



Faculty of Engineering

**Development of Deep Learning-based Approach for Robust Fault
Diagnosis and Failure Prediction of Bearings**

Li Xiuyu

**Doctor of Philosophy
2024**

Development of Deep Learning-based Approach for Robust Fault Diagnosis and Failure Prediction of Bearings

Li Xiuyu

A thesis submitted

In fulfillment of the requirements for the degree of Doctor of Philosophy

(Mechanical Engineering)

Faculty of Engineering
UNIVERSITI MALAYSIA SARAWAK
2024

DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Malaysia Sarawak. Except where due acknowledgements have been made, the work is that of the author alone. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

.....

Signature

Name: Li Xiuyu

Matric No.: 21010282

Faculty of Engineering

Universiti Malaysia Sarawak

Date : 10/12/2024

ACKNOWLEDGEMENT

I would like to express my gratitude to those who have made direct or indirect contributions to the successful completion of my PhD thesis.

I would like to extend my heartfelt gratitude to my family for their unwavering support throughout this journey. While pursuing my studies in Malaysia, they took exceptional care of our two children, enabling me to completely immerse myself in my academic pursuits. Their selflessness and understanding have been instrumental in my ability to focus on my research. Their love and encouragement have been a constant source of motivation, and I am profoundly grateful for their sacrifices and belief in me.

Additionally, I am deeply indebted to Anhui Business and Technology College for providing the essential funding that enabled me to establish the bearing fault data acquisition platform. Their financial support paved the way for the successful completion of this research project, facilitating the acquisition of necessary resources and materials for experimentation and analysis.

I would like to express my sincere appreciation to my research supervisor, Dr. Shirley Johnathan Tanjong, for her invaluable guidance, expertise, and encouragement throughout this endeavour. Her mentorship has been pivotal in shaping the direction of my research and enhancing its quality.

Lastly, I extend my gratitude to all those who have contributed to this research in various capacities, whether through support, advice, or encouragement. Your contributions have been deeply appreciated and have played a significant role in the completion of this study. Thank you all.

ABSTRACT

Bearings play a crucial role in rotating machinery. Once a bearing failure occurs, it may lead to severe casualties and economic losses. Therefore, diagnosing bearing faults and ensuring their smooth operation are essential steps in ensuring the safe and stable operation of modern mechanical equipment. However, traditional methods for diagnosing bearing faults typically involve complex manual denoising, filtering, and feature extraction processes, which are both tedious and lack intelligence, making them particularly challenging when dealing with large volumes of data. The operating conditions of bearings are complex and involve various factors, making manual extraction of fault features even more difficult. Thus, this motivates research in improvements of end-to-end bearing fault diagnosis and life prediction. Furthermore, the rapid development of information technology has made deep learning a powerful tool for automatically learning and diagnosing bearing faults under complex conditions. This research aimed to investigate the end-to-end bearing fault diagnosis and life prediction under complex scenarios such as high noise and different operating conditions, explore various deep neural networks for effective bearing degradation prediction and fault diagnosis. To achieve the aim, this research was segmented to 3 parts: (i) Firstly, to achieve end-to-end diagnosis of bearing fault types and severity, this thesis proposed a Multi-scale processing, Channel attention, Feature enhancement, Convolutional Neural Network (MCFCNN) network architecture. The MCFCNN network combines multi-channel parallel convolution, effectively capturing spatial information, and introduces channel attention mechanisms to adaptively recalibrate channel-level feature responses. Moreover, secondary neurons are introduced to enhance the model's ability to capture complex nonlinear patterns related to bearing faults. In a high-noise environment, the proposed model outperforms existing models in fault

diagnosis, with accuracy greater than 80%. (ii) Secondly, a bearing fault test platform was built to validate the effectiveness of the MCFCNN model. Rolling bearings vibration signals were generated using the test platform. In the experiments, the performance of the MCFCNN model in high-noise conditions and the results of different network structures and data augmentation strategies on the final results were verified. In the cross operating scenario, the average resolution in the Bearing Fault Test System (BFTS) dataset is 0.7595, compared to WDCNN with 0.6492 and QCNN with 0.7266. (iii) In the third and final part of this research, the problem of predicting the remaining life of bearings was approached as a classification problem. Rather than predicting the remaining life, the priority was to identify the stage of a bearing's life cycle for better maintenance and management. Bearing life was divided into four different stages and used the bearing data to predict the current stage of the bearing's life. Based on this, a method for early prediction of bearing degradation, referred to as Fourier Transform Autoencoder with K-means (FAEK) automatic labeling and Multi-scale Processing Channel Attention Classification (MCC) classification prediction model, was proposed. This method combines a convolutional autoencoder segmentation network with a high-performance classifier, effectively distinguishing bearing degradation processes and achieving accurate classification recognition. In the degradation stage prediction task, our model architecture achieves an accuracy of 0.9665. This thesis focuses on deep learning research applied to bearing fault diagnosis and prediction, which helps to reduce the workload of manual maintenance personnel, enhance the efficiency and accuracy of fault detection, and mitigate losses caused by equipment failures. Indirectly, this reduces maintenance costs, enhances equipment reliability and improves production efficiency, through timely maintenance and fault prevention of equipment.

Keywords: Bearing, fault diagnosis, degradation prediction, deep learning, MCFCNN, automatic label assignment, classification prediction model

***Pembangunan Aturcara Berasaskan Pembelajaran Mendalam untuk Kemantapan
Diagnosis Kesalahan dan Ramalan Kegagalan Bebola Galas***

ABSTRAK

Bebola gelas memainkan peranan penting dalam mesin berputar. Kegagalan gelas boleh menyebabkan kemalangan dan kerugian ekonomi yang besar. Oleh itu, diagnosis kerosakan gelas dan memastikan operasi gelas berjalan lancar adalah langkah penting untuk menjamin keselamatan dan kestabilan operasi peralatan mekanikal moden. Namun begitu, kaedah tradisional untuk mendiagnosis kerosakan gelas biasanya melibatkan proses penyingkiran hingar, penapisan, dan pengekstrakan ciri dengan cara manual yang kompleks. Proses ini bukan sahaja membebankan tetapi juga kurang pintar, menjadikannya lebih mencabar apabila berurusan dengan jumlah data yang besar. Keadaan operasi gelas yang kompleks dan melibatkan pelbagai faktor turut menjadikan pengekstrakan ciri kerosakan secara manual lebih sukar. Oleh itu, ini mendorong penyelidikan dalam peningkatan diagnosis kerosakan terkini dan ramalan hayat gelas. Selain itu, pembangunan pesat teknologi maklumat menjadikan pembelajaran mendalam (deep learning) teknik yang ampuh untuk pembelajaran dan diagnosis automatik kerosakan gelas dalam keadaan kompleks. Penyelidikan ini bertujuan untuk mengkaji diagnosis kerosakan gelas dan ramalan hayat dalam senario kompleks seperti hingar tinggi dan keadaan operasi yang berbeza, serta meneroka pelbagai rangkaian neural mendalam untuk ramalan degradasi dan diagnosis kerosakan gelas yang berkesan. Untuk mencapai objektif ini, kajian ini dibahagikan kepada tiga bahagian: (i) Pertama, untuk mencapai diagnosis jenis kerosakan gelas dan tahap kerosakan, tesis ini melaporkan proses Multi-scale processing, Channel attention, Feature enhancement, Convolutional Neural Network (MCFCNN). Rangkaian MCFCNN menggabungkan konvolusi selari

berbilang saluran, menangkap maklumat ruang dengan efektif, dan memperkenalkan mekanisme perhatian saluran untuk menyesuaikan semula respons ciri pada peringkat saluran secara adaptif. Selain itu, neuron sekunder diperkenalkan untuk meningkatkan keupayaan model menangkap corak tidak linear yang kompleks berkaitan dengan kerosakan gelas. Dalam persekitaran hingar tinggi, model yang dicadangkan menunjukkan prestasi yang lebih baik daripada model sedia ada dengan ketepatan lebih daripada 80%. (ii) Kedua, platform ujian kerosakan gelas dibina untuk mengesahkan keberkesanan model MCFCNN. Isyarat getaran gelas berguling dijana menggunakan platform ujian ini. Dalam eksperimen, prestasi model MCFCNN dalam keadaan hingar tinggi serta keputusan pelbagai struktur rangkaian dan strategi pengayaan data terhadap hasil akhir telah disahkan. Dalam senario operasi silang, resolusi purata dalam set data Bearing Fault Test System (BFTS) adalah 0.7595, berbanding WDCNN dengan 0.6492 dan QCNN dengan 0.7266. (iii) Dalam bahagian ketiga dan terakhir penyelidikan ini, masalah ramalan hayat gelas ditangani sebagai masalah klasifikasi. Daripada meramalkan baki hayat gelas, keutamaan diberikan kepada mengenal pasti tahap kitaran hayat gelas untuk penyelenggaraan dan pengurusan yang lebih baik. Hayat gelas dibahagikan kepada empat peringkat berbeza, dan data keadaan gelas digunakan untuk meramalkan tahap semasa kitaran hayat gelas. Berdasarkan ini, kaedah ramalan awal degradasi gelas, yang dikenali sebagai Fourier Transform Autoencoder dengan K-means (FAEK) pelabelan automatik dan Multi-scale Processing Channel Attention Classification (MCC), digunakan di dalam penyelidikan ini. Kaedah ini menggabungkan rangkaian segmen pengekod automatik konvolutional dengan pengekklasifikasi berprestasi tinggi, membezakan proses degradasi bearing dengan efektif dan mencapai pengenalan kelas yang tepat. Dalam tugas ramalan tahap degradasi, model binaan penyelidikan ini

mencapai ketepatan 0.9665. Tesis ini memfokuskan kepada penyelidikan pembelajaran mendalam yang diterapkan dalam diagnosis dan ramalan kerosakan gelas, yang berpotensi membantu mengurangkan beban kerja kakitangan penyelenggaraan manual, meningkatkan kecekapan dan ketepatan pengesanan kerosakan, serta mengurangkan kerugian akibat kegagalan mesin. Secara tidak langsung, ia membantu mengurangkan kos penyelenggaraan, meningkatkan kebolehpercayaan peralatan, dan meningkatkan kecekapan pengeluaran melalui penyelenggaraan tepat pada masanya dan pencegahan kerosakan peralatan.

Kata kunci: *Bebola gelas, diagnosis kesilapan, ramalan degradasi, pembelajaran mendalam, MCFCNN, pengenalan label automatik, model ramalan klasifikasi*

TABLE OF CONTENTS

	Page
DECLARATION	i
ACKNOWLEDGEMENT	ii
ABSTRACT	iii
<i>ABSTRAK</i>	vi
TABLE OF CONTENTS	ix
LIST OF TABLES	xiv
LIST OF FIGURES	xv
LIST OF ABBREVIATIONS	xvii
CHAPTER 1 INTRODUCTION	1
1.1 Research Overview	1
1.2 Research Background	2
1.3 Research Problem and Hypotheses	4
1.4 Research Questions	8
1.5 Research Objectives and Research Scope	10
1.6 Significance of the Research	11
1.7 Thesis Outline	12
CHAPTER 2 LITERATURE REVIEW	14
2.1 Chapter Overview	14
2.2 Traditional Fault Diagnosis Methods	14

2.2.1 Time Domain-based Bearing Fault Diagnosis Method	14
2.2.2 Frequency Domain-based Bearing Fault Diagnosis Method	16
2.2.3 Time-Frequency Domain-based Bearing Fault Diagnosis Method	18
2.3 Deep Learning Methods in Bearing Fault Diagnosis	22
2.3.1 Convolutional Neural Network (CNN)	23
2.3.2 Auto-Encoder	28
2.3.3 Deep Belief Network (DBN)	34
2.3.4 Recurrent Neural Network (RNN)	39
2.4 Chapter Summary	45
CHAPTER 3 METHODOLOGY	46
3.1 Overview	46
3.2 Research Design and Procedures	46
3.3 The Basic Theory of Rolling Bearing Fault Diagnosis	48
3.3.1 Rolling Bearing Structure	48
3.3.2 Vibration Frequency of Bearing Fault Features	49
3.4 Bearing Dataset	52
3.4.1 Case Western Reserve University (CWRU) Bearing Dataset	52
3.4.2 BFTS Bearing Dataset	54
3.4.3 FEMTO-ST Bearing Dataset	56
3.4.4 Data Preprocessing	58

3.5 Deep learning model knowledge used in the model of the thesis	63
3.5.1 Convolutional Neural Network	63
3.5.2 Auto Encoder	71
3.6 Model Training and Optimisation	72
3.6.1 Loss Functions	73
3.6.2 Optimisation Algorithm	75
3.7 Feature Extraction	78
3.7.1 K-Means	78
3.7.2 T-SNE Visualisation Technique	79
3.8 Model Evaluation Indicator	81
3.9 Chapter Summary	82
CHAPTER 4 FAULT DIAGNOSIS TASK IN HIGH NOISE ENVIRONMENT	
USING MCFCNN MODEL	84
4.1 Overview	84
4.2 MCFCNN Network Architecture	85
4.2.1 Multi-scale Processing Layer	86
4.2.2 Channel Attention Layer	87
4.2.3 Feature Enhancement Layer	89
4.2.4 Anomaly Classification Layer	90
4.3 Data Pre-processing and Model Parameter Settings	91

4.4 Comparison with Other Methods	94
4.4.1 Comparison of Different Models	95
4.4.2 T-SNE Visualization	97
4.4.3 Confusion Matrix Comparison	99
4.5 Analysis of MCFCNN	101
4.6 Chapter Summary	103
CHAPTER 5 VALIDATION OF MCFCNN MODEL ON BFTS TEST	
BENCH	104
5.1 Chapter Overview	104
5.2 The BFTS Test Bench	105
5.2.1 Machinery System	105
5.2.2 Accelerometer	107
5.2.3 Data Collection and Control Unit	108
5.3 Data Sets Collected by the BFTS Test Platform	113
5.4 Experimental Results	115
5.5 Ablation Experiment	118
5.6 Chapter Summary	121
CHAPTER 6 BEARING DEGRADATION PREDICTION BASED ON FAEK	
AND MCC	123
6.1 Chapter Overview	123
6.2 Theoretical Knowledge of Bearing Degradation	125

6.3 Network Architecture	128
6.3.1 Bearing Degradation Model	131
6.3.2 Prediction Model	133
6.4 Experimental Results	137
6.4.1 Data Labeling	139
6.4.2 Classifier	141
6.5 Chapter Summary	145
CHAPTER 7 CONCLUSION AND RECOMMENDATIONS	147
7.1 Conclusion	147
7.2 Recommendations	149
REFERENCES	151
APPENDICES	176

LIST OF TABLES

	Page
Table 2.1: Common Time-Domain Features in Bearing Fault Diagnosis	15
Table 2.2: Comparison of CNN techniques for fault-bearing FDP research	27
Table 2.3: Comparison of auto-encoder techniques for FDP research	33
Table 2.4: Comparison of FDP studies using Deep Belief Network	38
Table 2.5: Recurrent Neural Network utilized for FDP research comparison	43
Table 3.1: Specific parameters of rolling bearing (DE) (Zhu et al., 2023)	54
Table3.2: Rolling bearing fault information (Zhu et al., 2023)	54
Table 3.3: Rolling bearing fault information	55
Table3.4: FEMTO-ST data set information	57
Table 4.1: Ten classes in CWRU data set	92
Table 4.2: The architecture configuration of MCFCNN	93
Table 4.3: Accuracy of Different Models at Different Signal-to-Noise Ratios	95
Table 4.4: Comparison of model results (SNR=-6)	102
Table 5.1: Motor parameters	107
Table5.2: Bearing parameter	107
Table 5.3: Acceleration sensor parameter	108
Table 5.4: HFUSB-4004 parameters	110
Table 5.5: Bearing failure type and label	114
Table 6.1: Bearing degradation stage classification	130
Table 6.2: Distribution of FEMTO-ST bearing dataset	138

LIST OF FIGURES

	Page
Figure 2.1: An architecture of convolutional neural network (Chen et al., 2015)	23
Figure2.2: An architecture of an Auto-Encoder (Yao et al., 2016)	29
Figure 2.3: An architecture of Deep Belief Network (Lin et al., 2016)	35
Figure 2.4: An architecture of Recurrent Neural Network (Li et al., 2018)	40
Figure 3.1: Research methodology flow diagram	47
Figure 3.2: Rolling bearing (Zhang et al., 2020)	49
Figure3.3: Principle schematic diagram of bearing fault vibration signal generation	50
Figure 3.4: CWRU rolling bearing test bench (Zhu et al., 2023)	53
Figure 3.5: BFTS Fault Acquisition Platform	55
Figure 3.6: Experimental setup for FEMTO-ST (Zhang et al., 2018)	57
Figure 3.7: Data augmentation with overlap(Zhang et al., 2017)	60
Figure 3.8: Comparison of clean and noise-disturbed vibration time series plots	62
Figure 3.9: Schematic Comparison of Parameter Count between (a) Fully Connected Neural Network and (b) CNN (Qiu, 2016)	64
Figure 3.10: Schematic Diagram of Max Pooling with Pooling Kernel of 2 and Stride of 2	70
Figure 4.1: MCFCNN network architecture	86
Figure 4.2: Channel Attention Module (Shen et al., 2020)	88
Figure 4.3: Average accuracy of the model under different signal-to-noise ratios	96
Figure 4.4: Loss of Different Models (1HP)	97
Figure 4.5: Different models' t-SNE visualization results on the CWRU dataset (1HP, SNR = -4)	99
Figure 4.6: Shows the confusion matrices for MCFCNN and QCNN at SNR = -4.	101

Figure 4.7: The output feature map of the first layer of the MCFCNN	103
Figure 5.1: Mechanical Part of BFTS Test Bench	106
Figure 5.2: Bearings	106
Figure 5.3: Acceleration sensor	108
Figure 5.4: Dynamic signal analyzer	109
Figure 5.5: Dynamic signal analyzer software interface	109
Figure 5.6: BFTS Test Platform	113
Figure 5.7: Comparison between original signals of rolling element faults and signals with added noise (SNR=0)	116
Figure 5.8: Contrasting Different Methods on Experimental Data set Across Various Signal-to-Noise Ratios	118
Figure 5.9: Comparison on experimental data set with different network structures at different signal-to-noise ratios	119
Figure 5.10: Comparison between different models under different loads on the experimental data set	121
Figure 6.1: Bearing Life Stage Workflow	129
Figure 6. 2: FAEK Network Architecture	133
Figure 6.3: MCC architecture	134
Figure 6.4: Training set degradation labels	141
Figure 6.5: Accuracy of Different Classification Models	142
Figure 6.6: Test Set Accuracy of the Models (based on FAEK labels)	143
Figure 6.7: Overlap Rate of Adjacent Stages for the Models	144
Figure 6.8: The prediction results for bearings 1_5 (a), 2_4 (b), and 3_3 (c)	145

LIST OF ABBREVIATIONS

Adam	Adaptive Moment Estimation
ALSCN	Adaptive Long Sequence Convolutional Network
AR	Auto-Regressive
AResNet	Aggregated Residual Network
ARMA	Autoregressive Moving Average
ATLSAE	Adversarial Transfer Learning Based On Stacked Auto-Encoder
BFTS	Bearings Fault Test System
BN	Batch Normalization
BPFI	Ball Pass Frequency Inner
BPFO	Ball Pass Frequency Outer
BSF	Ball Spin Frequency
CNN	Convolutional Neural Network
CWRU	Case Western Reserve University
CWT	Continuous Wavelet Transform
DBN	Deep Belief Network
DSD	Double Sparse Dictionary
DTCWPT	Dual-Tree Complex Wavelet Packet Transform
DTL	Deep Transfer Learning
DWT	Discrete Wavelet Transform
EDLSTM	Ensemble of Deep, Long-Term, and Short-Term Memory
EEMD	Ensemble Empirical Mode Decomposition
ELU	Exponential Linear Unit
EMD	Empirical mode decomposition

FAEK	Fast Fourier Transform, Auto-Encoder, K-Mean
FDP	Fault Diagnosis And Prediction
FEMTO-ST	Franche-Comté Électronique Mécanique Thermique et Optique - Sciences et Technologies
FFT	Fast Fourier Transform
FN	False Negative
FP	False Positive
FTF	Fundamental Train Frequency
GRU	Gated Recurrent Unit
HI	Health Indicators
IMFs	Intrinsic Mode Functions
LMD	Local Mean Decomposition
LSTM	Long Short-Term Memory
MA	Moving Average
MAPE	Mean Absolute Percentage Error
MCC	Multi-Scale, Channel Attention Classification Model
MCFCNN	Multi-scale processing, Channel attention, Feature enhancement, Convolutional Neural Network
MLGTCN	Multi-Layer Gated Time Convolutional Network
MSE	Mean Squared Error
MTA-CNN	Multitask Attention Convolutional Neural Network
NAS	Neural Architecture Search
PHM	Predictive Health Management
QCNN	Quadratic Convolutional Neural Network
ReLU	Rectified Linear Unit

RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
RNN-HI	Recurrent Neural Network-Based Health Indicator
RUL	Remaining Useful Life
SGD	Stochastic Gradient Descent
SNR	Signal-To-Noise Ratio
STFT	Short-Time Fourier Transform
TN	True Negative
TP	True Positive
t-SNE	t-Distributed Stochastic Neighbor Embedding
VCN	Vision Convnet
VMD	Variational Mode Decomposition
W1DCNN	Width-1 Dilated Convolutional Neural Network
WPT	Wavelet Packet Transform
WT	Wavelet Transform
WVD	Wigner-Ville Distribution

CHAPTER 1

INTRODUCTION

1.1 Research Overview

Mechanical equipment plays a crucial role in human society, having facilitated numerous advancements throughout history. The emergence of steam engines revolutionized production methods, while subsequent developments in electrical and information technologies laid the foundation for modern society. Mechanical equipment is widely utilized across various industries, from basic agriculture to cutting-edge fields such as high-speed rail, aerospace, robotics, and military applications, playing a significant role in modern civilization.

However, with the advancement of modern mechanical equipment, the structures have become increasingly complex. Faults in components of mechanical equipment may render the entire system inoperable, and failures of critical components can lead to significant casualties and economic losses (Rai & Upadhyay, 2016). For instance, in 2015, a collapse incident occurred in a Repower wind turbine in France due to a main shaft failure, resulting in substantial economic losses. In 2018, the US aircraft carrier “Ford” experienced main thrust bearing failure during an exercise, leading to the interruption of the exercise and requiring a return for repairs. In February 2020, a Canadian Aviation A319 flight lost its main landing gear wheel shortly after takeoff due to suspected bearing failure, fortunately resulting in no casualties but still causing severe economic losses (He & He, 2017). These tragic incidents remind us that prevention is better than cure.

China attaches great importance to mechanical fault diagnosis, as outlined in the “National Medium and Long-Term Science and Technology Development Plan”, which emphasizes the development of fault diagnosis and prediction technologies using information technology to enhance equipment reliability and safety (Zhang et al., 2018). Therefore, timely detection of mechanical equipment anomalies, quantitative evaluation of system health status, and implementation of effective maintenance measures are crucial for ensuring the operational safety, reliability, and economic viability of systems.

1.2 Research Background

Fault diagnosis is an indispensable step in maintaining the normal operation of modern mechanical equipment, and bearings are essential components, especially in rotating machinery. Bearing failures account for a significant proportion of mechanical equipment failures (Qiu et al., 2023). Therefore, ensuring the normal operation of bearings is crucial. Early bearing fault diagnosis used to rely on manual contact with the bearing using a listening rod, followed by auditory judgment of the bearing’s health. However, this method is highly subjective and easily affected by environmental interference (Zhong et al., 2022).

With the development of fault diagnosis methods, various bearing fault detection methods have emerged, and bearing fault diagnosis technology is constantly evolving. Traditional algorithms for bearing fault diagnosis require complex signal filtering and noise reduction processing, as well as feature design and selection processes. However, with the advent of Industry 4.0 and the era of big data, manually extracting fault features becomes increasingly impractical in the face of massive data. In recent years, with the advancement of information technology, data-driven bearing fault diagnosis methods have

gradually become a research hotspot (Cerrada et al., 2018). However, most data-driven methods typically deal with relatively simple scenarios of bearing fault diagnosis and focus more on identifying types of bearing faults.

In actual situations, bearings operate under different conditions, and the criticality of different bearings in equipment varies. Simply detecting the types of bearing faults is insufficient. It is also necessary to understand the severity of bearing faults to better identify bearing faults and estimate bearing life. Particularly, bearing fault diagnosis under complex conditions such as high noise and different operating conditions poses significant challenges. The data distribution of bearings under different conditions is not consistent, which violates the common assumption of consistent data distribution between training and testing sets in machine learning. At this point, the difficulty of bearing fault diagnosis based on manually designed features will dramatically increase. With the advancement of information technology, data-driven bearing fault diagnosis methods have become a solution, and deep learning, as one of the frontrunners, demonstrates more powerful feature extraction and processing capabilities, making bearing fault diagnosis and prediction no longer an unattainable goal.

With deep learning, we can break free from the constraints of manually designing features and directly learn bearing fault features from big data to address challenges such as high noise and different operating conditions. This automation not only improves the accuracy of bearing fault diagnosis but also makes bearing life prediction more intelligent. The potential application and practical value of this technology are exciting. Furthermore, knowledge of predicting the health stages of bearings is also essential. Understanding the health status of bearings helps take preventive measures in advance to avoid losses caused

by sudden failures. Therefore, data-driven methods for predicting bearing health stages are also one of the current research hotspots.

1.3 Research Problem and Hypotheses

Fault diagnosis and prediction are effective methods for equipment operation and maintenance management, initially mainly applied in aerospace, military equipment, and other fields. With the continuous development of fault diagnosis and related technologies such as data collection and analysis, fault diagnosis methods have gradually been applied in civil aviation, equipment manufacturing, and other fields, and further promoted and popularized in multiple industries. Traditional fault diagnosis methods include empirical model-based methods, data-driven methods, and physics-based methods. However, these methods have many limitations.

Traditional methods rely on manual feature extraction, and the preprocessing model design of signals is complex, leading to heavy reliance on expert experience. When the operating conditions are complex, these methods cannot solve the problem of differences in data distribution, posing significant challenges to fault diagnosis. The diagnostic models need to undergo tedious design steps, different fault types require different model structure parameters, and in complex operating conditions, the diagnostic effect is poor when the signal is contaminated by noise.

Therefore, current research on bearing fault diagnosis is focusing on deep learning methods. Deep learning possesses deep structures and powerful non-linear feature extraction capabilities, allowing for the direct extraction of fault features from raw data, thereby achieving end-to-end fault diagnosis and avoiding the cumbersome feature extraction and classification steps in traditional methods.

Deep learning methods can handle large-scale datasets and complex signal situations, demonstrating robustness and generalization performance. Currently, various types of sensors can be installed to collect bearing operational signals, including vibration (Chen & Lee, 2020), temperature (Gunerkar et al., 2019), and electrical current (Bouras et al., 2020), among others.

Vibration signal monitoring has received extensive research and application due to its independence from mechanical structures and simplicity of testing. The collected vibration signals are generally one-dimensional time series signals. Based on the input dimension of the network, there are currently two types of bearing diagnostic methods: one directly processes one-dimensional data, with the One-Dimensional Convolutional Neural Network (1DCNN) model proposed in literature (Eren et al., 2019; Ozcan et al., 2021; H. Wang et al., 2020) proving the high generalization performance of the 1DCNN model. There are other proposed methods that convert one-dimensional data into two-dimensional data for processing (Wan et al., 2020; C. Wang et al., 2021; Xu et al., 2019). These approaches transform the time-domain vibration signals of bearings into two-dimensional grayscale images containing fault features, which are then input into 2DCNN models. The trained models can effectively identify fault types and severity.

Comparatively, the 1DCNN network ability to directly process one-dimensional data is more convenient, while converting 1D to 2D may lose some information. Therefore, Zhang et al. (2017) proposes a Width 1-Dimensional Convolutional Neural Network (W1DCNN) model with two main characteristics: the first layer uses convolutional kernels with a width of 64, and the other layers use a network structure with 3x3 small convolutional kernels, achieving high accuracy

in experiments. Due to the good diagnostic performance of the W1DCNN network, researchers have proposed many improved network models based on it (Shenfield & Howarth, 2020; C. Wang et al., 2021). With the rise of attention mechanisms, introducing attention mechanism modules into traditional Convolutional neural network (CNN) architectures can enhance the network's expression ability. Therefore, some researchers (H. Wang et al., 2020; Y. Wang et al., 2021) (Wang et al., 2020; Wang et al., 2021) have introduced attention mechanism modules into fault diagnosis tasks, enabling traditional CNN models to achieve higher classification accuracy.

Although deep learning has achieved good results in bearing fault diagnosis research, current deep learning models still have some issues. At present, many deep learning models perform well on clean signal and consistent data distribution datasets. However, when the signal contains a lot of noise, the model's classification performance will sharply decline. For example, in tasks where noise is added to the CWRU dataset with a signal-to-noise ratio of -6, the classification accuracy of AResNet (Zhong et al., 2022), W1DCNN (W. Zhang et al., 2017), and Quadratic Convolutional Neural Network (QCNN) (Liao et al., 2022) is 61.2%, 66.5%, and 71.1%, respectively. The classification accuracy of these models in noisy environments still needs to be improved. Noise is ubiquitous in actual industrial environments, making it crucial to efficiently diagnose signals containing multiple noise sources.

The research presented in this thesis takes rolling bearing vibration signals as the focus, using deep neural networks to establish end-to-end models that directly use raw time-domain signals as inputs, without the need for complex filtering and

feature selection steps, and directly output the final results. Different deep neural networks are proposed for high-noise and different operating conditions in complex environments. Through experimental comparisons, the superiority and effectiveness of the proposed methods are verified, providing a new solution for end-to-end intelligent bearing fault diagnosis in complex conditions, with practical significance. At the same time, in Predictive Health Management (PHM) of bearings, this thesis transforms the regression problem of predicting the remaining life of bearings into a classification problem. Instead of simply predicting the remaining life of bearings, we are more concerned about which stage of its life cycle the bearing is in, in order to plan and perform maintenance and management more effectively. To achieve this goal, the entire life of the bearing can be divided into different stages, and by using the given bearing data, the current stage of the bearing's life can be predicted. This will help determine appropriate maintenance strategies and operational measures, enabling suitable preventive maintenance or repair measures to be taken at different stages.

Based on the above-mentioned problem, the thesis proposes the following three hypotheses:

- i. The proposed MCFCNN model incorporates data augmentation and preprocessing methods which improve robustness and generalization, enabling accurate bearing fault diagnosis in noisy environments.
- ii. The MCFCNN model demonstrates consistent high diagnostic performance across varying operating conditions, when tested on actual bearing data.

- iii. Combining the FAEK segmentation approach with the MCC classification model enhances the prediction accuracy of bearing life stages, providing a reliable framework for preventive maintenance.

1.4 Research Questions

In recent years, there have been many studies on bearing fault diagnosis in the field of deep learning. However, most existing research focuses only on identifying the types of bearing faults, and research on end-to-end solutions for complex scenarios remains to be further improved. For instance, early studies primarily focused on diagnosing the types of bearing faults, with less attention given to diagnosing the severity of faults and addressing bearing fault diagnosis under complex conditions such as high noise and different operating conditions. Thus, this research posed the following three questions:

- i. How to achieve end-to-end intelligent bearing fault diagnosis?
- ii. How effective is the proposed model in diagnosing bearing fault, under strong noise and different operating conditions?
- iii. How to predict bearing degradation for effective bearing maintenance management?

Formulation of research question (i) arised from the need to improve traditional fault diagnosis method which heavily relies on manual processes such as complex denoising, filtering, and feature extraction. This process requires extensive expertise and consumes significant manpower. Moreover, the quality of feature extraction directly affects the effectiveness of fault diagnosis. When faced with complex conditions, the difficulty of manual feature extraction increases sharply. Deep learning methods can effectively utilize massive data to automatically learn bearing fault features without the

need for manual extraction, holding the potential for achieving end-to-end intelligent diagnosis. However, existing deep learning methods for bearing fault diagnosis still need improvement in feature processing capabilities, adaptability, computational efficiency, and interpretability, requiring further research and refinement.

A test platform was built to answer research question (ii). To validate the proposed model, the test platform was used to collect bearing fault signals, acquiring bearing operation data, and then validating the diagnostic effectiveness of the proposed model. Emphasis is placed on discussing the diagnostic effectiveness under strong noise and different operating conditions, which is the novel contribution of this research. Noise is an inevitable interference in industrial environments, and it is essential to extract information related to bearing faults effectively from signals containing noise interference. In current research, there is still considerable room for improvement in the actual effectiveness of end-to-end bearing fault diagnosis under noise interference, especially in diagnosing faults under strong noise interference. Additionally, commonly used fault diagnosis methods assume that training and testing data come from the same operating conditions, which may not always be the case in different operating condition scenarios. Furthermore, labeling data for each new operating condition for retraining is time-consuming and expensive. Therefore, bearing fault diagnosis under different operating condition scenarios is one of the urgent challenges to be addressed.

Research question (iii) was formulated to address bearing degradation prediction, which is crucial for precision maintenance. This thesis focused on predictive maintenance of bearings, which proposed a method that categorised bearings life into four degradation stages. Then, prediction on which stage of degradation the bearing is in at a given time

point was done, hence provides corresponding maintenance strategies based on different degradation stages.

1.5 Research Objectives and Research Scope

To address the formulated research questions, the objectives of this research was outlined as follows:

- i. To build a deep learning model that performs bearing fault diagnosis tasks in the presence of strong noise interference. In achieving this objective, the investigation of effective data augmentation and preprocessing methods to improve the robustness and generalization ability of the model was also done. Validation on the performance of the proposed model, named Multi-scale Convolutional Neural Network with Attention Mechanism for Feature Enhancement (MCFCNN), was done using public datasets.
- ii. To test the proposed model on actual bearing performing under different working conditions and strong noise environments. To achieve this objective, a bearing fault diagnosis test bench named Bearing Fault Test System (BFTS) was built. Bearing fault data was collected on the test bench and the advantages and effectiveness of the MCFCNN model was further verified.
- iii. To develop a deep learning approach for bearing degradation prediction. To achieve this objective, a model consisting of two parts was designed. The first part, named Fourier Transform Autoencoder with K-means (FAEK), combines a convolutional autoencoder segmentation network to effectively distinguish the degradation process of bearings. The latter part, a high-performance classifier called Multi-scale Processing Channel Attention Classification (MCC) was proposed, which proven to achieve accurate

classification of bearing life stages. This provides more comprehensive and reliable support for bearing health assessment and maintenance decisions.

After clarifying the research objectives, this section will provide a detailed explanation of the scope of this thesis to better define the methods employed, the research subjects, and their specific conditions. This thesis is primarily based on deep learning networks, focusing on the fault identification and prediction of rolling bearings. In terms of fault identification, the research covers 10 types of faults, with a particular emphasis on recognition rates under high signal-to-noise ratios and across varying operating conditions. The datasets used include Case Western Reserve University (CWRU) and BFTS. In the fault prediction task, the study divides the entire lifecycle of bearings into four stages, using the An Experimental Platform for Bearings Accelerated Degradation Tests (FEMTO-ST) dataset for sampling.

1.6 Significance of the Research

This research applied the deep learning approach on bearing fault diagnosis and prediction. The proposed models, MCFCNN improves the accuracy of bearing fault diagnosis while the FAEK and MCC models enhance the ability of bearing fault prediction. The proposed deep learning models can automatically learn and update, adapt to the status of rolling bearings under different working conditions and noisy environments, which provide as a novel contribution in this field of research. Within the industry, the proposed approach potentially reduces the workload of maintenance personnel, improve the efficiency and accuracy of fault detection, and reduce the losses caused by equipment failures. At the same time, it can reduce maintenance costs and improve production

efficiency, achieve timely maintenance and fault prevention of equipment, and improve equipment reliability.

1.7 Thesis Outline

The thesis is divided into seven chapters. Chapter 1 serves as an introductory section, providing a brief overview of the research topic background and problem statement, outlining the content and objectives of the study. This chapter offers a comprehensive outline of the research.

