

**MULTI-SCALE AND MULTI-RATE NEURAL NETWORKS FOR
INTELLIGENT BEARING FAULT DIAGNOSIS SYSTEM**

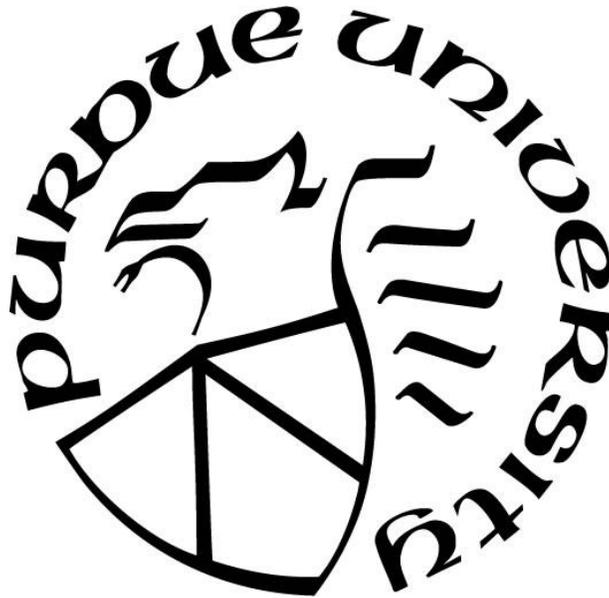
by
Xiaofan Liu

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science in Electrical and Computer Engineering



Department of Electrical and Computer Engineering

Hammond, Indiana

December 2022

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Lizhe Tan, Chair

Department of Electrical and Computer Engineering

Dr. Quamar Niyaz

Department of Electrical and Computer Engineering

Dr. Khair AI Shamaileh

Department of Electrical and Computer Engineering

Approved by:

Dr. Lizhe Tan

ACKNOWLEDGMENTS

I would like to thank the chair of the committee, Professor Lizhe Tan, who has provided crucial theoretical and practical guidance to my research over the course of a year and half. Under his guidance, I have further strengthened my research skills and theoretical foundation. I would also like to thank committee members Professor Quamar Niyaz and Professor Khair Al Shamaileh for their patient guidance and assistance over the recent years.

A part of this thesis has been previously published in IEEE Access, vol. 10, pp. 86998-87007, Aug. 2022, and the 2022 IEEE International Conference on Electro Information Technology (EIT), pp. 80-85, May 2022. Please see the section 8 publications for more information.

TABLE OF CONTENTS

LIST OF TABLES	5
LIST OF FIGURES	6
ABSTRACT	7
1. INTRODUCTION	8
1.1 Literature Review.....	8
1.2 Research and Motivation	10
1.3 Contribution of Thesis	11
1.4 Organization of Thesis	12
2. BEARING FAULT DATASET	13
2.1 Case Western Reserve University Bearing Dataset	13
2.2 Self-collect Dataset	16
3. SIGNAL FEATURE EXTRACTION	18
3.1 Signal Preprocessing	18
3.2 Signal Processing Methods	19
3.2.1 Downsample Scheme.....	19
3.2.2 Empirical Mode Decomposition.....	19
3.2.3 Wavelet Packet Transform	20
3.2.4 Sparse Wavelet Packet Transform.....	21
4. SIGNAL FEATURE LEARNING	29
4.1 One-Dimensional Convolutional Neural Network	29
4.2 Residual Neural Network.....	30
4.3 Gated Recurrent Unit	32
4.4 Multi-Scale and Multi-Rate Convolutional Neural Network.....	33
5. COMPARISON OF PERFORMANCE	35
5.1 Data Mining for Classification.....	37
6. CONCLUSION AND FUTURE WORK	39
REFERENCES	40
PUBLICATIONS.....	45

LIST OF TABLES

Table 2-1 Details of the data selected from the CWRU dataset	15
Table 2-2 Details of the bearing parameters and self-collected dataset.....	17
Table 3-1 Bearing defect frequencies of bearing fault.....	23
Table 3-2 Characteristic frequency bands for training process	23
Table 3-3 Complete wavelet decomposition.....	25
Table 4-1 Information of layers in 1DCNN-based multi-scale network	30
Table 4-2 Information of layers in 1D-ResNet-34.....	31
Table 4-3 Information of layers in GRU-based multi-scale network	33
Table 4-4 Information of layers in multi-scale and multi-rate convolutional neural network.....	34
Table 5-1 Accuracy and std of different methods.....	35
Table 5-2 F1-score and recall of different methods.....	35
Table 5-3 Comparison with other published methods	37

LIST OF FIGURES

Figure 2-1 The experimental platform for the CWRU dataset	13
Figure 2-2 Location of different parts for a typical bearing	14
Figure 2-3 Typical acquired vibration signals for CWRU dataset	15
Figure 2-4 The test platform of self-collected dataset	16
Figure 2-5 The location and shape of the bearing defect.....	16
Figure 2-6 Typical acquired vibration signals for various bearing faults.....	17
Figure 3-1 Two methods for data preprocessing	18
Figure 3-2 The principle of down sampled processing.....	19
Figure 3-3 Wavelet packet selection.....	20
Figure 3-4 The structure of a typical bearing.....	21
Figure 3-5 The bearing characteristic frequency zones	22
Figure 3-6 The subband tree structuring using WPT	24
Figure 3-7 WPT decomposition path using gray codes	25
Figure 3-8 Simplification process in the coding table and subband tree	27
Figure 3-9 Processing raw signal segments using optimized route coding	28
Figure 4-1 Multi-scale network based on 1D-CNN.....	29
Figure 4-2 Multi-scale network based on 1D-ResNet-34.....	31
Figure 4-3 Multi-scale network based on GRU	32
Figure 4-4 Multi-Scale and Multi-Rate Convolutional Neural Network.....	34
Figure 5-1 Loss function curves and Confusion matrix for 10 independent trainings	36
Figure 5-2 The visualization result using t-SNE.....	38

ABSTRACT

Roller bearing is one of the machine industry's common components. The roller bearing operation status is usually related to production efficiency. Failure of bearings during operation will cause downtime and severe economic losses. To prevent this situation, the proposal of effective bearing fault diagnosis methods has become a popular research topic. This thesis research first validates several popular bearing diagnosis methods based on signal processing and machine learning. Second, a novel signal feature extraction method called sparse wavelet packet transform (WPT) decomposition and a corresponding feature learning model called multi-scale and multi-rate convolutional neural network (MSMR-CNN) are proposed. Finally, the proposed method is verified using both Case Western Reserve University (CWRU) dataset and the self-collected dataset. The results demonstrate that our proposed MSMR-CNN method achieves higher performance of bearing fault classification accuracy in comparison with the methods which are recently proposed by the other researchers using machine learning and neural networks.

1. INTRODUCTION

1.1 Literature Review

Based on a large number of studies on the causes of production accidents, many research publications attribute the main cause of rotating machinery failures to the damage of bearing components [1]-[5]. If the corresponding characteristic signals can be accurately captured and identified at the early stage of bearing component damage, and the damaged bearings can be replaced, serious economic losses due to accidents can be greatly avoided [2] [6]-[9]. In this regard, a variety of bearing defect detection methods have been proposed by different researchers. In general, these methods can be classified as conventional signal analysis-based [10]-[25], machine-learning-based [26]-[30], and deep learning-based [31]-[38] methods.

First, for the methods based on conventional signal analysis. These methods tend to collect regular periodic vibration signals generated by the bearings during rotation and use some common signal processing methods to extract signal features and later determine whether the bearings are defective by human-specified rules. Signal processing methods can also be broadly classified into the time domain, frequency domain, and time-frequency domain-based analysis. For example, these signal processing methods include, but are not limited to, fast Fourier transform (FFT), wavelet transform (WT), empirical mode decomposition (EMD), and wavelet packet transform (WPT) [14], [15]-[18]. Among them, time domain analysis determines the bearing condition by capturing a specific scalar feature of the vibration signal located distinctly in the time domain [19]. These scalar features include, but are not limited to, signal peaks, root mean square (RMS), crest factor, kurtosis, spectral kurtosis, and gap factor [20]-[23]. However, for some nonsmoothed signals, time-domain-based signal analysis is difficult to find suitable scalar features for judgment among the noisy signal features. Therefore, some researchers have proposed using frequency domain analysis to extract signal features. For example, M. Cocconcelli et al. [24] used the short-time Fourier transform (STFT) to process the raw vibration signal. They analyzed the transformation results in the frequency domain for bearing defect detection. J.-H. Lee et al. [25] used the Wigner-Ville distribution in their study to perform frequency domain analysis of vibration signals for mechanical fault detection. However, the methods mentioned above based on signal analysis are only effective in extracting the distinct features of the bearing due to faults. The signal

analysis-based methods often need to identify inconspicuous features resulting from early failures. In addition, the signal features resulting from damage to bearing components are usually not limited to scalar changes in the time or frequency domain. Some of the relevant features are often difficult to identify artificially. This means that some data analysis methods and automated means are needed to enable the computer to spontaneously summarize potential signal features associated with bearing defects from the data.

To enable computers to spontaneously extract signal features associated with bearing failures from the data, some researchers have proposed that machine learning can directly or indirectly learn the vibration signal features generated during bearing operation. For example, M.-Y. Chow et al. [26] used artificial neural networks (ANNs) to detect early mechanical faults occurring in the rotors of small and medium-sized motors. A. Malhi and R. X. Gao [27] used principal component analysis (PCA) to automatically extract the vibration signal features most relevant to the defect state and classify these features using feedforward neural networks (FFNs). Y. S. Wang et al. [28] proposed an intelligent engine fault diagnosis method by using Hilbert-Huang transform (HHT) with a support vector machine (SVM) algorithm. Similarly, D. H. Pandya et al. [29] used HHT to process bearings' acoustic signals, with the difference that the K-nearest neighbor algorithm (KNN) was used to determine the bearing defects. F. Shen et al. [30] used a transfer learning-based singular value decomposition (SVD) model to diagnose bearing defects. Although the machine-learning-based bearing defect detection method succeeded in automatically learning the relationship between relevant features from vibration signals and bearing operating conditions. However, the learning capability of the method is still limited. This limits the continued improvement in the performance of this type of method.

