Barba-González, J. García-Nieto, M. del Mar Roldán-García, I. Navas-Delgado, A. J. Nebro, and J. F. Aldana-Montes, "Bigowl: Knowledge centered big data analytics," *Expert Systems with Applications*, vol. 115, pp. 543–556, 2019.

[8] L.-P. Berczi, I. Posner, and T. D. Barfoot, "Learning to assess terrain from human demonstration using an introspective gaussian-process classifier," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 3178–3185.

[9] S. Boluki, M. S. Esfahani, X. Qian, and E. R. Dougherty, "Constructing pathway-based priors within a gaussian mixture model for bayesian regression and classification," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 16, no. 2, pp. 524–537, 2017.

[10] D. Cabrera, A. Guamán, S. Zhang, M. Cerrada, R.-V. Sánchez, J. Cevallos, J. Long, and C. Li, "Bayesian approach and time series dimensionality reduction to lstm-based model-building for fault diagnosis of a reciprocating compressor," *Neurocomputing*, vol. 380, pp. 51–66, 2020.

[11] B. Cai, Y. Liu, Q. Fan, Y. Zhang, Z. Liu, S. Yu, and R. Ji, "Multi-source information fusion based fault diagnosis of ground-source heat pump using bayesian network," *Applied energy*, vol. 114, pp. 1–9, 2014.

[12] X. Che, J. Mi, and D. Chen, "Information fusion and numerical characterization of a multi-source information system," *Knowledge-Based Systems*, vol. 145, pp. 121–133, 2018.

[13] J. Chen, Z. Li, J. Pan, G. Chen, Y. Zi, J. Yuan, B. Chen, and Z. He, "Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review," *Mechanical systems and signal processing*, vol. 70, pp. 1–35, 2016.

[14] L. Chen, G. Xu, Q. Zhang, and X. Zhang, "Learning deep representation of imbalanced scada data for fault detection of wind turbines," *Measurement*, vol. 139, pp. 370–379, 2019.

[15] X.-W. Chen and X. Lin, "Big data deep learning: challenges and perspectives," *IEEE access*, vol. 2, pp. 514–525, 2014.

[16] Y. Chen, X. Liang, and M. Zuo, "An improved singular value decomposition-based method for gear tooth crack detection and severity assessment," *Journal of Sound and Vibration*, vol. 468, p. 115068, 2020.

[17] Y. Chen, X. Liang, and M. J. Zuo, "Sparse time series modeling of the baseline vibration from a gearbox under time-varying speed condition," *Mechanical systems and signal processing*, vol. 134, p. 106342, 2019.

[18] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 7, pp. 1693–1702, 2017.

[19] C. Cheng, G. Ma, Y. Zhang, M. Sun, F. Teng, H. Ding, and Y. Yuan, "A deep learning-based remaining useful life prediction approach for bearings," *IEEE/ASME Transactions on Mechatronics*, 2020.

[20] X. Cheng, J. Mao, J. Li, H. Zhao, C. Zhou, X. Gong, and Z. Rao, "An eemd-svd-lwt algorithm for denoising a lidar signal," *Measurement*, p. 108405, 2020.

[21] X. Dai and Z. Gao, "From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 2226–2238, 2013.

[22] J. M. B. de Lázaro, A. P. Moreno, O. L. Santiago, and A. J. da Silva Neto, "Optimizing kernel methods to reduce dimensionality in fault diagnosis of industrial systems," *Computers & Industrial Engineering*, vol. 87, pp. 140–149, 2015.

[23] L. Diener, M. Janke, and T. Schultz, "Direct conversion from facial myoelectric signals to speech using deep neural networks," in *Neural Networks (IJCNN), 2015 International Joint Conference on*. IEEE, 2015, pp. 1–7.

[24] K. Feng, K. Wang, Q. Ni, M. J. Zuo, and D. Wei, "A phase angle based diagnostic scheme to planetary gear faults diagnostics under non-stationary operational conditions," *Journal of Sound and Vibration*, vol. 408, pp. 190–209, 2017.

[25] K. Feng, Z. Jiang, W. He, and B. Ma, "A recognition and novelty detection approach based on curvelet transform, nonlinear pca and svm with application to indicator diagram diagnosis," *Expert Systems with Applications*, vol. 38, no. 10, pp. 12 721–12 729, 2011.

[26] L. Feng, X. Zhang, and B. Liu, "A high-dimensional spatial rank test for two-sample location problems," *Computational Statistics & Data Analysis*, vol. 144, p. 106889, 2020.

