공학박사 학위논문

딥러닝 기반 베어링 고장 진단 방법 개발

DEVELOPMENT OF DEEP LEARNING-BASED BEARING FAULT DIAGNOSIS METHODS

지도교수 강희준

이 논문을 공학박사 학위논문으로 제출함

2020 년 11 월

울산대학교 대학원
전기전자보시스템공학부
Duy Tang Hoang

UNIVERSITY OF ULSAN

DOCTORAL THESIS

Development of Deep Learning-based Bearing Fault Diagnosis Methods

Author:
Duy Tang Hoang

Supervisor:
Professor Hee Jun Kang

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy

in the

Department or Electrical Engineering

November 27, 2020

Duy Tang Hoang 의 공학박사 학위논문을 인준함

심사위원장 **조강현** 

심사위원 **김한실** 

심사위원 **서영수** 

심사위원 **강현덕** 

심사위원 **강희준** 

울산대학교 대학원

2020 년 11 월

Declaration of Authorship

I, Duy Tang Hoang, declare that this thesis titled, "Development of Deep Learning-based Bearing Fault Diagnosis Methods" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

_____

Date: November 27, 2020

_____

Acknowledgements

My great gratitude goes primarily to my supervisor Professor Hee Jun Kang for his support, patience, guidance, and enthusiastic assistance to my research. I deeply appreciate his time, effort, suggestions, and comments during my life in Korea.

I would like to give my special thanks to professor Han Sil Kim, professor Kang Hyun Jo, professor Young Soo Suh, and professor Hyun Duk Kang for serving as my supervisory committee members.

A special note of thanks goes to the alumni members of Intelligent Robotic System Lab: Dr. Mien Van, Dr. Xuan Toa Tran, Dr. Ngoc Bach Hoang, and Dr. Minh Duc Tran for their help when I first came to Korea.

A special word is sent to Dr. Phu-Nguyen Le, my very first lab mate, who started and finished the research career with me during such a long time at University of Ulsan.

Many thanks are given to lab members and friends in University of Ulsan.

Abstract

This thesis is a systematic research on bearing fault diagnosis using machine learning algorithms, especially on deep neural network or deep learning, an emerging topic of machine learning.

In the machine health monitoring area, bearing fault diagnosis is an important part because rolling element bearings are indispensable elements in rotary machines. Bearings are not only the most critical components but also the main contributor to the system failures, 45 - 55 % of equipment failure cases caused by broken of bearings. Any unexpected failure of bearings may cause sudden breakdown of the machine, even of the entire system, leading to huge financial losses.

The condition monitoring of a bearing can be considered as a pattern recognition task which has been successfully solved by intelligent diagnosis methods. According to the current literature, a general intelligent diagnosis methodology includes four steps as follows: data acquisition, feature extraction, feature selection, and feature classification.

Deep learning algorithms can learn multiple layers of representations from input data by deep architectures with many layers of data processing units. The output from a layer will be the input for its successive layer. Each layer can learn a higher level of data presentations from its preceding layer output. Therefore, DL architectures can automatically extract multiple complex features from the input data without human engineers.

The ultimate goal of this research is to develop fault diagnosis systems using existing deep neural networks and also to find novel deep learning algorithms.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Machine condition monitoring

Machine condition monitoring is a critical part of every industry to help production run smoothly and efficiently. In the 1960's condition monitoring has developed extensively and rapidly and became popular [1]. When operating under any conditions, electrical and mechanical systems always generate some kind of signal. If there are any changes in the intrinsic properties of the system, even if only slightly, the generated signals also alter. As a result, the operating condition and health status of a machine system can be monitored indirectly by monitoring and analyzing the generated signals from that system. Condition monitoring is the process of monitoring a parameter of conditions in machinery, in order to identify a significant change which is indicative of a developing fault.

The installation of condition monitoring systems in machinery systems results in major improvement in plant availability and in reduced costs. The condition monitoring may help to reduce the maintenance costs, extend the life and reliability of the machine, and avert the catastrophic failure [2].

A great number of methods have been proposed and extensively studied for machine condition monitoring. Some popular methods can be listed here are: analysis of lubricating oil [3], analysis of acoustic emission [4], electrical motor current analysis [5], and vibration signal analysis [6]. Among the existing condition monitoring methods, vibration analysis has been the most popular method. This method has been extensively

used in industry, since it can accurately identify 90% of all machine failures [2].

1.2 Bearing condition monitoring

Rolling element bearing (REB) is one of the most critical elements in industry. REBs are used extensively in every rotary machinery because of their low rolling friction. Bearings allow relative motion of joined parts in rotary machines by utilizing rolling contacts between the rolling elements and raceways to support the load while permitting constrained motion of one part to another. Bearing has two key functions: transfer motion and transmit force.

FIGURE 1.1: Ball bearing

Bearing are not only one of the most important components but also one of the first to fail [7]. Literature reviews by a lot of researchers show that:

- fewer than 20 % of bearings achieve of their design life [8],
- bearing failures account for more than 50 % of all motor failures [9],
- 45 - 55 % of equipment failures caused by breaking bearings [10, 11].

Any failure of bearings may cause a sudden breakdown of the machine, even of the entire system, leading to great financial losses. In some cases, the safety of human life can be threaten. Therefore, fault diagnosis of REBs is a major concern and has drawn substantial consideration of researchers.

1.3 Basic components of bearing

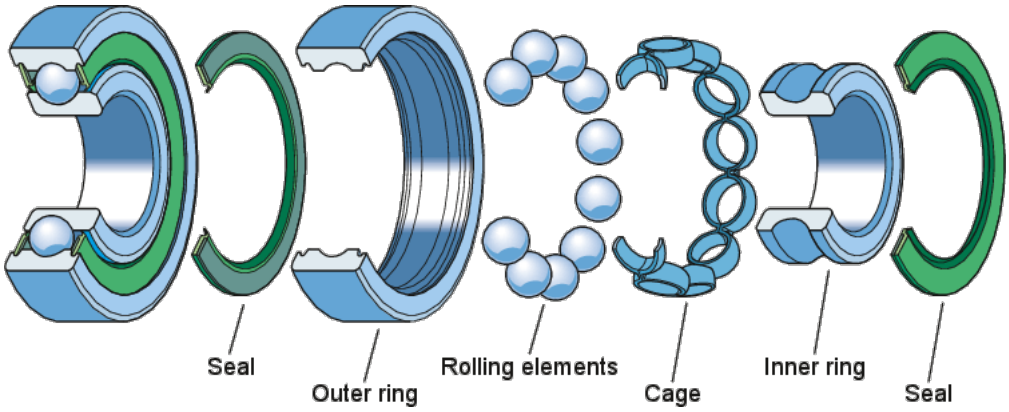

FIGURE 1.2: Components of Bearing

A typical rolling and its components are shown in Figure 1.2. Most bearings have four basic components: two rings, inner and outer, with raceways; rolling elements - rollers or balls; and a cage which keeps the rolling elements separated and guides motion. [12]

The inner and outer races are often made of high-purity chrome alloy steel. The inner race is mounted on the shaft of the machine and is the rotating part. The outer race is mounted in the machine housing and will not usually rotate.

The rolling elements can be balls, rollers, cones, spheres or needles. They are usually made from a special high-purity chrome alloy steel. The rolling elements roll on the raceway of the race and are separated and guided by the cage.

The cage keeps the rolling elements at an appropriate distance from each other and to prevent direct contact between neighboring rolling elements, in order to keep friction and thus heat generation as small as possible. The materials used to make the cages include steel, brass, and plastic.

1.3.1 Bearing fault

The proper description and categorization of bearing damages and their causes is an easy task as bearing damages occur as an interaction of different causes and conditions [13]. There are many reasons for bearing failure, including: excessive loading, inadequate lubrication, insufficient internal bearing clearance due to an excessively tight fit, etc [7]. The types of mechanical bearing failure and their relative frequencies are listed in Table 1.1 [14].

TABLE 1.1: Causes of bearing fault and corresponding frequency

Failure mechanism	Failure frequency
Fatigue	2 %
Lubrication	59 %
Dimensional discrepancies	29 %
Other	10 %

According to ISO 15243, bearing damages are classified into six main modes are: *fatigue*, *wear*, *corrosion*, *electrical erosion*, *plastic deformation*, and *fracture and cracking* [13]. The brief explanations are as follows.

- The primary failure mode for bearings is the spalling of the bearing elements due to local fatigue caused by the repeated application of stresses [15].
- Wear is a surface deterioration due to sliding friction at the surface of the raceway, rolling elements, and roller end faces. Wear can be caused by abrasive particles due to the lack of cleanliness during operation. Inadequate lubrication is also a reason of wear defects on bearings.
- Water and moisture in the corrosive operating environment are the factors that lead to bearing corrosion. Water and moisture form the rust, after that the lubricant fail to provide protection for the bearing surface.
- When an electric current passes through a bearing, which then proceeds from one ring to the other via the rolling elements, damage will occur. When the damage is existed, the material is heated to temperatures ranging from tempering to melting levels.

- There are many reasons to cause cracks occur in bearing rings. The most common cause is rough treatment when the bearings are being mounted or dismounted. The hammer blows can cause fine cracks which can grow and cause fragments of the ring to break-off when the bearing is put into operation. Another cause of ring cracking is excessive drive up on a tapered seating or sleeve.

1.3.2 Characteristics of bearing fault

In a mechanical system, when a rolling element strides a localized defect, impulses are created. The impulses are generated periodically and their characteristics depend on the location of the defect. The frequencies corresponding to the dynamic behaviors of each bearing component, they are f_o, f_i, f_r, and f_c corresponding to the outer race characteristic frequency, the inner race characteristic frequency, the roller characteristic frequency and the cage characteristic frequency. These bearing fundamental frequencies are computed as follows [7].

Outer race frequency

$$f_o = \frac{n}{2} f_s \left[1 - \frac{BD}{PD} \cos\beta \right] \tag{1.1}$$

Inner race frequency

$$f_i = n(f_s - f_c) \tag{1.2}$$

Roller frequency

$$f_r = \frac{PD}{BD} f_s \left[1 - \left(\frac{BD}{PD} \right)^2 \cos\beta \right] \tag{1.3}$$

Cage frequency

$$\frac{1}{2} f_s \left[1 - \frac{BD}{PD} \cos\beta \right] \tag{1.4}$$

where f_s is the shaft rotation frequency, n is the number of rolling elements, BD and PD are the ball diameter and the pitch circle diameter respectively, and β is the contact angle.

Computation of the bearing fundamental frequencies is helpful since they can indicate the damage parts in the bearing, i.e., on which component (inner race, outer race, roller, or cage) the fault is occurring.

1.4 Literature review on bearing fault diagnosis approaches

The signals generated from bearings under different conditions are complex and hard to extract useful information that help indicate the fault type and fault location. The signals are complex, non-stationary and contaminated by background noised. There are many signal processing techniques which can be divided into three categories: time-domain, frequency-domain, and time-frequency domain.

1.4.1 Time domain analysis

In this method, the raw signals measured from the machined are used directly to analyze. This method is fast, simple, and but not viable for noise signals [16]. Among the features of signal in time domain, the statistical features are extensively employed because they have close relations with bearing damage characteristics [16]. Short explanations and computation equations of time domain features are as follows [17, 18, 19].

Mean - The average of all values of the signal

$$x_{mean} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1.5}$$

Variance - The average of quadratic summation, which sums the square value of the difference value of each data and the mean. This value reflects the stability level of data.

$$x_{var} = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - x_{mean}\right)^2 \tag{1.6}$$

Standard deviation - The deviation from the mean of the signal. Standard deviation is the square root of the variance.

$$x_{std} = \sqrt{x_{var}} \tag{1.7}$$

Root mean square - Indicates the energy of the signal.

$$x_{rms} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{1.8}$$

Maximum amplitude - Maximum of all values in the signal.

$$x_{max} = \max(x_i) \tag{1.9}$$

Minimum amplitude - Minimum of all values in the signal.

$$x_{min} = \min(x_i) \tag{1.10}$$

Peak to peak value - Difference between maximum and minimum values.

$$x_{ppv} = x_{max} - xmin \tag{1.11}$$

Square root of amplitude - Value of the root of amplitude.

$$x_{sra} = \left(\frac{1}{N}\sum_{i=1}^{N}\sqrt{|x_i|}\right)^2 \tag{1.12}$$

Skewness - The characteristic parameter to attribute the asymmetry degree of the probability density curve relative to the mean. Skewness is the order three standard moments of the signal.

$$x_{skew} = E\left[\left(\frac{x-\mu}{\delta}\right)^3\right] \tag{1.13}$$

Kurtosis - Measure of the spikiness of the signal relative to a normal distribution.

$$x_{kurt} = \frac{\sum_{i=1}^{N}(x_i - x_{mean})}{(N-1)(x_{std})^4} - 3 \tag{1.14}$$

Clearance factor - Ratio of the maximum amplitude value to the square of the mean of the root of the absolute values.

$$x_{clf} = \frac{x_{max}}{x_{sra}} \tag{1.15}$$

Shape factor - Value of how much the shape of the signal is affected, other than shifting or scaling.

$$x_{sf} = \frac{x_{rms}}{(\frac{1}{N}\sum_{i=1}^{N}|x_i|)} \tag{1.16}$$

Impulse factor - Ratio of maximum amplitude value to the mean of absolute values.

$$x_{if} = \frac{x_{max}}{\frac{1}{N}\sum_{i=1}^{N}|x_i|} \tag{1.17}$$

Crest factor - Ratio of the maximum amplitude value to the root mean square value of the signal.

$$x_{cf} = \frac{x_{max}}{x_{rms}} \tag{1.18}$$

1.4.2 Frequency domain analysis

Frequency domain analysis is widely used in bearing condition monitoring. Investigation established that frequency domain analysis is generally more sensitive and reliable than time domain analysis [20]. In frequency domain signal analysis, the Fourier transform is by far the most popular technique [7]. As mention in 1.3.2, each component of a bearing has its own characteristic frequency, so does any fault associated with that component. The fundamental frequencies are a reliable indicator of the bearing conditions. The most popular approach in frequency analysis is envelope analysis [21]. Consider a localized defect hitting a raceway, each time the defect hits the raceway, an impulsive force (wide frequency content) is generated. This will excite structural resonances in the transmission bath between the point of impact and the point of measurement. Envelope analysis is a mechanism for extracting the periodic excitation or the amplitude modulation of the resonance, allowing the presence and location of a defect to be detected.

1.4.3 Time-frequency domain analysis

The frequency domain analysis transforms the signal into frequency domain. As a result, the time domain information is loss. On the contrary, the time domain analysis can not observe frequency domain information. The time-frequency domain analysis can conserve both time and frequency information. This approach of analysis is considered as the most effective. Some popular techniques in this approach can be listed are Short Time Furrier Transform (STFT), Wavelet Transform (WT) and Wigner-Ville Distribution (WVD) [22]. These techniques are reported having the ability to detect and diagnose bearing faults where the Signal-to-Noise Ratio (SNR) is low [23].

1.4.4 Intelligent fault diagnosis

In the intelligent diagnosis approach, the condition monitoring of a REB can be considered as a pattern recognition task. According to the current literature, as shown in Figure 1.3 a general intelligent diagnosis methodology includes four steps as follows: signal measurement, feature extraction, feature selection, and feature classification [24].

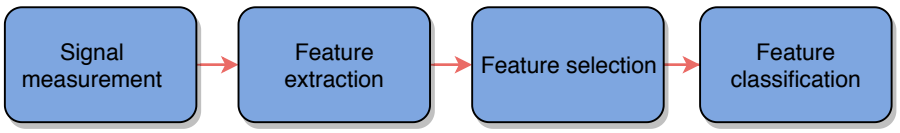

FIGURE 1.3: Fundamental steps of intelligent fault diagnosis

Data acquisition step collects signals which reflect the health status of bearings from sensor systems. These signals are vibration signals [25], acoustic emission signals [26], or electrical motor currents [27].

Feature extraction maps the original signals onto the statistical parameters which convey the information about the machine status. In the problem of pattern recognition, to obtain a high accuracy recognition result, the design of feature extractor takes an essential role [28]. Features of signals can be extracted from time domain [29], frequency domain [30], and time-frequency domain [31, 32]. The original bearing fault signals collected from rotary machines are in time domain. Bearing fault signals are also can be investigated in frequency domain and time-frequency domain by using the appropriate tools to transform them into the corresponding domains. Fourier Transform (FT)

is the most popular signal processing tool for transforming the signal into frequency domain. Time-frequency domain features can be extracted by Wavelet Packet Transform (WPT) [33], Dual-tree Complex Wavelet Transform (DT-CWT) [34], and Short-time Fourier Transform (STFT) [35].

The extracted feature set often has high dimensions so that it may contain some redundant and irrelevant features. High dimension feature set with many features may make it difficult to identify bearing faults. A large feature set often requires high computation effort, increases the learning time, and reduces the performance of classifiers. Thus, selecting the most discriminant features is an important step. Feature selection not only reduces the computation time but also increases the classification accuracy. Generally, there are two popular approaches for selecting features. The first approach is to generate a new feature set with lower dimension from the extracted feature set. Principal Component Analysis (PCA) [30] and Independent Component Analysis (ICA) [36] are two well-known methods of this approach. The second approach is to eliminate non-sensitive or useless features based on certain benchmarks. Sequential Selection [37] is a popular method of this approach.

Feature classification: once the salient features are selected, they are fed into an ML-based classifier such as k-Nearest Neighbor (kNN) [38], Artificial Neural Network (ANN) [39, 40], and Support Vector Machine (SVM) [41] to identify the bearing fault. Among the current ML based classification algorithms, ANN was proved as a powerful tool with high accuracy. There are many types of ANN were proposed such as Multilayer Perceptron (MLP), Recurrent Neural Network (RNN), Radial Basic Function Neural Network (RBFNN), and Probabilistic Neural Networks (PNN).

1.5 Data sources used for the research

In this research, two types of signals are the vibration signal and the motor current signal are considered. The vibration signal along with the vibration analysis are the most popular approaches in bearing fault diagnosis. The vibration signal data source used in this research is obtained from the well-known and popular Case Western Reserve University bearing data center [42].

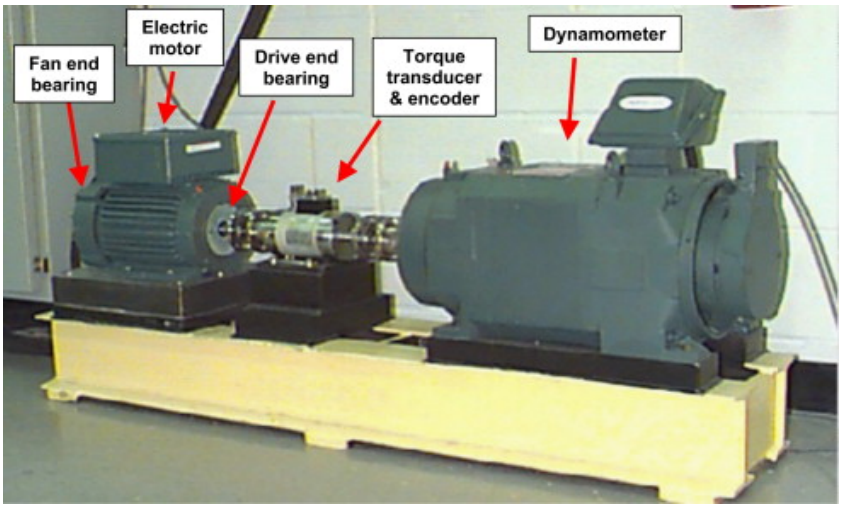

FIGURE 1.4: Case Western Reserve University bearing test-bed

The test-bed shown in Figure 1.4 includes a dynamometer (right), a 2 HP motor (left), a torque transducer/encoder (center). The test-bed also consists of a control electronics but was not shown in the figure. The motor shaft is supported by the test bearings. Single point faults were introduced to these bearings using electro-discharge machining with fault diameters of 7 mils, (1 mil = 0.001 inches). Vibration data are collected by using accelerometers, which are attached to the housing with magnetic bases. Accelerometers are placed at the 12 o'clock position at both the drive housing. Vibration signals are collected using a 16 channel DAT recorder including three operating conditions. The operating conditions are considered with bearing 6205-2RS JEM SKM, which is a deep groove ball bearing type.

The second type of signal used in this research is the motor current signal. The motor current signal analysis is considered as an non-destructive method because this method often does not require additional sensors. The motor current data source used in this research is obtained from the Kat-Data Center website of the Chair of Design and Drive Technology, Paderborn University, Germany [13]. The test-bed is shown in Figure 1.5.

Ball bearings with different types of defects were installed in the bearing test module to generate the experimental data. Bearing defects are real damages that were obtained from accelerated life test. Three types of ball bearing were used in the test-bed: bearings with fault at the outer race, bearings with fault at the inner race, and healthy bearings.

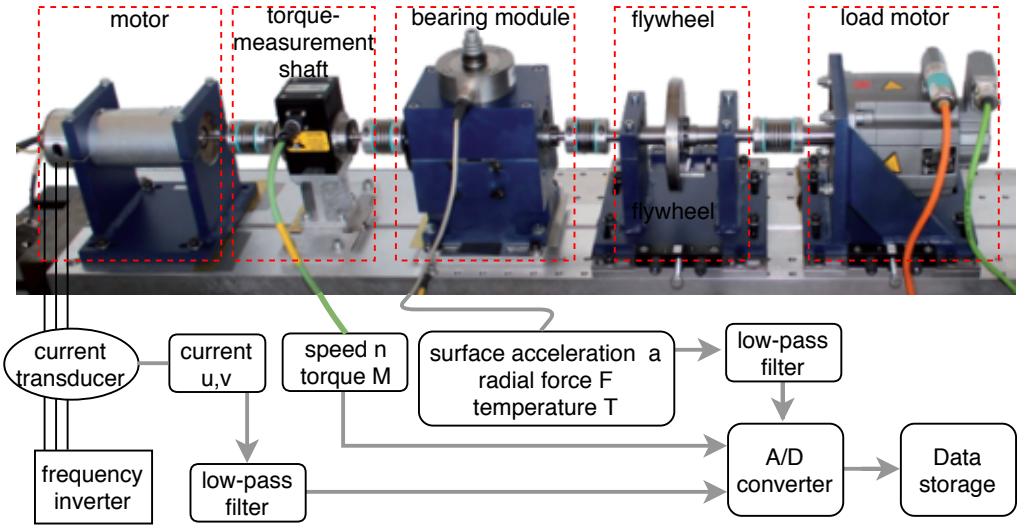

FIGURE 1.5: Kat-Data Center bearing test-bed

The test-bed operations could be varied by changing the rotational speed of the electric motor (S), the load torque of the load motor (M), and the radial force on the test bearings (F). The radial force on the bearings, the load torque at the torque-measuring shaft, the rotary speed and the oil temperature in the bearing module are measured synchronously to the motor currents and vibration signals.

