Spring 2023

Deep Learning Based Fault Diagnosis and Prognosis for Bearing

Guangxing Niu

Deep Learning based fault diagnosis and prognosis for Bearing

by

Guangxing Niu

Bachelor of Science
North University of China 2014
Master of Science
Harbin Institute of Technology 2016

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Electrical Engineering

College of Engineering and Computing

University of South Carolina

2023

Accepted by:

Bin Zhang, Major Professor

Xiaofeng Wang, Committee Member

Seongtae Bae, Committee Member

Paul Ziehl, Committee Member

Cheryl L. Addy, Interim Vice Provost and Dean of the Graduate School

Abstract

Rolling element bearings are critical components in industrial rotating machines. Faults and failures of bearings can cause degradation of machine performance or even a catastrophe. Therefore, it is significant work to perform bearing fault diagnosis and prognosis (FDP) reliably and effectively. Diagnosis aims to detect and estimate the fault state of bearing in real-time. Prognosis aims to conduct a long-term prediction to predict the fault evolution and estimate the remaining useful life (RUL) of the bearing.

Deep learning (DL) based FDP methods have become an important branch of bearing FDP methods due to their powerful capabilities in feature automatic learning, fault modeling, and fault identification. Although they have made significant achievements in bearing FDP, there still exist some challenges and open problems, especially for modern industrial systems that are often designed with multiple complicated functions and operated under varying operating conditions and environments. Most existing DL-based FDP methods are designed for stationary operating conditions and are purely data-based, which often cannot guarantee good results and high efficiency for FDP of modern industrial systems.

To overcome the challenges and improve the performance of the existing DL-based bearing FDP methods, this thesis proposes some improvements of DL-based bearing FDP methods in terms of structure optimization, adaptive learning strategy, FDP algorithm execution strategy, etc. The proposed DL-based FDP methods can generate better results in accuracy, efficiency, and robustness for bearing FDP tasks.

Deep belief network (DBN) and convolutional neural network (CNN) are two

mainstream DL structures. They have different advantages and unique network characteristics in various applications. This thesis is conducted based on these two networks to improve their performance in FDP applications. This thesis first proposes a deep belief network (DBN) and principal components analysis (PCA) based approach for bearing fault classification. A particle swarm optimization (PSO) based DBN adaptive training procedure is employed to optimize the DBN structure. This approach provides an automatic, accurate, and effective bearing fault classification solution.

To improve the accuracy and learning efficiency of the existing CNN-based bearing diagnosis methods, a deep residual CNN is proposed for multi-task bearing fault diagnosis. In the proposed approach, domain knowledge is integrated with monitoring data to build the information map. Two attention modules are introduced to enhance the discriminate feature learning ability. Two classifiers are employed for multi-task diagnosis. This diagnosis method has significant improvements in terms of diagnostic accuracy and training efficiency.

For the bearing diagnosis under varying speeds, a novel multi-scale discriminate CNN based bearing fault diagnosis is proposed to deal with the challenges caused by the varying speeds. The varying speed information is integrated with monitoring data to build the information map. A multi-scale discriminate convolutional neural network-based method is designed to enhance the learning ability for signals with some specific characteristics. The extracted features are then employed to identify bearing fault modes. Experimental results and comparisons show that the proposed approach can achieve better performance in terms of accuracy and efficiency than some state-of-the-art methods.

For bearing fault prognosis, this thesis proposes a hybrid Bayesian estimation-based FDP framework with fault detection and multiple model fusion. In the proposed approach, convolutional neural network (CNN) is used to detect

fault and select appropriate fault dynamic model. To improve the performance, continuous wavelet coefficient matrices (CWCM) power spectrum of vibration are fused with operating conditions to build information maps for fault detection and model selection. After a fault is detected, Bayesian estimation based FDP method is triggered to estimate the fault state and predict the remaining useful life. In the FDP process, Dempster-Shafer theory (DST) is employed to fuse prediction results from different models if necessary. The bearing state and RUL can be estimated based on PF-based estimation and fusion results.

The proposed methods are verified with different bearing case studies. The experimental results are analyzed and compared with other state-to-art approaches, which demonstrate that the proposed approaches have better performance. This thesis provides a successful exploration and attempt of DL algorithms in dealing with fault diagnosis and prognosis problems.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and motivation

Rotating machinery systems play important roles in modern industries, such as wind turbines, vehicles, helicopters, ships, and many other complex mechanical systems. Bearings are pivotal mechanical components in rotating systems, as illustrated in Fig. 1.1. Bearings typically operate under tough conditions, such as daily basis, heavy load, high speed, and high temperature. Moreover, bearings often work under ill-fitted conditions such as excessive loads, misalignment caused by improper installation, waviness of raceway, and off-size rolling elements caused by unqualified manufacturing. The harsh and ill-fitted working conditions can lead to bearing faults, degradation, and complete failure. The degradation and faults of bearings, if not processed in a timely and effective means, will affect machine production, performance, reliability and may eventually lead to system failure or safety issues. Unexpected faults and failures also cause significant economic loss.

Bearing faults are top contributors to the failure of rotating machinery systems. For example, in wind energy systems, about 80% gearbox failures are caused by bearing faults [1], about 41% total faults of induction motors are related to bearings [2]. To ensure the reliability, as well as the effectiveness of machine maintenance and logistics planning, and to meet the increasing demands for system safety, reliable and accurate fault diagnosis drew more and more studies in recent years. Condition-based maintenance (CBM) is an up-to-date maintenance strategy that is conducted based

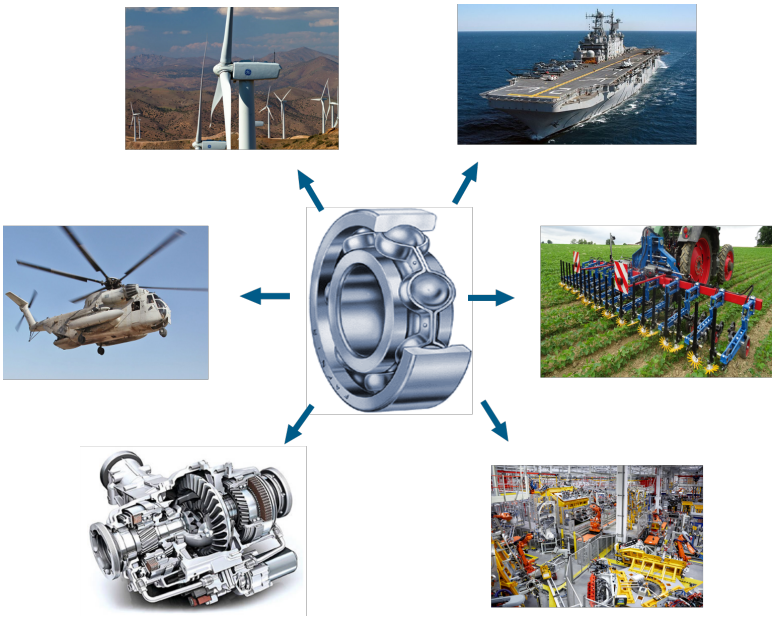

Figure 1.1: Rotating machinery system

on sensor measurements (such as vibration, acoustic emission, temperature, etc.) from bearings and advanced algorithms. The maintenance plan can be optimized by the availability of bearing health condition [3]. Bearing fault diagnosis and prognosis (FDP) is an essential technique of CBM, which provides evidence of bearing fault state in current and future time instants and bearing remaining useful life (RUL) for maintenance decision-making. Besides, production planning can also be optimized using the diagnosis and prognosis output to achieve a high productivity rate and execution efficiency.

Rolling element bearings are one of the most widely used bearings in rotating machinery systems. A rolling element bearing consists of an inner race, an outer race, a number of rolling elements, and a cage to hold and evenly separate the rolling elements. Bearing fault modes can be roughly divided as local faults and distributed faults [4, 5]. Local faults are defined as single localized faults, such as pitting, scratch, crack, hole, etc. Local faults can occur in any component of a bearing. Distributed faults are defined as irregularities of bearing structures, such as misalignment of shaft

or races, eccentric races, off-size rolling elements, roughness, etc. These faults can be caused by excessive operational conditions, improper installation, and imperfect manufacturing. In operation, distributed faults often cause excessive contact force and friction, which will eventually lead to local faults. When bearings operate under fault conditions, they cause certain characteristic signals that are excited in the form of sound, vibration, energy, or acoustic emission.

Bearing FDP are systematic approaches to evaluate the health condition and the RUL in real-time. Among the explored bearing sensing modalities, vibration signals and acoustic emission (AE) measurement and analysis are the two well-known, cost-effective, and most employed signals in bearing FDP. Bearing fault diagnosis includes two tasks: fault detection and fault classification. Fault detection aims to evaluate the bearing and determine whether a fault occurs. Fault classification aims to identify or recognize the type of fault. The fault type includes fault location, fault dimension, and other faults related information. Since the two tasks are all pattern recognition tasks, they are generally achieved at the same time. Bearing fault prognosis aims to predict the fault evolution process based on a dynamic fault model and predict the RUL based on the current health condition and the use of bearing.

1.2 Research state of bearing FDP

Numerous successes have been achieved for bearing FDP in the past few decades in many rotating machinery systems [6]. The performance of bearing FDP, which includes accuracy, efficiency, robustness, extensibility, depends not only on the quality of bearing monitoring techniques, but also on the applied fault FDP techniques. This section summarizes the research state of bearing FDP.

1.2.1 Bearing fault diagnosis

The existing fault diagnosis techniques can be summarized as signal processing methods, machine learning based methods, and deep learning based methods as shown in Fig. 1.2. The procedure of bearing fault diagnosis consists of data pre-processing, feature extraction, and fault classification. Data pre-processing is applied on raw data to do some feature extraction, which is also known as feature engineering. Since most bearing faults cannot be measured directly, feature extraction aims to extract fault indicators or features from bearing monitoring data to indicate the fault state. Therefore, it often requires that the extracted feature and fault state has a strong correlation. The extracted features are then used for fault classification manually or through intelligent algorithms.

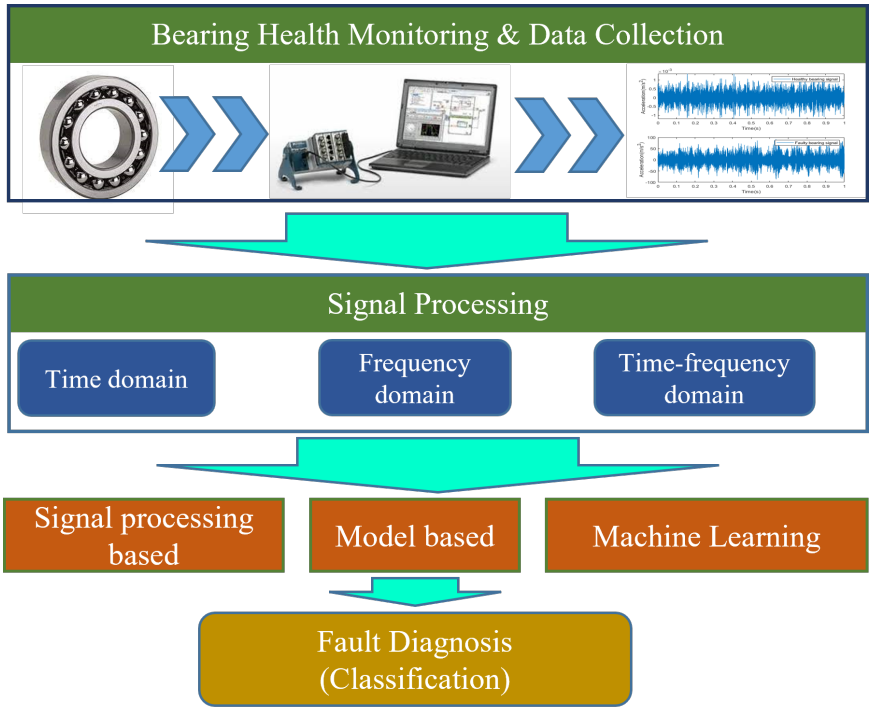

Figure 1.2: Bearing fault diagnosis method overview

1.2.1.1 Signal processing methods

Signal processing approaches mainly include two steps: feature extraction and decision-making. Bearing monitoring signals, such as vibration and acoustic emission, contain abundant information. However, the information is always submerged with background noise, which makes it difficult to distinguish. Signal processing methods rely on all kinds of single processing techniques to extract scalar indices to evaluate bearing health conditions. The signal can be analyzed in the time-domain, frequency-domain, and time-frequency domain. Time-domain signal processing methods are the most straightforward techniques for detecting bearing faults. Many of the time-frequency signal processing methods are conducted around the analysis on statistical factors [7], like peak-to-peak value, kurtosis, root-mean-square (RMS), crest factor, impulse factor, etc. Besides, some other time-domain analysis techniques like time-synchronous averaging [8], time-shifting [9], are also employed for bearing fault diagnosis. The bearing condition can be estimated from these statistical factor values by comparing them with the values when the bearing is in the normal condition. For example, RMS is compared with a recommended threshold value to estimate the condition of a bearing in [10]. Multiple time-domain features, including waveform length, slope sign changes, Wilson amplitude, etc., are used in [11]. Time-domain signal processing methods are the most basic and easy-to-use methods for bearing fault diagnosis. They show many disadvantages in practical applications. For example, some of the fault characteristics will not show up in the time-domain, especially for incipient faults. Besides, bearings are usually operated in noise environments, and fault signals are submerged into noise. Therefore, time-domain signal processing can just be employed in some elementary bearing fault diagnosis tasks.

The occurrence of some bearing faults introduces some specific frequency components in bearing monitoring signals. Frequency-domain signal processing

methods can reveal the frequency characteristics effortlessly for bearing fault diagnosis. Common frequency domain analysis techniques include fast Fourier transform (FFT), short time Fourier transform (STFT), energy spectrum analysis (ESA), etc. These techniques convert time-domain signals into frequency domain spectrum to reveal the frequency information. Many frequency-domain signal processing techniques have been proposed for bearing fault diagnosis. For example, Henning et al. [12] analyzed the dominant mechanical resonance frequencies of rolling bearings for fault detection. Wang et al. [13] adopted the kurtosis of temporal signals that are filtered by STFT.

Bearing operating signals contain rich periodic components, which show different frequency-related characteristics as non-stationary, cyclic, time invariance, etc. Time-domain or frequency-domain methods alone are often difficult to reveal these characteristics. Therefore, some time-frequency domain signal processing methods are developed to analyze the signal in both domains. The common used time-frequency techniques include wavelet transform (WT), empirical mode decomposition (EMD), Wigner Ville distribution (WVD), etc. These time-frequency methods have proven effectiveness in exploring non-stationary characteristics. For example, an empirical wavelet transform (EWT) was proposed for fault diagnosis of wheel bearing [14]. The method combined classic wavelet with EMD, which is verified suitable for processing non-stationary vibration signals.

Although signal processing methods have some advantages and are successfully applied in many practical rotating machinery systems, the disadvantages are also very evident. In signal processing methods, bearing faults are manually diagnosed based on the empirically determined and selected features from signal processing. However, there exist many challenges in this process. For example, temporal analysis cannot determine very similar fault-related components. Besides, fault-related signals are not strong enough to be distinguished in terms of amplitudes, energy, etc., especially

at the initial stages. It is a challenging work to build a relationship between selected features and fault modes. The challenge lies in that manually selected features may not be the optimal one for bearing fault diagnosis. Besides, the patterns or distinctive characteristics of the features that can reveal the occurrence of a bearing fault may be difficult or impossible for human being to identify, especially for the fault at early stages and the bearing monitoring signal with a small signal-to-noise ratio. Signal processing methods can only be used to detect a few well-defined fault modes, while faults in real rotating machinery systems are often complicated. Therefore, signal processing methods cannot guarantee satisfying performance for most bearing fault diagnosis tasks. To find the exceptional fault introduced patterns, many machine learning techniques are applied to learn and explore the patterns and relationships hidden in features and bearing fault modes.

1.2.1.2 Machine learning based methods

Machine learning techniques are derived from artificial intelligence. They are employed to find patterns from data. The broadly practiced machine learning algorithms for bearing fault diagnosis include artificial neural network (ANN) [15], support vector machine (SVM) [16], principle component analysis (PCA) [17], k-nearest neighbor (KNN), etc. Machine learning algorithms can learn subtle relationships between features and bearing fault modes. The trained machine learning algorithms can make decision for bearing fault diagnosis based on the learned relationship.

Another important branch of machine learning based bearing fault diagnosis is fuzzy formalisms. In practical bearing diagnosis fields, the knowledge about rotating machinery systems is necessarily incomplete and ambiguous. The reasons can be summarized as follows: first, fault-related signals or characteristics cannot be guaranteed to be fully observed; second, the volume of faulty data is always smaller

than that of healthy data; third, bearing monitoring data are noisy, redundant, and non-stationary; fourth, the relationship between bearing monitoring data and fault types are not clear and hard to identify. Fuzzy formalisms can be used to explore the inherent uncertainty and provide an insight into the linguistic relationships between features and faults. The most applied fuzzy formalisms include neuro-fuzzy inference system [18], neural-fuzzy inference system (ANFIS) [19], approximate entropy [20], etc.

The general procedure of machine learning based methods also includes data pre-processing, feature extraction, fault classification. In machine learning methods, bearing faults are identified and classified using machine learning algorithms based on the selected features. The machine learning based methods have achieved satisfying performance for different applications. This benefits the development of complex signal processing techniques and the intelligent recognition capacity of machine learning algorithms. The feature extraction techniques in machine learning methods are similar to those in signal processing methods. To improve the recognition ability of machine learning methods, machine learning methods are often integrated with various signal processing techniques, which include WT [21, 22], EMD [23, 24], EWT [25, 26], ensemble empirical mode decomposition (EEMD) [27, 28], variational mode decomposition (VMD) [29, 30], linear discriminant analysis (LDA) [31], local characteristic scale decomposition (LCD) [32], etc. All these techniques attempt to decompose the bearing monitoring signals in different scales to reveal some fault-related hidden characteristics.

Many research works have studied the effectiveness of machine learning based methods for bearing fault diagnosis. Most of the works are conducted based on the above-mentioned feature extraction techniques and machine learning algorithms. An intelligent EMD and ANN based bearing fault diagnosis approach was proposed in [15], in which the most significant intrinsic mode functions (IMFs) are selected to

train ANN. In [33], EEMD energy entropy and singular values are extracted using EEMD as the features and an optimized SVM is applied to identify the fault type. A composite multiscale fuzzy entropy and ensemble SVM method was proposed in [34]. In [19], an ANFIS was trained using the features extracted from wavelet transform under the presence of load variations. More machine learning based approaches are reviewed in [35–38].

Based on the above discussion, both signal processing methods and machine learning methods rely heavily on feature extraction. The performance of bearing diagnosis depends on the quality of feature extraction and selection. In this process, complex signal processing and domain expertise are always involved in selecting reasonable features for diagnosis. Besides, traditional machine learning algorithms are often assembled with shallow structures. There are some challenges when dealing with fault diagnosis tasks for some complex systems and the faults at their early stages. Even the features are effectively and professionally selected with expert knowledge, the fault patterns are still not obvious due to the interference of noise and fault-independent signals. Therefore, the diagnosis accuracy and efficiency cannot be guaranteed for modern rotating machinery systems that working in versatile operating conditions and noisy environments.

1.2.1.3 Deep learning based methods

Deep learning (DL) is a new branch of machine learning. DL can learn features automatically from raw data and find the potential patterns hidden in data. DL techniques have achieved great achievements in various fields, such as natural language processing, computer visual, pattern recognition, etc, in different applications, such as unmanned vehicles [39], automatic speech recognition (ASR) [40], as well as bearing fault diagnosis [41]. DL-based methods are drawing increasing attention to overcome the aforementioned challenges and further improve

the performance of bearing fault diagnosis.

DL-based bearing fault diagnosis methods have witnessed rapid development in the past about five years. DL algorithms perform excellently in fault feature extraction, classification, and can offer some new functionalities that cannot be accomplished by traditional methods. The widely studied DL-based bearing fault diagnosis methods include deep belief network (DBN), convolutional neural network (CNN), long short term memory (LSTM), deep auto-encoder (DAE), etc. DBN is a deep network constructed by multilayer Restricted Boltzmann Machines (RBMs), DBN and its variation structures are widely used for bearing fault diagnosis [36]. The DBN-based bearing fault diagnosis can be summarized in two categories. One is variation of the network, which includes adaptive learning [42–45], training and fine-tuning strategy [46, 47], etc. An improved hierarchical adaptive DBN is developed for bearing fault diagnosis in [42], in which the learning adjustment strategy is optimized by Nesterov momentum. The results show that the network training convergence can be improved with the optimized method. In [46], an improved ensemble method is introduced to improve DBN feature learning ability. Another one is developed with the combination of different signal processing techniques [7, 45, 48]. A dual-tree complex wavelet packet (DTCWPT) and DBN based method is proposed in [45] to extract the fault characteristics information. An adaptive DBN is constructed to identify bearing fault mode based on the extracted features. In [49], the bearing monitoring is processed using variational mode decomposition to highlight the fault related characteristics. The fault mode is diagnosed by ensemble DBN, and the experimental results demonstrate that this method is effective for bearing fault diagnosis.

CNN plays a dominant role in DL-based bearing fault diagnosis. A typical structure of CNN is mainly composed of several alternating convolutional layers, pooling layers, activation layers, fully connected layer, and output layer. CNN

is designed with the idea of location connection and convolution kernel parameter sharing [50]. The features can be extracted by the convolution and pooling operation layer by layer. Classification or regression task is performed based on the extracted features.

Most of the CNN-based bearing diagnosis methods are designed based on one-dimension (1-D) and two-dimensional (2-D) CNN. 1-D CNN can be used to act on bearing monitoring data directly, while 2-D CNN is designed for learning 2-D local correlations and extracting hierarchical correlations from 2-D images [51]. To fully investigate CNN's powerful feature learning ability for bearing fault diagnosis, CNN-based methods are mainly studied and explored on variant structures and combination with various signal processing techniques.

The variants of CNN include multiscale CNN [52], adaptive CNN [53, 54], residual network [55], densely connected convolutional network (DCNN) [56], etc. CNN and its variants based bearing fault diagnosis methods have achieved promising achievements in terms of accuracy and efficiency. A multiscale kernel residual CNN structure is proposed in [55] to better capture fault characteristics under non-stationary operating conditions. Residual CNN is also introduced to deal with the degradation problem in the network with high depth structures. To maximizing bearing fault diagnosis accuracy without machinery expertise and overcomplicating network structure, a novel hierarchical learning rate adaptive CNN is studied in [54]. A attentive DCNN based method is proposed by combination of DCNN and an attention mechanism [57], the method is verified to outperform competitive results with less training data in bearing fault detection tasks. More CNN-based bearing fault diagnosis methods are reviewed in [51, 58, 59]. The variant structures of CNN have improved the bearing diagnosis performance in different tasks. Another mainstream CNN-based bearing fault diagnosis is explored based on the combination with complex signal processing techniques, which are similar to the applied techniques in machine learning based

ones as introduced in the last section.

Apart from the aforementioned DL based bearing fault diagnosis, there are also many other DL based methods, including transfer learning, generative adversative network (GAN), capsule neural network, etc., that are widely studied to deal with some of the open challenges in bearing fault diagnosis. Transfer learning based methods are studied for bearing diagnosis problems under unknown or partially known area (target domain) or using known knowledge (source domain) [60]. Generative adversarial network (GAN) based methods can provide promising performance in bearing fault diagnosis with imbalanced datasets [59, 61]. A hybrid DL method that combine capsule network with stochastic delta rule is proposed in [62] for bearing fault diagnosis under varying operating conditions and environments.

Based on the review of DL-based methods, the domain knowledge and system operating conditions are not considered and utilized, which are very important information in bearing FDP. Besides, most works focus on the variations of network structures and the employment of signal processing techniques. More important, most of the existing methods conduct in traditional manners, without considering the discrimination of the features. Therefore, accuracy and efficiency of these existing methods can be further improved.

1.2.2 Bearing fault prognosis

Accurate fault diagnosis can detect and identify the presence of bearing faults. After a fault is detected, it is desirable to accurately predict the evolution or growth of the fault and estimate the RUL for optimal maintenance and schedule. Prognosis is the technique of predicting bearing future conditions using the measured or inferred features, as well as fault dynamic models and estimating the RUL of the damaged systems. Numerous prognosis techniques have been proposed for bearing. They can be categorized into three types of fundamental approaches: model-based methods,

data-driven methods, and hybrid methods as shown in Fig. 1.3.

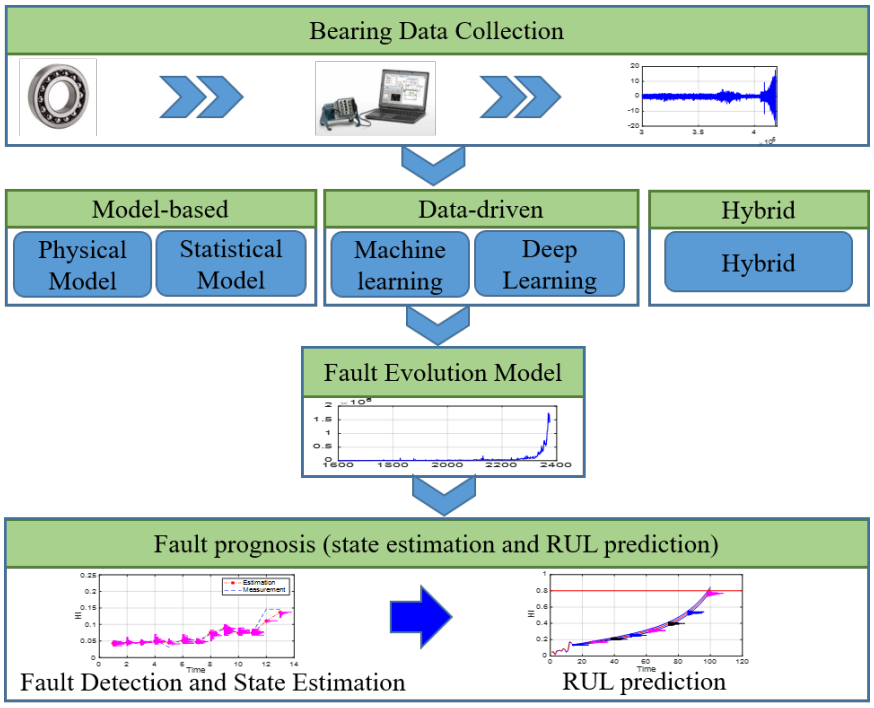

Figure 1.3: Bearing fault prognosis method overview

1.2.2.1 Model-based methods

Model-based approaches are developed based on fault dynamic models. The model can be mathematical physical models, empirical models, statistical models. Physical models can be a set of algebraic or partial differential equations [63], which accurately describe the physics phenomena underlying bearing faults. Significant understanding of mechanistic knowledge and theories relating to the systems is often needed to build models. The review of physical models in rotating machinery prognostics can be found in [64, 65]. For rolling element bearings, the most common behavioral models include crack growth models [66–69], spall progression model [70], Forman crack growth model [71], stress-based fatigue model [72], etc., With different physical models, many different approaches have been developed for bearing fault prognosis. A creep model based prognostic approach is presented for RUL prediction in [73]. In

[72], two stress-based fatigue life models are developed for ball and roller bearings fault prognosis. Statistical models, which include proportional hazards model [74, 75], proportional intensity model [76], exponential model [77, 78], etc, are also widely studied and applied for bearing fault diagnosis. An ensemble model with genetic algorithm and Weibull proportional hazards model is proposed in [74] for bearing RUL prediction. In [79], an exponential model is optimized by gradient descent to estimate the model parameters, and the RMS of bearing vibration is used as the model input to predict bearing RUL.

Besides, some Bayesian estimation techniques are often categorized as model-based methods since the methods are also dependent on pre-constructed fault models [80]. The commonly used techniques include Sequential Monte Carlo (Particle Filter (PF)), Kalman Filter(KF), Extended Kalman Filter (EKF), etc. This kind of prognosis methods employs dynamic state models and measurement models to predict fault states. Bayesian estimation frameworks have been widely used in bearing prognosis due to the uncertainty management ability, which is one of the key factors in prognosis.

Uncertainty is a quantitative indication of the quality of FDP results. It allows decision-makers to assess the reliability of FDP results. The uncertainties in the Bayesian estimation framework can be from fault dynamic modeling, model prediction, measurement, etc. PF based methods are effective in uncertainty management, which provides probability distribution of state that can be used to improve the method stability. The PF based methods have been widely studied and applied in industries. For example, an enhanced PF based method, which combines adaptive importance density function selection and neural network re-sampling smoothing, is developed in [81] for rolling bearing RUL prediction. The EKF based bearing RUL prediction method is studied in [82]. The proposed method is verified effective in bearing fault prognosis. More research works about stochastic modeling

based prognosis methods are reviewed in [83, 84].

Model-based approaches can deliver promising prognostic results when accurate mathematical models can be constructed. However, the structure and operating conditions of real systems are always complex, and the degradation mechanisms are generally stochastic [85]. Therefore, it is difficult to get the physical or analytical system models in forms of analytical equations. Besides, different systems have different physical mechanisms. The constructed model-based methods for a specific system cannot be generalized to others. Model-based method are cost expensive in general. Consequently, the applicability of model-based approaches may be limited in real applications.