[27] R. Ghimire, C. Zhang, and K. R. Pattipati, "A rough set-theory-based fault-diagnosis method for an electric power-steering system," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 5, pp. 2042–2053, 2018.

[28] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1.

[29] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai *et al.*, "Recent advances in convolutional neural networks,"

Pattern Recognition, vol. 77, pp. 354–377, 2018.

[30] J. Guo, D. Zhen, H. Li, Z. Shi, F. Gu, and A. D. Ball, "Fault feature extraction for rolling element bearing diagnosis based on a multi-stage noise reduction method," *Measurement*, vol. 139, pp. 226–235, 2019.

[31] T. Guo and Z. Deng, "An improved emd method based on the multi-objective optimization and its application to fault feature extraction of rolling bearing," *Applied Acoustics*, vol. 127, pp. 46–62, 2017.

[32] Y. Guo, S. J. Barnes, and Q. Jia, "Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation," *Tourism Management*, vol. 59, pp. 467–483, 2017.

[33] Z. Haiyang, W. Jindong, J. Lee, and L. Ying, "A compound interpolation envelope local mean decomposition and its application for fault diagnosis of reciprocating compressors," *Mechanical Systems and Signal Processing*, vol. 110, pp. 273–295, 2018.

[34] J. Han, D. Zhang, G. Cheng, N. Liu, and D. Xu, "Advanced deep-learning techniques for salient and category-specific object detection: a survey," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 84–100, 2018.

[35] M. M. Hassan, M. G. R. Alam, M. Z. Uddin, S. Huda, A. Almogren, and G. Fortino, "Human emotion recognition using deep belief network architecture," *Information Fusion*, vol. 51, pp. 10–18, 2019.

[36] W. Hu, Y. Qian, and F. K. Soong, "A new neural network based logistic regression classifier for improving mispronunciation detection of l2 language learners," in *Chinese Spoken Language Processing (ISCSLP), 2014 9th International Symposium on*. IEEE, 2014, pp. 245–249.

[37] O. Janssens, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufier, S. Verstockt, R. Van de Walle, and S. Van Hoecke, "Convolutional neural network based fault detection for rotating machinery," *Journal of Sound and Vibration*, vol. 377, pp. 331–345, 2016.

[38] O. Janssens, R. Van de Walle, M. Loccufier, and S. Van Hoecke, "Deep learning for infrared thermal image based machine health monitoring," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 1, pp. 151–159, 2017.

[39] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "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, pp. 303–315, 2016.

[40] M. Jiang, Y. Liang, X. Feng, X. Fan, Z. Pei, Y. Xue, and R. Guan, "Text classification based on deep belief network and softmax regression," *Neural Computing and Applications*, vol. 29, no. 1, pp. 61–70, 2018.

[41] X. Jiang, C. Shen, J. Shi, and Z. Zhu, "Initial center frequency-guided vmd for fault diagnosis of rotating machines," *Journal of Sound and Vibration*, vol. 435, pp. 36–55, 2018.

[42] Y. Jiang and S. Yin, "Recent advances in key-performance-indicator oriented prognosis and diagnosis with a matlab toolbox: Db-kit," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2849–2858, 2018.

[43] Y. Jiang, S. Yin, and O. Kaynak, "Data-driven monitoring and safety control of industrial cyber-physical systems: Basics and beyond," *IEEE Access*, vol. 6, pp. 47 374–47 384, 2018.

[44] S. Kim and J.-H. Choi, "Convolutional neural network for gear fault diagnosis based on signal segmentation approach," *Structural Health Monitoring*, p.

1475921718805683, 2018.

[45] J. R. Kolodziej and J. N. Trout, "An image-based pattern recognition approach to condition monitoring of reciprocating compressor valves," *Journal of Vibration and Control*, vol. 24, no. 19, pp. 4433–4448, 2018.

[46] S. J. Koopman, N. Shephard, and J. A. Doornik, "Statistical algorithms for models in state space using ssfpack 2.2," *The Econometrics Journal*, vol. 2, no. 1, pp. 107–160, 1999.

[47] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[48] H. Lee, P. Pham, Y. Largman, and A. Y. Ng, "Unsupervised feature learning for audio classification using convolutional deep belief networks," in *Advances in neural information processing systems*, 2009, pp. 1096–1104.

[49] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, 2016.