The motor currents are measured at the two current phases by the current transducers of the type LEM CKSR 15-NP with an accuracy of 0.8 % of $I_{PN} = 15\ A$. The motor currents are then filtered by a 25 kHz low-pass filter and then converted to a digital signal with sampling rate 64 kHz. The vibration signal is also measured for comparing the vibration-signal-based fault diagnosis and current-signal-based fault diagnosis. The vibration signals are filtered by a 30 kHz low-pass filter and then converted to digital signals with sampling rate 64 kHz.

Chapter 2

Fundamental of deep learning

2.1 Introduction

Deep Learning (DL) is a branch of Machine Learning (ML), where multiple layers of data processing units are assembled to form deep architectures to extract multiple levels of data abstraction. The concept of deep learning appeared in the 1980s, but it has become more popular recently because of two main reasons [43]:

- The increase of the computational ability of processing units, especially Graphics Processing Units (GPUs), while the cost is reducing: DL algorithms often require strong computation efforts. The low cost and high computational ability help to implement, train, and perform DL algorithms more easily and quickly.
- Recent advances in ML research: Before the year of 2006, training deep architectures was very difficult and had unsuccessful results. A rational reason explaining for this problem is that the gradient-based optimization starting from random initialization of weight matrices often has poor results [44]. In 2006, Hinton et al. proposed an efficient method for training networks with deep structures, called greedy layer-wise training [45]. This work can be considered as the breakthrough which opened the fascinating era of ML research with deep architectures.

DL architectures such as Convolutional Neural Network (CNN), Stacked Autoencoder (SAE), Deep Belief Network (DBN), Deep Boltzmann Machine (DBM), and Recurrent Neural Network (RNN) have been applied successfully in many areas, which includes computer vision [46, 47], natural language processing [48, 49], medical image analysis

[50, 51], and machine health monitoring [52, 53]. DL algorithms can learn multiple layers of representations from input data by deep architectures with many layers of data processing units [54]. The output from a layer will be the input for its successive layer. Each layer can learn a higher level of data presentations from its preceding layer output. Therefore, DL architectures can automatically extract multiple complex features from the input data without human engineers: layers of features are extracted from raw data by a general purpose learning procedure [55].

2.2 Fundamental deep neural networks

Up to date, there are a lot of DL structures applied in various areas. In this section, we supply a brief introduction about the three most popular and fundamental structures in DL.

2.2.1 Autoencoder and Its Deep Models

Autoencoder

Autoencoder (AE) is an unsupervised DL algorithm. It was first proposed by Rumelhart et al. [56]. As shown in Figure 2.1, an AE is a special neural network that consists of three layers: input layer, hidden layer, and output layer. The difference is that in the structure of AE, the input and output layer have the same number of neurons.

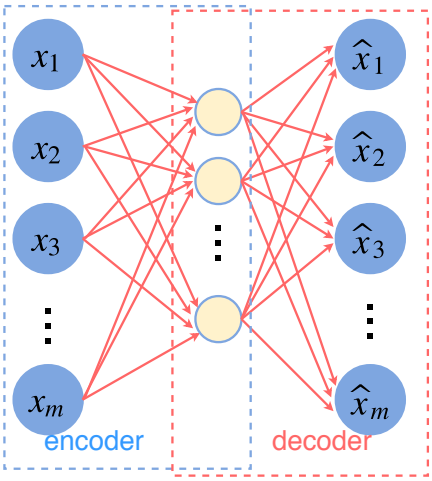

FIGURE 2.1: Structure of AE

The structure of an AE can be considered as an encoder which integrates with a decoder. The encoder includes the input layer and the hidden layer, mapping the input vector to the hidden layer. The decoder takes the output of the hidden layer to recreate the input values. Consider an input vector x, the forward direction computation of AE includes two steps: encoding and decoding. The encoding step maps the input vector into the hidden layer:

$$a^i = f(W_e x^i + b_e) \tag{2.1}$$

and the decoding step tries to reconstruct the input values from hidden values:

$$\hat{x}^i = f(W_d a^i + b_d) \tag{2.2}$$

where W_e, b_e and W_d, b_d are respectively the weight matrix - bias vector of the encoder and decoder. $f(.)$ denotes the activation function. With an input set including m samples $x^i, i = 1 : m$, the AE will produce m output samples $\hat{x}^i, i = 1 : m$. The loss function is defined as the following equation with squared error as:

$$J(W_e, W_d, b_e, b_d) = \frac{1}{2m} \sum_{i=1}^{m} (\hat{x}^i - x^i)^2 \tag{2.3}$$

or with the cross-entropy as:

$$J(W_e, W_d, b_e, b_d) = \frac{1}{m} \sum_{i=1}^{m} x^i \log(\hat{x}^i) + (1 - x_i) \log(1 - \hat{x}^i) \tag{2.4}$$

The training process minimizes the loss function and optimizes the AE parameters to reconstruct the output vector $\hat{x}$ so that the reconstruct error $(x - \hat{x})$ is as small as possible. After being trained, the AE can reconstruct the output from the original input with an arbitrary accuracy [56]. Since the output is reconstructed from the hidden vector, we can say that the hidden vector is a representation of the input data, i.e., the AE has learned representative features from the original input data and mapped into the hidden vector.

Denoising Autoencoder

Denoising Autoencoder (DAE) which first proposed by Vincent et al. [57] is another type of AE. A DAE has the same structure with the original AE, the difference is the way of feeding the input data into the network. First, the original input data is corrupted before being passed to the network. The procedure of encoding and decoding is similar to that of a typical AE. However, the DAE is trained to reconstruct the original input from the corrupted data version. The corrupted data can be produced by adding Gaussian noise, Masking noise, and Salt-and-pepper noise. The schematic diagram of the procedure of DAE is shown in Figure 2.2.

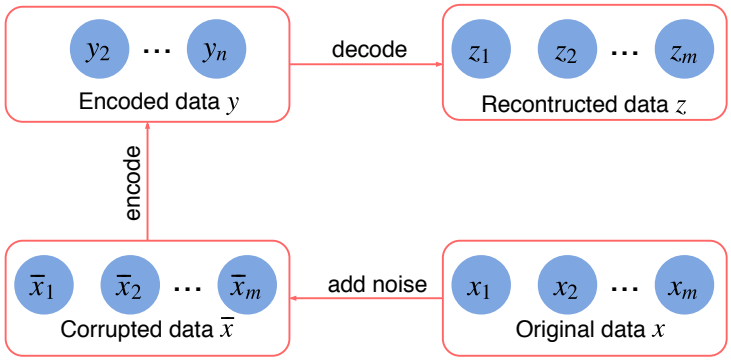

FIGURE 2.2: Structure of DAE

The original input data x^i is added noise to obtain the corrupted data $\bar{x}^i$. The corrupted input data are mapped into the hidden layer by:

$$y^i = f(W_e \bar{x}^i + b_e) \tag{2.5}$$

The the decoder phase of the AE reconstructs the output by:

$$z^i = f(W_d y^i + b_d) \tag{2.6}$$

The loss function is calculated by:

$$J(W_e, W_d, b_e, b_d) = \frac{1}{2m} \sum_{i=1}^{m} (z^i - x^i)^2 \tag{2.7}$$

Stacked Autoencoder

The architecture of a SAE is constructed by stacking multiple AEs together to form a deep model with many layers. The training process of an SAE called greedy layer-wise training proposed by Hinton et al. [45]. Consider a SAE constructed by n AE as shown in Figure 2.3.

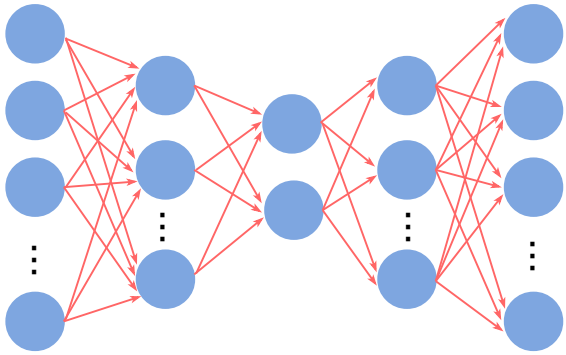

FIGURE 2.3: Stacked Autoencoder

The encoding operation is described by the following equation:

$$a^k = f(W_e^k a^{k-1} + b_e^k),\ k = 1 : n \tag{2.8}$$

where k denotes the k^{th} AE. a^k denotes the encoding result of k^{th} AE. When $k = 1$, $a^0 = \mathbf{x}$ is the input data. The decoding operation is described by the following equation:

$$c^k = f(W_d^{n-(k-1)} c^{k-1} + b_d^{n-(k-1)}),\ k = 1 : n \tag{2.9}$$

when $k = 1$, $c^0 = a^n$, when $k = n$, $c^n = \hat{\mathbf{x}}$ is the reconstructed data of the input data $\mathbf{x}$.

2.2.2 Restricted Boltzmann Machine and its deep models

Restricted Boltzmann Machine

Restricted Boltzmann Machine (RBM) is an unsupervised learning algorithm introduced by Smolensky [58]. The structure of a RBM is shown in Figure 2.4.

In the visible layer, the visible nodes are denoted by v_i. In the hidden layer, the hidden nodes are denoted by h_j. Nodes in the same layers are not connected. The weight which

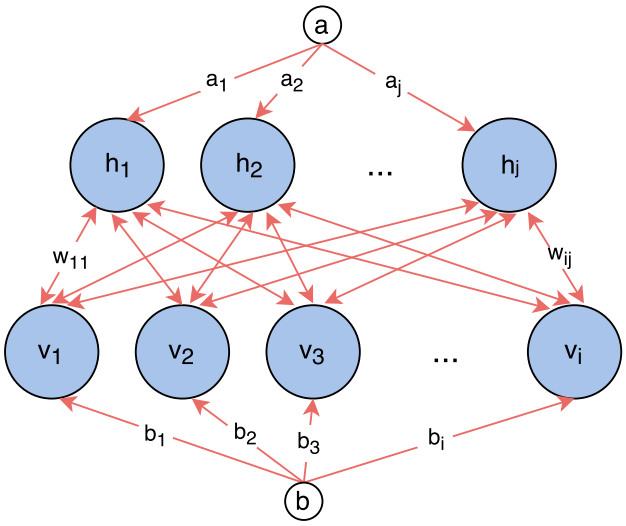

FIGURE 2.4: Structure of RBM

connects node v_i and h_j is denoted by w_{ij}. Each node has its own bias value. Visible node v_i has the corresponding bias value b_i; the hidden node h_j has the corresponding bias value c_j. The relationship of the visible layer and hidden layer is defined by the energy function given by:

$$E(v,h) = -\sum_i v_i b_i - \sum_j h_j c_j - \sum_{i,j} v_i h_j w_{ij} \tag{2.10}$$

The eventual goal of RBM training is to optimize the parameter set $\theta = w_{ij}, b_i, c_j$ that minimizes the model energy and balances the model at a finite state. Every possible case of a pair visible-hidden nodes is assigned a probability by the energy function:

$$p(v,h) = \frac{1}{Z} e^{-E(v,h)} \tag{2.11}$$

where Z is the partition function, calculated by summing all possible visible-hidden node pairs:

$$Z = \sum_{v,h} e^{-E(v,h)} \tag{2.12}$$

Since nodes in the same layer are not connected, the conditional probability distributions of each unit are described by the following equations:

$$p(v_i = 1|h) = \frac{1}{1 + exp(-c_j - \sum_j w_{ij} h_j)} \tag{2.13}$$

$$p(h_j = 1|v) = \frac{1}{1 + exp(-b_i - \sum_i w_{ij} v_i)} \tag{2.14}$$

The training process for RBM is to maximize the joint probability. The parameters can be trained by Contrastive Divergence (CD) algorithm proposed by Hinton [59].

Deep models of Restricted Boltzmann Machine

DBN proposed by Hinton [60] is a deep model constructed by stacking multiple RBMs, where the input of a layer is the output of the preceding layer. DBNs can also be trained with the greedy layer-wise training [44].

DBM introduced by Salakhutdinov and Hinton [61] is a deep model with many hidden layers stacked into a hierarchy structure. The difference between a DBM and a DBN is that a DBM is an indirect model while a DBN is a direct model. DBMs can also be trained efficiently by the greed layer-wise algorithm [62].

2.2.3 Convolutional Neural Network

Convolutional Neural Network (CNN) is a type of feed-forward neural network which constructed by three types of layers: Convolutional Layer (CL), Pooling Layer (PL), and Fully-connected Layer. The Fully-connected Layer has the same structure and the way of operation with the conventional feed-forward neural network. Advantages of CNN come from the differences in the structures and operations of CLs and PLs. An illustration of CNN structure for processing 2-D data is shown in Figure 2.5.

A CL consists of multiple learnable kernels. Each kernel has a trainable weight and bias. The CL convolves the input data with kernels in that layer. The result of the convolution operation then will be fed into an activate function to produce the final output of that CL. The math operation in the l^{th} layer between the specific j^{th} and the

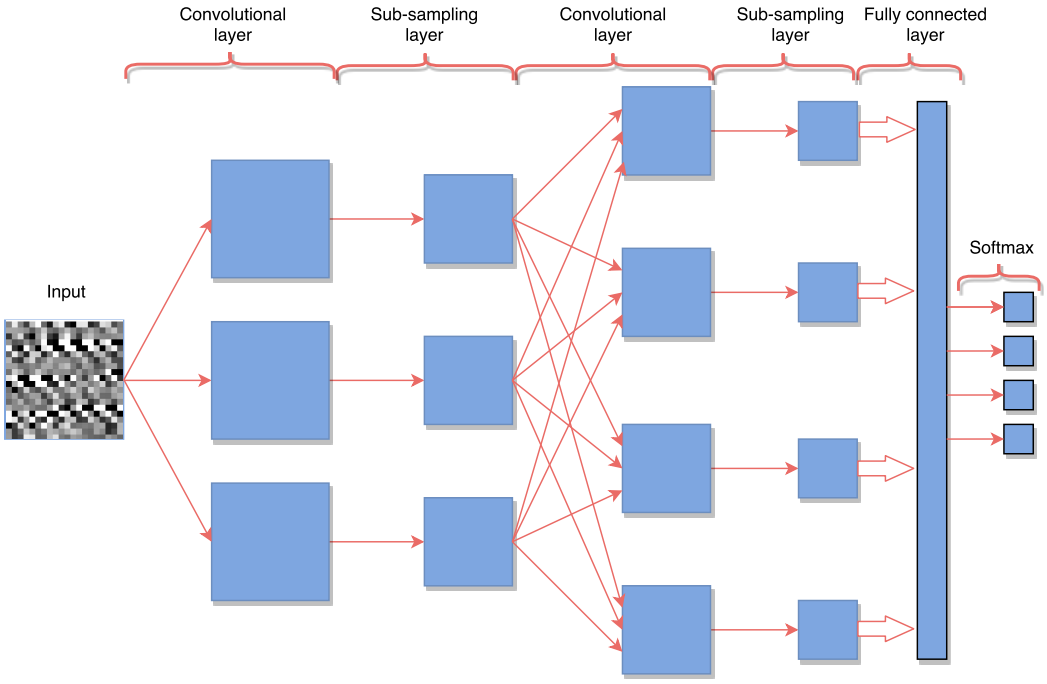

FIGURE 2.5: Convolutional Neural Network

input data x^{l-1} can be described by the following equation:

$$x_j^l = f\Big(\sum_{i \in M_j} x_i^{l-1} * k_j^l + b_j^l\Big) \tag{2.15}$$

where $(*)$ represents the convolution operation. The explanation for the above equation is as follows. Assume that the input data x^{l-1} includes m 2-D matrices. Every input matrix $x_i^{l-1}(\ i \in m)$ is convolved with the kernel k_j. Then the sum of all convolution operation results will be added to the bias. Finally, the result will be fed into the activate function f to produce the final output of kernel j. The purpose of CL is to extract local features from the input data. Weight sharing is a term often used for the way of using kernels for all input maps of data.

After each CL, there is a PL. The purpose of a PL is to reduce the spatial size of the feature maps produced by the preceding layer. The operation of a PL is a down-sampling operation which exploits max-sampling or average-sampling.

2.3 Deep Learning applications in bearing fault diagnosis

For a long time, conventional ML has been widely applied in bearing fault diagnosis. However, the performance of this approach highly requires hand-craft feature extraction, expert knowledge, and human labor. That makes the task of creating an automatic fault diagnosis model impossible. DL with the ability of automatically learning multiple levels of features and data abstraction has been considered to possess the potential to solve the drawbacks of conventional ML. Until now, AE has been the most popular algorithm of DL applied for bearing fault diagnosis. The reason is that AE has a simple structure as a typical NN. Moreover, even the deep models of AE such as SAE and SDA can be simply constructed and well trained by the greedy layer-wise training method. RBM is also an unsupervised learning algorithm like AE in the family DL. RBM has become popular since the invention of CD algorithm. RBM based deep models can also be trained by the greedy layer-wise training. AE and RBM based deep models were used as unsupervised learning, using unlabeled data to learn high-level salient features from fault data. CNN, a supervised learning algorithm is the third DL architecture widely applied in bearing fault diagnosis. Using CNN models, there was a novel approach when bearing vibration signals are presented in 2-D form beside the approach working directly with 1-D signals. CNN models are integrated both feature extractors and feature classifiers in their structures. Their training process is supervised learning which requires labeled data. In the cases of AE and RBM, their deep models are used as feature extractors. The training processes of AE and RBM based deep models are unsupervised learning which has two phases: pre-training phase and fine-turning phase. RNN is also an important deep structure in the DL family. RNN has memory and can process arbitrary sequences of input patterns, in a sense, RNN is the deepest model [63]. RNN models have been applied in many areas. However, until now, there are not much publications of RNN on bearing fault diagnosis. That is the reason we did not mention about RNN in the previous review section. Signals from bearings are time series data in nature, so RNN is also a promising tool for bearing fault diagnosis.

There are two approaches to process raw signals before feeding into the DL models. The first approach uses deep models with raw signals in time domain directly, while the

second one employs signal processing technique to extract low-level features or transforming raw signals into other forms before feeding into a deep model. The low-level features can be extracted from the time domain, frequency domain, and time-frequency domain. In some publications using models constructed from CNN, raw signals were transformed into 2-D forms, such as square matrix or image form. Obviously, this approach still requires signal processing techniques, expert knowledge, and human labor. In other words, this approach cannot create an automatic process for diagnosing bearing fault, although DL architectures were exploited and the performance is better than traditional ML based methods. The first approach which deals with raw signals successfully created the end-to-end systems which can automatically diagnose bearing fault.

Compared with conventional ML architectures, DL models can learn salient features from data more easily, but it is not easy to design an appropriate deep model for any specific diagnosis task. Since each hyper-parameter has a reasonable effect on the model performance. Through the above-reviewed publications, we can see that hyper-parameters were almost chosen by trial and error method. This is a time-consuming work and requires experiences. Designing a deep model is still a challenge in DL research when there is still no standard way to select appropriate hype-parameters.

Chapter 3

Vibration Image and Convolutional Neural Network

3.1 Introduction

In machine health monitoring, since 2015, a lot of researchers have tried to exploit DL models to diagnose bearing faults. F. Jia et al. used SAE to extract features from signals in the frequency domain [64]. In an introductory paper, Z. Chen et al. proposed three deep models, includes DBN, DBM, and SAE for bearing fault diagnosis [65]. They applied two approaches to learning fault features, the first one is to use the deep model with raw signals in time domain directly. In the second approach, low-level features were extracted from the time domain, frequency domain, or time-frequency domain, then deep models were employed to learn higher-level features from those low-level features. L. Eren used a one-dimensional CNN to diagnose bearing faults, using raw signals directly in the time domain [66].

In the family of DL algorithms, DBM and DBN are based on Restricted Boltzmann Machine (RBM), SAE is based on Autoencoder (AE), all of them are unsupervised learning algorithms, while CNN is a supervised learning method. Initially, with three key architectural ideas: local receptive fields, weight sharing, and sub-sampling in the spatial domain, CNN is suitable for processing 2-D data [67]. In machine health monitoring, some researchers have tried to apply one-dimensional CNN models [64, 65, 66]. However, it is much easier to extract information from data in a high dimension [68]. Being

motivated by this fact, in this chapter, a CNN model called (VI-CNN) is proposed for diagnosing bearing faults using 2-D form of vibration signals. First, the vibration signals in the time domain are transformed into 2-D form, called vibration images. After that, a CNN will be used to identify the faults of the bearing through vibration image classification.

3.2 Method development

3.2.1 Vibration image construction

The vibration signals of bearings are 1-D data form. In this section, we explain in detail the method to transform the vibration signals into gray-scale images. We call this type of image vibration image. Figure 3.1 shows the process of vibration image construction. The amplitude of each sample in the vibration signal is normalized into the range $[-1, 1]$. After that, the normalized amplitude of each sample becomes the intensity of the corresponding pixel in the corresponding image. The conversion between the normalized amplitude of the sample and the corresponding pixels can be described by the following equation [69].

$$P[i, j] = A[(i - 1) * M + j] \tag{3.1}$$

where $i = 1 : N; j = 1 : M; P[i, j]$ is the intensity of the corresponding pixel (i, j) in the $M \times N$ vibration image. $A[.]$ is the normalized amplitude of the sample in the vibration signal. The number of pixels in the vibration image equals to the number of samples in the vibration signal.

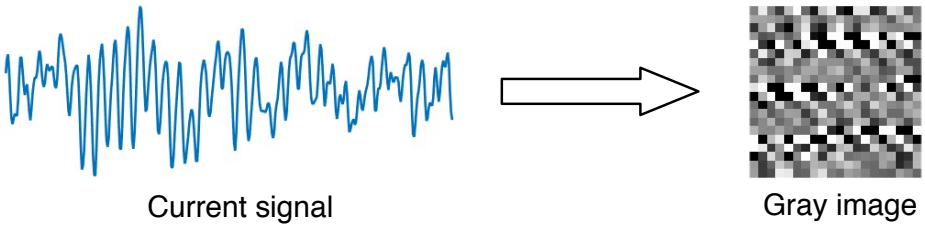

FIGURE 3.1: Vibration Image Construction

3.2.2 Image Classification with CNN

The proposed fault diagnosis method in this chapter has the diagram shown in Figure 3.2.