1.2.3 Data-driven methods

Data-driven methods use monitoring data to explore hidden system degradation laws and predict the evolution or degradation process of faults in the absence of physical nature of failure mechanism and system degradation models. The widely studied data-driven based techniques include traditional machine learning methods and DL methods. The most well-studied traditional machine learning methods include ANN, support vector regression (SVR), fuzzy logic based methods, Markov model, etc. Since traditional machine learning algorithms often have shallow structures, complicated signal processing are often integrated to improve the prognosis capability. Traditional machine learning based bearing prognosis methods consists of offline learning stage and online prediction process. The offline learning stage includes health indicator (HI) construction and fault degradation learning. HI construction is to build an indicator that can accurately reflect the health state of bearings through historical monitoring data. HI can be the features in any analysis domains or hybrid factors that are constructed based on multiple extracted features. In the HI construction process, various signal processing techniques are often involved. The process is labor

expensive. Besides, adequate prior knowledge and domain expertise are necessary for handcrafted feature extraction. Then, some machine learning techniques are applied to learn the degradation behaviors for online prognosis. In [86], Weibull hazard rates of RMS and kurtosis are used to train ANN based models to predict RUL for rotating machinery. A parsimonious network based on a fuzzy inference system is proposed and tested in [87] for low-speed slew bearing prognosis. In [88], SVM is used to evaluate the degradation stage of bearings, then a hybrid degradation tracking model is employed to exploit the optimal RUL prediction for bearings.

DL techniques have demonstrated powerful automatically representative feature extraction and learning ability. Different types of DL techniques have been successfully applied in bearing prognosis tasks. Compared with traditional machine learning based methods, DL can learn bearing degradation features automatically rather than extracting handcrafted features with domain expertise. Among the emerged DL techniques, DBN, LSTM, CNN are the dominant models that have been widely studied. A DBN-based bearing condition prediction method is proposed in [89] using the RMS of bearing vibration data. In [90], a hybrid LSTM based model is built using bearing data from multi-sources for RUL prediction. An improved 1D-CNN and simple recurrent unit based bearing RUL prediction method is investigated in [91].

DL-based prognosis methods have achieved great achievements. However, the existing DL-based works still have shortcomings, which hinder the application and development of methods in some prognosis tasks. The shortcomings mainly exist in HI construction, fault dynamic modeling, and algorithm execution mechanisms, etc. The HIs in most of the existing works are defined as one of the bearings features, which sometimes cannot reflect the real bearing health conditions in some cases. Besides, the fault dynamic modeling process is conducted without considering the system operating conditions and domain knowledge. Furthermore, the traditional FDP are often executed periodically without consideration of bearing fault growth

rate. They will result in large computation and low prognostic accuracy.

1.2.3.1 Hybrid methods

The hybrid methods are developed with the integration of model-based methods and data-driven methods. The hybrid method can take advantage of both methods to further improve the performance of prognosis. For example, the existing DL-based methods have difficulties in predicting the theoretical distribution of the predicted RUL and quantifying the prognosis uncertainties. In contrast, some of model-based methods have natural capability of dealing with prediction uncertainties but suffering from the degradation feature extraction and model construction. To overcome the limitations from both sides, PF and DL based methods are investigated in many research works. A DBN and particle filter based method is proposed in [92] for hybrid ceramic bearings.

The hybrid methods can overcome some limitations of the separate methods. However, with the increasing system operating conditions and functional complexities, some of the challenges and limitations in model-based methods or data-driven methods are still not well solved effectively.

1.3 Bearing diagnosis and prognosis challenges

Based on the above discussion, DL-based methods have been widely studied and have become one of the most important branches for bearing FDP. They have achieved significant achievements in the past few decades and have been successfully applied in industries. However, some challenges and open problems still exist that hinder the development of DL-based FDP methods. In the FDP tasks for modern industrial systems, the challenges of the existing DL related methods are mainly caused by the following factors:

- Increasing system complexity

  Modern systems are becoming more and more complicated by integrating more sensors and actuators to accomplish sophisticated manufacturing, navigation, transportation missions. The fault mechanisms and modes change a lot with the increasing system complexity. It leads to challenges for existing methods to detect and identify the occurred faults accurately and efficiently. Besides, it is difficult or even impossible to develop an accurate and efficient system physical model for these complicated engineering systems. The reasons can be summarized as follows: 1) the lack of full understanding of the physical mechanism and knowledge of the system, especially those complicated ones; 2) the current modeling approach is not capable of learning and describing the complex relationships in systems; 3) complex signal processing and professional knowledge are always involved in system modeling processes. All these limitations hinder the development of effective FDP.

- Varying working environment and operating condition

  Modern multi-function systems are required to work under different environments and operating conditions. These factors affect the performance, reliability, and lifetime of systems. The system responses to these factors are generally stochastic and irregular, which are difficult to study based on the system mechanism. However, the research on varying working environments and operating conditions are quite limited in the existing FDP works.

- Data collection and data volume

  With the development of sensing techniques, systems can be monitored in different degrees with different sensors. The explosion of monitoring data also brings challenges for bearing FDP. Rich information is behind the collected

data. It is desirable to make better use of monitoring data and explore the hidden fault-related information for bearing FDP.

The above-discussed factors show challenges for the existing methods, which often cannot generate satisfying results in bearing FDP tasks. Motivated by these challenges, this thesis is conducted to improve the performance of the existing methods in the following aspects:

- Network structure and training strategy

  The structures of the existing DL-based methods are not determined in an efficient way. Most of the structures are determined by trial-and-error. This makes the structure estimation process time-consuming and difficult. There is no general and theoretical guidance for determining the DL structure. Besides, the training efficiencies of DL-based methods are significantly affected by the learning rate and activation layers. However, traditional DL methods often use constant learning rates and saturated activation layers, such as *sigmoid* and *tanh*, which lead to low diagnosis accuracy and slow convergence rate in training.

- Discriminate feature learning

  For bearing monitoring data, some features are not informative or irrelevant to faults, and they will result in low training efficiency or large diagnosis errors. This is especially true for bearing monitoring with multiple and different types of sensors mounted at different locations. The data of these sensors often show different characteristics, and each sensor provides its unique perspective of bearing health conditions, which needs different levels of attention. Unfortunately, most of the existing works lack explicit discriminate feature learning mechanisms so that they give equal attentions to all features.

This not only introduces unnecessary computation cost, but also leads to low training efficiency, especially for diagnosis with data from multi-sensors.

- System domain knowledge and operating condition integration

  Most industrial systems are working in varying operating conditions and environments. The information of operating conditions, such as load profile, rotating speed, and environmental factors, etc., and domain knowledge, such as fault mode, fault mechanisms, and fault characteristic frequencies, etc., have significant influence on the accuracy and performance of diagnosis. However, this kind of important information is not utilized in most of the existing DL-based approaches. As a result, these approaches often cannot guarantee good results for diagnosis with varying operating conditions, environments, and unexpected fault modes. To make full use of these critical information, it is desirable to integrate the operating conditions and domain knowledge to improve the performance, reliability, and robustness of diagnosis.

- Multi-task fault diagnosis

  Most bearing diagnosis methods are developed as single task methods for detecting and identifying fault modes. To improve the maintenance efficiency and reduce maintenance cost, more fault information, including fault mode, fault location, and operating condition at the time of fault detection, are essential. This leads to multi-task bearing diagnosis methods in which a single network is able to provide multiple fault-related information, such as fault mode, fault location, etc.

- Prognosis start-to-prognosis (STP) detection

  In bearing whole service life, bearings usually experience different stages, which are defined as run-in stage, steady-state stage, and faulty stage [93], as shown in Fig. 1.4. Most of the service time is in the steady-state stage, which is

a long steady process before a fault occurs. In the run-in and steady-state stages, bearings are in healthy condition. The time duration of healthy stages of different bearings vary a lot due to manufacturing, operating condition, assembly quality, etc. In steady-state stage, only on-line anomaly detection is needed, while bearing FDP is not necessary as it leads to large computation and unreliable prognostic results. However, at the end stage of their service life, bearings usually experience a fast and severe degradation phase. Therefore, it is necessary and significant to identify the bearing operating stages and accurately estimate the start-to-prognosis (STP) time instant for RUL prediction.

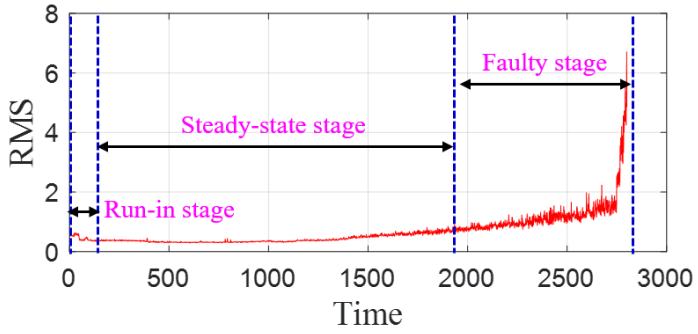

Figure 1.4: Bearing running stages

- Fault dynamic modeling, selection, and fusion

  Bearings often show different fault modes, and they present different degradation trends in the faulty stage. This makes it difficult to describe different degradation dynamics using one fault dynamic model, even with parameter adaptation. It is, therefore, desirable to select the most appropriate fault dynamic model to conduct FDP for different fault modes. However, the model selection methods in most existing works are designed using threshold setting or machine learning methods [94]. The measurements (features or fault indicators extracted from raw data) often show random fluctuations due to system noise or other interference, which may yield false detection and then

affect the accuracy of model selection. Besides, it is sometimes difficult to select an appropriate fault dynamic model for FDP, especially at the initial stage of fault. In this case, it is desirable to employ and fuse the results from multiple models to obtain robust FDP results.

1.4 Thesis organization

To address the above-mentioned challenges and open problems, this thesis aims to propose novel DL-based frameworks for bearing FDP. The thesis is organized as follows: Chapter 2 introduces the research contents and contributions; To improve the accuracy and training efficiency of the DBN based bearing fault diagnosis method, an optimized adaptive DBN method is proposed in Chapter 3; In Chapter 4, an enhanced discriminate feature learning based CNN is proposed for bearing fault diagnosis with the fusion of multi-sensor data, bearing domain knowledge, and operating conditions; Chapter 5 presents a multi-scale discriminate CNN based fault diagnosis method for the bearing under varying speed. Chapter 6 discusses a hybrid Bayesian estimation-based bearing FDP, which includes fault detection, fault model selection, PF-based prognosis, and Dempster Shafer Theory (DST) based prognosis fusion; Chapter 7 gives the conclusions and future work.

Chapter 2

Research Content and Contribution

2.1 Research content and contribution

DL-based bearing methods have made significant achievements in bearing FDP. They show powerful automatic feature extraction capability from raw data. Among the techniques, DBN and CNN are two of the most representative networks that have been widely used in industries. DBN is a representation of the feed-forward network structure, while CNN is a type of space invariant network structure. Each of them has its own unique advantages for different applications. This thesis mainly focuses on these two DL networks to improve the FDP performance in terms of accuracy, efficiency, robustness. This thesis is conducted from the following aspects:

2.1.1 Optimized adaptive PReLU-DBN for bearing fault diagnosis

DBN performs well in time series prediction, natural language processing, etc. The performance of DBN-based methods is mainly affected by the network structure, activation function, and training strategy. Based on the challenges and open problems that discussed in Section.1, the first work of this thesis is conducted to improve the performance of DBN-based fault diagnosis methods in terms of accuracy and efficiency. The research content and contribution of the improvements can be summarized as:

- Propose a hybrid flexible diagnosis framework for rolling element bearings that integrates feature extraction, dynamic training, and structure optimization.
- Propose a Parametric Rectified Linear Unit-DBN (PReLU-DBN) with dynamic training activation layers and adaptive learning rate that enables high accuracy and fast convergence speed.
- Develop an adaptive Restricted Boltzmann Machine (RBM) training procedure and Particle Swarm Optimization (PSO) based structure optimization, which achieves not only an optimal network structure, but also a fast training speed.

2.1.2 Multi-task Bearing Fault Diagnosis with Information Fusion

Deep learning-based diagnosis methods currently face some challenges and open problems. First, domain knowledge of fault modes and operating conditions is not integrated in most existing approaches, which results in low diagnosis accuracy and training efficiency. Second, existing methods treat all features with indiscriminate attention, which causes unnecessary computation and even false diagnosis results in some cases. Third, multi-task diagnosis becomes more important for health maintenance.

Based on the motivations and challenges summarized in Section 1, the second work of this thesis aims to develop an enhanced discriminate feature learning based deep residual CNN for multi-task bearing fault diagnosis with information fusion. In the proposed method, a variant structure CNN, deep residual convolutional neural network (DR-CNN), is employed [95]. For traditional CNN, when the depth of the CNN structure is increased to deal with monitoring data in complex operating conditions, it causes two problems. One is that a network with a deeper structure has many layers and trainable parameters, which makes training time-consuming

and difficult. The other one is that the performance may get saturated, which means a CNN structure with more layers cannot always lead to a better performance of diagnosis [95]. DR-CNN is used in this work as it overcomes performance degradation when the depth of the network increases.

The proposed CNN-based diagnosis method in this thesis fuses the domain knowledge and operating conditions with data from multiple sensors. Although the domain knowledge are considered in some of the existing works [96–98], these existing DL-based approaches often lack discriminate feature learning capability and do not consider human-involved feature extraction. In this work, two feature attention modules, channel attention module (CAM) and non-local attention module (NLAM), are adopted to improve the discriminate feature learning ability. Besides, the constructed and initialized network is trained with a dynamic training procedure to further improve the training efficiency. The main contributions of this work are summarized as follows:

- Explore the attention mechanisms in multi-task DR-CNN to improve bearing diagnosis performance in terms of training efficiency and accuracy.
- Domain knowledge is integrated and fused with multiple sensor data to make the diagnosis network converge in a fast and accurate way.
- Two classification tasks are achieved simultaneously by one network with a single training process.

2.1.3 Varying Speed Bearing Fault Diagnosis with Multi-scale Discriminate CNN

Most industrial bearing systems operate in varying conditions and environments, especially the operating speed and load. Compared with stationary working conditions, bearings under varying speed conditions throw more challenges for fault

diagnosis as fault-related signals will be affected by varying operating speed and load. Therefore, it is desirable to integrate the critical bearing information under varying conditions to further improve the accuracy, efficiency, and extensibility of fault diagnosis. Although operating conditions are considered in some existing works, they are designed for stationary operating conditions.

Besides, some fault-related signals of bearings under varying speeds show similar characteristics with bearing intrinsic signals, such as periodicity and repeatability. Therefore, it is more challenging for bearing fault diagnosis under varying speeds compared with that in stationary operating conditions. However, the fault-related characteristics are rarely considered in the existing methods, they often cannot guarantee satisfying results for the tasks under varying speeds.

Due to the complex working conditions, varying speeds and loads, the features of bearing faults often present unique characteristics in different spaces and scales. DL network can project bearing monitoring data into different feature spaces. However, simple projection and transform cannot guarantee the feature extraction performance for bearings under varying speeds, resulting in low diagnosis accuracy. Multi-scale analysis and feature extraction are effective methods to increase the feature learning ability of network to capture the hidden features in different spaces.

Based on the above-mentioned limitations and motivations for bearing fault diagnosis under varying speeds, this thesis proposes a multi-scale discriminate CNN-based method. The method is developed with complete consideration of bearing operating conditions and fault signal characteristics under varying speeds. The multi-scale feature extraction architecture can effectively capture the features under varying speeds. Besides, the bearing operating speed information is integrated into the input for feature extraction, and the correlations between operating speed and raw data are also extracted to enhance network feature extraction ability for bearings under varying speeds. To improve the effectiveness of feature learning and

further explore the hidden important information for fault diagnosis, the multi-head self-attention-based architecture is deevloped to learn important information in a parallel manner with discriminate attention. With the extracted multi-scale discriminate attention features, bearing fault diagnosis under varying speeds can achieve satisfying results in terms of accuracy and efficiency. The main contributions of this method are concluded as follows:

1. A multi-scale discriminate feature learning network is proposed to improve the performance of bearing fault diagnosis under varying speeds.

2. It sufficiently takes bearing fault characteristics into account. Bearing varying speed information is integrated with bearing raw data as the input of the network to achieve a high feature learning efficiency and diagnosis accuracy, which also provides a solution to integrate more system domain knowledge to get a robust fault diagnosis method.

3. The efficiency and accuracy of the proposed method are validated with two experiments with different settings. Experimental results show that it outperforms the state-to-the-art DL-based methods.

2.1.4 Hybrid bearing fault prognosis with fault detection and multiple model fusion

Accurate bearing fault diagnosis and prognosis (FDP) is critical for optimal maintenance schedules, safety, and reliability. The existing methods face some problems and challenges in detecting the starting time for prognosis and using a single model to describe fault dynamics. This thesis proposes a hybrid Bayesian estimation-based bearing FDP approach. The proposed method consists of two major offline components (Health Indicator and model training) and four major online components (CNN-based STP detection, CNN-based model selection, PF-based

prognosis, and DST-based prediction fusion). This work combines the powerful feature learning and pattern recognition ability of deep learning, and state estimation and uncertainty management ability of Bayesian estimation. In this work, CNN is adopted for STP detection and model selection due to its powerful feature learning and pattern recognition ability. In the offline part, two CNN-based models are trained using bearing information maps. The fault dynamic models are built using the extracted HI. In the online prognosis process, the fault state is identified using CNN-based fault detection model. The prognosis process is executed after the detection of a fault with the integration of fault dynamic models, which are selected by model selection CNN. A Dempster-Shafer Theory (DST) based fusion mechanism is proposed to further improve the prognostic performance. The bearing state and RUL can be estimated based on PF-based estimation and fusion results. The main contributions of the proposed method are summarized as follows:

1. Propose a hybrid Bayesian estimation-based bearing FDP framework with CNN-based STP detection and model selection to improve FDP efficiency, accuracy, and applicability.

2. Build an information map that integrates continuous wavelet coefficient matrices (CWCM) power spectrum of vibration and operating condition as the input of CNN for STP detection and fault model selection to improve the accuracy and training convergence speed of CNN.

3. Design a CNN-based automatic model selection and DST-based fusion in PF-based bearing FDP to improve accuracy and robustness of prognosis.

With the consideration of the challenges in DL-based bearing FDP methods, this thesis presents new explorations of DL-based FDP methods in terms of structure optimization, adaptive training, domain knowledge and operating condition integration, fault prognostic model selection and fusion. The proposed methods are

verified accurately and effectively in bearing FDP, and can achieve better performance than the existing methods.

Chapter 3

Optimized adaptive PReLU-DBN for bearing fault diagnosis

This chapter proposes a DBN based framework that integrates principal component analysis (PCA) and Parametric Rectified Linear Unit (PReLU) activation layers as a general approach for reliable and effective bearing fault diagnosis. First, PCA is first applied to raw vibration signals to reduce the signal dimension. Then, PReLU-based DBN (PReLU-DBN) is introduced to improve the learning performance in term of accuracy and efficiency. In the training process, the evolution strategy of particle swarm optimization (PSO) with an adaptive training procedure is utilized to obtain the optimal structure of PReLU-DBN. The rest of this chapter is organized as follows: Section 3.1 presents a brief introduction of PReLU-DBN, PCA, and PSO. Section 3.2 provides details of the design and implementation of the proposed fault diagnosis method, including PSO for optimal DBN structure and adaptive training process. Section 3.3 presents experiments and analysis of results to demonstrate the effectiveness of the proposed approach, which is followed by concluding remarks in Section 3.4.

3.1 Theoretical background

This section briefly introduces the principles of PCA, PReLU-DBN, adaptive training strategy, and PSO optimization for the development of the proposed approach.

3.1.1 Principle Component Analysis

Principal component analysis is a traditional statistical analysis approach that is widely used in data dimension reduction, feature extraction and fusion, high-dimensional data visualization, and data regression, etc. PCA can be regarded as a transform that projects the original data to a new space with lower dimension. Given original data vector x_i ($i = 1, \cdots, m$), the covariance matrix of the data vector can be calculated as:

$$Cov = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu)(x_i - \mu)^T \tag{3.1}$$

where $\mu = \frac{1}{m}\sum_{i=1}^{m} x_i$ is the mean value of the vector. With a n dimensional data vector, the eigenvalue of the covariance matrix can be described as:

$$\lambda_j \cdot u_j = Cov \cdot u_j \quad (j = 1, \cdots, n) \tag{3.2}$$

where λ_j are the eigenvalues of the covariance matrix sorted in descending order and u_j are their corresponding eigenvectors.

To get the first k eigenvectors ($k < n$) that corresponds to the first k largest eigenvalues, let

$$U_k = [u_1, u_2, \cdots, u_k] \tag{3.3}$$

$$\Lambda_k = diag[\lambda_1, \lambda_2, \cdots, \lambda_k] \tag{3.4}$$

The principal components of the original data can be computed as the orthogonal transformations of x_i:

$$P = U_k^T x_i \tag{3.5}$$

The obtained components P are principal components, which describe the original data vector with a reduced dimension. This work uses the characteristic of dimensional reduction of PCA to process the raw data.

3.1.2 PReLU-DBN and Adaptive Training Strategy

DBN can be regarded as a special neural network constructed from multiple RBMs [89]. Fig. 3.1 shows the schematic representation of DBN with two hidden layers. Due to this stacked structure, DBN has a strong capability in capturing representative information from raw time series data. The output of the learned information from the output layer is extracted features, which can be utilized as the input of a supervised learning algorithm in classification or regression for fault diagnosis.

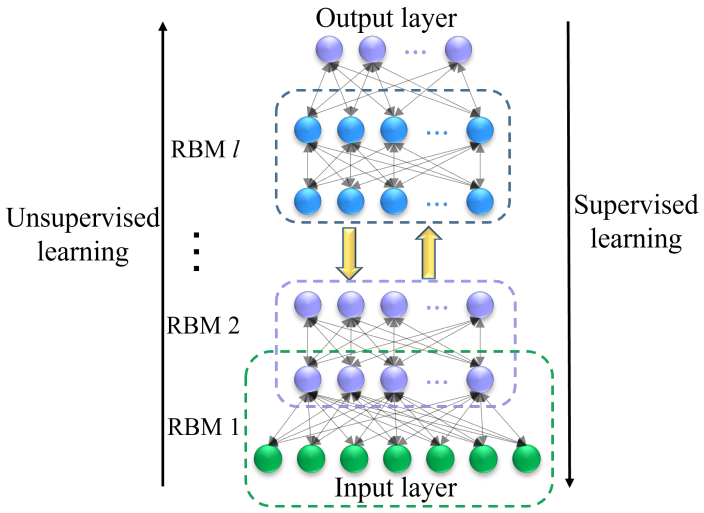

Figure 3.1: The structure of a n hidden layer DBN

In the DBN structure shown in Fig. 3.1, RBM is a special probabilistic model of Boltzmann machine consisted of a visible input vector v and a hidden vector h that are connected by weighting factors. The joint configuration (v, h) is given by the following energy function:

$$E(v,h) = -\sum_{j=1}^{m} a_j v_j - \sum_{i=1}^{n} b_i h_j - \sum_{i=1}^{n}\sum_{j=1}^{m} v_j w_{ij} h_i \tag{3.6}$$

where v_i and a_i are binary states and bias of the i-th element of the visible vector, respectively, h_j and b_j are the binary states and bias of the j-th element of the hidden vector, respectively, w_{ij} is the weight of the connection between the i-th neuro in the

visible layer and the j-th neuro in the hidden layer. The joint distribution over the visible layer and hidden units is defined as:

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

where Z is a partition function defined as:

$$Z = \sum_{v,h} e^{-E(v,h)} \tag{3.8}$$

In DBN structure, the connections just exist between adjacent layers. The neurons in the same layer are independent to each other. The conditional probabilities of the hidden layer and the visible units are given as:

$$p(h_i = 1|v) = \frac{1}{1 + e^{-b_i - \sum_{i=1}^{m} v_i w_{ij}}} \tag{3.9}$$

$$p(v_i = 1|h) = \frac{1}{1 + e^{-a_i - \sum_{j=1}^{m} h_j w_{ij}}} \tag{3.10}$$

The training is divided into two stages: pre-training and fine-tuning [99]. The forward pre-training can be regarded as a construction and reconstruction process using Eq. (3.6), which aims to extract features based on its learning rules automatically. In pre-training, the stacked RBMs are trained layer by layer using a greedy learning algorithm [99]. This is an unsupervised training process that is elaborated as follows:

Given training input data and initial DBN parameters, the hidden layer can be trained greedily by (3.9) and (3.10). This is a positive phase in which the gradient of the log probability of the given training data can be described as:

$$\frac{\partial \ log \ p(v)}{\partial \ \omega_{ij}} = \langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model} \tag{3.11}$$

where $\langle \cdot \rangle_p$ represent averages with respect to distribution p. Learning rule aims at maximizing the log probability of the data, which is equivalent to minimizing the

divergence of the distribution defined by the model and the given training data. Based on the Contrastive Divergence (CD) algorithm [99, 100], the parameters of DBN can be adjusted by:

$$\omega_{ij}^{n} = \omega_{ij}^{n-1} + \gamma(\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_k) \tag{3.12}$$

$$a_{ij}^{n} = a_{ij}^{n-1} + \gamma(\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_k) \tag{3.13}$$

$$b_{ij}^{n} = b_{ij}^{n-1} + \gamma(\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_k) \tag{3.14}$$

where $\gamma \in [0, 1]$ is the learning rate that can be used to adjust the learning speed, n is the iterations of training, k is the step of Contrastive Divergence.

After pre-training, the fine-tuning step is implemented with a back propagation algorithm [100] to optimize the parameters and structure of the pre-trained network to further improve the classification accuracy. This is a supervised learning process using a conjugate gradient algorithm and labeled data. In fine-tuning, the weights and biases of every layer are adjusted continuously at the same time until the iteration reaches the threshold. The trained DBN model after fine-tuning can then be used in describing the fault dynamics.

As mentioned early, fixed learning rate often results in slow convergence in training. To address this problem, an adaptive training method is introduced. In this method, a small learning rate is initialized and is adaptively adjusted with reconstruction errors. The learning rate is decreased (increased) when reconstruction error increases (decreases). This adaptive training method can be described as:

$$\Delta E = E(k) - E(k-1);\ E_R = |\Delta E / E(k)| \tag{3.15a}$$

$$\gamma^k = f_d \gamma^{k-1};\ \Delta E \geq 0. \tag{3.15b}$$

$$\gamma^k = f_i \gamma^{k-1};\ \Delta E < 0\ \&\ E_R < \sigma. \tag{3.15c}$$

$$\gamma^k = \gamma^{k-1};\quad \Delta E < 0\ \&\ E_R \geq \sigma. \tag{3.15d}$$

where $E(k)$ is the current epoch reconstruction error, $f_i \in [1, 10]$ and $f_d \in [0.1, 1]$ are increasing factor and decreasing factor, respectively, $\sigma \in [0.1, 10]$ is the threshold of relative error change rate.

In the training process, activation layers play key roles in learning complex functional mappings from data. The well-known activation functions like *sigmoid* and *tanh* are saturated activation functions and have been widely used in deep learning. However, their derivatives become small when the absolute value of the training input becomes large. This leads to the vanishing gradient problem [101] in back-propagation training algorithm and results in a low training speed.

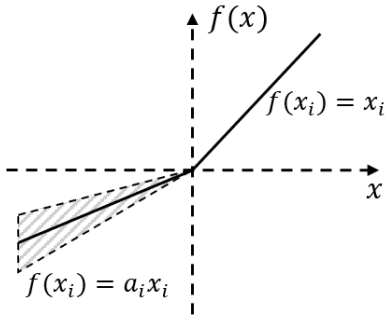

Figure 3.2: PReLU activation function

To address aforementioned problem, this chapter develops a DBN model with an adaptive activation function of PReLU. This non-saturated activation function allows to adaptively learn the slope of its input and, at the same time, contributes to accelerating the learning process [102]. The activation function of the proposed PReLU-DBN model is defined in Eq. (3.16) and shown in Fig. 3.2.

$$f(x_i) = \begin{cases} a_i x_i, & x_i < 0 \\ x_i, & x_i \geq 0 \end{cases} \tag{3.16}$$

where a_i are learnable parameters for all channels of the input training data. Its

gradient is given as:

$$\frac{df(x_i)}{dx_i} = \begin{cases} a_i, & x_i < 0 \\ 1, & x_i \geq 0 \end{cases}; \quad \frac{df(x_i)}{da_i} = \begin{cases} x_i, & x_i < 0 \\ 0, & x_i \geq 0 \end{cases} \tag{3.17}$$

3.1.3 Particle Swarm Optimization

The performance of DBN based fault diagnosis depends heavily on its structure. In order to obtain an optimal structure with high accuracy and convergence rate of training, a PSO based structure optimization approach is proposed. PSO is an evolutionary algorithm that has been widely used for optimization of weights and architecture of machine learning.