The use of deep learning models is a major solution to the problem of insufficient learning capability of traditional machine learning models. Deep learning-based methods for bearing defect detection can be further divided into end-to-end methods that learn directly using the original signal and methods that learn the signal features after processing by signal processing methods. For the former, some researchers, such as M. Zhao et al. [31], proposed a novel deep neural network structure, called deep residual reduction network, to improve fault diagnosis accuracy. However, the end-to-end learning model may learn some signal features in the raw data that are not strongly correlated with the bearing operating state. These poorly correlated features may cause the trained deep neural network to identify only specific bearing fault types effectively. To address

this shortcoming, pre-processing raw signal using some signal processing methods becomes an effective approach. Our research group has conducted a long-term study in this area. In previous work [32], we improved classification results based on a public dataset using input features generated by STFT and employing a convolutional neural network (CNN) structure optimized by a particle swarm optimization (PSO) algorithm. Similarly, a combination of STFT and CNN with tunable input size was explored in [33] for bearing fault diagnosis. In addition, our early research [34] reported evaluated the classification performance using the conventional deep learning neural networks, such as CNN and long short-term memory (LSTM) by using various signal transforms, including STFT, Cepstrum, WPT, and EMD, to generate input features. Despite the good classification results achieved by the above work, the signal processing methods mentioned above [32]-[34] are performed at a single scale level. To further increase the effective features that deep learning models can learn, multiscale features are introduced to improve the training of the models further. For example, W. Chen and K. Shi [35] used an end-to-end multiscale convolutional neural network (MSCNN) to classify the types of bearing defects. In their approach, the original, smoothed and downsampled signals constitute the network inputs at three different scales. G. Jiang et al. [36] also used MSCNNs to study the health status detection of wind turbine gearboxes. In fact, in [36], MSCNN achieved good detection results by using only downsampled signals at different scales as inputs. Furthermore, Yao et al. [37] used a one-dimensional convolutional layer to decompose the original signal into multiscale features that operate in an end-to-end format for improved practicality. In addition, their method introduces an attention mechanism to improve the final detection. Our research team noted the potential of multi-scale features in improving the accuracy of bearing defect detection and summarized the performance of different deep learning models in handling multi-scale features as reported in [38].

1.2 Research and Motivation

By revisiting the methods proposed in recent years in the literature, it can be found that signal processing combined with deep learning is well-established in bearing fault detection methods. The excellent performance of these methods on multiple datasets proves their effectiveness. However, these methods only use single-scale or simple multi-scale signals to complete the training of deep learning models. Our work in [32], [39]-[41] demonstrated that the signal features in bearings due to defects are distributed in multiple discontinuous frequency regions. These

frequency regions are designated as low-frequency region, bearing defect frequency region, inherent bearing frequency region, and high-frequency region, which can be determined by shaft speed and bearing physical parameters. It is noted that these frequency regions are very narrow and sparse with respect to the signal sampling frequency in the data acquisition system. Based on our previous work in speech recognition [42]-[44], we propose the basic framework of sparse wavelet packet transform (WPT) and use this method to extract the target frequency region from the original signal in a targeted manner and construct it as a multi-resolution feature to be applied to a neural network for learning. In this thesis research, a part of which is also published in [45], we further improved the sparse WPT architecture so that the extracted features have multi-scale and multi-rate features. The excellent performance of the deep learning model trained with these features on the CWRU dataset and self-collected dataset further validates the superiority of our proposed method over the methods published by other research groups.

1.3 Contribution of Thesis

The research work in thesis has the following contributions:

- (1) A framework using signal transform for multi-scale 1-D convolutional neural network (CNN) is proposed. The best performance on classification accuracy using discrete wavelet packet decomposition is validated.
- (2) A framework using multi-resolution signal features in the bearing characteristic frequency bands and multi-scale 1-D CNN is proposed.
- (3) An improved sparse WPT decomposition method for extracting signals in the sparse multi-frequency bands is proposed. This method first uses gray codes to generate all the paths required for completed signal decomposition for each frequency band. Then a sparse decomposition path is found based on the lowest cost of the decomposition operations.
- (4) The proposed method is validated using CWRU dataset and self-collected dataset from our designed experiments.

1.4 Organization of Thesis

This thesis work is organized as follows. Chapter 2 introduces the concept of the rolling bearing fault and the dataset used in the experiment. Chapter 3 describes the signal preprocessing method, five signal processing methods, and data mining for classification. Chapter 4 applies the commonly used deep learning algorithm for signal feature learning. Chapter 5 compares the performances of the proposed methods on the CWRU and self-collect datasets. Chapter 6 presents the conclusions and future work.

2. BEARING FAULT DATASET

When training machine learning models, sufficient training data is an essential prerequisite. For supervised training, it is even more important that all training data have the correct classification labels. Generally, it is often more expensive and time-consuming to collect and label data than design and train models. This often limits the speed of research in machine learning. However, the development of public datasets in recent years has provided new ideas for developing the machine learning research field. Any research group can test the proposed method on a public dataset with high confidence. Therefore, the research work covered in this paper focuses on training and validating the deep learning model of the proposed method using the Case Western Reserve University Bearing Dataset and the self-collected dataset. In the following sections, these two datasets are described in detail.

2.1 Case Western Reserve University Bearing Dataset

The CWRU dataset [46], one of the few most recognized public datasets in the field of bearing defect detection, has become very popular among various research groups in recent years. The data from this dataset which was collected from the experimental platform is shown in Figure 2-1.

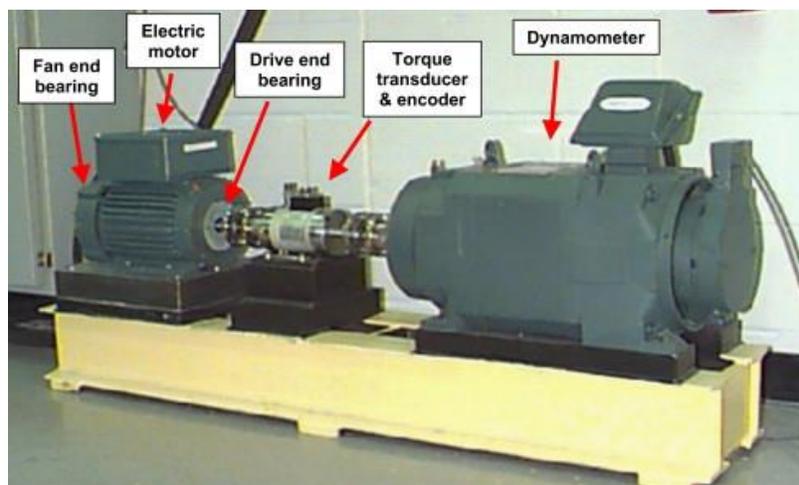


Figure 2-1 The experimental platform for the CWRU dataset

The platform uses an electric motor to drive a dynamometer and a bearing connection between the two. The platform has two main data collection points, one at the Fan end bearing and the other at the Drive end bearing. The data are collected by vibration sensors. All faulty bearings are artificially created using EDM to create single-point defects of various depths at different locations. Figure 2-2 shows the distribution of these three locations in a typical bearing.

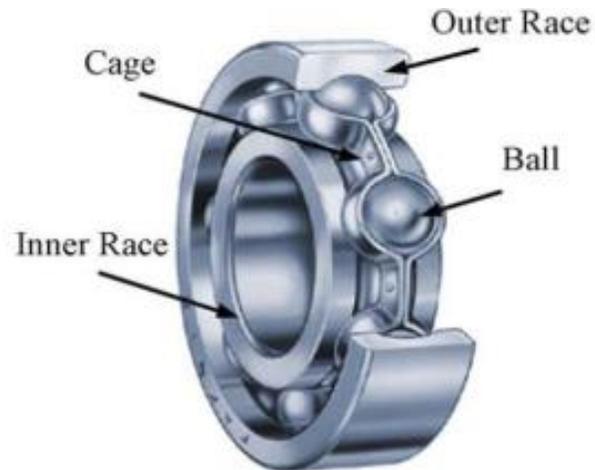


Figure 2-2 Location of different parts for a typical bearing

In our work, we selected training data samples with fault depths of 0 (normal), 0.07, 0.014, and 0.021 inches at the drive end. The reason for not choosing samples with a depth of 0.028 inches in the data set is that the specific parameters of the bearing cannot be determined for this type of sample. The above samples were driven by an electric motor with a speed of 1772 RPM. The vibration sensors were positioned at 12 o'clock in each bearing. All samples were sampled at 12 kHz, except for the 0-inch depth of failure (normal) sample, which was sampled at 24 kHz. In summary, the samples selected for this work are shown in Table 2-1 and Figure 2-3.

Table 2-1 Details of the Data Selected from the CWRU Dataset

Fault Depth	Fault Name	Class Name
0.021	Ball	No. 1 (21-Ball)
	Inner Race	No. 2 (21-IR)
	Outer Race 12	No. 3 (21-OR)
0.014	Outer Race 6	
	Outer Race 3	
	Ball	No. 4 (14-Ball)
0.07	Inner Race	No. 5 (14-IR)
	Outer Race 6	No. 6 (14-OR)
	Ball	No. 7 (07-Ball)
0	Inner Race	No. 8 (07-IR)
	Outer Race 3	No. 9 (07-OR)
	Outer Race 6	
	Outer Race 12	
0	Normal	No. 0 (Normal)

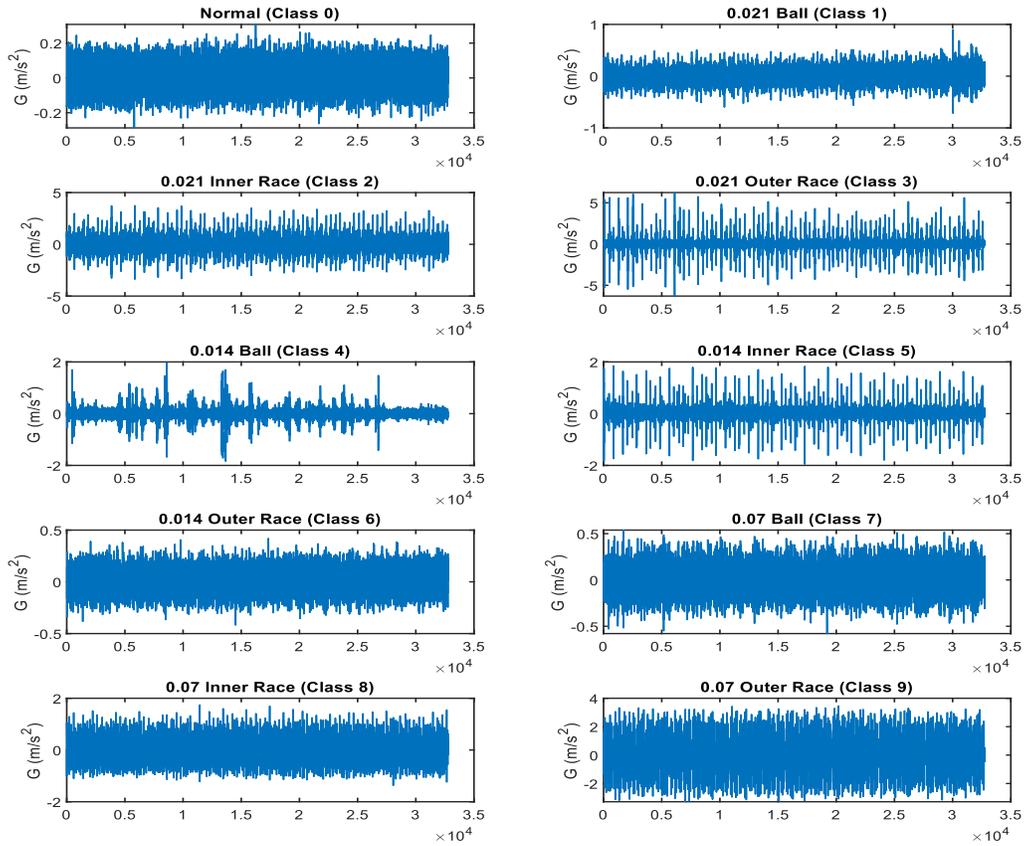


Figure 2-3 Typical acquired vibration signals for CWRU dataset

2.2 Self-collect Dataset

In addition to the public data set we also collected and labeled a bearing failure data set on our own bearing test platform to increase the data diversity. As shown in Figure 2-4, the data acquisition platform we built consists of a modified press borer (Roybi), joint assembly, bearing unit, a vibration sensor (CTC AC 102-1A accelerometer, 100 mV/G, $\pm 10\%$), signal condition circuit (homemade PCB circuit), data acquisition board (NI USB-6356), and a desktop workstation. We mounted a vibration sensor at the 6 o'clock position of the bearing and acquired data at 12 kHz. The bearing will be driven by the rig at 630 rpm.