[50] H. Li, H.-Z. Huang, Y.-F. Li, J. Zhou, and J. Mi, "Physics of failure-based reliability prediction of turbine blades using multi-source information fusion," *Applied Soft Computing*, vol. 72, pp. 624–635, 2018.

[51] J. Li, D. Deng, J. Zhao, D. Cai, W. Hu, M. Zhang, and Q. Huang, "A novel hybrid short-term load forecasting method of smart grid using mlr and lstm neural network," *IEEE Transactions on Industrial Informatics*, 2020.

[52] X. Li, X. Peng, Z. Zhang, X. Jia, and Z. Wang, "A new method for nondestructive fault diagnosis of reciprocating compressor by means of strain-based

p–v diagram," *Mechanical Systems and Signal Processing*, vol. 133, p. 106268, 2019.

[53] Y. Li, L. Zou, L. Jiang, and X. Zhou, "Fault diagnosis of rotating machinery based on combination of deep belief network and one-dimensional convolutional neural network," *IEEE Access*, vol. 7, pp. 165 710–165 723, 2019.

[54] Y. Li, M. Xu, Y. Wei, and W. Huang, "An improvement emd method based on the optimized rational hermite interpolation approach and its application to gear fault diagnosis," *Measurement*, vol. 63, pp. 330–345, 2015.

[55] D. Lin, Z. Lin, L. Sun, K.-A. Toh, and J. Cao, "Llc encoded bow features and softmax regression for microscopic image classification," in *Circuits and Systems (ISCAS), 2017 IEEE International Symposium on*. IEEE, 2017, pp. 1–4.

[56] T. R. Lin, E. Kim, and A. C. Tan, "A practical signal processing approach for condition monitoring of low speed machinery using peak-hold-down-sample algorithm," *Mechanical systems and signal processing*, vol. 36, no. 2, pp. 256–270, 2013.

[57] H. Liu, X. Wang, and C. Lu, "Rolling bearing fault diagnosis based on lcd–teo and multifractal detrended fluctuation analysis," *Mechanical Systems and Signal Processing*, vol. 60, pp. 273–288, 2015.

[58] S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, D. Feng, M. J. Fulham *et al.*, "Multimodal neuroimaging feature learning for multiclass diagnosis of alzheimer's disease," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 4, pp. 1132–1140, 2015.

[59] T. Liu, H. Liu, Y. Li, Z. Zhang, and S. Liu, "Efficient blind signal reconstruction with wavelet transforms regularization for educational robot infrared vi-

sion sensing," *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 1, pp. 384–394, 2018.

[60] Y. Liu, X. Fan, C. Lv, J. Wu, L. Li, and D. Ding, "An innovative information fusion method with adaptive kalman filter for integrated ins/gps navigation of autonomous vehicles," *Mechanical Systems and Signal Processing*, vol. 100, pp. 605–616, 2018.

[61] Z. Liu, Z. Wu, T. Li, J. Li, and C. Shen, "Gmm and cnn hybrid method for short utterance speaker recognition," *IEEE Transactions on Industrial Informatics*, 2018.

[62] Z. Liu, Z. Jia, C.-M. Vong, S. Bu, J. Han, and X. Tang, "Capturing high-discriminative fault features for electronics-rich analog system via deep learning," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1213–1226, 2017.

[63] Z. Liu, Z. He, W. Guo, and Z. Tang, "A hybrid fault diagnosis method based on second generation wavelet de-noising and local mean decomposition for rotating machinery," *ISA transactions*, vol. 61, pp. 211–220, 2016.

[64] O. Luaces, J. Díez, T. Joachims, and A. Bahamonde, "Mapping preferences into euclidean space," *Expert Systems with Applications*, vol. 42, no. 22, pp. 8588–8596, 2015.

[65] P. Luukko, J. Helske, and E. Räsänen, "Introducing libeemd: A program package for performing the ensemble empirical mode decomposition," *Computational Statistics*, vol. 31, no. 2, pp. 545–557, 2016.

[66] B. Ma, Y. Zhao, Y. Zhang, Q. L. Jiang, and X. Q. Hou, "Machinery early fault detection based on dirichlet process mixture model," *IEEE Access*, vol. 7, pp. 89 226–89 233, 2019.

[67] M. Ma, X. Chen, X. Zhang, B. Ding, and S. Wang, "Locally linear embedding on grassmann manifold for performance degradation assessment of bearings," *IEEE Transactions on Reliability*, vol. 66, no. 2, pp. 467–477, 2017.