PSO simulates the process using a set of particles with different positions and velocities. The position of each particle corresponds to a fitness degree, which is used to evaluate the performance of the current structure. In this work, the fitness degree g is defined as the error of fault classification, which is given as:

$$g = 1 - A \tag{3.18}$$

where A is the accuracy of diagnosis defined as the ratio of correct diagnosis samples over all testing samples. With this fitness function, particle i is defined by position P_i^k and velocity V_i^k, which can be updated as follows:

$$V_i^{k+1} = \omega V_i^k + c_1 r_1 (p_{B_i}^k - P_i^k) + c_2 r_2 (g_{B_i}^k - P_i^k) \tag{3.19}$$

$$P_i^{k+1} = P_i^k + V_i^{k+1} \tag{3.20}$$

where ω, c_1, and c_2 are inertia weight, personal learning coefficient, global learning coefficient, respectively, k represents the k-step of PSO, r_1 and r_2 are random values

between 0 and 1, p_B and g_B denote the best particle position and best group position, respectively.

Table 3.1 outlines the pseudocode of the PSO algorithm. First, the number of decision variables and fitness function are defined based on the bearing diagnosis problem. The decision variables are the numbers of the hidden layers of the DBN, and fitness function is defined in (3.18). Then, PSO is randomly initialized with a swarm including N particles based on the defined encoding strategy. The particles start to evolve until the process reaches the pre-defined threshold, which is defined as the error of the diagnosis with the evolved structure. In this step, g_B and p_B are updated based on the fitness degree. At the end of optimization, the optimized DBN structure p_B can be obtained, while g_B particle can be picked up for estimating the final performance of the optimized variables.

Table 3.1: PSO algorithm

Algorithm: Particle Swarm Optimization

Input:

Number of decision variables (swam size);

Lower and Upper bound of variables;

Maximum generation;

Fitness function;

Output:

Global best position;

Problem definition:

Define decision variables;

Define fitness function;

Parameter initialization of PSO:

for i to Number_Particle **do**

Initialize position and velocity;

Evaluate fitness value of each particle;

Determine global best position;

end

Evolution:

for i to Max_Iteration **do**

for j to Number_Particle **do**

Update velocity;

Update position;

Evaluate the fitness degree(classification error);

Determine g_B from PSO_Swarm;

Determine p_B from PSO_Swarm;

end

end

Return p_B and g_B;

3.2 Overview of the proposed solution

Fig. 3.3 shows the structure and information flow of the proposed PReLU-DBN-based fault diagnosis solution. It consists of two parts: an offline training part and an online test part. The offline training part determines the optimum adaptive DBN structure to maximize the performance and efficiency. The online testing part takes real-time data for online diagnosis. Detailed implementation process is described as follows:

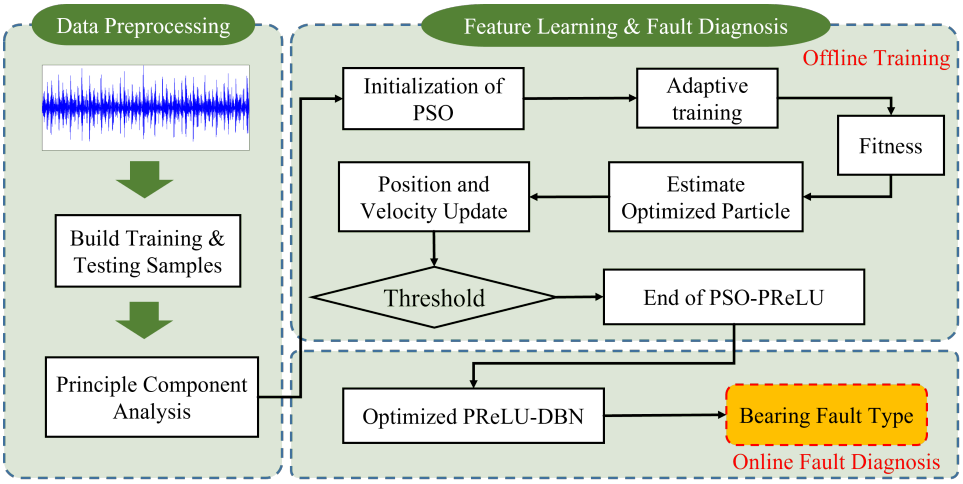

Figure 3.3: Framework of proposed approach

Step 1 Partition the time series data into segments based on sampling rate and bearing operation conditions. Then, the data segments are divided into training and testing sets.

Step 2 Apply PCA on raw signal segments $[x_1, x_2, ..., x_n]$ to reduce their dimensionality and obtain low-dimensional features $[z_1, z_2, ..., z_m], m < n$. Note that this step also generates a projection matrix that will be used in testing.

Step 3 Initialize the structure of PReLU-DBN and use PSO algorithm to optimize the number of neurons in each hidden layer. In this step, the adaptive training strategy is integrated into the PReLU-DBN training process to evaluate the fitness degree of particle swarm.

Step 4 Terminate the training process when the performance meets pre-defined requirements or the iteration number reaches the threshold. The optimized PReLU-DBN structure can guarantee high accuracy and fast convergence rate.

Step 5 Use testing data set to verify the trained optimal PReLU-DBN model. The performance on the test data is analyzed and compared with other DBN structures with different activation layers.

Compared with existing methods, this proposed PReLU-DBN based bearing diagnostic approach has an efficient learning efficiency and a high diagnosis accuracy. It avoids complex signal processing and minimizes human involvement, which makes it more generic, robust, and easier to be extended to other applications.

3.3 Experimental Results

This section presents a case study of bearing fault diagnosis and comparison with some existing methods to demonstrate the performance and efficiency of the proposed method.

3.3.1 Data Description and Pre-processing

The structure of the rolling element bearing in this study is shown in Fig. 3.4 [4]. Table 3.2 lists the main geometric parameters. The data was collected from accelerometers with a sampling rate of 50 KHz. The experiments were conducted with different fault sizes under three different rotating speeds and three different loads, which makes nine different data sets for each fault size.

Table 3.2: bearing geometric parameters

Parameter	Pitch Diameter	Roller Number	Roller Diameter	Contact Angle
Value	5.715 cm	19	0.784 cm	13.13°

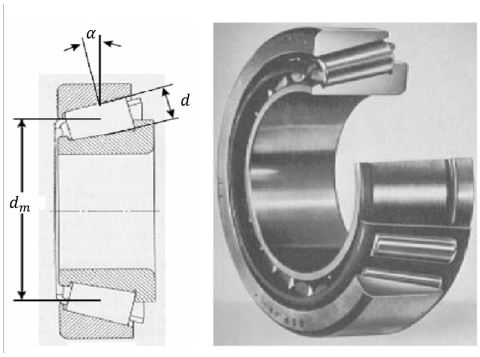

Figure 3.4: Tapered Roller Bearing Structure and Photo[4]

Table 3.3: Experimental data description

Test No.	Width (micron)	Depth (micron)	Fault	Speed (RPM)	Load (PSI)
C0	0	0	0		
C1	35.33	2.46	37.79	800	200
C2	37.67	10.56	48.23		
C3	48.33	2.38	50.71	1200	400
C4	49.33	4.88	54.21		
C5	61.00	5.80	66.8	1600	600
C6	64.00	11.00	75		
C7	131.3	1.40	132.7		
D1	64	11	75	800	200
D2	64	11	75	1200	400
D3	64	11	75	1600	600

The fault dimensions are measured by depth and width in microns. The overall fault size is defined as the sum of width and depth. For example, the bearing C1 in Table 3.3 with fault dimension width of 35.33 and depth of 2.46 has an overall fault dimension of 35.33+2.46=37.79. Based on overall fault dimension, fault modes are defined from F-1 (Health) to F-6 as shown in Table 3.4.

To capture the diagnostic information in each data segment, it should include at least one full rotation of the bearing at the lowest rotating speed. For this reason,

according to the lowest rotating speed of 800 rpm and sampling rate of 50 KHz, the size of data samples is set as 3750. This results in 1575 samples for each bearing fault mode, and 6 classes of bearing data in each data set. To train the DBN in a fair way, each raw data set is separated into several segments, which are randomly selected to construct the training and testing sets.

Table 3.4: Fault mode discretization

Dataset No.	Fault Dimension	Mode
C0,G2	0	F-1
C1	37.79	F-2
C2,C3,C4	48.23∼54.21	F-3
C5	66.8	F-4
C6,D1,D2,D3	75	F-5
C7	132.7	F-6

Table 3.5: Parameters of PSO

PSO Parameter	Value
Personal learning coefficient	1.5
Global learning coefficient	2
Inertia weight damping ratio	0.99
Population size	10
Inertia weight	1
Maximum/Minimum velocity	4.7/-4.7

Due to high dimensionality and large size of input samples, PCA is applied to the original signal for dimensionality reduction. The reduced dimension size of PCA is set as 50. The input size changes from 3750 to 50. This step can also be regarded as feature fusion, which greatly reduces the computational cost for the following training process and improve the efficiency of the structure optimization process.

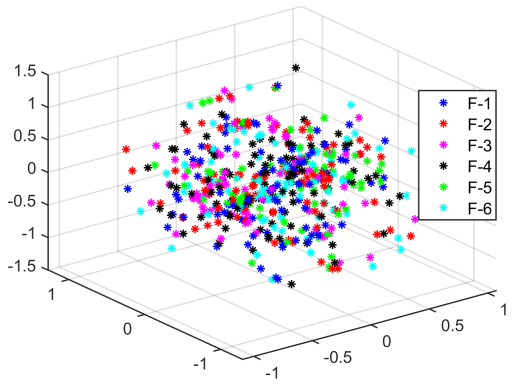

Figure 3.5: 3D-PCA visualization of raw data

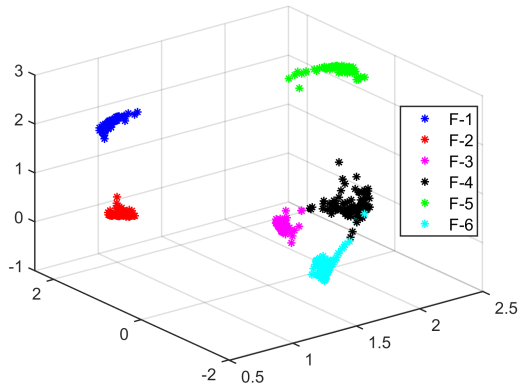

Figure 3.6: 3D-PCA visualization of extracted features

To demonstrate the feature extraction capability of DBN, the raw vibration signals and the processed data at each layer are visualized. To make the visualization clear, the first three main components are shown in the 3-D space. Figs. 3.5 and 3.6 visualize the first three main components of the raw data and the extracted features from DBN, respectively. The comparison shows that the extracted feature can better describe fault conditions and is easy to classify.

3.3.2 DBN structure optimization

PSO algorithm is used to determine the optimal structure of DBN, which leads to increase of diagnosis accuracy and decrease of training time. In this work, the defined DBN structure consists of three RBMs, denoted as RBM_1, RBM_2, RBM_3, and they have n_1, n_2, n_3 neurons, respectively. The decision matrix of PSO is designed as $X(n_1, n_2, n_3)$. Then, PSO is performed to find the optimal structure of DBN as described in Table 3.1. By following the evolution strategy of PSO, the positions and velocities of particles can be updated through the quantitative comparison to its fitness degree.

In the designed PSO algorithm, PSO related parameters are determined trial-and-error. The parameters are given in Table 3.5. In order to intuitively

investigate the efficiency of the PSO-DBN algorithm in optimizing DBN structure, Fig. 3.7 shows the evolutionary trajectories of the optimization process. It is clear that the fitness can converge to 0. In this process, the swarm may be trapped into a local minimum, which will lead to a stepwise decrease of the fitness. The positions of particles denote the optimum value of the decision matrix. From the evolutionary process of the positions, the optimal positions of particles are [500, 362, 147]. Therefore, the optimal DBN structure given by PSO is [500, 362, 147], i.e., n_1 = 500, n_2 = 362, and n_3 = 147.

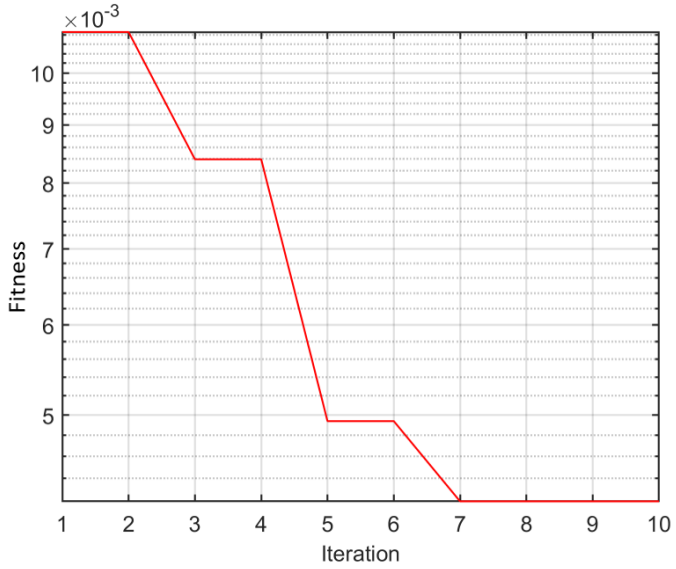

Figure 3.7: PReLU-DBN structure optimization

In the DBN structure optimization process, an adaptive training method is introduced to improve the training convergence speed. In the training process, learning rate, iteration number of RBM, and fine-tuning epoch are set as 0.05, 50, 30, respectively. Figure 3.8 describes the training process. At the end of training, the training accuracy of the proposed adaptive training approach achieves 100%. It is worth noting that the classification accuracy of the adaptive training converges to a constant at about 20 iterations. After the diagnostic model is established, it is used to testing data.

To further demonstrate the effectiveness of the proposed fault diagnosis approach, 3-fold cross-validation (CV) is conducted, in which the data are equality divided into three datasets {A, B, C}. Each dataset has 4050 samples. Figure 3.8 shows the training process of 3-fold CV. At each time, two are used for training and the last one is used for testing. The results show that the training performance of 3-fold CV is comparable, which indicates the proposed method is robust.

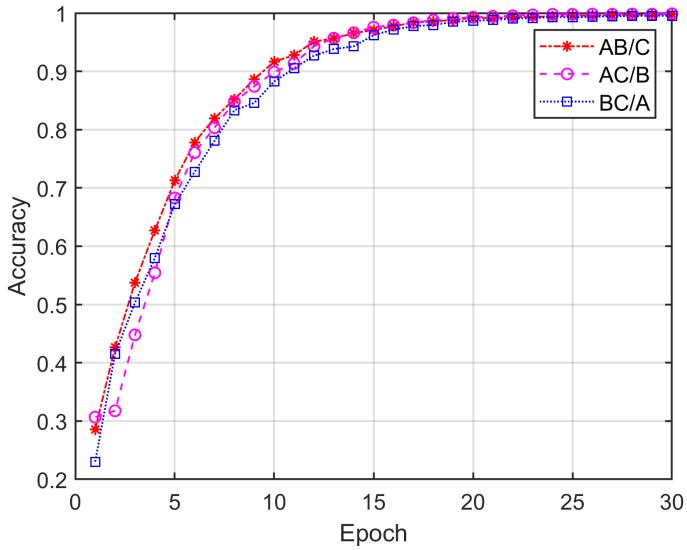

Figure 3.8: Accuracy of training process

3.3.3 Results discussion and comparison

3.3.3.1 Accuracy

To verify the effectiveness of the training process of the proposed approach, it is compared with other DBN structures with different activation layers, which include *sigmoid*, *tanh*, *ELU*, and *ReLU*.

Fig. 3.9 shows the accuracy of the training process of different DBNs at different training epochs. It is clear that the proposed approach has the best diagnosis accuracy. More importantly, the results show that the proposed PReLU-DBN has

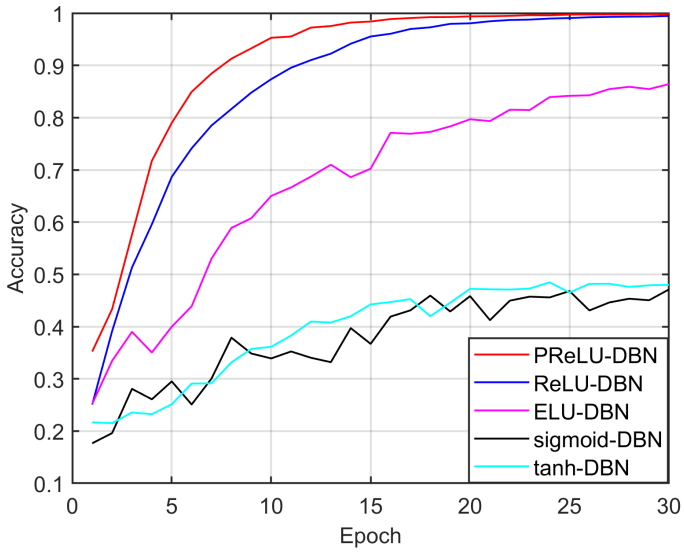

Figure 3.9: Comparison of DBN structures with different activation layers

the fast training speed and it converges at epoch 20. ReLU-DBN converges to similar accuracy at epoch 25, which is a little slower than the proposed PReLU-DBN. All other DBNs have much slower convergence speed and lower accuracy.

The quantitative comparison results of 3-fold CV diagnosis accuracy are also tabulated in Table 3.6. The results show that the average diagnosis accuracy of the proposed method can reach 99.73%, which is higher than all other DBN structures.

Table 3.6: Classification accuracy of 3-fold CV

Training /Testing	Sigmoid-DBN	tanh-DBN	ELU-DBN	ReLU-DBN	**Proposed**
A,B/ C	46.74%	67.62%	79.65%	99.4%	**99.78%**
A,C/ B	45.92%	70.61%	33.13%	99.35%	**99.88%**
B,C/ A	45.33%	62.07%	36.09%	98.64%	**99.53%**
Average	46.00%	66.77%	49.62%	99.13%	**99.73%**

3.3.3.2 Convergence analysis

To show the training efficiency of the proposed approach, Fig. 3.10 compares the convergence curves of PReLU-DBN with different DBN structures. The result shows

the errors of training processes after each training epoch. It is clear that the training process of the proposed PReLU-DBN is quick and smooth and converges to the highest accuracy with the fastest convergence rate at about epoch 25. Moreover, the convergence curve does not fluctuate, which indicates that the proposed PReLU-DBN has good stability.

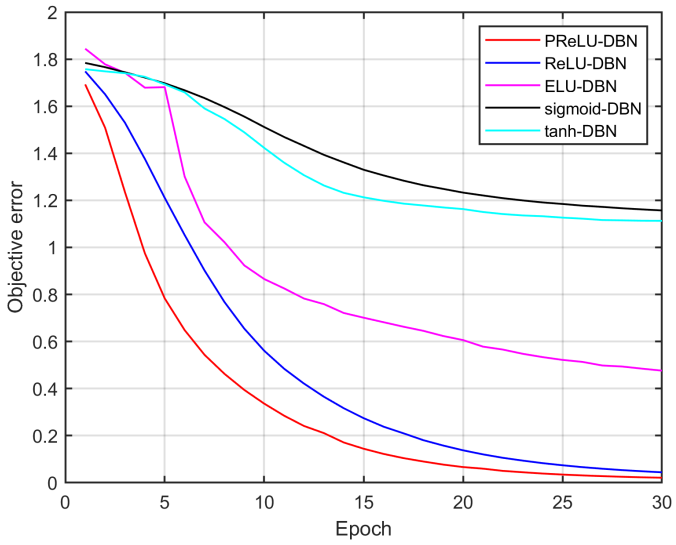

Figure 3.10: Comparison of Error of training process for DBN structures with different activation layers

3.4 Conclusions

This chapter proposes a DBN-based approach, which integrates PSO and adaptive PReLU-DBN, for accurate and efficient rolling element bearing diagnosis. In the proposed method, PCA is applied to the raw data to reduce the dimension of the input data. The optimal structure of DBN is determined by PSO. In the optimization of PSO, a PReLU activation layer and an adaptive training strategy are developed to speed up the training process. Experiments on a tapered roller bearing are conducted to demonstrate the effectiveness and efficiency of the methodology. Experimental results show that the proposed approach is effective for accurate

bearing fault diagnosis. It avoids manually feature extraction and selection that requires extensive human involvement. Compared with traditional DBN structures, the proposed PReLU-DBN based fault diagnosis approach is able to achieve higher accuracy and faster convergence speed. More important, the proposed approach is a generic solution that can be applied to a variety of systems.

Chapter 4

Enhanced Discriminate Feature Learning Deep Residual CNN for Multi-task Bearing Fault Diagnosis with Information Fusion

This chapter proposes an enhanced discriminate feature learning based deep residual CNN for multi-task bearing fault diagnosis with information fusion. The rest of this section is organized as follows: Section 4.1 provides details of the design and implementation of the proposed method, including information fusion, discriminate feature attention, and multi-task fault diagnosis. The theoretical knowledge of the proposed approach is detailed in Section 4.2. Section 4.3 presents experiments, analysis, discussion, and comparison studies to demonstrate the effectiveness of the proposed approach, which is followed by concluding remarks and future works in Section 4.4.

4.1 Implementation Procedure

Fig. 4.1 shows the framework of the proposed multi-task CNN based fault diagnosis method with enhanced discriminate feature learning. The proposed method consists of data pre-processing, information fusion, feature learning with discriminate feature attention mechanisms, and fault diagnosis. The detailed implementation process is described as follows:

Step 1 Partition the time series data into segments based on the sampling rate

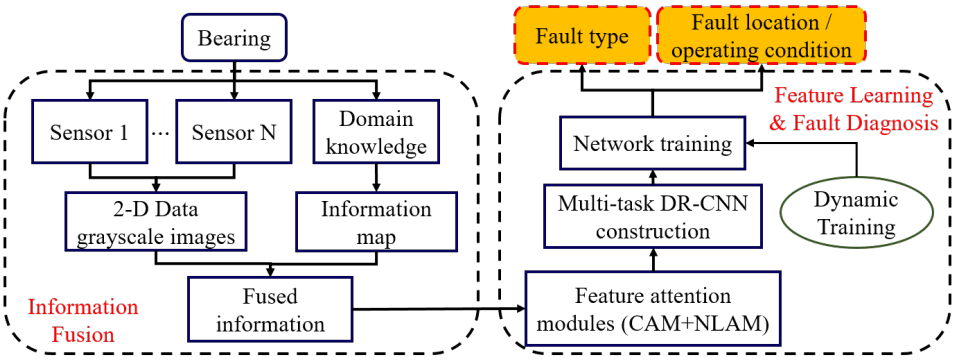

Figure 4.1: Framework of the proposed approach

and the bearing rotating speed. The data segments are then divided into training and testing sets.

Step 2 Convert the bearing monitoring data segments from different sensors into 2-D (dimensional) grayscale images and transform the domain knowledge into information maps. The data images are then integrated with information maps to build the fused information images, which are used as the input of the proposed multi-task CNN.

Step 3 Construct the discriminate feature attention (Channel Attention Module (CAM) and Non-local Attention Module (NLAM)) based DR-CNN network with multiple classifiers for multi-task diagnosis. The structure and parameters of the network are first initialized based on the input and tasks, and then trained with the fused information images by using a dynamic training strategy.

Step 4 Terminate the training process when the performance meets the pre-defined requirements or the iteration number reaches the threshold.

Step 5 Verify the trained network on the test data sets. Performance is analyzed and compared with other state-of-the-art approaches to demonstrate the effectiveness of the proposed method.

4.2.1 Discrimination Feature Attention

4.2.1.1 Channel Attention Module (CAM)

Convolution operation can be used to recognize fault features from the input images or feature maps from previous layers in different scales by different kernels. A wealth of feature information is provided in bearing data. However, some features are not related to the fault or even indicate false information due to interference signals. To address this problem, CAM is introduced to enhance the fault related information extraction ability. In CAM, feature attention map are obtained by exploring the inner-channel relationship of features [103].

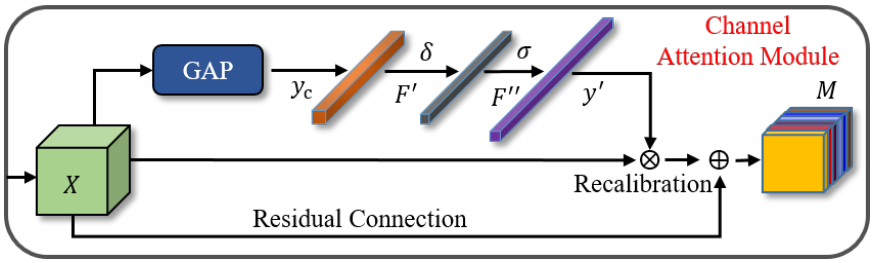

Figure 4.2: The Channel Attention Module

Fig. 4.2 shows the structure of CAM. The main idea is to distinguish different features among different channels of convolution operation and improve the network sensitivity by explicitly modeling the importance of channels. In this figure, X is a combination of convolution channels, which is denoted as $X = [x_1,\ x_2,\ \cdots,\ x_c]$, where x_i is the convolution map obtained by the i-th convolution channel and c is the number of convolution kernels. Note that the convolution map X is obtained by applying convolution operation on the input images, which are the combination of bearing monitoring data, operating conditions, and domain knowledges. Convolution operation is used to extract features in different scales from the input information

images. The features are compressed by a global average pooling (GAP) layer with the pooling filter size $H \times W$, where H and W are the width and height of the convolution map x_i, respectively. The channel-wise statistics vector y_c can be obtained as:

$$y_c = GAP(x_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i,j) \tag{4.1}$$

The channel recalibration weight vector y' is defined as:

$$y' = \sigma(F''(\delta(F'(y)))) \tag{4.2}$$

where δ is the Rectified Linear Unit (ReLU) activation function, σ is the Sigmoid function, F' and F'' are the convolution operations with convolution kernel size of 1×1. The output vector can be regarded as channel-wise dependencies with the size of $1 \times 1 \times c$, which are used to indicate the importance of each channel of the convolution map. The recalibrated feature map M can then be obtained by:

$$M = x \cdot y' = [x_1 y'_1, x_2 y'_2, \cdots, x_c y'_c] \tag{4.3}$$

In the obtained recelibrated feature map M, the discriminate feature information is highlighted by assigning the weight vector y'. In this module, the GAP is performed on all convolution channels of X, some of the useful information will also be diminished. To retain the original information, residual connection is introduced, which also improves the effectiveness of this design.

4.2.1.2 Non-local Attention Module (NLAM)

In bearing diagnosis, each input information map is converted and built from a segment of bearing data. Some of the features in the data show long-range dependencies, which include intrinsic bearing operating signals, fault related impulse excitations, system periodic signals, etc. Since these bearing signals are not short burst signals, and the bearing data are collected during a long time range, the

features are long-range dependent. Different fault modes show different long-range dependencies. In the output of NLAM, the long-range dependent features, such as some periodic excitation signals caused by faults and operation behaviors, are emphasized, while the useless information are suppressed. Enhancing the long-range dependencies related feature learning ability is critical for improving bearing fault diagnosis efficiency. Besides, the bearing data are always collected with noises, which significantly affect the performance of bearing diagnosis. How to reduce the effect of noise involvement in the learning process of DL is also important for bearing fault diagnosis.

To capture long-range dependencies and reduce the effect of data noise in DL, different non-local blocks are proposed and analyzed in [104]. The non-local blocks are efficient in capturing useful information and omitting the effect of noise. In this work, a simplified NLAM is adopted to construct the network, through which the fault feature learning ability can be improved with slight increase in computation cost. The NLAM is assigned following CAMs to enhance the network learning ability of this kind of long-range dependencies features.

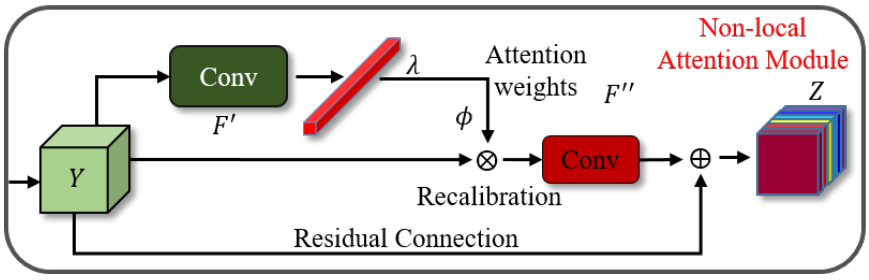

Figure 4.3: The Non-local Attention Module

The core idea of NLAM is non-local operation, which computes a weighted mean of all signal points from different sensors in the input of NLAM module Y [105]. In bearing diagnosis tasks, the fault related features are supposed to exist in each sample. The non-local operation is used to capture this kind of fault related features

that do not just exist in some bearing samples. The long-range dependencies can be captured based on the appearance similarity, which is estimated by computing interactions between features at any two positions. The receptive field of the NLAM is the whole input. The general non-local operation can be defined as:

$$\phi_i = \frac{1}{C(y)} \sum_{\forall j} f(y_i, y_j) g(y_j) \tag{4.4}$$

where i is the feature index of an output position whose response is to be computed, and j is the index that enumerates all possible positions, f is the function that is used to calculate the relationship between y_i and all y_j, and $C(y)$ is the normalization factor.