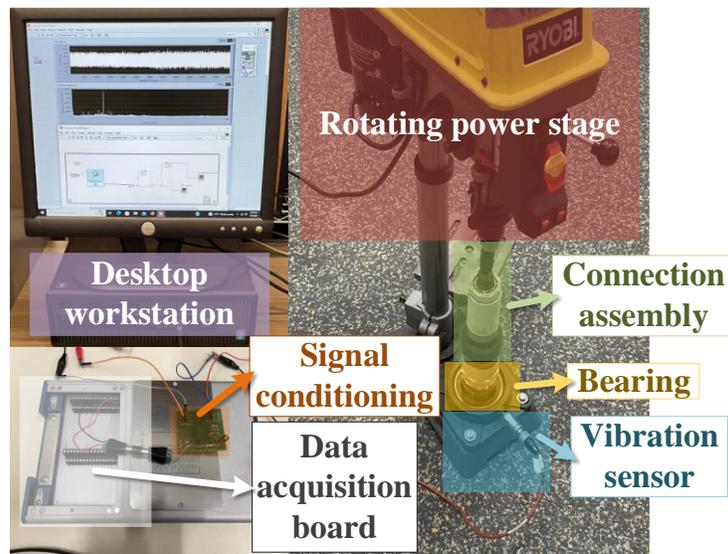


Figure 2-4 The test platform of self-collected dataset

We artificially created defects on the inside of the outer ring, outside of the inner ring and ball surface of each bearing. The locations of these defects are shown in Figure 2-5.



Figure 2-5 The location and shape of the bearing defect

After collecting data segments of 10 seconds in length, we organized and labeled these data. Table 2-2 shows the detailed parameters of the bearings we used and the data we collected. Typical plots of the vibration signals for each class from our own dataset are displayed in Figure 2-6.

Table 2-2 Details of the Bearing Parameters and Self-Collected Dataset

Bearing parameters		Fault location	Class name
Ball diameter	13.838 mm	Ball	No.1
Pitch diameter	75.4 mm	Inner Race	No.2
Rotor speed	630 rpm	Outer Race	No.3
Sampling rate	12000 Hz	Normal	No.0

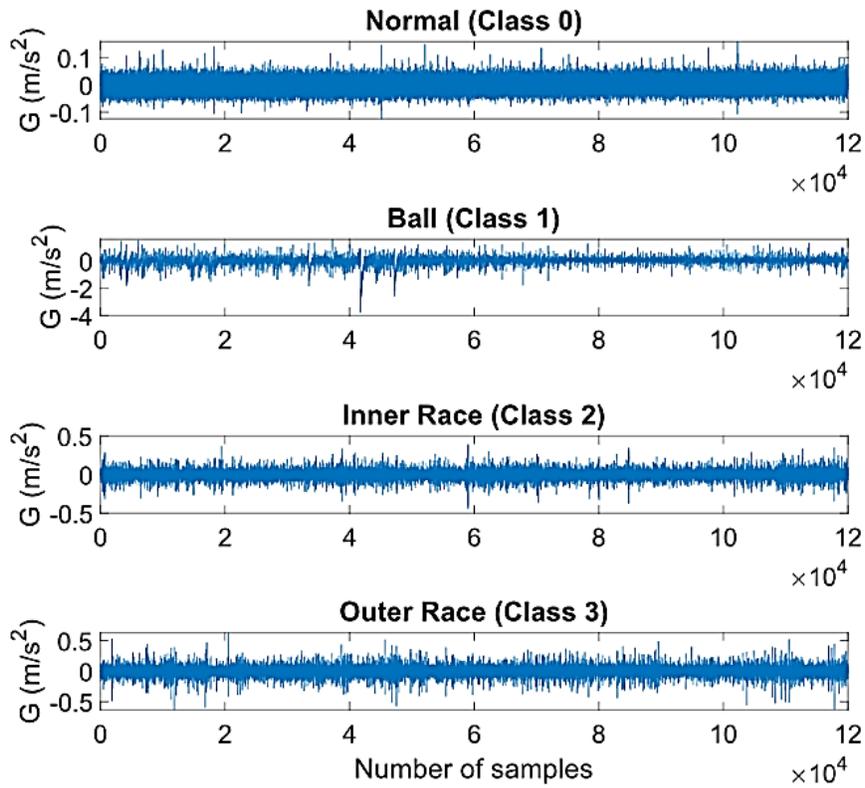


Figure 2-6 Typical acquired vibration signals for self-collected dataset

3. SIGNAL FEATURE EXTRACTION

3.1 Signal Preprocessing

Before applying signal processing method, the data first needs to be pre-processed. The preprocessing steps include down sampling the signal and randomly intercepting segments of the signal. The reason for down sampling is that the signals are sampled at 12 kHz, except for the normal bearing operation, which is sampled at 24 kHz. The inconsistency of the sampling frequency may lead to deviations in the signal processing results and affect the final results of feature learning later. Therefore, we downsampled the normal bearing samples from the sample rate of 24 kHz to 12 kHz to keep the total frequency domain of all the sample rate consistent. Then, we use two pre-processing methods to obtain the signal segmentation. The first method is to acquire data segments one by one from the first data point of the raw signal via sliding windows [38]. The second method is to obtain data segments at random locations in the raw signal [45]. The method of random segmentation is feasible for generating a compromised number of training samples over the entire raw signal range. The processing details of these two methods are shown in Figure 3-1.

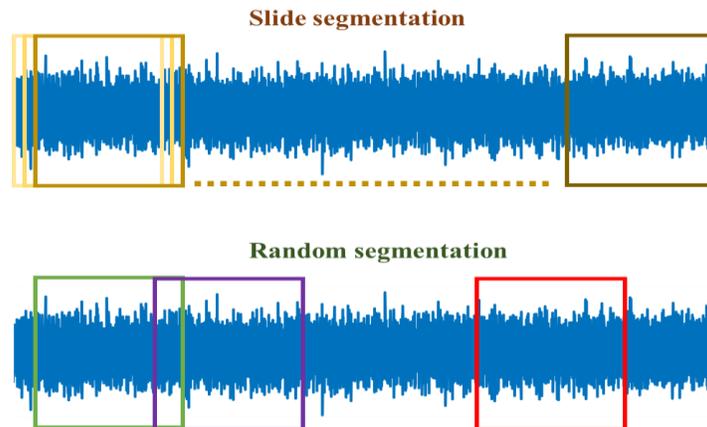


Figure 3-1 Two methods for data preprocessing

3.2 Signal Processing Methods

3.2.1 Downsample Scheme

In order to obtain multi-scale features, several signal processing methods are studied. The first method is the down sample scheme (DS), which averages N adjacent signals to produce a new sequence of signals arranged by averaging. The process can be expressed as follows:

$$y_j^{(n)} = \frac{(x_i + x_{i+n})}{n} \quad (1)$$

where n determines how many neighboring signals to average between. i is the original signal length. j is the signal length after down sampled processing. The relationship between the above three is $j = i/n$. This process results in several multiscale signals with different overall details. Figure 3-2 shows the details of this method.

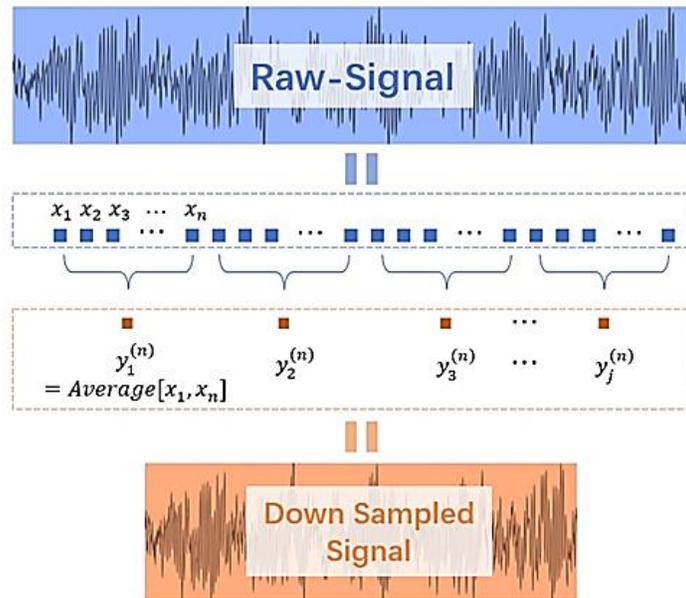


Figure 3-2 The principle of down sampled processing

3.2.2 Empirical Mode Decomposition

Empirical mode decomposition [47] allows the gradual decomposition of nonlinear and non-smooth signals into a finite number of eigenmode functions (IMFs) and residual signals. The process can be expressed by the following equation:

$$I(n) = \sum_{m=1}^M IMF_m(n) + Res_M(n) \quad (2)$$

where $I(n)$ denotes the input signal. $IMF_m(n)$ denotes the m -th intrinsic mode function. $Res_M(n)$ denotes the residual signal. Notice that the IMF is obtained by repeated iterations. After the EMD decomposition of the signal, the obtained multiple IMFs together with a single residue signal are adopted to form a multiscale signal as the input to the feature learning model.

3.2.3 Wavelet Packet Transform

Wavelet packet transform [48] is a special form of wavelet variation. It uses multiple low-pass and high-pass filters to decompose the high and low frequency components of the original signal. The decomposition process is shown in Figure 3-3.

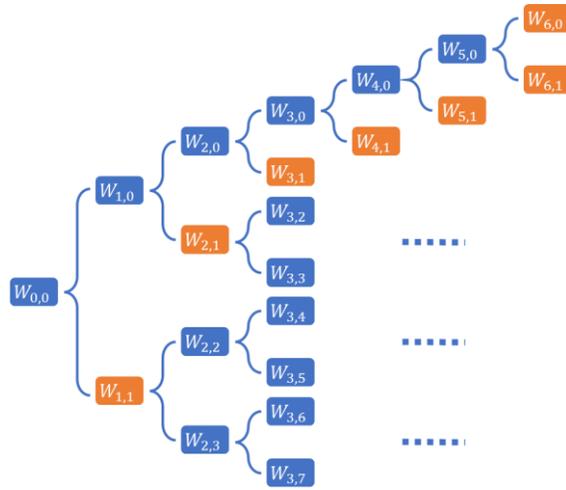


Figure 3-3 Wavelet packet selection

The wavelet packets of $W_{j,m}(n)$, $j = 0,1,2, \dots$ and $m < 2^{j-1}$, are given as:

$$W_{j,2m}(n) = \sqrt{2} \sum_{k=0}^{2N-1} h(k)W_{j-1,m}(2n - k) \quad (3)$$

$$W_{j,2m+1}(n) = \sqrt{2} \sum_{k=0}^{2N-1} g(k)W_{j-1,m}(2n - k) \quad (4)$$

where m is designated as the decomposition level. Equations (3) and (4) represent the low and high frequency components wavelet packet respectively. After the successive decomposition of wavelet packets level by level, we keep the high frequency part of each level of decomposition

together with the low frequency part of the final level to form a multi-scale signal. In this way, we are able to obtain a multiscale feature covering all frequency domains. Note that unlike conventional multiscale features that tend to be built on multiscale in the time domain, the multiscale obtained using WPT is built on multiscale in the frequency domain. This approach enables a more effective integration of signal features in different frequency domains allowing better training of the feature learning model. However, since the frequency regions affected by bearing defects are often located in several sparse and narrow regions, a method is needed to extract features only for specific frequency regions. A novel class of our proposed multi-scale signal feature extraction methods will be presented in the following section.