FIGURE 3.2: Proposed bearing fault diagnosis method

Figure 3.3 shows the architecture of the VI-CNN for bearing fault diagnosis. Vibration images are given to CNN, firstly, successive CLs and SLs extract features from vibration images. Multiple layers help acquire good representations of the input image and improve the performance of the network. Features extracted from the previous layers are classified by a full connection layer with the softmax function.

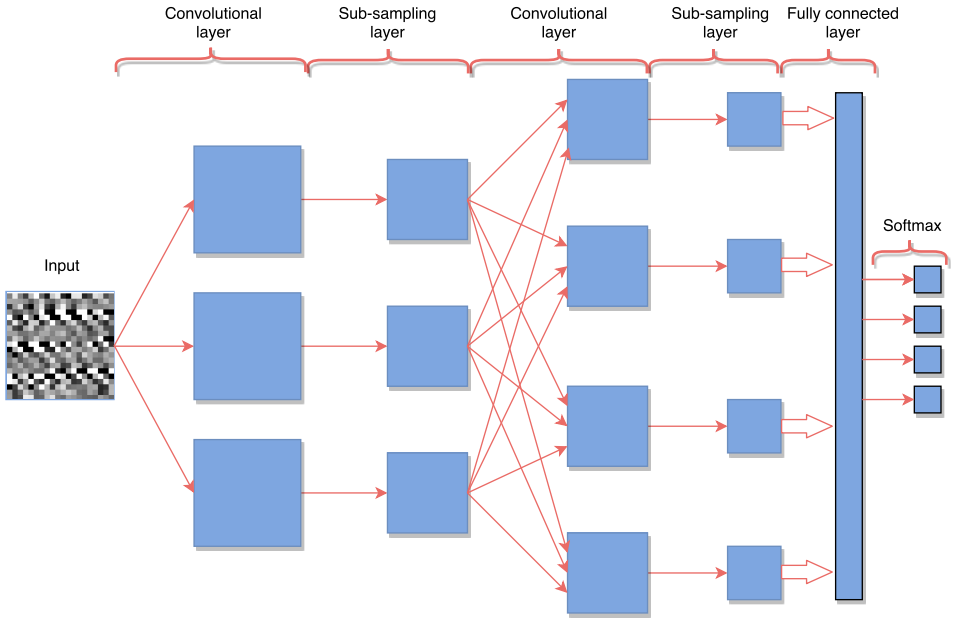

FIGURE 3.3: Convolutional Neural Network

Loss function of the proposed model is the cross-entropy between the estimated softmax output probability distribution and the target class probability distribution. Let denote the target distribution as $p(x)$ and the estimated distribution as $q(x)$. The cross

entropy between $p(x)$ and $q(x)$ is defined as follows:

$$H(p,q) = -\sum_{x} p(x) log q(x) \tag{3.2}$$

Once the loss function is derived, numerous optimization algorithms can be applied to train the network, such as stochastic gradient descent, gradient descent with momentum and variable learning rate, Adam Stochastic optimization [70]. In this paper, stochastic gradient descent is applied to train our CNN because it is easy to implement, memory-saving and computationally effective.

3.3 Experimental study

3.3.1 Data pre-processing

Four health status of the bearing are considered, includes normal status, fault in inner race, fault in ball, and fault in outer race as shown in Figure 3.4. The bearing data for each bearing condition is supplied a single Matlab file. In order to have enough samples for training and testing classifiers, vibration signals are split into segments with the same length. The process of dividing signals into segments is shown in Figure 3.5. Each segment is normalized to have zero mean in the range $[-1, 1]$. One segment of the vibration signal will be transformed into one corresponding vibration image with size $M = 20, N = 20$ by Equation 3.1. Correspondingly, we obtain four types of vibration images as shown in Figure 3.6. The test-bed is operated under four load conditions, includes 0 hp, 1 hp, 2 hp and 3 hp. Four corresponding datasets $A, B, C, and\ D$ are built as described in Table 3.1. Each dataset corresponds to one operating condition of the test-beb and contains four types of vibration images, with the number of images in each type is 606.

TABLE 3.1: Vibration Image Dataset

Name	Number of Images	Load Condition
A	2424	0 hp
B	2424	1 hp
C	2424	2 hp
D	2424	3 hp

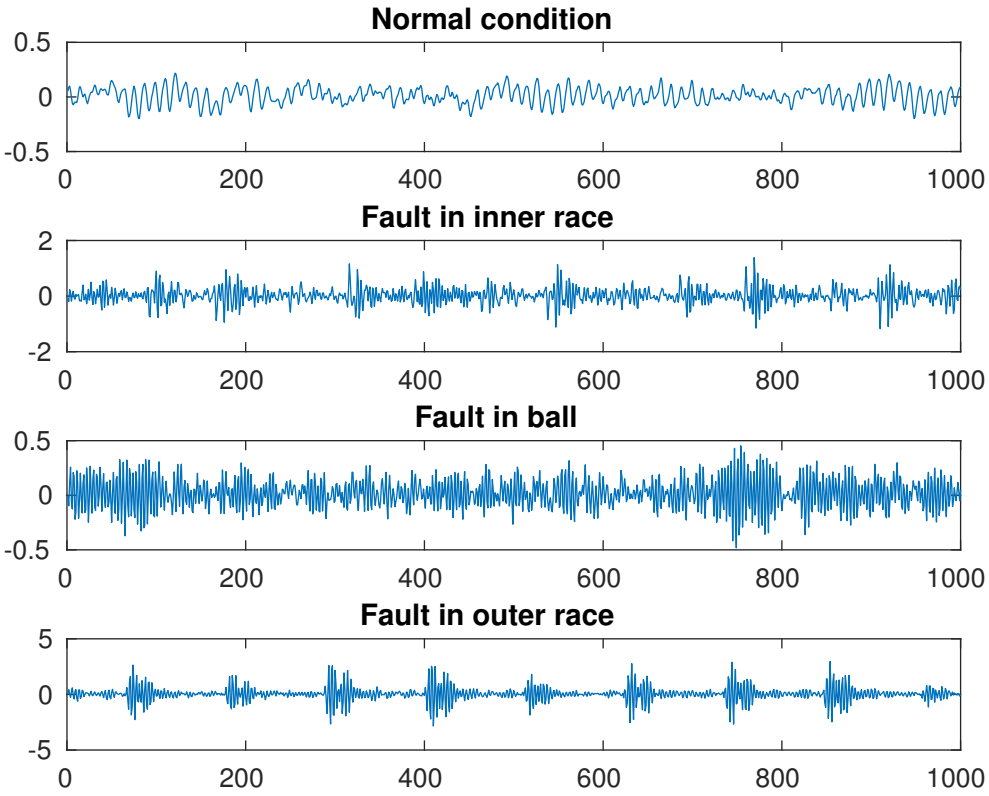

FIGURE 3.4: Vibration signals

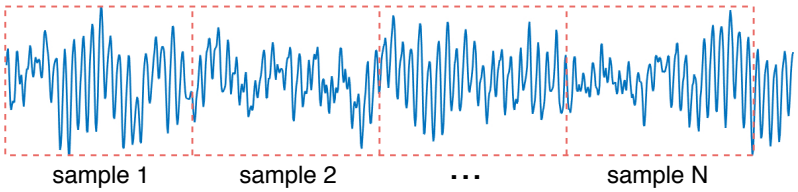

FIGURE 3.5: Signal Segmentation

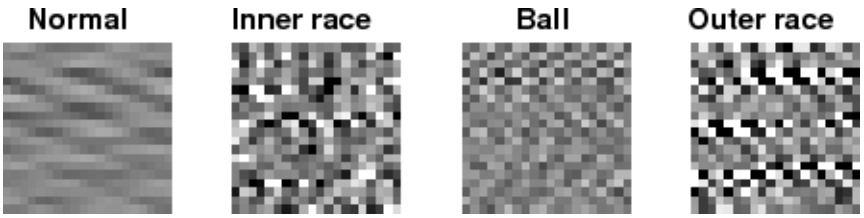

FIGURE 3.6: Vibration images

3.3.2 Hyper-parameters selection

Each hyper-parameter has a big effect on not only the classification accuracy but also the training time of CNN model. Unfortunately, until now, there is no standard method

for selecting appropriate hyper-parameters. In this section, we explain our simple way to select hyper-parameters based on experiments.

The first parameter should be selected is the size of signal segments. The segment size must be long enough to capture the localized features of vibration signals. However, too long segments will make the classifier model more complex. Moreover, to easily transform signal segments into vibration images, the length of the segments must be a square number. We choose the segment size is 400, i.e., each segment contains 400 samples. Correspondingly, each vibration image has a size of 20×20 and contains 400 pixels.

A deeper structure does not ensure a better classification performance, and a too shallow structure limits the capacity of the model in learning complex features of data. Because the size of vibration images is small (20×20), we only use two CLs and two SLs. Kernel size is selected as follows. The next CL has a kernel size smaller than that of the previous CL because after each CL and SL, the size of output data is reduced. The size of kernels in the first CL is 5×5, for the next CL, the kernels have a size of 3×3. In the FC, we fix the number of neurons equal to the number of fault types. The output layer uses the softmax function.

After fixing the number of layers and the kernel size in each layer, we select the number of kernels in each layer by experiments as follows. We start with a small number of kernels. In the second CL, the number of kernels is two times higher than in the first CL. The initial CNN model will be trained by 2024 vibration images and tested by 400 vibration images from the dataset A (load 0 hp). After each time of training - testing, the number of kernels will be increased by 5 and retrained with the same dataset. This process is continuously conducted until the satisfied performance is reached. Table 3.2 shows the classification accuracy of CNN with different kernel sizes.

TABLE 3.2: Classification Accuracy with Different Kernel Sizes

CL1 kernel size	CL2 kernel size	Accuracy (%)
10	20	96.75
15	30	96.75
20	40	99
25	50	99.75
30	60	100

Based on the result of this experiment, we can see that the CNN model with 30 kernels in the first CL and 60 kernels in the second CL achieves 100% classification accuracy. As a result, we select the configuration of CNN model as shown in Table 3.3.

TABLE 3.3: Structure of CNN Model

Layer	Kernel Size	Kernel Number	Padding Type	Output Size
CL	5×5	30	SAME	20×20
SL	2×2	30	no	10×10
CL	3×3	60	SAME	10×10
SL	2×2	60	no	5×5

3.3.3 Evaluate under different load conditions

In Table 3.2, we can see that when being trained and tested with vibration images from the same load conditions (dataset A), the CNN classifiers can achieve 100% classification accuracy. However, in the real application or in the industry, machines and their bearings have to work under various types of conditions. When the working condition changes, the measured vibration signals also change. That makes the fault diagnosis more difficult, and the classifier usually must be retrained before being applied in a different working condition. In this section, we conduct experiments to evaluate the performance of the proposed diagnosis method under different load conditions without retraining the diagnosis system.

As mentioned in Section 3.3.1, we build four datasets of vibration images corresponding to four working conditions of the test-bed: 0 hp (dataset A), 1 hp (dataset B), 2 hp (dataset C), and 3 hp (dataset D). The selected CNN model as in Section 3.3.2 is trained by 2424 images from one dataset and then is evaluated by 1200 images from three other datasets. We compare our proposed method with the one-dimensional CNN proposed in [66] and the SAE model proposed in [64]. Figure 3.7 shows the comparison results.

In case 1, we use 2424 vibration images from the dataset A to train three models, includes 1-D CNN [66], SAE [64] and our proposed VI-CNN model. After that, 400 vibration image samples from each data set B, C, and D (1200 in total) are used to test the trained models. Case 2, 3, and 4 are conducted in the same way with case 1, but the datasets are alternated.

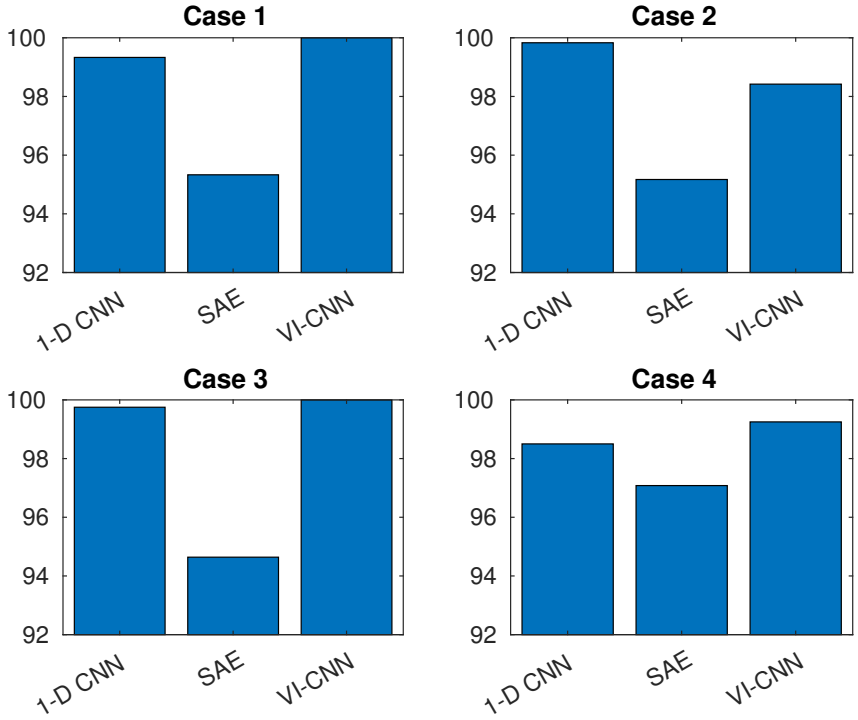

FIGURE 3.7: Performance under different loads

In all four cases, the SAE model has the worst performance with the lowest classification accuracy. In case 2, the 1-D CNN model achieves the best performance with the classification accuracy is 99.83%. In cases 1, 3, and 4, the VI-CNN model achieves the best performance with the highest classification accuracy. Especially, in cases 1 and 3, VI-CNN model reaches 100% classification accuracy. Through the comparisons, we can see that without retraining the whole model, our proposed method still works well when the working condition changes.

3.3.4 Evaluate under different noise conditions

Besides the change of working conditions, the effect of noise is also a big problem which decreases the performance of fault diagnosis. In real industrial environment, the sensory signals are contaminated by noise. In this section, we consider the robustness and accuracy of the proposed scheme under low signal-to-noise ratio (SNR) conditions, i.e., we try to detect the faults of bearings with noisy signals. The additive Gaussian white noise (AGWN) with various standard variances are added to the original vibration signals to mimic the low SNR. The SNR is defined as follows:

$$SNR = 10log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \tag{3.3}$$

where P_{signal} and P_{noise} are, respectively, the power of the signal and the noise in that signal. Figure 5.9 shows the noisy signal made by adding an original signal with Gaussian white noise.

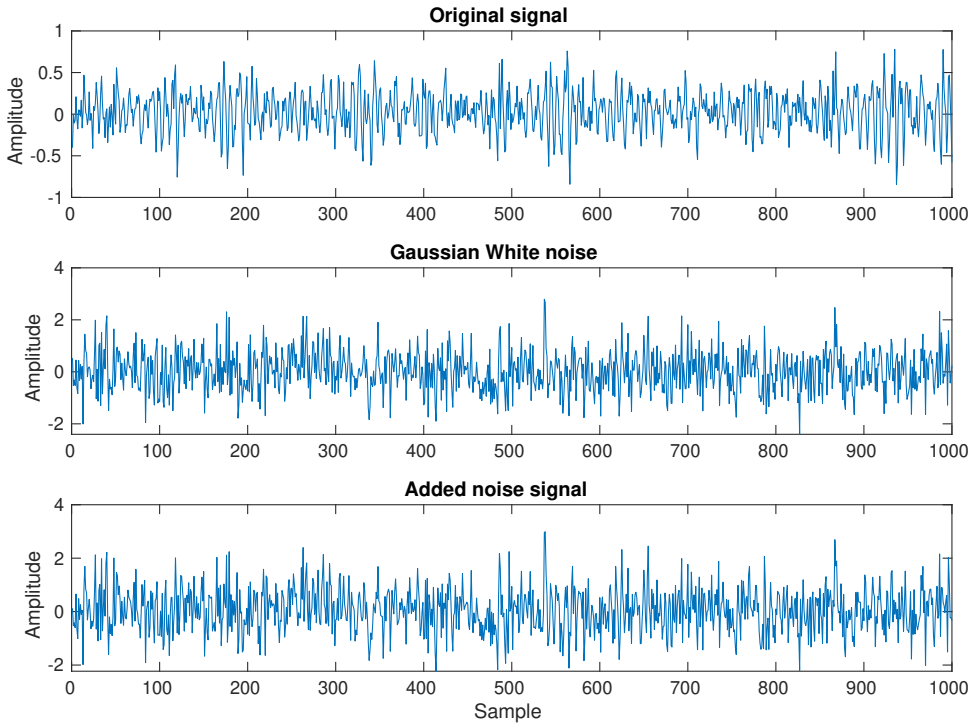

FIGURE 3.8: Noisy Signal with $SNR = -10dB$

The experiment is conducted as follows. The original vibration signal is added Gaussian white noise to form the noisy signals. Then these noisy vibration signals are preprocessed, split into segments, and transformed into vibration images by the same way as mentioned in section 3.3.1. The classifier model will work with vibration image datasets built from noisy signal.

Figure 3.9 shows the fault classification accuracy of the proposed scheme with the SNR value in decibel (dB) varying from -10 dB to 0 dB. Obviously, the bigger noise power is, the more difficult to diagnose faults. In these experiments with noisy signals, the 1-D CNN shows the worst performance with very low accuracy. Under the noise condition $SNR = -4dB$, three diagnosis models still achieve 100% accuracy. However, when the

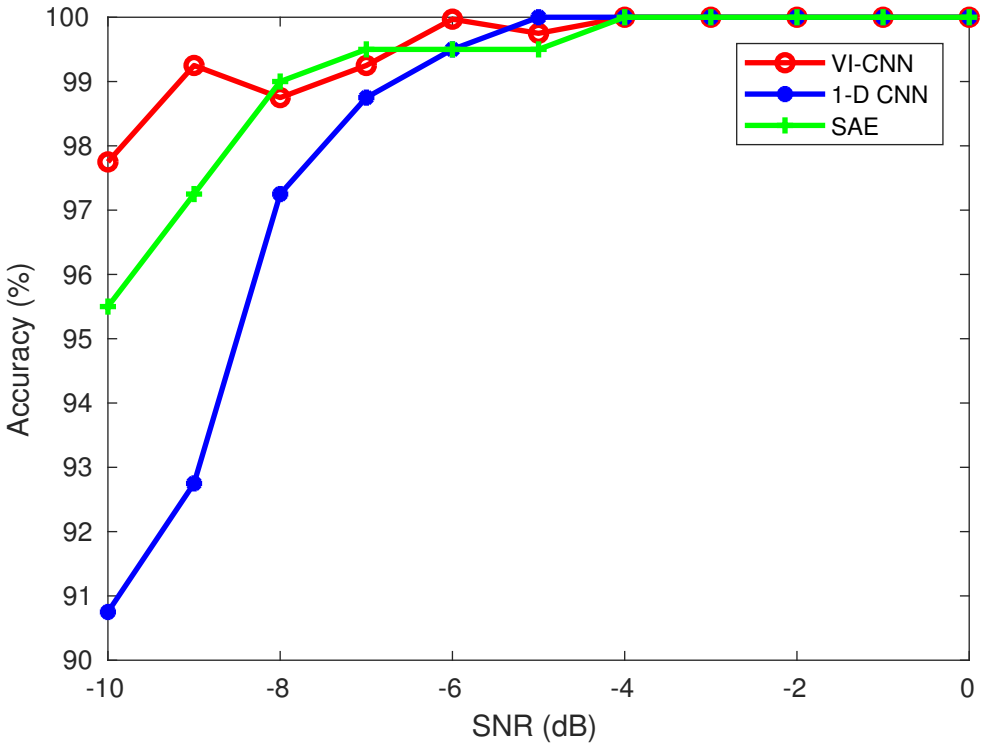

FIGURE 3.9: Performance under different noise conditions

noise condition becomes $SNR = -10dB$, the performances degrade significantly. With SAE model, the accuracy is 95.5% and in the case of 1-D CNN is 90.75%. However, our proposed CNN model using vibration image data still achieves satisfied performance with the accuracy of 97.74%. The comparison results in this experiment show that the proposed VI-CNN model has good robustness and can work well with noisy vibration signals.

3.4 Conclusion

In this chapter, a new approach based on CNN is proposed for diagnosing faults of rolling element bearings. By transforming 1-D vibration signals into 2-D images and exploiting the effectiveness of CNN in image classification, the proposed method can achieve 100% accuracy in CWRU bearing data set.

Compared to traditional machine learning based fault diagnosis, the main advantage of the proposed method is that it does not require the feature extraction step, but still achieves high classification accuracy. Furthermore, when the working load condition

is changed, without retraining the classifier, our proposed method still achieves satisfied performance with high accuracy. Moreover, without any denoising process, the proposed method has the robustness and the capacity of tolerating noisy environment.

A simple procedure based on experiments was used to select hyper-parameters of the CNN model. However, selecting appropriate hyper-parameters to design DL algorithms for fault diagnosis is still a challenge.

Chapter 4

Deep neural network and information fusion

4.1 Introduction

Currently, vibration signals are the most popular method for machine health monitoring because of their ability to convey intrinsic information on the health conditions of mechanical systems [71]. Vibration-signal measurement requires external vibration sensors such as accelerometers to be mounted around the bearing housing. Normally, the cost of vibration signal sensors is high, and the installation of vibration sensors requires direct access to the machine [72]. With these disadvantages, vibration-signal-based bearing fault monitoring is restricted to systems that are easily accessible. In cases where process monitoring is necessary for nearly in-accessible locations or remote locations, such as for centrifugal pumps, cryogenic pumps, etc., vibration-signal-based fault diagnosis cannot be applied [27]. In industry, many induction motor-driven equipment already have current monitoring for control purpose by frequency inverters or protection purposes by current transformers [73]. Hence, the stator current of the motor is often readily for fault diagnosis purpose, not require the installation of other types of sensors. Even in the case of there are no existing frequency inverters or current transformers, the stator current of motors can be easily measured by current transducers, which are accurate, noninvasive and cost-effective [74]. Therefore, motor current signal analysis has been considered as a promising condition monitoring technique because of its low-cost requirement and noninvasive nature.