In the simplified NLAM, the attention weight ϕ is obtained by F' and $softmax$ function. The simplification details can be found in [104]. The attentional feature maps are obtained by coupling the attention weight ϕ to the input of NLAM module Y by matrix multiplication. The transformed features are aggregated to the input of each position by element-wise addition operation. The residual connection enables inserting the attention module to any network without breaking or losing any original information. Then the network feature learning efficiency and reliability can be enhanced.

Fig. 4.3 illustrates the structure of NLAM in which the input feature maps Y are flatten by 1×1 convolution kernels. The attention weight vector ϕ of non-local channels is defined as:

$$\phi = \lambda(F'(Y))) \tag{4.5}$$

where λ is the $softmax$ function, $F^{'}$ is the convolution operation with a 1×1 kernel. The enhanced non-local fault related features can be obtained by:

$$Z = Y + F''(Y \cdot \phi) \tag{4.6}$$

where Z is the output of the attention module, F'' is the convolution operation with 1×1 kernel, which is used to transform the size of the obtained attentional feature map to the size of the NLAM module input Y.

4.2.2 Deep Residual Convolutional Neural Network

CNN is a variant structure of neural networks, which is mainly composed of alternating convolutional layers, pooling layers, activation layers, and a fully connected layer. CNN is designed to learn spatial hierarchies of features by applying multiple types of functional layers [106]. The convolutional layers and pooling layers are always along with the activation layers. The features are extracted layer by layer from lower to upper. The convolution operation is performed by applying filters to the input images or the output of lower layers to extract features. The pooling layer is a down-sampling process, which is usually used to compress the output feature maps from the convolutional layers. Maximum pooling is the most common used pooling technique, which uses the maximum value of a region of neurons in the lower layer. The fully connected layer is followed by the $softmax$ classifier to identify fault modes.

For applications under varying operating conditions and complex working environments, networks with deeper structures are often used to get better results for diagnosis tasks. This makes the training of the networks difficult. More important, existing research reveals that the accuracy saturates and even degrades when the depth of the network increases to a certain value [55, 95, 107]. The causes of the accuracy degradation remain an open problem. In other words, a deeper network cannot guarantee better performance.

The accuracy degradation problem in traditional CNN can be avoided by introducing residual learning units to form the so-called residual networks [107], which are also known as directed acyclic graph (DAG) networks. Residual networks have

residual connections that bypass some layers of the CNN structure. The residual connection provides an easy way to propagate the parameter gradients from the output layer to the previous layers. Based on this structure, a deeper structure can be designed to achieve high accuracy for complex diagnosis tasks.

With this consideration, this work adopts the residual learning unit (RLU) consists of two convolutional layers, two Batch Normalization (BN) laybers, and one ReLU activation layer, as shown in Fig. 4.4, to construct the residual networks. Note that a shortcut pathway is employed in the structure to connect the input and the output of the stacked layers directly. The saturation problem of training very deep networks can be alleviated with the shortcut path structure. With this structure, the integrated RLU can be defined as:

$$y = F(x, W_i) + x \tag{4.7}$$

where x and y are the input and output of the residual structure, respectively, F is the residual function, which represents the residual mapping to be learned, $F(x, W_i) + x$ is operated by a shortcut connection and element-wise addition.

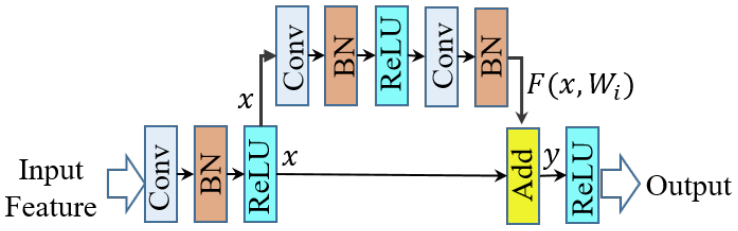

Figure 4.4: The Residual Learning Unit

4.2.3 Data Conversion and Information Map Construction

4.2.3.1 Data conversion and combination

The raw bearing monitoring data is 1-D time series data. To meet the requirements of the inputs of proposed network, the first task is to convert the data to 2-D images.

In the conversion process, the raw data are randomly divided into training set and testing set with the length of M^2, where M is the length and width of the converted grayscale images. The sample length is estimated based on the sampling rate and the bearing rotating speed with the principle of including all sampling data in one rotation of the bearing. Fig. 4.5 describes the data conversion process. For each segment of data sample, a sliding window is applied on the raw data. The gray pixels of the first segment M_1 can be calculated based on Eq. (4.8). The gray pixels are then used as the first row of the grayscale image. Based on this conversion procedure, the segments of the raw data can be converted into grayscale images with the size of $M \times M$. The pixel intensity values of the images are obtained by normalizing the raw data to the range of $(0, 255)$. By defining $I(i, j)$ as the raw time series data, the pixel intensity value of the images $G(i, j)$ can be calculated as:

$$G(i,j) = round\left\{I(i,j) - I_{min} \times \frac{255}{I_{max} - I_{min}}\right\} \tag{4.8}$$

where $i = 1, \cdots, M$, $j = 1, \cdots, M$ is the coordinate of the images, I_{max} and I_{min} represent the maximum and minimum of the raw data I, and $round(\cdot)$ is the round function.

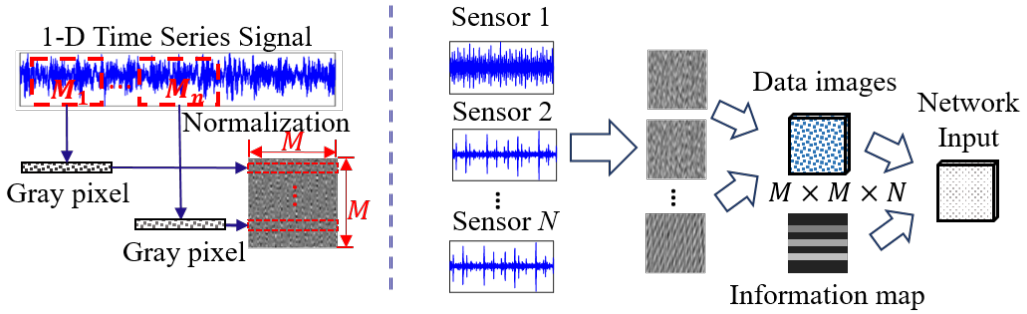

Figure 4.5: Data Conversion and Combination Procedure

The data from multi-sensors are used to improve the diagnosis performance. For each fault mode, the raw data from N sensors are used in this work, the data images can be built with the size of $M \times M \times N$ as Fig. 4.5.

4.2.3.2 Information map construction

Domain knowledge is the knowledge of a specific, specialized discipline or field [96]. In bearing fault diagnosis, domain knowledge includes the system mechanisms and expert empirical knowledge, such as bearing rotating mechanism, fault characteristics, and some common fault patterns that can be extracted from historical monitoring data, etc. These domain knowledge are critical information for diagnosis in practical applications. The powerful feature learning and fusion ability of deep learning based approaches provide a solution for integrating the complex heterogeneous data in diagnosis process.

The bearing operating conditions, such as rotating speeds and loads, have great effects on the monitoring data in terms of signal energy, amplitude, frequency, and noise term. These effects throw some challenges on fault diagnosis, especially for systems that operate in varying conditions and working loads. Therefore, it is desirable to integrate the information of operating conditions in the diagnosis process.

In this work, an operating information map is built to integrate the domain knowledge and operating condition information in the input of the networks. In order to keep the data format of the operating information map be consistent with the obtained data images, different operating conditions are represented with different grayscales. First, an empty information map with the size of $M \times M$ is built. The background of the map is defined as the operating condition information represented by gray levels that are estimated as:

$$o_{ij} = k^p \tag{4.9}$$

where o_{ij} is the gray value of the background at the coordinate (i, j), k is the discretized rotating speeds or load levels, $p = 1$ when k is the rotating speed level, $p = 2$ when k is the load level. To distinguish the operating conditions represented

by different backgrounds, the value of k is selected to avoid overlapping of the backgrounds with different grayscales.

Fault characteristics frequency (FCF) is one of the most important domain knowledge of rolling element bearings, which is a representation of the rolling element fault mechanism. Bearing faults can occur at inner ring, outer ring, or rolling ball. When a fault occurs, the related fault characteristics frequency will appear in the monitoring data. The FCFs can be calculated as [108]:

$$f_{IR} = f_r \times \frac{N(1 - \frac{d_1}{d_2}\cos\alpha)}{2} \tag{4.10}$$

$$f_{OR} = f_r \times \frac{N(1 + \frac{d_1}{d_2}\cos\alpha)}{2} \tag{4.11}$$

$$f_{BA} = f_r \times \frac{d_2(1 - (\frac{d_1}{d_2}\cos\alpha)^2)}{2d_1} \tag{4.12}$$

where f_r is the bearing rotating frequency, f_{IR} ,f_{OR}, f_{BA} are the fault characteristics frequency of inner race fault, outer race fault, and ball fault, respectively, N is the number of rolling elements, α is the contact angle, d_1 is the rolling element diameter, d_2 is the pitch diameter. The FCFs are represented using gray bands in the operating information map.

After estimating the FCFs, the ratio between the frequencies can be obtained as $f_{IR} : f_{OR} : f_{BA} = a : b : c$. Assume the dimension of the built operating information map is $M \times M$, the relative location of the three frequencies can be estimated on the information map. For example, the vertical location $i_{f_{IR}}$ of f_{IR} is estimated as:

$$i_{f_{IR}} = \frac{a}{a + b + b} \cdot M \tag{4.13}$$

The FCFs are integrated into the operating information map using three gray bands. Since the real FCFs may not exactly be the obtained values due to influence of operating noise, data acquisition error, and sampling error, the real FCFs may

show some deviation and vibrations. To reduce the influence and cover the FCFs as much as possible, the gray band centered at the calculated frequency values, which include the neighboring areas, are used in the information map. The built operating information map is then combined with the data images to build bearing information map with the size of $M \times M \times (N+1)$ as shown in Fig. 4.5.

4.2.4 Multi-task CNN based Fault Diagnosis

The overall architecture of the proposed fault diagnosis approach is shown in Fig. 4.6. It includes data pre-processing, discriminate feature attention, feature learning, and multi-task fault diagnosis. The domain knowledge are integrated into an information map, which is fused with bearing data images from multi-sources as the input of the network. The discriminate feature attention section is composed of a sequentially connected channel attention module and a non-local attention module, which enhance the discriminate features learning ability with different mechanisms. The DR-CNN based feature learning network mainly consists of several alternating convolutional layers, BN layers, and ReLU activation layers. In this work, the RLUs are introduced by two shortcut connections as shown in Fig. 4.4. The dropout layer is also introduced in this structure to avoid the over-fitting problem. To achieve the classification for multitasks using one single network, the network is designed with multiple classifiers after the fully connected layer. In Fig. 4.6, two classifiers, denoted as S_1 and S_2, are used as an example. The features are extracted layer by layer for both tasks. For each task, there are one independent fully connected layer and one classifier. The independent part can be trained separately based on the extracted features from the shared structure.

Using the fused information image as the input of the network, the constructed discriminate feature attention based DR-CNN network can be trained for multi-task fault diagnosis. The training process is conducted based on Adam optimization

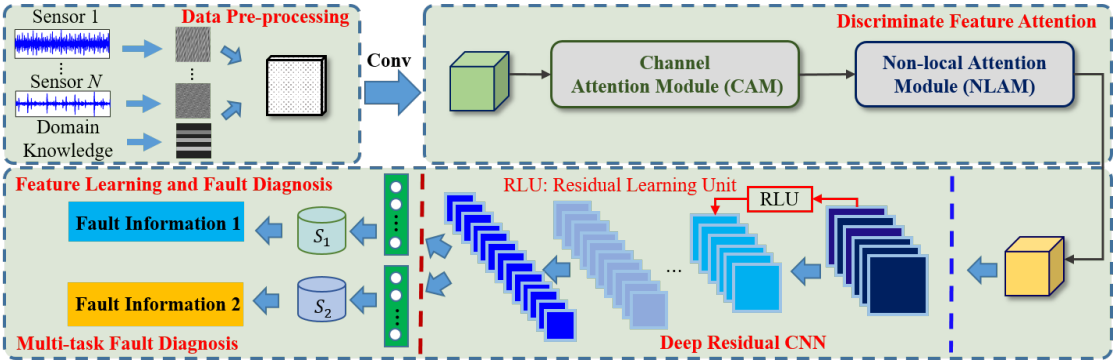

Figure 4.6: The proposed multi-task bearing fault diagnosis method

algorithm [109]. To achieve high training accuracy and quick convergence for both tasks, an exponential decay based learning rate schedule is introduced in the training process. It seeks to adjust the learning rate dynamically during the training process by reducing the learning rate according to a pre-defined schedule. The learning rate decreases to lower values as the training progresses. The dynamic learning rate can be defined as:

$$\lambda = \beta_0 \cdot exp(-\varphi \cdot t) \tag{4.14}$$

where β_0 is the initial learning rate, t is the iteration number, and φ is a hyperparameter, β_0 and φ are estimated empirically by trial-and-error based on the training performance of the initial training stage.

4.3 Experimental Results

To validate the performance of the proposed multi-task CNN based bearing fault diagnosis method, two case studies with different experiment settings are conducted. The experiments are performed in MATLAB R2020b environment running on a computer with Intel(R) Core i7-6700 CPU@3.40GHz (8CPUs) processor, 3.4GHz 16G RAM.

4.3.1 Case i: Case Western Reserve University Data

4.3.1.1 Data Description and Pre-processing

The first experiment uses the bearing data collected from the machinery fault simulator-rotor dynamics simulator[96] as shown in Fig. 4.7. Three bearings are mounted on the drive end, middle, and non-drive end of the testbed. For each bearing, three faults, which include outer race fault, inner race fault, ball fault, are injected in the bearings. The bearings are tested at ten different rotating speeds from 300 to 3000 r/min with a step increase of 300 r/min. The data are collected with a sampling frequency of 256 times the rotating speed. The bearings are tested with multi-sensors that are placed on the top of each bearing house. In this work, the data from two sensors at the drive end and middle are used.

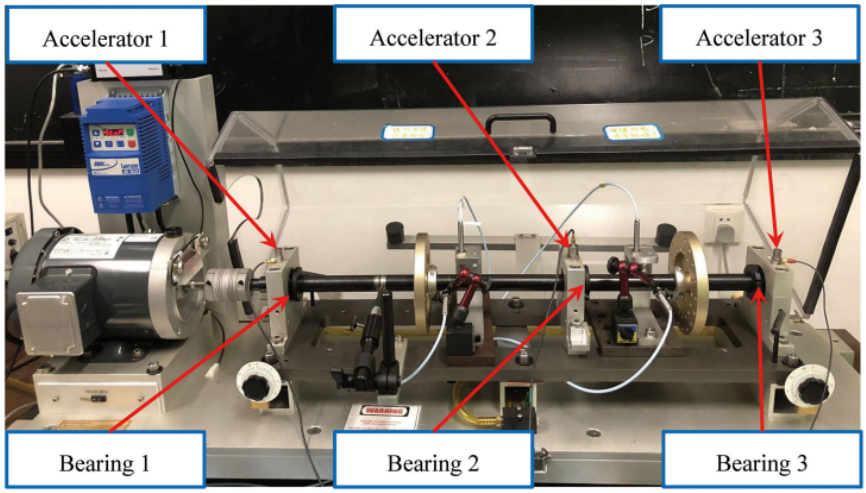

Figure 4.7: Machinery fault simulator-rotor dynamics simulator (MFS-RDS) testbed [96]

The raw data are randomly split into training sets and testing sets with a length of 2500 data points. For the data from each sensor, the 1-D time series data is converted into 2-D grayscale images with the size of 50×50. The network input is built by combining the data images with the information map. The size of the the network

input is $50 \times 50 \times 3$. For each fault mode, 1500 samples are used, and there are totally 6000 samples. To verify the accuracy and effectiveness of the proposed approach, 3-fold cross-validation (CV) is conducted. The samples are randomly divided into three datasets $\{A, B, C\}$. In this case study, Task 1 and Task 2 are assigned as fault mode diagnosis and fault localization, respectively. For each data sample, it comes with two labels. One is 4 classes of fault mode (inner-race fault, outer-race fault, ball fault, normal) while the other one is 4 classes of fault locations (drive end bearing, middle bearing, non-drive end bearing, normal).

4.3.1.2 Multi-task DR-CNN based fault diagnosis

The training process is conducted based on the Adam optimizer [110]. Table 4.1 shows the detailed structure of the discriminate feature attention based DR-CNN. The proposed network adopts one RLU whose structure is highlighted with bold text. In the training process, the initial learning rate and dropout rate are set at 0.01 and 0.3, respectively. The mini-batch size of the input is 125. The training process is terminated when the epochs reach the pre-defined maximum epoch. Then the trained network is used on testing set to validate the performance.

Figs. 4.8 shows the training progress of Tasks 1 and 2, both with comparison of results without domain knowledge and without attention modules. It is clear that the proposed approach can achieve high accuracy with fast convergence speed for both tasks, which also demonstrate the benefits of integration of domain knowledge and attention modules. The convergence of the training loss in the whole process for the 3-fold CV is shown in Fig. 4.9. Note that the convergence of the 3-fold CV processes are all stable and fast.

Table 4.2 lists the testing accuracy for the three CV combinations. The average training accuracies of both Tasks 1 and 2 can achieve 98.41% and 98.38%, respectively. To show the superiority of the proposed approach, the results are compared with

Table 4.1: Structure of the proposed network in Case I

Layer	Type	Filter size	Count	Output
1	Input	-	-	50×50×3
2	Convolution	5×5×1	100	50×50×100
3	Channel Attention	-	-	50×50×100
4	Non-local Attention	-	-	50×50×100
5	ReLU	1×1	100	1×1×100
6	**Convolution**	5×5×3	30	5×5×3
7	**Max pooling**	2×2×3	30	5×5×3
8	**BN**	-	30	5×5×3
9	**Convolution**	5×5×3	30	5×5×3
10	**Max pooling**	2×2×3	30	5×5×3
11	ReLU	-	-	56×56×3
12	Max Pooling	4×4×3	30	53×53×3
13	BN	-	-	53×53×3
14	ReLU	-	-	53×53×3
15	Dropout	-	-	1×1×100
16	Fully connected 1	1×1×100	4	1×1×4
17	Softmax 1	-	-	1
18	Fully connected 2	1×1×100	4	1×1×4
19	Softmax 2	-	-	1

Table 4.2: Accuracy of diagnosis with MFS-RDS data

Task	Training /Testing	Without Attention(%)	No Domain Knowledge(%)	Proposed(%)
Task1	$\{A, B\}\|C$	95.2	93.2	98.25
	$\{A, C\}\|B$	95.8	93.1	98.49
	$\{B, C\}\|A$	95.11	92.45	98.48
	Average	**95.37**	**92.9**	**98.41**
Task2	$\{A, B\}\|C$	94.53	96.4	98.27
	$\{A, C\}\|B$	96.9	94.26	98.14
	$\{B, C\}\|A$	95.44	96.46	98.72
	Average	**95.6**	**95.7**	**98.38**

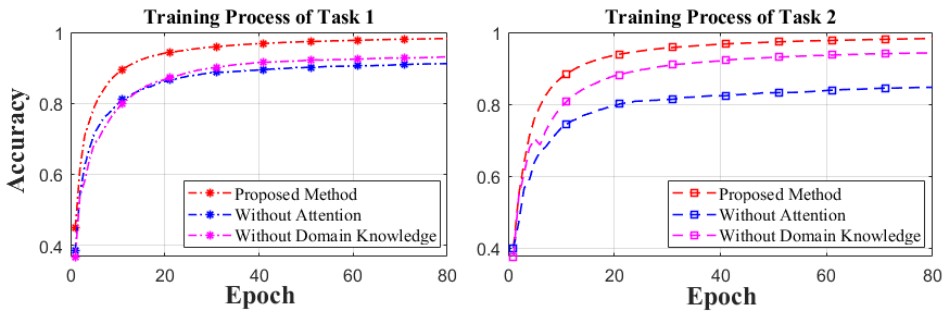

Figure 4.8: Training accuracy of multitask CNN of Case I

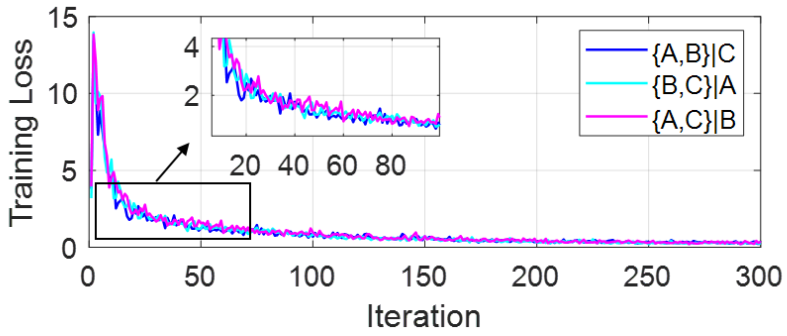

Figure 4.9: Training loss of multitask CNN of Case I,{A,B}|C indicates A and B are used as training, C is used as testing

the structure without domain knowledge and without feature attention mechanisms, respectively. Based on Table 4.2, the proposed approach can improve the accuracies by about 5.5% and 2.7%, respectively, for fault diagnosis and localization compared with the structure without domain knowledge. The comparison with the structure without attention mechanism shows that the accuracies can improve about 3% and 2.8% for fault diagnosis and localization, respectively.

The method is then compared with a latest multitask CNN based bearing fault diagnosis method [96]. The method uses bearing information map and the continuous wavelet coefficient matrices (CWCMs) of bearing data as the input of multitask CNN, in which the domain knowledge are also integrated in the diagnosis process and a dynamic CNN is applied to classify the fault modes. Table 4.3 shows the accuracies of fault diagnosis and fault location. The proposed method can improve the accuracies by about 2% in both diagnosis and localization. The results demonstrates that the

proposed method can achieve better performance than CWCM-CNN with less data processing.

Table 4.3: Comparison of multitask CNN

Method	CWCM-CNN	CWCM-CNN with information map	Proposed
Task 1 (%)	91.92	96.53	98.41
Task 2 (%)	93.09	96.36	98.38

To further validate the effectiveness of the proposed approach, it is compared with some recent DL-based approaches, which include 1-Dimensional CNN (1-D CNN) [111], hierarchical CNN (HCNN) with a matrix reconstruction method [112], deep CNN with wide first layer kernels (WDCNN) [113], dislocated time series CNN (DTS-CNN) [114]. The comparison results in Table 4.4 show that the average accuracy of the proposed approach has improvements about 20.17%, 15.61%, 11.58%, and 10.07%, respectively. Since these methods are single-task methods, only the accuracy of fault mode is compared.

Table 4.4: Comparison of Fault mode accuracy with different methods

Method	1D-CNN	HCNN	WDCNN	DTS-CNN	Proposed
Accuracy(%)	78.24	82.80	86.83	88.34	98.41

4.3.2 Case ii: Rolling Element Bearing Testing Data

4.3.2.1 Data Description and Pre-processing

The bearings in this case study are tapered ball bearings as described in Section. 3.3. In this work, the proposed approach is performed on vibration and AE acoustic emission signals. The data from X channel and Y channel of the triaxial sensors and the AE sensor are used. The multi-sensor data are converted and combined with the information map to build the input of network with the size of $60 \times 60 \times 4$. For each

fault mode, 1800 samples are used. There are totally 9600 samples in the training set and 4800 samples in the test set. The samples are also divided into three datasets $\{D, E, F\}$ to cross-verify the proposed approach.

4.3.2.2 Multi-task CNN based fault diagnosis

Using the 2-D information map built from the data of two types of sensors and the corresponding domain knowledge, the discriminate feature attention based DR-CNN is trained for multi-task fault diagnosis. In this case study, Task 1 is defined to diagnose the fault mode, while Task 2 is defined as identifying the bearing operating condition (speed) at the occurrence of the fault. The labels for Task 1 and Task 2 are fault dimension and rotating speed, which include 8 classes and 3 classes, respectively. The training procedure and parameters are consistent with that in Case i. The detailed structure is described in Table 4.5. The main differences are the input and output of the network. With the input information maps and the initialized structure, the multi-task fault diagnosis is trained. Then the trained network is implemented on the testing set to validate the performance. The training processes of Task 1 and Task 2 are shown in Fig. 4.10. Clearly, the training accuracy of both tasks converges at about 6 epochs. The training process has fast convergence and good accuracy for both tasks. The cross validation experiments indicate the robustness of the proposed approach.

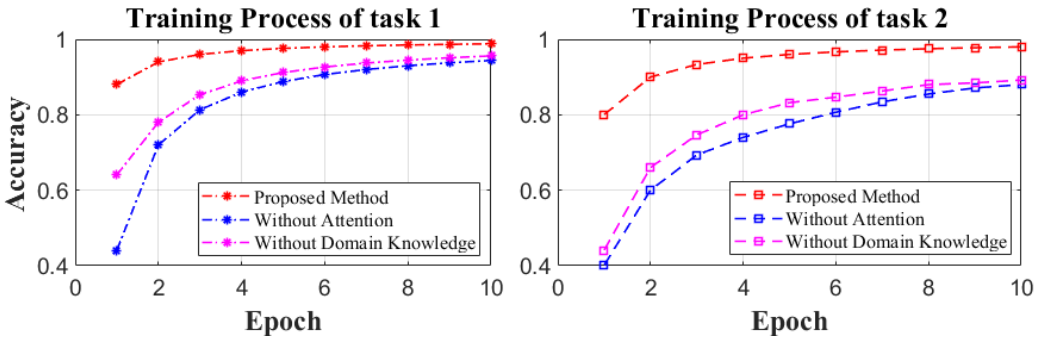

Figure 4.10: Training accuracy of multitask CNN of case 2

Table 4.5: Structure of the proposed network

Layer	Layer Type	Filter size	Count	Output
1	Input	-	-	60×60×4
2	Convolution	5×5×1	100	60×60×100
3	Channel Attention	-	-	60×60×100
4	Non-local Attention	-	-	60×60×100
5	ReLU	1×1	100	1×1×100
6	**Convolution**	5×5×3	30	5×5×3
7	**Max pooling**	2×2×3	30	5×5×3
8	**BN**	-	30	5×5×3
9	**Convolution**	5×5×3	30	5×5×3
10	**Max pooling**	2×2×3	30	5×5×3
11	ReLU	-	-	56×56×3
12	Max Pooling	4×4×3	30	53×53×3
13	BN	-	-	53×53×3
14	ReLU	-	-	53×53×3
22	Dropout	-	-	1×1×100
23	Fully connected 1	1×1×100	4	1×1×4
24	Softmax 1	-	-	1
26	Fully connected 2	1×1×100	4	1×1×4
27	Softmax 2	-	-	1

The experimental results are also compared with the structure without domain knowledge and feature attention mechanism, respectively. Table 4.6 shows the testing results of cross validation and quantitative comparisons. The average diagnosis accuracies of Task 1 and Task 2 of the proposed method can achieve 99.98% and 99.46%. Compared with the structure without attention mechanism and without domain knowledge, the proposed method has about 4% and 3% improvement in Task 1 and about 11% and 9% improvement in Task 2, respectively.

To show the convergence of the whole training process, Fig. 4.11 presents convergence curves of the 3-fold CV. It shows the overall loss after the training of each mini-batch. Based on Fig. 4.11, all the training processes are fast and smooth, the

Table 4.6: Accuracy of fault diagnosis (Fault diagnosis/Operating speed identification)

Task	Training /Testing	Without Attention(%)	No Domain Knowledge(%)	Proposed(%)
Task1(Fault)	$\{D, E\}\|F$	94.4	96.4	99.97
	$\{D, F\}\|E$	96.4	96.8	99.99
	$\{E, F\}\|D$	96.2	95.6	99.97
	Average	**95.9**	**96.2**	**99.98**
Task2(Speed)	$\{D, E\}\|F$	88	93.2	99.96
	$\{D, F\}\|E$	91.2	87.6	99.52
	$\{E, F\}\|D$	86	89.2	98.91
	Average	**88.4**	**90**	**99.46**

loss converges to close 0 at about iteration 200. The fluctuations of all convergence curves are small, which demonstrates that the proposed approach is stable and robust.

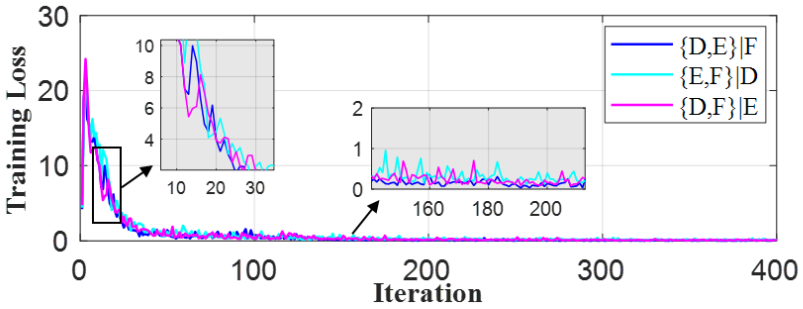

Figure 4.11: Training loss of multitask CNN of Case 2, {D,E}|F indicates D and E are used as training, F is used as testing

4.4 Conclusions

This chapter proposes a novel enhanced discriminate feature learning based DR-CNN multi-task bearing fault diagnosis. The raw data from multi-sensors are converted and fused with domain knowledge to build the information maps. Two different attention

modules are employed to enhance the fault related discriminate features learning ability. The constructed discriminate feature attention DR-CNN structure with two classifiers is assembled and trained with dynamic training procedure using the fused information maps. Two case studies are presented to verify the performance of the proposed approach. The first case study uses the data from the same type of sensors at different locations while the second case study uses the data from different types of sensors. Experimental results and comparisons show that the proposed approach can achieve high training accuracy, fast convergence speed, and high diagnostic accuracy for multitasks with a single network. The results demonstrate that the proposed multi-task bearing diagnosis network with the integration of domain knowledge and fusion of data from multi-sources is stable and robust.