3.2.4 Sparse Wavelet Packet Transform

For a typical bearing as shown in Figure 3-4, if the parameters of Pitch Diameter (PD) and Ball Diameter (BD) are given, the frequency region in which the signal characteristics of the bearing caused by defects are located can be obtained from Equation (1)-(4).

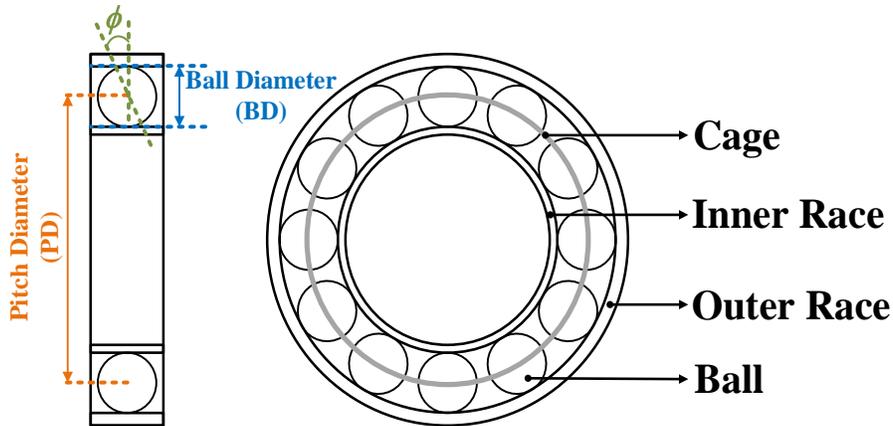


Figure 3-4 The structure of a typical bearing

$$f_{OD} = \frac{n}{2} f_{rm} \left(1 - \frac{BD}{PD} \cos \phi \right) \quad (5)$$

$$f_{ID} = \frac{n}{2} f_{rm} \left(1 + \frac{BD}{PD} \cos \phi \right) \quad (6)$$

$$f_{BD} = \frac{PD}{2BD} f_{rm} \left(1 - \left(\frac{BD}{PD} \right)^2 \cos^2 \phi \right) \quad (7)$$

$$f_{cage} = \frac{1}{2} f_{rm} \left(1 - \frac{BD}{PD} \cos \phi \right) \quad (8)$$

where f_{OD} , f_{ID} , f_{BD} , and f_{Cage} are the outer race, inner race, ball, and cage frequencies, respectively. n and f_{rm} are the number of balls and shaft speed, respectively. Note that ϕ is the contact angle between the ball and bearing, which is approximately to be 0~10° [13]. From the above equations, we can derive the bearing defect frequency zone, however, in practice, the frequency change caused by the defect is not only limited to this zone. The effect of the defect on the overall rotation of the bearing will result in a change in the low frequency characteristics. Similarly, when a ball or other component of the bearing comes into contact with the defective surface, it will cause a complex series of shocks that will lead to changes in the high frequency characteristics. Overall, the distribution of these features should be as shown in Figure 3-5.

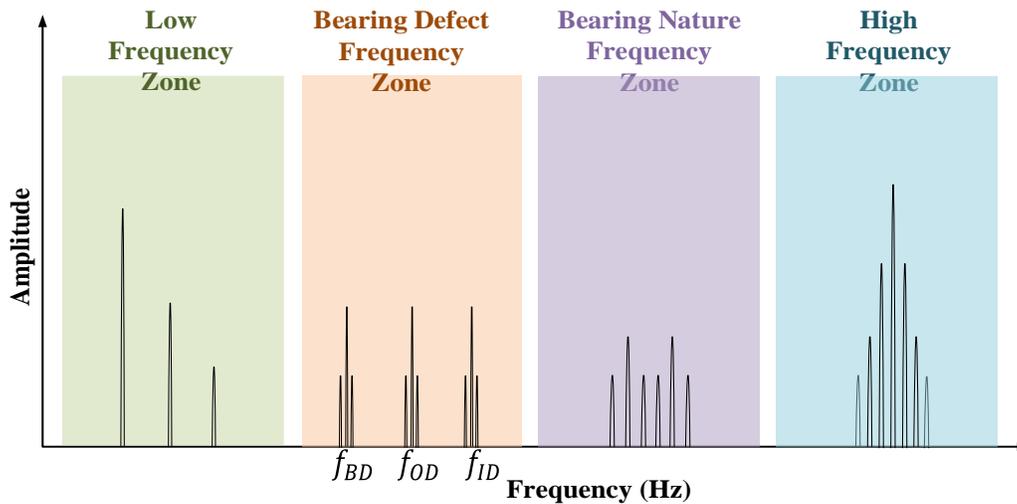


Figure 3-5 The bearing characteristic frequency zones

As an illustration using the physical parameters of test bearing provided by CWRU dataset, the shaft speed f_{rm} is set to 30 Hz (1800 rpm) since a slight variation between 29.93 and 29.95 Hz in the rotational speed of the bearing rolls during data acquisition and the contact angle is used as $\phi \approx 10^\circ$ [13], the corresponding fault frequency values for f_{OD} , f_{ID} , f_{BD} , and f_{Cage} are shown in Table 3-1.

Table 3-1 Bearing Defect Frequencies of Bearing Fault

Name	Fault frequencies (Multiple of shaft speed)	Fault frequencies (Hz)
f_{OD}	3.585	107.55
f_{ID}	5.415	162.45
f_{BD}	2.357	70.71
f_{cage}	0.3983	11.949

From Table 3-1, the band between f_{BD} and f_{ID} is a target band for a single-point fault, which is between 70.71 Hz and 162.45 Hz. However, this band only contain the information of bearing defect frequency zone. We still need to decide other bands to contain the information from other frequency zones. In our research, a frequency band from $0.5f_{rm}$ to f_{rm} , which is 15 Hz to 30 Hz, is considered as the low frequency zone. An additional frequency band from 162.46 Hz to 500 Hz is considered to contain information about the defect characteristics in the high-frequency band. The reason for choosing 500 Hz as the maximum frequency is to prevent the invalid information from the high frequency noise which may affect the validity of the training data. After obtaining the required three frequency bands, each band is expanded by $0.25f_{rm}$ at its band frequency edges in order to ensure the sufficient frequency margins to cover the required frequency range. Three frequency bands and their respective frequency edges for extracting training data samples are listed in Table II.

Table 3-2 Characteristic Frequency Bands for Training Process

Band Name	Frequency Range (Hz)
Band 1	7.50 – 37.50
Band 2	63.21 – 169.95
Band 3	169.95 – 500.00

To extract the corresponding information of the characteristic and relevant frequency bands from the full-band information of the original signal, a complete wavelet packet decomposition (WPT) is first applied. It decomposes the original signal into two sub-bands at low and high frequencies and continues the decomposition for each subband at the next level. Figure 3-6

displays the complete decomposition structure from decomposing a signal with a frequency band of 0~200 Hz up to three levels. However, the complete decomposition structure is not necessary. In fact, only the result of a finite number of decomposition paths is needed in the structure. To illustrate the process, we first need to encode the paths efficiently. Second, we need to find a way to automatically find a list of encodings corresponding to our desired frequency region and subsequently optimize the list to obtain the optimal decomposition path.

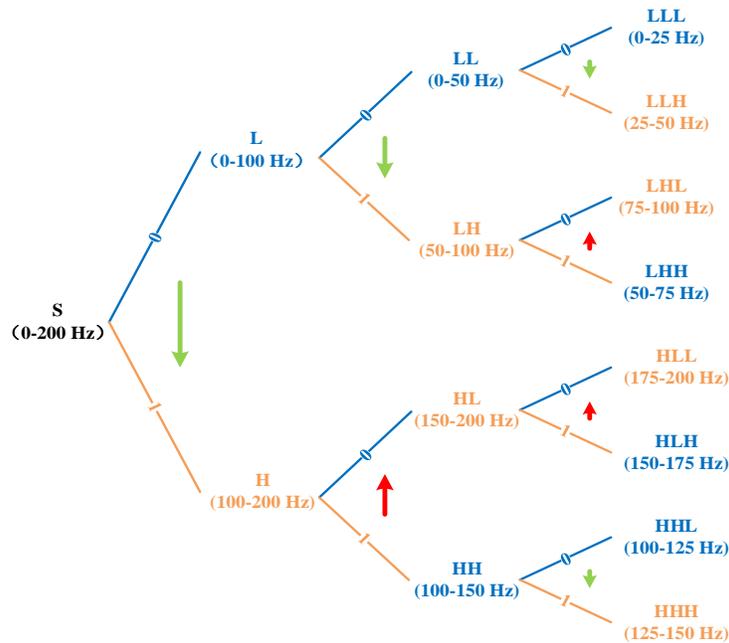


Figure 3-6 The subband tree structuring using WPT

As shown in Figure 3-6, symbol “0” represents the low-frequency decomposition while “L” represents the resulting low-frequency subband. Symbol “1” denotes the high-frequency decomposition while “H” represents the high-frequency subband. The arrows represent the direction of the frequency distribution at the current decomposition level, with the arrows pointing from low to high frequencies. A green arrow indicates no change in the direction of frequency distribution, while a red arrow indicates a direction change of the frequency distribution. According to Figure 3-6, the complete decomposition paths follow gray codes [41]-[43] and can be easily seen in Table 3-3 for a three-level decomposition. Note that the results in Table 3-3 can be extended to n-level decomposition in which 2^n sub-bands can be arranged in an order from low to high frequencies according to n-bit gray codes in the generated order. At the same time,

observing the results of Table 3-3, it is not difficult to find that there are multiple frequency band decomposition paths corresponding to the gray code that actually have the same first decomposition code. This phenomenon actually reflects that multiple decomposition paths actually share the same decomposition path, that is duplicated Gary codes. The proposed decomposition optimizes the decomposition paths by merging and simplifying the duplicated codes.

Table 3-3 Complete Wavelet Decomposition

Frequency Band (Hz)	Gray Code	Subband Name
0 – 25	000	LLL
25 – 50	001	LLH
50 – 75	011	LHH
75 – 100	010	LHL
100 – 125	110	HHL
125 – 150	111	HHH
150 – 175	101	HLH
175 - 200	100	HLL

Using the approach mentioned above, the decomposition paths for the target frequency band can be easily obtained. As an illustration, if a target band has the frequency range from 75 to 125 Hz, the required decomposition paths found from Table 3-3 or Figure 3-6 are 010 and 110, as well as shown in Figure 3-7.