Chapter 2 is a literature review, summarizing the current research status of bearing fault diagnosis and prediction.

Chapter 3 is the methodology section, primarily introducing the key technologies of deep learning, the introduction of bearing datasets, evaluation metrics for fault diagnosis and prediction, visualization techniques, and the research steps for fault diagnosis and prediction.

Chapter 4 constructs a bearing fault diagnosis model for bearing fault diagnosis tasks in high-noise scenarios, with the dataset utilizing publicly available bearing data (CWRU).

Chapter 5 establishes a bearing fault data collection experimental platform, validating the effectiveness of the model proposed in Chapter 4 with the collected BFTS dataset. The focus is on diagnostic performance in high-noise and cross-condition task scenarios.

Chapter 6 focuses on bearing degradation prediction tasks. It constructs another deep learning model for bearing degradation prediction and maintenance. The dataset utilizes publicly available bearing data (FEMTO-ST).

Chapter 7 provides a summary and conclusion of the entire thesis, as well as proposing future research directions.

CHAPTER 2

LITERATURE REVIEW

2.1 Chapter Overview

This chapter reviews topics relevant to the background of this research. There are two main topics discussed in this chapter i.e. traditional fault diagnosis methods and deep learning methods in bearing fault diagnosis.

2.2 Traditional Fault Diagnosis Methods

2.2.1 Time Domain-based Bearing Fault Diagnosis Method

The vibration signals of bearings contain rich fault features, and early bearing fault diagnosis methods mainly utilize the time-domain features of vibration signals. The statistical features of time-domain signals can characterize the characteristics of the signals to a certain extent. When a bearing fails, the time-domain statistical features of the vibration signal also change. Therefore, the fault condition of the bearing can be judged by analyzing the time-domain features of the signal. Slope and kurtosis can reflect the symmetry and pulse fluctuation characteristics of time-domain signals. When there is a fault in the bearing, the pulse component in the vibration signal usually increases, so they are commonly used in bearing fault diagnosis (Wang et al., 2016). However, dimensional parameters are sensitive to changes in operating conditions and are easily affected. Therefore, they can be normalized to dimensionless parameters to reduce the impact of changes in operating conditions. Common dimensionless parameters include clearance index, waveform index, pulse index, peak index, and kurtosis index (El-Thalji & Jantunen, 2015). Among them, the clearance index is commonly used to indicate wear conditions;

peak index, pulse index, and kurtosis index are commonly used to indicate the impact characteristics in the signal, describing early fault conditions. The waveform index is typically used to describe the late-stage features of faults. Common feature and index definitions are shown in Table 2.1.

Table 2.1: Common Time-Domain Features in Bearing Fault Diagnosis

Sequence	Parameter name	Parameter definition
1	Peak	$x_p = \max x_i $
2	Average	$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$
3	Average amplitude	$ \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i $
4	variance	$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$
5	Average square root	$x_{rms} = \sqrt{1/N \sum_{i=1}^N x_i^2}$
6	Fang root amplitude	$x_r = (1/N \sum_{i=1}^N \sqrt{ x_i })^2$
7	Slope	$\alpha = 1/N \sum_{i=1}^N x_i^3$
8	Cliffness	$\beta = 1/N \sum_{i=1}^N x_i^4$
9	Yield indicator	$L = \frac{x_p}{x_r}$
10	Waveform indicators	$S = \frac{x_{rms}}{ \bar{x} }$
11	Pulse indicator	$I = \frac{x_p}{\bar{x}}$
12	Peak indicators	$C = \frac{x_p}{x_{rms}}$
13	Cliffness indicator	$K = \frac{\beta}{x_{rms}^4}$

In addition to the method of using statistical features of time-domain signals for bearing fault diagnosis mentioned above, the autoregressive moving average model (ARMA) (Bayer et al., 2018) is also a commonly used approach. The ARMA method combines auto-regressive (AR) and moving average (MA) methods and can be used to model time-correlated sequences. Duan et al. (2019) combined ARMA with fuzzy C-means to achieve fault diagnosis of rolling bearings. Tang et al. (2021) utilized ARMA to assess the trend of gearbox faults. Han and Jiang (2016) combined AR models with variational mode decomposition and random forest classifiers to diagnose faults in non-stationary, non-Gaussian characteristic vibration signals of bearings. However, methods solely based on time-domain signals may not effectively utilize the characteristic frequencies of faults and are susceptible to interference. Therefore, many fault diagnosis methods based on frequency-domain and hybrid algorithms have been developed subsequently.

2.2.2 Frequency Domain-based Bearing Fault Diagnosis Method

Due to the different fault characteristic frequencies of various components in bearings, analyzing the frequency domain characteristics of vibration signals can provide a more intuitive understanding of the types of faults. Fast Fourier Transform (FFT) is a commonly used method for transforming time-domain signals into the frequency domain and is often used in bearing fault diagnosis. However, the fault characteristic signals of bearings are often modulated, and using only the spectrum cannot directly obtain the fault features. The impact vibration signals generated when the rolling elements pass through the crack are short in time and rich in spectrum, mixed with other vibration signals, and cannot be directly detected. Typically, this characteristic signal needs to be separated before the fault characteristic signal of the bearing can be found. Resonance demodulation,

also known as envelope analysis, is a commonly used method for demodulating signals. This method can separate the low-frequency impact vibration signals attached to the high-frequency amplitude-modulated signals, thereby enabling bearing fault diagnosis using fault characteristic frequencies (Leite et al., 2015). Wang et al. (2017) combined kurtosis and envelope analysis to locate bearing faults. (Yan et al., 2020) utilized fractional envelope analysis to extract weak bearing faults. They combined kurtosis with non-local mean weight envelope spectrum to enhance bearing impact fault features. Although resonance demodulation can effectively separate fault characteristic signals from modulated original signals, it is susceptible to noise, prone to mixing frequency effects, and requires manual specification of the bandpass filtering range, so it is often combined with other methods.

Cepstrum (Sousa et al., 2019) is another commonly used bearing fault diagnosis method that utilizes signal spectra. When computing the cepstrum, the power spectrum of the signal is first calculated, then the logarithm of the power spectrum is taken, and finally, the inverse Fourier transform is performed to obtain the final result. Using cepstrum, periodic components on the original spectrum can be obtained, separating the original two multiplied signals and achieving signal demodulation. Cepstrum is not sensitive to the power size of the signal and does not need to consider the attenuation effect of the signal too much. Borghesani et al. (2013) applied pre-whitening to improve the diagnostic accuracy of bearing cepstrum under variable speed conditions. Nishat Toma and Kim (2020) first extracted fault features using cepstrum and then used gradient boosting algorithm to identify bearing fault types. Randall (2017) summarized the development history of cepstrum and its application in mechanical fault diagnosis.

2.2.3 Time-Frequency Domain-based Bearing Fault Diagnosis Method

Bearing fault signals exhibit non-stationary characteristics, and frequency-domain-based methods cannot handle time-varying, non-stationary signals well. Additionally, the spectrum is easily disturbed by noise with low spectral resolution, making it difficult to identify bearing fault signals under complex conditions. Modern signal processing methods are capable of handling time-varying non-stationary signals and are therefore widely used in bearing fault diagnosis. Common algorithms include Short-time Fourier transform (STFT) (Dai et al., 2020), Wigner-Ville Distribution (WVD) (Rongjie et al., 2011), Wavelet Dransform (WT) (Rioul & Vetterli, 1991), Empirical Mode Decomposition (EMD) (Chen et al., 2019), Hilbert-Huang transform (Elbouchikhi et al., 2017), etc.

STFT operates on non-stationary signals using a sliding window, producing signals that can be considered stationary. Therefore, Fourier spectrum transformation can be applied. However, the selection of the window function and the width of the window greatly influence subsequent fault analysis, making it relatively difficult to choose a suitable window width. Cocconcelli et al. (2012) utilized encoder segmentation to handle bearing signals at different speeds, followed by fault diagnosis using STFT. Wang and He (2015) utilized time-frequency manifolds and STFT to achieve denoising and resolution enhancement of fault signals, thereby improving bearing fault diagnosis performance. Cheng et al. (2017) first extracted time-frequency domain features using STFT, and then inputted the dimensionality-reduced features into support vector machines to enhance fault classification accuracy.

The WVD has high time-frequency resolution and time-frequency aggregation characteristics, making it a powerful feature representation for time-varying signals.

However, when applied to multi-component signal analysis, the WVD may introduce cross-terms without any physical meaning. Therefore, many studies combine it with other methods to eliminate the influence of cross-terms. Zhou et al. (2011) combined cyclic spectrum density and WVD to suppress cross-term interference while improving algorithm noise resistance. Ge et al. (2020) first suppressed signal noise using empirical mode decomposition, then utilized smoothed pseudo WVD features to realize bearing fault diagnosis.

WT can be applied to non-stationary signals, enabling the detection of transient components within the signal. Therefore, it is commonly used in fault diagnosis. WT employs finite-length wavelet basis functions instead of trigonometric basis functions, allowing not only the extraction of signal frequency but also obtaining the corresponding time, thereby obtaining the signal's time-frequency spectrum. Compared to the fixed-width window functions used in STFT, WT can achieve variable time-frequency windows by using orthogonal wavelet bases. Continuous Wavelet Transform (CWT) can directly utilize wavelet transform parameters as the basis for fault diagnosis. Huo et al. (2017) used CWT to decompose the original signal, then combined kurtosis indicators to reconstruct the signal and utilized envelope analysis for early diagnosis of bearing faults.

Discrete wavelet transform (DWT) simplifies the information redundancy of CWT using the Mallat algorithm, improving computational speed. Khanam et al. (2014) utilized DWT to decompose vibration signals to detect impact signals generated by rolling elements entering and leaving faulty areas, estimating fault size based on the time difference of impact signals. Patel and Giri (2015) first extracted vibration signal features

using probability density distribution and DWT, then further analyzed features using Fourier transform for bearing fault type identification.

Wavelet Packet Transform (WPT) (Ezhilarasan et al., 2007) decomposes both low-frequency and high-frequency features of signals compared to wavelet transform, thus retaining more signal features. For bearing vibration signals, high-frequency components often include fine details such as pulse characteristics of faults, making WPT more prominent in bearing fault diagnosis. Zhang et al. (2017) first denoised the original signal by determining WPT decomposition parameters based on signal-to-noise ratio and mean square error, then achieved higher bearing fault diagnosis accuracy by selecting optimal fault indicators using random forest method. Wang et al. (2018) further extracted features from WPT results using sparse representation, then classified the extracted features using nearest neighbor classifier for bearing fault classification.

Second-generation wavelet transform (Simoens & Vandewalle, 2003) considers the entire scale of the signal, achieving better denoising effects. Li et al. (2011) first extracted signal features using second-generation wavelet transform, then manually selected some features for subsequent support vector machine input, achieving fault diagnosis of outer and inner races. Although methods based on wavelet transform and its improvements can effectively extract features from non-stationary signals, especially those containing impact response bearing fault signals, the selection of wavelet basis function parameters has a significant impact on subsequent results. The ability to extract fault features is stronger when selecting wavelet basis functions similar to signal features, and vice versa.

EMD is one of the most powerful time-frequency analysis techniques in modern signal processing methods. It decomposes signals into nearly orthogonal components

called Intrinsic Mode Functions (IMFs). IMFs, akin to vibration modes in physics, represent the inherent characteristics of the signal, making EMD adaptive. Compared to WT, EMD does not require manual specification of filter kernel parameters, thus exhibiting better adaptability. However, the basis functions in EMD are not strictly orthogonal, leading to mode mixing and endpoint effects.

To address these issues, Ensemble Empirical Mode Decomposition (EEMD) (Wu & Huang, 2009) introduces Gaussian white noise into the original signal, projecting different scale components of the mixed signal onto a uniform distribution reference scale established by the white noise. This compensates for the missing scale information in the original signal and makes the distribution of signal extrema more uniform. Since Gaussian white noise added is zero-mean and multiple averaging calculations are required during IMF computation, the noise gradually diminishes its influence on the final result during the averaging process.

Hilbert-Huang Transform combines EMD and Hilbert Transform. It initially decomposes the original signal into nearly orthogonal IMFs and then extracts envelopes from these IMFs, further extracting the signal's features (Luo et al., 2017). Combining kurtosis index with Hilbert-Huang Transform makes fault features more apparent in the reconstructed signal envelope spectrum. Moreover, Local Mean Decomposition (LMD) (Smith, 2005) and Variational Mode Decomposition (VMD) (Dragomiretskiy & Zosso, 2014) have also gained attention and application in bearing fault diagnosis.

Ma et al. (2015) proposed using LMD for signal decomposition, followed by envelope extraction for fault feature extraction. Compared to EMD, each component in LMD has practical physical significance and suppresses negative frequencies, mode

mixing, and endpoint issues. VMD (Mohanty et al., 2014) addresses the exponential decay DC offset problem in EMD and has clear theoretical support. Compared to EMD, VMD achieves faster convergence and higher diagnostic accuracy.

2.3 Deep Learning Methods in Bearing Fault Diagnosis

Neural networks are capable of performing fault diagnosis and prediction (FDP) tasks without the need for expert experience because they bypass the cumbersome step of manually extracting features, instead using raw vibration signals as model inputs, which is a typical method of intelligent feature extraction. However, when coping with complex bearing FDP problems, the network structure of conventional neural networks is relatively simple and the number of network layers is low, so they can only extract shallow features and their level of nonlinear operation is relatively low. To address this issue, the concept of deep learning has been proposed. Unlike traditional neural networks, deep learning usually has more complex structures and deeper network layers. The bottom-layer features are mapped to high-dimensional space through layer-by-layer nonlinear mapping, and feature representation gradually becomes hierarchical and abstract with the increase of network layers, ultimately forming hierarchical and abstract nonlinear features (Janiesch et al., 2021). In addition, increasing the number of network layers, the nonlinear fitting ability also increases, which enables deep learning to handle more complex high-dimensional data. Convolutional neural network (CNN), auto-encoder (AE), deep belief network (DBN), and recurrent neural network (RNN) are the most common neural networks now applied to handling FDP since the rapid growth of deep learning.

2.3.1 Convolutional Neural Network (CNN)

CNN was a traditional variety of multilayer neural networks (Chen et al., 2015) that were first proposed in 1994. Due to their characteristics of weight sharing, CNN have a simple structure and powerful feature extraction ability, making them commonly employed in the area of bearing FDP. CNN is an effective end-to-end learning system due to its hierarchical structure and the fact that each layer of the network harvests two-dimensional data as inputs, which enables intelligent learning of the defect aspects of vibration signals. Figure 2.1 displays the fundamental structure of CNN.

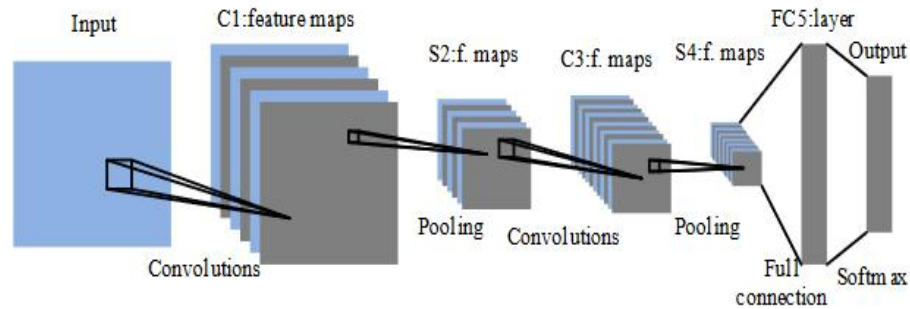


Figure 2.1: An architecture of convolutional neural network (Chen et al., 2015)

Ince et al. (2016) proposed a novel motor condition monitoring system that integrates feature extraction and classification using a one-dimensional CNN model for end-to-end fault monitoring. Proper training enables the model to be applied to any motor data for fault detection, achieving high generalization capability and simplifying monitoring operations by avoiding manual parameter tuning and feature extraction. Eren et al. (2019) applies the improved 1DCNN model to the CWRU dataset, showing the 1DCNN model's outstanding generalisation performance while obtaining 93.9% accuracy. The model is evaluated for its simple architecture, making it appropriate for monitoring and detecting faults in real-time, its ability to perform without signal pre-processing, and its

potential limitations in training iterations and dataset size. To deal with the problem of small data sets, Ozcan et al. (2022) proposes a multi-channel data enhancement carrying fault identification system. This technique makes advantage of vibration data gathered from various accelerometers mounted on the bearing, and data fusion greatly improves the model's robustness. By contrasting the results with conventional fault identification results based on single-axis data using the IMS bearing vibration data set, the proposed method is confirmed. Based on the literature above, a better WDCNN model is proposed in (Zhang et al., 2017), where the size of the convolutional kernels is kept constant during the convolutional operation. The model has two main features: the first layer uses 64 wide convolutional kernels, and the other layers use a network structure with 3x3 small convolutional kernels. Similarly, Wang et al. (2021) suggested a network architecture known as vision ConvNet (VCN) that used a single layer big convolutional kernel network and a multilayer small convolutional kernel network to extract low- and high-frequency characteristics, respectively. Finally, to achieve precise fault classification under challenging operating conditions, the features were flattened and stitched together to create a feature map. The VCN model achieved higher fault classification accuracy than traditional CNN under different operating conditions and ran more easily throughout the training phase, demonstrating increased model robustness. In bearing fault diagnosis tasks, researchers focused on both preprocessing the data and optimizing deep learning models. Literature (Toma et al., 2022; Xu et al., 2019) belong to the former.

In Xu et al. (2019), CWT was used to transform the time-domain vibration signals of the bearings into two-dimensional grayscale images including rich fault information. Then, the image data was fed into the LeNet-5 model. The trained model could effectively recognize the types and severity of faults. In Toma et al. (2022), a new

technology based on the Graphical Application Framework image encoding was proposed. By segmenting the current signal and converting it into image data, it was input to the CNN model. The model had a short training time, fast convergence, and an overall accuracy of over 99% in experiments.

Literature (Wan et al., 2020; Wang et al., 2020; Wang et al., 2022) belong to the latter. The conventional LeNet-5 network has been enhanced by Wan et al. (2020). After each convolutional layer, batch normalisation layers were added to speed up convergence. Dropout strategy was added after the fully connected layer to prevent overfitting and enhance generalization ability. The outcomes demonstrated that the enhanced model's performance had increased, especially in small sample and strong noise environments. Sun et al. (2020) suggested a multi-scale maximum pooling module to minimise information loss during maximum pooling. By parallelizing three pooling layer branches, using 4×4 , 8×8 , and 16×16 pooling layers for downsampling respectively, and finally stitching the three branches together to obtain the feature map, the original signal's spatial correlation between data points spaced at various intervals was preserved. When the signal-to-noise ratio (SNR) was -5, experimental findings indicated that the Attention-based Local Structure Convolutional Network (ALSCN) model had a greater diagnosis accuracy and had stronger generalisation capabilities for 5120-dimensional sequences. Wang et al. (2022) introduced an attention mechanism module to the traditional CNN model, which was the core module of Multi-Task Attention Convolutional Neural Network (MTA-CNN). The model might decide which features ought to be kept and which ought to be removed. This architecture could achieve information sharing for many activities by automatically learning task-specific features from globally shared features. Experiments were conducted on bearing datasets and motor bearing datasets.

Bearing fault prediction and diagnosis are closely related. Fault prediction involves modeling and analyzing historical data and operating conditions of bearings to anticipate potential faults in advance. Predicting the Remaining Useful Life (RUL) of bearings is a major emphasis in the research on bearing defect prediction. Wang et al. (2021) proposed a method that combines CNN with AE to predict the RUL of equipment. Firstly, an autoencoder network was used to create the bearings' Health Indicators (HI), which captured the features of the original data and obtained the degradation state model of the bearings. Then, the HI was used as the label for the original vibration data and input into the CNN prediction model. The model was tested on a dataset of deteriorating bearings and was successful in forecasting the RUL of the bearings. The gap between the bearings' actual and expected lifespan was assessed using the Root Mean Square Error (RMSE). The experiments conducted on the FEMTO-ST dataset achieved an optimal value of 0.054. Sanakkayala et al. (2022) used STFT for data preprocessing, followed by feature extraction using the VGG16 network for classifying the health condition. Finally, regression was employed for RUL prediction. The achieved RMSE on the NASA dataset was 10.71. Zhou et al. (2020) used a similar network framework (STFT+CNN) that can effectively predict the RUL of bearings. With an accuracy of about 99.45%, it can predict the bearing's lifespan over the course of its whole existence. With the continuous development of deep learning technology, deep neural network models have become increasingly complex. To address this issue, Ruan et al. (2023) employed neural architecture search (NAS) and weight-based ranking pruning techniques to reduce the complexity of the network in bearing RUL prediction. NAS is used to search for the optimal neural network structure by continuously optimizing the parameters and architecture of the network to improve its performance. Weight-based ranking pruning is a pruning technique that removes

unimportant connections and parameters from the network, reducing its complexity while maintaining model accuracy. The R-square parameter was utilised in the tests for bearing RUL prediction to assess how well the regression model fit the observed data. The experimental findings revealed that the suggested approach had a bearing RUL prediction task accuracy of 0.93, demonstrating that the model can successfully fit the observed data.

Overall, these approaches demonstrate the successful application of deep learning techniques in bearing FDP, further comparison of articles is discussed in Table 2.2.

Table 2.2: Comparison of CNN techniques for fault-bearing FDP research

Refs	Year	Method	Result	Data set
(Ince et al., 2016)	2016	1D-CNN	Acc:97.4%	Real-time motor
(Eren et al., 2019)	2019	1D-CNN	Acc:93.2%, 93.9%	CWRU, IMS
(Ozcan et al., 2022)	2021	1D-CNN	Acc:100%	IMS
(Zhang et al., 2017)	2017	WDCNN	Acc:100%	CWRU
(Wang et al., 2021)	2021	VCN	Average acc: 97.54%, 85.3%	CWRU, PU
(Xu et al., 2019)	2019	LeNet-5	Acc:97.38%	CWRU
(Sun et al., 2020)	2020	ALSCN	Average acc: 98.62%, 97.46%	CWRU,MFPT

Table 2.2: continued

(Toma et al., 2022)	2022	GAF+2D CNN	Acc: 99.58%	KAT bearing
(Wan et al., 2020)	2020	WPT-CNN	Average acc:99.81%	CWRU
(Wang et al., 2022)	2022	MTA-CNN	SNR=-5,acc:92.33% SNR=0,acc:98.64% SNR=5,acc:99.72%	CWRU
(Wang et al., 2021)	2021	DCAE+CNN	RMSE: 0.054	FEMTO-ST
(Sanakkayala et al., 2022)	2022	VGG16	RMSE: 10.71	NASA
(Zhou et al., 2020)	2020	STFT+CNN	Acc: 99.45%	IMS
(Ruan et al., 2023)	2023	CNN	R-square: 0.93	FEMTO-ST

2.3.2 Auto-Encoder

The Auto-Encoder (Yao et al., 2016) is a self-learning neural network with the same input and output, consisting of an encoding layer and a decoding layer. The encoding layer is composed of an input layer and a hidden layer, while the decoding layer is composed of a hidden layer and a reconstruction layer. The network model's encoding process involves mapping the high-dimensional input features into the low-dimensional features of the hidden layer using an activation function, and the decoding process entails reconstructing the low-dimensional features of the hidden layer into the output target using an activation function. The training process of the auto-Encoder is to build different

optimization functions to minimize the loss value between the target output and input.

Figure2.2 depicts the network architecture of the auto-Encoder.

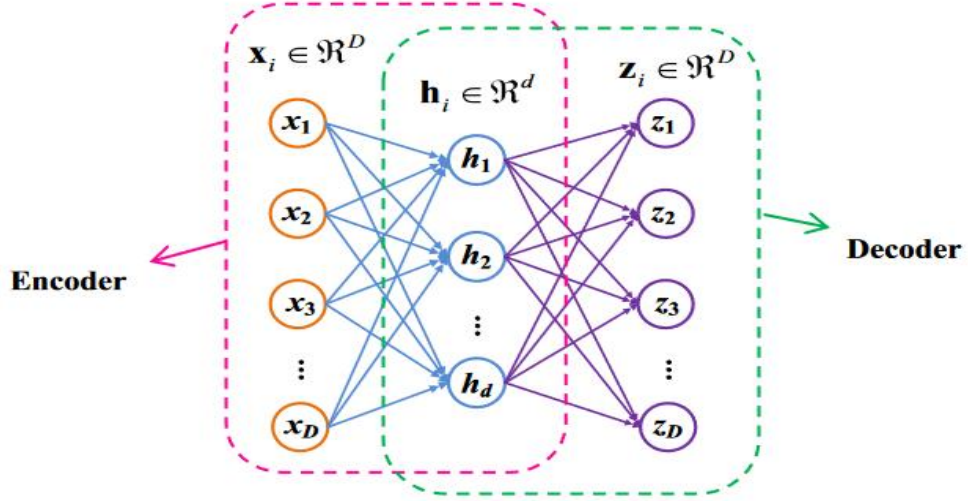


Figure2.2: An architecture of an Auto-Encoder (Yao et al., 2016)

The bearing fault dataset contains a wealth of information, involving diverse operating settings and health situations. How to better extract information from the dataset using deep learning models is a research direction in this field. In Jia et al. (2016), a DNN-AE model was proposed with a five-layer Auto-Encoders at the core for efficient classification of massive data. The approach demonstrated a 99.6% accuracy in fault diagnosis when adaptively mining existing fault features from the original signal. In Shao et al. (2017), a brand-new deep auto-Encoder system for diagnosing rotating equipment faults was proposed. In order to improve feature learning, maximum correlation entropy was utilised as the loss function. At the same time, the auto-Encoder's essential parameters were adjusted to meet the signal properties using an artificial fish swarm algorithm (Shao et al., 2018). The different series of features could be used to diagnose the information under different dimensions of the fault, and finally a combination strategy was designed to

be able to discriminate the type and severity of the fault. The method obtained good results with CWRU dataset.

In the cited papers above, the same distribution applied to both the training and test sets. However, in real-world applications, bearings frequently function in a variety of operating environments, and the distributions of training and testing sets obtained under such circumstances are typically different. Based on this, a new DTL model was proposed by Wen et al. (2019), which minimised the distributional difference between the feature of training data and testing data by applying the greatest mean discrepancy to the three-layer sparse auto-encoders used to extract features from the raw data. The results showed that DTL had a prediction accuracy of up to 99.82%. An effective fault diagnosis model adaptable to changes in working conditions and ambient noise was also proposed in Lu et al. (2017), namely Stacked Denoising Auto-Encoders (SDA). This model's primary idea involved creating a deep hierarchical structure with a greedy training transfer rule, sparse representation, and data erasure to produce high-order features with good robustness. Similarly, in Tong et al. (2020), a similar network architecture Deep Collaborative Stacked Auto-Encoder (DCSAE) was proposed, for better the network's capacity for learning new features, the loss function of the Auto-Encoders was primarily regularised using the sparse penalty and contraction penalty, and the cuckoo search algorithm (CS) was used to automatically optimise the hyperparameters. This model could effectively diagnose the fault categories and severity under different working conditions in the Anhui University of Technology bearing dataset.

In Xiong et al. (2020), a deep Auto-Encoders network consisting of multiple Auto-Encoder was proposed to address the inconsistency of fault types between training and

testing sets in actual working conditions of bearings. Moreover, the Wasserstein gradient penalty generative adversarial network, a brand-new technique for intelligent fault diagnosis, was employed. This network may help with the imbalanced fault data issue, accelerate model convergence, and enhance network stability. In Li et al. (2020), an Attention-based Temporal Latent Space Autoencoder (ATLSAE) model was proposed. Through the use of FFT, the original vibration signal was converted into the frequency domain. Transferable characteristics were then extracted from the frequency domain using SAE as a feature extractor. This method was effectively applied to diagnostic tasks with different fault types in target and source domains, and could separate target domain samples from new fault types. To solve the diagnostic tasks of small sample data, which was a challenge for deep learning techniques that strongly depend on large amounts of measurement data, a deep Auto-Encoders model that integrated multiple types of fault discrimination information was proposed in Mao et al. (2021). The core of this method was to optimize the loss function and design a regularizer using the highest association entropy to improve the representation capability of features. Meanwhile, a relationship matrix of fault types was introduced to visualise structural data among various defect kinds. This model achieved efficient diagnosis on the IMS and CWRU datasets. In Shi and Shang (2022), to enhance the effectiveness of early bearing fault identification, a diagnostic model based on deep auto-encoders and second-order cyclic autocorrelation was presented. The model uses denoising and shrinkage Auto-Encoders to recover the bearing's initial defect characteristics and the second-order cyclic statistics of the predicted vibration signal from Soca to suppress random noise. The bearing's initial fault characteristics could be improved by the model with more diagnostic robustness and accuracy. In Kancharla et al. (2022), the Multi-Layer Convolutional Auto-Encoder, K-Nearest Neighbors model was

proposed, which used the potential value of the Auto-Encoders as a robust feature for inter-conditional fault classification, capable of cross-condition and cross-machine fault diagnosis.

In the study of RUL prediction for bearings, Ren et al. (2018) proposed a novel feature vector based on the joint time-frequency wavelet features to effectively characterize the degradation process of bearings. To achieve feature compression and computation integration, a deep autoencoder was employed, which could preserve relevant information without increasing the size of the deep neural network (DNN). This method achieved good performance in the RUL prediction of bearings. Mean squared error (MSE) was employed in the studies as the evaluation metric to assess how well the prediction model performed. The MSE obtained on the FEMTO-ST dataset was 0.2, indicating relatively high prediction accuracy of the model. The DA+AE technique based on domain adaptation and autoencoder network is suggested by Ding et al. (2021) to successfully tackle the problem of adapting RUL prediction models learned from historical data to new operating conditions for bearings. The feature extraction model and RUL regression model, which were trained in the labelled source domain, are transferred to the unlabeled target domain using two methodologies in the method: parameter transfer and feature representation transfer. In this way, the historical model can be fully utilized and quickly adapted to new operating conditions. The goal of the parameter transfer strategy is to minimise the difference in feature distribution between the source and target domains by transferring and adjusting the feature extraction model's parameters through adversarial training on the source and target domains. The mean absolute percentage error (MAPE) is used to evaluate the relative error of the prediction model in the regression problem, and the optimal value on the FEMTO-ST dataset is 0.26. She et al. (2020) suggested a

technique for constructing the health indicator (HI) of rolling bearings based on sparse autoencoder with regularization. The HI is constructed by compressing and extracting features and using the RUL prediction model based on maximum likelihood estimation and particle filter. The health status and remaining service life of bearings can be reliably predicted using this method. Chen et al. (2021) proposed a novel method based on quadratic function deep convolutional autoencoder (DCAE) for constructing the HI of bearings. The method first preprocesses the raw bearing vibration signals with low-pass filtering and then inputs the processed signals into the DCAE neural network based on quadratic function. The experimental results show that the proposed method has better comprehensive performance in constructing the health indicator of bearings. Table 2.3 summarizes the results obtained from the above studies on Auto-Encoders.

Table 2.3: Comparison of auto-encoder techniques for FDP research

Refs	Year	Method	Result	Data set
(Jia et al., 2016)	2016	DNN-AE	Average acc:99.6%	Gearbox Rolling element bearing
(Zhao et al., 2017)	2017	AFSA-SDAE	Average acc:94.05%	Gearbox Electrical Locomotive roller bearing
(Shao et al., 2018)	2018	EDAE	Average acc:97.33%	CWRU
(Wen et al., 2019)	2019	DTL	Acc:99.82%	CWRU
(Li et al., 2020)	2020	ATLSAE	Average acc:99.88%	CWRU
(Xiong et al., 2020)	2020	WGGAN+DAE	SNR=50, Acc :94% SNR=30, Acc :88%	CWRU

Table 2.3: continued

(Lu et al., 2017)	2017	SDA	SNR=12,Acc :91.67 % SNR=16,Acc :94.18 % SNR=20,Acc :99.83 %	CWRU
(Tong et al., 2020)	2020	DCSAE	Average acc:98.57%	Anhui University of Technology rolling bearing
(Mao et al., 2021)	2021	SDIAE	Acc:96.13%,99.16%	CWRU, IMS
(Shi & Shang, 2022)	2022	Soca-DCT	Average acc:99.3%	CWRU
(Kancharla et al., 2022)	2022	MLCAE-KNN	Average acc: 99.9%,73.9%	CWRU, PU
(Ren et al., 2018)	2018	AE-DNN	MSE: 0.2	FEMTO-ST
(Ding et al., 2021)	2021	DA+AE	MAPE: 0.26	FEMTO-ST
(She et al., 2020)	2020	SAEwR	Error Percentage:12%	FEMTO-ST
(Chen et al., 2021)	2021	DCAE	MAPE: 0.16	FEMTO-ST

2.3.3 Deep Belief Network (DBN)

With both generative and discriminative models and visible layers as well as hidden layers, the deep belief network (DBN) is a deep hybrid network model (Lin et al., 2016). The DBN model is very adaptable, has excellent mapping capabilities, and can accurately represent any continuous function. The network can automatically extract features through

unsupervised learning in the field of bearing fault diagnosis, map the original signal to the feature space, and transform the feature space into a discriminant space through supervised learning, thereby performing fault type discrimination and successfully completing the fault classification task. DBN is a not parametric based on data model that does not rely on rigorous presumptions on the underlying data-generation process. The architecture of DBN is shown in Figure 2.3.

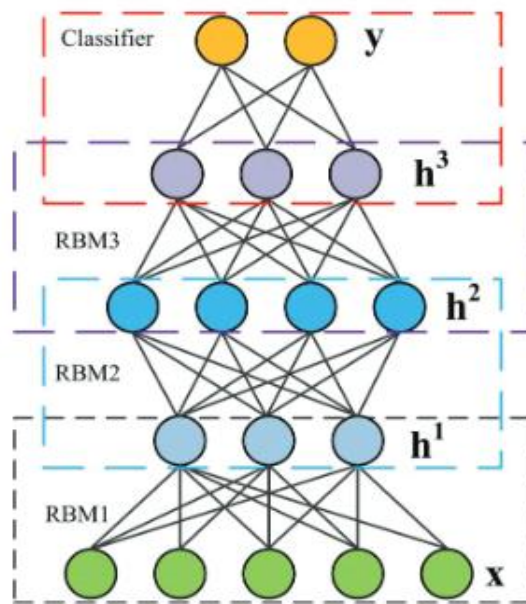


Figure 2.3: An architecture of Deep Belief Network (Lin et al., 2016)

It was first suggested to utilise DBN for jobs including bearing fault detection. The framework is composed of three stacked restricted Boltzmann machines, and hyperparameters are optimised using particle swarm optimisation (Shao et al., 2015). The accuracy of the experimental outcomes was good. In Shao et al. (2018), feature fusion technology was used to significantly enhance the standard of acquired features on the basis of a model pre-trained with DBN. The findings showed that the proposed method outperformed conventional approaches and standard DBN in terms of effectiveness and

dependability. In Wang et al. (2017), the dual-tree complex wavelet packet transform (DTCWPT) and adaptive DBN were used to create a novel model. By preparing data, DTCWPT can refine fault feature information, to enhance convergence speed and recognition precision, which is subsequently sent into the adaptive DBN. The CWRU dataset's bearing fault severity can be accurately identified by the model.

Yu et al. (2020) proposed a fault diagnosis scheme that combines DBN and the theory of Dempster-Shafer, and also achieves great degree of precision in assessing the seriousness of bearing defects. When it comes to parameter optimisation, the model employs a method that combines a hybrid genetic algorithm with particle swarm optimisation. To improve the generalization ability of the model, Guo and Zheng (2020) proposed a model that combines double sparse dictionary (DSD) with DBN. Every fault type is trained using the DSD algorithm, and then the extracted features are recombined and inputted into DBN. The model has been validated on two different datasets and has achieved good diagnostic accuracy and stability. For small sample bearing fault diagnosis tasks, a domain adaptive deep belief network was proposed by Che et al. (2020). The multi-kernel maximum mean discrepancy between the known operational state data and the new working state data is computed using the domain adaptive approach, and the loss function is then optimised. The findings demonstrate that the model is capable of above 92% accuracy in three distinct noise settings, with an average accuracy of 93.5% in a range of operating situations. Lv et al. (2021) proposed a multi-modal data fusion and deep belief network model. The time-domain, frequency-domain, and features learnt during unsupervised training of the original vibration signal samples were combined to create the multi-modal feature information of the model, increasing its robustness. Peng et al. (2022) proposed an Improved Sparrow Search Algorithm Deep Belief Network which used

Principal Component Analysis technique to make data less dimensional and achieved high diagnostic accuracy under small sample sizes. Jiao and Zheng (2022) presented a DBN-based model joint information fusion technology, which could diagnose joint bearing faults of industrial robots in real time. The technology combined wavelet energy entropy and DBN. The vibration signal was denoised, decomposed, and reconstructed through wavelet transform to increase the fault model's ability to accurately identify faults.