[68] A. Mejia-Barron, M. Valtierra-Rodriguez, D. Granados-Lieberman, J. C. Olivares-Galvan, and R. Escarela-Perez, "The application of emd-based methods for diagnosis of winding faults in a transformer using transient and steady state currents," *Measurement*, vol. 117, pp. 371–379, 2018.

[69] C. Mishra, A. K. Samantaray, and G. Chakraborty, "Rolling element bearing fault diagnosis under slow speed operation using wavelet de-noising," *Measurement*, vol. 103, pp. 77–86, 2017.

[70] A.-r. Mohamed, G. E. Dahl, G. Hinton *et al.*, "Acoustic modeling using deep belief networks," *IEEE Trans. Audio, Speech & Language Processing*, vol. 20, no. 1, pp. 14–22, 2012.

[71] J. Ngiam, Z. Chen, S. A. Bhaskar, P. W. Koh, and A. Y. Ng, "Sparse filtering," in *Advances in neural information processing systems*, 2011, pp. 1125–1133.

[72] T. H. Nguyen, K. Shirai, and J. Velcin, "Sentiment analysis on social media for stock movement prediction," *Expert Systems with Applications*, vol. 42, no. 24, pp. 9603–9611, 2015.

[73] K. Pichler, E. Lughofer, M. Pichler, T. Buchegger, E. P. Klement, and M. Huschenbett, "Fault detection in reciprocating compressor valves under varying load conditions," *Mechanical Systems and Signal Processing*, vol. 70, pp. 104–119, 2016.

[74] G. Qi, W.-T. Tsai, Y. Hong, W. Wang, G. Hou, Z. Zhu *et al.*, "Fault-diagnosis for reciprocating compressors using big data," in *2016 IEEE Second Inter-*

national Conference on Big Data Computing Service and Applications (BigDataService). IEEE, 2016, pp. 72–81.

[75] Y. Ren and Y. Wu, "Convolutional deep belief networks for feature extraction of eeg signal," in *2014 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2014, pp. 2850–2853.

[76] D. Reynolds, "Gaussian mixture models," *Encyclopedia of biometrics*, pp. 827–832, 2015.

[77] R. Sadeghi, H. Samet, and T. Ghanbari, "Detection of stator short-circuit faults in induction motors using the concept of instantaneous frequency," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 8, pp. 4506–4515, 2018.

[78] M. Safizadeh and A. Golmohammadi, "Prediction of oil whirl initiation in journal bearings using multi-sensors data fusion," *Measurement*, vol. 151, p. 107241, 2020.

[79] N. Sakthivel, B. B. Nair, M. Elangovan, V. Sugumaran, and S. Saravanmurugan, "Comparison of dimensionality reduction techniques for the fault diagnosis of mono block centrifugal pump using vibration signals," *Engineering Science and Technology, an International Journal*, vol. 17, no. 1, pp. 30–38, 2014.

[80] R. Sarikaya, G. E. Hinton, and A. Deoras, "Application of deep belief networks for natural language understanding," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 4, pp. 778–784, 2014.

[81] H. Shao, H. Jiang, H. Zhang, W. Duan, T. Liang, and S. Wu, "Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing," *Mechanical Systems and Signal Processing*, vol. 100, pp. 743–765, 2018.

[82] H. Shao, H. Jiang, H. Zhang, and T. Liang, "Electric locomotive bearing fault diagnosis using novel convolutional deep belief network," *IEEE Trans. Ind. Electron*, vol. 65, no. 3, pp. 2727–2736, 2018.

[83] V. Sharma and A. Parey, "Performance evaluation of decomposition methods to diagnose leakage in a reciprocating compressor under limited speed variation," *Mechanical Systems and Signal Processing*, vol. 125, pp. 275–287, 2019.

[84] M. Silva, A. Santos, R. Santos, E. Figueiredo, C. Sales, and J. C. Costa, "Deep principal component analysis: An enhanced approach for structural damage identification," *Structural Health Monitoring*, p. 1475921718799070, 2018.

[85] R. Slama, H. Wannous, M. Daoudi, and A. Srivastava, "Accurate 3d action recognition using learning on the grassmann manifold," *Pattern Recognition*, vol. 48, no. 2, pp. 556–567, 2015.

[86] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the case western reserve university data: A benchmark study," *Mechanical Systems and Signal Processing*, vol. 64, pp. 100–131, 2015.

[87] K. Sonkin, L. Stankevich, Y. Khomenko, Z. Nagornova, N. Shemyakina, A. Koval, and D. Perets, "Neurological classifier committee based on artificial neural networks and support vector machine for single-trial eeg signal decoding," in *International Symposium on Neural Networks*. Springer, 2016, pp. 100–107.