Chapter 5

Varying Speed Bearing Fault Diagnosis with Multi-scale Discriminate CNN

The preceding chapters present different DL-based methods for bearing fault diagnosis. Experimental results show that the proposed methods can achieve satisfying results for bearing fault diagnosis. However, the methods are mainly proposed for bearing fault diagnosis under stationary operating speed. In practice, most industrial bearing systems operate in varying conditions and environments, especially the operating speed and load. Compared with stationary working conditions, bearings under varying speed conditions throw more challenges for fault diagnosis as fault-related signals will be affected by varying operating speed and load. Therefore, it is desirable to develop the method for bearing fault diagnosis under varying conditions to improve the accuracy, efficiency, and extensibility of fault diagnosis. Although operating conditions are considered in some existing works [96–98], they are designed for stationary operating conditions.

Besides, some fault-related signals of bearings under varying speeds show similar characteristics with bearing intrinsic signals, such as periodicity and repeatability. Therefore, it is more challenging for bearing fault diagnosis under varying speeds compared with that in stationary operating conditions. However, the fault-related characteristics are rarely considered in the existing methods. Therefore, they often cannot guarantee satisfying results for the tasks under varying speeds.

Due to the complex working conditions, varying speeds and loads, the features

of bearing faults often present their unique characteristics in different spaces and scales. DL network can project bearing monitoring data into different feature spaces. However, simple projection and transform cannot guarantee the feature extraction performance for bearings under varying speeds, resulting in the low diagnosis accuracy. Multi-scale analysis and feature extraction are effective methods to increase the depth of network to capture the hidden features in different spaces.

Bearing monitoring data contain abundant information, which includes fault-related frequency components, operating environment conditions, system intrinsic signals, unexpected external interrupt signals, and various noises. The collected data are a superposition of individual vibration signals of the system from different parts. Some signals are not informative or irrelevant to faults. They will introduce more learning loads and will result in the degradation of accuracy and training efficiency. The signals of interest to fault diagnosis are the components that are related to faults. Besides, the operation of bearings is often monitored with multiple sensors that are placed at different locations or are of different types. How to learn useful information in an effective way is critical for fault diagnosis tasks. Most existing works lack explicit discriminate feature learning mechanisms for different signals. Therefore, some unnecessary computation costs will be introduced in the learning process and the diagnosis accuracy and training efficiency will also be affected. As a result, the generalization of existing methods is often not suitable for fault diagnosis tasks under varying speeds. To deal with these challenges, it is of great importance to develop an efficient fault diagnosis method with discriminate feature learning mechanisms to fully explore the learning ability of DL networks.

With the above-mentioned limitations and motivations, this chapter proposes a multi-scale discriminate CNN-based method. The method is developed with fully consideration of bearing operating conditions and fault signal characteristics under varying speeds. The multi-scale feature extraction architecture can effectively capture

the features under varying speeds. Besides, the bearing operating speed information is integrated into the input for feature extraction, and the correlations between operating speed and raw data are also extracted to enhance the network feature extraction ability for bearings under varying speeds. To improve the effectiveness of feature learning and further explore the hidden important information for fault diagnosis, the multi-head self-attention-based architecture enables the network to learn important information in a parallel manner with discriminate attention. With the extracted multi-scale discriminate attention features, bearing fault diagnosis under varying speeds can achieve satisfying results in terms of accuracy and efficiency.

5.1 Theoretical Background

5.1.1 Convolutional Neural Network

CNN is a typical deep learning network, which has been widely and successfully used in computer vision, pattern recognition, etc. It is a representative variant of neural networks. A typical structure of CNN is mainly composed of several alternating convolutional layers, pooling layers, activation layers, fully connected layers, and output layers. CNN is designed with the idea of location connection and convolution kernel parameter sharing [115]. The convolution operation is conducted by applying convolution kernels on the given inputs (input images or the output of lower layers) to learn features in different scales. The pooling layer is performed to compress the output feature maps from the convolutional layers. The features of the input images can be extracted by convolution and pooling operation layer by layer. Classification or regression task is performed based on the extracted features. The convolution operation is conducted by applying a convolution kernel on the input image or the previous layer output to generate a corresponding feature using a nonlinear activation function [106]. The pooling layer is carried out for the down-sampling operation.

The common pooling operations include maximum pooling and average pooling. Maximum pooling is operated by finding the maximum value of a region of the feature map while average pooling is to find the average value of a region. The extracted feature is fully connected in fully connected layers to perform the classification or regression tasks.

5.1.2 Multi-Head Self Attention Module

The multi-scale feature learning network is designed based on the self-attention mechanism [116]. The attention architecture provides a way to assign different weights to input information for feature extraction, which helps the network pay close attention to important information and reduces the role of irrelevant components. Multi-head self-attention (MHSA) is designed with a parallel attention head to extract features from different subspaces or different scale feature spaces.

5.1.2.1 Scaled dot-product attention

The scaled dot-product attention (SDPA) is the core part of multi-head self-attention architecture. It's designed to capture the internal correlation of the input signals. The key idea of the attention mechanism is to map a query and a set of key-value pairs to output with attention. The quarry Q, key K, and V are obtained by linear transformation of the input signals. The output of the attention can be described as a weighted sum of the value using the attention weights, where the weight is calculated by a compatibility function of the query with the corresponding key. Dot-product attention is adopted since it is much faster and more space-efficient in practice. Fig. 5.1(a) shows the structure of scale dot-product attention with the input of Q, K, and V. The attention weights which are applied on the values and the output of scale

dot-product are computed as:

$$w = softmax(\frac{QK^T}{\sqrt{d_k}}) \tag{5.1}$$

$$Attention(Q, K, V) = wV \tag{5.2}$$

where $Q = XW_q$ are the queries matrices, $K = XW_k$ and $V = XW_k$ represent the keys and values matrices respectively, W_q, W_k, W_v are learnable matrixes, and $\sqrt{d_k}$ is a scaling factor.

5.1.2.2 Multi-head self attention

Since a single attention function just provides one projection for the input features, it is beneficial to project the input features with different projections simultaneously. The core idea of multi-head attention is to perform the scale dot-product attention for the input features with different projections in parallel, which is shown in Fig. 5.1(b). The weighted feature from each scale dot-product attention is concatenated together to generate a final weighted feature matrix as:

$$MultiHead(Q, K, V) = Concat(H_1, \cdots, H_m)W^q \tag{5.3}$$

where $H_i = Attention(Q, K, V)$, W_q is the corresponding matrix, m is the number of attention head. Multi-head attention provides a parallel feature learning structure, which has a more powerful feature attention ability than single-head attention with similar computation. In this work, the multi-head self-attention is as the hidden layers in multi-scale branches.

5.2 The Proposed Approach

5.2.1 Overall Framework

This work presents a multi-scale discriminate CNN based fault diagnosis using multi-head attention. Fig. 5.2 shows the overall architecture of the proposed

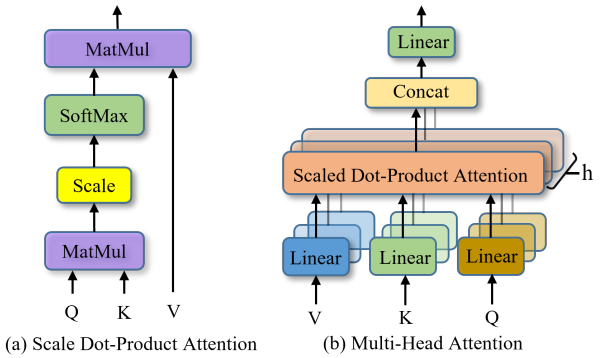

Figure 5.1: Self attention network [116]

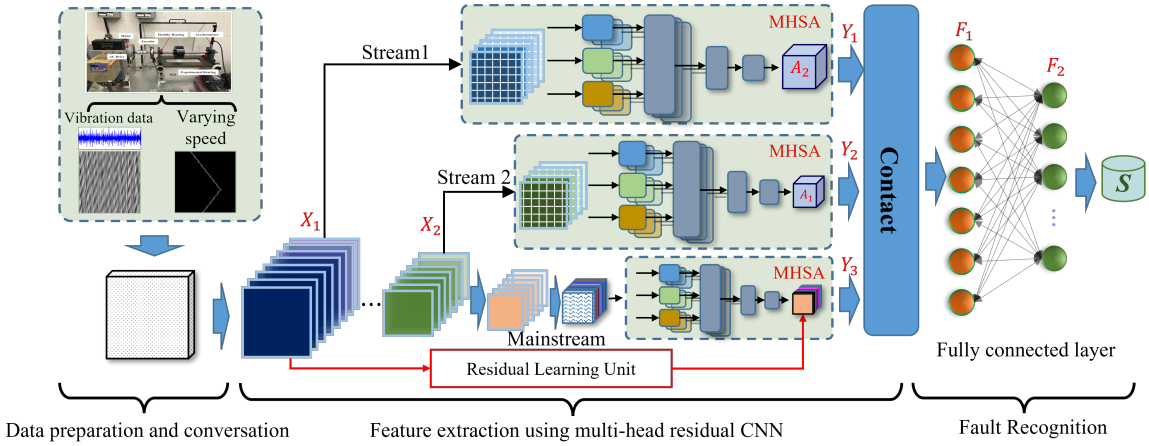

Figure 5.2: The Proposed multi-scale discriminate CNN based bearing fault diagnosis method

approach, which includes data pre-processing, discriminate feature learning, and fault diagnosis. The detailed implementation process is described as follows:

Step 1) Partition the 1-dimensional (1-D) monitoring data into segments based on the sampling rate and the bearing rotating speed. The obtained data samples are then split into training and testing sets.

Step 2) Convert raw bearing data segments into 2-dimensional (2-D) grayscale images and transform the varying speed information into information maps. The data images are then integrated with the information map to build the fused information maps, which are used for fault diagnosis.

Step 3) Design multi-scale discriminate feature learning network based on multi-head attention. The structure and parameters of the network are first initialized based on the input and tasks, and then trained with the build information images.

Step 4) If the performance satisfies the pre-defined requirements or the training epoch reaches the threshold, the training process is terminated.

Step 5) The trained network is tested using testing datasets. The advantages of the proposed network are further demonstrated with the comparisons of the state-to-art methods.

5.2.2 Data Conversion and Information Map Construction

5.2.2.1 Data conversion and combination

The raw 1-D time series vibration data are first converted to 2-D images. Fig. 5.4 is the data conversation and sample construction process. In data conversion, the raw data are divided into samples with length of M^2. To include at least one rotation of bearing data to each sample, the sample length is estimated with the consideration of the data sampling rates and bearing rotating speeds. These samples are then converted into 2-D grayscale images with the size of $M \times M$. To this end, a sliding window with length of M is employed to divide each sample into M segments. The pixel intensity values of the first segment M_1 are obtained based on Eq. (5.4), and then used as the first row of the 2-D image. Based on this conversion strategy, the other rows of the 2-D grayscale image can be obtained. In the end, a 2-D image with the size of $M \times M$ can be obtained from each raw data sample. The 2-D images are then randomly separated into training and testing sets.

The pixel intensity values of the images are calculated by normalizing the raw data to the grayscale value in the range of $(0, 255)$. By defining $G(i, j)$ as the raw

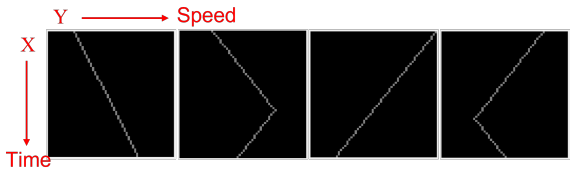

Figure 5.3: Bearing Varying Speed Information

data, the pixel intensity of the images K at the coordination (i, j), $i, j \in [1, \cdots, M]$ can be calculated as:

$$K(i,j) = round\left\{G(i,j) - G_{min} \times \frac{255}{G_{max} - G_{min}}\right\} \tag{5.4}$$

where G_{max} and G_{min} represent the maximum and minimum of the raw data G, and $round(\cdot)$ is the round function.

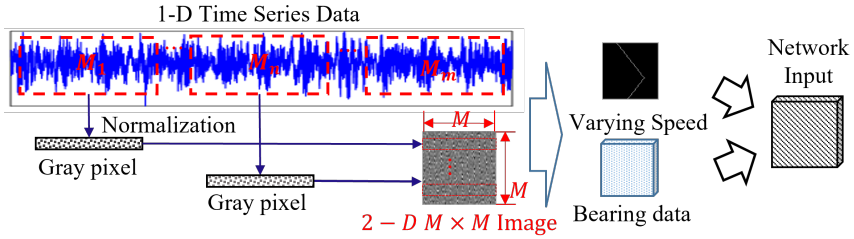

Figure 5.4: Data Conversion and Combination Procedure

5.2.2.2 Information map construction

Domain knowledge is the knowledge of a specific, specialized discipline or field [96]. In bearing fault diagnosis, domain knowledge includes fault mechanisms and expert empirical knowledge, such as bearing operating speed, rotating mechanism, fault characteristics, and some common fault patterns that can be extracted from historical monitoring data, etc. This domain knowledge is critical for diagnosis. The powerful feature learning and fusion ability of DL-based approaches provide a solution for integrating complex heterogeneous domain knowledge in the diagnosis process.

The bearing operating conditions, such as rotating speeds and loads, have great effects on the monitoring data in terms of signal energy, amplitude, frequency, and noise term. These effects throw some challenges on fault diagnosis, especially for systems that operate in variable conditions and working load. Therefore, it is desirable to integrate the information of the operating condition in the diagnosis process. Adding bearing operating information can enable the model to learn some key features for identifying faults.

In this work, an information map is built to integrate operating speed in the input to the networks. An empty information map with the same size of $M \times M$ is built. The built information maps are then combined with the data images to build the network input, which has the size of $M \times M \times (N + 1)$ as shown in Fig. 5.4.

5.2.3 Architecture of Multi-scale Discriminate Feature Extraction

The framework of the multi-scale feature extraction network is shown in Fig. 5.2. The main stream of the proposed network is an improved deep residual CNN network, which is composed of alternating convolutional and polling layers, a MHSA block, and a resiudal connection. The other two sub-streams are mainly composed of MHSA blacks. This network is designed with the fully consideration of bearing fault-related signal characteristics.

As mentioned above, the fault-related signal components are expected to contribute more to the training of the network. The most common bearing signal components are periodic impulses, which can be system intrinsic operating signals or system periodic signals. Besides, some bearing-related faults can also lead to repetitive and periodic impulses in vibration signals. Since mechanical systems are always operated in harsh working environments, the periodic impulses are often submerged by noises and it is difficult to identify these signal characteristics

from vibration signals. In addition, although the periodic signals from different sources or faults show differences in frequency and energy, they all show long-range dependencies as shown in Fig. 5.5. Therefore, enhancing the learning ability for the noise-submerged fault-related periodic impulses and long-range dependencies is critical for improving network feature extraction efficiency and fault diagnosis accuracy. The emphasis on learning these specific signal components can also reduce the computation cost of bearing data noise.

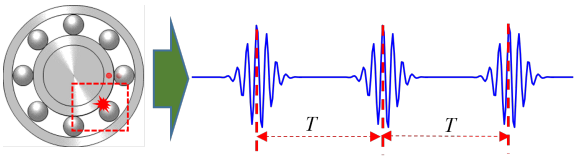

Figure 5.5: Long range dependency example

CNN extracts temporal features by scanning the input images through different convolution kernels, the different scale convolution operations allow the network to extract features in multiple scales as Eq. (5.5). The input can be projected into different spaces using convolution and pooling operations. The features from different layers can be regarded as different representations of the input.

$$S = \sum_{u} \sum_{v} X[u, v] H[i - u, j - v] \tag{5.5}$$

where X, H denote the input image and convolution kernel, i, j are the index of rows and columns of the input image. The input is projected into high-level representations in different scales by different convolution and pooling layers as shown in Fig. 5.6. However, this projection cannot guarantee a satisfying result for the diagnosis tasks of rotating machinery due to the increasing complexity of operating conditions.

The different scale features extracted in Fig. 5.6 are then used as the input for different streams, which are developed based on MHSA. In each feature extraction stream, a multi-scale and multi-head discriminate feature extraction network is

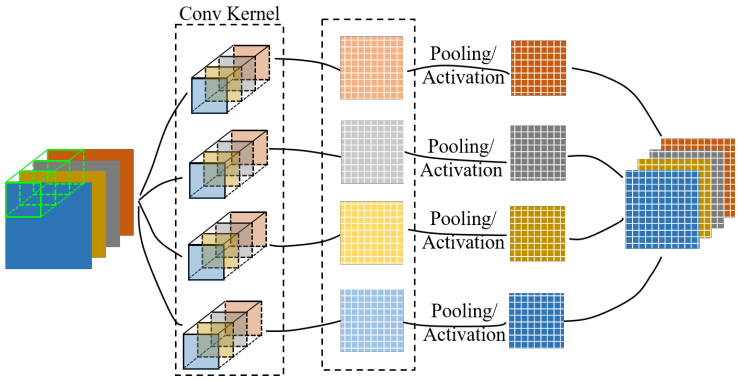

Figure 5.6: Multi-scale feature extraction network architecture

designed to further learn the hidden characteristics in the input feature maps. The network architecture of each stream is shown in Fig. 5.7.

In each stream, convolution operation with multiple kernels is first applied on the input to further project the input to different spaces in multi-scales. MHSA block is then performed on each feature representation to extract features in a parallel manner as illustrated in Section 4.1. As shown in Fig. 5.7, the input feature maps of each stream are projected to different feature matrices with linear transformers, and SDPA is then applied on the transformed matrix to get weighted feature maps in parallel. In the multi-head discriminate feature attention network, different attention heads focus on different features and will give the focused features different weights based on the contributions for fault mode identification. This design enables the network to focus on important information and reduce the effects of irrelevant information in terms of computation cost and diagnosis accuracy. Besides, multi-head self-attention also aims to capture the internal correlation of the input. The attention features from the different heads are concatenated together as weighted features of the current feature stream. With the weighted features, the performance of fault diagnosis can be further improved with high accuracy and efficiency.

Besides, bearing speed information is integrated into the input for fault diagnosis in this work. With the speed varying, the characteristics of fault-related signal

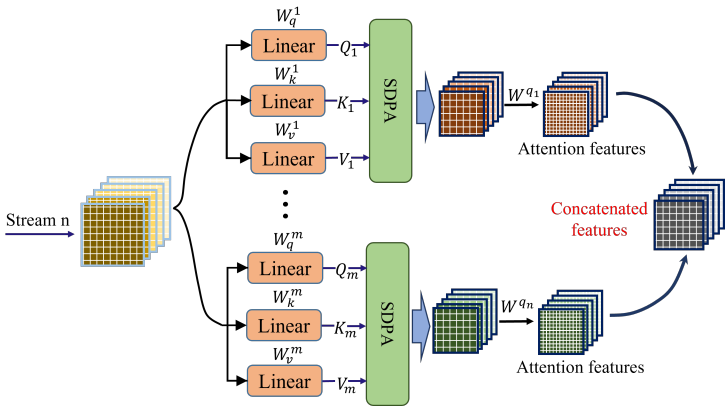

Figure 5.7: Multi-head discriminate feature extraction network architecture

components will also change in the time domain and frequency domain. There exists a correlation between bearing data and varying speed information. The multi-scale discriminate feature extraction network is also capable of learning this kind of spatial information.

Finally, the extracted features from each stream are contacted together as the input of a fully connected layer for fault diagnosis as Eq. (5.6). The training is carried out using the loss function that is defined in Eq. (5.7).

$$F = f(W \cdot \cup(\cup_{k=1}^{m} W^k y_k) + \beta) \tag{5.6}$$

$$L = -\frac{1}{n}\sum_{i=1}^{n} y_i \cdot log(p_i) \tag{5.7}$$

where $\cup$ stands for concatenation operation, f is $softmax$ function, y_i and p_i are the actual label and predicted probability for the i^{th} sample, respectively. The training process is conducted based on the Adam optimizer [110].

5.3 Experimental Results

To verify the effectiveness and advantages of the proposed method, it is performed on two benchmarks with different experiment settings for fault diagnosis tasks. The

experiments are performed in anaconda python environment running on a computer with Intel(R) Core i7-6700 CPU@3.40GHz (8CPUs) processor, 3.4GHz 16G RAM.

5.3.1 Case I: SpectraQuest machinery fault simulator

5.3.1.1 Data Description and Pre-processing

The datasets are from the University of Ottawa. The experiments are performed on a SpectraQuest machinery fault simulator (MFS-PK5M) [117], which is the experimental testbed shown in Fig. 4.7. The bearing is driven by a motor. Two bearings are installed at the drive-end and the non-drive end. The drive-end bearing is a healthy bearing, while the non-drive-end bearing is the experimental bearing, which can be replaced as a healthy bearing, inner fault bearing, or outer race fault bearing based on different health conditions. An accelerometer is installed on the house of the experimental bearing to collect the vibration signals. An encoder is placed near the drive-end to collect the speed information. The vibration signals and rotational speed are all collected with a sampling rate of 200kHz and a duration of 10s.

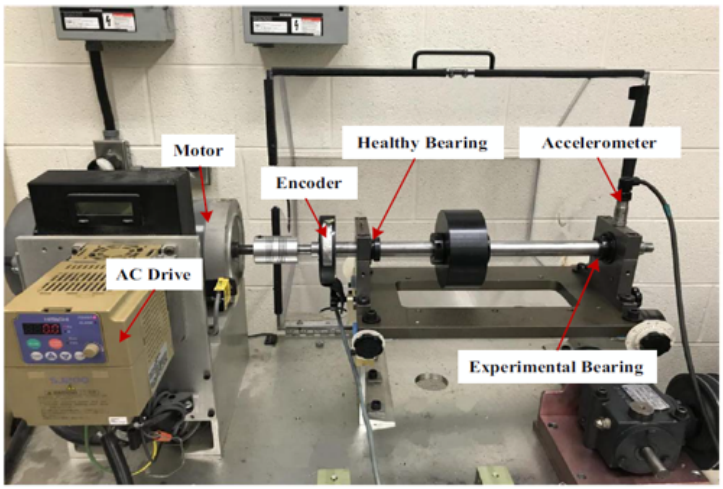

Figure 5.8: MFS-RDS testbed [117]

The experiments are conducted under four operating rotation speed conditions, which include increasing speed, decreasing speed, increasing then decreasing speed, and decreasing then increasing speed. Bearings in different health conditions are all operated under the four rotating speeds. For each fault mode, the raw data are randomly split into training sets and testing sets with a length of 3600 data points. The 1-D time series data of each sensor is converted into 2-D grayscale images with the size of 60×60. The network input is built by combining the data images with bearing rotating speed. The size of the network input is $60 \times 60 \times 2$. For each fault mode, 2000 samples are used, and there are total 6000 samples. To verify the accuracy and effectiveness of the proposed approach, 3-fold cross-validation (CV) is conducted. The samples are randomly divided into three datasets $\{A, B, C\}$. For each data sample, it comes with a label of 3 fault modes (healthy, inner-race fault, outer-race fault).

5.3.1.2 Multi-scale Attention based Fault Diagnosis

In this experiment, the initial learning rate and dropout rate are set at 0.01 and 0.3, respectively. The mini-batch size of the input is 125. The training process is terminated when the number of epoch reaches the pre-defined maximum epoch. Then the trained network is performed on the testing set to validate the performance.

Figs. 5.9 shows the training progress of 3-fold CV. It is evident that the proposed approach can achieve high accuracy with fast convergence speed, which demonstrates that the developed multi-scale CNN can yield promising results. The convergence of the training loss in the whole process for the 3-fold CV is shown in Fig. 5.10. It is obvious that the convergence of the 3-fold CV processes is all stable and fast.

Table 5.1 lists the testing accuracy for the three CV combinations. The average training accuracy can achieve 99.53%. To show the superiority of the proposed approach, the results are compared with the structure without domain knowledge

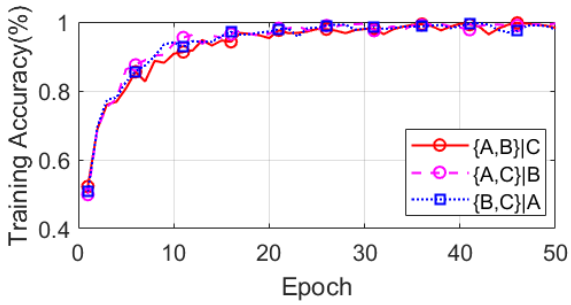

Figure 5.9: Training accuracy of varying speed

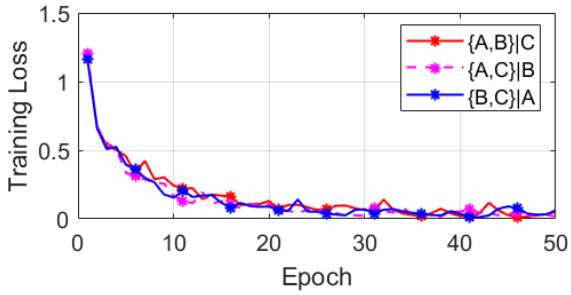

Figure 5.10: Training Loss of varying speed

Table 5.1: Accuracy of diagnosis with MFS-RDS data

3-fold CV	$\{A,B\}\|C$	$\{A,C\}\|B$	$\{B,C\}\|A$	**Average**
Accuracy	99.82	99.37	99.39	99.53

and without feature attention mechanisms, respectively.

To further highlight the effectiveness of the proposed approach, four state-of-the-art DL-based methods, i.e., MK-ResCNN [55], WD-CNN [113], DICNN [118], convolution enabled transformer (Con-ET) [119] are selected for comparison. The comparison results in Table 5.2 show that the average accuracy of the proposed approach has improvements of about 1.85%, 18.07%, 0.96%, and 1.11%, respectively.

Table 5.2: Comparison of fault diagnosis accuracy with different methods

Method	MK-ResCNN	WD-CNN	DICNN	Con-ET	Proposed
Accuracy(%)	97.67	81.45	98.56	98.41	99.53

5.3.2 Case II: Varying Speed Rotating Machine Test Bed

5.3.2.1 Experimental Platform Setup and Data Collection

To further verify the promising performance of the proposed method. The method is verified using the data collected from the rotating machine test bed (RMTB). The test rig consists of six major parts, including a 2 HP motor, three support bearing units, a test bearing unit, a force loading unit (hydraulic pump for applying load radially), transmission (shaft), variable frequency drive (VFD). All the components are placed in line, and the motor connects to the shaft by a shaft coupling. The physical structure is shown in Fig. 5.11. The bearings in this test rig can be divided into two types: support bearings and test bearings. The support bearings with plumber blocks are mainly used for supporting the shaft. The data is collected from the bearing housing with test bearing, and the load is applied on the testing bearing over housing.

The experiment was performed on a bearing from new to failure. Bearing is tested at 3200 pounds load. The rotational speed was cycled through 1-minute periods of 600, 900, and 1200 rpm to simulate a practical varying speed mechanical system. The experiment was terminated with a complete failure that happened in the inner race of the test bearing as shown in Fig. 5.11.

5.3.2.2 Multi-scale Attention based Fault Diagnosis

The bearing run-to-failure data are divided into 5 health stages based on the fault severity. For each health stage, the data are preprocessed to generate 800 samples. There is a total of 3000 samples in use, while 80% of the samples are used for training, and 20% are used for testing.