Based on the position of the installation, we can distinguish two types of bearings: the bearing which is installed inside the motor (internal bearing) and the bearing which is installed outside the motor (external bearing). Fault diagnosis for internal bearings has been extensively studied by researchers with both approaches: vibration-signal-based [75, 76] and motor current-signal-based approach [77, 78]. In the case of external bearing, the vibration signal-based approach achieves better accuracy compared to the current-signal-based approach [13]. The defects of external bearings affect the motor current in an indirect way because its signatures have to be transmitted along the drive train through torque vibrations [13]. Moreover, the effects of external bearings are damped and overlapped with disturbances from the powered process, leading to noisy and difficult to detect signals [79]. In [13], C. Lessmeier et al. used motor current signal for external bearing diagnosis. From the raw signal, features are extracted from time domain, frequency domain, and time-frequency domain. Fast Fourier Transform (FFT) and power spectral density (PSD) are used to extracted features in frequency domain. In the time-frequency domain, the wavelet packet decomposition is used to extracted features. Totally, 23 features are extracted from each raw signal. Then a feature selection technique which bases on the maximum separation distance between different health states is employed to select 9 out of 23 features. Finally, the selected feature set is classified by some machine learning algorithms. The classification accuracy is about 93.3%, which is an unsatisfactory performance. In [80], G. Karatzinis et al. proposed a method using fuzzy cognitive networks (FCN) with functional weights. Motor currents from two phases are extracted to obtain 8 features, all in time domain. The classification accuracy of the FCN is about 91.42%.

From the literature review, we can see that traditional signal processing and feature extraction techniques are unsuccessful when applied in external bearing fault diagnosis with motor-current-signal. In recent years, deep learning (DL) has been widely applied to expedite the task of traditional intelligent fault diagnosis. DL algorithms are machine learning ones that employ deep architectures with many layers of data processing units [43]. With the ability of automatically learning multiple levels of data abstraction, when applied to extract features from the raw signals in fault diagnosis

applications, DL achieves high accuracy and outperforms other traditional signal processing and feature extraction techniques [68, 81, 82]. Motivated by that observation, in this chapter, I propose an advanced technique of extracting features from current signal using DL.

Currently, single-sensor data are extensively used to diagnose bearing faults. However, in machine systems which are installed multiple sensors, multiple-sensor signals have been used for bearing fault diagnosis. In a system with multiple sensors are installed, the signals measured by these sensors are disordered and correlated with multiple sources [83]. Therefore, information fusion (IF) techniques are required to effectively fuse information from multiple sensors. In the case of induction motor, theoretically, the three phases of the motor are symmetrical. However, in real-time running, the three phases are not totally symmetrical because of noise, disturbance, and imperfect installation. Especially, when there faults or defects occur at any part of the system, the asymmetric of the three phases is more serious. Therefore, it is high potential that each phase of motor current can convey different information about the fault status of the system. Motivated by this fact, in this paper, the authors develop a technique to efficiently merge current features. This technique tries to eliminate duplicate information from the current phases and supply richer information about the status of the system.

The goal of this chapter is to develop a bearing fault diagnosis method for rotary machine systems in which multiple-phase motor current signals are readily available, using DL-based feature learning and decision-level IF. The proposed method uses directly raw signals from multiple phases of motor current as input, the signals from each phase of the current are extracted features and classified separately by a corresponding DL model. To enhance the classification accuracy, a novel decision-level IF technique is introduced to fuse information from all the used DL models. The problem of decision-level IF is transformed into a simple pattern classification task, which is effectively solved by familiar supervised learning algorithms such as multilayer perceptron (MLP), support vector machine (SVM), and k-nearest neighbor (kNN). The proposed bearing fault diagnosis method is verified through experiments carried out with a public bearing data set supplied by the Kat-Data Center website of the Chair of Design and

Drive Technology, Paderborn University, Germany [13].

4.2 Method development

4.2.1 Proposed decision-level information fusion method

IF is defined as a process of fusing information or data from different inputs to derive information that is better that would be derived from each of the sources independently [84]. This section explains the proposed decision-level IF algorithm. For the sake of simply explain, we consider the problem of identifying the fault of bearing in a rotary machine. Assuming that there are m types of faults, labeled as $F_i,\ i = 1 : m$. Assuming that we have n feature sets extracted from n signal sources measured simultaneously by n sensors on the machine. Each feature set consists of t samples as follows:

(1) Feature set $X^1 = [x_1^1,\ x_2^1,\ ...,\ x_t^1]$

(2) Feature set $X^2 = [x_1^2,\ x_2^2,\ ...,\ x_t^2]$

...

(n) Feature set $X^n = [x_1^n,\ x_2^n,\ ...,\ x_t^n]$

We need to classify m types of faults that may occur in that machine. At first, n feature sets are classified by n different Softmax classifiers (SCs). The probability of class $F_i,\ i = 1 : m$ based on the observation of $SC^k, k = 1 : n$ on the sample $X_j^k, j = 1 : t$ of the feature set X^k is:

$$P_i^{k,j} = \textbf{probability}(F_i | X_j^k, SC^k) \tag{4.1}$$

The output of SC^k with all n samples in the feature set X^k is arranged in the matrix P^k of the size $m \times t$:

$$P^k = \begin{bmatrix} p_1^{k,1} & p_1^{k,2} & \dots & p_1^{k,t} \\ p_2^{k,1} & p_2^{k,2} & \dots & p_2^{k,t} \\ \vdots & \vdots & \ddots & \vdots \\ p_m^{k,1} & p_m^{k,2} & \dots & p_m^{k,t} \end{bmatrix} \tag{4.2}$$

The outputs of all classifiers are merged to form a $(m \times n) \times t$ matrix P:

$$P = \begin{bmatrix} P^1 \\ P^2 \\ \vdots \\ P^n \end{bmatrix} \tag{4.3}$$

The P matrix is considered as a new feature set that is labeled as the same way with all feature sets X^k. Now we consider the task of fusing n classifiers SC^k as a task of classifying the new feature set P. This classification task is then solved by supervised learning algorithms such as MLP, SVM, and kNN. Finally, the output of the second classification task is the final conclusion about the type of bearing fault in the machine or the result of fusing n data sources.

4.2.2 Proposed bearing fault diagnosis method

In this chapter, the object is the external bearing which is distinguished from the internal bearing (inside the motor). The defects of external bearings affect the motor current in an indirect way because its signatures have to be transmitted along the drive train through torque vibrations. Moreover, the effects of external bearings are damped and overlapped with disturbances from the powered process, leading to noisy and difficult to detect signals. Therefore, to use the stator current as the input for the fault detector, a more effective feature extractor must be developed.

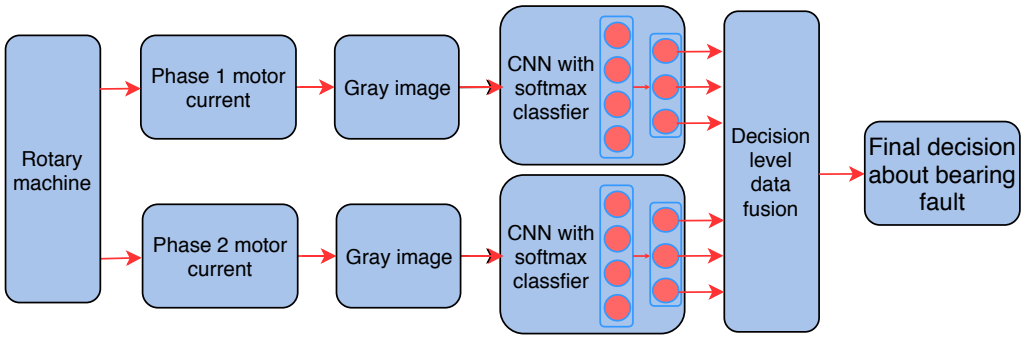

FIGURE 4.1: Proposed bearing fault diagnosis method

The schematic of the proposed bearing fault diagnosis method is shown in Figure 4.1.

First, signals from multiple phases (at least two phases) of the motor current are measured simultaneously. These measured signals from each motor current phase are saved as a separate data set, each data set will be processed independently in the next step. For a given raw current signal of a phase, a sliding window is used to split that signal into equal samples as shown in Figure 4.2. The length of the signal segments is selected as a square number for the sake of ease for successive steps.

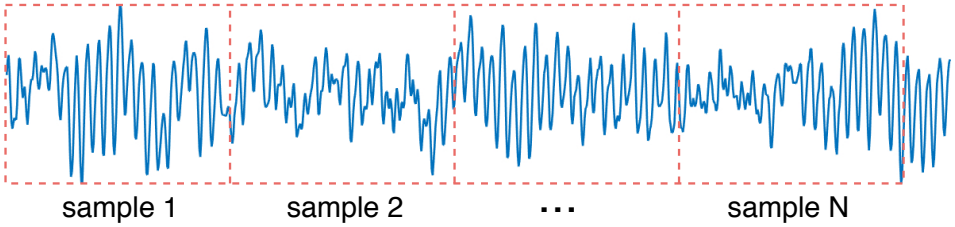

FIGURE 4.2: Signal segmentation

The original measured signals are in one-dimensional $(1-D)$ form. To be easily processed with a $2-D$ CNN, these signals are transformed into $2-D$ form by simply rearranging the array of signal amplitudes into a square matrix form. Assuming that each signal sample is represented as an array of amplitudes $1 \times n^2$, the result of the transformation is an $n \times n$ matrix as in Equation 4.4:

$$\begin{bmatrix} a_1 & \dots & a_{n^2} \end{bmatrix} \rightarrow \begin{bmatrix} a_1 & a_2 & \dots & a_n \\ a_{n+1} & a_{n+2} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{(n-1)n+1} & \dots & \dots & a_{n^2} \end{bmatrix} \tag{4.4}$$

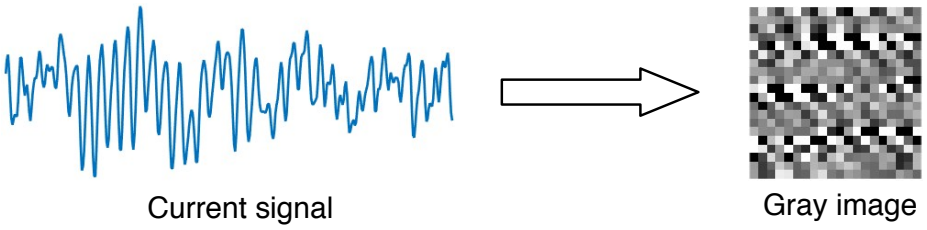

FIGURE 4.3: Gray image representation of current signal

In our proposed method, each segment of the current signal with the length of n^2 is rearranged in a square matrix $n \times n$. The transformation transforms the current signal

in 1-D form into 2-D form. This 2-D representation of the current signals can be considered as a gray image representation of the respective current signal. As a result, the task of classifying current signals becomes the task of classifying images.

The outputs of each CNN model are the probabilities of each type of occurring fault based on the observation of that CNN on the corresponding data source. Finally, the proposed IF uses the results of all CNN models to make the final conclusion about the bearing fault.

4.3 Experimental study

4.3.1 Signal pre-processing

TABLE 4.1: Three operating conditions

Working condition	S (rpm)	M (Nm)	F (N)
A	1500	0.1	1000
B	900	0.7	1000
C	1500	0.7	400

In this work, three working conditions are considered as shown in Table 4.1. As shown in Table 4.2, for each bearing condition, five different bearing codes were established for data measurement. For each bearing code, 20 measurements were performed. Each measurement was saved as a MATLAB file that contained the vibration signals (VS), the phase 1 current signals (CS1), and the phase 2 current signals (CS2).

TABLE 4.2: Bearing codes used for experiments

Class	Label	Used bearing code
NF	1	K001, K002, K003, K004, K005
OF	2	KA04, KA15, KA16, KA22, KA30
IF	3	KI04, KI14, KI16, KI18, KI21

The data files are split into equal-length segments to be used as input samples for designing the fault diagnosis models. The length of the signal segment (L) is selected as follows. Firstly, we can see that the two values of rotary speed are 900 rpm and 1500 rpm, and all signals are sampled at the same sampling frequency (fs = 64 kHz). For lower rotary speed, we will have less samples in each rotary circle. To select the length of the signal segment that is suitable for signals measured at both speed values,

the lower rotary speed value (S = 900 rpm) is considered. The number of samples per rotary circle (spc) is calculated as follows:

$$spc = \frac{fs \times 1000 \times 60}{S} = \frac{64 \times 1000 \times 60}{900} \approx 4266 \tag{4.5}$$

The number of rotary circles corresponding to each signal segment is selected at value 1.5. Then we have the length of the signal segment:

$$L = 1.5 \times spc = 1.5 \times 4266 = 6399 \tag{4.6}$$

To easily rearrange each sample into a $2-D$ square matrix form, the value of L should be a square number, we select the nearest square number is 6400 ($\sqrt{6400} = 80$).

TABLE 4.3: Data set corresponding to the operating condition A

Signal type	Training samples	Testing samples
Vibration	10500	750
Phase 1 current	10500	750
Phase 2 current	10500	750

Each signal sample was saved as a $1-D$ array with a size of 1×6400. The process described in Section 4.2.2 transformed the original signal samples into a $2-D$ form as a matrix with a size of 80×80. In total, we obtained 11250 samples for each data set corresponding to each type of signal. Then each data set was split into a training set and a testing set, the details of the data set A corresponding to the operating condition ($S = 1500\ rpm, M = 0.1\ Nm, F = 1000\ N$) are shown in Table 4.3.

4.3.2 Signal analysis

With the parameters of the test bearing of type 6203 are ball diameter ($d = 6.75\ mm$), pitch diameter ($D = 28.55\ mm$), number of balls ($n = 8$), bearing contact angle ($\phi = 15°$) and rotary speed ($S = 1500\ rmp$), we can compute the critical frequencies as follows [85].

The rotary frequency of the shaft:

$$Fr = \frac{S}{60} = 1500/60 = 25\ Hz \tag{4.7}$$

The ballpass frequency, outer race:

$$Fo = \frac{nF_r}{2}\left(1 - \frac{d}{D}\cos\phi\right) = \\ = \frac{8*15}{2} * \left(1 - \frac{6.75}{28.55} * \cos(15^\circ)\right) = 46.3\ Hz \tag{4.8}$$

The ballpass frequency, inner race:

$$Fi = \frac{nF_r}{2}\left(1 + \frac{d}{D}\cos\phi\right) = \\ = \frac{8*15}{2} * \left(1 + \frac{6.75}{28.55} * \cos(15^\circ)\right) = 73.7\ Hz \tag{4.9}$$

The envelope spectrum analysis is employed to analyze three types of current signals: the normal bearing signal, the outer race fault signal, and the inner race fault signal. The power spectrums of the envelope signals are shown in Figure 4.4. In can be observed that in all three cases, the power spectrum of the electrical supply frequency ($Fe = 100\ Hz$) is shown clearly. However, the envelope spectrums in all cases do not show any peaks at Fo and Fi. The weak indicators of the bearing faults in the current signals are mainly due to the fact that the damage signatures have to be transmitted indirectly through torque variations along the drive train. Therefore, they are damped and masked with disturbances and noises. Consequently, it is difficult to extract features from the current signal for external bearing fault diagnosis. As a result, to exploit current signals for diagnosing the bearing fault, effective feature learning methods need to be developed.

4.3.3 Experiment result

First of all, the configuration of the CNN is determined. To easily control the sizes of feature maps, we will keep intact the sizes of the feature maps in all CLs and decreased

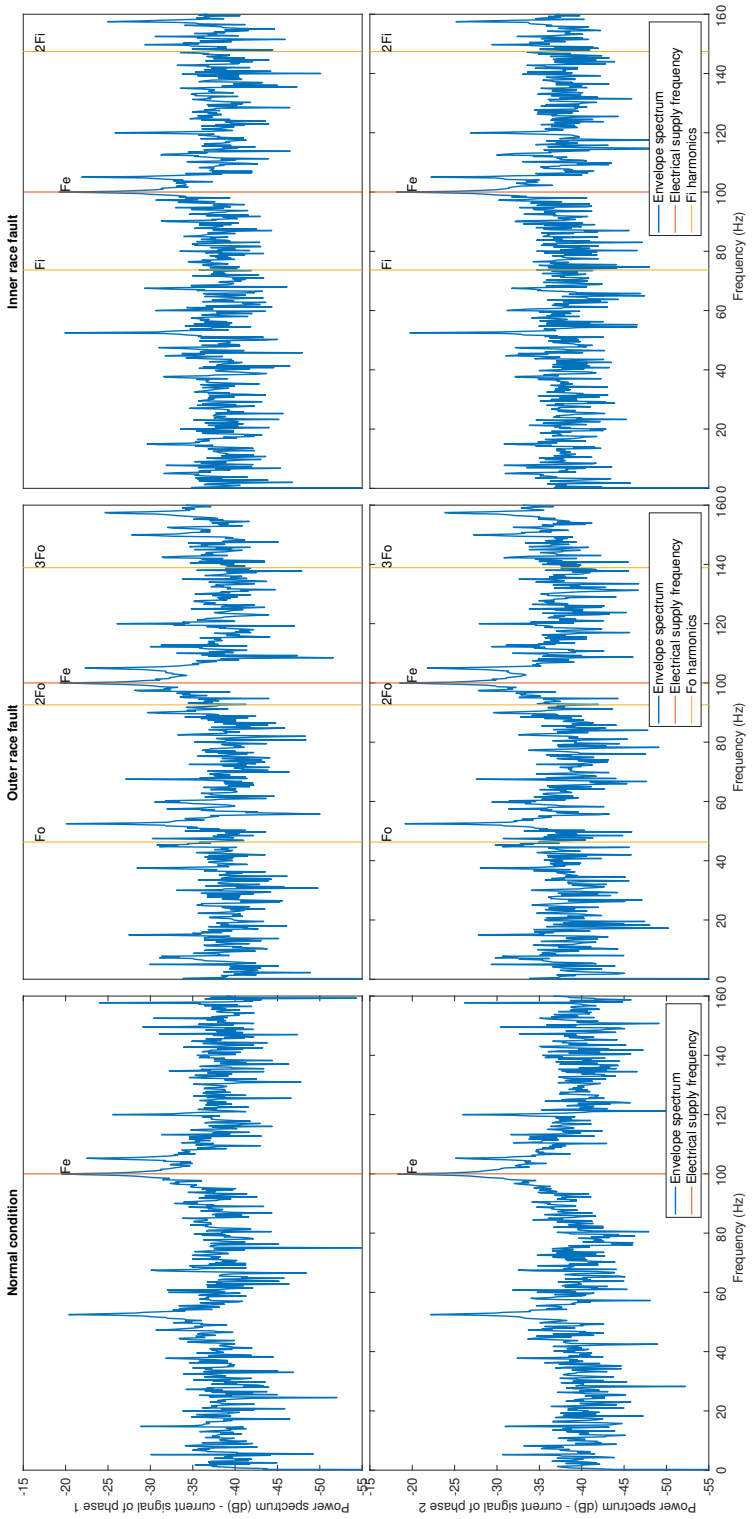

FIGURE 4.4: Power spectrum analysis of three types of current signals (phase 1 and 2)

2 times that sizes in all PLs. Therefore, we apply zero-padding technique and stride step 1×1 in all CLs and kernel size 2×2 with stride step 2×2 in all PLs. Hence, after going through each pair of CL and its corresponding PL, the size of the feature map will be reduced by 2. As determined in the previous section, the input data of CNN (first feature map) has size 80×80. We can see that $80 = 2^4 \times 5$. Thus we will design the CNN with 4 CLs and 4 corresponding PLs. Next, we select the kernel size in each CL. The kernel size 3×3 is a popular choice in designing CNN, this kernel size will be applied for the last CL. We want each CL extracts more local features from its preceding CL, so that from the first CL to the last CL, the kernel size is decreased gradually. Therefore, we have kernel sizes 9×9, 7×7, 5×5, and 3×3 in CL1, CL2, CL3, and CL4 respectively.

The CL1 is fixed with 5 kernels. We increase the number of kernels by 5 after each layer. However, we regulate the number of kernels in the CL4 30 kernels to obtain better training performance. The number of kernels in each PL is equal to that of its corresponding CL.

We use 1 FL to flatten the last feature map from 2-D form into 1-D form, therefore, the size of FL is 5*5*30 = 750. The Softmax layer has three outputs corresponding to 3 types of bearing fault need to be classified. The final configuration of the CNN model is shown in Table 4.4.

TABLE 4.4: CNN structure

Layer	Kernel size	Kernel number	Input size	Output size
CL1	9×9	5	80×80	80×80
PL1	2×2	5	80×80	40×40
CL2	7×7	10	40×40	40×40
PL2	2×2	10	40×40	20×20
CL3	5×5	15	20×20	20×20
PL3	2×2	15	20×20	10×10
CL4	3×3	30	10×10	10×10
PL4	2×2	30	10×10	5×5
FL			5×5	750
Softmax			750	3

As described in the Introduction, the main goal of this paper is to propose a bearing fault diagnosis method using CS that is accurate as the methods that use VS. First,

we evaluate the fault detection accuracy of a CNN with input VS as input data. The procedure of this experiment is shown in Figure 4.5.