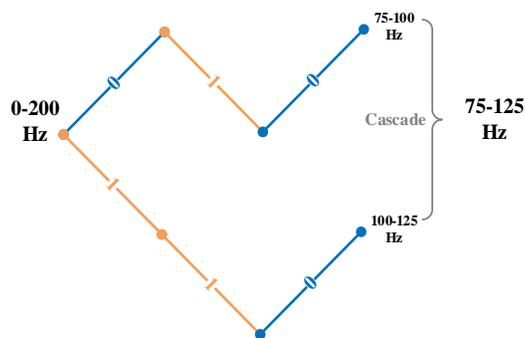


Figure 3-7 WPT decomposition path using gray codes

For bearing fault detection, we can use this approach to extract the multi-resolution features for a given bearing frequency characteristic band. As illustrative example, we can extract features from the band between 7.5 and 37.5 Hz with the given full frequency band from 0 to 6000 Hz at a sampling rate of 12,000 Hz. To obtain the multi-resolution signal features using a sparse WTP decomposition for characteristic frequency bands as listed in Table 3-2, our proposed method is divided into three steps: determining the optimal decomposition level; finding a sparse decomposition path structure; and performing decomposition execution. First, for given the sampling frequency of f_s , and target frequency band ideally ranging from f_L to f_H , the minimum number of decomposition levels N_{\min} is determined by

$$N_{\min} = \log_2 \frac{0.5f_s}{f_H - f_L} \quad (9)$$

Let N_{\max} be the maximum number of decomposition levels:

$$N_{\max} = N_{\min} + L_{\max} \quad (10)$$

where L_{\max} is the number of maximum allowable steps. We can search for the optimal N_{best} so that the following error is minimized:

$$err = \min \{ \max(f_L - f_{L_i}, f_{H_i} - f_H) \} \quad (11)$$

where f_{L_i} and f_{H_i} are the actual lower frequency and the upper frequency edges, respectively. They can be determined by

$$f_{L_i} = \left\lfloor \frac{f_L}{\Delta f} \right\rfloor \Delta f \quad (12)$$

$$f_{H_i} = \left\lceil \frac{f_H}{\Delta f} \right\rceil \Delta f \quad (13)$$

Note that Δf is the band resolution at decomposition level of N , given below:

$$\Delta f = \frac{f_s}{2^N}, \quad N \in [N_{\min}, N_{\max}] \quad (14)$$

Let us define the desired minimum error as

$$err_{\min} = \lambda(f_H - f_L) \quad (15)$$

where the empirical constant λ determines the minimum error as a percentage of the target frequency band. The algorithm for searching the optimal decomposition level N_{best} can be illustrated in the following pseudo-code.

Algorithm about Finding the Best Decomposition Level

Input: N_{min} ; N_{max} ; err_{min}

Output: N_{best}

```

1: for  $N = N_{min} : N_{max}$  do
2:   if  $err < err_{min}$  then
3:      $N_{best} = N$ 
4:   else
5:      $N_{best} = N_{max}$ 
6:   end if
7: end for

```

With N_{best} , the second step is to find a sparse decomposition structure. This can be illustrated using an example shown in Figure 3-8 for the case of extracting band 3. The following parameters are used: $f_s = 6,000$ Hz, $f_L = 169.95$ Hz, $f_H = 500.00$ Hz, $L_{max} = 6$, and $\lambda = 0.15$, the algorithm finds $L_{best} = 7$, $f_{L_t} = 140.63$ Hz and $f_{H_t} = 515.63$ Hz, respectively. Eight 7-bit gray codes covering this frequency band, beginning with 0000010 and ending with 0001111 are listed in Figure 3-8. Again, note that any consecutive two gray codes with least significant bit ending with 1 and 0 can be merged and the corresponding branch can be pruned by one level. Repeating such process leads to four gray codes as shown in Figure 3-8. Eventually, a sparse decomposition structure using the four gray codes is achieved and displayed in Figure 3-8. As can be seen, without the decomposition structure reduction, there are $7 \times 8 = 56$ decomposition operations. However, with the obtained sparse structure, only 13 operations are required.

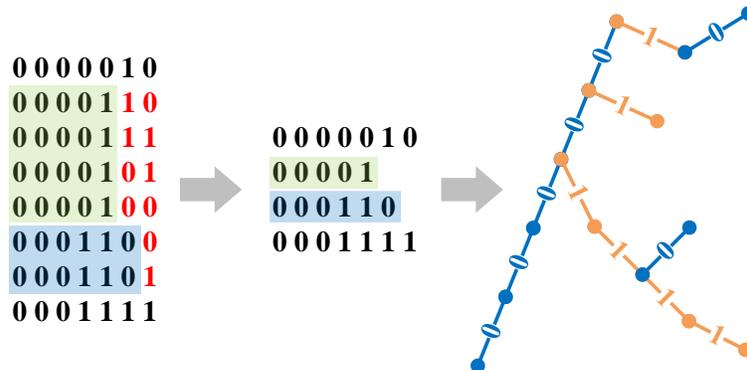


Figure 3-8 Simplification process in the coding table and subband tree

Note that the extracted signal features are in a nature of sparse multi-resolution. The algorithm can be repeated for the other two bands similarly. The combined and simplified final sparse structure for three bands with a further reduced number of decomposition operations is displayed in Figure 3-9. As shown in Figure 3-9, the highest level is 10 and there are 31 decomposition operations in total.

The third step is to perform signal extraction by applying the final sparse decomposition structure. The original raw signal with duration of 10 seconds at 12,000 Hz sampling rate is first pre-processed to have zero mean and then normalized. 512 segments with the same length are retrieved at the random starting locations. The length of each segment is determined to have no less than $2^{N_{best}+3}$ data samples in the segment. The requirement of the minimum length ensures that the original signal retains a data sufficient length to prevent loss of signal characteristics after N_{best} level decomposition. Each segment is decomposed into three frequency bands based on the final sparse structure as shown in Figure 3-9. After decomposition, the multi-resolution signal features from each band forms a feature vector. The feature vectors from all the three frequency bands are used as the input for the signal feature learning module in training and detection processes.

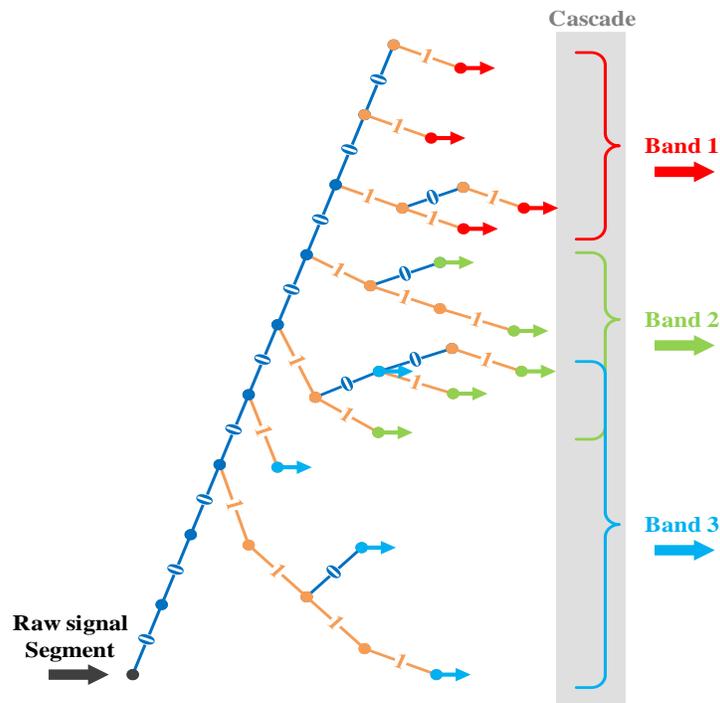


Figure 3-9 Processing raw signal segments using optimized route coding

4. SIGNAL FEATURE LEARNING

4.1 One-Dimensional Convolutional Neural Network

One-dimensional convolutional neural network (1D-CNN) is commonly used as a common deep neural network model for feature learning of one-dimensional signals. In this work, we design a multi-scale one dimensional convolutional neural network structure for learning signal features from multiple scales. In this, n parallel 1D-CNNs activated by ReLU are connected together in parallel by concatenation layer. The results of these channels form a complete feature vector in this way. Subsequently, this feature vector will be fully connected layer to learn the features in it. And finally, the final classification is done by the Softmax layer. In addition to this, the parameters of the first convolution of each parallel 1D-CNN are changed with the length of the multi-scale features. This ensures that the feature vector learned by the final fully connected layer contains a sufficient number of signal features at each scale. This model was trained using slide segmentation data. The structure of the multi-scale 1-D CNN is shown in Figure 4-1.

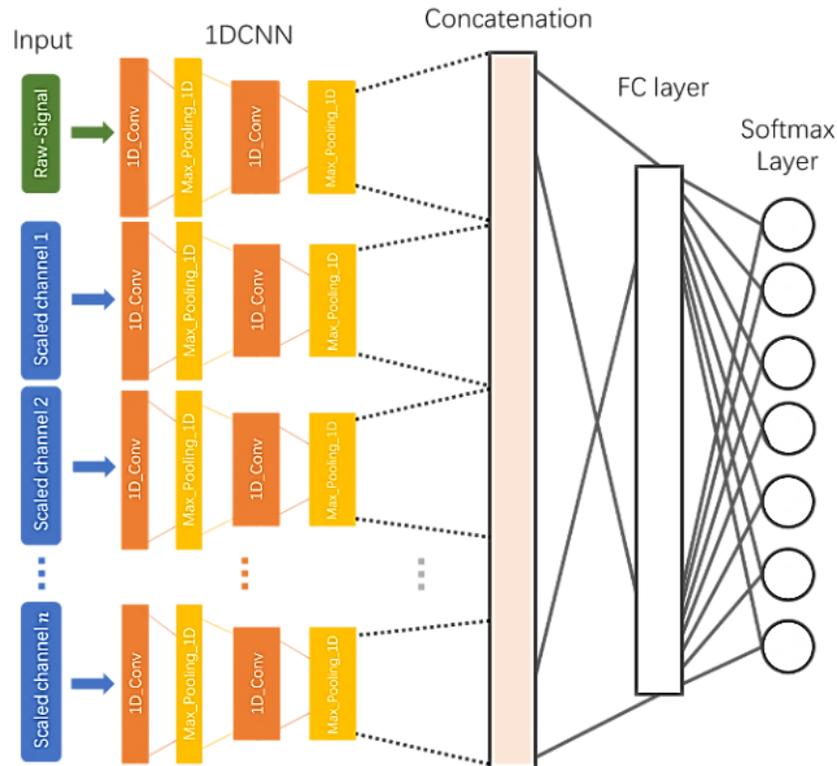


Figure 4-1 Multi-scale network based on 1D-CNN

Table 4-1 Information of Layers in 1DCNN-Based Multi-Scale Network

Layer name	Parameters (Kernel Size, No. of Feature Maps, Activation Function, Strides, Padding)
Conv0_1	Conv1D (9×1, 32, ReLU, 8, Same)
Conv1_1	Conv1D (9×1, 32, ReLU, 2, Same)
Conv2_1	Conv1D (9×1, 32, ReLU, 2, Same)
Conv3_1	Conv1D (9×1, 32, ReLU, 4, Same)
MXP0_1	MaxPooling1D (2, None, None, 1, None)
MXP1_1	MaxPooling1D (2, None, None, 1, None)
MXP2_1	MaxPooling1D (2, None, None, 1, None)
MXP3_1	MaxPooling1D (2, None, None, 1, None)
Conv0_2	Conv1D (9×1, 64, ReLU, 2, Same)
Conv1_2	Conv1D (9×1, 64, ReLU, 2, Same)
Conv2_2	Conv1D (9×1, 64, ReLU, 2, Same)
Conv3_2	Conv1D (9×1, 64, ReLU, 2, Same)
MXP0_2	MaxPooling1D (2, None, None, 1, None)
MXP1_2	MaxPooling1D (2, None, None, 1, None)
MXP2_2	MaxPooling1D (2, None, None, 1, None)
MXP3_2	MaxPooling1D (2, None, None, 1, None)
CONCATE	None
BN	None
FLATTEN	None
FC	Dense (128, ReLU, None, None)
OUTPUT	Dense (12, Softmax)

4.2 Residual Neural Network

Our second network is a deep residual network (ResNet). The base part of this network is also composed of one-dimensional convolutions. However, unlike a simple 1D convolutional neural network, this network introduces the concept of residual blocks to address the accuracy degradation caused by the increase in the number of layers of the deep neural network. Similar to Section 4-1, we use n parallel 1D-ResNet-34s to form a multiscale structure. These parallel branches are also connected by a concatenation layer to accomplish the classification task using the same fully connected layer structure as in Section 4-1. This model was trained using slide segmentation data. The structure of this network is shown in Figure 4-2.