To intuitively demonstrate the feature extraction capability of the proposed method, the raw data and the features extracted by the proposed method are

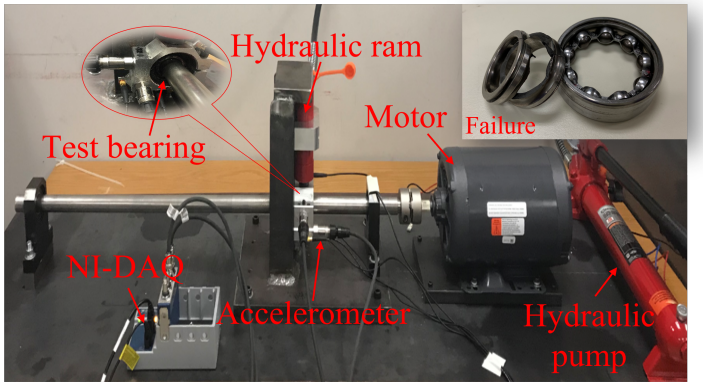

Figure 5.11: The rotating machine test bed

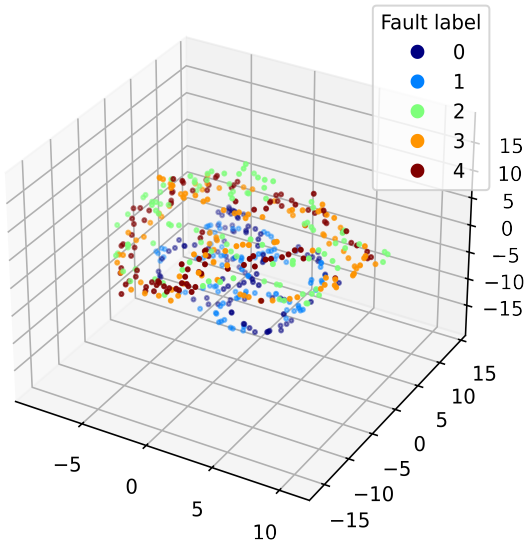

Figure 5.12: Feature visualization of raw data

visualized in 3-D space with t-distributed stochastic neighbor embedding (t-SNE) [120]. Fig. 5.12 and Fig. 5.13 visualize the first three main components of the raw data and extracted features, respectively. In Fig. 5.12, the samples from raw data are mixed together, and it is difficult to distinguish the corresponding fault mode. However, the extracted features from different fault modes are almost separable, which indicates that they can be clearly distinguished. The results show that the proposed method has a powerful feature extraction ability. To verify the effectiveness of the varying speed integration, the first three main components of the extracted

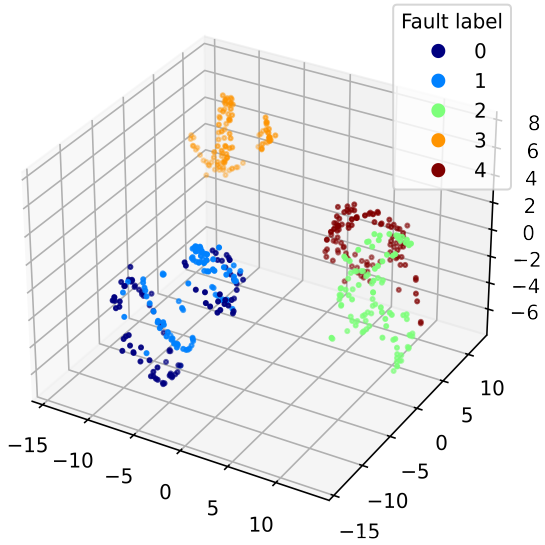

Figure 5.13: Feature visualization of extracted features

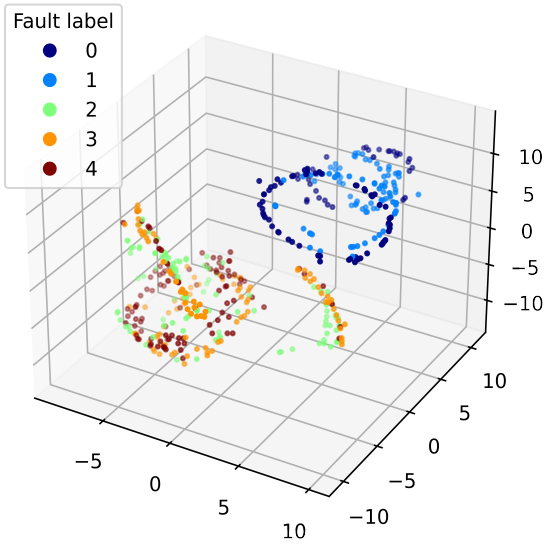

Figure 5.14: Feature visualization of extracted features without speed information

features without speed information involvement are visualized in Fig. 5.14. It can be observed that the extracted features of different faults are still mixed together, and this will further affect the accuracy of fault diagnosis. It demonstrates that speed information and domain knowledge integration can significant improve the network feature learning ability.

The training process is shown in Fig. 5.15. It can be seen that although the

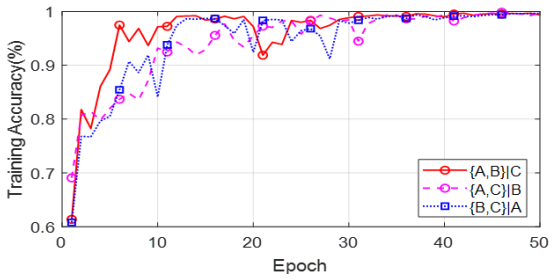

Figure 5.15: Training accuracy of RMTB

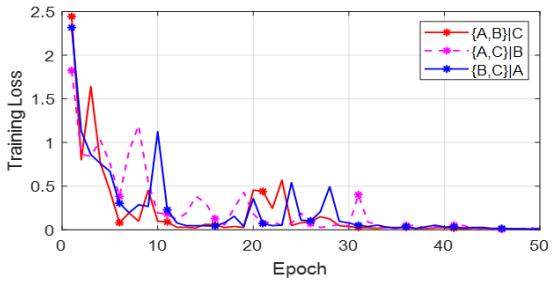

Figure 5.16: Training loss of RMTB

Table 5.3: Accuracy of diagnosis with RMTB

3-fold CV	$\{A,B\}\|C$	$\{A,C\}\|B$	$\{B,C\}\|A$	**Average**
Accuracy	99.2	98.3	99.8	99.1

training accuracy shows some fluctuations, the training accuracy can still reach a high accuracy after about 30 epochs. Table. 5.3 shows the testing accuracies of 3-fold CV. The average accuracy can achieve 99.1% on the RMTB, which further indicates that the proposed method has stable and robust performance.

5.4 Conclusions

This chapter proposes a novel multi-scale discriminate CNN based bearing fault diagnosis. The raw data from multi-sensors are converted and fused with varying speeds to build the information maps. Multi-head attention-based discriminate features learning framework is proposed for feature extraction. Two case studies are conducted to verify the performance of the proposed approach. The feature

extraction capability is visualized using t-SNE dimensionality reduction technique, which demonstrates the powerful feature extraction ability of the proposed method. Experimental results and comparisons show that the proposed approach can achieve better performance in terms of accuracy and efficiency compared with some state-to-the-art DL-based methods.

Chapter 6

Cost-Efficient CWT-CNN Based Bearing Fault Diagnosis and Prognosis

This chapter introduces a hybrid Bayesian estimation-based bearing FDP approach. The proposed method consists of two major offline components (Health Indicator (HI) and model training) and four major online components (CNN-based STP detection, CNN-based model selection, PF-based prognosis, DST-based prediction fusion). This work combines the powerful feature learning and pattern recognition ability of deep learning, and state estimation and uncertainty management ability of Bayesian estimation. In this work, Convolutional neural network (CNN) [121] is adopted for STP detection and model selection due to the powerful feature learning and pattern recognition ability [53, 121–124]. In the offline part, two CNN-based models are trained using bearing information maps. The fault dynamic models are built using the extracted HI. In the online prognosis process, the fault state is detected using CNN-based fault model detection. The prognosis process is executed after the detection of a fault with the integration of fault dynamic models, which are selected using CNN-based fault selection model. A Dempster-Shafer Theory (DST) based fusion mechanism is proposed to further improve the prognostic performance. The bearing state and RUL can be estimated based on PF based estimation and fusion results.

6.1 Proposed Approach for bearing fault diagnosis and prognosis

Fig. 6.1 illustrates the framework of the proposed method, which includes feature extraction, offline modeling, online fault detection, and Bayesian estimation based FDP. The implementation details are described as follows:

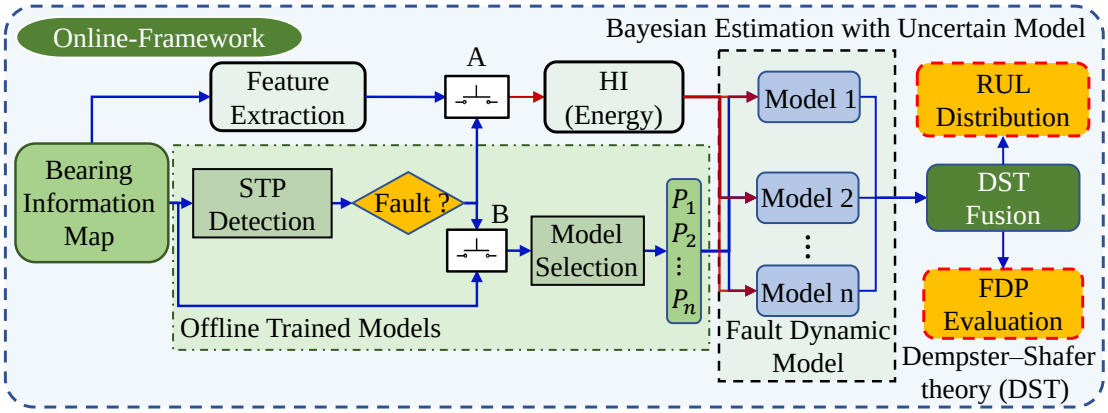

Figure 6.1: Proposed hybrid PF based FDP approach.

Step 1 Take a segment of vibration signals and transform it into CWCM energy spectrum images, which are combined with bearing operating conditions to build information maps.

Step 2 Offline CNN training: Construct and train two CNNs using the information maps. One is for automatic fault and STP detection. The other one is for fault model selection.

Step 3 Offline fault dynamic modeling: Extract HI from raw data and build different fault dynamic models based on the HI.

Step 4 Design PF based FDP framework with the selected fault dynamic model, which includes state estimation, RUL prediction, and DST based prognostic fusion if necessary.

Step 5 Conduct online experiments to verify the performance of the proposed approach.

6.1.1 CWT-CNN based STP identification

6.1.1.1 Theory of CWT

CWT is a time-frequency domain signal processing approach, which provides a time-scale view of signals [125]. The transformed signal is obtained by applying a family of wavelet functions, defined by translation τ and scale a, on raw signals. The translation defines the location of the wavelet, while the scale, defined as $a = 1/f$, relates to the stretching or compressing scale of the wavelet. The transform is defined as:

$$CWT(a,\tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) * \Psi(\frac{t-\tau}{a}) dt \tag{6.1}$$

where Ψ is the mother wavelet and the Morlet wavelet is used in this work. This decomposes raw signal $f(t)$ into CWCMs. Fig. 6.2 shows the wavelet power spectrum of a bearing in the run-to-failure experiments. For each subfigure, the horizontal axis is time and the vertical axis is frequency. The images show that the bearing vibration has apparent energy distribution in the degradation process.

6.1.1.2 Fusion of CWCM and system operating condition

Operating conditions, including rotating speed, loading profile, temperature, etc., have great effects on bearing degradation rate and fault state. For example, the service life may reduce a lot for a system operating in a high-stress condition, such as high load and high speed. Vibration signals at different operating conditions show different features in terms of energy, noise, frequency, amplitude, among others. Therefore, it is important information and should be integrated into the FDP process.

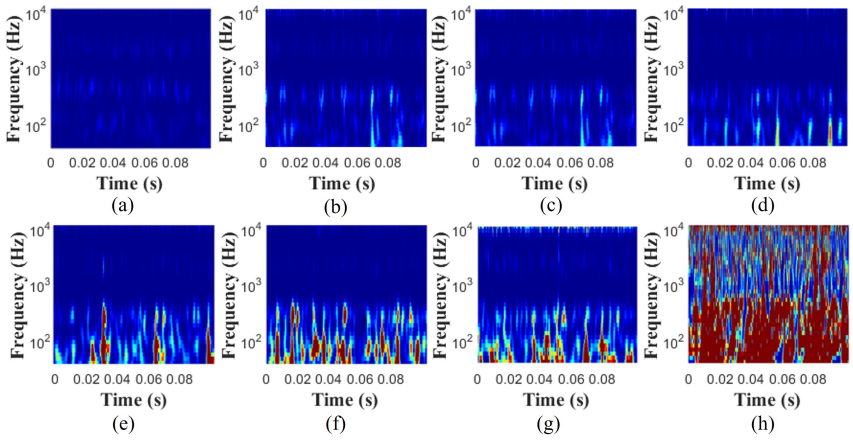

Figure 6.2: Power Spectrum of CWCM during the run-to-failure process.(a)-(h) are CWCM of samples at different time instants

In this work, two operating conditions, loading profile and rotating speed, are considered. To insure the uniformity of CWCM images with the operating information, an empty image with the same size as the CWCM image is built. The background color of the empty image is defined as the level of the loading profile. The rotating speed information is integrated into the empty image using color bands in different locations. According to CWT, the vertical axis of CWCM is the frequency that corresponds to scale a as:

$$F_a = \frac{F_c \cdot f_s}{a} \tag{6.2}$$

where F_c is the center frequency of the wavelet function, F_a is the related frequency of scale a. Based on the relationship, the location of the color band can be estimated as:

$$i_{f_r} = \frac{F_c \cdot f_s}{f_r} \tag{6.3}$$

where i_{f_r} is the location of color band, f_c is the signal sampling frequency, f_r is bearing rotating frequency, which can be obtained as $f_r = r_p/60$, r_p is the rotating speed with the unit of rpm. After estimating the location, a color band can be added to locations

that correspond to different rotating speeds. The built operating condition images are then fused with CWCM to build bearing information maps as Fig. 6.3. For each snapshot of raw data, an information map is built for the following FDP process.

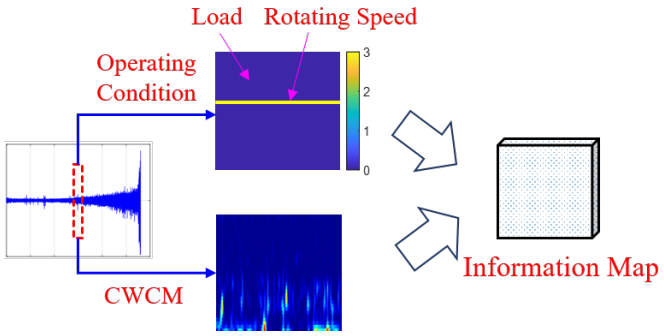

Figure 6.3: Information map construction.

6.1.1.3 CNN based fault detection and model selection

Accurate detection of the occurrence of a bearing fault and a good fault dynamic model are critical for FDP. CNN is a classical deep learning technology that has achieved promising results in bearing FDP as it is able to process original data with minimal amounts of pre-processing. A classic CNN is mainly composed of alternating convolutional layers, pooling layers, and fully connected layers. In the convolutional layer, a set of filters are applied on the input images to extract feature maps. The pooling layer down-samples the feature maps to reduce the dimension of convolution features. The extracted features are finally integrated in fully connected layer as the input of classifier to calculate the probabilities of bearing health condition (healthy or faulty).

Two 2-dimensional (2-D) CNNs, i.e., STP detection CNN and model-selection CNN, are trained separately for fault detection (STP detection) and fault dynamic model selection. STP detection aims to detect the occurrence of a fault, i.e., the STP for prognosis. The input of STP detection is the information maps that include the energy spectrum of CWCM and operating condition. The output is bearing

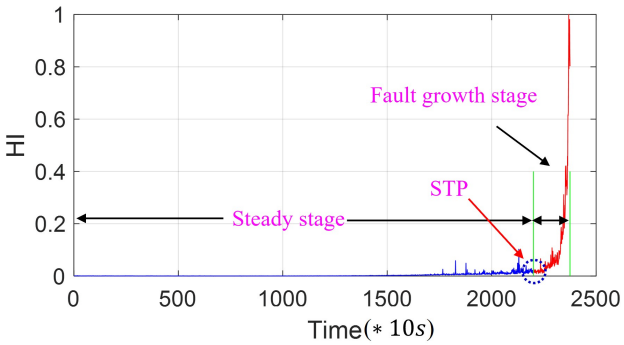

Figure 6.4: Bearing STP detection

health condition label which indicates the fault state (health or fault). Note that the STP detection CNN only uses the information map at the current time instant for detection. When a fault is detected, model-selection CNN selects the most appropriate fault dynamic model for prognosis. Since the degradation model for prognosis describes the fault growth trends, it requires several consecutive historical information maps as the input vector of the model selection CNN. The output of the model-selection CNN is the probabilities of candidate fault dynamic models. The model selection probabilities are used as evidence for decision-making of the prognosis with the single model or multiple models. Besides, the probabilities are also used to fuse the results of the prognosis with multiple models.

For STP detection CNN, the data formet is $D^F = \{X_t; Y_t^F\}$, where X_t and Y_t^F are the bearing information maps and the label indicating the bearing health stage at t, respectively. For each bearing dataset, the data are divided into healthy datasets and faulty datasets manually to train the CNN for fault detection.

For model selection CNN, the data format is $D^M = \{X_{t-k+1}, \cdots, X_t; Y_t^M\}$, where $X_{t-k+1}, \cdots, X_t$ are bearing information maps in the past k time instants, and Y_t^M is the labels of different models. Note that the model selection CNN is triggered after a bearing fault is detected. Then, the model selection CNN is trained using the information maps that are divided as faulty to select appropriate fault models.

6.1.2 Bearing FDP using Particle Filtering

6.1.2.1 HI Construction

A high-quality HI should accurately indicate the degradation of bearings. Based on the bearing analysis results in the time-domain, frequency domain, and time-frequency domain, energy is selected as HI since it presents clear degradation trends in the faulty stages. HI can indicate the fault severity of bearings, which is defined as the ratio of the power of CWCM to its maximum power.

6.1.2.2 Fault dynamic model

The fault growth dynamics can be generally described as:

$$x_k = f(x_{k-1}, \omega_k); \text{also denoted as} \quad p(x_k|x_{k-1}) \tag{6.4a}$$

$$y_k = h(x_k, v_k); \quad \text{also denoted as} \quad p(y_k|x_k) \tag{6.4b}$$

where k is time stamp, x is the fault state, $f(\cdot)$ depicts the bearing state transition, $h(\cdot)$ is the measurement equation, y is the state measurement, ω_k and v_k are process and measurement noises, respectively. Since bearing state cannot be measured, HI is used as the state. For this setting, Eq. (6.4b) reduces to $y_k = x_k + v_k$. To capture the uncertainty of different bearing degradation cases, the models are built as probability models, in which the parameters are subject to different distributions. Fig. 6.5 shows the comparison of model with the degradation data of different bearings.

The diagnostic algorithm is executed from the estimated STP t_{STP} to estimate the current health state. Prognosis is the procedure of long-term (multi-step) prediction and RUL calculation. The process involves two stages: the first stage is to calculate the fault state distribution at each future time instant by using the fault state model repeatedly. The prediction step is carried out with a fixed time interval from the current time t_k to the failure time instant t_f when the fault state reaches the failure

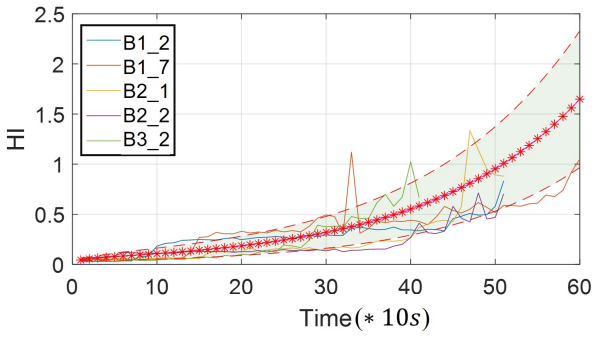

Figure 6.5: Fault modeling with uncertainty.

threshold F_f. The prediction steps are $\{t_k, t_{k+1}, \cdots, t_{f-1}, t_f\}$ and the predicted fault state mean value of the distribution at these time instants can be denoted as $\{F(t_k), F(t_{k+1}), \cdots, F(t_f)\}$.

The second stage is to compare the fault state pdf (probability density function) at all time instants with the failure threshold by using the law of total probabilities to get the time to failure (TTF) or RUL distribution. Here, RUL is calculated by comparing the distributions of fault state at all prediction steps with the failure threshold. The prognosis is conducted at every time instant to get a RUL distribution.

6.1.2.3 Bearing FDP using PF

In this long-term prediction of prognosis, uncertainty is a key factor that should be addressed. Bayesian estimation techniques [126] provide a general rigorous solution for uncertainty management. In this research, PF is employed for its advantages in nonlinear representation and uncertainty management.

Mathematically, the bearing fault states X can be described by a Markov process characterized by the initial distribution $p(x_0)$ and the transition probability $p(x_k|x_{k-1})$ defined in Eq. (6.4a). Define $x_{0:k} = \{x_0, \cdots, x_k\}$ and $y_{1:k} = \{y_1, \cdots, y_k\}$ as the state and measurement. It is of interest to estimate the posterior distribution $p(x_{0:k}|y_{1:k})$. Based on the Bayesian estimation theory, the task involves two steps, i.e., prediction and filtering.

Since many bearing fault dynamics are nonlinear or non-Gaussian, PF method [127] is used to approximate the optimal solution. Firstly, a set of N particles $\{x_{k-1}^i,\ w_{k-1}^i\},\ i = 1, 2, \cdots, N$ is assumed available at the time instant $(k-1)$, where x_{k-1}^i define the locations of particles in the fault state space and w_{k-1}^i are the weights of particles with the sum of 1. This set of particles can be used to approximate the desired state distribution $\psi_{k-1}(x_{k-1})$.

The objective is to get a new set of particles $\{\hat{x}_k^i,\ \hat{w}_k^i\}$ that can approximate the state distribution $\psi_k(x_k)$. Since the state distribution $\psi_k(x_k)$ is unknown, an importance distribution $q_k(\hat{x}_k^i|x_{k-1}^i)$ is defined and all current particles are extended using this important distribution. Since $q_k(\hat{x}_k^i|x_{k-1}^i)$ is different from the real state distribution $\psi_k(x_k)$, their difference is compensated by weights.

For bootstrap particle filtering, the importance distribution is taken as the transition prior, i.e., $q_k(\hat{x}_k^i|x_{k-1}^i) = p(x_k|x_{k-1})$. With this choice, the weights are updated as:

$$w(\hat{x}_k^i) = \psi_k(x_k)/q_k(\hat{x}_k^i|x_{k-1}^i) = w(x_{k-1}^i)p(y_k|x_k^i) \tag{6.5}$$

Then, the approximation at k can be estimated as:

$$p(x_k|y_k) = \sum_{i=1}^{N} \hat{w}_k^i \delta(x_k - \hat{x}_k^i) \ \text{ with } \ \hat{w}_k^i = \frac{w(\hat{x}_k^i)}{\sum w(\hat{x}_k^i)} \tag{6.6}$$

where δ is the Dirac-delta function, $\sum_{i=1}^{N} \hat{w}_k^i = 1$. When the particle number N is large, the posteriori estimation distribution $p(x_k|y_k)$ will converge to the exact solution $\psi_k(x_k)$.

In prognosis, the current state distribution is projected to future time instance. For each execution, the fault state x_{k+1} at the next time instant is predicted as $\tilde{x}_{k+1}^{(i)}$. The pdf $p(\tilde{x}_{k+1})$ can be approximated using the predicted particles.

Following this process, the predicted state distribution is recursively taken as the input of fault model for predicting the state pdf at the next time instant. Based on this strategy, the state pdf at each future time instant $\{p(\tilde{x}_{k+1}),\ p(\tilde{x}_{k+2}), \cdots, p(\tilde{x}_f)\}$

can be predicted. Finally, the TTF pdf can be obtained by comparing those state pdf with the failure threshold. The predicted RUL pdf $p(RUL)$ can be obtained by calculating the time interval between the current time instant and the predicted TTF as:

$$p(RUL) \approx \frac{1}{N} \sum_{n=1}^{N} \delta(TTL^{(i)} - k) \tag{6.7}$$

6.1.2.4 RUL prediction fusion

Suppose n models are defined for fault dynamic description, the output of model-selection CNN can be described as:

$$[P_1, P_2, \cdots, P_n] = softmax(x) \tag{6.8}$$

where $P_1 \geq P_2 \geq \cdots \geq P_n$ are the ranked probabilities of model selection, x is the vector of the extracted features in the fully connected layer of model-selection CNN. In practice, n often is a small number to reduce the design effort. An adaptive model selection mechanism is developed to avoid equivocal decision-making for model selection. The model selection with high confidence and weak confidence will be distinguished for different operations. A high probability P_1 means that the model is selected with strong confidence. The FDP is conducted with the single selected fault model. On the contrary, if the selection probability is small, the fault model is selected with weak confidence. In this case, the two models with higher probabilities P_1 and P_2 are used for prognosis parallelly. The final RUL is estimated by fusing predictions from two models using DST. The selection of two models will ensure the real-time implementation. The model selection threshold is set as 0.7 in this work.

DST is an extension of Bayesian methodology [128], which utilizes belief uncertainty intervals based on evidence of multiple observations to represent the belief of assumptions (boa) [128]. It is an effective decision fusion algorithm that exploits the probabilities of multiple pieces of uncertain evidence enclosed within the

prediction process. It can provide the confidence of the occurrence of a specific event. The DST fusion of prognosis results from multiple models can be more sufficient and accurate to support decision-making.

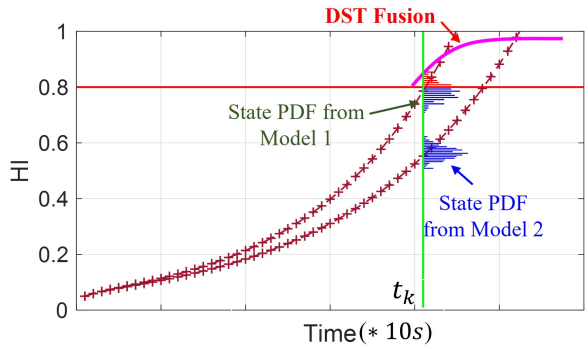

Figure 6.6: RUL prediction fusion

In DST fusion, the frame of discernment is composed of bearing condition probabilities (Faulty (μ_1), Failure (μ_2)) and two source uncertainties (model selection uncertainty (μ_3), and prediction uncertainty (μ_4)) as $\Theta = \{\mu_1, \mu_2, \mu_3, \mu_4\}$. The faulty probability, blue part of state pdf in Fig. 6.6, is defined as the probability that the state is detected as faulty but not yet reached the failure threshold, suppose two models are used. The failure probability (red part of pdf) is defined as the probability that the state has reached failure threshold. The basic probability assignment (BPA) for faulty, failure, model selection uncertainty, and prediction uncertainty are denoted by $m(\mu_1), m(\mu_2), m(\mu_3), m(\mu_4)$, respectively. For example, $m(\mu_2)$ is bearing failure probability.

FDP process considers uncertainties of model selection and prediction. Model selection uncertainty is evidenced by the probability of model selection, which is quantified as a function of model selection probabilities:

$$m(\mu_3) = \lambda \cdot (P_2/P_1) \tag{6.9}$$

where P_1 and P_2 are model selection probabilities of two models, λ is a parameter

determined by trial-and-error to adjust the effect of model selection in FDP. Prediction uncertainty is evidenced by the state pdf at each time instant. It is defined as:

$$m(\mu_4) = S/S_{max} \tag{6.10}$$

where S represents the spread function, which is represented as $2\sigma^2$, and σ^2 is the variance of the state pdf, and S_{max} is the spread limit. Then, the BPA for failure is assigned as:

$$m(\mu_2)_k = p(f|x_k > \xi) \cdot (1 - m(\mu_3)_k - m(\mu_4)_k) \tag{6.11}$$

where ξ is the bearing failure threshold, k is the current time instant, $p(\cdot)$ is the probability function that describes the failure probability of the state when the prediction reaches the pre-defined failure threshold. Based on the quantification rule, the values of different masses are described in Table 6.1.

Table 6.1: The mass combination of prediction using DST

Model 1 \ Model 2		Faulty	Failure	Model Selection Uncertainty	Prediction Uncertainty
		$m_2(\mu_1)$	$m_2(\mu_2)$	$m_2(\mu_3)$	$m_2(\mu_4)$
Faulty	$m_1(\mu_1)$	$m_1(\mu_1) \cdot m_2(\mu_1)$	$m_1(\mu_1) \cdot m_2(\mu_2)$	$m_1(\mu_1) \cdot m_2(\mu_3)$	$m_1(\mu_1) \cdot m_2(\mu_4)$
Failure	$m_1(\mu_2)$	$m_1(\mu_2) \cdot m_2(\mu_1)$	$m_1(\mu_2) \cdot m_2(\mu_2)$	$m_1(\mu_2) \cdot m_2(\mu_3)$	$m_1(\mu_2) \cdot m_2(\mu_4)$
Model selection Uncertainty	$m_1(\mu_3)$	$m_1(\mu_3) \cdot m_2(\mu_1)$	$m_1(\mu_3) \cdot m_2(\mu_2)$	$m_1(\mu_3) \cdot m_2(\mu_3)$	$m_1(\mu_3) \cdot m_2(\mu_4)$
Prediction Uncertainty	$m_1(\mu_4)$	$m_1(\mu_4) \cdot m_2(\mu_1)$	$m_1(\mu_4) \cdot m_2(\mu_2)$	$m_1(\mu_4) \cdot m_2(\mu_3)$	$m_1(\mu_4) \cdot m_2(\mu_4)$
			m(F)	*m(U)*	

Fig. 6.6 is the DST based prognostic fusion process. At each prognostic cycle, the mass function of each part can be obtained as Table 6.1. The final prognostic results are the BPA combinations of failure $m(\mu_2)$, model selection uncertainty $m(\mu_3)$, and prediction uncertainty $m(\mu_4)$, the combination is the fused failure probability. The blue box represents $m(F)$ where the prognostic from two models agree with each other. The magenta box is the combined uncertainty $m(U)$.

$$m(F) = m_1(\mu_2)m_2(\mu_2) + m_1(\mu_2)m_2(\mu_3) + m_1(\mu_2)m_2(\mu_4) + m_1(\mu_3)m_2(\mu_2) + m_1(\mu_4)m_2(\mu_2)$$

$$m(U) = m_1(\mu_3)m_2(\mu_3) + m_1(\mu_3)m_2(\mu_4) + m_1(\mu_4)m_2(\mu_3) + m_1(\mu_4)m_2(\mu_4)$$

After all BPA functions are estimated, the fused failure probability and uncertainty can be estimated as:

$$P_F(k) = \frac{m(F)}{1 - (m_1(\mu_1)m_2(\mu_2) + m_2(\mu_1)m_1(\mu_2))}$$

$$P_U(k) = \frac{m(U)}{1 - (m_1(\mu_1)m_2(\mu_2) + m_2(\mu_1)m_1(\mu_2))}$$

6.2 Experiments and Analysis

In this section, a series of bearing experimental results are presented to verify the proposed method. The experiments are implemented in MATLAB R2018a environment running on a computer with Intel(R) Core i7-6700 CPU @ 3.40GHz (8CPUs) processor, 3.4GHz 16G RAM.