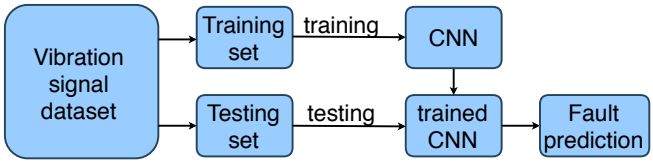

FIGURE 4.5: Vibration signal-based fault diagnosis

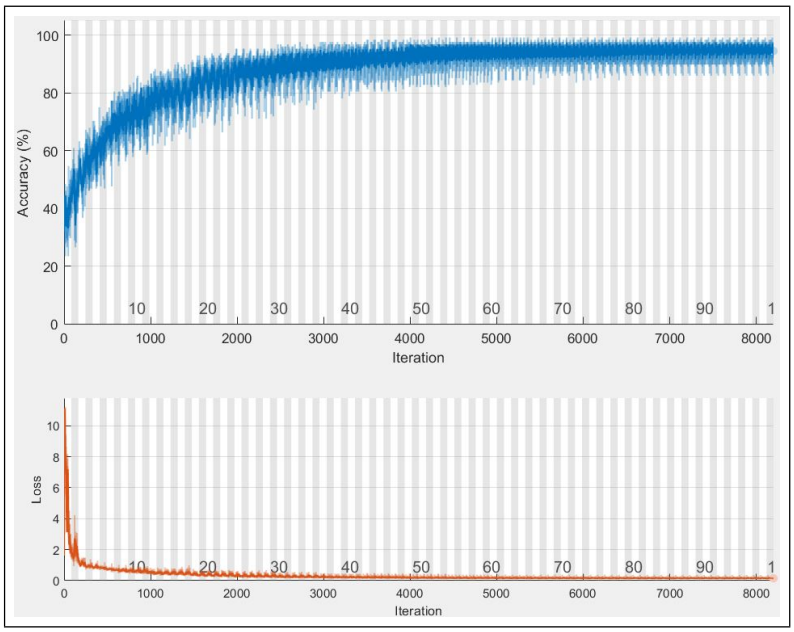

FIGURE 4.6: Training process of CNN with current signals

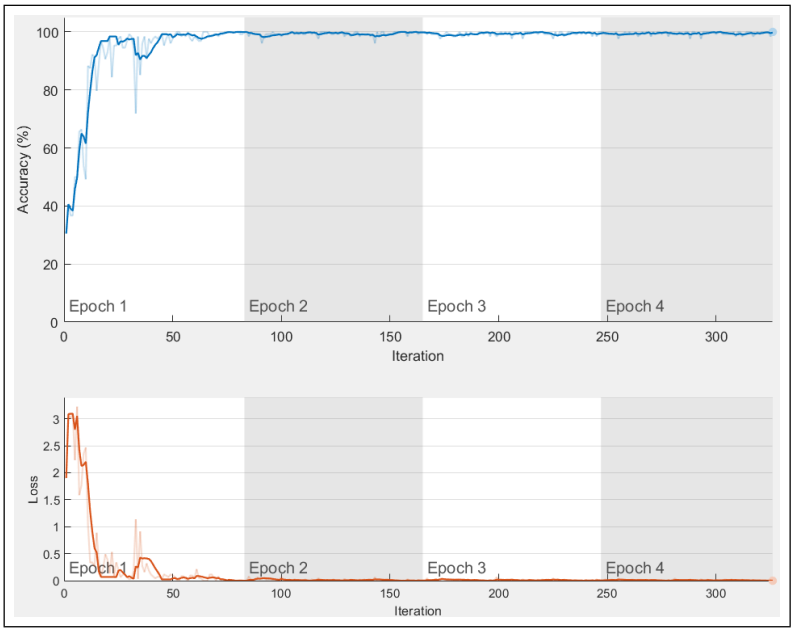

FIGURE 4.7: Training process of CNN with vibration signals

The training processes of the CNNs with current signals and vibration signals are shown in Figure 4.6 and Figure 4.7, respectively. Both networks have the same structure and configuration. In the case of the CNN with vibration signal input, the training process is implemented with the stochastic gradient descent with momentum (SGDM), the learning rate $\alpha = 0.0012$, and the momentum $\gamma = 0.9$. The CNN quickly converges and easily achieves the training accuracy of 100 %. In the case of the current signal, the training process is implemented with the adaptive moment estimation (Adam) optimizer, the learning rate $\alpha = 0.01$, the gradient decay factor $\beta_1 = 0.9$, and the squared gradient decay factor $\beta_2 = 0.99$. Training the CNN with current signal input takes much time to converge and the accuracy is not as high as the training accuracy of the CNN with vibration signal input.

Initially, 10500 signal samples of the training set were used to train the CNN. Then the trained CNN was evaluated by testing with 750 samples of the testing set. The accuracy of this vibration signal-based fault diagnosis experiment was 99.47 % - a high accuracy and a satisfying result that can have real applications in industry.

Then we investigated the CS-based fault diagnosis method. Two different CNN models were used to deal with phase 1 and phase 2 current-signals (CS1 and CS2). Each CNN was trained by 10500 samples of the corresponding training set. After the training phase, two testing sets were used to evaluate the accuracy of the trained CNNs. With the CS1 and CNN1, the classification accuracy was 95.2 %. With the CS2 and CNN2, the classification accuracy was 93.07 %. In both cases, the accuracy results were very low compared to the results of the VS-based method in the previous experiment.

To improve the performance of the CS-based methodology, we then used the proposed IF technique to enhance the fault detection accuracy. Since the proposed IF technique is based on supervised learning, we built separate training and testing sets for this problem. When all samples of the training set 1 (CS1 data set) were fed into the corresponding diagnosis model CNN1, we obtained the following probability matrix:

$$P_{training\ 1} = \begin{bmatrix} p_1^{1,1} & p_1^{1,2} & \dots & p_1^{1,10500} \\ p_2^{1,1} & p_2^{1,2} & \dots & p_2^{1,10500} \\ p_3^{1,1} & p_3^{1,2} & \dots & p_3^{1,10500} \end{bmatrix} \tag{4.10}$$

Similarly, we obtained the probability matrix of training set 2 and CNN2:

$$P_{training\ 2} = \begin{bmatrix} p_1^{2,1} & p_1^{2,2} & \dots & p_1^{2,10500} \\ p_2^{2,1} & p_2^{2,2} & \dots & p_2^{2,10500} \\ p_3^{2,1} & p_3^{2,2} & \dots & p_3^{2,10500} \end{bmatrix} \tag{4.11}$$

With the testing set, we had:

$$P_{testing\ 1} = \begin{bmatrix} p_1^{1,10501} & p_1^{1,10502} & \dots & p_1^{1,11250} \\ p_2^{1,10501} & p_2^{1,10502} & \dots & p_2^{1,11250} \\ p_3^{1,10501} & p_3^{1,10502} & \dots & p_3^{1,11250} \end{bmatrix} \tag{4.12}$$

and:

$$P_{testing\ 2} = \begin{bmatrix} p_1^{2,10501} & p_1^{2,10502} & \dots & p_1^{2,11250} \\ p_2^{2,10501} & p_2^{2,10502} & \dots & p_2^{2,11250} \\ p_3^{2,10501} & p_3^{2,10502} & \dots & p_3^{2,11250} \end{bmatrix} \tag{4.13}$$

Merging $P_{training\ 1}$ with $P_{training\ 2}$, and $P_{testing\ 1}$ with $P_{testing\ 2}$, we obtained the following matrices:

$$P_{training} = \begin{bmatrix} p_1^{1,1} & p_1^{1,2} & \dots & p_1^{1,10500} \\ p_2^{1,1} & p_2^{1,2} & \dots & p_2^{1,10500} \\ p_3^{1,1} & p_3^{1,2} & \dots & p_3^{1,10500} \\ p_1^{2,1} & p_1^{2,2} & \dots & p_1^{2,10500} \\ p_2^{2,1} & p_2^{2,2} & \dots & p_2^{2,10500} \\ p_3^{2,1} & p_3^{2,2} & \dots & p_3^{2,10500} \end{bmatrix} \tag{4.14}$$

$$P_{testing} = \begin{bmatrix} p_1^{1,10501} & p_1^{1,10502} & \cdots & p_1^{1,11250} \\ p_2^{1,10501} & p_2^{1,10502} & \cdots & p_2^{1,11250} \\ p_3^{1,10501} & p_3^{1,10502} & \cdots & p_3^{1,11250} \\ p_1^{2,10501} & p_1^{2,10502} & \cdots & p_1^{2,11250} \\ p_2^{2,10501} & p_2^{2,10502} & \cdots & p_2^{2,11250} \\ p_3^{2,10501} & p_3^{2,10502} & \cdots & p_3^{2,11250} \end{bmatrix} \tag{4.15}$$

This resulted in the training set $P_{training}$ of size 6×10500 and the testing set $P_{testing}$ of size 6×750. The problem of the fusing decisions of the two classifiers CNN1 and CNN2 was considered as a new classification problem as follows:

- Three classes need to be classified: NF, OF, and IF corresponding to labels: 1, 2, and 3.
- Input: $P_{training}$ and $P_{testing}$

This classification problem is simple, with only three classes and the dimension of the feature size is 6. We solved this problem with three familiar supervised learning algorithms are MLP, SVM, and kNN. The results are shown in Table 4.5. By using the proposed IF technique, the diagnosis accuracy increased from 95.2 % (CS1 based experiment) and 93.07 % (CS2 based experiment) to 98.3 %. Compared to the performance of the VS-based method in the first experiment (99.47 %), this result (98.3 %) was quite high and acceptable.

TABLE 4.5: IF based diagnosis results

Algorithm	Configuration	Accuracy (%)
MLP	1 hidden layer with 5 neurons	98.3
SVM	linear kernel function	98.0
kNN	$k = 2$	97.07

To show the effectiveness of the proposed feature learning method, the t-distributed Stochastic Neighbor Embedding (tSNE) [86] which is a data visualization technique is employed. The idea of this algorithm is to embed high-dimensional data points in low dimensions in a way that respects similarities between points. The tSNE technique is applied to three data sets: the raw current signal phase 1, the raw current signal phase

2, and the feature set extracted by the proposed method. The 2-D embedding of three data sets are shown in Figure 4.8. The visualization of the raw current signals phase 1 and phase 2 do not show the clear distinction of three types of labels: NF, OF, and IF. Those data points are almost mixed up and they are unable to be distinguished. The visualization of the features extracted from the raw current signals phase 1 and phase 2 are shown in the third figure. It can be observed that the data points in this visualization are clearly separated. Consequently, using the features extracted by the proposed network will be easy to recognize the labels of the data points.

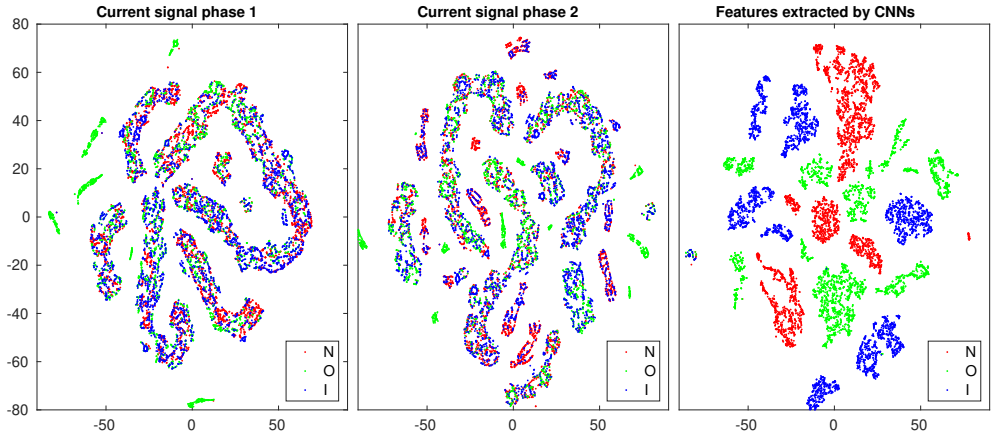

FIGURE 4.8: Feature visualization

To more deeply evaluate this proposed bearing fault diagnosis method, we considered three more operating conditions of the test-bed. Three data sets A, B, and C are obtained from three operating conditions as in Table 4.1. Data set A corresponding to the operating condition No. 1 was used in the previous experiments. We then conducted the same experiments with data sets B and C. The experimental results are shown in Figure 4.9 and Table 4.6.

TABLE 4.6: Results of fault diagnosis in different data sets

Diagnosis method	Accuracy (%)		
	A	B	C
Use CS1 signal	95.2	90.93	87.6
Use CS2 signal	93.07	86.53	93.06
Use CS1 & CS2 signal, MLP based IF	98.3	96.7	96.8
Use CS1 & CS2 signal, SVM based IF	98	96.13	97.86
Use CS1 & CS2 signal, kNN based IF	97.07	96.13	96.8
Use vibration signal	99.47	98.8	98.93

It can be observed that, in the current-signal-based approach, if the single current phase

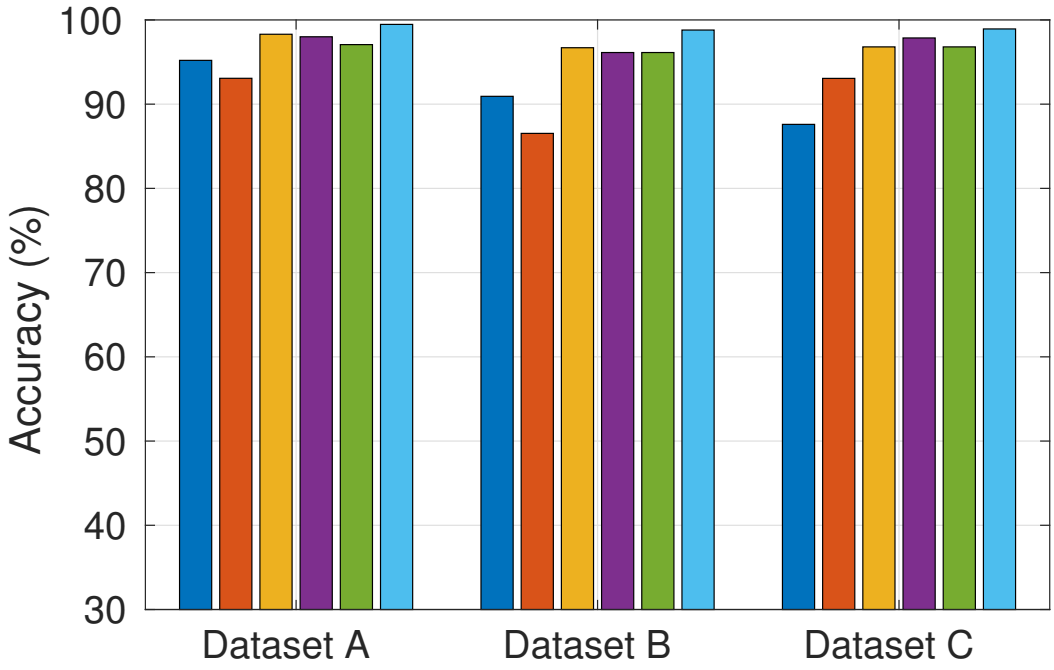

FIGURE 4.9: Fault diagnosis accuracy in three data sets

1 or phase 2 is used separately, the fault detection accuracy is low, the average value is just about 91.07 %. However, if the IF technique is used to merge information from two phases of the current signal, the accuracy increases significantly, up to 98.3%. Compared to the vibration signal-based method whose accuracy is up to 99.47 %, the performance of our proposed method is slightly lower.

Compared to other existing works based on the motor current signal in [13] and [80], our proposed method has much better performance. As mentioned in section 4.1, using the same bearing data source as that one used in this paper, these works employed the traditional feature extraction methods and could not achieve satisfactory performance (the fault detection accuracy are 93.3 % and 91.42 % respectively).

4.4 Conclusion

Since the existing current-signal-based works can not supply satisfactory performance, external bearing fault diagnosis often employed the vibration-signal-based approach. However, the vibration-signal-based approach has two major drawbacks: requiring external expensive vibration sensors and requiring direct access to the machine. This

chapter presents a current-signal-based bearing fault diagnosis using a DL algorithm - CNN and a novel decision-level IF technique. The proposed method is not only cost-effective and noninvasive but also achieves satisfactory performance. Moreover, the use of DL to extracting features from the raw signals does not require hand-craft feature extraction. With these advantages, the proposed bearing fault diagnosis method is promising to be applied in actual applications.

Chapter 5

Wide Deep Neural Network Structure

5.1 Introduction

In traditional signal-based fault diagnosis, extracting fault features from multiple domains of signals is a popular approach. In can be observed that one common point in current DL based fault diagnosis methods is that their DNN models only attempt to extract features from a single domain of vibration signals, such as time domain [87], frequency domain [64], or time-frequency domain [22]. This point may be a drawback since the goal of feature extraction is to extract as much information as possible from the signal. Moreover, it is shown that multiple domain feature models can achieve better performance compared to single domain feature models [88]. Based on this observation, in this paper, a bearing fault diagnosis method based on the multiple domain feature model approach is proposed. First, a new method transforming the vibration signals into high dimensional data is proposed. In this method, the vibration signals are simultaneously converted into time domain images and time-frequency images. This representation type of vibration signal is named here as multiple-domain image-representation (MDIR) data. A CNN based deep model named wide deep neural network (WDNN) is proposed to classify the MDIR data. The proposed DNN has two main characteristics. Firstly, it inherits the deep learning ability of DL. Secondly, it has a wide structure to extract features from multiple domains of signals simultaneously. More details of the MDIR data and the WDNN are explained in sections 5.2.1 and 5.2.2

of this paper. The task of signal-based fault diagnosis is now considered as the task of image classification.

5.2 Method development

5.2.1 Multiple-domain image-representation of vibration signal

Originally, the vibration signal is time-series data, which is a 1-D data form. A new method to represent vibration signals in high dimension is proposed. It is motivated by the three following facts:

1. It may be easier to understand and mine information in high dimensional data [68].
2. CNN and its variants are suitable for the task of recognizing two-dimensional visual patterns [67].
3. By transforming the signal into visual data, the task of fault diagnosis can be converted into the task of image classification.

The vibration signal will be transformed into a time domain image by a simple method proposed by D. Nguyen et al. in [69]. Consider a signal $x(i)$ with the number of samples is n^2, all samples in the signal are rearranged into a square matrix with a size of $n \times n$ as follows:

$$\begin{bmatrix} x_1 & \dots & x_{n^2} \end{bmatrix} \rightarrow \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ x_{n+1} & x_{n+2} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{(n-1)n+1} & \dots & \dots & x_{n^2} \end{bmatrix} \tag{5.1}$$

Then the obtained square matrix is normalized to range $[0.0 - 1.0]$ by the linear normalization:

$$I_{ij} = \frac{X_{ij} - min(X)}{max(X) - min(X)} \tag{5.2}$$

The normalized sample I_{ij}, $i, j \in [0, n]$ is placed at the row i, column j of the matrix. The corresponding time domain image consists of pixels that are normalized samples of the corresponding signal as shown in Figure 5.1.

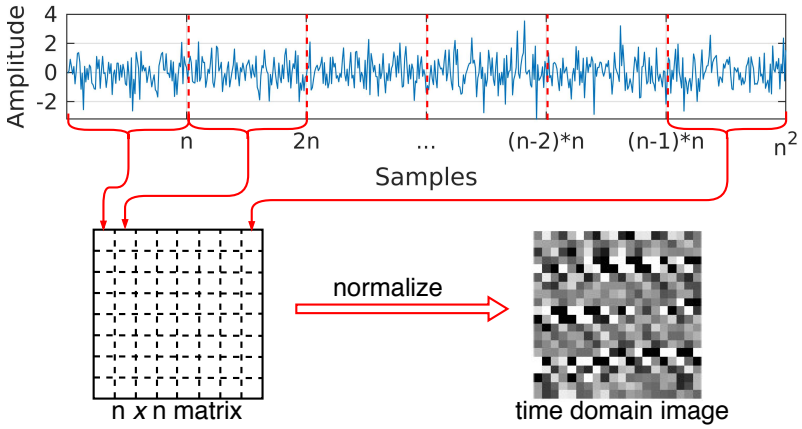

FIGURE 5.1: Time domain image transformation

The time-frequency image representation is based on the continuous wavelet transform (CWT). A mother wavelet is a function $\psi(t)$ with zero average (i.e. $\int_{\mathbf{R}} \psi - 0$), normalized (i.e. $||\psi|| = 1$), and centered in the neighborhood of $t = 0$ [89]. Scaling $\psi(t)$ by a positive quantity s, and translating it by $y \in \mathbf{R}$, a wavelet family can be defined as:

$$\psi_{u,s}(t) := \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right),\ u \in \mathbf{R},\ s > 0 \tag{5.3}$$

Given $x(t) \in L^2(\mathbf{R})$, the continuous wavelet transform of $x(t)$ at time u and scale s (which inversely relate to frequency) is defined as:

$$W(s,u) := \langle x(t), \psi_{s,u} \rangle = \frac{1}{\sqrt{s}} \int x(t)\psi^*\left(\frac{t-u}{s}dt\right) \tag{5.4}$$

where ψ^* denotes the complex conjugate of ψ. CWT decomposes the input signal $x(t)$ into a series of wavelet coefficients. The scalogram of $x(t)$ is defined by the function:

$$\mathbf{S}(s) := ||W(s,u)|| = \sqrt{\int_{-\infty}^{+\infty} |W(s,u)|^2 du} \tag{5.5}$$

If a time interval $[t_0, t_1]$ needs to be considered, the corresponding windowed scalogram is defined by the function:

$$\mathbf{S}_{[t_0,t_1]}(s) := ||W(s,u)||_{[t_0,t_1]} = \sqrt{\int_{t_0}^{t_1} |W(s,u)|^2 du} \tag{5.6}$$

In other words, the scalogram is the absolute value of the CWT of a signal, plotted as a function of time and frequency, as shown in Figure 5.2.

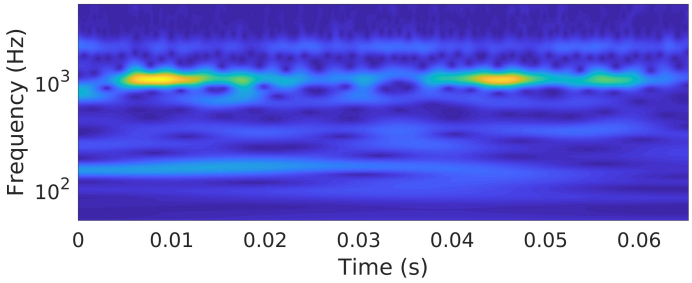

FIGURE 5.2: Scalogram of vibration signal

By combining two transformation, each vibration signal is transformed simultaneously into image form: the time domain image and the time-frequency domain image is shown in Figure 6.2. This representation of vibration signal is named as multiple-domain image-representation (MDIR). By using MDIR data, the problem of fault diagnosis based on vibration signal now can be considered as a task of image classification.