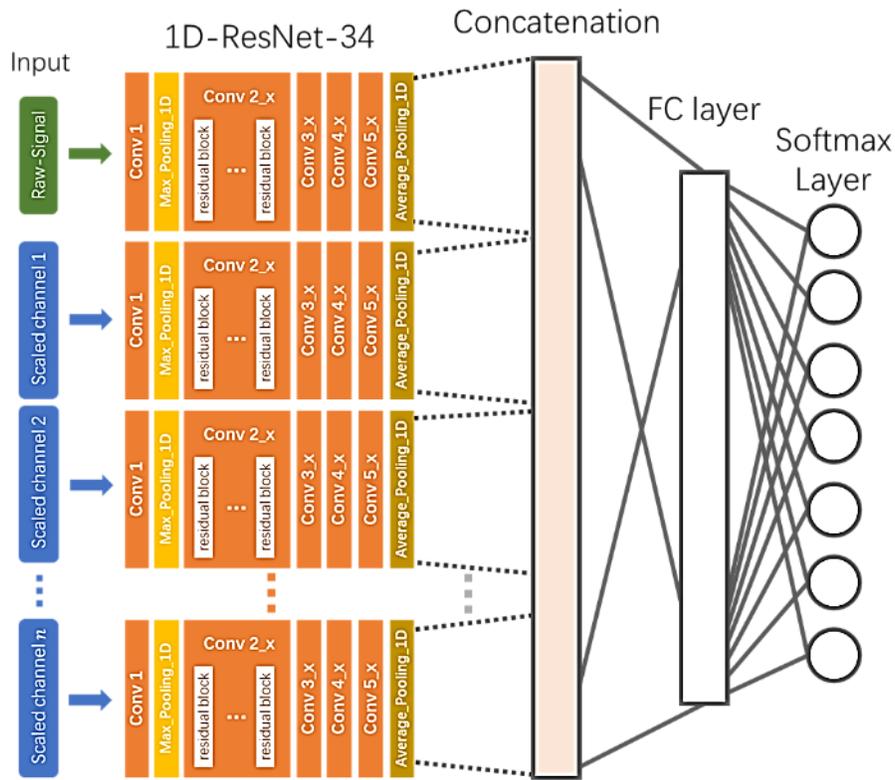


Figure 4-2 Multi-scale network based on 1D-ResNet-34

Table 4-2 Information of Layers in 1D-ResNet-34

Layer name	Parameters (Kernel Size, No. of Feature Maps, Activation Function, Strides, Padding)
Conv1	Conv1D (3×1, 32, Tanh, 3, Valid)
MPX	MaxPooling1D (3, None, None, 3, Same)
Conv2_x	Conv1D (3×1, 32, Tanh, 1, Same) ×3
Conv3_x	Conv1D (3×1, 64, Tanh, 1, Same) ×4
Conv4_x	Conv1D (3×1, 128, Tanh, 1, Same) ×6
Conv5_x	Conv1D (3×1, 128, Tanh, 1, Same) ×3
AVGP	AveragePooling1D (1, None, None, 1, None)
CONCATE	None
BN	None
FLATTEN	None
FC	Dense (128, ReLU, None, None)
OUTPUT	Dense (12, Softmax)

4.3 Gated Recurrent Unit

The multi-scale structure based on recurrent neural network (RNN) is employed for investigation. This is due to the fact that, similar to the role of one-dimensional convolution, RNNs are also good at feature extraction of one-dimensional signals. Typical representatives of deep neural networks based on RNN principles are gate recurrent Unit (GRU) and long-short term memory (LSTM). Here, we use GRU instead of LSTM for the reason that GRU can reduce the number of parameters needed for computation while obtaining similar training results. The composition of the GRU-based multiscale model is generally similar to that described in Sections 4-1 and 4-2. The difference, however, is that the RNN type of deep neural network represented by GRU requires more cache space than a one-dimensional convolution when processing signals of excessive length. To prevent training failure due to cache overflow, we shrink the input received by all branches to a suitable size first through the maxpooling layer and then use GRU to learn the features. This model was trained using slide segmentation data. The structure and parameters of the multi-scale GRU network is shown in Figure 4-3 and Table 4-3.

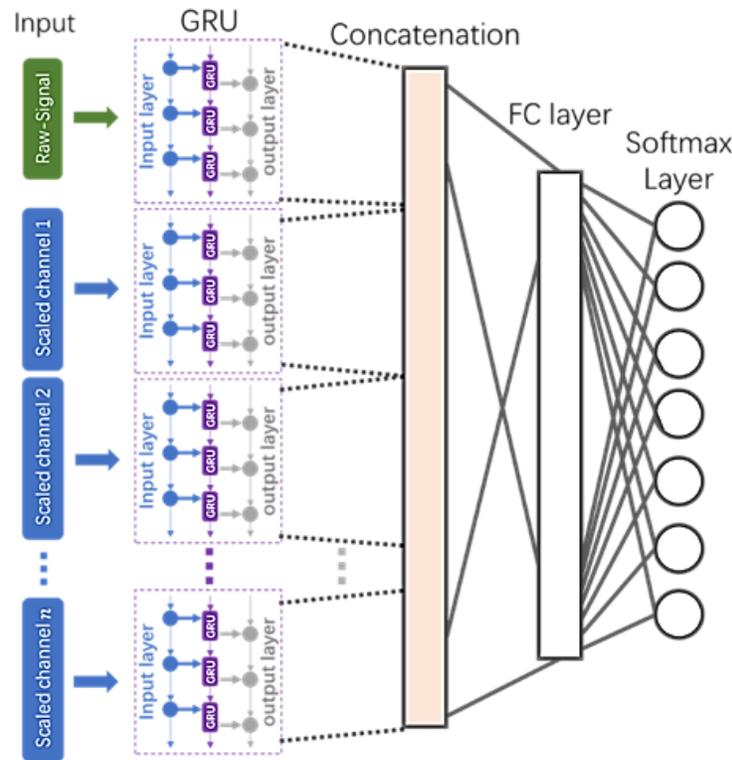


Figure 4-3 Multi-scale network based on GRU

Table 4-3 Information of Layers in GRU-based Multi-Scale Network

Layer name	Parameters (Kernel Size, No. of Feature Maps, Activation Function, Strides, Padding)
MXP0_1	MaxPooling1D (256, None, None, 1, None)
MXP1_1	MaxPooling1D (256, None, None, 1, None)
MXP2_1	MaxPooling1D (256, None, None, 1, None)
MXP3_1	MaxPooling1D (256, None, None, 1, None)
GRU0_2	GRU (8, None, tanh, 1, None)
GRU1-2	GRU (8, None, tanh, 1, None)
GRU2-2	GRU (8, None, tanh, 1, None)
GRU3-2	GRU (8, None, tanh, 1, None)
CONCATE	None
BN	None
FLATTEN	None
FC	Dense (128, ReLU, None, None)
OUTPUT	Dense (12, Softmax)

4.4 Multi-Scale and Multi-Rate Convolutional Neural Network

As described in Section 4-1, the multiscale networks only create different deep neural network branches within a single scale for the purpose of learning multiscale features. However, for the Multi-scale and Multi-rate features eventually obtained by the decomposition method we proposed in Section 3.2.4, it is obvious that the simple multi-scale structure cannot handle such complex structure of features. Therefore, in this research work, we design the corresponding multi-scale and multi-rate convolutional neural network (MSMR-CNN) to handle this class of features. The output obtained using sparse wavelet transform is a combination of sub-signals decomposed from multiple sub-branches. In order to process these complex feature forms into a form suitable for deep neural network processing. We concatenate the sub-signals decomposed by each subbranch into groups of the same frequency region to form a new 1D feature vector. This feature vector reflects the multi-rate characteristics of the frequency region. Each frequency region constitutes a separate feature vector. In this work, a total of three feature vectors representing different frequency regions are generated. This model was trained using random segmentation data. We acquired the 512 data segments from each 10 second raw signal. Figure 4-4 illustrates the

structure of our proposed MSMR-CNN algorithm. Table 4-4 shows the detailed structure for each 1-D CNN.

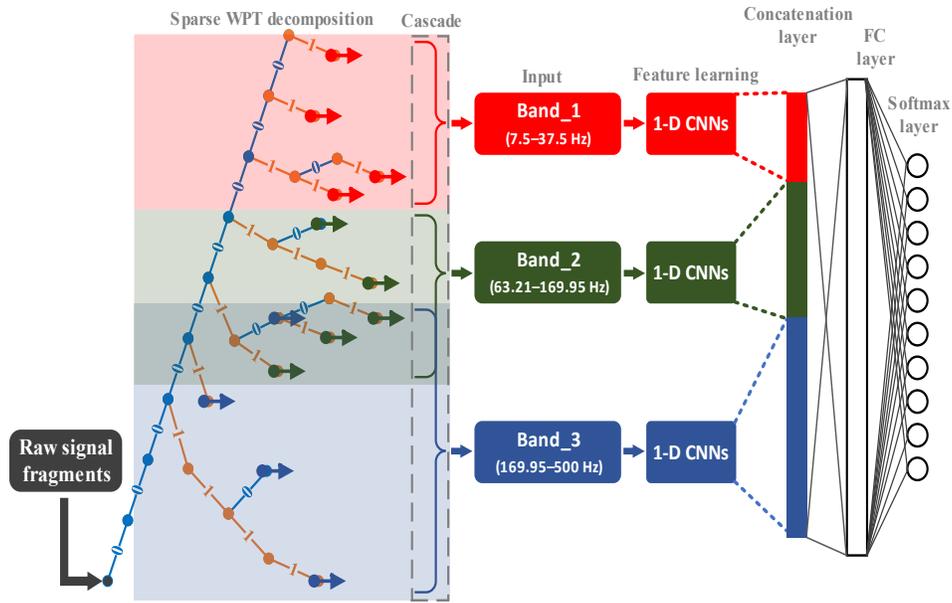


Figure 4-4 Multi-Scale and Multi-Rate Convolutional Neural Network

Table 4-4 Information of Layers in Multi-Scale and Multi-Rate Convolutional Neural Network