DBN can also be used in fault prediction. Peng et al. (2019) developed a Deep Belief Networks-based approach to building an unsupervised health index and paired it with particle filtering for RUL prediction. First, DBN is trained to extract system fault features. Then, a health index is constructed based on the separation between degradation states and fault states. A fuzzy inference method is used to enhance the particle filtering process by predicting the equipment's remaining usable life. The efficiency of the suggested strategy was proved through experimental validation on data from aircraft engines. Hu et al. (2020) introduced a bearing RUL prediction model based on Deep Belief Networks (DBN) and Diffusion Process (DP), considering prediction uncertainty. DBN is used to extract deep hidden features from monitoring signals during the feature extraction phase. The probability density function of the projected RUL is derived using the First-Hit Time representation in the RUL prediction section utilising a DP-based health index evolution model. This enables taking the forecast results' uncertainty into account. Experiments were used to validate the effectiveness and superiority of the suggested approach on bearing deterioration data. Xu et al. (2021) proposed a unique model for extracting bearing health index curves and enhancing their monotonicity that combines DBN with a new sigmoid zero local minimum point. In this approach, the initial health index curve is extracted using DBN, and oscillation is removed and the monotonicity of

the curve is improved by using the new sigmoid zero local minimum point. To limit the dependence of weights on the present health index value and hence eliminate oscillation, the average value based on the sigmoid function is specifically used. Experimental validation demonstrated the superiority of the model.

Overall, these studies highlight the application of DBN networks in FDP. Table 2.4 summarizes the results obtained from the above studies on Deep Belief Network.

Table 2.4: Comparison of FDP studies using Deep Belief Network

Refs	Year	Method	Result	Dataset
(Shao et al., 2015)	2015	DBN	Average acc:87.45%	CWRU
(Shao et al., 2018)	2018	DBN+LPP	Average acc:90.6%	Rolling bearing fault
(Shao et al., 2017)	2017	DBN	Acc:95.22%	CWRU
(Yu et al., 2020)	2020	DBN+DS	Acc:99.69%	Qingdao University of Technology rolling bearing
(Guo & Zheng, 2020)	2020	DBN+DSD	Acc:98.12%,95.63%	CWRU, IMS
(Che et al., 2020)	2020	DA-DBN	Average acc:92%(Noise) 93.5%(Variable working condition)	CWRU
(Lv et al., 2021)	2021	MDF-DBN	Acc: 98.5%	CWRU
(Peng et al., 2022)	2022	ISAA-DBN	Average acc:93.89%	Mill failure bearings
(Jiao & Zheng, 2022)	2022	WT-DBN	Average acc:98.01%	joint bearings of robots

Table 2.4: continued				
(Peng et al., 2019)	2019	DBN+HI	Accuracy:63%	aircraft engine
(Hu et al., 2020)	2020	DBN	RMSE:0.29	FEMTO-ST
(Xu et al., 2021)	2021	DBN – SZLMP	RMS:0.03	FEMTO-ST

2.3.4 Recurrent Neural Network (RNN)

The application of recurrent neural network (RNN) (Mikolov, n.d.) has demonstrated outstanding performance in various fields, for instance, natural language processing (Jelodar et al., 2020; Puiu & Brabete, 2019), speech recognition (Agarwal et al., 2022; R. Tang et al., 2018), and image description (Li et al., 2018; Yang et al., 2021). Due to its unique network architecture, RNN is capable of effectively capturing the temporal dependencies in sequential data. However, the issue of gradient vanishing arises when RNN processes long sequences, limiting its capacity to understand long-term dependencies. In order to address this problem, long short-term memory (LSTM) and gated recurrent unit (GRU) are two RNN versions that have been developed. These variants possess specialized structures such as memory cells and gating units, enabling them to better capture the long-term dependencies in sequential data. Figure 2.4 depicts the RNN's architecture.

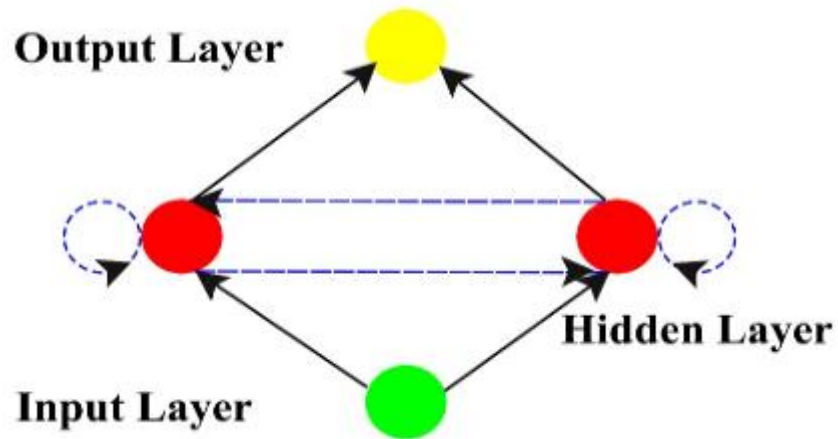


Figure 2.4: An architecture of Recurrent Neural Network (Li et al., 2018)

Due to the fact that bearing fault diagnosis data often exhibits time series characteristics in the form of vibration signals, RNN have been frequently used for bearing fault identification applications. In Li et al. (2018), a deep recursive neural network intelligent deep learning method was proposed. The input is a spectrum sequence signal, and the DRNN is formed by a series of recursive hidden layers that automatically acquire features from the input spectrum sequence. This model achieved good accuracy. In Shenfield and Howarth (2020), a novel RNN-WDCNN was proposed, which merged parts of RNN and CNN to extract long-range relationships in time series data and decrease high-frequency noise in input data. The results of experiments on the CWRU dataset demonstrated that the model outperformed classic CNN approaches in domain adaption and noise suppression tasks. Moreover, Zhiwei (2022) designed a 1DCNN-GRU model that combined the spatial processing capabilities of CNN with the time series processing capabilities of GRU. The CNN technique was utilized to extract feature information out of the bearings, which was then diagnosed using GRU. The recognition precision of the model exceeded 99%. Nie and Xie (2021) introduced a structure called Normalized Recurrent Neural Network for robust fault detection in noisy situations. The model

incorporated a forward cross-entropy loss to handle noisy labels effectively. The usefulness of the technique was validated on four separate datasets, each including increasing amounts of noisy label proportions.

In Liu et al. (2018), an approach using a denoising autoencoder based on gated recurrent units (GRUs) was proposed to forecast the vibration values of the bearings in the future cycle based on the preceding cycle. Experimental results reveal that the suggested technique demonstrates robustness and high accuracy in classification. Specifically, it achieved over 99% accuracy in three different operating conditions on the CWRU dataset. In Liu et al. (2019), a low-latency and lightweight recursive neural network model was proposed, which analyzed the impact of data dimensionality, data length, and network depth on the model's accuracy, disturbance resistance, and computation latency. Experimental findings indicated that the framework achieved the maximum accuracy and noise resistance when the network depth was 2, the data dimensionality was 64, and the data length was 128. The model was validated on the CWRU and IMS datasets, and achieved at least 10% reduction in computation latency compared to LSTM or GRU models with comparable performance, making it more appropriate for real-time defect detection systems. In An et al. (2020), an improved LSTM model was suggested for defect diagnosis under time-varying speed and varied load. The model extracts different types of data at different stages of training and inputs them into the classifier to improve training efficiency and diagnostic capability. The fault diagnosis accuracy achieved 95.3% in their experimental setup. In Zhang et al. (2021), a recursive neural network structure called FDGRU was proposed, which converts a time-series vibration data into a two-dimensional picture and incorporates gated recurrent units (GRUs) to fully extract characteristics in the temporal information. The proposed method achieved over 99% accuracy on two bearing

datasets and showed robustness to noise. In Gao et al. (2021), a model suitable for extracting weak fault features and diagnosis was proposed, which first conducts multiple channels continuous wavelet transform (MCCWT) on the vibration signal of the bearing, and then inputs it into a convolutional based on features recurrent neural network (CFRNN). The model also incorporates several remaining blocks and LSTM blocks to better mine time features and localised vibration information. Tests demonstrate that the accuracy can approach 97.7% in diagnosing faulty bearings under severe noise. In Gu et al. (2021), a model based on multi-sensor fault diagnosis using LSTM and discrete wavelet transform was proposed. Firstly, DWT was utilized to collect detailed defect information at varied frequencies and time periods. Then, LSTM network was utilized to define the long-term reliance hidden in the defect information time series. The model incorporates multi-scale data from numerous sensors to augment fault features, therefore boosting the diagnostic accuracy of the DWT-LSTM model.

Guo et al. (2017) proposed a Recurrent Neural Network-based Health Indicator (RNN-HI) method for predicting the RUL of bearings. The study combined six pertinent similarity features and eight conventional time-frequency features to create an original feature set in order to account for the uneven contributions of various statistical features in conventional health indicators. The RNN-HI was then built using the most sensitive features that met the criteria for monotonicity and correlation. The experiment was validated on the FEMTO-ST dataset, resulting in an average error of 23.24%. Due to the limitations of traditional RNN in handling long-term dependencies and computational efficiency, Li et al. (2021) proposed a Multi-Layer Gated Time Convolutional Network (MLGTCN) for predicting the RUL of rotating machinery. The model innovatively introduced multiple layers of temporal convolutional networks and Gated Linear Units

(GLU). It enlarged the receptive field of each layer and captured long-term historical information to address the issue of long-term dependencies. Experimental results demonstrated that MLGTCN achieved high prediction accuracy and computational efficiency in the RUL prediction of double-row roller bearings. Chui et al. (2021) presented an RUL prediction algorithm that combines RNN and LSTM. RNN leveraged the advantage of short-term prediction, while LSTM performed well in long-term prediction. The weights of the model were determined using the Non-Dominated Sorting Genetic Algorithm II. The model was validated on a simulated dataset of a turbofan engine, yielding an average RMSE of 17.2. Jiang and Xiang (2023) proposed an Ensemble of Deep, Long-Term, and Short-Term Memory (EDLSTM) method. First, the features required for predicting RUL were extracted, including time-domain features and frequency-domain energy features. Then, the EDLSTM model for bearing RUL prediction was constructed. The model integrated multiple DLSTM blocks using the Maximum Information Coefficient criterion to address the instability of model predictions. Experimental results showed a prediction accuracy exceeding 0.8. Table 2.5 summarizes the results obtained from the above studies on Recurrent Neural Network.

Table 2.5: Recurrent Neural Network utilized for FDP research comparison

Refs	Year	Method	Result	Data set
(Li et al., 2018)	2020	DRNN	Acc:94.75%	CWRU
(Shenfield & Howarth, 2020)	2020	RNN-WDCNN	Average acc:96.2%	CWRU
(Liu et al., 2019)	2019	LLRNN	Average acc:99.6%,99.5%	CWRU,IMS

Table 2.5: continued

(Zhiwei, 2022)	2022	1DCNN-GRU	Acc:99%	CWRU
(Liu et al., 2018)	2018	GRU-NP-DAE	hp=0 Acc:99.65% hp=1 Acc:99.85% hp=3 Acc:99.55%	CWRU
(An et al., 2020)	2020	LSTM	Average acc:95.3%	Private data set
(Zhang et al., 2021)	2021	FDGRU	Average acc:99.96%,99.99%	CWRU,Self-priming Centrifugal Pump
(Nie & Xie, 2021)	2021	NRNN	Average acc:99.18%	CWRU
(Gao et al., 2021)	2021	CFRNN	Average acc:97.7%	CWRU
(Gu et al., 2021)	2021	DWT-LSTM	hp=0 acc:99% hp=2 acc:93.3%	CWRU
(Guo et al., 2017)	2017	RNN-HI	Mean of Er:23.24%	FEMTO-ST
(Li et al., 2021)	2021	MLGTCN	RMSE:0.0435	IMS
(Chui et al., 2021)	2021	RNN-LSTM	RMSE:17.2	CMAPSS-TD
(Jiang & Xiang, 2023)	2023	DLSTM	Acc:81%	XJTU-SY

Despite the progress made in the literature on fault diagnosis, there are still some notable shortcomings. First, many studies have not achieved ideal fault recognition rates under high signal-to-noise ratio conditions, failing to effectively address the impact of signal interference on recognition accuracy. Additionally, current research often does not

sufficiently discuss the issue of fault recognition across different operating conditions, which is particularly important in practical applications. Specifically, in the context of bearing fault prediction, while many papers focus on the prediction of remaining useful life, there is a lack of in-depth discussion on predicting the degradation stages of bearings. Therefore, this study aims to address these research gaps by exploring strategies to improve fault recognition rates under high signal-to-noise conditions and placing a particular emphasis on predicting the degradation stages of bearings, in order to provide a more comprehensive solution for the field of fault diagnosis.

2.4 Chapter Summary

This chapter introduced the traditional methods of bearing fault diagnosis, followed by an elucidation of the advantages of deep learning in bearing fault diagnosis and prediction. These advantages include the ability to extract fault features directly and perform pattern recognition, thus achieving end-to-end fault diagnosis, owing to their deep structure and powerful nonlinear feature extraction capabilities. Furthermore, a comprehensive review of various deep learning models applied to bearing fault diagnosis and prediction is presented, including convolutional neural networks, recurrent neural networks, autoencoders, and deep belief networks.

CHAPTER 3

METHODOLOGY

3.1 Overview

This chapter describes the research methods applied in this study. It elaborates the design approach for bearing fault diagnosis and prediction using deep learning technology. Additionally, it covers the methods used in this study for preprocessing vibration data related to bearings, the theoretical foundation of deep learning models, training and optimization methods for the models, techniques for feature extraction and dimensionality reduction, as well as evaluation metrics for the models.

3.2 Research Design and Procedures

The research methodology flow diagram is illustrated in Figure 3.1, consisting of three main components.

The first step involves constructing a bearing fault diagnosis model suitable for high-noise environments to achieve qualitative and quantitative diagnostic tasks of bearing faults. The model's accuracy is then validated using publicly available bearing datasets. The study provides insights into end-to-end bearing fault diagnosis research under strong noise interference.

The second step entails establishing a bearing fault signal acquisition experimental platform to collect fault signals. The adaptability of the model under strong noise environments and across different operating conditions is then verified using the model developed in the first step.

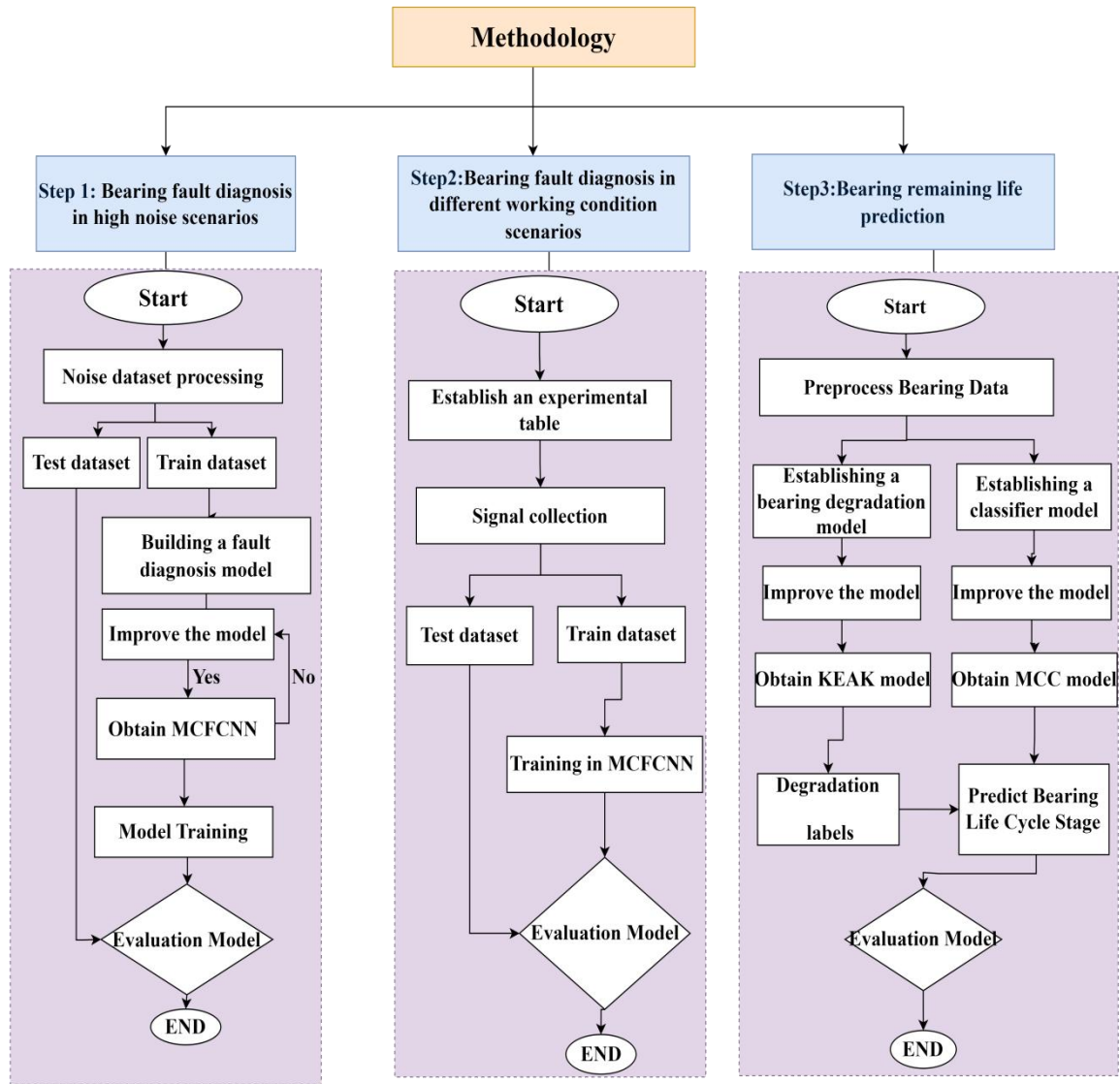


Figure 3.1: Research methodology flow diagram

The third step focuses on building a predictive model for bearing degradation maintenance. By selecting appropriate features, designing suitable model architectures and parameters, and undergoing steps such as data preprocessing, training, and optimization, the model predicts the degradation stage of bearings. Finally, the model's accuracy and reliability are validated using datasets.

The successful completion of these three steps will contribute to improving the accuracy and reliability of bearing fault diagnosis and prediction, providing robust support

for practical bearing maintenance applications. Finally, the results were discussed, summarized, and reported.

3.3 The Basic Theory of Rolling Bearing Fault Diagnosis

3.3.1 Rolling Bearing Structure

A typical rolling bearing consists of four components: the outer ring, rolling elements, cage, and inner ring, as illustrated in Figure 3.2. The outer ring is usually fixed to the bearing housing, while the inner ring is fixed to the shaft. The rolling elements are positioned between the outer and inner rings, with the cage used to maintain the relative distance between the rolling elements. Bearings can be classified into radial and thrust bearings based on the direction of the load they support. Depending on the application scenario, the rolling elements in bearings can be spheres or cylinders, among others.

Although different bearings may have different types and sizes, the basic components of a bearing remain the same as the aforementioned four parts. Therefore, the types of bearing faults can be classified according to the structure of the bearing (El-Thalji & Jantunen, 2015), namely, outer ring faults, rolling element faults, cage faults, and inner ring faults. These faults can further be categorized into wear faults and surface damage faults (Cerrada et al., 2018). Wear faults typically refer to surface roughness, irregularities, or extensive degradation of the bearing. The causes of such faults include internal contamination, lack of lubrication, shaft currents, misalignment, etc. Usually, wear faults do not exhibit obvious fault characteristic signals and pose relatively smaller risks compared to other types of faults. Therefore, the majority of research focuses on surface damage faults, which manifest as localized faults such as cracks, spalling, pitting, and flaking. These damages produce specific fault characteristics and pose significant risks.



Figure 3.2: Rolling bearing (Hoang & Kang, 2019)

3.3.2 Vibration Frequency of Bearing Fault Features

When mechanical equipment operates, vibrations are typically generated, especially when the equipment contains rotating components. Bearings, crucial components in rotating machinery, are commonly utilized to support the operation of rotating parts. Being constantly in motion themselves, bearings generate specific vibrations under normal operating conditions. However, when a bearing experiences a fault, its smooth motion is disrupted, resulting in additional vibrations depending on the type of fault.

The generation of characteristic vibration signals due to bearing faults is illustrated in Figure 3.3(a). For instance, if there is a crack on the outer race of the bearing, impact vibration signals occur as the rolling elements pass through the crack, as depicted in Figure 3.3(b). Since cracks have two edges, when the rolling elements move away from the crack area, they collide with the other edge, resulting in another impact vibration signal. When overlapped with the previous impact vibration signal, as shown in Figure 3.3(c), there are two vibration peaks visible. Theoretically, utilizing the time difference between these two peaks can be employed to calculate the width of the crack, thereby determining the size of the fault.

However, the aforementioned analysis represents an ideal scenario for bearing fault diagnosis. In reality, the working conditions of bearings are highly complex, and the vibrations induced by the rotation of bearings themselves make it challenging to directly analyze time-domain signals to determine the type and severity of faults.

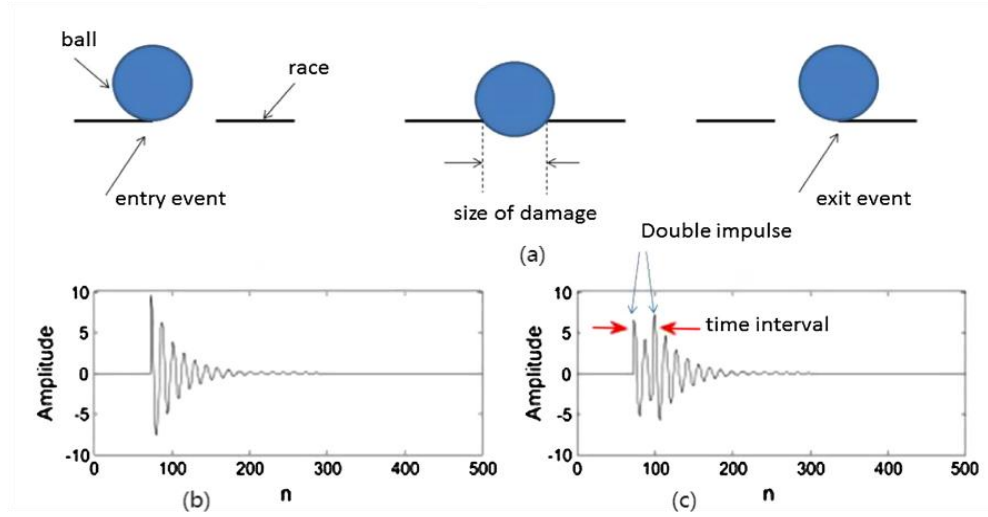


Figure3.3: Principle schematic diagram of bearing fault vibration signal generation: (a) Rolling element passing through crack on outer race or inner race; (b) Vibration signal generated as rolling element just enters crack; (c) Double impact vibration signal generated as rolling element exits crack zone (Cerrada et al., 2018).

Since bearings typically rotate at certain frequencies, the rolling elements passing through the faulty area generate vibrations at specific frequencies. In other words, different components' faults produce distinct fault characteristic frequencies. Therefore, by analyzing the characteristic frequencies in the vibration signals, it is possible to diagnose the location of the fault within the bearing. Assuming the bearing is a rigid body, installed with the outer ring fixed, and the rolling elements are purely rolling, the theoretical fault vibration frequency of the bearing can be obtained using the following formula.

$$\text{Outer circle: BPFO} = \frac{nf_r}{2} \left(1 - \frac{d}{D} \cos \phi \right)$$

$$\text{Inner circle: BPFI} = \frac{nf_r}{2} \left(1 + \frac{d}{D} \cos \phi \right)$$

$$\text{Cage: FTF} = \frac{f_r}{2} \left(1 - \frac{d}{D} \cos \phi \right) \quad \text{Equation 3.1}$$

$$\text{Roller: BSF} = \frac{Df_r}{2d} \left(1 - \left(\frac{d}{D} \cos \phi \right)^2 \right)$$

Where d and D represent the diameter of the rolling elements and the pitch diameter respectively, f_r represents the shaft rotational speed, ϕ represents the contact angle, and n represents the number of rolling elements. BPFO stands for Ball Pass Frequency Outer race, indicating the frequency at which vibrations occur when a fault is present on the outer race of the rolling bearing, as the rolling elements pass over the fault. BPFI stands for Ball Pass Frequency Inner race, indicating the frequency at which vibrations occur when a fault is present on the inner race of the rolling bearing. FTF represents Fundamental Train Frequency, indicating the frequency at which vibrations occur due to faults in the bearing cage. BSF represents Ball Spin Frequency, indicating the frequency at which vibrations occur due to faults on the rolling elements. When a fault occurs on a rolling element, vibrations occur at twice the BSF because the rolling element makes contact with both the outer and inner races. Moreover, the fault frequencies of the rolling elements are typically modulated by the BPFO or BPFI signals. Therefore, diagnosing faults on rolling elements is often more challenging than on inner and outer races. The probability of faults in different components of the bearing varies, with faults in the outer and inner races typically accounting for over 90% of bearing failures, while faults in the rolling elements and cage are relatively less common (Zhao et al., 2013). By understanding the characteristic frequencies of each component, it is possible to identify the source of the fault by analyzing the vibration signals of the bearing.

In practical scenarios, acceleration sensors are typically installed on the bearing pedestal or bearing housing to indirectly obtain the vibration signals of the bearing. Since the vibration signals are not directly collected from the bearing, the acquired vibration signals often contain vibration information from other components of the mechanical device. Additionally, the bearing fault signals may also be interfered with by the rotational signals of the bearing itself. Therefore, extracting fault-related features from the vibration signals containing interference becomes crucial for bearing fault diagnosis.

3.4 Bearing Dataset

Currently, there are abundant datasets available for bearing fault diagnosis and prediction. This section will introduce three bearing datasets that used in the thesis, two of which are for bearing fault diagnosis tasks, while the other is for bearing life stage prediction tasks.

3.4.1 Case Western Reserve University (CWRU) Bearing Dataset

The Case Western Reserve University Bearing Dataset (CWRU) (Zhu et al., 2023) is a publicly available mechanical fault diagnosis dataset created in 2003 by Dr. James T. Allison and colleagues from the Department of Mechanical Engineering at Case Western Reserve University in the United States. This dataset consists of accelerometer signals from four bearings, each containing data from normal operation as well as various types of fault conditions. The fault types include inner race faults, outer race faults, and rolling element faults.

This dataset is utilized for research and experimentation in mechanical fault diagnosis and predictive analytics. It contains extensive time-series signals suitable for studying the performance of vibration analysis and signal processing algorithms.

Additionally, the dataset includes basic parameters and fault characteristic information of the bearings, which aids researchers in analyzing the fault mechanisms of the bearings.

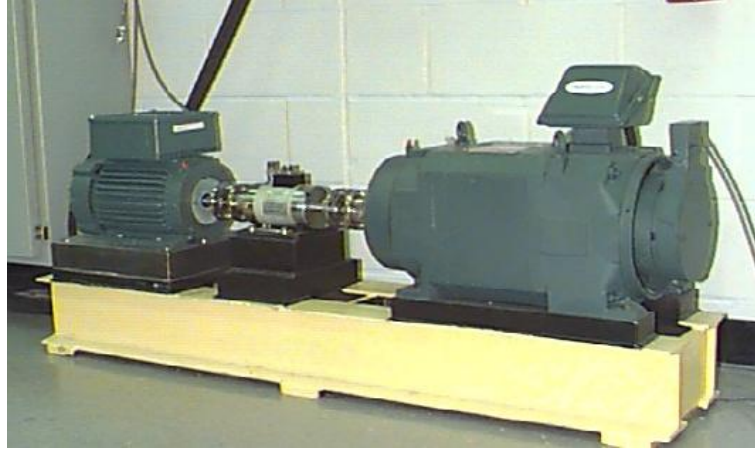


Figure 3.4: CWRU rolling bearing test bench (Zhu et al., 2023)

As shown in Figure 3.4, the test rig consists of a 2-horsepower motor (left), a torque sensor/encoder (middle), and a dynamometer (right). The test bearing supports the motor shaft. Artificial faults are introduced using electrical discharge machining on the inner race, rolling elements, and outer race, resulting in various fault types. The fault diameters are 7, 14, 21, and 28 milliliters, respectively, where 1 milliliter equals 0.001 inches. Data are sampled at frequencies of 12 kHz and 48 kHz. The collected data mainly include fault data under different speeds and conditions, as well as vibration data under normal conditions. The CWRU data set is used for fault diagnosis tasks. Relevant parameters and fault information are presented in Tables 3.1 and 3.2. Each data set is provided as a .mat file containing acceleration data for both drive-end and fan-end bearings.

Table 3.1: Specific parameters of rolling bearing (DE) (Zhu et al., 2023)

Inside Diameter(in)	Outside Diameter(in)	Thickness(in)	Ball Diameter(in)	Pitch Diameter(in)
0.9843	2.0472	0.5906	0.3126	1.537

Table3.2: Rolling bearing fault information (Zhu et al., 2023)

Fault Location	Diameter (in) Four sizes	Depth (in) Two sizes	Defect frequencies (Hz)
Inner Raceway	.007,.014,.021,.028	.011,.050	4.9469
Outer Raceway	.007,.014,.021,.028	.011,.050	3.0530
Ball	.007,.014,.021,.028	.011,.050	3.9874

3.4.2 BFTS Bearing Dataset

This data set was collected from a self-built bearing fault signal acquisition platform. The data set comprises accelerometer signals from four bearings, each containing data from both normal and various fault states. The types of faults include inner race faults, outer race faults, and rolling element faults.

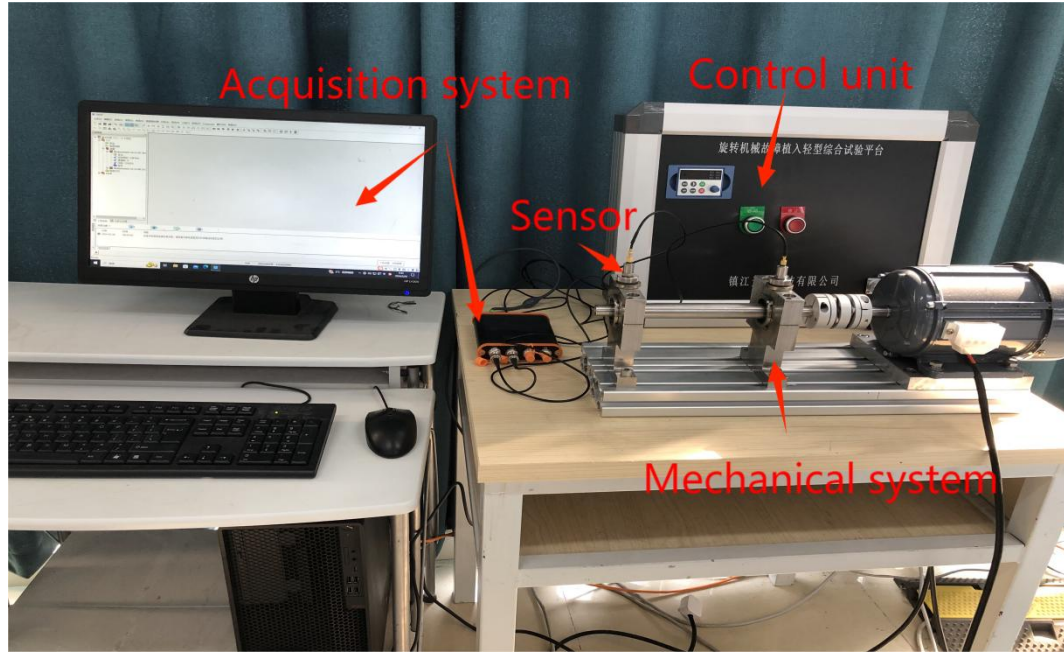


Figure 3.5: BFTS Fault Acquisition Platform

As shown in Figure 3.5, the test rig consists of a motor, an electrical control cabinet, and a signal analysis system. The test bearing supports the motor shaft, with the bearing model being 6203. Artificial treatment using electric spark technology is applied to the inner race, rolling elements, and outer race of the bearing, resulting in various forms of crack faults. The crack faults vary in depth and width, with three different sizes for each, distinguishing three different levels of fault types. The sampling frequency of the signal acquisition system is 32 kHz. The collected data mainly include fault data under different operating conditions and vibration data under normal conditions. The relevant parameters and fault information are shown in Table 3.3.

Table 3.3: Rolling bearing fault information

Fault Location	Crack Width (mm)	Crack Depth (mm)
	Three sizes	Three sizes
Inner Raceway	0.2,0.3,0.5	0.2,0.3,0.5

Outer Raceway	0.2,0.3,0.5	0.2,0.3,0.5
Ball	0.2,0.3,0.5	0.2,0.3,0.5

The data set comprises acceleration signals from 10 bearings, out of which 9 are faulty. The types of faults include inner race faults, outer race faults, and rolling element faults. Each type of fault has three different crack sizes. Therefore, the data set consists of nine faulty bearing datasets and one healthy bearing data set.

3.4.3 FEMTO-ST Bearing Dataset

The FEMTO-ST research centre at the University of Franche-Comté in France created and made available the publicly accessible FEMTO-ST (Zhang et al., 2018) bearing dataset. It is frequently employed for predicting and diagnosing bearing faults. The dataset contains vibration data of four different types of simulated bearing fault states under laboratory conditions. These fault categories include rolling element fault, inner race fault, outer race fault, and normal state fault. It also includes degradation data of the bearings, which are presented as vibration and temperature data. The vibration sensor is made up of two tiny accelerometers that are 90 degrees apart .

The data set collected the entire process of bearings from initial operation to failure, including vibration data in both horizontal and vertical directions of the bearings. To accelerate the degradation process, the bearings were subjected to extreme stress conditions. A total of 17 bearings were collected under three different load and speed conditions, as detailed in Table 3.4. For safety reasons, stopping the experiment after the accelerometer reading exceeds 20 times the gravitational acceleration indicates that the bearing has failed. The FEMTO-ST bearing test rig is shown in Figure 3.6. During the

experiment, samples were taken from the horizontal and vertical directions of the bearings every 10 seconds at a frequency of 25,600 Hz, each lasting for 0.1 seconds, resulting in each sample containing 2560X2 points.