[88] J. T. Springenberg and M. Riedmiller, "Improving deep neural networks with probabilistic maxout units," *arXiv preprint arXiv:1312.6116*, 2013.

[89] C. Sun, P. Wang, R. Yan, R. X. Gao, and X. Chen, "Machine health monitoring based on locally linear embedding with kernel sparse representation for neighborhood optimization," *Mechanical Systems and Signal Processing*, vol. 114, pp. 25–34, 2019.

[90] C. Sun, Z. Zhang, X. Luo, T. Guo, J. Qu, and B. Li, "Support vector machine-based grassmann manifold distance for health monitoring of viscoelastic sandwich structure with material ageing," *Journal of Sound and Vibration*, vol. 368, pp. 249–263, 2016.

[91] H. Tan, Z. Ma, S. Zhang, Z. Zhan, B. Zhang, and C. Zhang, "Grassmann manifold for nearest points image set classification," *Pattern Recognition Letters*, vol. 68, pp. 190–196, 2015.

[92] S. Tao, T. Zhang, J. Yang, X. Wang, and W. Lu, "Bearing fault diagnosis method based on stacked autoencoder and softmax regression," in *Control Conference (CCC), 2015 34th Chinese*. IEEE, 2015, pp. 6331–6335.

[93] A. Terenin, M. Magnusson, L. Jonsson, and D. Draper, "Polya urn latent dirichlet allocation: a doubly sparse massively parallel sampler," *IEEE transactions on pattern analysis and machine intelligence*, 2018.

[94] M. Z. Uddin, M. M. Hassan, A. Almogren, A. Alamri, M. Alrubaian, and G. Fortino, "Facial expression recognition utilizing local direction-based robust features and deep belief network," *IEEE Access*, vol. 5, pp. 4525–4536, 2017.

[95] M. V. van der Seijs, D. de Klerk, and D. J. Rixen, "General framework for transfer path analysis: History, theory and classification of techniques," *Mechanical Systems and Signal Processing*, vol. 68, pp. 217–244, 2016.

[96] N. K. Verma, R. K. Sevakula, S. Dixit, and A. Salour, "Intelligent condition based monitoring using acoustic signals for air compressors," *IEEE Transactions on Reliability*, vol. 65, no. 1, pp. 291–309, 2016.

[97] C. Wang, M. Gan *et al.*, "Non-negative emd manifold for feature extraction in machinery fault diagnosis," *Measurement*, vol. 70, pp. 188–202, 2015.

[98] D. Wang and A. Al-Rubaie, "Incremental learning with partial-supervision based on hierarchical dirichlet process and the application for document classification," *Applied Soft Computing*, vol. 33, pp. 250–262, 2015.

[99] F. Wang, T. Xu, T. Tang, M. Zhou, and H. Wang, "Bilevel feature extraction-based text mining for fault diagnosis of railway systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 1, pp. 49–58, 2017.

[100] J. Wang, P. Fu, L. Zhang, R. X. Gao, and R. Zhao, "Multilevel information fusion for induction motor fault diagnosis," *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 5, pp. 2139–2150, 2019.

[101] J. Wang, G. Du, Z. Zhu, C. Shen, and Q. He, "Fault diagnosis of rotating machines based on the emd manifold," *Mechanical Systems and Signal Processing*, vol. 135, p. 106443, 2020.

[102] R. Wang, M. Xiang, and C. Li, "Denoising fmcw ladar signals via eemd with singular spectrum constraint," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 6, pp. 983–987, 2019.

[103] X.-B. Wang, Z.-X. Yang, and X.-A. Yan, "Novel particle swarm optimization-based variational mode decomposition method for the fault diagnosis of complex rotating machinery," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 1, pp. 68–79, 2017.

[104] C.-K. Wen, S. Jin, K.-K. Wong, J.-C. Chen, and P. Ting, "Channel estimation for massive mimo using gaussian-mixture bayesian learning," *IEEE Transactions on Wireless Communications*, vol. 14, no. 3, pp. 1356–1368, 2015.

[105] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT Press Cambridge, MA, 2006, vol. 2, no. 3.

[106] R.-T. Wu and M. R. Jahanshahi, "Data fusion approaches for structural health monitoring and system identification: Past, present, and future," *Structural Health Monitoring*, p. 1475921718798769, 2018.