6.2.1 Data description

The data is collected from the PROGNOSTIA experimental testbed [129]. The testbed is composed of three main components: the rotating system, the loading part, and the measurement part. The bearings are operated in different loads and speeds. The load is generated by a force actuator, and the speed is changed by switching different couplings. Two accelerometers, which are installed radially on the external race, are used to collect the bearing vibration data with a sampling rate of 25.6 kHz. Table 6.2 presents an overview of bearing datasets.

Table 6.2: Bearing data description

Operating Condition	Rotating Speed (rpm)	Load (N)	Dataset
Condition 1	1800	4000	B1_1, B1_2, B1_3 B1_4, B1_5, B1_6
Condition 2	1650	4200	B2_1, B2_2, B2_3 B2_4, B2_5, B2_6
Condition 3	1500	5000	B3_1, B3_2, B1_3

6.2.2 HI construction and fault modeling

As mentioned earlier, HI is used as feature to describe fault dynamics. Fig. 6.7 shows the extracted HIs for different bearings. Note that, the steady-state phases are removed in this figure to highlight the degradation process after the fault occurs. It is clear that these degradation processes of bearings can be described by three fault models. For this reason, three models are built for bearing FDP. To cover most bearing degradations for each fault mode, probabilistic models are used. The three probabilistic models are built as:

$$\text{Model 1, 2: } f_{k+1} = f_k + p_1 e^{p_2 k} + \omega_k \tag{6.12a}$$

$$\text{Model 3: } f_{k+1} = f_k + p_1 k^3 + p_2 k^2 + p_3 k + p_4 + \omega_k \tag{6.12b}$$

where f is bearing HI, k is time index, $p_1 \sim p_4$ are the model parameters. Fault models 1 and 2 are built as (6.12a), fault model 3 is built as (6.12b). The model parameters are identified for different models as shown in Table 6.3. Note that the model parameters are subject to a Gaussian distribution to accommodate the uncertainty of fault dynamics.

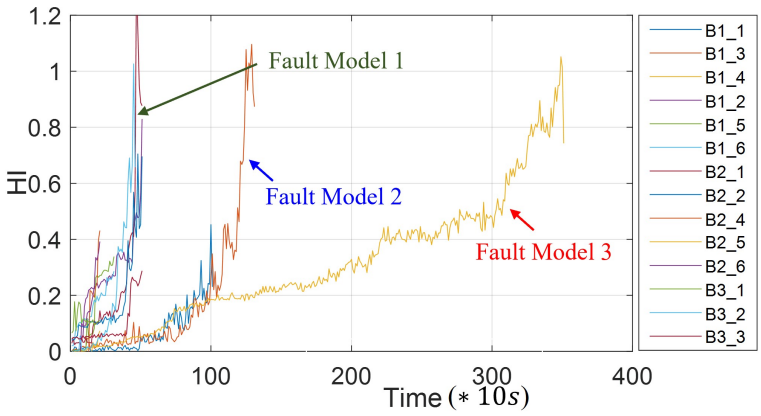

Figure 6.7: The HIs for different bearings.

Table 6.3: Model parameters for different fault models

Parameter	Value	
Model 1	$p_1 \sim N[9.11e^{-5}, 0.0005]$	$p_2 \sim N[0.0565, 0.005]$
Model 2	$p_1 \sim N[0.01139, 0.0005]$	$p_2 \sim N[0.9199, 0.005]$
Model 3	$p_1 \sim N[9.86e^{-10}, 0.0005]$	$p_3 \sim N[4.1e^{-5}, 0.005]$
	$p_2 \sim N[-3.87e^{-7}, 0.0005]$	$p_4 \sim N[0.0006, 0.005]$

6.2.3 STP detection and fault model selection

Two CNNs are trained for STP detection and model selection, respectively. STP detection CNN aims to detect bearing fault and estimate the STP for prognosis. The output of STP detection CNN is bearing health condition (healthy, faulty). Model selection CNN outputs the probabilities for three fault models. The training process is conducted based on the Adam optimizer. The initial learning rate and dropout rate are set as 0.01 and 0.3, respectively. The detailed structure of the trained model selection CNN is described in Table 6.4. The mini-batch size of the input is 125. The training process is terminated when the epochs reach the pre-defined training threshold of maximum epoch or identification accuracy. The overall average offline testing accuracy for STP detection and model selection are illustrated in Table 6.5.

Based on Table 6.5, two CNN models can accurately detect bearing faults and select the most appropriate fault models for FDP.

Table 6.4: Structure of the proposed network in model selection

Layer No.	Layer Type	Filter Size	Filter Count	Output
1	Input	-	-	60×60×6
2	Convolution	5×5×1	100	60×60×100
3	Maxpooling	-	-	60×60×100
4	Convolution	-	-	60×60×100
5	ReLU	1×1	100	1×1×100
6	Convolution	5×5×3	30	5×5×3
7	Max pooling	2×2×3	30	5×5×3
8	BN	-	30	5×5×3
9	Convolution	5×5×3	30	5×5×3
10	Max pooling	2×2×3	30	5×5×3
11	ReLU	-	-	56×56×3
12	Max Pooling	4×4×3	30	53×53×3
13	BN	-	-	53×53×3
14	ReLU	-	-	53×53×3
15	Dropout	-	-	1×1×100
16	Fully connected	1×1×100	3	1×1×3
17	Softmax	-	-	1

Table 6.5: Accuracy of stage identification and model selection

Task	Stage Identification	Fault Model Selection
Accuracy(%)	98.27	96.04

Fig. 6.8 is the fault detection result of Bearing 1_3. Although the bearing is detected as faulty at 1827s and 1877s, they are not effective detection since the state is identified as faulty only after three consecutive states are detected as failure. The

fault is finally detected at 2143s. It is clear that the fault detection results are almost consistent with the actual bearing condition. After a fault is detected, three consecutive historical information maps are combined as the input of model selection CNN to select fault prognostic model. The probabilities of the three models are [0.98, 0.01, 0.01]. Since $P_1 > 0.7$, Model 1 in Table 6.3 is selected to be implemented with PF to conduct prognosis for Bearing 1_3.

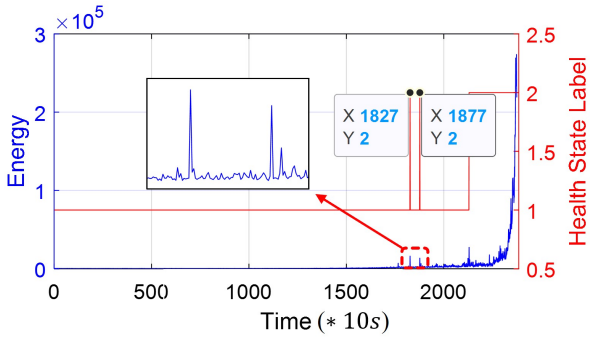

Figure 6.8: Bearing failure detection results for Bearing 1_3.

6.2.4 Fault diagnosis and prognosis with single fault model

In the implementation, the selected single model is integrated in the PF-based FDP algorithm. In the diagnosis stage, the particle filter is configured with 100 particles. The failure threshold of HI is set as 0.8. Fig. 6.9 shows the diagnostic results for Bearing 1_3 at the 13th cycle after the detection of STP. The figure shows the comparison of the mean of bearing state pdf (red) and the measurements (HI values) (blue). The pdf of each estimation are also given to show estimation uncertainty.

After the bearing state distribution is obtained from diagnosis, it is used as the initial condition for prognosis to estimate the TTF. Since there is no measurement in prognosis, it is conducted based only on fault dynamic models. The prognosis is also configured with 100 particles. Fig. 6.10 shows the prognostic result at the 13th cycle after the STP detection. At each time instant, the mean value and the 95% confidence interval of the predicted fault state distributions are plotted. To make

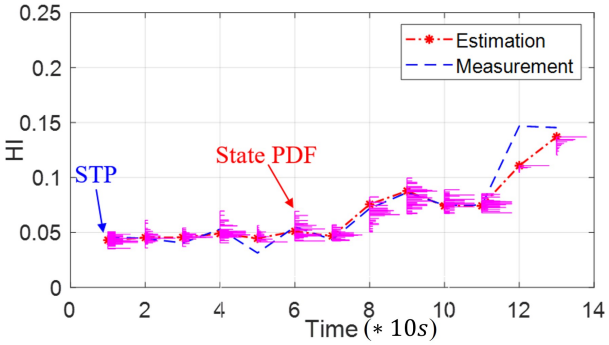

Figure 6.9: Bearing diagnosis results.

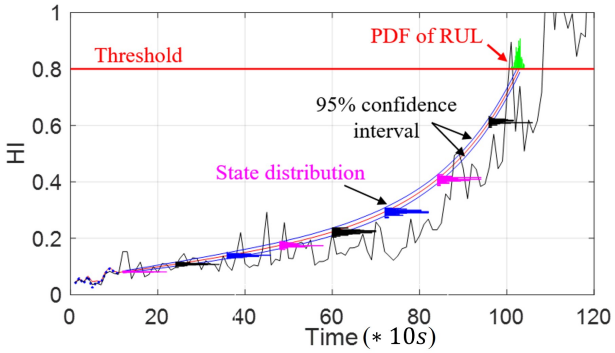

Figure 6.10: Bearing prognosis results.

the figure clear, it only shows the state pdf on some selected time instants. Each prognosis process is terminated after all particles reach the failure threshold. The RUL pdf, which is given in green color, is obtained by collecting the time instants when the particles reach the failure threshold.

To demonstrate the RUL prediction accuracy in the whole bearing life, $\alpha - \lambda$ metrics [130] with $\alpha = 0.3$ is used. This metric shows whether the predicted RUL at any particular time instant falls into a defined precision range. Fig. 6.11 shows the RUL prediction results for Bearing 1_3. It is clear that 91% of the predicted RUL falls into the defined accuracy zones.

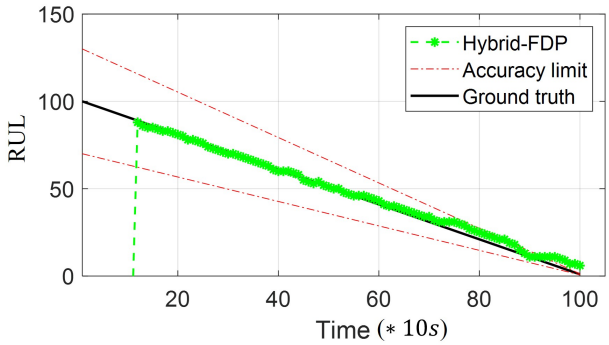

Figure 6.11: The RUL prediction result of Bearing 1_3.

6.2.5 DST fusion based FDP with multiple fault models

As mentioned earlier, DST-based fusion process will be triggered when necessary. Fig. 6.12 shows a prognosis process using multiple models of Bearing 1_3 at the 28th time instant, which triggers the DST-based fusion process for prognosis. At the current time instant, the model selection CNN yields probabilities of $[0.63, 0.32, 0.05]$. Since the probabilities of all models are less than 0.7, Models 1 and 2 are selected to run the prognosis in parallel, i.e., two state pdfs are obtained for each prognosis cycle. The mass function for the faulty state and failure state can be calculated based on the DST details described in Section 6.1. For the failure prediction from two models, DST provides a fused failure probability. Bearing TTF is estimated using a detection probability 90%. In the figure, Model 1 has an early failure alarm, while Model 2 has a late one. The fused failure probability falls between the two failure probabilities. For example, at the 122nd time instant, the probability of failure (PoF) of Model 1 is 0.85, while the PoF of Model 2 is 0.08. The fused PoF is 0.53, which provides a more reasonable failure alarm than single model. Table 6.6 shows the average errors of the predicted RUL at several model fusion time instants from single model and DST fusion. Clearly, DST based fusion method can provide an RUL prediction with better performance.

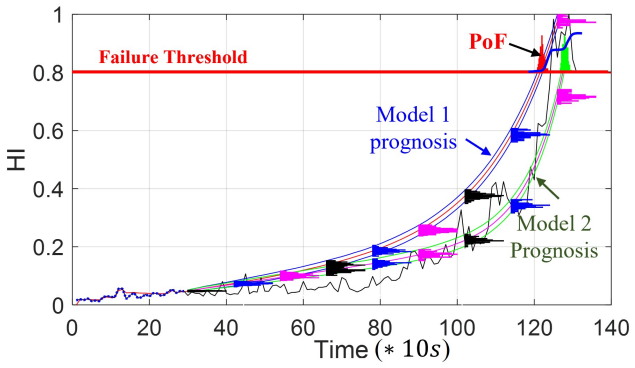

Figure 6.12: Fusion of prognosis with two fault models

Table 6.6: RUL Prediction Error Comparison

Time Instant(s)	Model 1	Model 2	DST
25	11	6	4
42	3	3	2
57	5	3	2
73	8	4	3
Average Error	6.75	4	2.75

6.2.6 Results comparison and analysis

To further demonstrate the effectiveness of the proposed method, the results are analyzed and compared with some state-to-the-art methods in terms of accuracy and application economical efficiency using two different evaluation metrics.

6.2.6.1 Accuracy analysis

Cumulative Relative Accuracy (CRA) [131] comprehensively assesses the accuracy of a prognostics method by aggregating the relative prediction accuracy at all prediction times. It has the definition of:

$$CRA_\lambda = \frac{1}{|\ell_{\lambda|}} \sum_{i=1}^{\ell_\lambda} w(r(i)) RA_\lambda \tag{6.13}$$

where ℓ_λ is the set of all time index of the predictions, $w(r(i))$ is a weight factor as a function of RUL at all prediction time indices. The results are compared with particle

filter and relevance vector machine based hybrid prognosis method (PF-RVM)[132], and a multi-objective deep belief networks ensemble (MODBNE) [133], as shown in Table 6.7, in which the best CRA values are highlighted. The proposed method has higher prediction accuracy than the other three methods for most bearing cases.

Table 6.7: Comparison of Different methods using CRA

Bearing No.	PF-RVM	MODBNE	Proposed
Bearing 1_1	0.5384	0.4318	**0.7313**
Bearing 1_2	0.6369	0.6248	**0.7403**
Bearing 1_3	0.7306	0.5571	**0.8000**
Bearing 1_4	0.6278	0.4085	**0.7636**
Bearing 1_5	0.6352	0.6636	**0.7140**
Bearing 2_1	0.6237	0.5518	**0.7557**
Bearing 2_2	0.5081	0.2564	**0.6946**
Bearing 3_2	0.5825	0.5167	**0.7797**
Bearing 3_3	0.6518	0.6050	**0.7024**

6.2.6.2 Evaluation of application economical efficiency

The potential economic loss caused by bearing failure is a significant factor in real applications. If the predicted failure time falls later than the actual failure time, it will cause unexpected breakdown and should be avoided. Based on the consideration, a modified CRA (MCRA) is proposed. In MCRA, the weights are assigned based on the prediction horizon and the absolute error (ahead or later) between the predicted RUL and the ground truth. It is given as:

$$MCRA_\lambda = \frac{1}{|\ell_{\lambda|}} \sum_{i=1}^{\ell_\lambda} w^*(r(i), D) RA_\lambda \tag{6.14}$$

where $D = sign(r_*(t_\lambda) - r^l(t_\lambda))$ is the sign of the prediction error, and weight w^* is defined as:

$$w^* = g(l, D) = \begin{cases} k_1 \cdot (t_f - l) + k_2, & D > 0 \\ k_3 \cdot (t_f - l) + k_4, & D < 0 \end{cases} \tag{6.15}$$

where l is the length of the prediction horizon, g is the function to estimate the unique weight for each prediction, $k = [0.005, 0.5, 0.008, 0.2]$ are the parameters for the two different linear weight functions, and t_f is the failure time. MCRA takes more application-related factors into account and gives award (penalty) to early (late) failure time prediction.

Table 6.8: Comparison of Different methods using RMSE and MCRA

Bearing No.	PF-RVM		Proposed	
	RMSE	MCRA	RMSE	MCRA
Bearing 1_1	4.3621	0.5136	**3.5322**	**0.6372**
Bearing 1_2	6.2031	0.5287	**5.2219**	**0.5973**
Bearing 1_3	4.6324	0.6637	**3.6304**	**0.7036**
Bearing 1_4	4.5682	0.6521	**3.8261**	**0.6713**
Bearing 1_5	6.4271	0.5931	**5.8595**	**0.6394**
Bearing 1_6	3.9271	0.6254	**3.0608**	**0.7530**

Table 6.8 presents and compares the RUL prediction performance with root mean square error (RMSE) and MRCA score, in which we program the algorithm of PF-RVM [132]. It can be seen that the proposed method has smaller RMSE and higher MCRA than the PF-RVM method for all the bearings, which fully demonstrates the performance of the proposed method.

6.3 Conclusions

Motivated by the challenges or open problems in bearing FDP, this chapter presents a systematic bearing FDP framework that integrates CNN-based fault detection and model selection, Bayesian estimation-based FDP, and DST-based prognostic fusion. This proposed method makes the most of the strengths and circumvents the demerits of deep learning methods and model-based methods, which combines the advantage of the powerful learning and pattern identification ability of CNN, uncertainty representation ability of Bayesian estimation, and information fusion of different resources. Two CNN models are trained to detect the STP and select

appropriate models for prognosis, which guarantees the computation efficiency and accuracy of FDP. DST fusion is applied to fuse the prognostic results when necessary. Experiments and comparisons show that the proposed method has high performance in accurate STP detection and RUL prediction.

Chapter 7

Conclusion and future work

This chapter concludes the main works of this thesis and summarizes the future work of this research.

7.1 Conclusions

Early and accurate bearing fault diagnosis is critical for condition-based maintenance and logistics, which can significantly reduce the cost of operating and maintenance. Current industrial systems often work in variable and complex operating conditions, which bring some challenges for existing DL-based FDP methods. The challenges can result in a low FDP accuracy and efficiency. To overcome the challenges, this thesis improves the existing DL-based fault diagnosis (classification) approach in terms of structure optimization, adaptive learning strategy, domain knowledge, and FDP algorithm execution strategy. Moreover, this thesis proposes a hybrid Bayesian estimation-based method for prognosis, which includes fault detection, fault model selection, PF-based prognosis, and DST-based prediction fusion.

First, this thesis proposes a DL-based approach, which integrates PSO and adaptive PReLU-DBN, for accurate and efficient rolling element-bearing diagnosis. In the proposed method, PCA is applied to the raw data to reduce the dimension of the input data. The optimal structure of DBN is determined by PSO. In the optimization of PSO, a PReLU activation layer and an adaptive training strategy are developed to speed up the training process. Experiments on a tapered roller bearing are conducted to demonstrate the effectiveness and efficiency of the methodology.

It avoids manual feature extraction and selection that requires extensive human involvement. Compared with traditional DBN structures, the proposed approach is able to achieve higher accuracy and faster convergence speed. More important, the proposed approach is a generic solution that can be applied to a variety of systems.

Second, a novel enhanced discriminate feature learning-based multi-task DR-CNN is proposed to improve the accuracy and efficiency of the bearing under varying operating conditions. The raw data from multi-sensors are converted and fused with domain knowledge to build information maps. Two different attention modules are employed to enhance the fault-related discriminate features learning ability. The constructed discriminate feature attention DR-CNN structure with two classifiers is assembled and trained with a dynamic training procedure using the fused information maps. Two case studies are conducted to verify the performance of the proposed approach. The proposed approach can achieve high training accuracy, fast convergence speed, and high diagnostic accuracy for multitasks with a single network. It demonstrates that the proposed multi-task bearing diagnosis network is stable and robust in different applications.

Third, this thesis proposes a novel multi-scale discriminate CNN-based bearing fault diagnosis for bearing fault diagnosis under varying speed. The raw data are converted and fused with varying speeds to build the information maps. Multi-head attention-based discriminate features learning framework is proposed for feature extraction. Two case studies are conducted to verify the performance of the proposed approach. The feature extraction capability is visualized using t-SNE dimensionality reduction technique, which demonstrates the powerful feature extraction ability of the proposed method. Experimental results and comparisons show that the proposed approach can achieve better performance in terms of accuracy and efficiency than some state-of-the-art methods.

Fourth, this thesis presents a hybrid bearing FDP framework that integrates

CNN-based fault detection and model selection, Bayesian estimation-based FDP, and DST-based prognostic fusion. This approach combines the advantage of the strong learning and pattern identification ability of CNN, the uncertainty representation ability of PF. Two CNN models are trained to detect the STP and select appropriate models for prognosis, which guarantees the computation efficiency and accuracy of FDP. DST fusion is applied to fuse the prognostic results when necessary. The proposed method can accurately estimate the real-time bearing fault condition and predict the RUL.

7.2 Future work

Future work will focus on transfer learning based on the proposed DL-based networks to improve the applicability and efficiency of the methods. The proposed DL-based methods have achieved satisfying performance on different bearing cases. The training and testing datasets that are used in this thesis are all from the same source, and the characteristic distributions of the training and testing are similar. However, in practice, the working conditions of bearings often change a lot in the service life, which makes the data show different characteristic distributions. The fault diagnosis accuracy cannot be guaranteed when the developed network is trained under one working condition or task but applied on the bearings under other working conditions or tasks. Besides, there are some other problems in real applications, such as the training and testing data accessibility, imbalance problem of collected data, low data volume, etc. These problems will all hinder the application of the proposed methods. Therefore, future work will focus on transfer learning based on the proposed networks to improve their practicability, and allow the proposed methods to be applicable to different bearing cases.

Transfer learning-based fault diagnosis is a cross-domain learning problem. The model is trained with source domain data and tested with the data from the

target domain. Future work will focus on developing novel cross-domain adaption mechanisms based on the developed DL-based networks to improve the usability of the methods.

Chapter 8

Appendix A: Publications

1. **Niu, G.**, Wang, X., Golda, M., Mastro, S. and Zhang, B., 2021. An optimized adaptive PReLU-DBN for rolling element bearing fault diagnosis. **Neurocomputing**, 445, pp.26-34.

2. **Niu, G.**, Liu, E., Wang, X., Ziehl, P. and Zhang, B., 2022. Enhanced Discriminate Feature Learning Deep Residual CNN for Multitask Bearing Fault Diagnosis With Information Fusion. **IEEE Transactions on Industrial Informatics**, 19(1), pp.762-770.

3. **Niu, G.**, Wang, X., Liu, E. and Zhang, B., 2021. Lebesgue sampling based deep belief network for lithium-ion battery diagnosis and prognosis. **IEEE Transactions on Industrial Electronics**, 69(8), pp.8481-8490.

4. Lyu, D., **Niu, G.**, Liu, E., Zhang, B., Chen, G., Yang, T. and Zio, E., 2022. Time space modelling for fault diagnosis and prognosis with uncertainty management: A general theoretical formulation. **Reliability Engineering & System Safety**, 226, p.108686.

5. Weddington, J., **Niu, G.**, Chen, R., Yan, W. and Zhang, B., 2021. Lithium-ion battery diagnostics and prognostics enhanced with Dempster-Shafer decision fusion. **Neurocomputing**, 458, pp.440-453.

6. Liu, E., **Niu, G.**, Wang, X. and Zhang, B., 2021, October. SOH Diagnostic and Prognostic Based on External Health Indicator of Lithium-ion Batteries.

In IECON 2021–47th Annual Conference of the IEEE Industrial Electronics Society (pp. 1-6). IEEE.

7. Liu, E., Wang, X., **Niu, G.**, Lyu, D., Yang, T. and Zhang, B., 2021. Lebesgue sampling-based li-ion battery simplified first principle model for soc estimation and rdt prediction. **IEEE Transactions on Industrial Electronics**, 69(9), pp.9524-9534.

8. Lyu, D., **Niu, G.**, Liu, E., Yang, T., Chen, G. and Zhang, B., 2021. Uncertainty management and differential model decomposition for fault diagnosis and prognosis. **IEEE Transactions on Industrial Electronics**, 69(5), pp.5235-5246.

9. **Niu, G.**, Liu, E., Zhang, B., Golda, M. and Mastro, S., 2021, May. A deep residual convolutional neural network based bearing fault diagnosis with multi-sensor data. In 2021 4th IEEE International Conference on Industrial Cyber-Physical Systems (ICPS) (pp. 655-660). IEEE.

10. Lyu, D., **Niu, G.**, Yang, T., Gang, C. and Zhang, B., 2021, May. Uncertainty Analysis in the Application of Fault Diagnosis and Prognosis. In 2021 4th IEEE International Conference on Industrial Cyber-Physical Systems (ICPS) (pp. 686-690). IEEE.

11. Zhang, H., **Niu, G.**, Zhang, B. and Miao, Q., 2021. Cost-effective lebesgue sampling long short-term memory networks for lithium-ion batteries diagnosis and prognosis. **IEEE Transactions on Industrial Electronics**, 69(2), pp.1958-1967.

12. Zhao, G., Wu, Z., Gao, Y., **Niu, G.**, Wang, Z.L. and Zhang, B., 2020. Multi-layer extreme learning machine-based keystroke dynamics identification for intelligent keyboard. **IEEE Sensors Journal**, 21(2), pp.2324-2333.

13. Lyu, D., **Niu, G.**, Zhang, B., Chen, G. and Yang, T., 2020. Lebesgue-time–space-model-based diagnosis and prognosis for multiple mode systems. **IEEE Transactions on Industrial Electronics**, 68(2), pp.1591-1603.

14. **Niu, G.**, Zhang, B., Ziehl, P., Ferrese, F. and Golda, M., 2019, September. Rolling element bearing fault diagnosis based on deep belief network and principal component analysis. In Proceedings of the Annual Conference of the PHM Society (Vol. 11, No. 1, pp. 1-9).

15. Liu, E., **Niu, G.**, Tang, S., Zhang, B., Williams, J., Martin, R. and Moore, C., 2019, September. Permanent magnet synchronous motor winding fault simulation and diagnosis. In Annual Conference of the PHM Society (Vol. 11, No. 1).

16. Lyu, D., **Niu, G.**, Hu, D., Chen, G., Yang, T. and Zhang, B., 2019, August. A Lebesgue-Time-Space-Model and Particle Filter based Diagnosis and Prognosis Method. In 2019 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC) (pp. 278-283). IEEE.

17. Zhao, G., Yang, J., Chen, J., Zhu, G., Jiang, Z., Liu, X., **Niu, G.**, Wang, Z.L. and Zhang, B., 2019. Keystroke dynamics identification based on triboelectric nanogenerator for intelligent keyboard using deep learning method. **Advanced Materials Technologies**, 4(1), p.1800167.

18. **Niu, G.**, Tang, S. and Zhang, B., 2018, October. Machine condition prediction based on long short term memory and particle filtering. In IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society (pp. 5942-5947). IEEE.

19. Mayfield, E.W., **Niu, G.**, Zhang, B., Ziehl, P. and Golda, M., 2018, September. An Integrated Health Management System Approach: Application to Shipboard Rotating Machinery. In Annual Conference of the PHM Society (Vol. 10, No. 1).

20. **Niu, G.**, Tang, S., Liu, Z., Zhao, G. and Zhang, B., 2018, September. Fault diagnosis and prognosis based on deep belief network and particle filtering. In Annual Conference of the Prognostics and Health Management Society.

21. Yan, W., Zhang, B., Zhao, G., Tang, S., **Niu, G.** and Wang, X., 2018. A battery management system with a Lebesgue-sampling-based extended Kalman filter. IEEE transactions on industrial electronics, 66(4), pp.3227-3236.

22. Zhao, G., Liu, X., Zhang, B., Liu, Y., **Niu, G.** and Hu, C., 2018. A novel approach for analog circuit fault diagnosis based on deep belief network. Measurement, 121, pp.170-178.

23. Yan, W., **Niu, G.**, Tang, S. and Zhang, B., 2017, October. State-of-charge estimation of Lithium-ion batteries by Lebesgue sampling-based EKF method. In IECON 2017-43rd Annual Conference of the IEEE Industrial Electronics Society (pp. 3233-3238). IEEE.