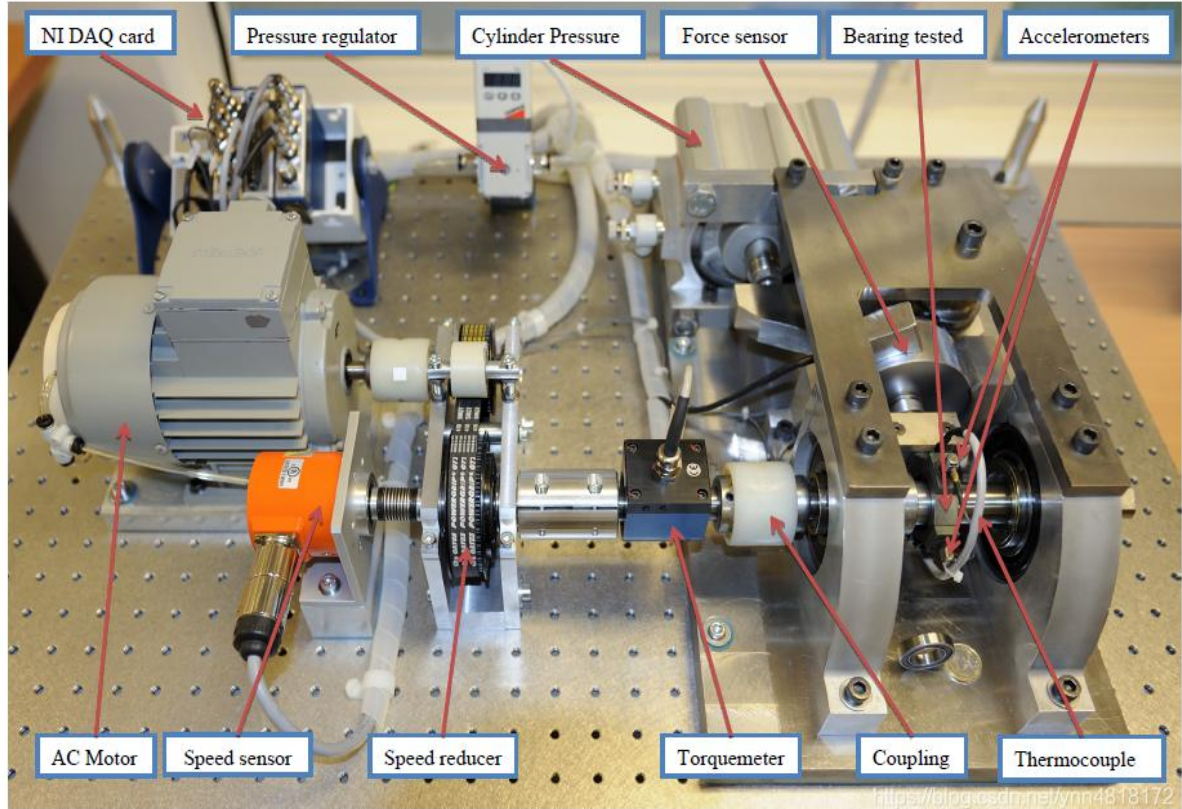


Figure 3.6: Experimental setup for FEMTO-ST (Zhang et al., 2018)

Table3.4: FEMTO-ST data set information

Work Conditions	Load Force(N)	Speed (r/min)	No. of train sets	No. of test sets
1	4000	1800	2	5
2	4200	1650	2	5
3	5000	1500	2	1

3.4.4 Data Preprocessing

The vibration signal collected by the vibration accelerometer is a one-dimensional signal. Before utilizing deep learning models for processing, data preprocessing is essential. In this research, the preprocessing methods employed for fault vibration tasks mainly include data segmentation, data augmentation, and noise addition. Meanwhile, for the bearing degradation stage prediction task, the primary preprocessing methods involve frequency domain and time domain transformations of the signal.

1. Signal segmentation

Due to the lengthy nature of the collected raw vibration signals and the fixed input data length required by the model, it is necessary to segment the original signals into fixed lengths. Considering that vibration signals exhibit periodicity, which is correlated with the motor speed, the length of signal segmentation should be related to the motor speed. In the subsequent experiments of this research, the input length used is the number of samples collected when the motor completes one revolution. This segmentation length ensures the completeness of the period while minimizing data redundancy as much as possible. In the CWRU database, the sampling rate of bearing signals is 12000 Hz, and the motor speed is 1797 rpm, so approximately 402 data points can be collected per revolution of the motor. Therefore, in the subsequent experiments, the signal input length needs to be greater than 402. Additionally, to capture fault signals, the length of each sample should not be too short, and it is common practice to include some overlap between segments during segmentation to augment the data (Zhang et al., 2018). Thus, in our model, we choose 2048 as the length of each sample. Furthermore, the amplitude of the original signals is

normalized to $[-1, 1]$. This normalization approach is also applied to the BFTS dataset for data processing.

2. Data augmentation

Data augmentation is a powerful signal processing technique aimed at increasing the quantity of sampled data to enhance the resolution of signals and improve the analysis capability of specific frequency components. In the analysis of bearing vibration signals, data augmentation plays a critical role by effectively adapting to datasets with different sampling frequencies, enabling more flexible data handling and analysis. Such processing is crucial for the assessment of bearing health and fault prediction. To address datasets with varying sampling frequencies, we choose time length as the metric unit to describe the length of data, the size of sliding windows, and the length of overlaps. The advantage of using time length as the metric unit lies in eliminating dependence on specific sampling rates, allowing for more flexible data handling without being restricted by specific hardware or instruments. By measuring data in time length, we can achieve consistent processing of data at different sampling frequencies, which is crucial for data analysis and algorithm application. The schematic diagram of data augmentation is shown in Figure 3.7.

In this study, we specifically adopt a data overlap sampling method with a time step of 28. Data overlap sampling is a common processing technique that effectively increases the sampling density by introducing overlap between adjacent time intervals. This approach helps preserve the dynamic characteristics and transient information of signals, resulting in more accurate and reliable analysis results.

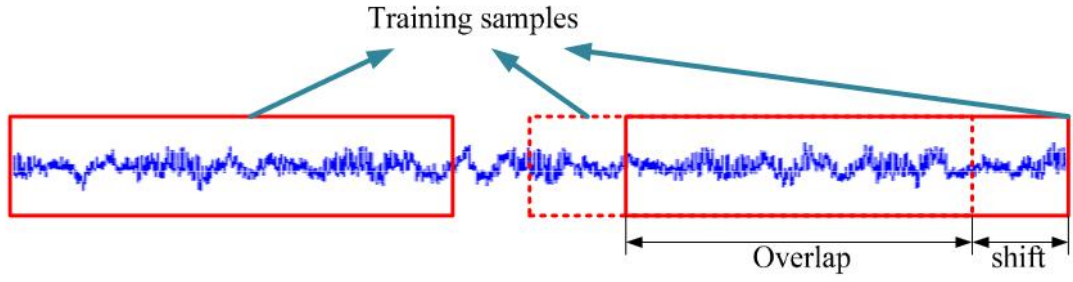


Figure 3.7: Data augmentation with overlap (Zhang et al., 2017)

3. Noise addition

Noise is commonly present in various environments, among which additive Gaussian white noise is one of the most representative and easily quantifiable types of noise. In signal processing, additive Gaussian white noise is typically used to simulate noise. In this research, in order to investigate the impact of noise on the classification of bearing fault signals, additive Gaussian white noise is also used as additional noise. The magnitude of noise is usually measured using the Signal-to-Noise Ratio (SNR). SNR is defined as the ratio of the power of the useful signal to the power of the noise in the signal, and its unit is typically expressed in decibels (dB):

$$\text{SNR}_{\text{dB}} = 10 \log_{10} \left(\frac{P_{\text{signal}}}{P_{\text{noise}}} \right) \quad \text{Equation 3.2}$$

The symbol P denotes the average power of the signal $X(t)$ over a period T . The average power of a signal is defined as follows:

$$P_{\text{avg}} = \frac{1}{T} \int_0^T p(t) dt = \frac{1}{T} \int_0^T |x(t)|^2 dt \quad \text{Equation 3.3}$$

When the signal is in discrete form, its average power is defined as follows:

$$P_{\text{avg}} = \frac{1}{\text{length}(x)} \sum_{t \in \text{length}(x)} |x(t)|^2 \quad \text{Equation 3.4}$$

For a signal with zero mean and known variance, its power can be expressed by its variance. For standard distributed noise, i.e., Gaussian noise, the signal power is 1. Therefore, to generate a signal with a specified signal-to-noise ratio (SNR), we first calculate the power, P_{signal} , of the original signal using Equation (3.4). Then, we calculate the power, P_{noise} , of the noise signal to be generated using Equation (3.2). Finally, we generate the required additive Gaussian white noise signal with the specified power using Equation (3.5).

$$N = \sqrt{P_{\text{noise}}} \times \text{randn}(\text{length}(x)) \quad \text{Equation 3.5}$$

In this context, $\text{randn}()$ represents a function used to generate noise with a standard distribution, while $\text{length}()$ denotes the length of the input signal. By superimposing the generated noise signal with the original signal, a noisy signal with a specified signal-to-noise ratio (SNR) can be generated.

The comparison of signals before and after noise addition is shown in Figure 3.8. The signals in the figure are sourced from the outer race fault signals in the CWRU database. The original signal is depicted in the first row of the figure, while the signal with added noise at -4 dB SNR is illustrated in the second row.

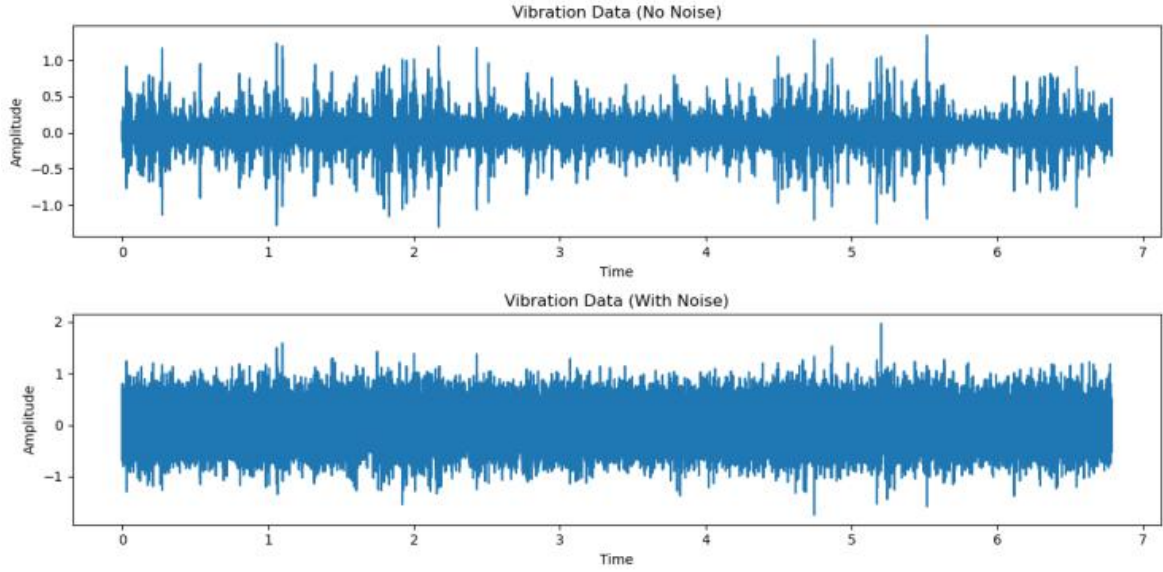


Figure 3.8: Comparison of clean and noise-disturbed vibration time series plots

When comparing time-domain signals, it is evident that the time-domain signal of the outer race fault without noise exhibits significant periodicity and noticeable amplitude fluctuations. However, signal identification becomes relatively challenging when noise is added. Considering that noise is inevitable in real-world environments, a robust model must possess sufficient adaptability to recognize fault signals with different levels of noise. To simplify the complex and variable noise in real-world environments, we only consider the level of noise in the experiments, testing the performance of the proposed algorithm under signals with different signal-to-noise ratios (SNRs).

4. The time domain and frequency domain conversion of the signal

In the task of predicting the degradation stages of bearings in this research, we adopted a method that comprehensively utilizes both time-domain and frequency-domain features. Each bearing in the data set underwent a feature extraction process, which included 13 time-domain features. These time-domain features were detailed in Chapter 2 and help describe the variation of signals over time. Additionally, to analyze the vibration

signals of bearings more comprehensively, we also employed frequency-domain transformation methods, specifically using the Fourier transform. Through the Fourier transform, signals can be transformed from the time domain to the frequency domain, revealing the signal's characteristics in terms of frequency, known as the signal's spectrum. In the frequency domain, the signal is represented as a spectrum showing how the amplitude and phase change with frequency. This allows the model to detect specific frequency components in the bearing signal, such as fault frequency peaks, thereby better extracting features of bearing degradation. Finally, to fully utilize the information from both time and frequency domains, these features were fused in the model. This fusion enables the model to capture the features of bearing degradation more comprehensively, thereby improving the accuracy and reliability of the predictions.

3.5 Deep learning model knowledge used in the model of the thesis

In Chapter 2, we have reviewed the current research status of deep learning models in bearing fault diagnosis and life prediction. This section will specifically explain the theoretical knowledge of deep learning models involved in the model proposed by us.

3.5.1 Convolutional Neural Network

Inspired by the hierarchical working principle of the visual cortex in the biological neural system, Convolutional Neural Networks (CNNs) (Y, 1989) also consist of multiple convolutional layers, each containing several neurons with shared weights, with each neuron responsible for processing a small portion of the receptive field. The lower layers are typically used to extract common features from the input, such as edges, corners, and colors; as the depth increases, the network layers can recognize textures, shapes, and other features by combining low-level features; and the higher layers of the network can achieve

recognition of objects and other high-level abstractions. By using this local-to-global recognition approach, CNNs can reuse common features, thereby reducing the difficulty of recognizing different features and reducing the number of network parameters (Najafabadi et al., 2015).

Furthermore, CNNs utilize convolutional kernels with the same weights to extract features from different positions of the input, enabling the extraction of the same features from the input and significantly reducing the number of parameters. Typically, there are many convolutional kernels in each layer of the network, each of which is different, meaning each kernel is used to extract a separate feature from the input, resulting in the combination of multiple different features as the output of the network layer. This weight-sharing mechanism can extract all similar features from the input to facilitate the subsequent recombination and utilization of features by network layers, while also greatly reducing the number of network parameters.

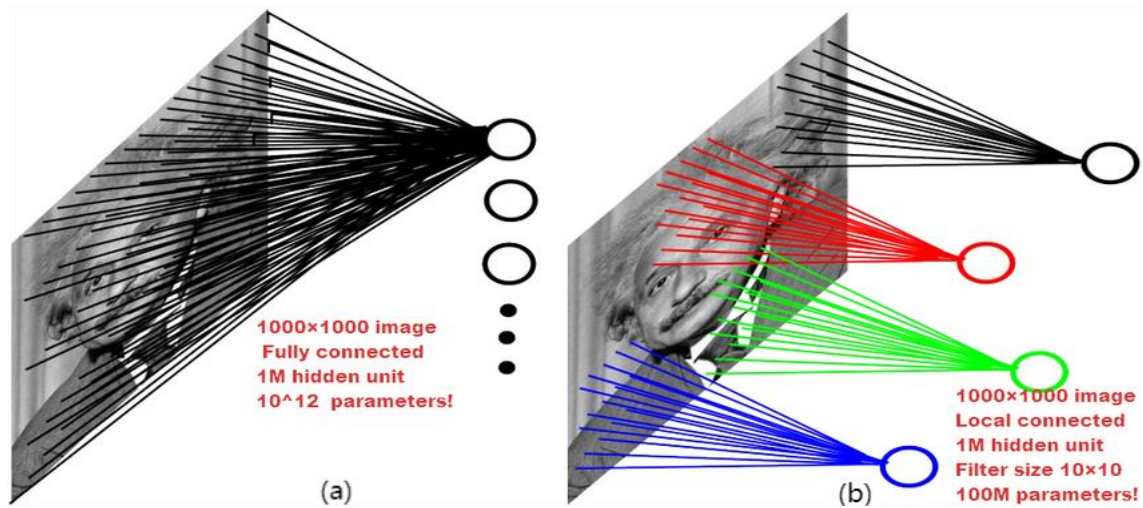


Figure 3.9: Schematic Comparison of Parameter Count between (a) Fully Connected Neural Network and (b) CNN (J. Qiu, 2016)

In Figure 3.9 (a), for a 1000x1000 size image input, if fully connected layers are used, assuming there are 100 neurons in the hidden layer, the total number of parameters

needed would be on the order of 10^8 . However, in Figure 3.9 (b), assuming a convolutional kernel size of 10×10 and using 100 convolutional units, employing a convolutional network would require parameters on the order of 10^4 , significantly reducing the parameter volume of the network.

Additionally, convolutional neural networks possess a certain degree of translational invariance (Qiu, 2016). When a feature in the input is shifted by a certain distance, the network still produces the same response output. Due to the weight-sharing property of convolutional kernels, the same kernel extracts features at all positions in the input, resulting in the extraction of identical features. Furthermore, the combination of convolution and max-pooling enables the feature to remain relevant even if it shifts by a few pixels at the current position, as long as it remains within the receptive field of the next layer's neurons, thereby achieving translational invariance of features. Another advantage of convolutional neural networks is that their input size is variable because the number of neurons in the input layer does not necessarily correspond to the input length. In contrast, traditional neural networks cannot handle variable input lengths, making convolutional neural networks more flexible.

Typically, a convolutional neural network consists of convolutional layers, batch normalization layers, activation function layers, pooling layers, and dropout layers. The following sections will provide separate introductions to each layer.

1. Convolutional layers

The basic operation in CNN is convolution, where each convolutional unit corresponds to a small portion of the receptive field. The convolution operation in neural networks differs slightly from the mathematical definition of convolution; it does not

involve a flip operation and therefore lacks commutativity. The convolution described later in this text follows the definition in neural networks. For one-dimensional operations, the convolution in neural networks is defined as follows:

$$k * f[i] = \sum_{m=0}^W k[m]f[i + m] \quad \text{Equation 3.6}$$

In the context of convolutional neural networks, $*$ denotes the convolution operation, where k and f respectively represent the convolution kernel function and the convolution input, and W represents the width of the convolution kernel. Due to the fact that the width of the convolution kernel is much smaller than the length of the input signal, convolutional networks exhibit sparsity in connections. The convolution kernel k can be regarded as a filter, and the convolution operation can be seen as the process of filtering the input signal. For the two-dimensional case, the convolution in neural networks is defined as follows:

$$k * f[i, j] = \sum_{m=0}^{W_1} \sum_{n=0}^{W_2} k[m, n]f[i + m, j + n] \quad \text{Equation 3.7}$$

Similar to the one-dimensional case, where W_1 and W_2 represent the widths of convolutional kernels in different dimensions. The convolutional layer is a fundamental operation layer in convolutional neural networks, typically composed of several convolutional layers in a deep neural network.

2. Batch normalization layers

The introduction of Batch Normalization (BN) (Ioffe & Szegedy, 2015) aims to address the issue of internal covariate shift caused by the changing distributions of inputs in convolutional networks. In multi-layer neural networks, the distribution of inputs in each layer varies with changes in the network's input, necessitating constant adjustments during training to adapt to these distribution changes. This dynamic adjustment process can slow

down the convergence of the network during training, affecting its overall performance. Moreover, in deep networks, the accumulated distribution shifts from earlier layers to deeper layers can potentially increase the risk of gradient vanishing or exploding, thereby complicating the training process. By incorporating BN layers, the outputs of each layer can be normalized, reducing the impact of data distribution shifts at each layer. This normalization process helps accelerate the training speed of neural networks and simultaneously improves their accuracy.

Suppose a BN layer has a d -dimensional input, where $X \in \mathbb{R}^{(n \times d)}$, with n representing the batch size. Then, for each feature dimension $j \in \{1 \dots d\}$, the transformation in the BN layer can be expressed as follows:

$$\hat{x}_j = \frac{x_j - E[X_j]}{\sqrt{\text{Var}[X_j]}} \quad \text{Equation 3.8}$$

$$y_j = \gamma_j \hat{x}_j + \beta_j \quad \text{Equation 3.9}$$

In this context, y_j represents the transformed feature, x_j denotes the j -th column vector in the input features, corresponding to the j -th feature. γ_j and β_j are two parameters in BN that need to be learned, representing the scale and shift parameters, respectively. The BN layer first standardizes each feature channel separately for each batch, and then scales and shifts the standardized features to increase the network's feature representation capability and flexibility. By applying BN, the feature distributions of each layer are adjusted to positions where the activation function gradients are large, thereby alleviating the vanishing gradient problem to some extent in deep neural networks. Since its introduction, BN has been widely adopted in deep neural networks to enhance network stability, accelerate neural network training speed, and improve accuracy.

3. Activation function layers

The activation function defines the relationship between the output and input of a neuron, typically exhibiting nonlinearity, thereby ensuring that the neural network has sufficient representational capacity. If a neural network lacks an activation function or the activation function is a linear function, then regardless of the number of layers in the network or the number of neurons in each layer, the output of the network is merely a linear combination of the inputs, and the network is equivalent to a single-layer model. Below are common activation functions used in neural networks.

The Rectified Linear Unit (ReLU) (Nair & Hinton, 2010) employs a simple piecewise function to form the overall nonlinear function, typically applied after convolutional layers, defined as follows:

$$ReLU = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases} \quad \text{Equation 3.10}$$

The ReLU consists of two simple linear functions, resulting in relatively low computational complexity both during forward propagation and backward propagation for gradient computation. Furthermore, compared to the sigmoid function, ReLU's non-saturation property effectively alleviates the vanishing gradient problem (Glorot & Bengio, 2010). Additionally, due to ReLU's activation value being 0 for the negative half-axis, this structure easily encourages sparsity (Glorot et al., 2011), thereby accelerating network computation and training speed. These advantages have made ReLU one of the most popular activation functions in modern deep networks.

The softmax function, originally used to describe the Boltzmann distribution in statistical mechanics, was later applied in machine learning to normalize the probabilities of different items in the output while preserving the order of magnitudes. In neural networks, the softmax function is defined as:

$$Softmax(x_i) = \frac{e^{x_i}}{\sum_{j=1}^K e^{x_j}} \quad \text{Equation 3.11}$$

In this equation, K represents the number of categories. For each input x_i , it is first transformed into a positive number through exponentiation, then normalized to the interval $(0,1]$, thus representing the probability of each item. Moreover, all probabilities sum up to 1, and the order of probabilities remains unchanged. Softmax is commonly used in the final layer of neural networks to represent the ultimate output, particularly in scenarios involving multi-class classification.

The Exponential Linear Unit (ELU) activation function, introduced by (Clevert et al., 2016), is used in deep learning to introduce non-linearity. The definition of the ELU function is as follows:

$$ELU(x) = \begin{cases} \alpha(e^x - 1) & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases} \quad \text{Equation 3.12}$$

In this equation, α is a hyperparameter, typically chosen as a small positive value, used to control the slope in the negative region.

4.Pooling layers

The concept of the pooling layer (Graham, 2015) originates from models of visual recognition in biology and is associated with the concept of local unordered images. Pooling operations replace a region with its adjacent regions' overall statistical features. Depending on the type of statistical feature, average pooling and max pooling are the most commonly used pooling operations. Taking max pooling as an example, its definition is as follows:

$$y_j = \max(r_j^{n \times n} u(n, n)) \quad \text{Equation 3.13}$$

In this equation, y_j represents the output value of region j , where $r_j^{n \times n}$ denotes the j -th $n \times n$ region, and $u(n,n)$ represents an $n \times n$ window function. Typically, this window function is set to 1. Region r_j is usually traversed using a sliding window approach over the feature map, thereby implementing pooling operation across the entire image.

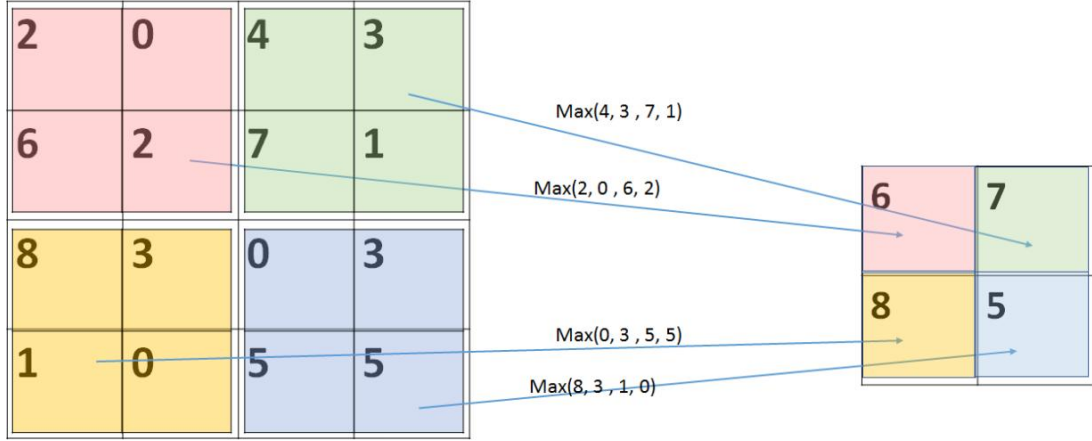


Figure 3.10: Schematic Diagram of Max Pooling with Pooling Kernel of 2 and Stride of 2

The schematic diagram of max pooling is shown in Figure 3.10, with a pooling kernel size of 2 and a stride of 2. It can be observed that the maximum value within the top-left region of the image is 6. Therefore, in max pooling, 6 is used to replace the values within the top-left region. Pooling operation can be regarded as a nonlinear downsampling method. By pooling, the size of feature maps can be significantly reduced, thereby reducing the number of parameters and computational workload of neural networks and lowering the risk of overfitting. Additionally, the combination of convolutional layers and pooling layers can bring about local invariance and enhance the network's receptive field, thus improving the computational efficiency and feature representation capability of neural networks.

5.Dropout layers

The purpose of the Dropout layer is similar to that of the BN layer, both aimed at optimizing the entire neural network. To prevent redundant relationships between neurons, the Dropout layer places some neurons in a non-activated state. The inclusion of the Dropout layer is akin to training many additional subnetworks, which share the same parameters. These subnetworks are not trained explicitly, but rather integrated. Through the integration of multiple subnetworks, the entire network gains greater robustness. This method is relatively convenient, reducing redundant relationships between neurons, enhancing training speed, and preventing overfitting. Additionally, it strengthens the adaptability of the network, effectively improving its ability to resist noise.

3.5.2 Auto Encoder

Deep learning algorithms consist of multiple single-layer networks connected in series, capable of performing complex learning tasks. However, they often encounter problems such as network convergence difficulties and gradient vanishing. The greedy unsupervised layer-by-layer training algorithm can decompose a multi-hidden-layer neural network into many independent networks during the training process, thus solving the problem of overall training (Mao et al., 2021). In this process, the labeled information of the data is not used, hence referred to as unsupervised learning.

Auto encoder (AE) is a typical unsupervised neural network model, which consists of two main parts: the encoder and the decoder. The encoder compresses the input to form a spatial representation, while the decoder reconstructs this spatial representation. Similarly, the training of weights in AE is achieved through back propagation. The encoding process of AE can be represented as follows:

$$h = f(wx + b) \quad \text{Equation 3.14}$$

In this equation, w and b are weights and biases, respectively, x representing the input to the network, and h is the compressed spatial representation vector. The decoding process is represented as follows:

$$\hat{x} = f(w'h + b') \quad \text{Equation 3.15}$$

w' and b' represent the weights and biases of the decoding layer, and they are used for the reconstructed output.

According to the preceding discussion, CNNs exhibit outstanding feature extraction capabilities, with sparse connections between layers, resulting in fewer parameters to train and greatly enhancing the overall training speed of the network. However, CNNs belong to supervised learning and require a sufficient amount of labeled data to be effective. In contrast to CNNs, AEs belong to unsupervised learning algorithms. Masci et al.(2011) proposed Convolutional Autoencoders (CAEs), which not only extract deep features from unlabeled data but also require fewer parameters to train, making them adaptable to various task requirements.

3.6 Model Training and Optimisation

Deep learning has been widely applied in the diagnosis and prediction of bearing faults. In the process of model training, it is crucial to choose appropriate loss functions and optimization algorithms to improve model performance and generalization ability. Below are some common loss functions and optimization algorithms introduced.

3.6.1 Loss Functions

The loss function is a performance metric for deep learning models, used to measure the difference between the model's predicted values and the actual values. Common loss functions for bearing fault diagnosis and prediction tasks include:

Cross Entropy Loss Function (Martinez & Stiefelhagen, 2018): It is suitable for classification problems, particularly multi-class classification tasks. Cross entropy measures the difference between the model's output and the true labels. Assuming the model's output is an N-dimensional vector p , representing the probability distribution of each class, and the true label's probability distribution is represented by an N-dimensional one-hot encoded vector q (i.e., only the element corresponding to the true label is 1, while the rest are 0). Its formulation is:

$$\text{CrossEntropy} = - \sum_{i=1}^N q_i \log(p_i) \quad \text{Equation 3.16}$$

In this context, q_i represents the i -th element of the true label (0 or 1), while p_i represents the predicted probability of the i -th element by the model. Cross-entropy measures the difference between the probability distributions of the model output and the true labels. When the two distributions are identical, the cross-entropy reaches its minimum value of 0. Therefore, by minimizing cross-entropy, the model can better fit the true label distribution.

The Root Mean Squared Error (Hodson, 2022) (RMSE) is a common loss function used to measure the difference between predicted and true values in regression problems. Suppose there are n samples, where the model's predicted values are represented by y_i , and

the true label values are represented by y_i . The mathematical formula for RMSE is as follows:

$$\text{RMSE} = \sqrt{1/n \sum_{i=0}^n (\hat{y}_i - y_i)^2} \quad \text{Equation 3.17}$$

In the tasks of bearing fault diagnosis and bearing degradation stage prediction, the categorical cross-entropy loss function was chosen as the optimization objective. The fault diagnosis task involves ten different fault categories and aims to accurately identify the types of faults in bearings for timely maintenance and repair. In contrast, the degradation stage prediction is a four-class problem, with the goal of determining the current degradation state of the bearing to predict its remaining useful life.

The reason for selecting the categorical cross-entropy loss function is that it is particularly well-suited for handling multi-class problems. This loss function effectively measures model performance by comparing the output probability distribution of the model with the encoded labels of the actual categories. Using this loss function helps the model adjust its parameters more effectively during the training process, thereby improving the accuracy of identifying various faults and degradation states. Additionally, the categorical cross-entropy loss function demonstrates a certain robustness to class imbalance issues, allowing it to maintain good learning performance even when sample distributions across different classes are uneven. This choice ensures the effectiveness and reliability of the model in both fault diagnosis and degradation prediction tasks, laying a solid foundation for implementing intelligent monitoring and predictive maintenance.

3.6.2 Optimisation Algorithm

Optimization algorithms used to update parameters of deep learning models to minimize loss functions. Common optimization algorithms in bearing fault diagnosis and prediction include:

Stochastic Gradient Descent (SGD) (Sweke et al., 2020): SGD is one of the most fundamental optimization algorithms. Assuming we have a loss function $L(\theta)$, where θ represents the parameters of the model, the update formula for SGD is as follows:

$$\theta_{t+1} = \theta_t - \eta \nabla L(\theta_t; x^{(i)}; y^{(i)}) \quad \text{Equation 3.18}$$

Where θ_t represents the value of the model parameters at the t -th iteration. η is the learning rate, which is a hyperparameter controlling the step size of updates, used to adjust the magnitude of parameter updates. The learning rate is typically a small positive number, such as 0.001 or 0.01. $\nabla L(\theta_t; x^{(i)}; y^{(i)})$ denotes the gradient of the loss function L with respect to the parameter θ_t . It is the gradient of the loss function at the current parameter value θ_t with respect to the sample $x^{(i)}$ and its corresponding label $y^{(i)}$.

In SGD, each iteration only utilizes one sample (or a small batch of samples, called stochastic gradient), instead of using all samples as in traditional gradient descent (batch gradient descent). This leads to faster update rates but may introduce some randomness in the direction of parameter updates. The main drawback of SGD is that the stochastic nature of the update direction may result in unstable convergence. To address this issue, improved variants of SGD such as Mini-batch SGD and Momentum SGD can be employed. Mini-batch SGD uses a small batch of samples in each iteration, while Momentum SGD

introduces the concept of momentum, leveraging the direction of previous updates to influence the current update direction, thus achieving more stable parameter updates.

Adaptive Moment Estimation (Zhang et al., 2019) (Adam): Adam is an optimization algorithm with adaptive learning rates, combining momentum optimization and the RMSProp algorithm. Adam utilizes first moment estimates (mean) and second moment estimates (variance) of gradients to adjust the learning rates.

Suppose we have a loss function $L(\theta)$, where θ is the model's parameter. The update formula for Adam is as follows:

First, compute the first-order moment estimate m_t and the second-order moment estimate (variance) v_t of the gradients.

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \quad \text{Equation 3.19}$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \quad \text{Equation 3.20}$$

Where, g_t is the gradient of the loss function L with respect to parameter θ_t . m_t is the first moment estimate of the gradient, i.e., the mean of the gradient. β_1 is the exponential decay rate used to control the first moment estimate. v_t is the second moment estimate of the gradient, i.e., the variance of the gradient. β_2 is the exponential decay rate used to control the second moment estimate.

The next step is to perform bias correction on the gradients.

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad \text{Equation 3.21}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad \text{Equation 3.22}$$

Where t represents the current iteration number, bias correction can eliminate the bias at the initial iteration. Finally, update the model parameters θ_t :

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}} \cdot \hat{m}_t \quad \text{Equation 3.23}$$

Where η is the learning rate, which controls the magnitude of parameter updates. ϵ is a very small value used to prevent division by zero.

Adam's update formula combines the momentum term m_t with the adaptive learning rate term $\frac{1}{\sqrt{\hat{v}_t + \epsilon}}$, allowing for more stable convergence during training and relative insensitivity to learning rates. It performs well in many deep learning tasks, making it one of the primary choices for optimization algorithms in deep learning.

When selecting an optimization algorithm, it is crucial to base the choice on the specific requirements of the task. In my bearing fault diagnosis model, employed SGD because this task requires accurate identification of multiple fault categories. SGD offers stable convergence properties and can effectively update parameters when dealing with large-scale data, helping the model gradually adjust to the optimal solution. In contrast, the bearing degradation prediction model uses the Adam optimization algorithm due to the nature of the task, which involves predicting the degradation state of the bearings. The data includes full-life data, is large in volume, and contains two-dimensional information. Adam can adapt more quickly to these changes through its adaptive learning rate, enhancing the model's learning efficiency and prediction accuracy. Therefore, selecting the

appropriate optimization algorithm based on the characteristics and requirements of the task can significantly improve the model's performance and reliability.

3.7 Feature Extraction

Feature extraction methods for bearing fault diagnosis and prediction will be introduced in this chapter. Here are the feature selection and dimensionality reduction techniques used in my model.

3.7.1 K-Means

K-means (Yoder & Priebe, 2017) is a common clustering algorithm used to partition a dataset of samples into K distinct categories or clusters. It is a distance-based clustering algorithm aimed at minimizing the sum of squared distances between data points and their corresponding cluster centers, so that data points within the same cluster are as close to each other as possible, while being as far apart as possible from data points in different clusters. The objective of K-means is to minimize the sum of squared errors (SSE), which is the sum of squared distances between data points and their corresponding cluster centers.

$$SSE = \sum_{i=1}^n \sum_{j=1}^K (x_i - \mu_j)^2 \quad \text{Equation 3.24}$$

Where x_i is the i -th data point, K is the number of clusters, and μ_j is the j -th cluster center.

In the prediction of bearing degradation stages, K-means can be utilized to cluster the vibration characteristic data of bearings, thereby forecasting the degradation status of the bearings. The purpose of doing so is to partition the vibration data into different

clusters, where each cluster represents a distinct bearing degradation status or level of malfunction.

3.7.2 T-SNE Visualisation Technique

When facing high-dimensional data, it is often challenging to intuitively understand the relationships between data points. t-SNE (t-distributed Stochastic Neighbor Embedding) is a nonlinear dimensionality reduction and visualization technique that helps us better understand the similarity and structural relationships between data points by mapping high-dimensional data to a lower-dimensional space. T-SNE (Kruiger et al., 2017) was initially proposed by Laurens van der Maaten and Geoffrey Hinton in 2008 and has since become a widely used tool for data visualization.

In t-SNE, we first need to compute the similarity of data points in high-dimensional space. A Gaussian kernel function is typically used to measure the similarity between two data points. For data points x_i and x_j , the similarity using the Gaussian kernel function can be represented as:

$$p_{ij} = \frac{\exp(-||x_i - x_j||^2 / 2 \sigma^2)}{\sum_{k \neq i} \exp(-||x_i - x_k||^2 / 2 \sigma^2)} \quad \text{Equation 3.25}$$

In this equation, $||x_i - x_j||^2$ represents the squared Euclidean distance between data points, and σ^2 is the variance of the Gaussian kernel, which controls the degree of similarity.