[107] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in adaptive data analysis*, vol. 1, no. 01, pp. 1–41, 2009.

[108] M. Xia, T. Li, L. Xu, L. Liu, and C. W. De Silva, "Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 1, pp. 101–110, 2017.

[109] W. Xu and J. Yu, "A novel approach to information fusion in multi-source datasets: a granular computing viewpoint," *Information sciences*, vol. 378, pp. 410–423, 2017.

[110] Y. Xu, Y. Sun, J. Wan, X. Liu, and Z. Song, "Industrial big data for fault diagnosis: Taxonomy, review, and applications," *IEEE Access*, vol. 5, pp. 17 368–17 380, 2017.

[111] X. Yan and M. Jia, "A novel optimized svm classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, vol. 313, pp. 47–64, 2018.

[112] J. Yang, X. Liao, X. Yuan, P. Llull, D. J. Brady, G. Sapiro, and L. Carin, "Compressive sensing by learning a gaussian mixture model from measurements," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 106–119, 2015.

[113] J. Yang, Y. Bai, F. Lin, M. Liu, Z. Hou, and X. Liu, "A novel electrocardiogram arrhythmia classification method based on stacked sparse auto-encoders

and softmax regression," *International Journal of Machine Learning and Cybernetics*, pp. 1–8, 2017.

[114] J.-F. Yeh, Y.-S. Tan, and C.-H. Lee, "Topic detection and tracking for conversational content by using conceptual dynamic latent dirichlet allocation," *Neurocomputing*, vol. 216, pp. 310–318, 2016.

[115] S. Yin, J. J. Rodriguez-Andina, and Y. Jiang, "Real-time monitoring and control of industrial cyberphysical systems: With integrated plant-wide monitoring and control framework," *IEEE Industrial Electronics Magazine*, vol. 13, no. 4, pp. 38–47, 2019.

[116] Z. Yin and J. Hou, "Recent advances on svm based fault diagnosis and process monitoring in complicated industrial processes," *Neurocomputing*, vol. 174, pp. 643–650, 2016.

[117] J. Yu and X. Zhou, "One-dimension residual convolutional auto-encoder-based feature learning for gearbox fault diagnosis," *IEEE Transactions on Industrial Informatics*, 2020.

[118] K. Yu, T. R. Lin, and J. W. Tan, "A bearing fault diagnosis technique based on singular values of eemd spatial condition matrix and gath-geva clustering," *Applied Acoustics*, vol. 121, pp. 33–45, 2017.

[119] A. Yunusa-Kaltungo, J. K. Sinha, and K. Elbhbah, "An improved data fusion technique for faults diagnosis in rotating machines," *Measurement*, vol. 58, pp. 27–32, 2014.

[120] M. Žarković and Z. Stojković, "Analysis of artificial intelligence expert systems for power transformer condition monitoring and diagnostics," *Electric Power Systems Research*, vol. 149, pp. 125–136, 2017.

[121] H. Zhang and Y. Deng, "Engine fault diagnosis based on sensor data fusion considering information quality and evidence theory," *Advances in Mechanical Engineering*, vol. 10, no. 11, p. 1687814018809184, 2018.

[122] X. Zhang, Y.-C. Liang, and J. Fang, "Novel bayesian inference algorithms for multiuser detection in m2m communications," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 7833–7848, 2017.

[123] X. Zhang, Z. Liu, Q. Miao, and L. Wang, "An optimized time varying filtering based empirical mode decomposition method with grey wolf optimizer for machinery fault diagnosis," *Journal of Sound and Vibration*, vol. 418, pp. 55–78, 2018.

[124] Z. Zhang and J. Zhao, "A deep belief network based fault diagnosis model for complex chemical processes," *Computers & chemical engineering*, vol. 107, pp. 395–407, 2017.

[125] H. Zhao, J. Wang, H. Han, and Y. Gao, "A feature extraction method based on hlmd and mfe for bearing clearance fault of reciprocating compressor," *Measurement*, vol. 89, pp. 34–43, 2016.

[126] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mechanical Systems and Signal Processing*, vol. 115, pp. 213–237, 2019.

[127] J.-H. Zhong, P. K. Wong, and Z.-X. Yang, "Fault diagnosis of rotating machinery based on multiple probabilistic classifiers," *Mechanical Systems and Signal Processing*, vol. 108, pp. 99–114, 2018.

[128] Y. Zhong, Q. Zhu, and L. Zhang, "Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing

imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 11, pp. 6207–6222, 2015.