Layer name	Parameters (Kernel Size, No. of Feature Maps, Activation Function, Strides, Padding)
Conv0_1	Conv1D (3×1, 32, ReLU, 1, Same)
Conv1_1	Conv1D (3×1, 32, ReLU, 1, Same)
Conv2_1	Conv1D (3×1, 32, ReLU, 1, Same)
MXP0_1	MaxPooling1D (2, None, None, 1, None)
MXP1_1	MaxPooling1D (2, None, None, 1, None)
MXP2_1	MaxPooling1D (2, None, None, 1, None)
Conv0_2	Conv1D (3×1, 64, ReLU, 1, Same)
Conv1_2	Conv1D (3×1, 64, ReLU, 1, Same)
Conv2_2	Conv1D (3×1, 64, ReLU, 1, Same)
MXP0_2	MaxPooling1D (2, None, None, 1, None)
MXP1_2	MaxPooling1D (2, None, None, 1, None)
MXP2_2	MaxPooling1D (2, None, None, 1, None)
CONCATE	None
BN	None
FLATTEN	None
FC	Dense (128, ReLU, None, None)
OUTPUT	Dense (10, Softmax)

5. COMPARISON OF PERFORMANCE

To verify the effectiveness of our proposed method, we validated it on CWRU and self-collected datasets, respectively. Table 5-1 and Table 5-2 show the average classification accuracy, standard deviation, F1-score, and Recall values after ten rounds of independent testing, respectively.

Table 5-1 Accuracy and STD of Different Methods

ACC STD	1DCNN	ResNet	GRU
Raw data	94.87% 2.12	96.04% 2.12	78.64% 3.54
Down Sample	98.12% 1.25	88.57% 9.27	88.77% 2.74
EMD	98.64% 0.74	94.28% 3.19	90.39% 1.08
WPT	98.7% 0.92	96.69% 2.14	90.91% 1.39

Table 5-2 F1-Score and Recall of Different Methods

F1-score Recall	1DCNN	ResNet	GRU
Raw data	0.9632 0.9614	0.7754 0.7970	0.7271 0.7617
Down Sample	0.9864 0.9861	0.9806 0.9788	0.8663 0.8823
EMD	0.9860 0.9847	0.9929 0.9947	0.8631 0.8764
WPT	0.9916 0.9917	0.9833 0.9848	0.8952 0.8950

It can be seen that the best performance is obtained using the combination of WPT and 1D convolutional neural network. The combination of WPT and ResNet also obtains better performance in some values, but the training cost is higher because ResNet contains more parameters compared to 1D CNN. In addition, more parameters also affect the forward inference speed of the model, which will be detrimental to the processing speed of the algorithm when

running in real-time. Figure 5-1 illustrates the loss function curve during training and the confusion matrix after ten independent training sessions using the CWRU dataset (upper) and self-collected dataset (lower).

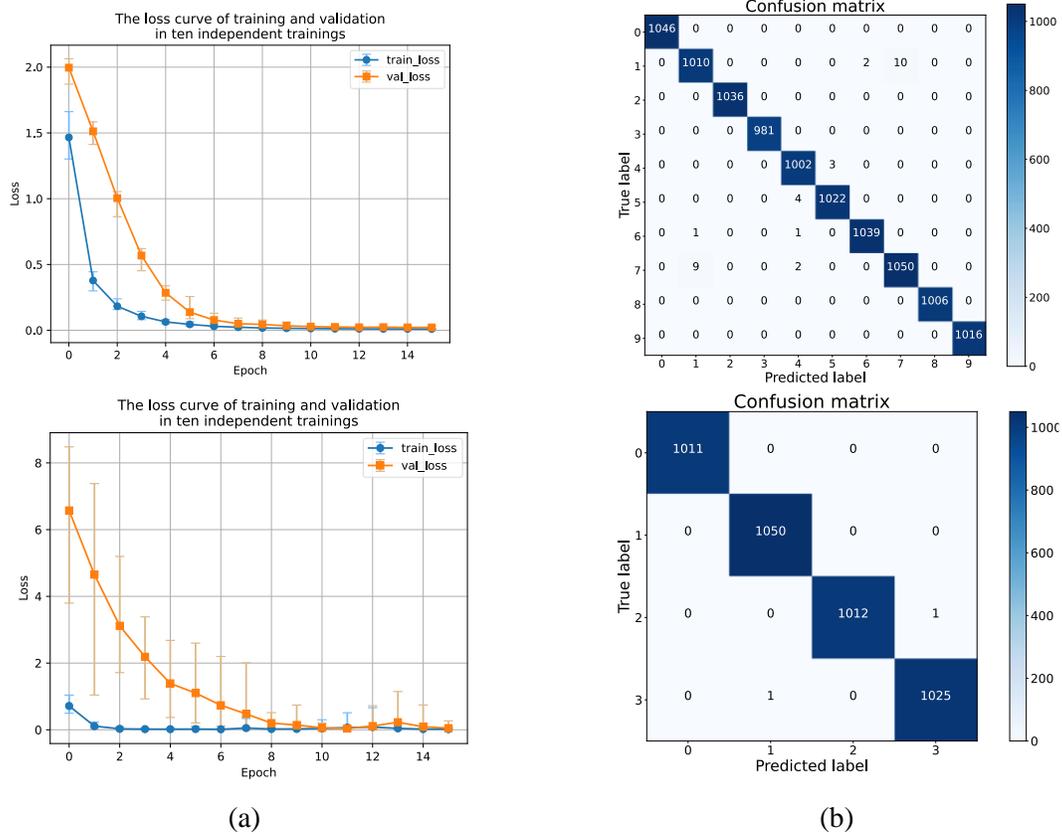


Figure 5-1 (a) Loss function curves for 10 independent trainings; (b) Confusion matrix

Also, we compared our proposed method with some popular bearing defect diagnosis methods proposed by other research groups in recent years. The comparison results are shown in Table 5-3. The results in Table 5-3 show that our method performs better on the CWRU and self-collected dataset.

Table 5-3 Comparison with Other Published Methods

Method	Accuracy (%)
ANN [26]	95.00
APF-KNN [29]	96.67
SVD [30]	85.00
DRSN-CW [31]	89.57
MSCNN [36]	98.53
Multi-scale CNN [37]	93.30
Proposed Method Multi-scale Multi-rate CNN	99.85

5.1 Data Mining for Classification

As a machine learning algorithm for dimensionality reduction, t-distributed stochastic neighbor embedding (t-SNE) was first proposed by Laurens van der Maaten and Geoffrey Hinton [49] in 2008. The method was developed from Stochastic Neighbor Embedding (SNE). While effectively downscaling high-dimensional data to two or three dimensions for visualization, it can avoid the shortcomings of other visualization methods such as Principal Component Analysis (PCA) that have difficulty in explaining the complex polynomial relationships between features and thus lead to underfitting. For each data point, t-SNE first calculates a Gaussian distribution centered on the target data point, and then selects the neighboring points in proportion to the probability density according to this distribution. The conditional probability of each neighboring point's distance from the target data point is the similarity of that point to the target data point. Subsequently, t-SNE uses a gradient descent method to find the minimum value of the similarity between all data points relative to each other in the high-dimensional space and the low-dimensional space. Through this process, the data points with a higher similarity will exhibit aggregation in the visualization space in two or three dimensions. In order to verify the performance of our proposed method, we have visualized and labeled the feature vectors output from the last dense layer, respectively. The visualized results are shown in Figure 5-2.

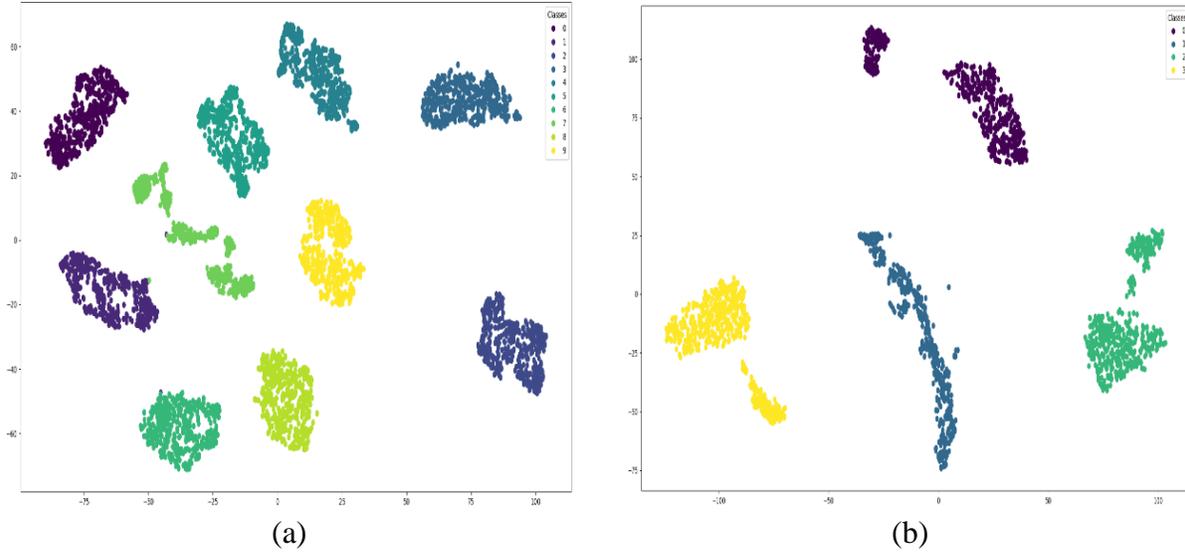


Figure 5-2 The visualization result using t-SNE (a) CWRU dataset (b) self-collected dataset

Figure 5-2 (a) shows the clustering effect of the CWRU dataset after processing by the proposed method while Figure 5-2 (b) shows the clustering effect of the self-collected dataset after the proposed method. From all the visualization results in Figure 5-2, it can be seen that the final dense layer of the neural network shows good learning results on both CWRU and self-collected datasets. The neural network successfully extracts the most relevant features to the type of bearing defects. This also reflects that the three frequency regions do contain effective information to extract learning features. Since our proposed feature extraction method removes a large amount of invalid information such as noise, higher harmonics, etc. from the original data, it is only necessary to extract the significant features from the original data. Only a simple neural network structure is required to learn the processed data efficiently. This greatly reduces the number of parameters and layers in the neural network. Therefore, our proposed bearing defect detection method can run on low-performance embedded devices at the computing speed which meets real-time processing.

6. CONCLUSION AND FUTURE WORK

In this thesis, we validate the performance of a combination of multiple signal processing algorithms and deep learning methods for bearing defect detection. In particular, we innovatively propose a sparse wavelet packet decomposition method. This method is able to decompose the original signal into feature vectors with multi-scale and multi-rate by encoding the optimal decomposition path. To learn this feature vector, we design a deep neural network called MSMR-CNN. Our algorithm eventually achieves 99.85% classification accuracy on the CWRU dataset and 99.95% classification accuracy on the self-collected dataset. The excellent performance on multiple datasets demonstrates the advantages of our proposed method over other bearing defect diagnosis methods. The future work will include using more novel neural networks including graph neural networks for the feature learning task. We believe that by incorporating the relationship between frequency regions into the learning range of the deep learning model will further increase the accuracy and robustness of our proposed method in more complex environments.