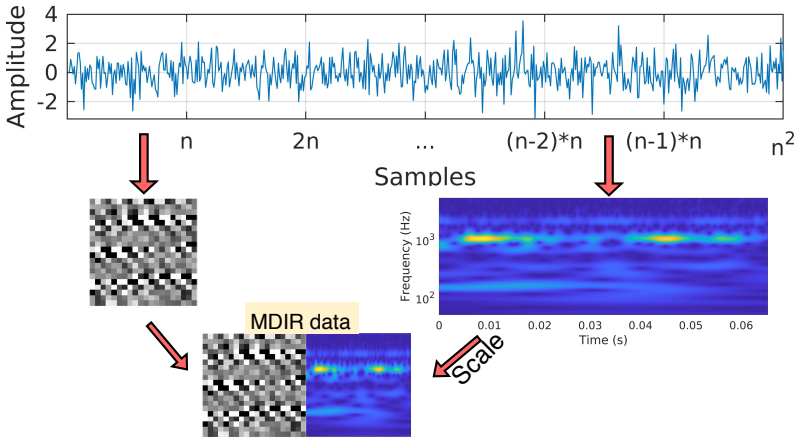

FIGURE 5.3: MDIR data of vibration signal

5.2.2 Proposed wide deep neural network

In the previous section, the multiple-domain image-representation (MDIR) of the vibration signal has been described. To handle the MDIR data, a wide deep neural network (WDNN) is proposed. The structure of WDNN is shown in Figure 5.4.

The proposed WDNN consists of several types of layers, includes convolutional layer, batch normalization layer, pooling layer, feature fusing layer, dense layers, and softmax

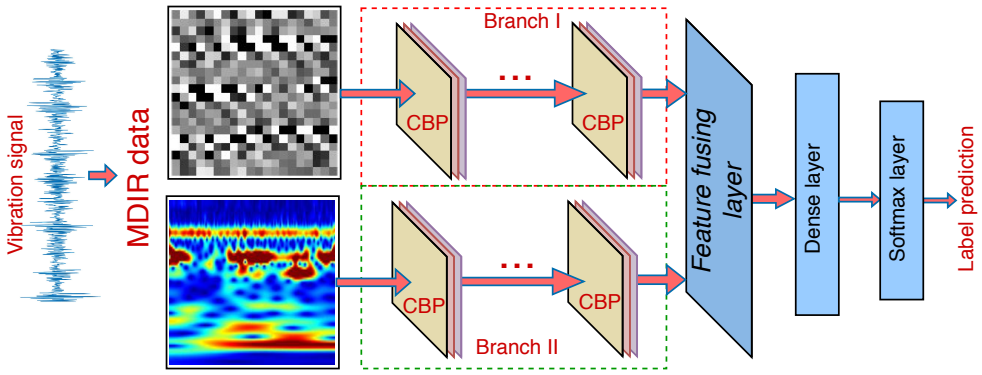

FIGURE 5.4: Wide deep neural network

layer.

Convolutional layer based on the convolution operation is the most important layer type which has been employed in many well-known DNN models such as LeNet [90], AlexNet [46], VGGNet [91], ResNet [92], and DenseNet [93]. The convolutional layer convolves the input with its kernels and feeds the obtained result into the activate function to generate the output. Consider a convolutional layer with m kernels, the input has n feature maps. The output of that layer can be calculated as follows:

$$y_i = f\left(\sum_{j=1}^{n} w_i \otimes x_j + b_i\right), i = 1:m \tag{5.7}$$

where f denotes the activate function. Nowadays, Rectified Linear Unit (ReLU) is extensively used in DNNs as activate function since it is simple and easy to compute. The equation of ReLU function is as follows

$$f(x) = \boldsymbol{max}(0, x) \tag{5.8}$$

Batch normalization layer exploited batch normalization technique proposed in [94] to improve the training process of DNNs. Consider the input with n feature maps $x = (x_i, ..., x_n)$. First, the input feature maps are normalized independently to have the zero mean and one variance.

$$\hat{x}_i = \frac{x_i - \boldsymbol{E}[x_i]}{\sqrt{\boldsymbol{Var}[x_i]}} \tag{5.9}$$

then the output is calculated by:

$$y_i = \gamma_i \hat{x}_i + \beta_i \tag{5.10}$$

where γ_i and β_i are introduced to scale and shift the normalized the input feature maps. Batch normalization ensures the transformation inserted in the network by representing the identity transform.

Pooling layer reduces the dimension of the input feature maps. The dimension reduction operation can be conducted by **max** or **average** operation. Pooling layer computes the max value or average value of a group of neurons at the previous layer.

Three layers including one convolutional layer, one batch normalization layer, and one pooling layer are grouped successively to construct a CBP module. Assume that a CBP module with a convolutional layer with m kernels. With the input feature map $x^{n\times k\times k}$, the output of the CBP module will be $y^{m\times t\times t}$ where $t = k/2$.

As shown in Figure 5.4, the proposed WDNN has two branches. While branch I handles the time domain images, branch II handles the time-frequency domain images of the vibration signals. Each branch consists of several CBP modules. Two branches simultaneously extract features from the input MDIR data, generate two types of feature maps. After that, the feature fusing layer will fuse these feature maps to generate a single one and forward it to the next part of the network. The operation of the feature fusing layer is described as follows:

$$x = \boldsymbol{C}(x_t, x_f) \tag{5.11}$$

where $\boldsymbol{C}$ denotes the fusing operation of the layer; $x_t,\ x_f$ are the feature maps extracted by branches I and II, respectively. Assume that x_t and x_f has sizes $m \times (n \times n)$ and $a \times (b \times b)$, respectively. First, the two feature maps will be flattened to have size $1 \times (m * n^2)$ and $1 \times (a * b^2)$, respectively. The output of this layer x will have size $1 \times (m * n^2 + a * b^2)$.

Dense layer is a traditional perception neural network. The purpose of using a dense layer is to collect all features from the previous feature map. The generated feature map

of this layer will be used for the classification task, which will be done by the softmax layer. With an input feature map x, the softmax layer computes the probabilities:

$$p_j = \frac{\exp(w_i * x)}{\sum_{j=1}^{N} \exp(w_i * x)} \tag{5.12}$$

where N is the number of classes in the classification task. The softmax function calculates the probabilities of each target class over all target classes.

The loss function of WDNN is calculated by cross entropy loss as follows:

$$L(q, p) = -\sum_{j}^{N} q_j \log(p_j) \tag{5.13}$$

where q is the true label of the input data, p is the output of softmax function. N is the number of classes in the classification task.

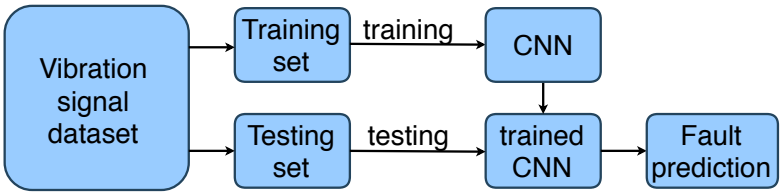

FIGURE 5.5: Diagram of proposed bearing fault diagnosis method

Using the MDIR data and WDNN, the proposed fault diagnosis method is illustrated as in Figure 6.1. As shown in the diagram, the original fault signals are transformed into MDIR data. Then the obtained dataset is split into training set and testing set. The WDNN is trained with the training set by back-propagation algorithm. Using the stochastic gradient descent with momentum, the weights of WDNN are updated by the equations

$$v_t = \beta v_{t-1} + (1 - \beta)\nabla_w L \tag{5.14}$$

$$w = w - \alpha v_t \tag{5.15}$$

where α denotes the learning rate, β is the momentum parameter.

5.3 Experimental study

5.3.1 Signal pre-processing

In the data source supplied by CWRU, each bearing condition fault signal is measured and stored in a single file. Then there are 10 different signal files corresponding to 10 bearing conditions. In the intelligent fault diagnosis approach, classifiers require data samples to be trained. Therefore, the original signal files are split into equal signal samples. Each signal sample must contain enough sampling points to convey the information of the bearing status, that is, if the length is too short, the signal sample can not reflect the bearing health status. Normally, the sample length is selected to be equal to one revolution of the rotary shaft. In this work, the rotary speed of the shaft is $S = 1796\ rpm$. Accordingly, the rotary frequency is $S_f = \frac{1796}{60} \approx 30\ Hz$. In the CWRU test-bed, the sampling frequency $F = 12000\ Hz$. The minimum value of the sample length is $\frac{F}{S_f} = 400\ (sampling\ point)$. As mentioned in Section 5.2.2, in the proposed method, there is a step where signal samples are transformed into gray images and apply CNN-liked neural network to classify. Therefore we aim to make a MNIST-liked data set where each image has a size of $1 \times 28 \times 28$. Therefore, the length of signal samples is selected at the value of $28 \times 28 = 784$. The time domain image has the size of $1 \times 28 \times 28$.

The time-frequency domain image conversion exploits CWT using the Morse wavelet function. To deal with the time-frequency domain image, the obtained time-frequency images are scaled to a size of $3 \times 224 \times 224$. The MDIR data of the vibration signals with ten labels are obtained as shown in Figure 5.6.

From each original signal files, 300 data samples are obtained. Accordingly, the 10-class classification task is balance since each class have 300 data samples. For each label, the image data are split randomly with the ratio 7:3 for the training set and the test set.

5.3.2 Design and train the proposed DNN

The structure of the proposed WDNN is designed as follows. First of all, the number of branches is 2 corresponding to two types of input images. The input sizes of branches

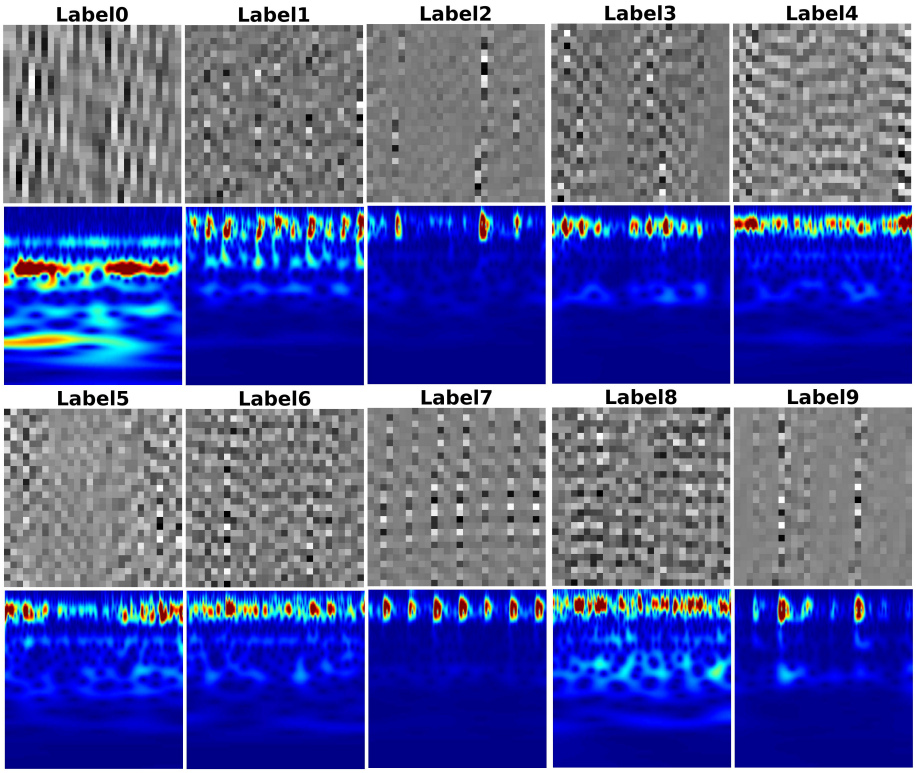

FIGURE 5.6: MDIR data of vibration signals

have to be suitable for the input images. Therefore, the first branch has the input size of $1 \times 28 \times 28$, the second branch has the input size of $3 \times 224 \times 224$. The kernel sizes of convolutional layers and pooling layers are 3×3 and 2×2 respectively. The number of kernels in convolutional layers are selected by a simple rule. Start with a small number of kernels in the first layer and double that number in the next layer. The first convolutional layer has 8 kernels, the second layer has 16 kernels, the third layer has 32 kernels, and so on. Each module CBP consists of 1 convolutional layer, one batch normalization layer, and one pooling layer. Since we use the zero padding method in convolutional layers and the kernel size of 2×2 in pooling layers, after each module CBP, the size of data will decrease by the factor of 2. The number of CBP modules is increased by 1 until the size of the output data is an odd number. In the first branch, 2 module CBPs are used. In the second branch, 5 module CBPs. The proposed WDNN has the configuration as shown in Figure 5.7. Branch I that consists of two CPB modules handles the time domain images. Branch II uses five CBP modules to handles the time-frequency domain images. The output of branch I has a size of $16 \times 7 \times 7$. The output of branch II has a size of $128 \times 7 \times 7$. These two outputs are fed into the feature fusing layer to

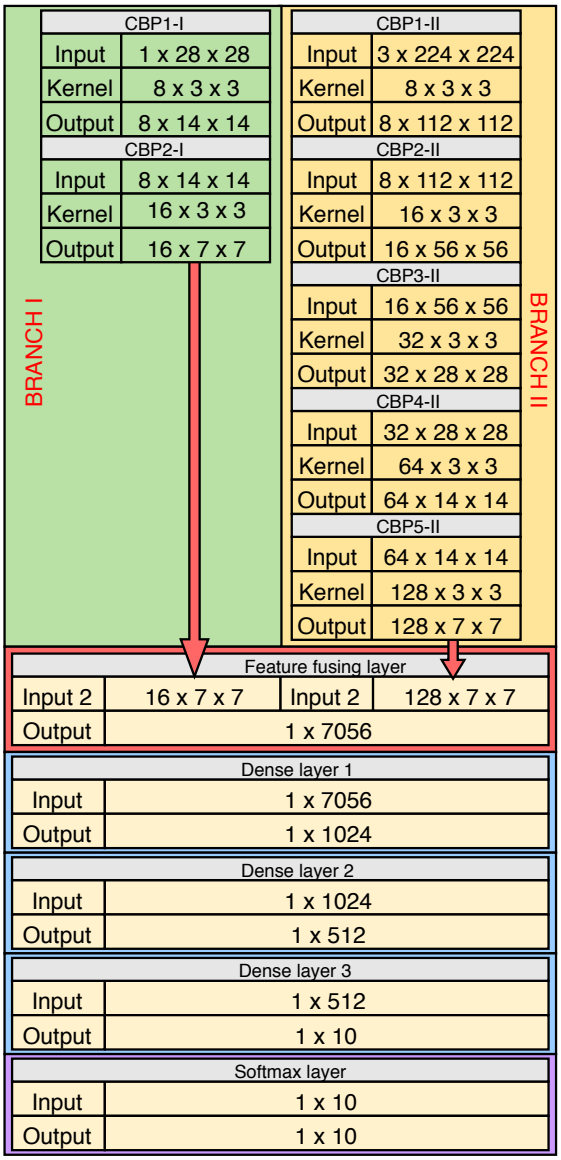

FIGURE 5.7: Configuration of WDNN

generate the output data with size of 1×7056. Then three successive dense layers are used to learn from the fused feature map. Finally, a softmax layer with ten outputs is used to classify the feature map generated by the third dense layer.

The WDNN is trained by mini batch stochastic gradient descent with momentum algorithm, the learning rate $\alpha = 0.001$, the momentum $\beta = 0.9$, and the batch size $b = 10$. The training-testing process of the network is conducted using the k-fold validation method where $k = 5$.

5.3.3 Fault diagnosis result

Three other bearing fault diagnosis methods are adopted to make comparisons with the proposed method as follows. The first method is published in [22]. In this method, the deep transfer learning technique is utilized to transfer the very deep neural network (Alexnet) pre-trained in the image classification domain into the domain of bearing fault diagnosis. The vibration signals are transformed into image form by using CWT. The second method is published in [87]. This method using Lenet-5 which is a classical type of CNN in [95]. In this method, each sample of vibration signal is rearranged into a square matrix. The third method is published in [66]. This method utilizes the 1-D form of CNN, which can process the vibration signal samples directly in 1-D form without any transformation. All methods are trained and tested with the 5-fold validation scenario with the same bearing data source. Under each operating condition of the bearing test-bed, one corresponding bearing data set is obtained. Accordingly, a total four bearing data sets are prepared to evaluate the bearing fault diagnosis methods.

TABLE 5.1: Accuracy of Compared Methods Using 5-fold Validation Under Different Load Conditions

	Proposed method				Method in [87]				Method in [22]				Method in [66]			
Fold	0 hp	1 hp	2 hp	3 hp	0 hp	1 hp	2 hp	3 hp	0 hp	1 hp	2 hp	3 hp	0 hp	1 hp	2 hp	3 hp
1	99.67	100	100	100	100	99.67	100	99.83	99.17	99.83	100	99.0	95.83	99.17	99.67	98.33
2	100	99.67	99.67	100	100	99.83	100	100	98.5	99.83	100	99.83	99.5	96.83	98.5	96.83
3	100	99.33	100	99.67	100	100	99.83	100	99.83	99.67	99.67	97.83	95.67	95.5	95.5	99.33
4	100	99.67	99.33	100	100	99.67	99.5	100	99.67	99.83	99.33	99.67	98.83	96.33	95.5	99.33
5	99.83	99.83	100	100	99.0	100	100	100	99.33	99.33	98.67	99.5	98.55	95.67	98.67	98.33
Mean	99.9	99.7	99.8	99.93	99.8	99.83	99.87	99.97	99.3	99.7	99.53	99.17	97.68	96.7	97.57	98.43
Std	0.13	0.22	0.27	0.13	0.4	0.15	0.19	0.07	0.46	0.19	0.5	0.72	1.6	1.32	1.73	0.92

The accuracy of all methods is shown in Table 5.1 and Figure 5.8. It can be observed that in all cases of load conditions, the proposed method and the Alexnet-based method in [22] achieve the best performance. The two methods have high mean accuracy and small standard deviation. The Lenet5-based method in [87] has a little lower accuracy. The CNN1D-based method in [66] has the poorest performance with low mean accuracy and big standard deviation.

Deep neural networks often consist of a huge number of trainable parameters. Therefore, the training and validating of deep neural networks require a considerable amount of time. However, the training and validating time is not a big problem thanks to the popularity of powerful computing hardware and this process can be conducted offline.

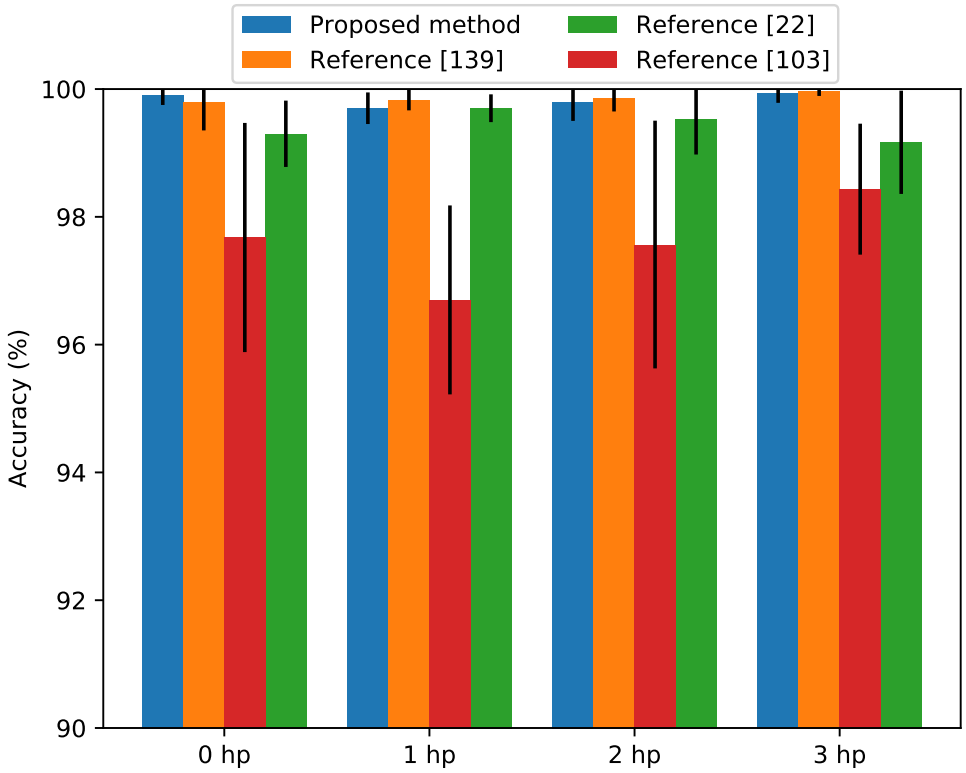

FIGURE 5.8: Mean accuracy (%) of methods under different load conditions

The more important aspect should be considered is the processing time. This time is defined as the amount of time for a trained model processes a signal sample to predict the label of that sample. The processing time reflects the ability of the model in detecting bearing faults in real time. As shown in Table 5.2, the proposed model consists of 2 branches with over 7.3 millions trainable parameters. The model in [22] has the most parameters - over 61 millions. The processing time measured in millisecond (ms) of all methods are not much different. Obviously, the model in [22] requires the longest processing time since it consists of a huge number of parameters.

TABLE 5.2: Comparison of Model Complexity

Model	Proposed model	[87]	[22]	[66]
Branch	2	1	1	1
Parameter	7,338,842	61,100,840	3,028,426	566,670
Time (ms)	2.98	3.59	2.12	2.23

5.3.4 Evaluation under noisy conditions

The bearing data set supplied by CWRU has been extensively employed as a benchmark for evaluating bearing fault diagnosis methods. Recently, the proposed fault diagnosis methods with advantage signal processing and feature learning techniques can achieve very high accuracy. It is not easy to highlight the performance of the newly proposed fault diagnosis methods. Therefore, noise signals are often added into the original signals to evaluate methods. This way can help to evaluate the robustness of fault diagnosis methods under more challenging conditions. In this scenario of evaluating diagnosis methods, Gaussian white noise (AGWN) is added into the original vibration signals as in Figure 5.9.