Next, we need to compute the similarity between data points in a low-dimensional space. To maintain similarity in the low-dimensional space, t-SNE uses the t-distribution to

represent the probability of similarity between data points. The similarity between data points y_i and y_j in the low-dimensional space can be represented as:

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|y_k - y_l\|^2)^{-1}} \quad \text{Equation 3.26}$$

T-SNE employs the Kullback-Leibler (KL) divergence to measure the disparity between the distribution of data points in low-dimensional space and their similarity in high-dimensional space. KL divergence gauges the relative entropy between two probability distributions, serving to quantify the extent of their differences. The definition of KL divergence is as follows:

$$KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad \text{Equation 3.27}$$

The goal of t-SNE is to minimize the KL divergence, making the distribution of data points in the low-dimensional space as similar as possible to that in the high-dimensional space. To achieve this goal, t-SNE uses optimization algorithms such as gradient descent to adjust the positions of data points in the low-dimensional space, thus minimizing the KL divergence.

Simultaneously, in t-SNE, perplexity is an important parameter that determines how many neighboring data points should be around each data point in the low-dimensional space. A lower perplexity leads to tighter clustering in the low-dimensional space, while a higher perplexity results in a more dispersed distribution. The choice of perplexity has a significant impact on the visualization effectiveness of t-SNE and requires experimentation and adjustment.

T-SNE is a powerful technique for data dimensionality reduction and visualization, aiding in better understanding the similarity and structural relationships among high-dimensional data. However, t-SNE also has some limitations, such as high computational complexity, which may require longer computation time when dealing with large-scale datasets. Therefore, in practical applications, selecting appropriate parameters and perplexity based on the characteristics of the data and visualization needs is crucial to achieving better visualization results.

3.8 Model Evaluation Indicator

In the task of bearing fault diagnosis, common evaluation metrics mainly involve performance assessment for classification problems. Since bearing fault diagnosis is typically a multi-class classification problem, common evaluation metrics include accuracy, recall, F1-score, and confusion matrix.

Suppose we have N samples, where the number of positives (indicating fault categories) is P , and the number of negatives (indicating normal categories) is $N-P$. The model's predictions can be described using the following four values:

True Positive (TP): The model correctly predicts fault samples as faults.

False Positive (FP): The model incorrectly predicts normal samples as faults.

True Negative (TN): The model correctly predicts normal samples as normal.

False Negative (FN): The model incorrectly predicts fault samples as normal.

Based on these four values, we can calculate the following evaluation metrics:

$$\text{Accuracy: } \text{Accuracy} = \frac{TP+TN}{N} \quad \text{Equation 3.28}$$

$$\text{Recall: } \text{Recall(TPR)} = \frac{TP}{TP+FN} \quad \text{Equation 3.29}$$

$$\text{Precision: } \text{Precision} = \frac{TP}{TP+FP} \quad \text{Equation 3.30}$$

$$\text{F1-score: } \text{F1 - score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad \text{Equation 3.31}$$

Confusion Matrix: A confusion matrix is a two-dimensional table with actual labels as rows and predicted labels as columns, used to compare the differences between model predictions and actual labels. The form of a confusion matrix is as follows:

$$\begin{bmatrix} \text{TN} & \text{FP} \\ \text{FN} & \text{TP} \end{bmatrix} \quad \text{Equation 3.32}$$

Through a confusion matrix, one can gain a more detailed understanding of a model's predictions, including the quantities of true positives, false positives, true negatives, and false negatives. This helps analyze the types of errors and performance of the model.

3.9 Chapter Summary

This chapter provides a detailed introduction to the research methods applied in this study, aiming to utilize deep learning techniques to achieve the design approach for bearing fault diagnosis and prediction. Firstly, we elucidate the methods for preprocessing bearing vibration data, including data cleaning, filtering, and normalization, to ensure data accuracy and consistency. Next, we introduce the theory of deep learning models, including neural network structures, activation functions, loss functions, etc. We select deep learning models suitable for bearing fault diagnosis and prediction tasks and provide

detailed explanations of their principles. Regarding model training and optimization methods, we delve into the mathematical formulas and pros and cons of optimization algorithms such as SGD, Adam, etc. We choose optimization algorithms suitable for this study and discuss methods for hyperparameter tuning to ensure that the model can learn the data features adequately and avoid overfitting. In terms of feature extraction and dimensionality reduction techniques, we introduce commonly used methods for feature extraction in bearing vibration data. Finally, we provide a detailed explanation of model evaluation metrics, including accuracy, recall, F1-score, etc., as well as the use of confusion matrices for a deeper analysis of the model's performance.

CHAPTER 4

FAULT DIAGNOSIS TASK IN HIGH NOISE ENVIRONMENT USING MCFCNN MODEL

4.1 Overview

Bearing fault diagnosis is an essential step in maintaining the smooth and safe operation of modern mechanical equipment, with vibration signal analysis being the most commonly used method for bearing fault diagnosis. Previous research has focused on manually designing filtering and feature selection processes, which require specialized background knowledge and lack intelligence in the diagnostic process. Despite achieving relatively good results, there is still room for improvement in diagnosing the severity of bearing faults and diagnosing bearing faults in complex situations such as noise. With the advent of the big data era, the massive amount of data makes manual extraction of fault features increasingly impractical. Therefore, there is a need for an end-to-end identification algorithm that directly uses raw data as input, processes it through the algorithm, and outputs the final result, reducing the steps of manual operation. The emergence of deep neural networks has made this possible. In recent years, deep neural networks have been widely applied in various fields due to their powerful feature representation and extraction capabilities, as well as relatively low computational costs, being widely used in tasks such as image recognition, classification, and detection.

Currently, research on bearing fault diagnosis in noisy environments focuses on deep learning methods. Deep learning has deep structures and powerful nonlinear feature extraction capabilities, which can directly extract fault features from raw data, thus achieving end-to-end fault diagnosis and avoiding the tedious feature extraction and

classification steps in traditional methods. Although deep learning has achieved good results in bearing fault diagnosis research, there are still some issues with current deep learning models. Many deep learning models perform well on clean signal and consistent data distribution datasets. However, when the signal contains a lot of noise, the model's classification performance will sharply decline. In the actual industrial environment, noise is ubiquitous, making it crucial to efficiently diagnose signals containing multiple noises.

Based on the above problems, this chapter proposes a new network architecture called MCFCNN. It includes a Multi-scale processing layer, Channel attention layer, Feature enhancement layer, and Anomaly classification layer network model, used for efficient diagnosis of bearing faults in multi-noise environments.

4.2 MCFCNN Network Architecture

This chapter utilizes time-series vibration data collected during the operation of bearings for fault diagnosis. The diagnostic problem can be defined as follows: input is $X_t \in \mathbb{R}^k$, $t=(0,1, \dots, T)$, where T is the time step and k is the data type, which includes the fault data and health data of the bearing. The corresponding output is the fault type Y_t for model fault diagnosis. Our goal is to achieve fault diagnosis by establishing a mapping relationship between Y_t and the dataset, The representation is as follows:

$$Y_t = f(X) \quad \text{Equation 4.1}$$

where X is the bearing operation data, f is the mapping function, and Y_t is the fault classification result of f diagnosis.

The bearing fault diagnosis network proposed in this research is illustrated in Figure 4.1. The entire network consists of four layers: the Multi-scale Processing Layer,

the Channel Attention Layer, the Feature Enhancement Layer, and the Anomaly Classification Layer . We will discuss each of these sub-structures in detail.

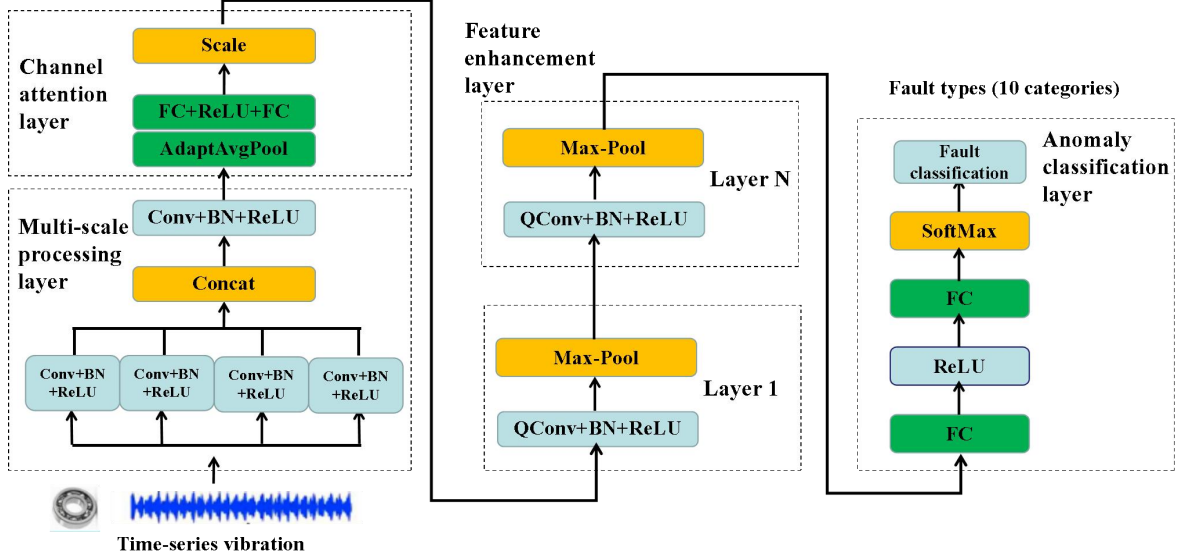


Figure 4.1: MCFCNN network architecture

4.2.1 Multi-scale Processing Layer

The multi-scale processing layer consists of a parallel convolution operation that utilizes multiple parallel convolution branches to extract different features. Each branch can use different convolution kernels to process the input data and generate a set of feature maps. These feature maps extract features from different aspects, which are then fused together to obtain a more comprehensive feature representation. The fusion can be done through simple addition, concatenation, or weighted fusion by learning the weights. The third approach is adopted in the MCFCNN model. This approach provides a richer and more diverse feature representation, helping the network better adapt to different input patterns and enhance its ability to extract complex features.

Assuming the input data is X , the number of convolution kernels is N , the size of the convolution kernel is $K \times K$, the number of channels in the input data is C , and the

number of channels in the output feature maps is M . For each convolution kernel i ($1 \leq i \leq N$), its weight matrix is W_i with a shape of $K \times K \times C$, and the bias term is b_i . The output of the multi-branch convolution operation for the m -th feature map of the input data X can be calculated using the following formula:

$$Y_m = \sum^n f(\text{conv}(W_n, X) + b_n) \quad \text{Equation 4.2}$$

In the given expression, $\text{conv}(W_n, X)$ represents the convolution operation between the input data X and the convolutional kernel W_n . b_n represents the bias term of the n th convolutional kernel. f represents the ReLU function.

In bearing fault diagnosis, different types of faults exhibit different fault characteristic frequencies, which results in different sensitivities of different scales of features to different types of faults. Through multi-branch convolutional operations, it is possible to simultaneously capture vibration features at different scales, thereby improving the detection and diagnostic capabilities of the model for bearing faults. This helps enhance the model's adaptability and generalization ability to bearing faults. Simultaneously, feature fusion at different scales can acquire a more comprehensive and rich feature representation. This aids the model in better understanding vibration signals at different scales and improves its discriminative power for bearing faults.

4.2.2 Channel Attention Layer

Channel Attention Networks (Bahdanau et al., 2016) won the championship in the ImageNet competition for the classification task in 2017. It can be used to enhance the attention mechanism of convolutional neural networks. It improves network performance by adaptively adjusting the importance of features across channels.

The Channel Attention Module Structure: The module consists of two key steps: squeeze and excitation. Squeeze performs global pooling on the input feature map along the channel dimension to obtain global information about the channels. Excitation then uses a learned non-linear function to transform the squeezed features into channel attention weights, which are used to adjust the importance of each channel in the input feature map. Figure 4.2 illustrates the structure of the channel attention module.

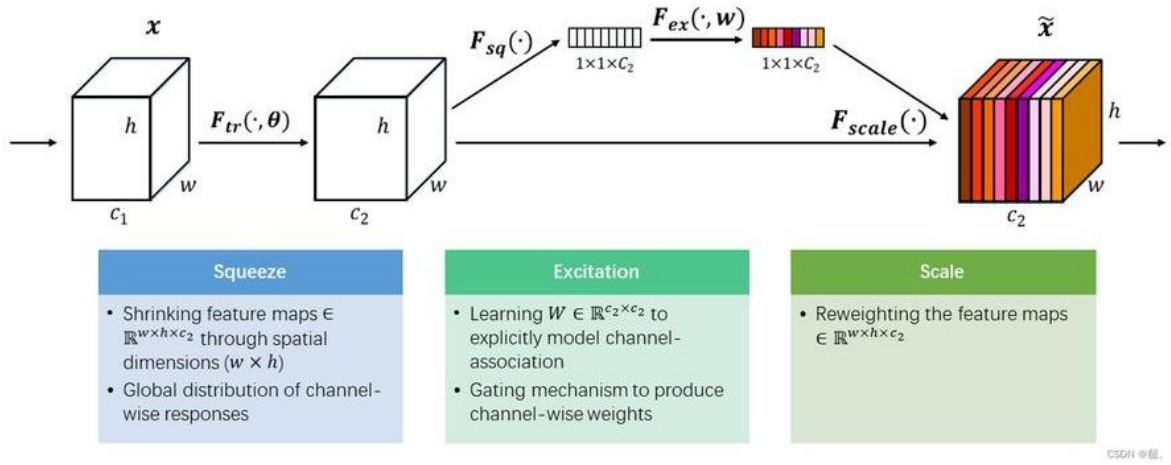


Figure 4.2: Channel Attention Module (Hu et al., 2020)

Squeeze: Assuming the size of the input feature map is (H, W, C) , the squeeze operation performs global average pooling on the input feature map, resulting in a C -dimensional vector that represents the global average value for each channel. This vector can represent the overall importance of each channel in the input feature map. The resulting output, Z , can be represented as:

$$Z = F_{sq}(X) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X(i, j) \quad \text{Equation 4.3}$$

Excitation: The excitation operation generates channel attention weights by learning a nonlinear mapping function. This mapping function can be learned using either

fully connected layers or one-dimensional convolutional layers. Given an input feature vector Z , the result of the excitation operation S can be represented as:

$$S = \sigma (W2 * \delta (W1 * Z)) \quad \text{Equation 4.4}$$

$W1$ and $W2$ are weight matrices, δ representing the ReLU activation function, and σ represents the sigmoid function. S is a C -dimensional vector, representing the attention weights for each channel.

Scale: By performing element-wise multiplication between S and X , the channels of X can be reweighted. Therefore, the value of the weighted feature map Y is:

$$Y = \text{Fscale}(S, X) = S * X \quad \text{Equation 4.5}$$

The channel attention layer can adaptively learn the correlations between feature channels and enhance important features in bearing vibration signals. In bearing fault diagnosis, there are usually limited samples of faulty cases, which means the model needs to learn from limited data and generalize to unseen fault scenarios. The channel attention module can enhance the sensitivity and discriminability of the model towards bearing fault samples. By adaptively adjusting the weights of feature channels, the model can focus more on the features relevant to bearing faults, enabling more accurate identification and classification of faulty samples.

4.2.3 Feature Enhancement Layer

The traditional neural network models commonly use activation functions such as ReLU and sigmoid to capture non-linear relationships. However, these functions can only capture first-order features. Fan et al. (2020) proposes a novel neural network that uses

quadratic neurons to enhance the expressive power of the model and capture higher-order non-linear features. This promotes neuron diversity in deep learning. The distinctive feature of this approach is the integration of two inner products of the input vector and a power term before the activation function. If the input vector is x , the output of the quadratic neuron, denoted as $f(x)$, is:

$$\begin{aligned}\sigma(f(x)) &= \sigma\left(\left(\sum_{i=1}^n w_i^r x_i + b^r\right)\left(\sum_{i=1}^n w_i^g x_i + b^g\right) + \sum_{i=1}^n w_i^b x_i^2 + c\right) \\ &= \sigma((x^\top w^r + b^r)(x^\top w^g + b^g) + (x \odot x)^\top w^b + c)\end{aligned}\quad \text{Equation 4.6}$$

σ represents the activation function, $\odot$ denoting the Hadamard product. w^r , w^g , w^b are weight matrices, b^r , b^g , c are bias vectors.

In previous studies (Fan et al., 2020), it was mentioned that quadratic neural networks have higher feature extraction capabilities. Liao et al. (2022) explains that the mathematical formulation of quadratic neurons can derive a module similar to an attention mechanism, which helps the network focus more on local information. Bearing fault signals, which are generally high-frequency, short-range, and high-amplitude, belong to local feature information. Therefore, in bearing fault diagnosis tasks, especially in complex noisy environments, we choose to replace traditional linear convolution operations with quadratic neural network convolutions.

4.2.4 Anomaly Classification Layer

The vibration signal of the bearing, after a series of feature extraction, needs to be mapped to the sample label space using an anomaly classification layer to achieve fault classification. The MCFCNN model employs two linear layers to reduce the

dimensionality of the features to 10, and then performs fault feature classification using SoftMax function. Specifically, assuming the input vector is x , the linear layer $h(x)$ can be expressed as:

$$h(x) = f(wx + b) \quad \text{Equation 4.7}$$

The output of the Anomaly classification layer $Y(x)$ of MCFCNN can be expressed as:

$$Y(x) = \text{SoftMax}(f(h(x)w_1 + b_1)) \quad \text{Equation 4.8}$$

w and w_1 are weight matrices, b and b_1 are bias vectors. f ReLU is a activation function. $h(x)$ represents the output of the first layer.

4.3 Data Pre-processing and Model Parameter Settings

We select the Case Western Reserve University Bearing Data Set (CWRU), which has been introduced in Chapter 3. In this thesis, we use vibration signals collected at a sampling rate of 12 kHz on the Drive End (DE). The dataset comprises acceleration signals from four bearings, each bearing including data from normal operation and various types of faults. These fault types encompass inner race faults, outer race faults, and rolling element faults. Each fault type is represented by three fault diameters. Hence, the dataset consists of nine faulty bearings and one healthy bearing data. The fault types and corresponding labels are shown in Table 4.1.

First, the original data will be preprocessed to prepare for the subsequent deep learning model training. Since the vibration signals in the CWRU dataset are quite long, we define each sample to have a sequence length of 2048 data points. To increase the

number of samples and enrich the dataset, we employ data augmentation techniques to randomly extract 1000 samples from each long sequence. This means that each type of vibration signal will be segmented into 1000 independent samples, resulting in a total of 10,000 samples.

In this process, all samples will undergo normalization to eliminate dimensional differences between various data features, ensuring that each feature has a similar range during training. This step is crucial for improving the training efficiency and convergence speed of the model.

Finally, to reasonably partition the dataset, all samples will be divided into training, validation, and test sets in a 0.5:0.25:0.25 ratio. Specifically, 50% of the samples (i.e., 5000 samples) will be used for training the model, 25% of the samples (i.e., 2500 samples) will be used for validating the model's performance, and the remaining 25% of the samples (i.e., 2500 samples) will be used for the final testing of the model's accuracy. By doing so, we ensure that the model has good generalization ability throughout the training, tuning, and evaluation processes.

Table 4.1: Ten classes in CWRU data set

Label	Fault Mode	Label	Fault Mode
0	Ball (slight)	5	Inner Race (serious)
1	Ball (medium)	6	Outer Race(slight)
2	Ball (serious)	7	Outer Race(medium)
3	Inner Race (slight)	8	Outer Race(serious)
4	Inner Race (medium)	9	Healthy

Table 4.2: The architecture configuration of MCFCNN

Layer	Type	Kernel	Channel	Stride	Padding	Output
0	input	-	-	-	-	1x2048
1	Conv+BN+ReLU(4)	3x1	4	1	1	4x2048
2	Conv+BN+ReLU	3x1	32	1	1	32x2048
3	Squeeze	-	-	-	-	32x1
4	Excitation	-	-	-	-	32x2048
5	QConv+BN+ReLU	64x1	16	8	28	16x256
6	MaxPool	2x1	-	2	-	16x128
7	QConv+BN+ReLU	3x1	32	1	1	32x128
8	MaxPool	2x1	-	2	-	32x64
9	QConv+BN+ReLU	3x1	64	1	1	64x64
10	MaxPool	2x1	-	2	-	64x32
11	QConv+BN+ReLU	3x1	64	1	1	64x32
12	MaxPool	2x1	-	2	-	64x16
13	QConv+BN+ReLU	3x1	64	1	1	64x16
14	MaxPool	2x1	-	2	-	64x8
15	QConv+BN+ReLU	3x1	64	1	-	64x6
16	MaxPool	2x1	-	2	-	64x3
17	Linear+ReLU	-	-	-	-	192x100
18	Linear	-	-	-	-	100x10
19	Out	10 classes				

The loss function used for training is cross-entropy, and the network is optimized using the SGD optimizer. Gradient clipping technique is applied during iteration to prevent the occurrence of gradient explosion. The batch size for all networks is set to 64, the learning rate is 0.1, and the momentum is 0.9. Through iterative training of the model, the specific architecture of MCFCNN network is defined as shown in Table 4.2. Feature enhancement layer consists of a total of 6 layers, where the first layer uses a large convolutional kernel with a width of 64, and the remaining 5 layers use small

convolutional kernels with a width of 3. This architecture is referenced from W1DCNN (Zhang et al., 2017).

All experiments were conducted on a Windows 10 operating system, equipped with an Intel i5-12490F CPU and an NVIDIA RTX 3060 12GB GPU. Our code was written using Python 3.8 and PyTorch.

In this chapter, we employ three performance metrics: accuracy, precision, and recall, to evaluate the fault diagnosis performance. Additionally, we introduce Gaussian white noise(Yadavar Nikraves et al., 2019) into the dataset to evaluate the model's performance metrics under noisy diagnostic conditions.

4.4 Comparison with Other Methods

The vibration signals of bearings are prone to interference from the surrounding environment and system noise during the data acquisition process. Therefore, the noise robustness of deep learning models is crucial. We evaluated the recognition performance of the MCFCNN model in different signal-to-noise ratio (SNR) environments and compared it with the W1DCNN and QCNN models using the same data preprocessing techniques. The results are presented in Table 4.3. The table compares the recognition accuracy under seven different SNR conditions, and all results are the average of 10 runs. The best results among the three models are highlighted in bold, indicating that the MCFCNN model outperforms the other two models in almost every SNR environment.

Table 4.3: Accuracy of Different Models at Different Signal-to-Noise Ratios

Operating Conditions	SNR	-6	-4	-2	0	2	4	6
CWRU (0HP)	WIDCNN	66.52%	77.36%	87.12%	94.20%	98.36%	99.52%	99.92%
	QCNN	71.12%	79.92%	89.28%	95.76%	98.52%	99.64%	99.96%
	MCFCNN	81.40%	91.04%	96.84%	99.72%	99.92%	99.96%	100%
CWRU (1HP)	WIDCNN	69.68%	72.68%	88.60%	92.28%	98.12%	97.92%	99.12%
	QCNN	70.80%	81.44%	89.82%	96.24%	98.72%	99.40%	99.64%
	MCFCNN	81.88%	92.32%	97.44%	96.64%	98.84%	99.16%	99.96%
CWRU (2HP)	WIDCNN	71.92%	83.72%	90.76%	97.80%	99.08%	99.84%	99.60%
	QCNN	75.88%	85.16%	93.52%	97.96%	99.64%	100%	100%
	MCFCNN	87.6%	97.28%	96.64%	99.72%	99.88%	100%	100%
CWRU (3HP)	WIDCNN	72.40%	82.64%	90.68%	96.60%	99.68%	99.88%	100%
	QCNN	73.96%	84.44%	92.68%	97.56%	99.52%	100%	100%
	MCFCNN	87.66%	92.84%	98.12%	99.72%	100%	100%	100%

4.4.1 Comparison of Different Models

According to the diagnostic results in Table 4.3, we can calculate the average accuracy of each model for the four operating conditions and plot them as line graphs in Figure 4.3. It can be observed that as the signal-to-noise ratio (SNR) increases, the diagnostic accuracy of all three models also improves. This means that when the noise in the signal is relatively low and the difference between the signal and background noise is more pronounced, the models are more likely to accurately identify and classify faults. Figure 5 demonstrates that the MCFCNN model achieves an average accuracy nearly 10% higher than the other two models when the SNR is -6. Particularly in the CWRU dataset, when the SNR is -4 and the load is 2HP, the MCFCNN model outperforms the QCNN model by 12.12% and the WIDCNN model by 13.56%. Additionally, when considering

the cases where the diagnostic results are 100% accurate across all seven SNR levels, W1DCNN only achieves this once, QCNN achieves it four times, and MCFCNN achieves it six times. Therefore, we conclude that MCFCNN consistently provides high performance across various noise environments.

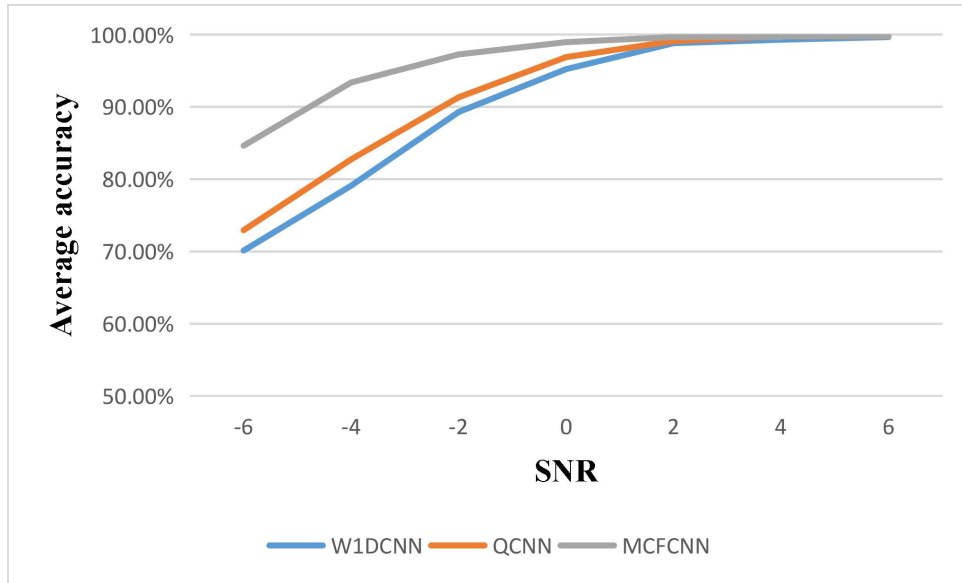


Figure 4.3: Average accuracy of the model under different signal-to-noise ratios

Figure 4.4 shows the training loss graphs for three models under the 1HP condition. From the graph, it can be observed that the loss exhibits a decreasing trend overall. As training progresses, the loss gradually decreases, indicating that the models are optimizing during the learning process. Specifically, it can be seen that the W1DCNN has the highest loss value, followed by QCNN, while MCFCNN has the lowest loss. This suggests that the MCFCNN model performs relatively well in terms of loss. By observing the fluctuation of the loss, the stability of the model can be preliminarily assessed. It can be observed that the MCFCNN has relatively smaller fluctuations in loss, indicating a more stable model with better adaptability to input data.

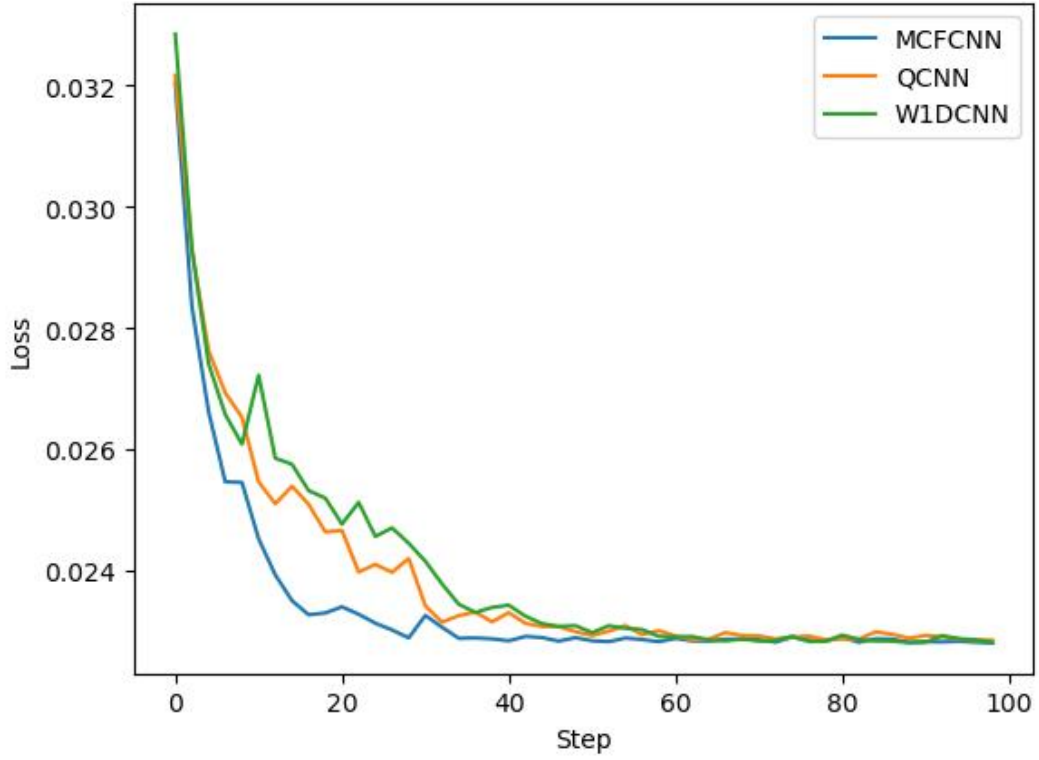
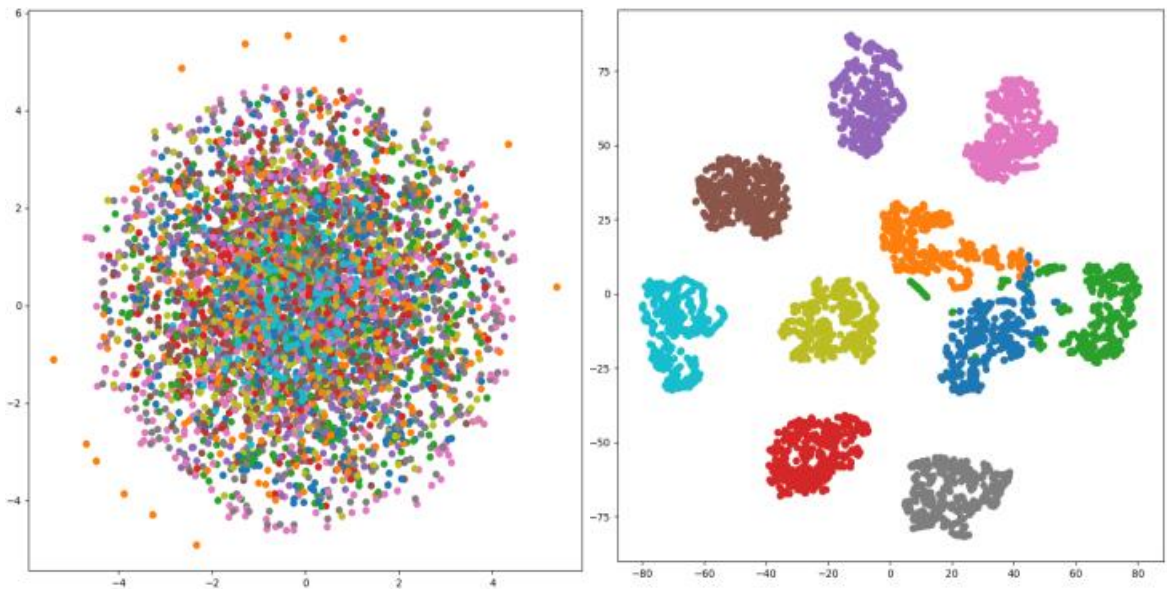


Figure 4.4: Loss of Different Models (1HP)

4.4.2 T-SNE Visualization

When the signal is heavily disturbed by noise, the fault recognition of bearings is also compromised. For example, when $\text{SNR}=-6$ and the load is 0HP, the classification accuracy of the W1DCNN model is 66.52%, QCNN is 71.12%, but MCFCNN still maintains a high recognition rate of 81.40%. Since QCNN and MCFCNN have similar network architectures and quadratic neuron networks, this suggests that the parallel convolution and SE modules in our proposed network play a positive role in recognizing fault features. To better compare the patterns learned by the three models in the training data, we output the results of the last convolutional layer and then reduce the high-dimensional data to 2D. We divide the data into 10 fault categories and use t-SNE (Pouyet et al., 2018) for visualization. t-SNE transforms the similarities in high-dimensional data into a structure in a lower-dimensional space, preserving local neighborhood relationships,

to visualize the ten types of bearing faults. As shown in Figure 4.5, different colors represent different categories. From the visualization results, it can be observed that all three models have learned patterns of different categories, and the same fault categories are clustered together. However, there may be some overlap in the classification of certain categories. For example, in the W1DCNN results, the orange, blue, and green categories are not well distinguished at the boundaries, especially the green category is scattered in clusters, indicating that this model did not learn well for the green category. In the QCNN results, it can be seen that there is some improvement, as the orange category is completely separated, but the blue and green categories are not completely distinct, and the purple and dark yellow categories are close to each other at the boundaries. In the MCFCNN results, although there are some points of the blue and green categories close to the orange category, their number is relatively small, and overall, the boundaries of the three categories are distinguished.