REFERENCES

- [1] Robert Bond Randall, *Vibration-based Condition Monitoring: Industrial, Automotive and Aerospace Applications*. Second Edition, Wiley, 2021.
- [2] Y. Lei, *Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery*. Butterworth-Heinemann, 2017.
- [3] R. N. Bell, D. W. McWilliams, P. O'Donnell, C. Singh, and S. J. Wells, "Report of large motor reliability survey of industrial and commercial installations, part I," *IEEE Trans. Ind. Appl.*, vol. IA-21, no. 4, pp. 853–864, Apr. 1985.
- [4] R. N. Bell, C. R. Heising, P. O'Donnell, S. J. Wells, and C. Singh, "Report of large motor reliability survey of industrial and commercial installations, part II," *IEEE Trans. Ind. Appl.*, vol. IA-21, no. 4, pp. 865–872, Jan. 1985.
- [5] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Deep learning algorithms for bearing fault diagnostics—A comprehensive review," *IEEE Access*, vol. 8, pp. 29857–29881, 2020.
- [6] W. Huang, J. Cheng, and Y. Yang, "Rolling bearing fault diagnosis and performance degradation assessment under variable operation conditions based on nuisance attribute projection," *Mech. Syst. Signal Process.*, vol. 114, pp. 165–188, Jan. 2019.
- [7] M. Benbouzid, M. Vieira, and C. Theys, "Induction motors' faults detection and localization using stator current advanced signal processing techniques," *IEEE Trans. Power Electron.*, vol. 14, pp. 14-22, Jan. 1999.
- [8] M. El Hachemi Benbouzid, "A review of induction motors signature analysis as a medium for faults detection," *IEEE Trans. Ind. Electron.*, vol. 47, no. 5, pp. 984-993, Oct. 2000.
- [9] P. Zhang, Y. Du, T. G. Habetler, and B. Lu, "A survey of condition monitoring and protection methods for medium-voltage induction motors," *IEEE Trans. Ind. Appl.*, vol. 47, no. 1, pp. 34-46, Jan./Feb. 2011.
- [10] M. S. Safizadeh and S. K. Latifi, "Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell," *Inf. Fusion*, vol. 18, pp. 1–8, Jul. 2014.
- [11] Boudiaf, A. Moussaoui, A. Dahane, and Atoui, "A comparative study of various methods of bearing faults diagnosis using the case western reserve university data," *J. Failure Anal. Prevention*, vol. 16, no. 2, pp. 271–284, Apr. 2016.

- [12]D. Neupane and J. Seok, "Bearing fault detection and diagnosis Using Case Western Reserve University dataset with deep learning Approaches: A Review," *IEEE Access*, vol. 8, pp. 93155-93178, June 2020.
- [13]W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mechanical Systems and Signal Processing*, vol. 64–65, pp. 100-131, Dec. 2015.
- [14]L. Tan and J. Jiang, *Digital Signal Processing: Fundamentals and Applications*. Third Edition, Elsevier/Academic Press, 2018.
- [15]V. K. Rai and A. R. Mohanty, "Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2607–2615, Aug. 2007.
- [16]Z. Peng, F. Chu, and Y. He, "Vibration signal analysis and feature extraction based on reassigned wavelet scalogram," *J. Sound Vibrat.*, vol. 253, no. 5, pp. 1087–1100, Jun. 2002.
- [17]C. Cheng, B. Zhou, G. Ma, D. Wu, and Y. Yuan, "Wasserstein distance based deep adversarial transfer learning for intelligent fault diagnosis with unlabeled or insufficient labeled data," *Neurocomputing*, vol. 409, pp. 35-45, Jun. 2020.
- [18]C. Castejón, M. Jesús, J. J. Gómez, J. Carlos García-Prada, A. J. O. Nez, and H. Rubio, "Automatic selection of the WPT decomposition level for condition monitoring of rotor elements based on the sensitivity analysis of the wavelet energy," *Int. J. Acoust. Vib.*, vol. 20, no. 2, pp. 95–100, June 2015.
- [19]P. Gupta and M. K. Pradhan, "Fault detection analysis in rolling element bearing: A review," *Mater. Today, Proc.*, vol. 4, no. 2, pp. 2085–2094, April 2017.
- [20]N. Tandon and B. C. Nakra, "Detection of defects in rolling element bearings by vibration monitoring," *J. Mech. Eng. Division*, vol. 73, pp. 271–282, Jan. 1993.
- [21]D. Dyer and R. M. Stewart, "Detection of rolling element bearing damage by statistical vibration analysis," *J. Mech. Des.*, vol. 100, no. 2, pp. 229–235, April 1978.
- [22]R. Dwyer, "Detection of non-Gaussian signals by frequency domain kurtosis estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, pp. 607–610, Dec. 1983.
- [23]Shrivastava and S. Wadhvani, "Development of fault detection system for ball bearing of induction motor using vibration signal," *Int. J. Sci. Res.*, vol. 2, no. 5, pp. 256–259, June 2012.

- [24] M. Cocconcelli, R. Zimroz, R. Rubini, and W. Bartelmus, "STFT based approach for ball bearing fault detection in a varying speed motor," in *Proc. Cond. Monit. Mach. Non-Stationary Oper.*, pp. 41-50, Oct. 2012.
- [25] J.-H. Lee, J. Kim, and H.-J. Kim, "Development of enhanced Wigner Ville distribution function," *Mech. Syst. Signal Process.*, vol. 15, no. 2, pp. 367–398, Mar. 2001.
- [26] M.-Y. Chow, P. M. Mangum, and S. O. Yee, "A neural network approach to real-time condition monitoring of induction motors," *IEEE Trans. Ind. Electron.*, vol. 38, no. 6, pp. 448–453, 1991.
- [27] Malhi and R. X. Gao, "PCA-based feature selection scheme for machine defect classification," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 6, pp. 1517–1525, Dec. 2004.
- [28] Y. S. Wang, Q. H. Ma, Q. Zhu, X. T. Liu, and L. H. Zhao, "An intelligent approach for engine fault diagnosis based on Hilbert–Huang transform and support vector machine," *Appl. Acoust.*, vol. 75, pp. 1–9, Jan. 2014.
- [29] D. H. Pandya, S. H. Upadhyay, and S. P. Harsha, "Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN," *Expert Syst. Appl.*, vol. 40, no. 10, pp. 4137–4145, Aug. 2013.
- [30] F. Shen, C. Chen, R. Yan, and R. X. Gao, "Bearing fault diagnosis based on SVD feature extraction and transfer learning classification," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM)*, pp. 1-6, Oct. 2015.
- [31] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, "Deep residual shrinkage networks for fault diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4681-4690, July 2020.
- [32] J. Chen, J. Jiang, X. Guo, and L. Tan, "A self-adaptive CNN with PSO for bearing fault diagnosis," *Systems Science & Control Engineering*, vol. 9, no. 1, pp. 11-22, Dec. 2020.
- [33] J. Chen, J. Jiang, X. Guo, and L. Tan, "An efficient CNN with tunable input-size for bearing fault diagnosis," *International Journal of Computational Intelligence Systems*, vol. 14, no. 1, pp. 625-634, Jan. 2021.
- [34] Y. Cai, L. Tan, and J. Chen, "Evaluation of deep learning neural networks with input processing for bearing fault diagnosis," in *2021 IEEE International Conference on Electro Information Technology (EIT)*, pp. 140-145, May 2021.

- [35] W. Chen and K. Shi, "Multi-scale attention convolutional neural network for time series classification," *Neural Networks*, vol. 136, pp. 126-140, April 2021.
- [36] G. Jiang, H. He, J. Yan, and P. Xie, "Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 4, pp. 3196-3207, April 2019.
- [37] Y. Yao, S. Zhang, S. Yang, and G. Gui, "Learning attention representation with a multi-scale CNN for gear fault diagnosis under different working conditions," *Sensors*, vol. 20, no. 4, pp. 1233-1253, Feb. 2020.
- [38] X. Liu, Y. Cai, Y. Song, and L. Tan, "Bearing fault diagnosis based on multi-scale neural networks," presented at the 2022 *IEEE International Conference on Electro Information Technology (EIT)*, pp. 80-85, May 2022.
- [39] El-Thalji and E. Jantunen, "Fault analysis of the wear fault development in rolling bearings," *Engineering Failure Analysis*, vol. 57, pp. 470-482, Nov. 2015
- [40] L. Eren, "Bearing fault detection by one-dimensional convolutional neural networks," *Mathematical Problems in Engineering*, vol. 2017, pp. 1-9, Jul. 2017.
- [41] Stack, T. Habetler, and R. Harley, "Fault classification and fault signature production for rolling element bearings in electric machines," *IEEE Transactions on Industry Applications*, vol. 40, no. 3, pp. 735-739, May/Jun. 2004.
- [42] N. Chimitt, W. Misch, L. Tan, A. Togbe, and J. Jiang, "Comparative study of simple feature extraction for single-channel EEG based classification," in *2017 IEEE International Conference on Electro Information Technology (EIT)*, pp. 166-170, May 2017.
- [43] Dai, V. Vijayarajan, X. Peng, L. Tan, and J. Jiang, "Speech recognition using sparse discrete wavelet decomposition feature extraction," in *2018 IEEE International Conference on Electro/Information Technology (EIT)*, pp. 0812-0816, May 2018.
- [44] J. Dai, Y. Zhang, J. Hou, X. Wang, L. Tan, and J. Jiang, "Sparse wavelet decomposition and filter banks with CNN deep learning for speech recognition," in *2019 IEEE International Conference on Electro Information Technology (EIT)*, pp. 098-103, May 2019.
- [45] X. Liu, J. Centeno, J. Alvarado, and L. Tan, "One dimensional convolutional neural networks using sparse wavelet decomposition for bearing fault diagnosis," *IEEE Access*, vol. 10, pp. 86998-87007, Aug. 2022.

- [46]Case Western Reserve University (CWRU) Bearing Data Center. Accessed: Dec. 2018.[Online].Available:<https://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-universitybearing-data-center-website>
- [47]N. Huang, Z., Shen, S. Long, et al, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings Mathematical Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903-995, 1998.
- [48]H. Bourlard and Y. Kamp, "Auto-association by multilayer perceptrons and singular value decomposition," *Biological Cybernetics*, vol. 59, no. 4-5, pp. 291-294, 1988.
- [49]L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of machine learning research*, vol. 9, no. 11, pp. 2579-2605, Nov. 2008.

PUBLICATIONS

X. Liu, J. Centeno, J. Alvarado, and L. Tan, "One dimensional convolutional neural networks using sparse wavelet decomposition for bearing fault diagnosis," *IEEE Access*, vol. 10, pp. 86998-87007, Aug. 2022.

X. Liu, Y. Cai, Y. Song, and L. Tan, "Bearing fault diagnosis based on multi-scale neural networks," presented at the *2022 IEEE International Conference on Electro Information Technology (EIT)*, pp. 80-85, May 2022.