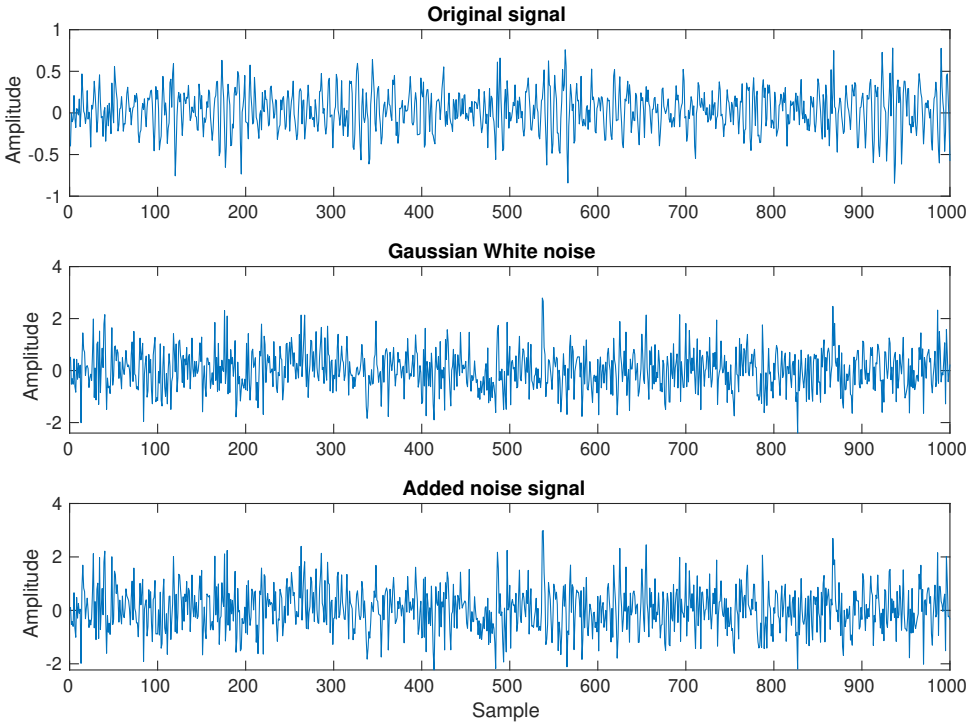

FIGURE 5.9: Noise signal

The signal-to-noise ratio (SNR) is defined to measure the level of the obtained noisy signal to the level of the additional Gaussian noise. The SNR of a noisy signal is computed as follows:

$$SNR = 10 \log \left(\frac{P_{signal}}{P_{noise}} \right) \tag{5.16}$$

where P_{signal} and P_{noise} are the power of signal and noise, respectively. Eight noise levels

(dB) in the range of $[-8, -7, ..., -1]$ are taken into account. The comparison is illustrated in Figure 5.10.

In the previous Section, all compared methods achieved very high diagnostic accuracy when the inputs are the original signals. However, when the input signals are transformed by adding noisy signals, their performance decreases dramatically. Obviously, the noise makes it harder to extract fault signatures from the signals. It can be observed that, at the severest case (-8 dB), the methods in [87] and [66] are totally fail since their accuracy are under 50%. The trend of all methods' performance is, higher the noise level, lower the accuracy of the diagnosis result. Among all methods, the proposed one achieves the best performance with good robustness against noise. Even under the worst noise case (-8 dB), it achieves the accuracy of 57 %. Comparing the structures of all methods, we can see that the method in [22] only takes care of the time-frequency domain features; the methods in [87] and [66] only consider the time domain features. The difference of the proposed methods is that it can receive multiple feature domains. As a result, the proposed method can extract more robust features from signals in noise conditions. So it can be concluded that the proposed network with a wide deep structure can extract fault features more effectively even under severe noise conditions. That leads to better diagnostic performance compared to other DNNs.

5.4 Conclusion

This chapter proposed a novel method of bearing fault diagnosis based on the vibration signals. By using simple transformation methods, time-series vibration signals are transformed into high dimensional data form (MDIR). By using this transformation, the task of fault diagnosis becomes the task of image classification. A novel DNN with wide structure is proposed to handle the MDIR data of vibration signal, named WDNN. The proposed WDNN inherits the advantages of CNN in processing high dimensional data. In addition, WDNN has a wide structure with two branches, which can simultaneously extract features from time domain and time-frequency domain.

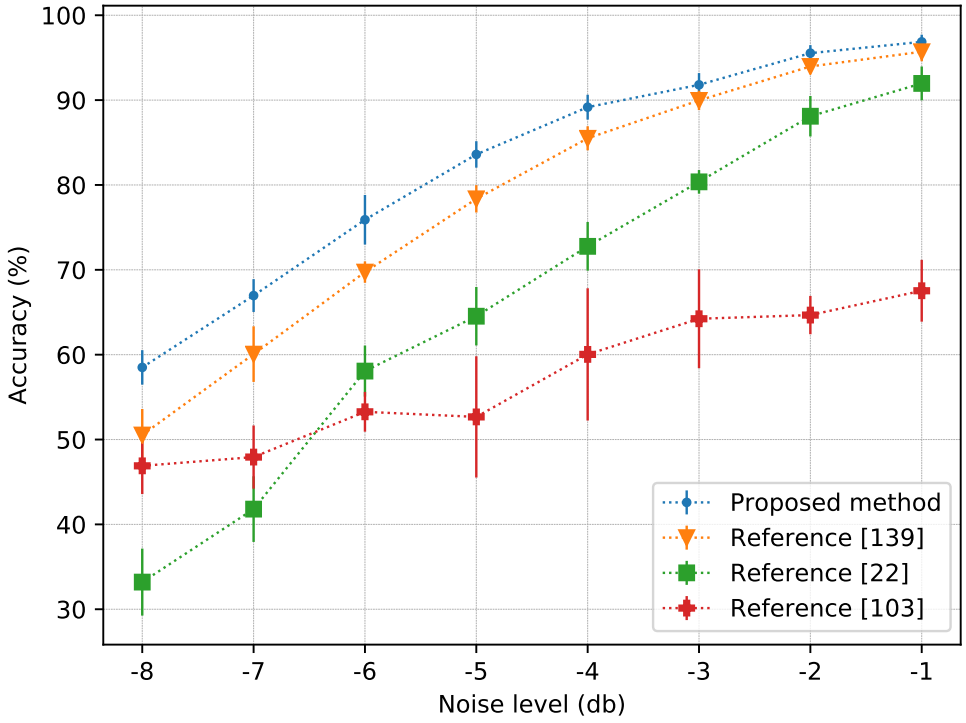

FIGURE 5.10: Diagnosis accuracy of 4 methods under different noise levels (load condition: 2 hp)

Chapter 6

Transfer Learning

6.1 Introduction

Transfer learning (TF) is a method in ML where a model designed for a task is reused for another task. Designing and training a new DNN model from scratch is difficult and requires a huge number of labeled data. Using TF in ML applications does not require much labeled data for training since it can utilize DNN models from another domain. As a result, TF helps to save time for designing and training DNNs from scratch.

Motivated by the advantages of TF, this chapter proposed a signal-based bearing fault diagnosis using TF. The well-known Squeezenet [96] which was trained on the image classification domain, is utilized to extract fault features from vibration signals measured by multiple sensors. From each signal source, a corresponding fault feature set is obtained. Then each feature set is classified to generate a conclusion about the fault type. The conclusions from multiple feature sets may be different. At the final step, the Dempster-Shafer theory is employed to combine all classification results to generate a final conclusion about the bearing type.

6.2 Method development

6.2.1 Transfer learning

Transfer learning is a ML technique that involves transferring the knowledge learned in a domain to another domain. In practice, it is common to pre-train a network on a

data set, then use that network either as an initialization or a fixed feature extractor for the task of current interest [97]. There are two main scenarios of transfer learning as follows:

(a) Fine-tuning the pre-trained network by continuing the back-propagation training process with the data set of the new task. It is possible to fine-tune all layers of the network or to fine-tune only some selective layers of the network.

(b) Only using the design of the network and train the whole network from scratch with the data set of the new task. This scenario often requires much more time for training compared to the fine-tuning scenario.

(c) Using the pre-trained network as a fixed feature extractor by freezing all layers of the network except the final layer but replacing the last one with a new one that respects the number of classes in the new task.

6.2.2 Dempster-Shafer evidence theory

Dempster-Shafer evidence theory (DSET) provides a method to combine information from multiple sources and measure the confidence that a given event occurs. Consider a set of n events $\Theta = \{\theta_1,\ \theta_2,\ ,...,\ \theta_n\}$. The power set of Θ which consists of all its subsets is denoted by 2^{Θ}. The basic probability assignment (**BPA**) of Θ is defined as a mapping $\mathbf{m} : 2^{\Theta} \rightarrow [0, 1]$ which satisfies:

$$\begin{aligned} \mathbf{m}(\emptyset) &= 0 \\ \sum_{A \subset \Theta} \mathbf{m}(A) &= 1 \end{aligned} \tag{6.1}$$

The BPA function $\mathbf{m}(A)$ reflects the confidence or the degree of evidence supporting the proposition A. Accordingly, the belief function $\mathbf{BEL}(A)$ and the plausibility function $\mathbf{PL}(A)$ reflect the trust and the doubted degree of A, respectively, are defined as follows:

$$\mathbf{BEL}(A) = \sum_{A \subset \Theta} \mathbf{m}(A) \tag{6.2}$$

$$\mathbf{PL}(A) = \sum_{A \not\subset \Theta} \mathbf{m}(A) \tag{6.3}$$

Consider two BPA functions $\mathbf{m}_1$ and $\mathbf{m}_2$ obtained from two information sources, the DSET based fusion algorithm is defined as:

$$\mathbf{m}(\emptyset) = 0 \tag{6.4}$$

$$\mathbf{m}(A) = \frac{1}{1-K} \sum_{B \cap C = A} \tag{6.5}$$

where K represents the amount of conflict between the two sources computed by:

$$K = \sum_{B \cup C = \emptyset} \mathbf{m}_1(B)\mathbf{m}_2(C) \tag{6.6}$$

6.2.3 The proposed bearing fault diagnosis method

In this work, a bearing fault diagnosis method using TF and DSET is proposed. The overall procedure is illustrated in Figure 6.1.

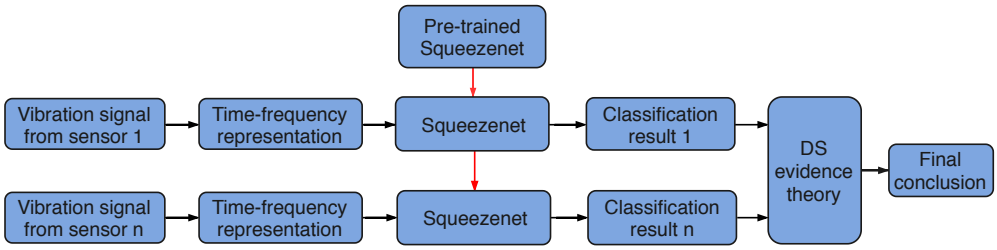

FIGURE 6.1: The proposed fault diagnosis method

As shown in Figure 6.1, Squeezenet is adopted to transfer to the fault diagnosis task. Squeezenet is a DNN model initially proposed for the image classification task [96]. The network consists of 18 layers can classify images into 1000 different labels. The network is trained on a data set of more than a million images from the ImageNet database [98]. As mentioned in 6.2.1, there are three ways to apply TF including training the network from scratch, fine-tuning the pre-trained network, and freezing the network to use it as a fixed feature extractor. In this current work, the source task of the network is

image classification, while the target task is fault diagnosis. The two tasks are quite different domains. Therefore, the scenario of fine-tuning the network by the data set bearing fault signal is selected. The last layer of Squeezenet is replaced with a new one that suitable for the number of fault types of bearing. Originally, the inputs of Squeezenet are images with a size of $224 \times 224 \times 3$. To use the network in the fault diagnosis task, the fault signals which are time-series data are pre-processed to have the size that suitable for the receptive input of Squeezenet. Wavelet transform (WT) is exploited to represent the time series data into 2-D form. The wavelet transform $WT_x(b,a)$ of a signal $x(t)$ can be described as [99]:

$$WT_x(b,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\phi(\frac{t-b}{a})dt, \tag{6.7}$$

where a is a scale parameter, b is the time parameter, and ϕ is the mother wavelet. Figure 6.2 shows a time-frequency representation of a bearing vibration signal.

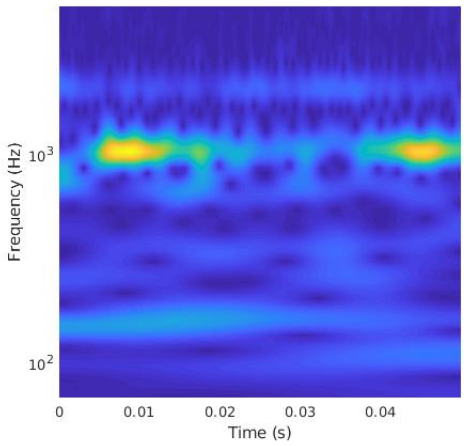

FIGURE 6.2: Time-frequency representation of vibration signal by wavelet transform

The new data sets obtained by WT are used to fine-tune the pre-trained Squeezenet. After being fine-tuned, each network has its own prediction about the label of the input data. For the sake of simplicity in explaining the proposed method, assume that there are two different sensors that measure the fault signals. From each signals source,

each fine-tuned network will have a prediction about the probability of fault occurring. Assume that there are N types of faults labeled as $F_1, F_2, ..., F_N$. The first fine-tuned network will generate the probabilities $p_1(F_i)$ that satisfy $\sum_{i=1}^{N} p_1(F_i) = 1$. Similarly, the second fine-tuned network will generate the probability $p_2(F_i)$ that satisfy $\sum_{i=1}^{N} p_2(F_i) = 1$. To employ DSET fusion, the probabilities p_1 and p_2 are considered as two BPA functions $\mathbf{m}_1$ and $\mathbf{m}_2$, respectively. Then the equations 6.4, 6.5, and 6.6 are used to compute the fused BPA functions $\mathbf{m}(F_i)$. The final conclusion is generated by finding the value $\max(\mathbf{m}(F_i))$.

6.3 Experimental study

6.3.1 Pre-processing vibration signal

Vibration signals are measured by accelerometers attached to the housing with magnetic bases. Obtained signals are sampled at frequency 12000 Hz. Ten types of vibration signals are recorded and labeled as in Table 6.1.

TABLE 6.1: Signal labels

Bearing condition	Fault size (mils)	Label
Normal		0
Inner race fault	7	1
Inner race fault	14	2
Inner race fault	21	3
Ball fault	7	4
Ball fault	14	5
Ball fault	21	6
Outer race fault	7	7
Outer race fault	14	8
Outer race fault	21	9

Two different accelerometers are installed at the fan end (FE sensor) and the driver end (DE) sensor to measure vibration signals. Ten types of bearing vibration signals of four types of bearing conditions with different fault sizes are shown in Figure 6.3. Each sample signal is transformed into time-frequency representation. This type of time-frequency representation can be considered as an RGB image with a size of $224 \times 224 \times 3$. After the transformation of all signal samples into time-frequency images, the image data set consisting of 3000 images is obtained. The image data set is split randomly

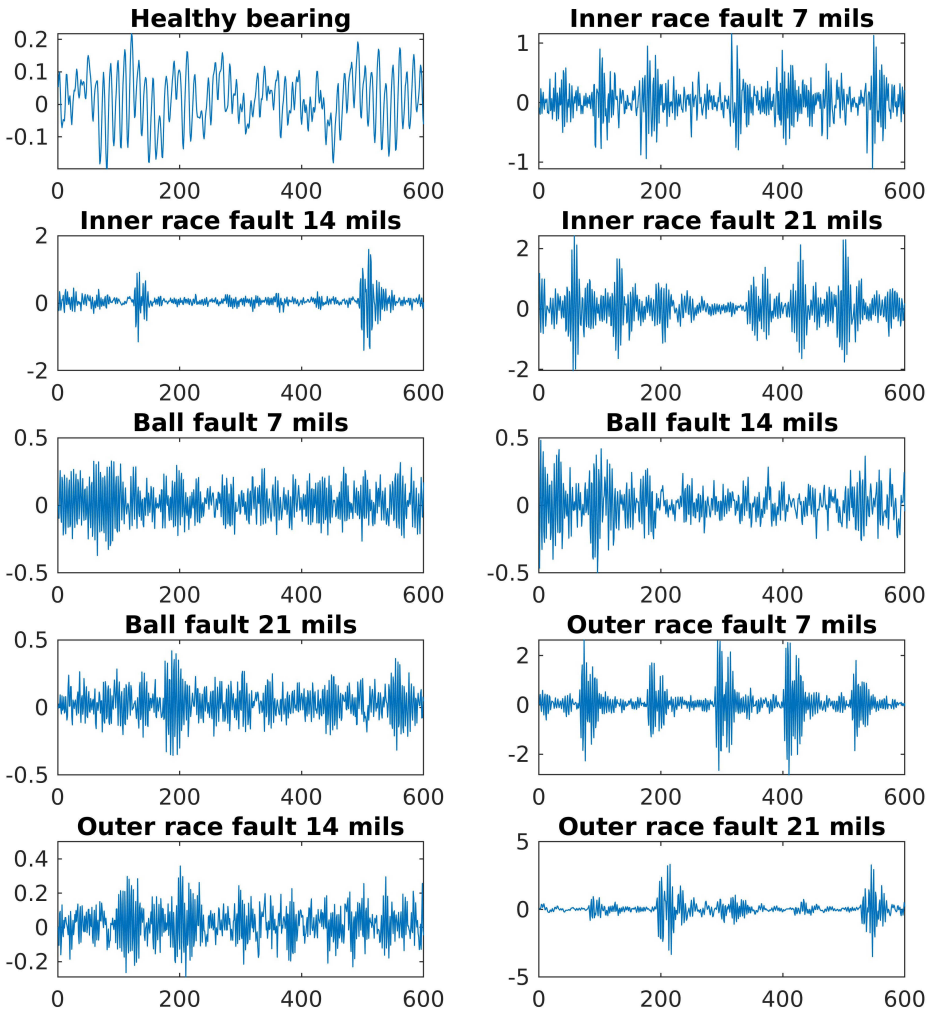

FIGURE 6.3: Vibration signals (DE sensor)

by each class into three subsets: the training data set (2400 images) and testing data set (600 images). The same time-frequency samples of vibration signals are shown in Figure 6.4.

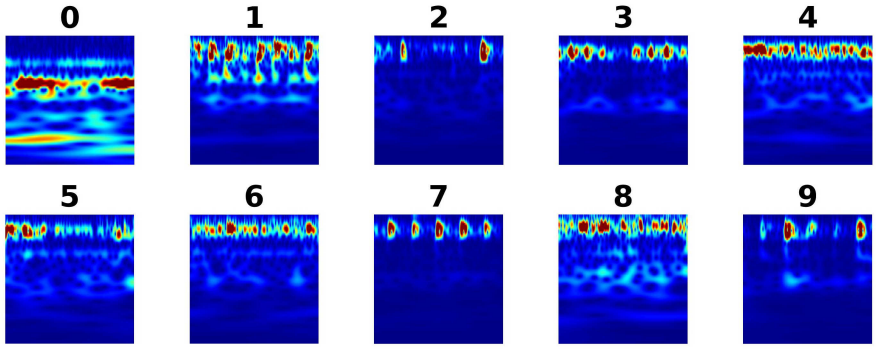

FIGURE 6.4: Time-frequency representation of vibration signals (DE sensor)

6.3.2 Fine-tuning pre-trained Squeezenet

The last layer of Squeezenet is replaced by a new layer with 10 outputs corresponding with 10 types of bearing signals need to be classified. Two scenarios are considered: fine-tuning the pre-trained network and training the network from scratch. Two scenarios use the same mini-batch stochastic gradient descent with the learning rate of 0.001, the momentum of 0.9, and the batch size of 10. The training processes are shown in Figure 6.5. It can be observed that, if the pre-trained network is fine-tuned, it quickly converges in both cases of using DE data and FE data. On the other hand, if the network is trained from scratch, the networks converge slowly.

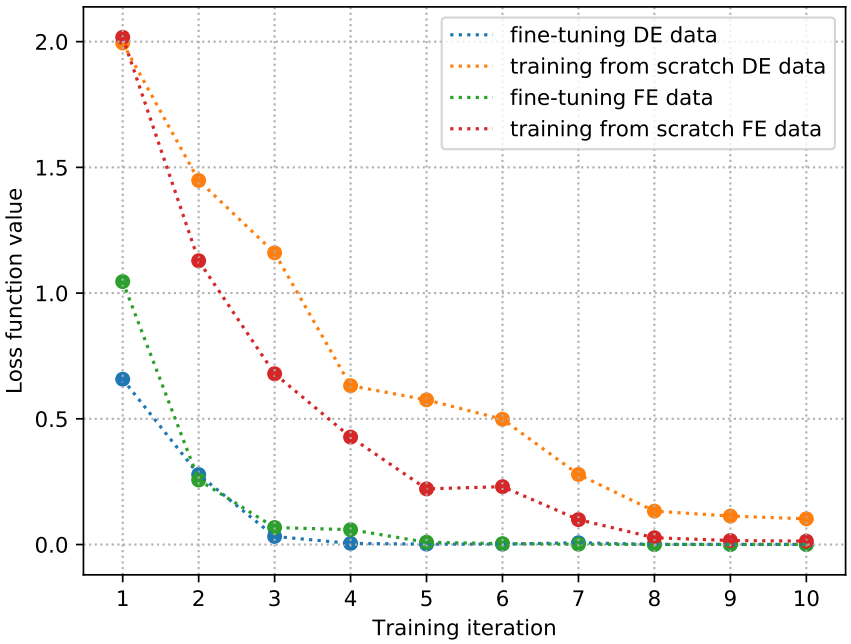

FIGURE 6.5: Fine-tuning and training loss

6.3.3 Experiment result

In actual applications, especially in the industry, the measured signals are severely contaminated by noise. Noise component in measured signals makes it is difficult to extract the feature and reduce the fault detection accuracy. There, the robustness in the noisy environment of fault diagnosis methods is worth to study. Therefore, in experiments, the additive Gaussian white noise (AGWN) is added into the original vibration signals to mimic the noisy environment. After being added AGWN, the signal-to-noise ratio

(SNR) of the signal is defined as:

$$SNR = 10 \log_{10} \left(\frac{P_{signal}}{P_{noise}} \right) \tag{6.8}$$

where P_{signal} and P_{noise} are the power the signal and the AGWN, respectively. Eight levels of SNR from $-8\ dB$ to $-1\ dB$ are considered.