(a) Raw Data

(b) W1DCNN

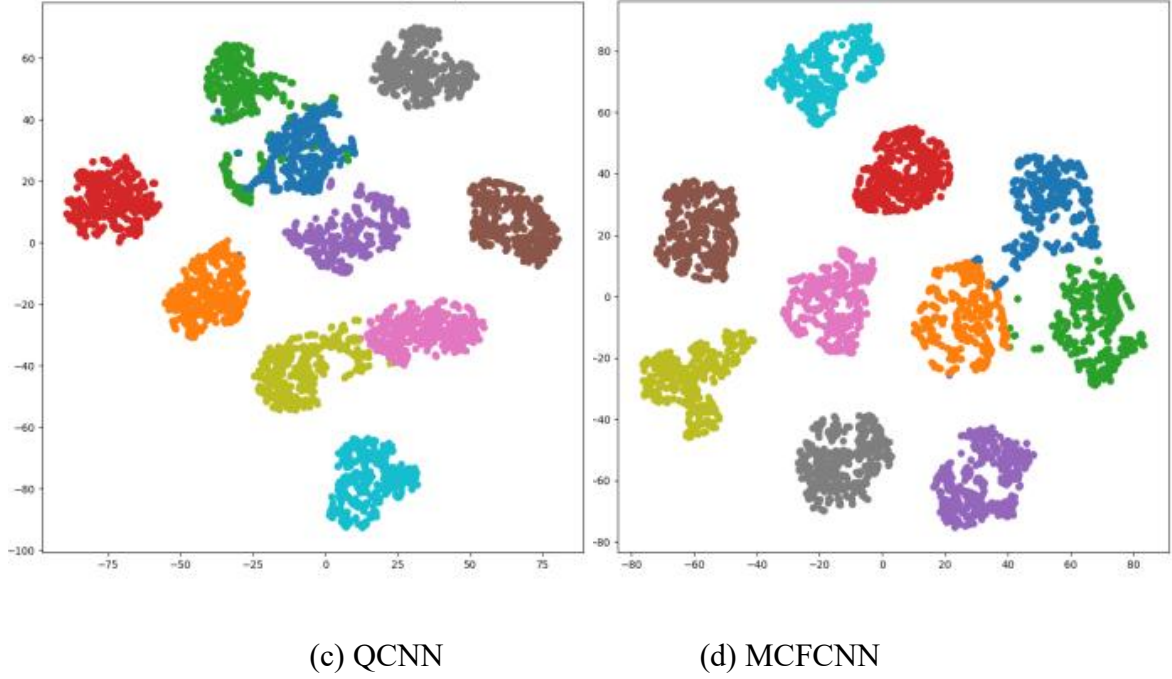


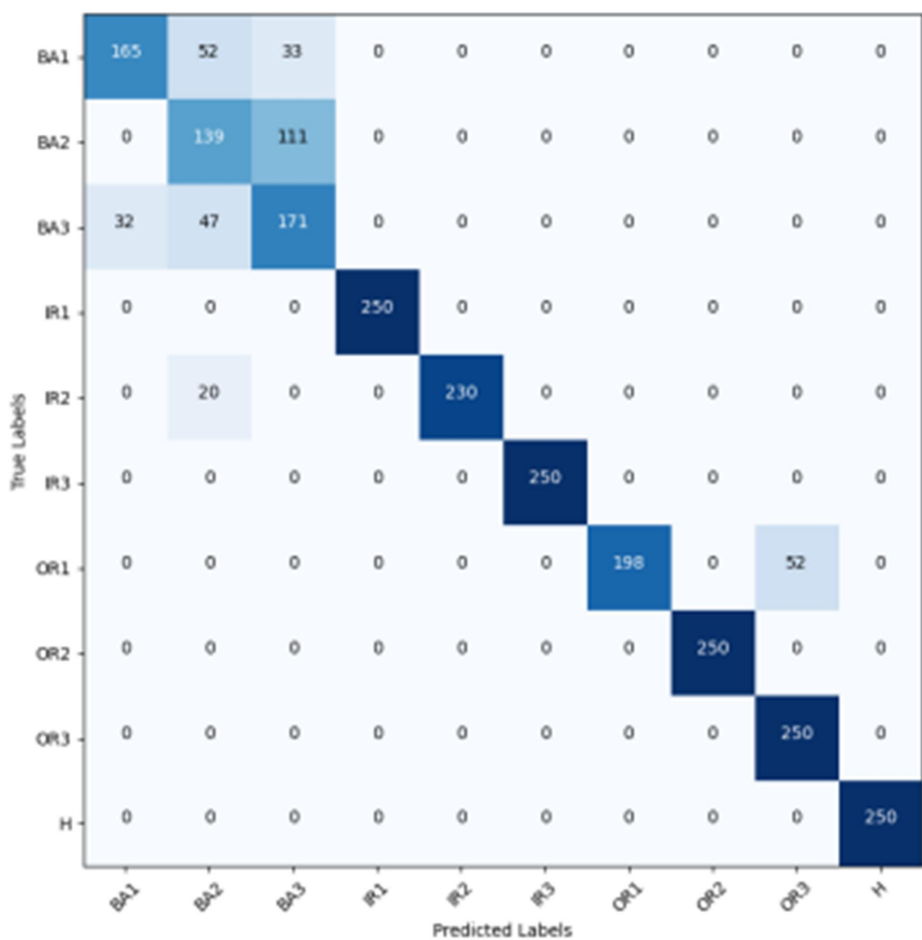
Figure 4.5: Different models' t-SNE visualization results on the CWRU dataset (1HP, SNR = -4)

4.4.3 Confusion Matrix Comparison

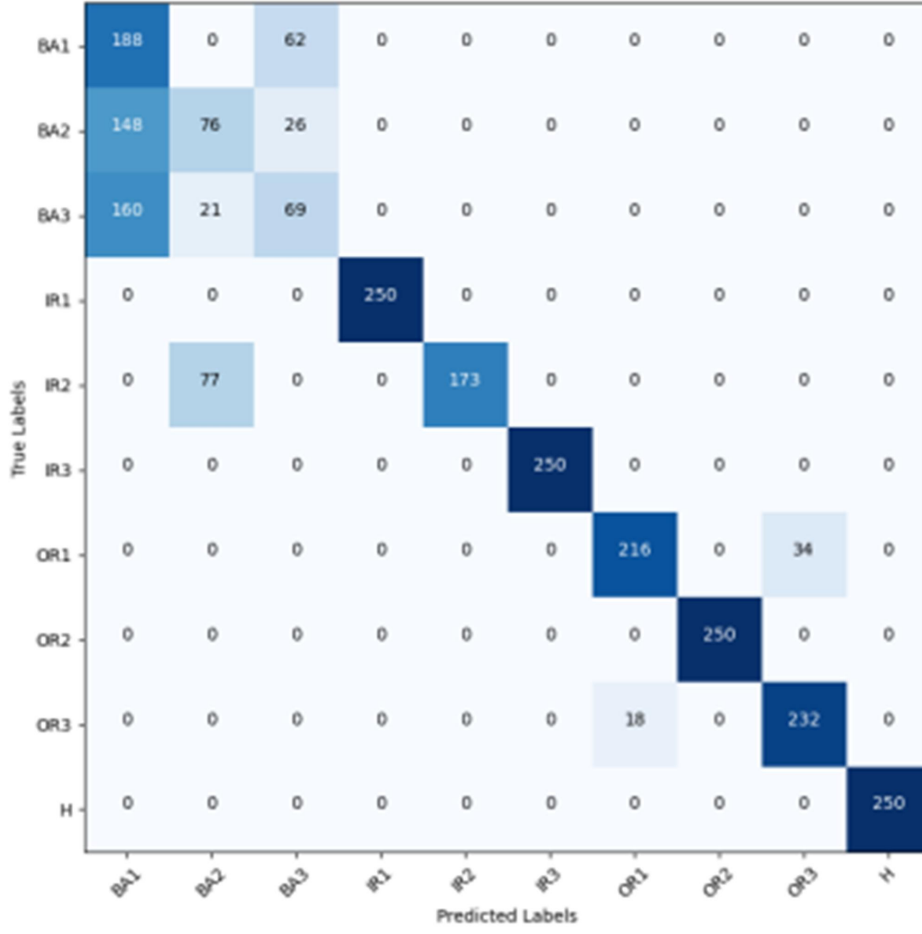
The classification performance of fault modes under strong noise is analyzed in detail by calculating the confusion matrices of the MCFCNN and QCNN models. The values on the diagonal of the confusion matrix represent the number of correctly identified instances. Numbers outside the diagonal represent the number of misclassified instances. Figure 4.6 shows the classification results under strong noise. It can be observed that, under the interference of strong noise, the QCNN model accurately identifies 4 categories, while the MCFCNN model identifies 5 categories correctly. Specifically, both models accurately identify the healthy bearing. Both models misclassify 52 instances as outer race faults, indicating errors in identifying the size of the outer race faults. In the case of inner race faults, the MCFCNN model misclassifies 20 instances, while the QCNN model

misclassifies 77 instances, with the QCNN model mistakenly classifying inner race faults as rolling element faults.

The performance of both models in classifying rolling element bearing faults is relatively low, primarily due to their inability to accurately identify the fault size of the rolling elements. In comparison, the MCFCNN model ($475/750=63.33\%$) accurately identifies more rolling element faults than the QCNN model ($333/750=44.4\%$). Lastly, we are more interested in the number of misclassified samples for fault type identification. Despite strong noise interference, the MCFCNN model achieves 100% accuracy in recognizing fault types, while the QCNN misidentifies 77 samples of inner race faults as rolling element faults. All results indicate that the MCFCNN model performs better.



(a) MCFCNN



(b) QCNN

Figure 4.6: Shows the confusion matrices for MCFCNN and QCNN at SNR = -4. BA, IR, OR, and H respectively represent rolling element fault, inner race fault, outer race fault, and healthy condition.

4.5 Analysis of MCFCNN

Our proposed model performs exceptionally well in a strong noisy environment. To determine which part of the model has a higher ability to extract fault signals from the bearing, we decomposed the MCFCNN model. We separately added the multi-scale processing layer and the channel attention layer to the network for comparison. For ease of identification in the future, the network with only the multi-scale processing layer is referred to as MFCNN, and the network with only the channel attention layer is referred to as CFCNN.

According to Table 4.4, it can be observed that adding these two layers separately improves the classification performance of the model. Specifically, adding the channel attention layer alone outperforms adding the multi-scale processing layer alone by 3.46%. This indicates that in a noisy environment, the channel attention layer has a better capability to extract fault features and assign more weights to them. However, in a strong noisy environment, the MCFCNN model has higher accuracy, precision, and recall rates.

Table 4.4: Comparison of model results (SNR=-6)

Methods	Accuracy	Precision	Recall
MCFCNN	0.8312	0.8899	0.8452
MFCNN	0.7716	0.7933	0.7908
CFCNN	0.8065	0.8215	0.8192

To investigate the patterns learned by the multi-scale processing layer and channel attention layer in network training, we define them as the first layer of the MCFCNN network. We apply visualization techniques to observe the information focused on by the first layer of the network when fed with noisy healthy and faulty bearing signals. Figure 4.7 presents the results, with the left side showing the healthy bearing signal and the right side showing the faulty bearing signal. We display the original signal, the signal with added noise, and the output signal after passing through the first layer of the MCFCNN network. A higher value on the y-axis indicates higher importance. Comparing the outputs of the healthy and faulty bearings, we observe that the network pays more attention to the faulty signal, as indicated by the smaller amplitude in the healthy bearing case. From Figure 4.7 (b), it is evident that despite severe noise interference in the original signal, the

network is still able to capture the location of the fault signal. This also explains why the MCFCNN maintains good classification performance even under strong noise conditions.

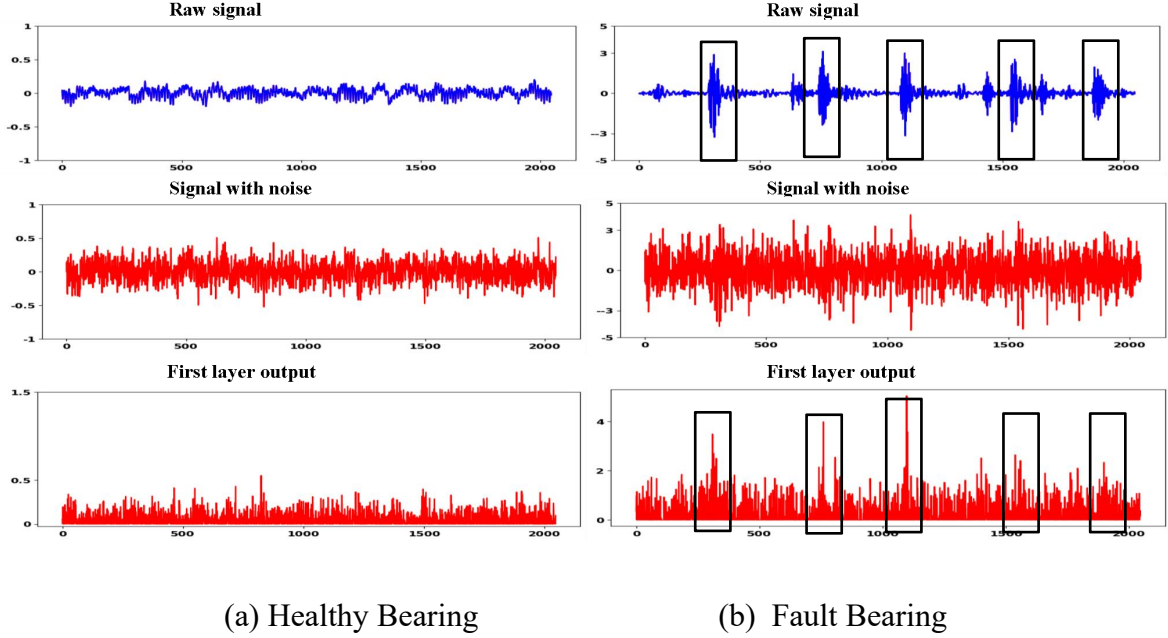


Figure 4.7: The output feature map of the first layer of the MCFCNN

4.6 Chapter Summary

In this chapter, we proposed a convolutional network called MCFCNN for end-to-end bearing fault diagnosis. Additionally, the network achieves good classification performance on the CWRU bearing dataset, maintaining above 80% classification accuracy even under strong noise interference. The proposed model incorporates secondary neuron convolutional network operations, alongside the inclusion of multiscale processing layers and channel attention layers to enhance the model's performance in adapting to strong noise for diagnostic tasks. Moreover, classification tests are conducted on the multiscale processing layers and channel attention layers to validate their significance in extracting bearing fault features. We employed visualization techniques to interpret the output feature maps. In the future, further refinement within the network is needed to enhance its practical applicability.

CHAPTER 5

VALIDATION OF MCFCNN MODEL ON BFTS TEST BENCH

5.1 Chapter Overview

In the previous chapter, we introduced a neural network named MCFCNN designed for end-to-end bearing fault diagnosis. The network demonstrates outstanding classification performance on the CWRU bearing data set, maintaining classification accuracy above 80% even under strong noise interference. To further validate the network's high generalization performance, this chapter constructs a bearing fault collection platform and conducts a more comprehensive performance evaluation on the network using the collected data set.

In this chapter, our goal is to verify the adaptability of MCFCNN to real bearing fault data and examine its performance under different environmental conditions. By building the bearing fault collection platform, we can simulate various fault scenarios that may occur in real industrial settings, thereby evaluating the network's robustness and generalization ability more comprehensively. During the data set collection process, we focus on covering various types and degrees of bearing faults to ensure the network can effectively respond to diverse fault patterns. Additionally, we consider noise and interference factors that may exist under different working conditions to ensure MCFCNN can maintain efficient performance in complex environments. Through training and testing on the newly collected data set, we will comprehensively assess the performance of MCFCNN and further optimize the network's structure and parameters to meet the bearing fault diagnosis requirements under various real operating conditions.

5.2 The BFTS Test Bench

5.2.1 Machinery System

The core component of the BFTS bearing fault simulation platform is a mechanical system equipped with bearings, which collaborates with an electrical control cabinet to enable adjustments of different speeds, loads, and operating conditions. This ensures the simulation of bearing operation under various working conditions during data acquisition, providing diversified input data for the MCFCNN network. In the mechanical part, as shown in the Figure 5.1, the far right comprises a three-phase asynchronous motor, with two bearing housings on the left, each containing a bearing. These two bearing housings and the motor are connected via a coupling, allowing the motor to drive the rotation of the bearings. The detailed information about the electric motor and bearings is shown in Figure 5.2, Table 5.1, and Table 5.2.

The design of this mechanical system permits the adjustment of the motor's operating parameters, including speed and load, through the electrical control cabinet. This controllability enables us to collect bearing operation data under different working conditions, providing the network with more complex and diverse input data, further validating the generalization performance of the MCFCNN network.

In experiments, we will manipulate the motor's operating parameters to simulate the operational status of bearings under various working conditions encountered in real industrial scenarios. This design ensures that we can comprehensively evaluate the performance of the MCFCNN network in bearing fault diagnosis tasks from multiple perspectives.

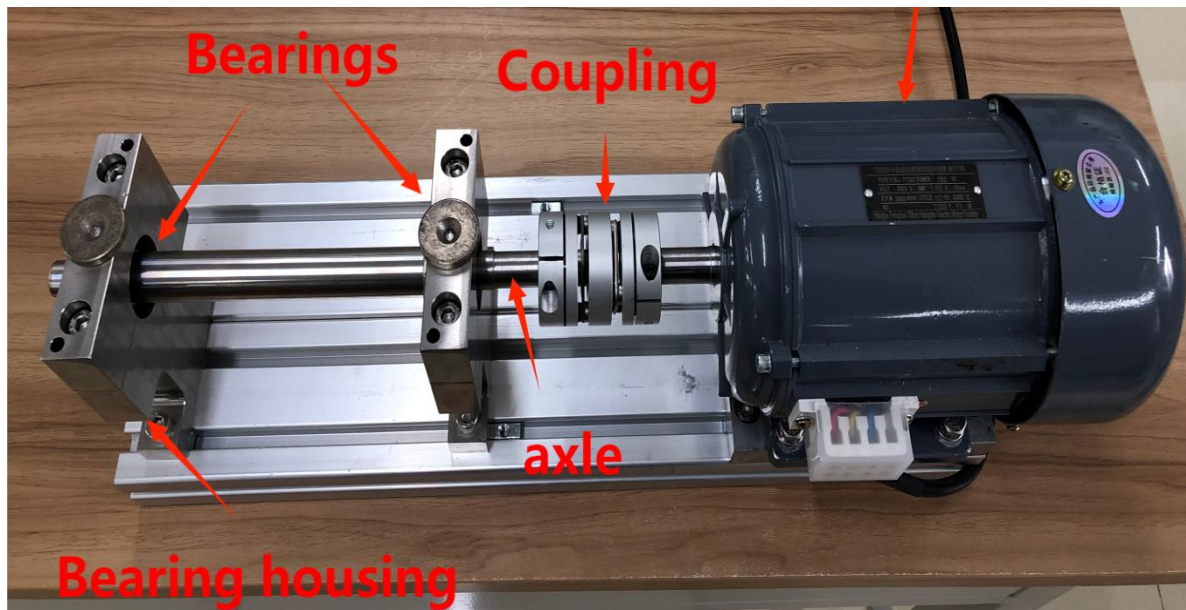


Figure 5.1: Mechanical Part of BFTS Test Bench



Figure 5.2: Bearings

Table 5.1: Motor parameters

Parameter	Specification
POWER	750W
VOLT	380V
AMP	1.8A
R.P.M	2800r/min
CYCLE	50Hz

Table 5.2: Bearing parameter

Type	Outer Diameter	Bore Diameter	Width
6203	40mm	17mm	12mm

5.2.2 Accelerometer

On our platform, each bearing housing is equipped with an accelerometer sensor, designed to comprehensively capture vibration information during bearing operation. The acquisition of this vibration signal is crucial for accurately simulating potential fault signals that may exist in real industrial environments. By deploying accelerometer sensors on each bearing housing, we are able to monitor and record vibration characteristics of the bearing operation in a timely and comprehensive manner. The sensors selected for the platform are shown in Figure 5.3 and the parameters are given in Table 5.3.

This collection of vibration information not only enables us to gain a deeper understanding of the bearing's operational condition but also provides more representative input data for the MCFCNN network. Changes and patterns in vibration signals can reflect various faults that may exist in bearings, such as ball damage, inner and outer race cracks,

or other anomalies. Therefore, the vibration information captured by accelerometer sensors provides critical diagnostic features in bearing fault diagnosis.



Figure 5.3: Acceleration sensor

Table 5.3: Acceleration sensor parameter

IEPE accelerometer	Sensitivity	Range	Frequency Response	External Dimensions	Weight
	50mv/g	100g	0.5-10kHz	Ø 12 * 20	10g

5.2.3 Data Collection and Control Unit

The platform is equipped with a dedicated data acquisition and analysis system along with a control unit for real-time collection of sensor data, controlling mechanical system parameters, and ensuring the accuracy and synchronicity of data acquisition. This unit also enables real-time monitoring of bearing status to promptly adjust experimental conditions.



Figure 5.4: Dynamic signal analyzer

The dynamic signal analyzer used by the platform is model HFUSB-4004, as shown in the Figure 5.4. The system is equipped with a signal acquisition and analysis software designed specifically for vibration, noise, and strain testing techniques, as shown in Figure 5.5. The software features dual functionality for signal acquisition and post-processing analysis, supporting seamless switching between two operating modes: online mode and offline analysis mode. The entire system is developed using C++ programming language, based on object-oriented programming principles, and incorporates a fully plugin-based architecture to ensure system stability and reliability.

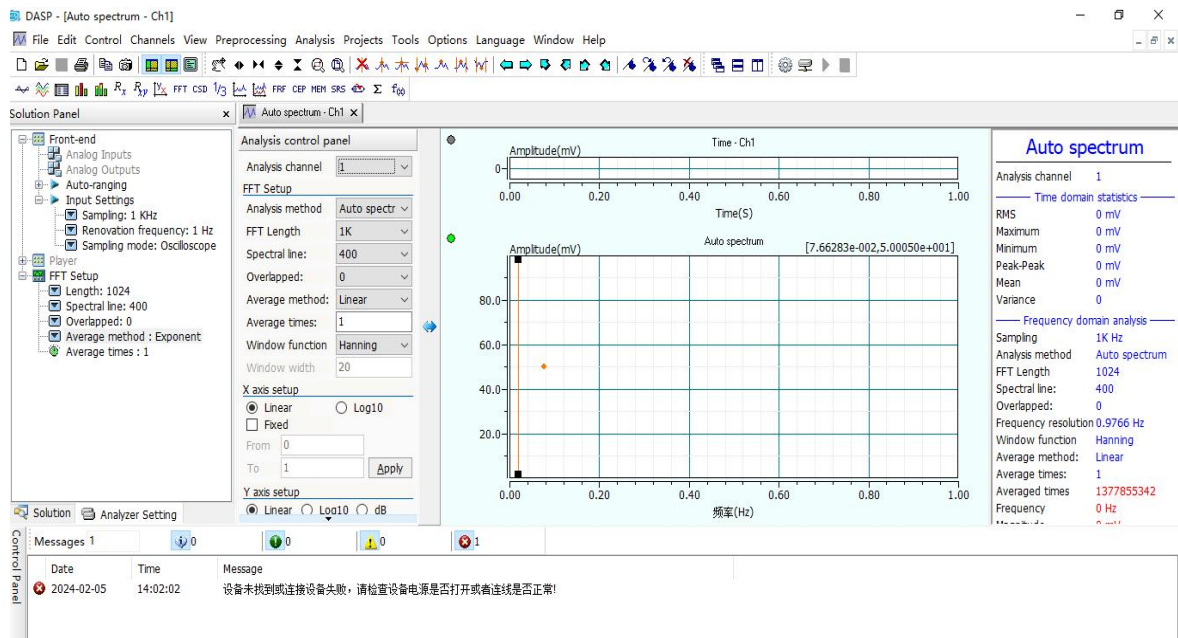


Figure 5.5: Dynamic signal analyzer software interface

The HFUSB-4004 system is designed with considerations for user-friendliness and convenient data management. It supports management functionalities for solutions and project documents, making the organization and retrieval of data extremely convenient. The system also supports virtual TEDS sensors, enabling unified management of sensors through a database, thereby enhancing usability.

Table 5.4: HFUSB-4004 parameters

Parameter	Specification	Parameter	Specification
Channel Count	4	Built -in gain	×1 、 ×10 、 ×100
A/D	24	Accuracy	< 0.5 %
Input Type	Voltage/IEPE	Maximum Sampling Rate	Parallel synchronization 128kHz /channel
IEPE Incentive Source	4mA /24V	Transmission Interface	USB2.0
Signal Input Range	±10VPEAK	Dimensions (mm)	110W×30H×126D
Signal Input Bandwidth	Voltage: DC~100kHz(-3dB) IEPE: 0.3Hz~100kHz(-3dB)	Weight (grams)	300
Filter	Multiple -grade simulation filter plus digital anti -mixing filtering	Power	USB interface self - power supply

This dynamic signal analyzer boasts a rich set of time-domain and frequency-domain analysis processing capabilities, including but not limited to single trace

oscilloscope view, multi-trace time-domain oscilloscope, Fourier analysis, autospectral analysis, power spectrum analysis, and power spectral density analysis, as well as transfer function analysis. These features empower users to delve deeply into the analysis and understanding of collected dynamic signals, providing powerful tools for fault diagnosis and performance evaluation.

Specific parameters of the dynamic signal analyzer can be found in Table 5.4. The introduction of this tool will contribute to a more comprehensive and accurate collection and analysis of vibration information during bearing operation, thereby enhancing the monitoring and diagnostic capabilities of bearing conditions.

The control unit, as a crucial component of the bearing fault simulation platform, incorporates elements such as the electrical control cabinet, motor frequency converter, and filters, playing a vital role in system management and control.

Electrical control cabinet: Throughout the entire system, the desktop electrical control cabinet serves as the core component for power supply, user interface operations, and electrical protection. Through its power unit, the system obtains the required stable power supply. The control panel provides an intuitive interface for operators to monitor and adjust system parameters. Additionally, the presence of electrical protection devices helps ensure the safe operation of the system under various working conditions.

Motor frequency converter: The motor frequency converter is tasked with adjusting the operating speed of the electric motor in the system. Its speed regulation function enables us to simulate different motor speeds under various working conditions. Through frequency control, this device facilitates flexible control of the electric motor. Its presence

ensures the system's response speed under different loads and speeds, providing a variety of working conditions for experiments.

Filters: Filters are critical equipment for ensuring the stability of power supply. Their filtering function helps remove high-frequency interference from the power source, providing smooth power output. The design of this component aims to ensure that the system is not affected by power fluctuations during operation, thereby guaranteeing the accuracy and reliability of data collection.

These components work together to ensure the efficient operation of the control unit and flexible control of system parameters. Through adjustments to the control unit, various working conditions can be simulated, providing more comprehensive and realistic bearing fault data for the MCFCNN network. This supports further research into the network's performance under actual operating conditions.

The design of this bearing fault simulation platform is distinctive, primarily manifested in several aspects: firstly, comprehensive collection of vibration signals through acceleration sensors. Secondly, the flexibility to adjust mechanical system parameters allows experiments under different speeds and loads, expanding the scope of data collection. The platform also implements real-time monitoring of bearing status and dynamic adjustment of experimental conditions to ensure the collected data is highly representative. The establishment of this bearing fault simulation platform provides robust support for exploring the potential application of the MCFCNN network in practical bearing fault diagnosis. Figure 5.6 shows a physical diagram of the platform we built.

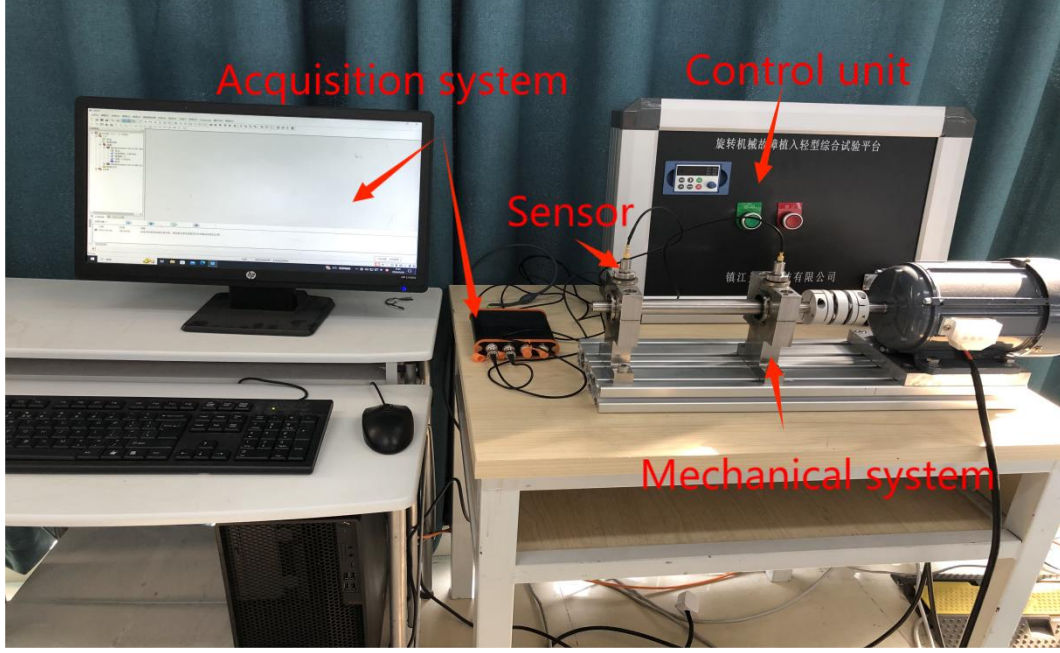


Figure 5.6: BFTS Test Platform

5.3 Data Sets Collected by the BFTS Test Platform

To test the vibration signals of faulty bearings, we created a total of nine faulty bearings and one healthy bearing of the same model. Each faulty bearing was artificially manufactured, with specific fault types including crack faults, further categorized into outer race faults, inner race faults, and rolling element faults. Each fault type was further subdivided into three different levels: mild, moderate, and severe.

This design allowed us to construct a test dataset containing a rich variety of fault patterns, covering different degrees and types of crack faults. In this manner, we could comprehensively evaluate the diagnostic performance of the MCFCNN network for various types of faults, thereby providing a deeper understanding and effective solutions for bearing fault diagnosis. See Table 5.5 for bearing fault labels and types. The slight, medium, and serious in the table respectively refer to the degree of failure of the faulty bearing at different positions. For example, Ball (slight) refers to the location of crack

damage on the rolling element, with a depth of 0.2mm and a width of 0.2mm; Inner Race (medium) refers to the crack breaking in the inner race, with a depth of 0.5mm and a width of 0.3mm; Outer Race (serious) refers to crack damage on the outer race, with a depth of 0.5mm and a width of 0.5mm.

Table 5.5: Bearing failure type and label

Label	Fault Mode	Label	Fault Mode
0	Ball (slight)	5	Inner Race (serious)
1	Ball (medium)	6	Outer Race(slight)
2	Ball (serious)	7	Outer Race(medium)
3	Inner Race (slight)	8	Outer Race(serious)
4	Inner Race (medium)	9	Health

The data set collected from the BFTS experimental platform exhibits similarities to the CWRU data set. To further validate the diagnostic performance of the MCFCNN model in practical applications, we collected datasets under three different operating conditions on this platform. In this study, we adopted a random setting for motor speeds during data collection. This approach allows us to simulate various working conditions, including different motor speeds. The randomness of this setup helps to comprehensively cover a wide range of possible real-world scenarios, enhancing the diversity of the dataset. This further ensures the robustness and generalization performance of the MCFCNN model in addressing bearing fault diagnostic tasks under different motor speeds.

Following the signal preprocessing methods outlined in Chapter Four, we segmented the raw signals based on the number of data points collected during a specific

time of bearing rotation. The length of each sample was set to 2048 data points. Additionally, we employed data augmentation techniques to process the raw data. There are a total of 10 different bearing datasets, with each category containing 1000 samples, resulting in a total of 10,000 samples. Subsequently, we randomly selected data for each fault type to ensure a ratio of 5:2.5:2.5 for the training, validation, and testing sets, respectively.

5.4 Experimental Results

In order to verify the performance of the network under high noise conditions, we added different levels of Gaussian white noise to the data in the same way as in Chapter 4. A comparison between the original rolling element fault signals and those with added Gaussian white noise is shown in Figure 5.7.

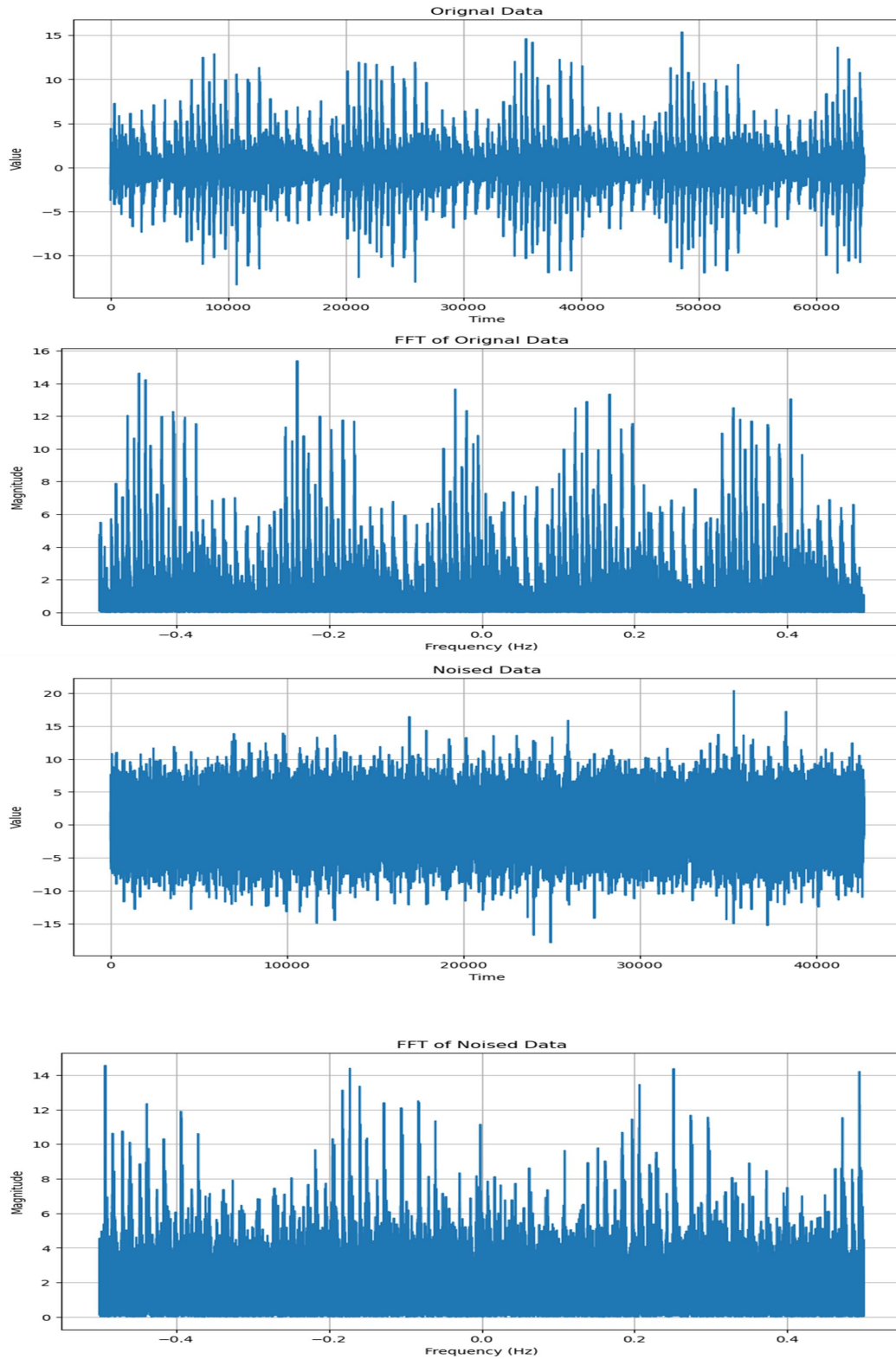


Figure 5.7: Comparison between original signals of rolling element faults and signals with added noise (SNR=0)

From the figure, it can be observed that the time-domain signals without noise exhibit clear periodicity and prominent amplitude modulation features, whereas the time-domain signals with added Gaussian white noise are difficult to distinguish manually. Comparing the frequency spectrum information, it is evident that the noise spectrum is uniformly distributed across the entire frequency domain, and the high-frequency components of the original spectrum are largely submerged within the noise spectrum.

In Chapter 4, we observed that the MCFCNN model achieved the highest recognition accuracy across various signal-to-noise ratios (SNRs) in the CWRU dataset. Similarly, as depicted in Figure 5.8, the MCFCNN model also exhibited the highest accuracy in the BFTS dataset. This indicates that the MCFCNN model possesses good robustness and stability in handling signal classification tasks, achieving optimal performance across different datasets and SNR conditions.

Specifically, in the BFTS dataset, the accuracy of MCFCNN was significantly higher than that of WDCNN and QCNN across all SNR conditions. For example, at SNR 0, the accuracy of MCFCNN was 97.95%, while that of WDCNN and QCNN were 90.73% and 96.91%, respectively. This suggests that the MCFCNN model exhibits better robustness and performance advantages in handling signal classification tasks. With increasing SNR, the accuracy of all models improved. However, the extent of performance improvement may vary among different models. For instance, under SNR 2 conditions, the accuracy of MCFCNN increased to 99.48%, while that of WDCNN and QCNN increased to 97.81% and 98.85%, respectively. This indicates that the performance improvement of MCFCNN under high SNR conditions is more significant.