24. Zhao, G., Liu, X., Zhang, B., Zhang, G., **Niu, G.** and Hu, C., 2017, October. Bearing health condition prediction using deep belief network. In Annual Conference of the Prognostics and Health Management Society.

25. Yan, W., Zhang, B., Zhao, G., Weddington, J. and **Niu, G.**, 2017. Uncertainty management in Lebesgue-sampling-based diagnosis and prognosis for lithium-ion battery. **IEEE Transactions on Industrial Electronics**, 64(10), pp.8158-8166.

Bibliography

[1] P. Qian, X. Ma, D. Zhang, and J. Wang. "Data-driven condition monitoring approaches to improving power output of wind turbines". In: *IEEE Transactions on Industrial Electronics* 66.8 (2018), pp. 6012–6020.

[2] Q. Han, Z. Ding, X. Xu, T. Wang, and F. Chu. "Stator current model for detecting rolling bearing faults in induction motors using magnetic equivalent circuits". In: *Mechanical Systems and Signal Processing* 131 (2019), pp. 554–575.

[3] M. S. Kan, A. C. Tan, and J. Mathew. "A review on prognostic techniques for non-stationary and non-linear rotating systems". In: *Mechanical Systems and Signal Processing* 62 (2015), pp. 1–20.

[4] B. Zhang, G. Georgoulas, M. Orchard, A. Saxena, D. Brown, G. Vachtsevanos, and S. Liang. "Rolling element bearing feature extraction and anomaly detection based on vibration monitoring". In: *Control and Automation, 2008 16th Mediterranean Conference on*. IEEE. 2008, pp. 1792–1797.

[5] M. Cerrada, R.-V. Sánchez, C. Li, F. Pacheco, D. Cabrera, J. V. de Oliveira, and R. E. Vasquez. "A review on data-driven fault severity assessment in rolling bearings". In: *Mechanical Systems and Signal Processing* 99 (2018), pp. 169–196.

[6] Y. Wei, Y. Li, M. Xu, and W. Huang. "A review of early fault diagnosis approaches and their applications in rotating machinery". In: *Entropy* 21.4 (2019), p. 409.

[7] X. Wang, Y. Zheng, Z. Zhao, and J. Wang. "Bearing fault diagnosis based on statistical locally linear embedding". In: *Sensors* 15.7 (2015), pp. 16225–16247.

[8] K Christian, N. Mureithi, A. Lakis, and M. Thomas. "On the use of time synchronous averaging, independent component analysis and support vector machines for bearing fault diagnosis". In: *First International Conference On Industrial Risk Engineering, Montreal*. 2007.

[9] F. Dalvand, S. Dalvand, F. Sharafi, and M. Pecht. "Current noise cancellation for bearing fault diagnosis using time shifting". In: *IEEE Transactions on Industrial Electronics* 64.10 (2017), pp. 8138–8147.

[10] N Tandon. "A comparison of some vibration parameters for the condition monitoring of rolling element bearings". In: *Measurement* 12.3 (1994), pp. 285–289.

[11] B. Nayana and P Geethanjali. "Analysis of statistical time-domain features effectiveness in identification of bearing faults from vibration signal". In: *IEEE Sensors Journal* 17.17 (2017), pp. 5618–5625.

[12] H. Zoubek, S. Villwock, and M. Pacas. "Frequency response analysis for rolling-bearing damage diagnosis". In: *IEEE Transactions on Industrial Electronics* 55.12 (2008), pp. 4270–4276.

[13] D. Wang, W. T. Peter, and K. L. Tsui. "An enhanced Kurtogram method for fault diagnosis of rolling element bearings". In: *Mechanical Systems and Signal Processing* 35.1-2 (2013), pp. 176–199.

[14] H. Cao, F. Fan, K. Zhou, and Z. He. "Wheel-bearing fault diagnosis of trains using empirical wavelet transform". In: *Measurement* 82 (2016), pp. 439–449.

[15] 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". In: *Applied Acoustics* 89 (2015), pp. 16–27.

[16] Z. Wang, L. Yao, and Y. Cai. "Rolling bearing fault diagnosis using generalized refined composite multiscale sample entropy and optimized support vector machine". In: *Measurement* 156 (2020), p. 107574.

[17] G. Niu, S. Tang, Z. Liu, G. Zhao, and B. Zhang. "Fault diagnosis and prognosis based on deep belief network and particle filtering". In: *Annual Conference of the PHM Society*. Vol. 10. 1. 2018.

[18] I. Attoui, N. Boutasseta, N. Fergani, B. Oudjani, and A. Deliou. "Vibration-based bearing fault diagnosis by an integrated DWT-FFT approach and an adaptive neuro-fuzzy inference system". In: *2015 3rd International Conference on Control, Engineering & Information Technology (CEIT)*. IEEE. 2015, pp. 1–6.

[19] X. Lou and K. A. Loparo. "Bearing fault diagnosis based on wavelet transform and fuzzy inference". In: *Mechanical systems and signal processing* 18.5 (2004), pp. 1077–1095.

[20] J. Ma, Z. Li, C. Li, L. Zhan, and G.-Z. Zhang. "Rolling Bearing Fault Diagnosis Based on Refined Composite Multi-Scale Approximate Entropy and Optimized Probabilistic Neural Network". In: *Entropy* 23.2 (2021), p. 259.

[21] R. Nishat Toma and J.-M. Kim. "Bearing fault classification of induction motors using discrete wavelet transform and ensemble machine learning algorithms". In: *Applied Sciences* 10.15 (2020), p. 5251.

[22] K. Yu, T. R. Lin, J. Tan, and H. Ma. "An adaptive sensitive frequency band selection method for empirical wavelet transform and its application in bearing fault diagnosis". In: *Measurement* 134 (2019), pp. 375–384.

[23] R. Li, C. Ran, B. Zhang, L. Han, and S. Feng. “Rolling bearings fault diagnosis based on improved complete ensemble empirical mode decomposition with adaptive noise, nonlinear entropy, and ensemble SVM”. In: *Applied Sciences* 10.16 (2020), p. 5542.

[24] A. Rai and S. Upadhyay. “Bearing performance degradation assessment based on a combination of empirical mode decomposition and k-medoids clustering”. In: *Mechanical Systems and Signal Processing* 93 (2017), pp. 16–29.

[25] Z. Liu, J. Ding, J. Lin, and Y. Huang. “A rolling bearing fault diagnosis-optimized scale-space representation for the empirical wavelet transform”. In: *Shock and Vibration* 2018 (2018).

[26] Y. Xu, Y. Deng, J. Zhao, W. Tian, and C. Ma. “A novel rolling bearing fault diagnosis method based on empirical wavelet transform and spectral trend”. In: *IEEE Transactions on Instrumentation and Measurement* 69.6 (2019), pp. 2891–2904.

[27] Y. Cheng, Z. Wang, B. Chen, W. Zhang, and G. Huang. “An improved complementary ensemble empirical mode decomposition with adaptive noise and its application to rolling element bearing fault diagnosis”. In: *ISA transactions* 91 (2019), pp. 218–234.

[28] M. Ali and R. Prasad. “Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition”. In: *Renewable and Sustainable Energy Reviews* 104 (2019), pp. 281–295.

[29] H. Habbouche, Y. Amirat, T. Benkedjouh, and M. Benbouzid. “Bearing Fault Event-Triggered Diagnosis using a Variational Mode Decomposition-based Machine Learning Approach”. In: *IEEE Transactions on Energy Conversion* (2021).

[30] X. Chen, Y. Yang, Z. Cui, and J. Shen. "Vibration fault diagnosis of wind turbines based on variational mode decomposition and energy entropy". In: *Energy* 174 (2019), pp. 1100–1109.

[31] A. Yang, Y. Wang, Y. Zi, and T. W. Chow. "An enhanced trace ratio linear discriminant analysis for fault diagnosis: An illustrated example using HDD data". In: *IEEE Transactions on Instrumentation and Measurement* 68.12 (2019), pp. 4629–4639.

[32] J. Zheng, J. Cheng, and Y. Yang. "A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy". In: *Mechanism and Machine Theory* 70 (2013), pp. 441–453.

[33] X. Zhang and J. Zhou. "Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines". In: *Mechanical Systems and Signal Processing* 41.1-2 (2013), pp. 127–140.

[34] J. Zheng, H. Pan, and J. Cheng. "Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines". In: *Mechanical Systems and Signal Processing* 85 (2017), pp. 746–759.

[35] C. Li, J. V. De Oliveira, M. Cerrada, D. Cabrera, R. V. Sánchez, and G. Zurita. "A systematic review of fuzzy formalisms for bearing fault diagnosis". In: *IEEE Transactions on Fuzzy Systems* 27.7 (2018), pp. 1362–1382.

[36] R. Liu, B. Yang, E. Zio, and X. Chen. "Artificial intelligence for fault diagnosis of rotating machinery: A review". In: *Mechanical Systems and Signal Processing* 108 (2018), pp. 33–47.

[37] J. Singh, M. Azamfar, F. Li, and J. Lee. "A systematic review of machine learning algorithms for prognostics and health management of rolling element

bearings: fundamentals, concepts and applications". In: *Measurement Science and Technology* 32.1 (2020), p. 012001.

[38] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi. "Applications of machine learning to machine fault diagnosis: A review and roadmap". In: *Mechanical Systems and Signal Processing* 138 (2020), p. 106587.

[39] A. Carrio, C. Sampedro, A. Rodriguez-Ramos, and P. Campoy. "A review of deep learning methods and applications for unmanned aerial vehicles". In: *Journal of Sensors* 2017 (2017).

[40] D. Yu and L. Deng. *Automatic Speech Recognition.* Springer, 2016.

[41] D.-T. Hoang and H.-J. Kang. "A survey on deep learning based bearing fault diagnosis". In: *Neurocomputing* 335 (2019), pp. 327–335.

[42] C. Shen, J. Xie, D. Wang, X. Jiang, J. Shi, and Z. Zhu. "Improved hierarchical adaptive deep belief network for bearing fault diagnosis". In: *Applied Sciences* 9.16 (2019), p. 3374.

[43] J. Xie, G. Du, C. Shen, N. Chen, L. Chen, and Z. Zhu. "An end-to-end model based on improved adaptive deep belief network and its application to bearing fault diagnosis". In: *IEEE Access* 6 (2018), pp. 63584–63596.

[44] S. Gao, L. Xu, Y. Zhang, and Z. Pei. "Rolling bearing fault diagnosis based on intelligent optimized self-adaptive deep belief network". In: *Measurement Science and Technology* 31.5 (2020), p. 055009.

[45] H. Shao, H. Jiang, F. Wang, and Y. Wang. "Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet". In: *ISA transactions* 69 (2017), pp. 187–201.

[46] T. Liang, S. Wu, W. Duan, and R. Zhang. "Bearing fault diagnosis based on improved ensemble learning and deep belief network". In: *Journal of Physics: Conference Series*. Vol. 1074. 1. IOP Publishing. 2018, p. 012154.

[47] H. Shao, H. Jiang, X. Zhang, and M. Niu. "Rolling bearing fault diagnosis using an optimization deep belief network". In: *Measurement Science and Technology* 26.11 (2015), p. 115002.

[48] X. Wang, J. Huang, G. Ren, and D. Wang. "A hydraulic fault diagnosis method based on sliding-window spectrum feature and deep belief network". In: *Journal of Vibroengineering* 19.6 (2017), pp. 4272–4284.

[49] C. Zhang, Y. Zhang, C. Hu, Z. Liu, L. Cheng, and Y. Zhou. "A novel intelligent fault diagnosis method based on variational mode decomposition and ensemble deep belief network". In: *IEEE Access* 8 (2020), pp. 36293–36312.

[50] S. Albawi, T. A. Mohammed, and S. Al-Zawi. "Understanding of a convolutional neural network". In: *2017 International Conference on Engineering and Technology (ICET)*. Ieee. 2017, pp. 1–6.

[51] D. Wang, Q. Guo, Y. Song, S. Gao, and Y. Li. "Application of multiscale learning neural network based on CNN in bearing fault diagnosis". In: *Journal of Signal Processing Systems* 91.10 (2019), pp. 1205–1217.

[52] J. Chen, R. Huang, K. Zhao, W. Wang, L. Liu, and W. Li. "Multiscale Convolutional Neural Network With Feature Alignment for Bearing Fault Diagnosis". In: *IEEE Transactions on Instrumentation and Measurement* 70 (2021), pp. 1–10.

[53] W. Fuan, J. Hongkai, S. Haidong, D. Wenjing, and W. Shuaipeng. "An adaptive deep convolutional neural network for rolling bearing fault diagnosis". In: *Measurement Science and Technology* 28.9 (2017), p. 095005.

[54] X. Guo, L. Chen, and C. Shen. "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis". In: *Measurement* 93 (2016), pp. 490–502.

[55] R. Liu, F. Wang, B. Yang, and S. J. Qin. “Multiscale kernel based residual convolutional neural network for motor fault diagnosis under nonstationary conditions”. In: *IEEE Transactions on Industrial Informatics* 16.6 (2019), pp. 3797–3806.

[56] J. Jiao, M. Zhao, J. Lin, and C. Ding. “Deep coupled dense convolutional network with complementary data for intelligent fault diagnosis”. In: *IEEE Transactions on Industrial Electronics* 66.12 (2019), pp. 9858–9867.

[57] S. Plakias and Y. S. Boutalis. “Fault detection and identification of rolling element bearings with Attentive Dense CNN”. In: *Neurocomputing* 405 (2020), pp. 208–217.

[58] N. F. Waziralilah, A. Abu, M. Lim, L. K. Quen, and A. Elfakharany. “A review on convolutional neural network in bearing fault diagnosis”. In: *MATEC Web of Conferences*. Vol. 255. EDP Sciences. 2019, p. 06002.

[59] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler. “Deep learning algorithms for bearing fault Diagnosticsx—A comprehensive review”. In: *IEEE Access* 8 (2020), pp. 29857–29881.

[60] C. Li, S. Zhang, Y. Qin, and E. Estupinan. “A systematic review of deep transfer learning for machinery fault diagnosis”. In: *Neurocomputing* 407 (2020), pp. 121–135.

[61] D. Neupane and J. Seok. “Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review”. In: *IEEE Access* 8 (2020), pp. 93155–93178.

[62] T. Chen, Z. Wang, X. Yang, and K. Jiang. “A deep capsule neural network with stochastic delta rule for bearing fault diagnosis on raw vibration signals”. In: *Measurement* 148 (2019), p. 106857.

[63] D. A. Tobon-Mejia, K. Medjaher, N. Zerhouni, and G. Tripot. "A data-driven failure prognostics method based on mixture of Gaussians hidden Markov models". In: *IEEE Transactions on reliability* 61.2 (2012), pp. 491–503.

[64] A. Cubillo, S. Perinpanayagam, and M. Esperon-Miguez. "A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery". In: *Advances in Mechanical Engineering* 8.8 (2016), p. 1687814016664660.

[65] X. Jin, J. Ni, et al. "Physics-based Gaussian process for the health monitoring for a rolling bearing". In: *Acta Astronautica* 154 (2019), pp. 133–139.

[66] Z. Tian, M. J. Zuo, and S. Wu. "Crack propagation assessment for spur gears using model-based analysis and simulation". In: *Journal of Intelligent Manufacturing* 23.2 (2012), pp. 239–253.

[67] R. Shao, P. Jia, and F. Dong. "Dynamic characteristics of cracked gear and three-dimensional crack propagation analysis". In: *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 227.6 (2013), pp. 1341–1361.

[68] W. Wu and C. Ni. "A study of stochastic fatigue crack growth modeling through experimental data". In: *Probabilistic Engineering Mechanics* 18.2 (2003), pp. 107–118.

[69] P. Rycerz, A. Olver, and A. Kadiric. "Propagation of surface initiated rolling contact fatigue cracks in bearing steel". In: *International Journal of Fatigue* 97 (2017), pp. 29–38.

[70] M. N. Kotzalas and T. A. Harris. "Fatigue failure progression in ball bearings". In: *J. Trib.* 123.2 (2001), pp. 238–242.

[71] C. H. Oppenheimer and K. A. Loparo. “Physically based diagnosis and prognosis of cracked rotor shafts”. In: *Component and Systems Diagnostics, Prognostics, and Health Management II*. Vol. 4733. International Society for Optics and Photonics. 2002, pp. 122–132.

[72] P. K. Gupta and E. V. Zaretsky. “New stress-based fatigue life models for ball and roller bearings”. In: *Tribology Transactions* 61.2 (2018), pp. 304–324.

[73] H. H. Chin. “Turbine engine hot section prognostics”. In: *Proceedings of the 59th meeting of the society for machinery failure prevention technology: essential technologies for successful prognostics*. 2005, pp. 45–53.

[74] G. Qiu, Y. Gu, and J. Chen. “Selective health indicator for bearings ensemble remaining useful life prediction with genetic algorithm and Weibull proportional hazards model”. In: *Measurement* 150 (2020), p. 107097.

[75] F. Wang, X. Chen, B. Dun, B. Wang, D. Yan, and H. Zhu. “Rolling bearing reliability assessment via kernel principal component analysis and weibull proportional hazard model”. In: *Shock and Vibration* 2017 (2017).

[76] Q. Zhang, C. Hua, and G. Xu. “A mixture Weibull proportional hazard model for mechanical system failure prediction utilising lifetime and monitoring data”. In: *Mechanical Systems and Signal Processing* 43.1-2 (2014), pp. 103–112.

[77] F. Di Maio, K. L. Tsui, and E. Zio. “Combining relevance vector machines and exponential regression for bearing residual life estimation”. In: *Mechanical systems and signal processing* 31 (2012), pp. 405–427.

[78] N. Li, Y. Lei, J. Lin, and S. X. Ding. “An improved exponential model for predicting remaining useful life of rolling element bearings”. In: *IEEE Transactions on Industrial Electronics* 62.12 (2015), pp. 7762–7773.

[79] G. Wang and J. Xiang. “Remain useful life prediction of rolling bearings based on exponential model optimized by gradient method”. In: *Measurement* 176 (2021), p. 109161.

[80] G. J. Vachtsevanos and G. J. Vachtsevanos. *Intelligent fault diagnosis and prognosis for engineering systems*. Vol. 456. Wiley Online Library, 2006.

[81] Y. Qian and R. Yan. “Remaining useful life prediction of rolling bearings using an enhanced particle filter”. In: *IEEE Transactions on Instrumentation and Measurement* 64.10 (2015), pp. 2696–2707.

[82] R. K. Singleton, E. G. Strangas, and S. Aviyente. “Extended Kalman filtering for remaining-useful-life estimation of bearings”. In: *IEEE Transactions on Industrial Electronics* 62.3 (2014), pp. 1781–1790.

[83] M. Jouin, R. Gouriveau, D. Hissel, M.-C. Péra, and N. Zerhouni. “Particle filter-based prognostics: Review, discussion and perspectives”. In: *Mechanical Systems and Signal Processing* 72 (2016), pp. 2–31.

[84] X. Jin, Y. Sun, Z. Que, Y. Wang, and T. W. Chow. “Anomaly detection and fault prognosis for bearings”. In: *IEEE Transactions on Instrumentation and Measurement* 65.9 (2016), pp. 2046–2054.

[85] N. Jammu and P. Kankar. “A review on prognosis of rolling element bearings”. In: *International Journal of Engineering Science and Technology* 3.10 (2011), pp. 7497–7503.

[86] S. Saon, T. Hiyama, et al. “Predicting remaining useful life of rotating machinery based artificial neural network”. In: *Computers & Mathematics with Applications* 60.4 (2010), pp. 1078–1087.

[87] W. Caesarendra, M. Pratama, B. Kosasih, T. Tjahjowidodo, and A. Glowacz. “Parsimonious network based on a fuzzy inference system (PANFIS) for time

series feature prediction of low speed slew bearing prognosis". In: *Applied Sciences* 8.12 (2018), p. 2656.

[88] M. Yan, X. Wang, B. Wang, M. Chang, and I. Muhammad. "Bearing remaining useful life prediction using support vector machine and hybrid degradation tracking model". In: *ISA transactions* 98 (2020), pp. 471–482.

[89] G. Zhao, X. Liu, B. Zhang, G. Zhang, G. Niu, and C. Hu. "Bearing Health Condition Prediction Using Deep Belief Network". In: *Proceedings of the Annual Conference of Prognostics and Health Management Society, Orlando, FL, USA*. 2017, pp. 2–5.

[90] M. Xia, X. Zheng, M. Imran, and M. Shoaib. "Data-driven prognosis method using hybrid deep recurrent neural network". In: *Applied Soft Computing* 93 (2020), p. 106351.

[91] D. Yao, B. Li, H. Liu, J. Yang, and L. Jia. "Remaining useful life prediction of roller bearings based on improved 1D-CNN and simple recurrent unit". In: *Measurement* 175 (2021), p. 109166.

[92] J. Deutsch, M. He, and D. He. "Remaining useful life prediction of hybrid ceramic bearings using an integrated deep learning and particle filter approach". In: *Applied Sciences* 7.7 (2017), p. 649.

[93] Y. Chen, Y. Li, S. Kurosu, Q. Meng, N. Tang, Y. Koizumi, and A. Chiba. "Analysis of run-in-stage wear behavior and contact mechanics of metal-on-metal hip joint bearings with different radial clearances". In: *Materials Transactions* (2015), p. M2014440.

[94] A. Rai and S. Upadhyay. "A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings". In: *Tribology International* 96 (2016), pp. 289–306.

[95] K. He, X. Zhang, S. Ren, and J. Sun. “Deep residual learning for image recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[96] S. Guo, B. Zhang, T. Yang, D. Lyu, and W. Gao. “Multitask convolutional neural network with information fusion for bearing fault diagnosis and localization”. In: *IEEE Transactions on Industrial Electronics* 67.9 (2019), pp. 8005–8015.

[97] L. Chen, Q. Li, C. Shen, J. Zhu, D. Wang, and M. Xia. “Adversarial domain-invariant generalization: a generic domain-regressive framework for bearing fault diagnosis under unseen conditions”. In: *IEEE Transactions on Industrial Informatics* (2021).

[98] J. Feng, Y. Yao, S. Lu, and Y. Liu. “Domain Knowledge-Based Deep-Broad Learning Framework for Fault Diagnosis”. In: *IEEE Transactions on Industrial Electronics* 68.4 (2020), pp. 3454–3464.

[99] G. E. Hinton, S. Osindero, and Y.-W. Teh. “A fast learning algorithm for deep belief nets”. In: *Neural computation* 18.7 (2006), pp. 1527–1554.

[100] G. Zhao, X. Liu, B. Zhang, Y. Liu, G. Niu, and C. Hu. “A novel approach for analog circuit fault diagnosis based on Deep Belief Network”. In: *Measurement* 121 (2018), pp. 170–178.

[101] M. M. Lau and K. H. Lim. “Review of adaptive activation function in deep neural network”. In: *2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*. IEEE. 2018, pp. 686–690.

[102] B. Ding, H. Qian, and J. Zhou. “Activation functions and their characteristics in deep neural networks”. In: *2018 Chinese Control And Decision Conference (CCDC)*. IEEE. 2018, pp. 1836–1841.

[103] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon. “Cbam: Convolutional block attention module”. In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 3–19.

[104] Y. Cao, J. Xu, S. Lin, F. Wei, and H. Hu. “Gcnet: Non-local networks meet squeeze-excitation networks and beyond”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*. 2019, pp. 0–0.

[105] X. Wang, R. Girshick, A. Gupta, and K. He. “Non-local neural networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 7794–7803.

[106] Z. Chen, A. Mauricio, W. Li, and K. Gryllias. “A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks”. In: *Mechanical Systems and Signal Processing* 140 (2020), p. 106683.

[107] K. He and J. Sun. “Convolutional neural networks at constrained time cost”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 5353–5360.

[108] S. A. McInerny and Y. Dai. “Basic vibration signal processing for bearing fault detection”. In: *IEEE Transactions on education* 46.1 (2003), pp. 149–156.

[109] D. P. Kingma and J. Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).

[110] E. Dogo, O. Afolabi, N. Nwulu, B Twala, and C. Aigbavboa. “A comparative analysis of gradient descent-based optimization algorithms on convolutional neural networks”. In: *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*. IEEE. 2018, pp. 92–99.

[111] F. Cong, J. Chen, G. Dong, and M. Pecht. “Vibration model of rolling element bearings in a rotor-bearing system for fault diagnosis”. In: *Journal of sound and vibration* 332.8 (2013), pp. 2081–2097.

[112] C. Lu, Z. Wang, and B. Zhou. “Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification”. In: *Advanced Engineering Informatics* 32 (2017), pp. 139–151.

[113] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang. “A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals”. In: *Sensors* 17.2 (2017), p. 425.

[114] R. Liu, G. Meng, B. Yang, C. Sun, and X. Chen. “Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine”. In: *IEEE Transactions on Industrial Informatics* 13.3 (2016), pp. 1310–1320.

[115] G. Niu, E. Liu, X. Wang, P. Ziehl, and B. Zhang. “Enhanced Discriminate Feature Learning Deep Residual CNN for Multi-task Bearing Fault Diagnosis with Information Fusion”. In: *IEEE Transactions on Industrial Informatics* (2022).

[116] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. “Attention is all you need”. In: *Advances in neural information processing systems* 30 (2017).

[117] H. Huang and N. Baddour. “Bearing vibration data collected under time-varying rotational speed conditions”. In: *Data in brief* 21 (2018), pp. 1745–1749.

[118] Y. Wang, X. Ding, R. Liu, and Y. Shao. “Conditionsensenet: A deep interpolatory convnet for bearing intelligent diagnosis under variational

working conditions". In: *IEEE Transactions on Industrial Informatics* 18.10 (2021), pp. 6558–6568.

[119] H. Zhou, X. Huang, G. Wen, S. Dong, Z. Lei, P. Zhang, and X. Chen. "Convolution enabled transformer via random contrastive regularization for rotating machinery diagnosis under time-varying working conditions". In: *Mechanical Systems and Signal Processing* 173 (2022), p. 109050.

[120] V Sinitsin, O Ibryaeva, V Sakovskaya, and V Eremeeva. "Intelligent bearing fault diagnosis method combining mixed input and hybrid CNN-MLP model". In: *Mechanical Systems and Signal Processing* 180 (2022), p. 109454.

[121] Z. Zhu, G. Peng, Y. Chen, and H. Gao. "A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis". In: *Neurocomputing* 323 (2019), pp. 62–75.

[122] L. Wen, X. Li, L. Gao, and Y. Zhang. "A new convolutional neural network-based data-driven fault diagnosis method". In: *IEEE Transactions on Industrial Electronics* 65.7 (2017), pp. 5990–5998.

[123] M. M. Islam and J.-M. Kim. "Automated bearing fault diagnosis scheme using 2D representation of wavelet packet transform and deep convolutional neural network". In: *Computers in Industry* 106 (2019), pp. 142–153.

[124] B. Yang, R. Liu, and E. Zio. "Remaining useful life prediction based on a double-convolutional neural network architecture". In: *IEEE Transactions on Industrial Electronics* 66.12 (2019), pp. 9521–9530.

[125] L. Aguiar-Conraria and M. J. Soares. "The continuous wavelet transform: Moving beyond uni-and bivariate analysis". In: *Journal of Economic Surveys* 28.2 (2014), pp. 344–375.

[126] B. Cai, L. Huang, and M. Xie. "Bayesian networks in fault diagnosis". In: *IEEE Transactions on industrial informatics* 13.5 (2017), pp. 2227–2240.

[127] L. Cui, W. Li, X. Wang, D. Zhao, and H. Wang. "Comprehensive remaining useful life prediction for rolling element bearings based on time-varying particle filtering". In: *IEEE Transactions on Instrumentation and Measurement* 71 (2022), pp. 1–10.

[128] J. Weddington, G. Niu, R. Chen, W. Yan, and B. Zhang. "Lithium-ion battery diagnostics and prognostics enhanced with Dempster-Shafer decision fusion". In: *Neurocomputing* 458 (2021), pp. 440–453.

[129] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, and C. Varnier. "PRONOSTIA: An experimental platform for bearings accelerated degradation tests." In: *IEEE International Conference on Prognostics and Health Management, PHM'12.* IEEE Catalog Number: CPF12PHM-CDR. 2012, pp. 1–8.

[130] A. Saxena, J. Celaya, E. Balaban, K. Goebel, B. Saha, S. Saha, and M. Schwabacher. "Metrics for evaluating performance of prognostic techniques". In: *2008 international conference on prognostics and health management.* IEEE. 2008, pp. 1–17.

[131] Z. Pan, Z. Meng, Z. Chen, W. Gao, and Y. Shi. "A two-stage method based on extreme learning machine for predicting the remaining useful life of rolling-element bearings". In: *Mechanical Systems and Signal Processing* 144 (2020), p. 106899.

[132] Y. Chang and H. Fang. "A hybrid prognostic method for system degradation based on particle filter and relevance vector machine". In: *Reliability Engineering & System Safety* 186 (2019), pp. 51–63.

[133] C. Zhang, P. Lim, A. K. Qin, and K. C. Tan. "Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics". In:

IEEE transactions on neural networks and learning systems 28.10 (2016), pp. 2306–2318.