The proposed method is compared with several other fault diagnosis methods. The first method uses a convolutional neural network (CNN) to classify a time-frequency image representation of vibration signals. The second method uses neural networks (NNs) to classify the feature sets extracted by WT in time-frequency domain of vibration signals. This method also uses DSET to fuse the results of two NNs to generate a final conclusion. The classification accuracy of all methods are shown in Figure 6.6. It can be observed that with the original signals (no adding AGWN), all methods achieve very high accuracy, almost 100 %. However, the more severe the noise level, the lower accuracy of all methods. It is easy to see that the CNN method has the worst robustness. In the worst case where $SNR = -8\ dB$, the CNN method with both case DE and FE sensors only achieves the accuracy of about 30 %. The proposed method using TF with single sensor DE of FE (without DSET based fusion) achieves quite good performance. In the case of using DSET, the best performance is achieved. In can be observed that, even in the worst case where $SNR = -8\ dB$, the proposed methods with TF and DSET achieve the accuracy of about 82 %.

6.4 Conclusion

In this chapter, a bearing fault diagnosis method based on transfer learning and DS evidence theory is proposed. The proposed fault diagnosis method achieves very high accuracy. In addition, the proposed method has very good robustness against noise signals. The experiment in this work is conducted with multiple vibration signal sources which are measured by two different accelerometers. However, the proposed method also has the ability to be applied for different types of signal together, such as acoustic emission signals, current signals, and so on.

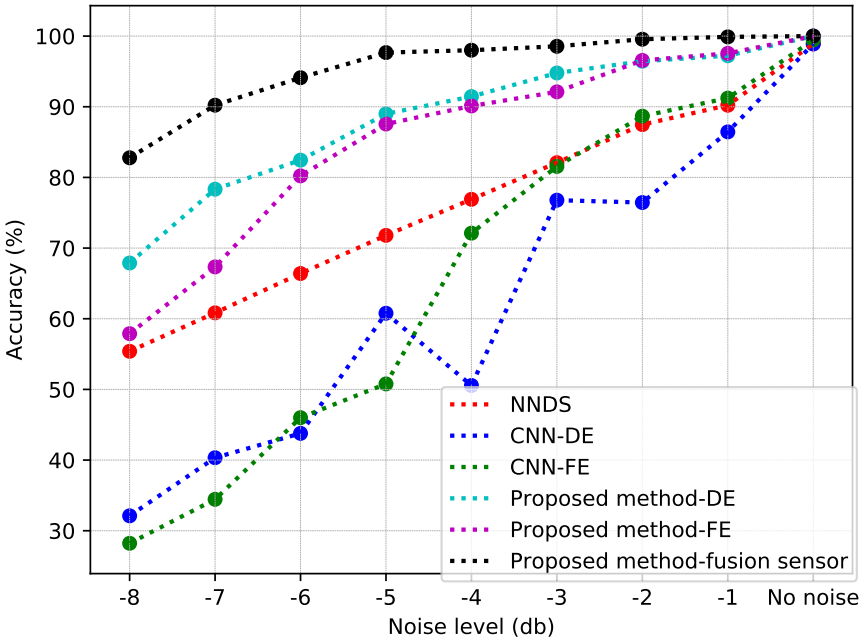

FIGURE 6.6: Fault diagnosis accuracy under noisy environments

Chapter 7

Conclusion

The ultimate goal of this study is to enhance the ability and performance of deep leaning and deep neural networks in bearing fault diagnosis. Several novel diagnosis methods are proposed. During this research, the following findings are successfully applied for bearing fault diagnosis.

- Automatically feature learning by deep neural network overcomes the conventional hand-crafted feature engineering in the accuracy aspect. Moreover, deep neural network feature learning does not require knowledge domain expert and human labor.
- By transforming bearing signals (vibration or current) into higher dimensional data (vibration image, multiple-domain image-representation), the task of bearing fault diagnosis is transformed into the task of visual data classification, which can utilize the power of convolutional neural networks.
- Not only deep neural network (more number of network layers), but also wide neural network (more number of input layers) can enhance the feature learning ability of neural networks.
- Information fusion can distill and fuse multiple neural networks. As a result, the fault diagnosis actuary is increased.
- Transfer learning can partly solve the difficulty of designing and training a new DNN model from scratch.

The following issues should be considered in future work:

- Extend the application of deep learning to another type of signals, such as acoustic emission, temperature signals.
- Condition monitoring is the process of monitoring the current condition and predicting the future condition of machines while in operation. In this thesis, only the fault detection and classification problems were studied. It means only the current condition is observed. Predicting the future condition is referred as fault prognostics. Prognostics of machine faults have the potential to give the greatest economic benefit from condition monitoring. Thus, the fault prognostic for machines should be focused on in future work.

Publication

Journal paper

- Hoang, Duy Tang, and Hee Jun Kang. "Rolling element bearing fault diagnosis using convolutional neural network and vibration image." Cognitive Systems Research 53 (2019): 42-50.
- Hoang, Duy Tang, and Hee Jun Kang. "A survey on Deep Learning based bearing fault diagnosis." Neurocomputing 335 (2019): 327-335.
- Hoang, Duy Tang, and Hee Jun Kang. "A Motor Current Signal Based Bearing Fault Diagnosis Using Deep Learning And Information Fusion." IEEE Transactions on Instrumentation and Measurement (2019).
- Van, Mien, Duy Tang Hoang, and Hee Jun Kang. "Bearing Fault Diagnosis Using a Particle Swarm Optimization-Least Squares Wavelet Support Vector Machine Classifier." Sensors 20.12 (2020): 3422.

Conference paper

- Hoang, Duy Tang, and Hee Jun Kang. "A bearing fault diagnosis method based on autoencoder, particle swarm optimization and support vector machine." International Conference on Intelligent Computing. Springer, Cham, 2018.
- Hoang, Duy Tang, and Hee Jun Kang. "Deep belief network and Dempster-Shafer evidence theory for bearing fault diagnosis." 2018 IEEE 27th International Symposium on Industrial Electronics (ISIE). IEEE, 2018.
- Hoang, Duy Tang, and Hee Jun Kang. "Rotary Machine Fault Diagnosis Using Scalogram Image and Convolutional Neural Network with Batch Normalization." International Conference on Intelligent Computing. Springer, Cham, 2019.
- Hoang, Duy Tang, and Hee Jun Kang. "A Bearing Fault Diagnosis Method using Transfer Learning and Dempster-Shafer Evidence Theory." Proceedings of the 2019 International Conference on Artificial Intelligence, Robotics and Control. 2019.

Bibliography

[1] BKN Rao. *Handbook of condition monitoring*. Elsevier, 1996.

[2] Robert Bond Randall. *Vibration-based condition monitoring: industrial, aerospace and automotive applications*. John Wiley & Sons, 2011.

[3] A Agoston, C Ötsch, and Bernhard Jakoby. "Viscosity sensors for engine oil condition monitoring—Application and interpretation of results". In: *Sensors and Actuators A: Physical* 121.2 (2005), pp. 327–332.

[4] David Mba and Raj BKN Rao. "Development of acoustic emission technology for condition monitoring and diagnosis of rotating machines; bearings, pumps, gearboxes, engines and rotating structures." In: (2006).

[5] Manjeevan Seera et al. "Fault detection and diagnosis of induction motors using motor current signature analysis and a hybrid FMM–CART model". In: *IEEE transactions on neural networks and learning systems* 23.1 (2011), pp. 97–108.

[6] E Peter Carden and Paul Fanning. "Vibration based condition monitoring: a review". In: *Structural health monitoring* 3.4 (2004), pp. 355–377.

[7] Shukri Ali Abdusslam. "Detection and diagnosis of rolling element bearing faults using time encoded signal processing and recognition". PhD thesis. University of Huddersfield, 2012.

[8] James E Berry. "How to track rolling element bearing health with vibration signature analysis". In: *Sound and Vibration* 25.11 (1991), pp. 24–35.

[9] R Moore et al. "The Reliability-based maintenance strategy: a vision for improving industrial productivity". In: *CSI Ind. Report, Knoxville, TN* (1993).

[10] Curtis Lanham. "Understanding the tests that are recommended for electric motor predictive maintenance". In: *Baker Instrument Company* (2002).

[11] Subhasis Nandi, Hamid A Toliyat, and Xiaodong Li. "Condition monitoring and fault diagnosis of electrical motors—A review". In: *IEEE transactions on energy conversion* 20.4 (2005), pp. 719–729.

[12] Tedric A Harris and Michael N Kotzalas. *Essential concepts of bearing technology*. CRC press, 2006.

[13] Christian Lessmeier et al. "Condition Monitoring of Bearing Damage in Electromechanical Drive Systems by Using Motor Current Signals of Electric Motors: A Benchmark Data Set for Data-Driven Classification". In: *Proceedings of the European Conference of the Prognostics and Health Management Society*. 2016, pp. 05–08.

[14] Jerry L Miller and Duka Kitaljevich. "In-line oil debris monitor for aircraft engine condition assessment". In: *2000 IEEE Aerospace Conference. Proceedings (Cat. No. 00TH8484)*. Vol. 6. IEEE. 2000, pp. 49–56.

[15] Rolling Bearings. "Dynamic Load Ratings and Rating Life". In: *International Organization for Standards: International Standard ISO* 281 (2007).

[16] Thi Phuong Khanh Nguyen et al. "Analysis and comparison of multiple features for fault detection and prognostic in ball bearings". In: 2018.

[17] Sheng Fu et al. "Rolling bearing diagnosing method based on time domain analysis and adaptive fuzzy-means clustering". In: *Shock and Vibration* 2016 (2016).

[18] James Kuria Kimotho and Walter Sextro. "An approach for feature extraction and selection from non-trending data for machinery prognosis". In: *Proceedings of the second european conference of the prognostics and health management society*. Vol. 5. 4. 2014, pp. 1–8.

[19] Sanyam Shukla et al. "Analysis of statistical features for fault detection in ball bearing". In: *Computational Intelligence and Computing Research (ICCIC), 2015 IEEE International Conference on*. IEEE. 2015, pp. 1–7.

[20] Mahdi Karimi. "Rolling element bearing fault diagnostics using the blind deconvolution technique". PhD thesis. Queensland University of Technology, 2006.

[21] PD McFadden and JD Smith. "Vibration monitoring of rolling element bearings by the high-frequency resonance technique—a review". In: *Tribology international* 17.1 (1984), pp. 3–10.

[22] Jianyu Wang et al. "A Deep Learning Method for Bearing Fault Diagnosis Based on Time-frequency Image". In: *IEEE Access* (2019).

[23] Ian Howard. *A Review of Rolling Element Bearing Vibration'Detection, Diagnosis and Prognosis'*. Tech. rep. Defence Science And Technology Organization Canberra (Australia), 1994.

[24] De-Shuang Huang. "Systematic theory of neural networks for pattern recognition". In: *Publishing House of Electronic Industry of China, Beijing* 201 (1996).

[25] Jafar Zarei, Mohammad Amin Tajeddini, and Hamid Reza Karimi. "Vibration analysis for bearing fault detection and classification using an intelligent filter". In: *Mechatronics* 24.2 (2014), pp. 151–157.

[26] Juan Luis Ferrando Chacon et al. "A novel approach for incipient defect detection in rolling bearings using acoustic emission technique". In: *Applied Acoustics* 89 (2015), pp. 88–100.

[27] Sukhjeet Singh, Amit Kumar, and Navin Kumar. "Motor current signature analysis for bearing fault detection in mechanical systems". In: *Procedia Materials Science* 6 (2014), pp. 171–177.

[28] Zhong-Qiu Zhao, De-Shuang Huang, and Bing-Yu Sun. "Human face recognition based on multi-features using neural networks committee". In: *Pattern Recognition Letters* 25.12 (2004), pp. 1351–1358.

[29] B Samanta and KR Al-Balushi. "Artificial neural network based fault diagnostics of rolling element bearings using time-domain features". In: *Mechanical systems and signal processing* 17.2 (2003), pp. 317–328.

[30] Arnaz Malhi and Robert X Gao. "PCA-based feature selection scheme for machine defect classification". In: *IEEE Transactions on Instrumentation and Measurement* 53.6 (2004), pp. 1517–1525.

[31] Gary G Yen and K-C Lin. "Wavelet packet feature extraction for vibration monitoring". In: *IEEE transactions on industrial electronics* 47.3 (2000), pp. 650–667.

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

[33] Farzad Hemmati, Wasim Orfali, and Mohamed S Gadala. "Roller bearing acoustic signature extraction by wavelet packet transform, applications in fault detection and size estimation". In: *Applied Acoustics* 104 (2016), pp. 101–118.

[34] Mien Van and Hee-Jun Kang. "Two-stage feature selection for bearing fault diagnosis based on dual-tree complex wavelet transform and empirical mode decomposition". In: *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 230.2 (2016), pp. 291–302.

[35] Chuan Li et al. "Rolling element bearing defect detection using the generalized synchrosqueezing transform guided by time–frequency ridge enhancement". In: *ISA transactions* 60 (2016), pp. 274–284.

[36] Aapo Hyvärinen and Erkki Oja. "Independent component analysis: algorithms and applications". In: *Neural networks* 13.4 (2000), pp. 411–430.

[37] Isabelle Guyon and André Elisseeff. "An introduction to variable and feature selection". In: *Journal of machine learning research* 3.Mar (2003), pp. 1157–1182.

[38] Piero Baraldi et al. "Hierarchical k-nearest neighbours classification and binary differential evolution for fault diagnostics of automotive bearings operating under variable conditions". In: *Engineering Applications of Artificial Intelligence* 56 (2016), pp. 1–13.

[39] Laxmikant S Dhamande and Mangesh B Chaudhari. "Bearing fault diagnosis based on statistical feature extraction in time and frequency domain and neural network". In: *International Journal of Vehicle Structures & Systems* 8.4 (2016), p. 229.

[40] De-Shuang Huang. "A constructive approach for finding arbitrary roots of polynomials by neural networks". In: *IEEE Transactions on Neural Networks* 15.2 (2004), pp. 477–491.

[41] Yongbo Li et al. "A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree". In: *Measurement* 77 (2016), pp. 80–94.

[42] Kenneth A Loparo. "Bearing data center". In: *Case Western Reserve University* (2013).

[43] Li Deng. "A tutorial survey of architectures, algorithms, and applications for deep learning". In: *APSIPA Transactions on Signal and Information Processing* 3 (2014).

[44] Yoshua Bengio et al. "Greedy layer-wise training of deep networks". In: *Advances in neural information processing systems*. 2007, pp. 153–160.

[45] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. "A fast learning algorithm for deep belief nets". In: *Neural computation* 18.7 (2006), pp. 1527–1554.

[46] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.

[47] Christian Szegedy et al. "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning." In: *AAAI*. 2017, pp. 4278–4284.

[48] Tomas Mikolov et al. "Recurrent neural network based language model." In: *Interspeech*. Vol. 2. 2010, p. 3.

[49] Ronan Collobert and Jason Weston. "A unified architecture for natural language processing: Deep neural networks with multitask learning". In: *Proceedings of the 25th international conference on Machine learning*. ACM. 2008, pp. 160–167.

[50] Hayit Greenspan, Bram van Ginneken, and Ronald M Summers. "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique". In: *IEEE Transactions on Medical Imaging* 35.5 (2016), pp. 1153–1159.

[51] Yan Xu et al. "Deep learning of feature representation with multiple instance learning for medical image analysis". In: *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE. 2014, pp. 1626–1630.

[52] Rui Zhao et al. "Deep learning and its applications to machine health monitoring". In: *Mechanical Systems and Signal Processing* 115 (2019), pp. 213–237.

[53] Prasanna Tamilselvan and Pingfeng Wang. "Failure diagnosis using deep belief learning based health state classification". In: *Reliability Engineering & System Safety* 115 (2013), pp. 124–135.

[54] Li Deng, Dong Yu, et al. "Deep learning: methods and applications". In: *Foundations and Trends® in Signal Processing* 7.3–4 (2014), pp. 197–387.

[55] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *nature* 521.7553 (2015), p. 436.

[56] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. "Learning representations by back-propagating errors". In: *nature* 323.6088 (1986), p. 533.

[57] Pascal Vincent et al. "Extracting and composing robust features with denoising autoencoders". In: *Proceedings of the 25th international conference on Machine learning*. ACM. 2008, pp. 1096–1103.

[58] Paul Smolensky. *Information processing in dynamical systems: Foundations of harmony theory*. Tech. rep. COLORADO UNIV AT BOULDER DEPT OF COMPUTER SCIENCE, 1986.

[59] Geoffrey Hinton. "A practical guide to training restricted Boltzmann machines". In: *Momentum* 9.1 (2010), p. 926.

[60] Geoffrey E Hinton. "Deep belief networks". In: *Scholarpedia* 4.5 (2009), p. 5947.

[61] Ruslan Salakhutdinov and Geoffrey Hinton. "Deep boltzmann machines". In: *Artificial Intelligence and Statistics*. 2009, pp. 448–455.

[62] Ruslan Salakhutdinov and Hugo Larochelle. "Efficient learning of deep Boltzmann machines". In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. 2010, pp. 693–700.

[63] Jürgen Schmidhuber. "Deep learning in neural networks: An overview". In: *Neural networks* 61 (2015), pp. 85–117.

[64] Feng Jia et al. "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data". In: *Mechanical Systems and Signal Processing* 72 (2016), pp. 303–315.

[65] Zhiqiang Chen et al. "Deep neural networks-based rolling bearing fault diagnosis". In: *Microelectronics Reliability* (2017).

[66] Levent Eren. "Bearing Fault Detection by One-Dimensional Convolutional Neural Networks". In: *Mathematical Problems in Engineering* 2017 (2017).

[67] Son Lam Phung and Abdesselam Bouzerdoum. "Matlab library for convolutional neural networks". In: *ICT Research Institute, Visual and Audio Signal Processing Laboratory, University of Wollongong, Tech. Rep.* (2009).

[68] Xiaoxi Ding and Qingbo He. "Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis". In: *IEEE Transactions on Instrumentation and Measurement* (2017).

[69] Dinh Nguyen et al. "Highly reliable state monitoring system for induction motors using dominant features in a two-dimension vibration signal". In: *New Review of Hypermedia and Multimedia* 19.3-4 (2013), pp. 248–258.

[70] D Kinga and J Ba Adam. "A method for stochastic optimization". In: *International Conference on Learning Representations (ICLR)*. 2015.

[71] Luyang Jing et al. "A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox". In: *Measurement* 111 (2017), pp. 1–10.

[72] Fardin Dalvand et al. "Current Noise Cancellation for Bearing Fault Diagnosis Using Time Shifting". In: *IEEE Transactions on Industrial Electronics* 64.10 (2017), pp. 8138–8147.

[73] Valeria Cristina Maria Nascimento Leite et al. "Bearing fault detection in induction machine using squared envelope analysis of stator current". In: *Bearing Technology*. InTech, 2017.

[74] Elhoussin Elbouchikhi et al. "Motor Current Signal Analysis Based on a Matched Subspace Detector". In: *IEEE Transactions on Instrumentation and Measurement* 66.12 (2017), pp. 3260–3270.

[75] BR 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.

[76] GS Maruthi and Vishwanath Hegde. "Application of MEMS accelerometer for detection and diagnosis of multiple faults in the roller element bearings of three phase induction motor". In: *IEEE Sensors Journal* 16.1 (2016), pp. 145–152.

[77] Jing Tian et al. "Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with K-nearest neighbor distance analysis". In: *IEEE Transactions on Industrial Electronics* 63.3 (2016), pp. 1793–1803.

[78] Christelle Piantsop Mbo'o and Kay Hameyer. "Fault diagnosis of bearing damage by means of the linear discriminant analysis of stator current features from the frequency selection". In: *IEEE Transactions on Industry Applications* 52.5 (2016), pp. 3861–3868.

[79] Thomas Herold, Christelle Piantsop Mbo'o, and Kay Hameyer. "Evaluation of the use of an electrical drive as a sensor for the detection of bearing damage". In: *AIA-DAGA Conference on Acoustics*. 2013.

[80] Georgios Karatzinis, Yiannis S Boutalis, and Yannis L Karnavas. "Motor Fault Detection and Diagnosis Using Fuzzy Cognitive Networks with Functional Weights". In: *2018 26th Mediterranean Conference on Control and Automation (MED)*. IEEE. 2018, pp. 709–714.

[81] Jiedi Sun, Changhong Yan, and Jiangtao Wen. "Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning". In: *IEEE Transactions on Instrumentation and Measurement* 67.1 (2018), pp. 185–195.

[82] Haidong Shao et al. "Electric locomotive bearing fault diagnosis using novel convolutional deep belief network". In: *IEEE Trans. Ind. Electron* 65.3 (2018), pp. 2727–2736.

[83] Zhuyun Chen and Weihua Li. "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network". In: *IEEE Transactions on Instrumentation and Measurement* 66.7 (2017), pp. 1693–1702.

[84] Pek Hui Foo and Gee Wah Ng. "High-level information fusion: An overview." In: *J. Adv. Inf. Fusion* 8.1 (2013), pp. 33–72.

[85] Robert B Randall and Jerome Antoni. "Rolling element bearing diagnostics—A tutorial". In: *Mechanical systems and signal processing* 25.2 (2011), pp. 485–520.

[86] Laurens van der Maaten and Geoffrey Hinton. "Visualizing data using t-SNE". In: *Journal of machine learning research* 9.Nov (2008), pp. 2579–2605.

[87] Long Wen et al. "A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method". In: *IEEE Transactions on Industrial Electronics* 65.7 (2018), pp. 5990–5998.

[88] Thomas W Rauber, Francisco de Assis Boldt, and Flávio Miguel Varejão. "Heterogeneous feature models and feature selection applied to bearing fault diagnosis". In: *IEEE Transactions on Industrial Electronics* 62.1 (2015), pp. 637–646.

[89] Vicente J Bolós and Rafael Benítez. "The wavelet scalogram in the study of time series". In: *Advances in Differential Equations and Applications*. Springer, 2014, pp. 147–154.

[90] Yann LeCun et al. "Object recognition with gradient-based learning". In: *Shape, contour and grouping in computer vision*. Springer, 1999, pp. 319–345.

[91] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014).

[92] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[93] Gao Huang et al. "Densely Connected Convolutional Networks." In: *CVPR*. Vol. 1. 2. 2017, p. 3.

[94] Sergey Ioffe and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift". In: *arXiv preprint arXiv:1502.03167* (2015).

[95] Yann LeCun et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

[96] Forrest N Iandola et al. "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size". In: *arXiv preprint arXiv:1602.07360* (2016).

[97] Andrej Karpathy et al. "Cs231n convolutional neural networks for visual recognition". In: *Neural networks* 1 (2016).

[98] Jia Deng et al. "Imagenet: A large-scale hierarchical image database". In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255.

[99] David Verstraete et al. "Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings". In: *Shock and Vibration* 2017 (2017).