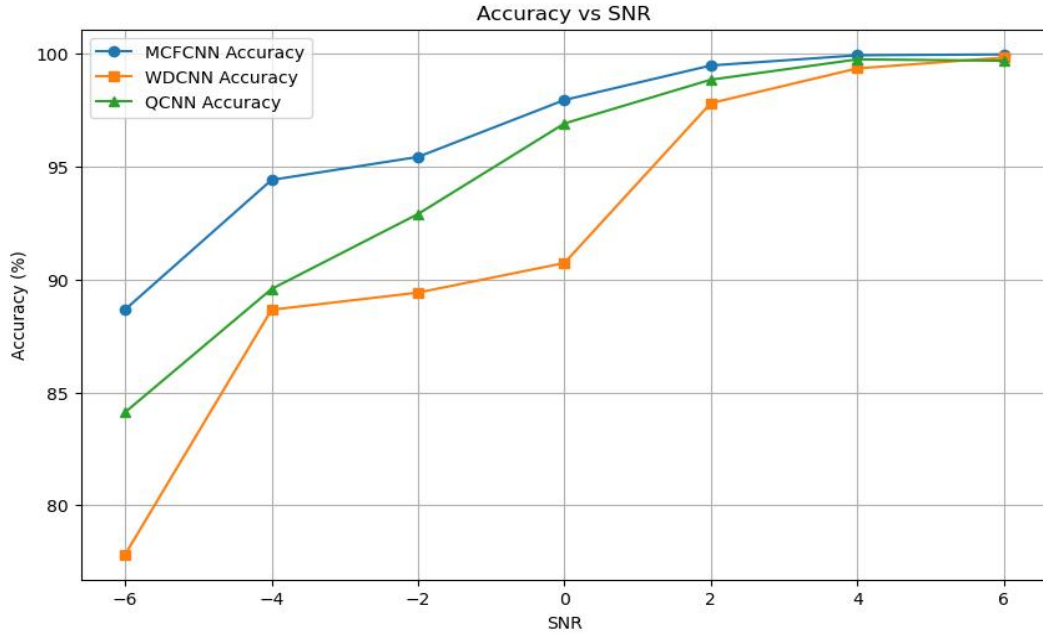


Figure 5.8: Contrasting Different Methods on Experimental Data set Across Various Signal-to-Noise Ratios

5.5 Ablation Experiment

In this section, we also utilize comparative experiments to assess the contributions of different augmentation methods in the proposed algorithm to the final results. The experimental results are depicted in Figure 5.9, where MCFCNN (ELU) denotes the utilization of ELU as the activation function in the feature extraction layer instead of ReLU, and MCFCNN (NoRS) indicates the absence of data augmentation strategies in the network.

From Figure 5.9, it can be observed that when there are no data augmentation strategies employed in the network, improving model recognition accuracy becomes challenging, demonstrating the effectiveness of data augmentation strategies in enhancing the network's noise robustness. When ELU is used as the activation function in the network's feature extraction layer instead of ReLU, the classification accuracy drops by a

maximum of 2.11 percentage points, indicating that ELU activation function can provide better noise adaptability to the network.

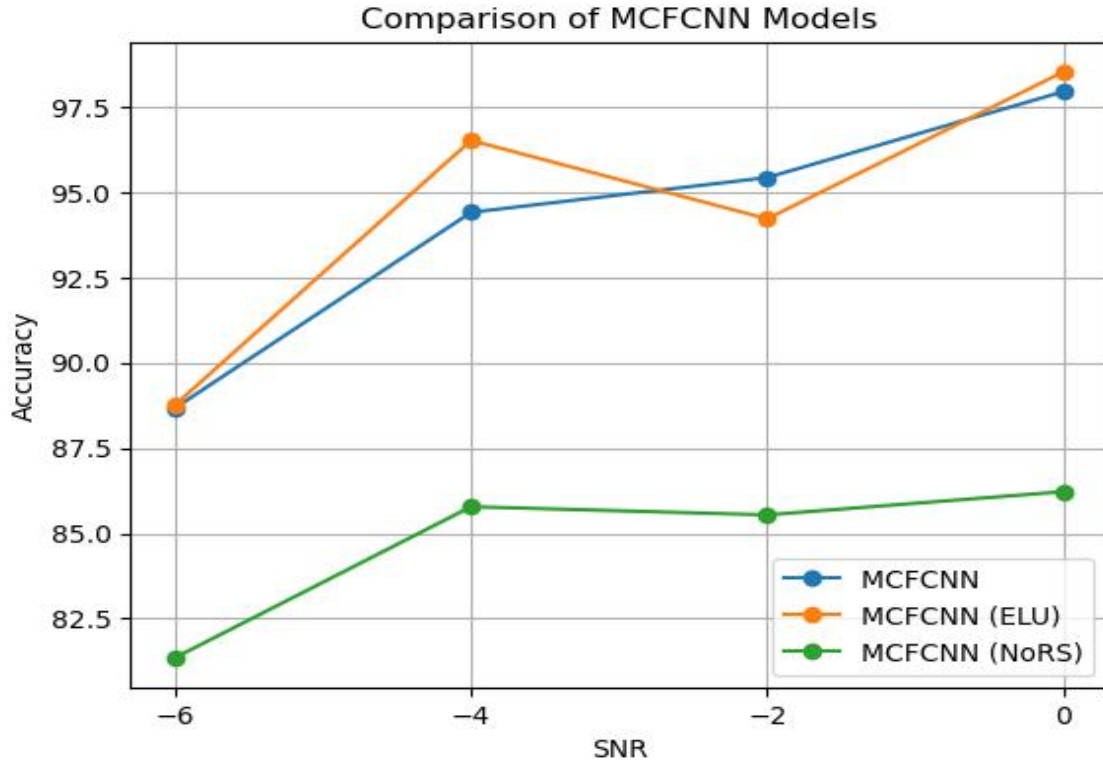


Figure 5.9: Comparison on experimental data set with different network structures at different signal-to-noise ratios

To validate the proposed MCFCNN network's model transferability under different operating conditions, we train the network under one condition and test its performance under other conditions. In the following experiments, we utilize the BFTS dataset. In these experiments, we continue to use time-domain signals as inputs without manually extracting features from the signals, and we discuss the detection accuracy at a signal-to-noise ratio of -6. The training and testing data are randomly selected with a ratio of 8:2. Comparative results are shown in Figure 5.10, with a total of six sets of comparative experiments corresponding to three load conditions. "1-2" in the figure indicates that the model is

trained under condition 1 and tested under condition 2. For further classification, condition 1 refers to speed of 500 r/min; condition 2 refers to speed of 1000 r/min; condition 3 refers to speed of 1500 r/min.

According to the provided bar chart data, the dashed line represents the average resolution of each model under all 6 cross operating conditions, we can draw the following conclusions: Overall, the MCFCNN model performs the best, with an average performance score of 0.7595, compared to 0.6492 for WDCNN and 0.7266 for QCNN. Specifically, MCFCNN performs best under condition 2-1, with a performance score of 0.8126, possibly due to better generalization of the model to data under this condition during training. In contrast, WDCNN performs notably worse under conditions 2-3 and 3-1, suggesting the need for more optimization or specific datasets for these conditions. In general, the MCFCNN model demonstrates high stability and generalization under different operating conditions, confirming its adaptability as a more capable network model.

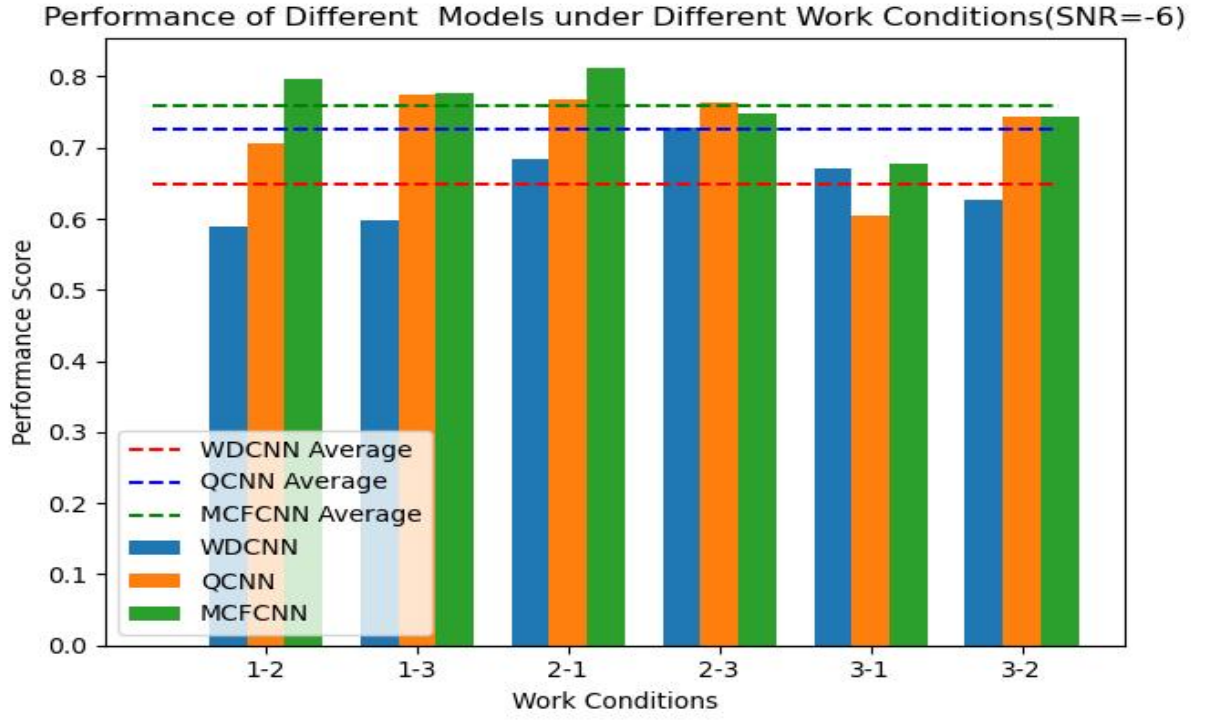


Figure 5.10: Comparison between different models under different loads on the experimental data set(condition 1 500r/min condition 2 1000r/min condition 3 1500r/min)

5.6 Chapter Summary

This chapter primarily introduces the BFTS test platform and the dataset collected on this platform, aiming to validate the adaptability of MCFCNN to real bearing fault data and examine its performance under different environmental conditions. Firstly, we provide a detailed description of the composition of the BFTS bearing fault simulation platform, including the mechanical system, accelerometer sensors, and data acquisition/control unit. This platform allows for the simulation of bearing operating states under various conditions and provides diverse input data, laying the foundation for a comprehensive evaluation of the MCFCNN network. Next, we present the experimental design conducted on the BFTS dataset, including the fabrication of faulty bearings, data collection methods, and setting of signal-to-noise ratios. Through comparative experimental results, we

validate the performance advantages of the MCFCNN model under high noise conditions, as well as the influence of different network structures and data augmentation strategies on the final results. Finally, we conduct model transfer experiments to verify the stability and generalization ability of the MCFCNN model under different operating conditions. Overall, the experimental results of this chapter demonstrate that the MCFCNN model exhibits excellent robustness and generalization ability, performing well in bearing fault diagnosis tasks.

CHAPTER 6

BEARING DEGRADATION PREDICTION BASED ON FAEK AND MCC

6.1 Chapter Overview

In today's society, with the rapid development of technology and the continuous advancement of industrialization, the reliability and performance of equipment and systems have become important focal points across various fields. Prognostics and Health Management (PHM) (Xia & Xi, 2019), as an advanced technology and approach, offers an effective means for achieving equipment failure prediction, optimizing maintenance, and enhancing reliability through real-time monitoring, diagnostics, and prognostics of equipment health.

PHM utilizes operational data from equipment to predict its health condition and enable precision maintenance. Compared to traditional methods involving routine inspections and regular checkups, PHM provides an intelligent maintenance approach. In conventional maintenance, determining the optimal timing for maintenance can be challenging, but with PHM, it becomes possible to accurately predict which stage of the life-cycle the equipment is in, allowing for the implementation of appropriate maintenance strategies. This precise maintenance approach significantly enhances the efficiency and cost-effectiveness of equipment maintenance while also extending its operational lifespan. In the context of managing large-scale equipment and complex operational requirements, PHM technology offers equipment managers a more intelligent and efficient maintenance approach.

This chapter will focus on bearing components, which are critical parts of many rotating devices. Their failure can lead to equipment downtime and production interruptions, causing significant losses to businesses. PHM technology can establish fault prediction models by collecting and analyzing vibration data from bearings. By monitoring changes in the bearing's condition, it becomes possible to predict its remaining lifespan. Currently, there are two methods for predicting the remaining lifespan of bearings, 1) Direct prediction method (Li et al., 2019; Wang et al., 2019; Ruan et al., 2023). This method relied on machine learning models to directly predict the bearing's lifespan. By inputting feature data into the model, it directly outputted the predicted value of the remaining lifespan. This approach required selecting and extracting appropriate features, demanding a high level of expertise in feature engineering. However, it could not provide explanations about the bearing's degradation process and failure mechanism. 2) Health indicator method (Wang et al., 2021; Ren et al., 2018; Sanakkayala et al., 2022): This method predicted the time when the bearing would reach its failure threshold by monitoring its degradation process. Subsequently, it calculated the remaining useful life. This approach required prior determination of suitable health indicators and thresholds, which relied on understanding the system's characteristics and degradation behavior, making it face certain challenges.

In Predictive Health Management of bearings, the regression problem of predicting the remaining life of a bearing is transformed into a classification problem. Rather than solely predicting the remaining life of the bearing, we are more interested in determining which stage of its lifecycle the bearing is currently in. This enables us to conduct more effective maintenance and management procedures. To achieve this goal, the entire lifespan of the bearing can be divided into different stages, and with the given bearing data,

we can predict the current stage of the bearing's life. This will help in identifying appropriate maintenance strategies and operational measures, allowing us to implement suitable preventive maintenance or repair actions at different stages.

Based on this, this chapter proposes a method for early prediction of bearing degradation. The method combines a convolutional auto-encoder segmentation network with a high-performance classifier to achieve effective differentiation and accurate classification recognition of bearing degradation processes. This provides more comprehensive and reliable support for bearing health assessment and maintenance decision-making.

6.2 Theoretical Knowledge of Bearing Degradation

Existing research on bearing fault prediction has largely focused on remaining useful life (RUL) prediction. However, in the field of PHM for bearings, we are more concerned about accurately determining the current life stage of the bearing at a given time point. Therefore, this section will concentrate on research related to the segmentation of bearing degradation stages.

In the PHM domain of bearings, accurate segmentation of the bearing's degradation stages is crucial for achieving precision maintenance and management. Unlike solely predicting RUL, accurately segmenting the degradation stages of bearings allows for the identification of appropriate maintenance strategies and actions, such as implementing preventive maintenance or corrective measures at different stages. In order to achieve precise segmentation of bearing degradation stages, researchers have proposed various methods and techniques. Among them, approaches based on image segmentation, deep

learning, and unsupervised learning have played a significant role in the research of bearing degradation stage segmentation.

Methods based on image segmentation (Liao et al., 2022; Wang, 2017) transformed vibration data into image format and utilized image segmentation algorithms such as threshold segmentation and edge detection to separate degraded regions from normal regions in the vibration images. However, this approach could only achieve segmentation for two stages and might not have provided accurate differentiation for complex bearing degradation processes, such as early stages of degradation. Deep learning-based methods, using CNN (Wang, et al., 2021; Zhou et al., 2019), and RULNet (Gamanayake et al., 2023), performed feature learning and segmentation on vibration signals. These approaches could automatically learn feature representations and complex relationships, leading to precise bearing degradation segmentation. Nonetheless, deep learning methods typically required a large amount of labeled data for supervised learning, which could have posed challenges in engineering applications. Manually annotating a large-scale vibration dataset was a time-consuming and labor-intensive task, demanding domain expertise and experience to correctly label samples of different degradation stages.

In the past, segmentation methods based on unsupervised learning allowed for the segmentation of bearing degradation stages without the need for labeled data. For instance, methods like Generative Adversarial Networks (Mylonas & Chatzi, 2020) and Variational Auto-encoders (Hong et al., 2014) could learn the latent representation and distribution of the data, enabling segmentation of bearing degradation and dividing the degraded areas into different categories. In unsupervised learning approaches, the key lies in modeling bearing degradation. Currently, based on the nature of the degradation process, bearing

degradation modeling methods can be categorized into continuous degradation and discrete degradation stage models.

The continuous degradation model aimed to establish a single model to describe the gradual deterioration process of bearing life. This type of model is typically based on methods such as physical principles, statistics, or machine learning. The continuous degradation model considers the progressive changes in bearing life and can capture the trends of bearing degradation based on various indicators (e.g., vibration features, temperature, load, etc.). Common continuous degradation models include physical models (Cui & Su, 2021; Li et al., 2023), Markov process models (Gu et al., 2023; Kou et al., 2022), regression models (Chen et al., 2022), etc. By estimating and predicting the parameters of the continuous degradation model, it is possible to forecast and assess the remaining life of the bearings. The establishment and prediction of the continuous degradation model were based on existing bearing degradation data and assumed that the degradation process is continuous and gradual. However, the actual bearing degradation process may be influenced by multiple factors, such as working conditions, load variations, environmental factors, etc. These factors may lead to certain discrepancies between the model and the real situation.

In the discrete degradation stage model, the bearing's lifespan was divided into multiple discrete stages, and based on these stages, the bearing's condition was modeled and predicted. This model could better adapt to the non-linearity and uncertainty of degradation processes in practical situations. Common discrete degradation stage models included Hidden Markov Models (Aggab et al., 2022), time-series models (Cao et al., 2021; Zhu et al., 2023), etc. These models could better describe and identify different stages in

the bearing degradation process, thereby assisting in determining appropriate maintenance strategies and operational measures.

The bearing discrete degradation stage modeling proposed in Juodelyte et al. (2022) transforms the life prediction into a classification problem. However, the model mentioned above failed to identify effective features, resulting in lower accuracy in predicting early-stage bearing degradation. In contrast, our proposed approach introduces a convolutional auto-encoder and combines it with a domain knowledge-based automatic data labeling strategy and a classifier based on a multi-scale attention network architecture. This model is well-suited for predicting bearing degradation throughout its entire lifespan.

6.3 Network Architecture

This chapter utilizes time-domain vibration data collected during the operation of bearings for bearing life stage prediction. The problem is divided into two stages. Assuming the input is $X \in R_k$, $t = (0, 1, \dots, T)$, where T is the time step. X contains the complete life cycle data of the bearings. The data transformed into the frequency domain is represented as $F(X) \in R_m$, where m is the dimension of frequency domain features. The model maps the frequency domain features to bearing life stage labels $L \in \{0, 1, 2, 3\}$ to form life labels, reflecting different degradation states of the bearings. Stage I can be represented as follows:

$$L = M(F(X)) \quad \text{Equation 6.1}$$

M represents the model function for Stage I. M represents FAEK model function.

Assuming the input data is $X \in R_k$, and $L \in \{0, 1, 2, 3\}$, representing different life stage labels of the bearings. The model takes input data X , which includes frequency

domain and time domain features of the bearings, along with the life stage labels L . The output is the predicted life stage label of the bearings, denoted as $P \in \{0, 1, 2, 3\}$. Stage II can be represented as follows:

$$P = G(X, L) \quad \text{Equation 6.2}$$

G represents the classifier for Stage II. G represents MCC model function.

By using the function M in stage I to obtain the degradation label of the bearing, the bearing data and label are input into the function G in stage II to predict the degradation stage of the bearing. This achieves the classification and prediction of the degradation stage of the bearing vibration data, determines the bearing life stage at a given time point, and provides corresponding maintenance strategies. Figure 6.1 illustrates the workflow of our model. Based on the description of the above problem, our model is divided into two parts. Below, we will discuss each sub-model in detail.

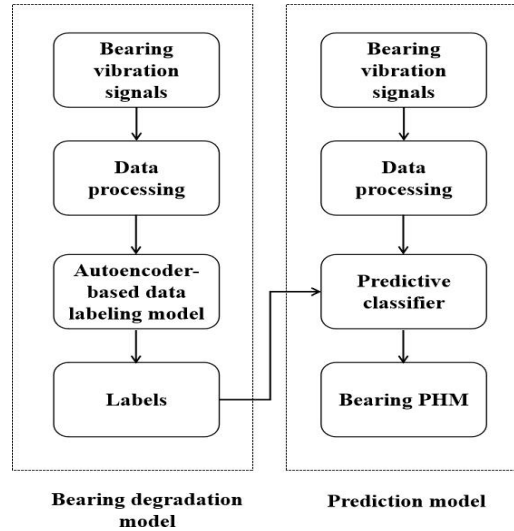


Figure 6.1: Bearing Life Stage Workflow

When bearings undergo degradation, we can categorize their degradation process into four stages by observing their physical manifestations and features in the frequency and time domains. This division of degradation stages aids in better understanding the trend of bearing life expectancy and state evolution. Different stages correspond to different degradation characteristics, thus requiring corresponding maintenance strategies to prolong the bearing's lifespan and ensure equipment's reliable operation. Table 6.1 summarizes the characteristics and maintenance strategies for the four stages.

Table 6.1: Bearing degradation stage classification

Stage	Physical Characteristics	Time Domain Characteristics	Frequency Domain Characteristics	Maintenance Strategy
Health	In normal operating condition	Manifested as stable periodic oscillations	Display the normal operating frequency components	Monitor and record the baseline condition and performance parameters of bearings
Early degeneration	Minor signs of degradation: scratches, tiny cracks, or slight wear	Minor non-periodic variations, slight changes in amplitude	High-frequency components related to early-stage faults appear	Monitor spectrum data, Perform regular vibration analysis and diagnosis, Timely lubricate and perform cleaning
Severe degeneration	Degradation has become more significant and severe	Non-periodic fluctuations, more amplitude variations and waveform distortion	More high-frequency components and wide-band noise	Vibration analysis and fault diagnosis, maintenance and replacement measures, Inspect lubrication conditions and the lubricants

Table 6.1: continued

Failure	Severe Fault State	Significant mutations, shocks, irregular fluctuations	More high-frequency components and wide-band noise, possibly causing impact frequencies	Inspection and replacement of bearings
---------	--------------------	---	---	--

6.3.1 Bearing Degradation Model

The bearing degradation model we propose is called FAEK, which consists of three components: data preprocessing, convolutional auto-encoder, and clustering operation. The FAEK network architecture is illustrated in Figure 6. 2.

Firstly, the bearing vibration data undergoes a Fast Fourier Transform (FFT) to determine the dominant frequency components and energy distribution of the vibration signal, thereby extracting features related to bearing degradation. These features can be used for the division of degradation stages. Assuming the bearing vibration data is represented as a time series $x(t)$, where t represents time, the frequency domain transformation can be expressed as:

$$X(f) = \text{ABS}(\text{FFT}(x(t))) \quad \text{Equation 6.3}$$

$X(f)$ represents the frequency domain representation of vibration data, and f denotes frequency. FFT stands for Fast Fourier Transform, ABS indicates taking the absolute value.

The frequency domain signals of the bearings are fed into the convolutional auto-encoder to learn the data's feature representation through unsupervised learning. The

convolutional auto-encoder consists of an encoder and a decoder. The encoder compresses the input data into a latent representation, while the decoder decodes the latent representation back into a reconstructed input data. In contrast to traditional fully connected auto-encoders, convolutional auto-encoders utilize convolutional operations in the encoding and decoding process, making them effective in handling the time-series data of bearings.

The encoder part takes the input data x and processes it through convolution and activation operations. It gradually reduces the size of the feature maps through max-pooling operations, extracting and compressing the feature information from the input data. The results of the convolutional layers and pooling layers can be represented as h and p :

$$h = \text{ReLU}(\text{Conv1D}(W, b)(x)) \quad \text{Equation 6.4}$$

$$p = \text{MaxPooling1D}(h) \quad \text{Equation 6.5}$$

Conv1D represents a one-dimensional convolution operation, and W and b are the weights and bias parameters corresponding to the convolutional layer; ReLU represents the activation function; MaxPooling1D represents a one-dimensional max-pooling operation.

The decoder part gradually restores the size of the feature maps through up-sampling operations, followed by convolution and activation operations, to obtain the reconstructed output. The up-sampling can be represented as u :

$$u = \text{UpSampling1D}(h) \quad \text{Equation 6.6}$$

UpSampling1D represents the one-dimensional up-sampling operation.

Through the use of a convolutional auto-encoder, we can capture meaningful features from the vibration data of bearings and transform them into a meaningful representation. The convolutional auto-encoder is trained in an unsupervised learning manner, eliminating the need for manual labeling of bearing data's life stages and avoiding reliance on handcrafted feature engineering. The design of the convolutional auto-encoder enables it to perform exceptionally well in processing time-series data of bearing vibrations, allowing it to extract temporal correlations and local features from the data.

Finally, by mapping the learned features into a clustering space, we can effectively group the bearing data into different degradation stages and generate corresponding labels. This provides essential guidance for subsequent maintenance decisions. This data-driven approach exhibits good generalization performance and can adapt to various types and scales of bearing data, making it a viable solution for practical engineering applications.

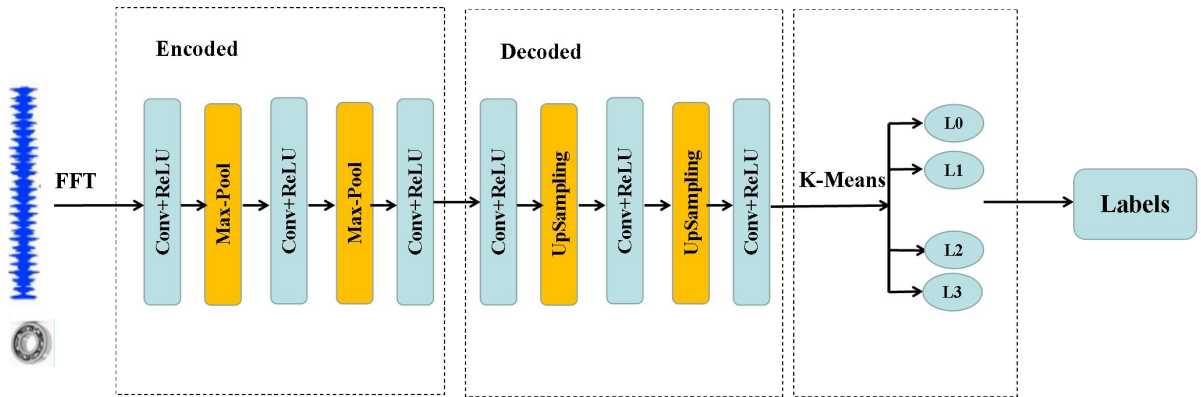


Figure 6. 2: FAEK Network Architecture

6.3.2 Prediction Model

The early signs of bearing degradation are only visible in the frequency domain (Qiu et al., 2023). In order to capture both the frequency domain features associated with

the early degradation signs and the time domain features that intensify with the progression of degradation, we propose a multi-input classification network architecture, which we refer to as the Multi-scale, Channel Attention Classification model (MCC). Figure 6.3 illustrates the schematic diagram of this architecture.

In this architecture, we simultaneously input time-domain signals, frequency-domain signals, and degradation labels. The time-domain signals provide more detailed temporal information, while the frequency-domain signals can better capture frequency characteristics. By combining these two inputs, we can fully leverage their complementarity. The network consists of three layers: Multi-scale processing layer, Channel attention layer, and Classification layer. We will discuss each substructure in detail.

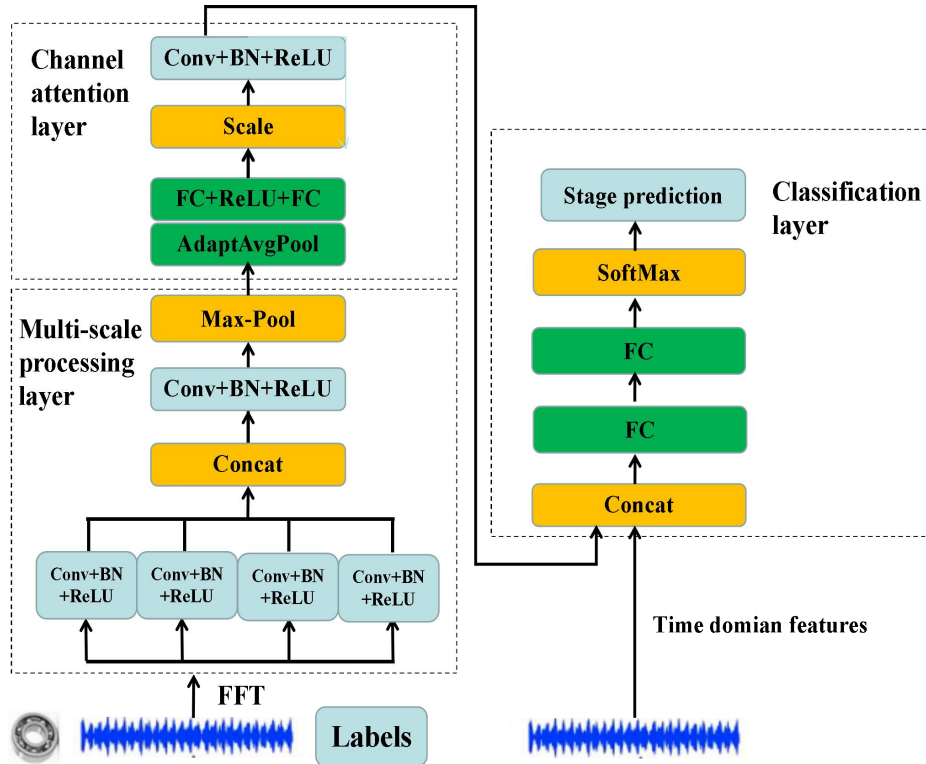


Figure 6.3: MCC architecture

Multi-scale processing layer: It consists of a multi-branch parallel convolution, utilizing four parallel convolutional branches to extract different frequency domain features. Each branch extracts a set of feature maps, capturing distinctive aspects of the input data. These feature maps are then fused to obtain a more comprehensive representation of the features. This approach provides a richer and more diverse feature representation, enabling the network to adapt to various bearing degradation patterns and enhance its ability to extract complex features.

Assuming the input data is denoted as X , the number of convolution kernels is N , the size of each convolution kernel is $K \times K$, the number of input data channels is C , and the number of output feature maps is M . For each convolution kernel i ($1 \leq i \leq N$), its weight matrix is represented as W_n with the shape $K \times K \times C$, and its bias term is b_n . The output of the multi-path convolution operation for the m -th feature map of the input data X can be calculated using the following formula:

$$Y_m = \sum^n f(\text{conv}(W_n, X) + b_n) \quad \text{Equation 6.7}$$

Here, $\text{conv}(W_n, X)$ denotes the convolution operation of the input data X with the convolution kernel W_n . b_n denotes the bias term of the n th convolution kernel. f denotes the ReLU function.

During the bearing degradation process, the frequency spectrum characteristics of the vibration signal undergo changes. Traditional single convolutional kernels often struggle to capture subtle variations across different frequency bands. In contrast, multiple parallel convolutions, known as multi-branch parallel convolution, employ several parallel convolutional kernels, each corresponding to a different frequency band. This enables

better extraction of signal features in the frequency domain. Moreover, multi-branch parallel convolution increases the depth and width of the network, providing more parameters and nonlinear expressive power. As a result, the model can better adapt to complex vibration signal features, thereby improving its performance in bearing degradation research.

Channel attention layer: The channel attention network was introduced in Chapter 4, where, in the task of bearing degradation prediction, it dynamically enhances important features by learning the weights for each channel. For vibration frequency domain signals of bearings, it can extract the most representative and discriminative frequency domain features, aiding in accurately analyzing the bearing degradation. Moreover, the channel attention network significantly boosts network performance without introducing additional parameters. This is particularly crucial for large-scale data processing in bearing degradation studies, as bearing vibration signals are typically high-dimensional time series data. The network can improve feature extraction efficiency and accuracy without increasing network complexity.

Classification layer: In bearing fault diagnosis, time-domain features are crucial for capturing the condition of faults. Slope and kurtosis are commonly used time-domain features that can reflect the symmetry and impulse fluctuation characteristics of signals. These features are indispensable when considering degradation prediction of bearings. When faults exist in bearings, the pulse components in the vibration signal typically increase, hence these features are often employed in bearing fault diagnosis. Additionally, there are dimensionless parameters that can reduce the influence of operating conditions and other factors on the model. Common dimensionless parameters include clearance

indicators, waveform indicators, pulse indicators, peak indicators, and kurtosis indicators. Clearance indicators are typically used to indicate wear conditions, while peak indicators, pulse indicators, and kurtosis indicators are commonly used to describe the impact characteristics of signals, aiding in better depicting early-stage degradation situations. Table 3.1 lists the 13 time-domain features we extracted (Zhu et al., 2014).

To more accurately identify the degradation of bearings, after the feature extraction through the above two-layer structure, the feature dimension of the vibration signal is reduced. Frequency-domain and time-domain features are merged, and then degradation stage prediction is performed through linear layers and SoftMax. This fusion approach helps in leveraging the information from both types of features to enhance the accuracy of diagnosis and prediction.

6.4 Experimental Results

In this chapter, we utilize a publicly available bearing dataset to validate the performance of our model. We evaluate the effectiveness of automatic labeling and compare the partitioning results for different bearings under various operating conditions. Additionally, we compare the accuracy of predicting bearing degradation stages. Our aim is to demonstrate the efficacy of our model in predicting bearing degradation, helping us understand its generalizability across different bearings and operating conditions. We use the FEMTO-ST bearing dataset (Zhang et al., 2018) for training and evaluation, as introduced in Chapter Three. The dataset contains 17 bearing data under three different operating conditions, with 6 bearings in the training set and 11 bearings in the testing set. As shown in Table 6.2.

Table 6.2: Distribution of FEMTO-ST bearing dataset

	Operating Conditions		
Data sets	Conditions 1	Conditions 1	Conditions 1
Learning set	Bearing1_1	Bearing2_1	Bearing3_1
	Bearing1_2	Bearing2_2	Bearing3_2
Test set	Bearing1_3	Bearing2_3	Bearing3_3
	Bearing1_4	Bearing2_4	
	Bearing1_5	Bearing2_5	
	Bearing1_6	Bearing2_6	
	Bearing1_7	Bearing2_7	

During experiments, the bearing's horizontal and vertical vibration signals are sampled at a frequency of 25,600 Hz every 10 seconds, with each sample lasting 0.1 seconds, resulting in 2560X2 data points per sample.

We begin by preprocessing the raw data. Since the FEMTO-ST dataset contains entire lifecycle data with vibration signals in both horizontal and vertical directions, we separate and process these signals individually. We then employ FFT to transform them into frequency domain signals to extract frequency domain features. Simultaneously, we extract time-domain features for each direction's vibration signals based on the list of time-domain features in Table 3.4 to describe their statistical and impulsive characteristics. Subsequently, we merge the features from horizontal and vertical directions to obtain more comprehensive bearing features. Moreover, since each bearing operates under different conditions leading to distinct degradation patterns, we perform all the aforementioned

operations separately for each bearing to capture its unique vibration features and degradation patterns.

We utilize cross-entropy as the loss function and the Adam optimizer for network optimization. Gradient clipping technique is applied during iteration to prevent gradient explosion. The batch size for all networks is set to 64, and the learning rate is 0.01 (Zhang et al., 2018). All experiments are conducted on a Windows 10 operating system with an Intel i5-12490F CPU and an NVIDIA RTX 3060 12GB GPU. Our code is written in Python 3.8 and PyTorch.

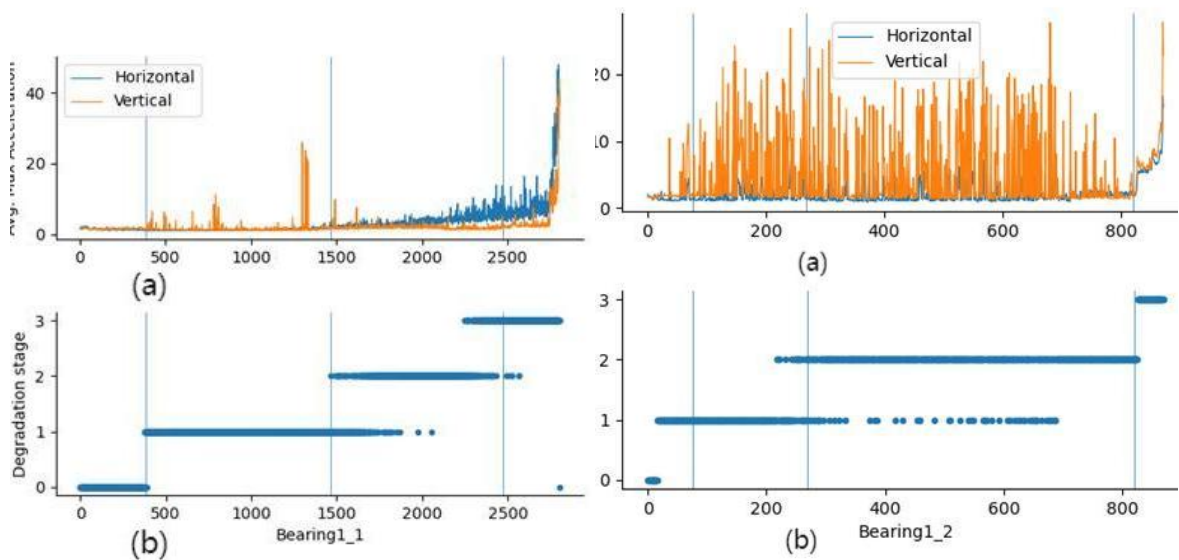
To evaluate the effectiveness of the model's automatic label partitioning, we compare the automatically generated phase labels with manually partitioned phases (for the training set) using the method proposed by (Sutrisno et al., 2012). A smaller difference indicates better accuracy of the model's partitioning. Additionally, to assess the performance of the bearing phase prediction classifier, we compare the predicted phases with the phases automatically labeled by our model (for the testing set) to evaluate prediction accuracy. For PHM in bearings, predictions at different time points should follow the real degradation sequence. Therefore, we also focus on the overlap between predicted phases and actual phases when predicting in chronological order to assess the model's reliability.

6.4.1 Data Labeling

We automatically assigned degradation stages to the 6 bearings in the training set and compared the results with manually labeled stages. As shown in Figure 6.4, the model's automatically assigned bearing degradation stages are in good agreement with the manually labeled ones, indicating that the model performs well in identifying bearing

degradation stages. Although there is some overlap between adjacent stages, this is reasonable due to the continuity of the degradation process and variations in the signals, which may lead to some degree of overlap between neighboring degradation stages. The model demonstrates excellent performance in distinguishing between healthy and faulty stages, which is crucial for early maintenance and timely identification of failed bearings. The FEMTO-ST dataset was collected under extreme operating conditions, resulting in relatively short healthy stages for all bearings, which aligns with real-world scenarios.

However, for bearing1_2, the model's performance is relatively poor. This is due to the excessive amplitude of the vertical direction's smoothest maximum acceleration, leading to strong interference in the signal. Despite this, overall, the model exhibits good accuracy and robustness in automatically assigning bearing degradation stages.



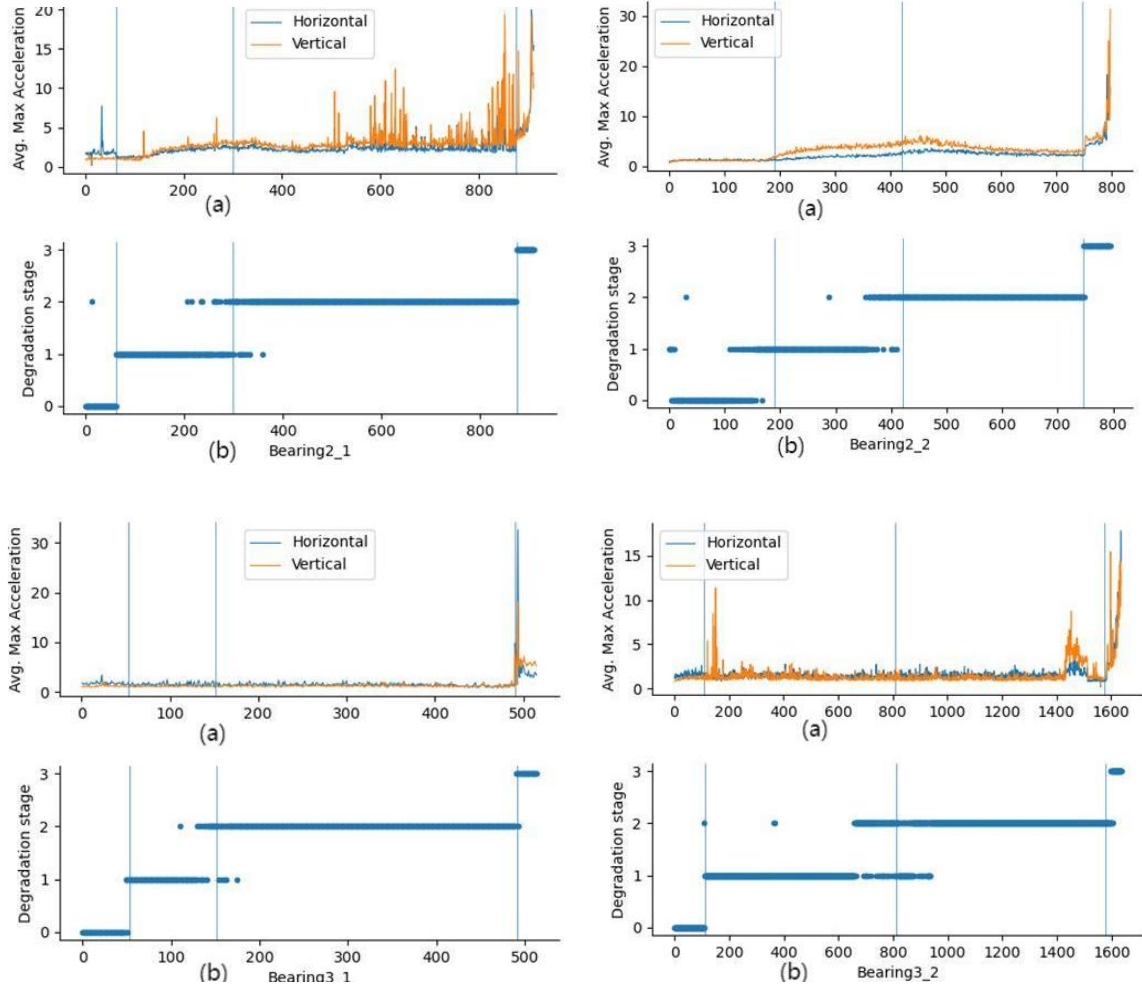


Figure 6.4: Training set degradation labels:(a): Smoothed Maximum Acceleration obtained by averaging the five highest absolute acceleration measurements in the time domain (Sutrisno et al., 2012).(b): Automatically assigned labels obtained through training the auto-encoder. The vertical lines represent the manually assigned results.

6.4.2 Classifier

After automatically labeling the training set for the 6 bearings using the FAEK model, we compared the performance of the AE model (Juodelyte et al., 2022), 1DCNN model (Wang et al., 2021), and our proposed MCC model in predicting bearing degradation stages, as shown in Figure 6.5. The results indicate that for the task of bearing degradation classification, the MCC model outperforms the other two models with remarkable classification accuracy, reaching up to approximately 0.9665. The superiority

of the MCC model can be attributed to its innovative model design and feature extraction techniques. The model adopts multi-path convolution and attention mechanisms, effectively extracting meaningful features. Additionally, we simultaneously train the model on both the horizontal and vertical data of the bearings and ultimately fuse frequency-domain and time-domain features.

The AE model and 1DCNN model exhibit similar performance, with maximum accuracies of 0.8471 and 0.8763, respectively. Notably, our proposed MCC model achieves accuracy above 0.9 after just 30 epochs, indicating its fast convergence speed. This efficiency makes the MCC model more effective in solving the bearing degradation classification task, providing timely and accurate solutions for bearing fault diagnosis and prediction.

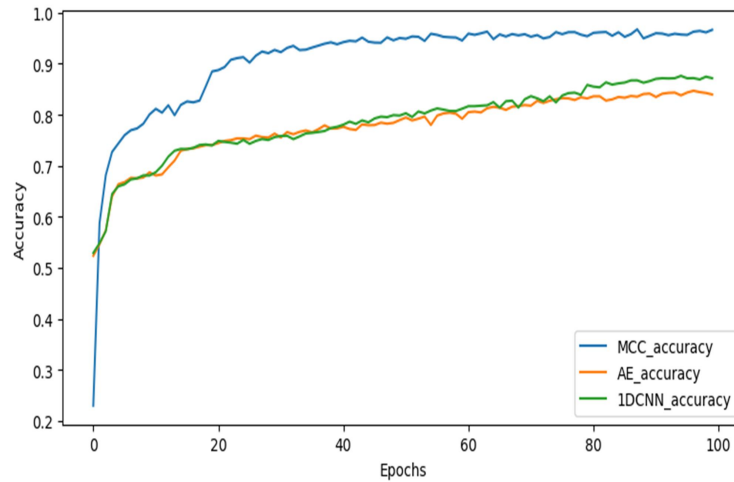


Figure 6.5: Accuracy of Different Classification Models

To further validate the accuracy of our model in predicting bearing degradation stages, we compared the results of the MCC model and the 1DCNN model on the test set, as shown in Figure 6.6. Figure 6.6 (a) presents the prediction accuracy of the 1DCNN model. The box plot reveals that the 1DCNN model has lower classification accuracy in

stages 1 and 3, with the median line of the boxes relatively low and the boxes narrow, indicating significant dispersion of data in these two stages. Particularly, the performance is poor in bearing degradation stage 1. Figure 6.6 (b) displays the results of our MCC model, indicating high classification accuracy in all stages. The median line of the boxes is higher, and the boxes are flatter, suggesting a higher concentration of data in each stage. Therefore, the MCC model demonstrates more stable performance and superior results.

Figure 6.7 provides statistics on the overlapping rates between adjacent stages of all 11 bearings in the test set. It can be observed that our model's predicted results have a low overlapping rate, indicating the MCC model's ability to distinguish different degradation stages and better identify the distinctive features of each stage in the bearing degradation classification task.

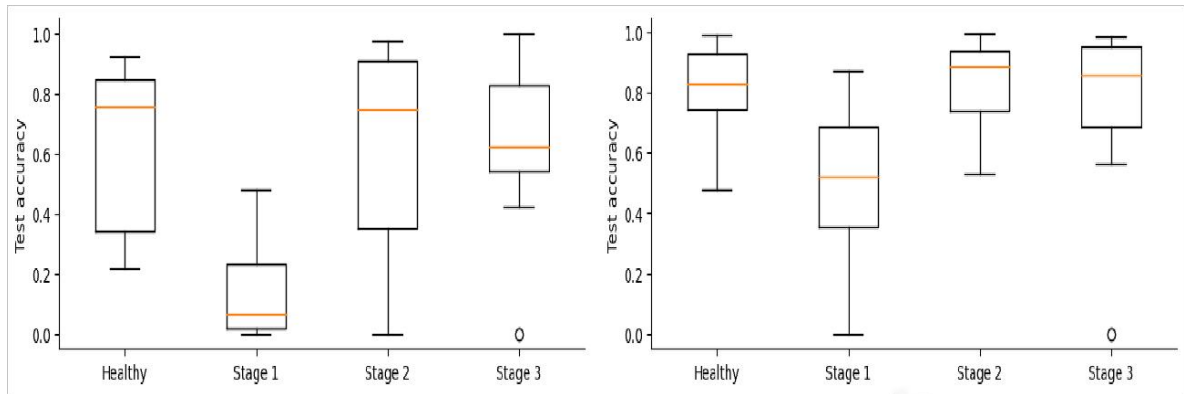


Figure 6.6: Test Set Accuracy of the Models (based on FAEK labels).(a): 1DCNN model. (b): MCC model

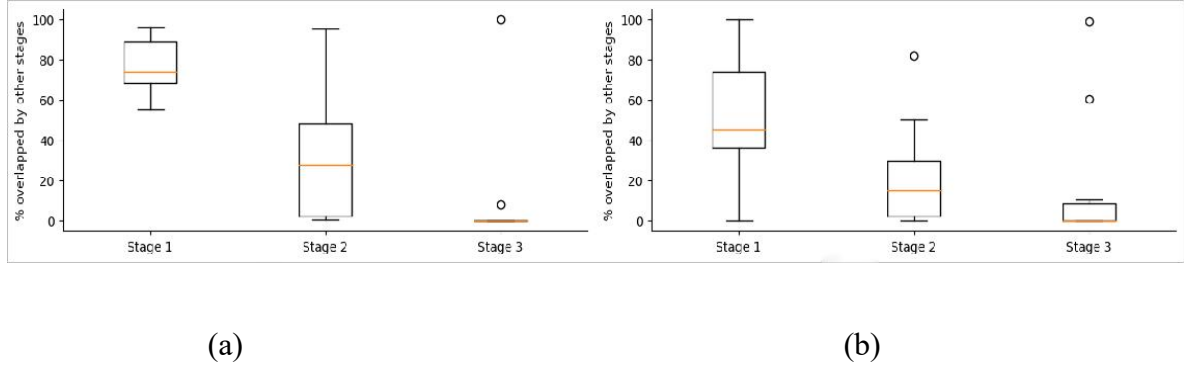


Figure 6.7: Overlap Rate of Adjacent Stages for the Models.(a): 1DCNN Model. (b): MCC Model

We used a smoothing classifier to present the prediction results, aiming to reduce prediction fluctuations at individual time points and obtain more stable predictions. Specifically, for each time point in the test set, we generated posterior probabilities for five predictions and calculated the average of these five predictions within a fixed sliding window, resulting in the smoothed classifier predictions.

In the test set containing 11 bearings, we showcased the prediction results for three different operating conditions, as shown in Figure 6.8. From the graphs, it can be observed that over time, the predicted bearing degradation stages align well with the actual degradation progress of the bearings. This indicates that our predictions are not only highly accurate but also consistent with the actual degradation processes of the bearings.

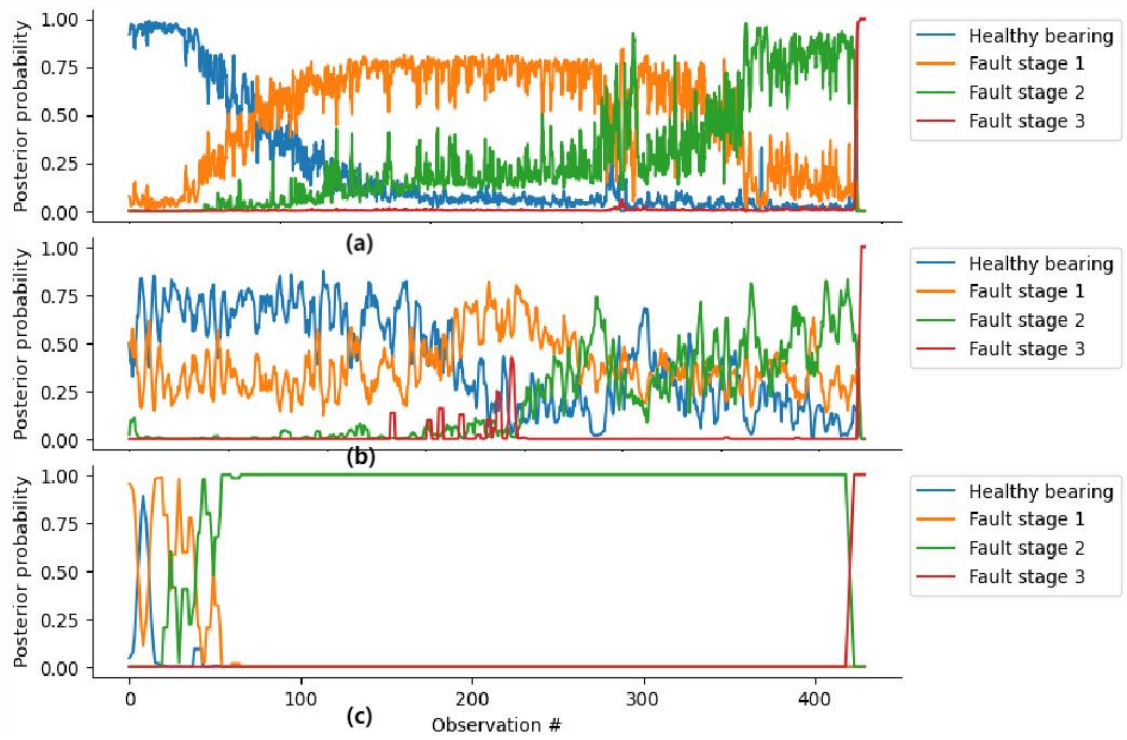


Figure 6.8: The prediction results for bearings 1_5 (a), 2_4 (b), and 3_3 (c)

6.5 Chapter Summary

This chapter focuses on the predictive maintenance of bearings, aiming to classify bearings into four degradation stages to predict at a given time point which degradation stage the bearing is in and provide corresponding maintenance strategies for different degradation stages. To achieve this, we propose a method based on FAEK automatic labeling and MCC classification prediction model. The model features: 1) Automatic division of degradation labels: We adopt a method based on convolutional autoencoder models to divide the vibration signals of bearings into stages. The FAEK model's automatic division effectively distinguishes different degradation stages of bearings, avoiding the tediousness and subjectivity of manual label division. 2) Prediction model: We input the time-domain and frequency-domain features of bearing signals into the MCC classification model, which includes multiple convolution modules and channel attention

modules, helping to better extract signal features and achieve precise prediction of bearing degradation stages. With such a prediction model, we can accurately predict the degradation status of bearings and take corresponding maintenance measures in a timely manner. We validate this network architecture using the FEMTO-ST dataset, and the results show excellent performance of this architecture. The automatically divided labels highly overlap with manually divided labels, successfully distinguishing different degradation stages. In the degradation stage prediction task, our model architecture achieves an accuracy of 0.9665.

However, we also recognize that this bearing dataset is relatively clean, while in actual industrial applications, bearing data may be subject to various interferences and noises. Additionally, the accuracy of automatically divided labels needs further discussion and improvement. Our subsequent research should focus on exploring better segmentation methods to improve label quality. Only by improving label quality can our classification predictions be more credible and reliable, providing more effective support for predictive maintenance of bearings.

CHAPTER 7

CONCLUSION AND RECOMMENDATIONS

7.1 Conclusion

In this study, we thoroughly investigate the issue of bearing fault diagnosis and prediction, proposing a series of solutions leveraging deep learning techniques.

For objective 1, we introduces a convolutional neural network named MCFCNN for achieving end-to-end bearing fault diagnosis. The network demonstrates good classification performance on the CWRU bearing dataset, maintaining over 80% classification accuracy even under strong noise interference. Our model incorporates secondary neuron convolutional operations and introduces multiscale processing layers and channel attention layers to enhance diagnostic performance in noisy environments. Classification tests are conducted to validate the significance of multiscale processing layers and channel attention layers in extracting bearing fault features. Additionally, visualization techniques are employed to interpret the feature maps output by the model, providing better insights into the model's workings and performance.

For objective 2, we focuses on introducing the BFTS test platform and the dataset collected on it, aimed at validating the adaptability of MCFCNN on real bearing fault data and examining its performance under different environmental conditions. The composition of the BFTS bearing fault simulation platform is detailed, including mechanical systems, acceleration sensors, and data acquisition and control units. The platform can simulate bearing operation under different conditions and provide diverse input data, laying the foundation for a comprehensive evaluation of the MCFCNN network. Experimental

designs on the BFTS dataset, including the production of faulty bearings, data collection methods, and signal-to-noise ratio settings, are described. By comparing experimental results, the performance advantages of the MCFCNN model under high-noise conditions, as well as the effects of different network structures and data augmentation strategies on the final results, are validated. Lastly, model transfer experiments are conducted to verify the stability and generalization ability of the MCFCNN model under different operating conditions.

For objective 3, we focus on addressing the issue of predictive maintenance for bearings, aiming to divide the lifespan of bearings into four degradation stages to accurately predict the degradation status of bearings at specific time points and provide corresponding maintenance strategies. To achieve this goal, we propose a method based on FAEK automatic labeling and MCC classification prediction model. This model has two main features: firstly, we employ the FAEK method based on convolutional autoencoder models to automatically segment the full-life vibration signals of bearings, effectively distinguishing different degradation stages of bearings and avoiding the complexity and subjectivity of manual label segmentation. Secondly, we input the time-domain and frequency-domain features of bearing signals into the MCC classification model, which includes multiple convolutional modules and channel attention modules, to better extract signal features and achieve accurate prediction of bearing degradation stages. With this predictive model, we can accurately predict the degradation status of bearings and take corresponding maintenance measures in a timely manner. We validate this model using the FEMTO-ST dataset, demonstrating excellent performance. The automatically segmented labels highly coincide with manually segmented labels, successfully distinguishing

different degradation stages. In the degradation stage prediction task, our model achieves an accuracy of 0.9665.

7.2 Recommendations

This thesis conducts a series of studies on end-to-end bearing fault diagnosis tasks in various scenarios using deep neural networks. Although certain achievements have been made, there are several aspects worthy of further investigation:

In this thesis, Gaussian white noise is used to simulate noise interference for studying bearing fault diagnosis under noise disturbance. However, the types of noise in real environments are complex and diverse, with noise distributions possibly being not only additive but also correlated with signals. Therefore, it is necessary in subsequent research to study different types of noise, noise models, and bearing fault diagnosis tasks under the assumption that noise and bearing vibration signals are multiplied together.

When studying bearing fault diagnosis under different operating conditions, the thesis assumes that the source domain and target domain tasks are identical. However, in reality, the operating conditions of the source and target domains may not completely match. For example, data may only be available for certain fault conditions in the target domain, or the severity classification of faults in the target domain may differ from that in the source domain. In such cases, the assumption of domain adaptation is not met, thus necessitating the consideration of bearing fault diagnosis problems under these circumstances in future research.

The thesis only utilizes information from a single sensor in the research. In future experiments, the fusion of information from multiple sensors will be considered. For

instance, the fusion of acceleration sensor information from different locations or the fusion of information from different types of sensors such as sound, current, torque, etc. It is expected that the fusion of multimodal sensor information can improve the performance of bearing fault diagnosis in complex situations, thus warranting further investigation.

In the bearing life prediction aspect of the thesis, the model is only validated using publicly available datasets. Due to the difficulty and cost of collecting full lifecycle data for bearings, we did not collect our own bearing data due to financial constraints. Subsequent research can involve validating the model using privately collected datasets. Additionally, further discussion and improvement are needed regarding the accuracy of automatically segmented labels. Our future research should focus on exploring better segmentation methods to improve the quality of labels.

REFERENCES

- Agarwal, G., Om, H., & Gupta, S. (2022). A learning framework of modified deep recurrent neural network for classification and recognition of voice mood. *International Journal of Adaptive Control and Signal Processing*, 36(8), 1835–1859. <https://doi.org/10.1002/acs.3425>
- Aggab, T., Vrignat, P., Avila, M., & Kratz, F. (2022). Remaining useful life estimation based on the joint use of an observer and a hidden Markov model. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 236(5), 676–695. <https://doi.org/10.1177/1748006X211044343>
- An, Z., Li, S., Wang, J., & Jiang, X. (2020). A novel bearing intelligent fault diagnosis framework under time-varying working conditions using recurrent neural network. *ISA Transactions*, 100, 155–170. <https://doi.org/10.1016/j.isatra.2019.11.010>
- Bahdanau, D., Cho, K., & Bengio, Y. (2016). *Neural Machine Translation by Jointly Learning to Align and Translate* (arXiv:1409.0473). arXiv. <https://doi.org/10.48550/arXiv.1409.0473>
- Bayer, F. M., Cintra, R. J., & Cribari-Neto, F. (2018). Beta seasonal autoregressive moving average models. *Journal of Statistical Computation and Simulation*, 88(15), 2961–2981. <https://doi.org/10.1080/00949655.2018.1491974>
- Borghesani, P., Pennacchi, P., Randall, R. B., Sawalhi, N., & Ricci, R. (2013). Application of cepstrum pre-whitening for the diagnosis of bearing faults under variable speed conditions. *Mechanical Systems and Signal Processing*, 36(2), 370–384. <https://doi.org/10.1016/j.ymssp.2012.11.001>

- Bouras, A., Bennedjai, S., & Bouras, S. (2020). Experimental detection of defects in variable speed fan bearing using stator current monitoring. *SN Applied Sciences*, 2(5), 891. <https://doi.org/10.1007/s42452-020-2687-2>
- Cao, Y., Jia, M., Ding, P., & Ding, Y. (2021). Transfer learning for remaining useful life prediction of multi-conditions bearings based on bidirectional-GRU network. *Measurement*, 178, 109287. <https://doi.org/10.1016/j.measurement.2021.109287>
- Cerrada, M., Sánchez, R.-V., Li, C., Pacheco, F., Cabrera, D., Valente de Oliveira, J., & Vásquez, R. E. (2018). A review on data-driven fault severity assessment in rolling bearings. *Mechanical Systems and Signal Processing*, 99, 169–196. <https://doi.org/10.1016/j.ymssp.2017.06.012>
- Che, C., Wang, H., Ni, X., & Fu, Q. (2020). Domain Adaptive Deep Belief Network for Rolling Bearing Fault Diagnosis. *Computers & Industrial Engineering*. <https://doi.org/10.1016/j.cie.2020.106427>
- Chen, D., Qin, Y., Wang, Y., & Zhou, J. (2021). Health indicator construction by quadratic function-based deep convolutional auto-encoder and its application into bearing RUL prediction. *ISA Transactions*, 114, 44–56. <https://doi.org/10.1016/j.isatra.2020.12.052>
- Chen, H.-Y., & Lee, C.-H. (2020). Vibration Signals Analysis by Explainable Artificial Intelligence (XAI) Approach: Application on Bearing Faults Diagnosis. *IEEE Access*, 8, 134246–134256. <https://doi.org/10.1109/ACCESS.2020.3006491>

- Chen, Q., Dai, S., & Dai, H. (2019). A Rolling Bearing Fault Diagnosis Method Based on EMD and Quantile Permutation Entropy. *Mathematical Problems in Engineering*, 2019, e3089417. <https://doi.org/10.1155/2019/3089417>
- Chen, Y., Duan, W., Ding, Z., & Li, Y. (2022). Battery Life Prediction Based on a Hybrid Support Vector Regression Model. *Frontiers in Energy Research*, 10. <https://www.frontiersin.org/articles/10.3389/fenrg.2022.899804>
- Chen, Z., Li, C., & Sanchez, R.-V. (2015). Gearbox Fault Identification and Classification with Convolutional Neural Networks. *Shock and Vibration*, 2015, e390134. <https://doi.org/10.1155/2015/390134>
- Cheng, Y., Yuan, H., Liu, H., & Lu, C. (2017). Fault diagnosis for rolling bearing based on SIFT-KPCA and SVM. *Engineering Computations*, 34(1), 53–65. <https://doi.org/10.1108/EC-01-2016-0005>
- Chui, K. T., Gupta, B. B., & Vasant, P. (2021). A Genetic Algorithm Optimized RNN-LSTM Model for Remaining Useful Life Prediction of Turbofan Engine. *Electronics*, 10(3), Article 3. <https://doi.org/10.3390/electronics10030285>
- Clevert, D.-A., Unterthiner, T., & Hochreiter, S. (2016). *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)* (arXiv:1511.07289). arXiv. <https://doi.org/10.48550/arXiv.1511.07289>
- Cocconcelli, M., Zimroz, R., Rubini, R., & Bartelmus, W. (2012). STFT Based Approach for Ball Bearing Fault Detection in a Varying Speed Motor. In T. Fakhfakh, W. Bartelmus, F. Chaari, R. Zimroz, & M. Haddar (Eds.), *Condition Monitoring of Machinery in Non-Stationary Operations* (pp. 41–50). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-28768-8_5

- Cui, L., & Su, Y. (2021). Contact fatigue life prediction of rolling bearing considering machined surface integrity. *Industrial Lubrication and Tribology*, 74(1), 73–80. <https://doi.org/10.1108/ILT-08-2021-0345>
- Dai, W., Mo, Z., Luo, C., Jiang, J., Zhang, H., & Miao, Q. (2020). Fault Diagnosis of Rotating Machinery Based on Deep Reinforcement Learning and Reciprocal of Smoothness Index. *IEEE Sensors Journal*, 20(15), 8307–8315. IEEE Sensors Journal. <https://doi.org/10.1109/JSEN.2020.2970747>
- Ding, Y., Ding, P., & Jia, M. (2021). A Novel Remaining Useful Life Prediction Method of Rolling Bearings Based on Deep Transfer Auto-Encoder. *IEEE Transactions on Instrumentation and Measurement*, 70, 1–12. IEEE Transactions on Instrumentation and Measurement. <https://doi.org/10.1109/TIM.2021.3072670>
- Dragomiretskiy, K., & Zosso, D. (2014). Variational Mode Decomposition. *IEEE Transactions on Signal Processing*, 62(3), 531–544. IEEE Transactions on Signal Processing. <https://doi.org/10.1109/TSP.2013.2288675>
- Duan, Y., Wang, C., Chen, Y., & Liu, P. (2019). Improving the Accuracy of Fault Frequency by Means of Local Mean Decomposition and Ratio Correction Method for Rolling Bearing Failure. *Applied Sciences*, 9(9), Article 9. <https://doi.org/10.3390/app9091888>
- Elbouchikhi, E., Choqueuse, V., Amirat, Y., Benbouzid, M. E. H., & Turri, S. (2017). An Efficient Hilbert–Huang Transform-Based Bearing Faults Detection in Induction Machines. *IEEE Transactions on Energy Conversion*, 32(2), 401–413. IEEE Transactions on Energy Conversion. <https://doi.org/10.1109/TEC.2017.2661541>

- El-Thalji, I., & Jantunen, E. (2015). A summary of fault modelling and predictive health monitoring of rolling element bearings. *Mechanical Systems and Signal Processing*, 60–61, 252–272. <https://doi.org/10.1016/j.ymssp.2015.02.008>
- Eren, L., Ince, T., & Kiranyaz, S. (2019). A Generic Intelligent Bearing Fault Diagnosis System Using Compact Adaptive 1D CNN Classifier. *Journal of Signal Processing Systems*, 91(2), 179–189. <https://doi.org/10.1007/s11265-018-1378-3>
- Ezhilarasan, M., Thambidurai, P., Praveena, K., Srinivasan, S., & Sumathi, N. (2007). A New Entropy Encoding Technique for Multimedia Data Compression. *International Conference on Computational Intelligence and Multimedia Applications (ICCIMA 2007)*, 4, 157–161. <https://doi.org/10.1109/ICCIMA.2007.123>
- Fan, F., Xiong, J., & Wang, G. (2020). Universal approximation with quadratic deep networks. *Neural Networks*, 124, 383–392. <https://doi.org/10.1016/j.neunet.2020.01.007>
- Fan, F.-L., Li, M., Wang, F., Lai, R., & Wang, G. (2022). *On Expressivity and Trainability of Quadratic Networks* (arXiv:2110.06081). arXiv. <http://arxiv.org/abs/2110.06081>
- Gamanayake, C., Qin, Y., Yuen, C., Jayasinghe, L., Tan, D.-E., & Low, J. (2023). A Hybrid Deep Learning Model-Based Remaining Useful Life Estimation for Reed Relay With Degradation Pattern Clustering. *IEEE Transactions on Industrial Informatics*, 19(6), 7401–7413. <https://doi.org/10.1109/TII.2022.3210250>
- Gao, D., Zhu, Y., Ren, Z., Yan, K., & Kang, W. (2021). A novel weak fault diagnosis method for rolling bearings based on LSTM considering quasi-periodicity.

Knowledge-Based Systems, 231, 107413.
<https://doi.org/10.1016/j.knosys.2021.107413>

Ge, J., Niu, T., Xu, D., Yin, G., & Wang, Y. (2020). A Rolling Bearing Fault Diagnosis Method Based on EEMD-WSST Signal Reconstruction and Multi-Scale Entropy. *Entropy*, 22(3), Article 3. <https://doi.org/10.3390/e22030290>

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 249–256.
<https://proceedings.mlr.press/v9/glorot10a.html>

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep Sparse Rectifier Neural Networks. *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, 315–323. <https://proceedings.mlr.press/v15/glorot11a.html>

Graham, B. (2015). *Fractional Max-Pooling* (arXiv:1412.6071). arXiv.
<https://doi.org/10.48550/arXiv.1412.6071>

Gu, K., Zhang, Y., Liu, X., Li, H., & Ren, M. (2021). DWT-LSTM-Based Fault Diagnosis of Rolling Bearings with Multi-Sensors. *Electronics*, 10(17), Article 17.
<https://doi.org/10.3390/electronics10172076>

Gu, M. Y., Ge, J. Q., & Li, Z. N. (2023). Improved similarity-based residual life prediction method based on grey Markov model. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 45(6), 294. <https://doi.org/10.1007/s40430-023-04176-z>

- Gunerkar, R. S., Jalan, A. K., & Belgamwar, S. U. (2019). Fault diagnosis of rolling element bearing based on artificial neural network. *Journal of Mechanical Science and Technology*, 33(2), 505–511. <https://doi.org/10.1007/s12206-019-0103-x>
- Guo, J., & Zheng, P. (2020). A Method of Rolling Bearing Fault Diagnose Based on Double Sparse Dictionary and Deep Belief Network. *IEEE Access*, 8, 116239–116253. IEEE Access. <https://doi.org/10.1109/ACCESS.2020.3003909>
- Guo, L., Li, N., Jia, F., Lei, Y., & Lin, J. (2017). A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing*, 240, 98–109. <https://doi.org/10.1016/j.neucom.2017.02.045>
- Han, T., & Jiang, D. (2016). Rolling Bearing Fault Diagnostic Method Based on VMD-AR Model and Random Forest Classifier. *Shock and Vibration*, 2016, e5132046. <https://doi.org/10.1155/2016/5132046>
- He, M., & He, D. (2017). Deep Learning Based Approach for Bearing Fault Diagnosis. *IEEE Transactions on Industry Applications*, 53(3), 3057–3065. IEEE Transactions on Industry Applications. <https://doi.org/10.1109/TIA.2017.2661250>
- Hoang, D.-T., & Kang, H.-J. (2019). A survey on Deep Learning based bearing fault diagnosis. *Neurocomputing*, 335, 327–335. <https://doi.org/10.1016/j.neucom.2018.06.078>
- Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development*, 15(14), 5481–5487. <https://doi.org/10.5194/gmd-15-5481-2022>

- Hong, S., Zhou, Z., Zio, E., & Wang, W. (2014). An adaptive method for health trend prediction of rotating bearings. *Digital Signal Processing*, 35, 117–123. <https://doi.org/10.1016/j.dsp.2014.08.006>
- Hu, C.-H., Pei, H., Si, X.-S., Du, D.-B., Pang, Z.-N., & Wang, X. (2020). A Prognostic Model Based on DBN and Diffusion Process for Degrading Bearing. *IEEE Transactions on Industrial Electronics*, 67(10), 8767–8777. IEEE Transactions on Industrial Electronics. <https://doi.org/10.1109/TIE.2019.2947839>
- Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2020). Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8), 2011–2023. IEEE Transactions on Pattern Analysis and Machine Intelligence. <https://doi.org/10.1109/TPAMI.2019.2913372>
- Huo, Z., Zhang, Y., Francq, P., Shu, L., & Huang, J. (2017). Incipient Fault Diagnosis of Roller Bearing Using Optimized Wavelet Transform Based Multi-Speed Vibration Signatures. *IEEE Access*, 5, 19442–19456. IEEE Access. <https://doi.org/10.1109/ACCESS.2017.2661967>
- Ince, T., Kiranyaz, S., Eren, L., Askar, M., & Gabbouj, M. (2016). Real-Time Motor Fault Detection by 1-D Convolutional Neural Networks. *IEEE Transactions on Industrial Electronics*, 63(11), 7067–7075. IEEE Transactions on Industrial Electronics. <https://doi.org/10.1109/TIE.2016.2582729>
- Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *Proceedings of the 32nd International Conference on Machine Learning*, 448–456. <https://proceedings.mlr.press/v37/loffel5.html>

- Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, 31(3), 685–695. <https://doi.org/10.1007/s12525-021-00475-2>
- Jelodar, H., Wang, Y., Orji, R., & Huang, S. (2020). Deep Sentiment Classification and Topic Discovery on Novel Coronavirus or COVID-19 Online Discussions: NLP Using LSTM Recurrent Neural Network Approach. *IEEE Journal of Biomedical and Health Informatics*, 24(10), 2733–2742. *IEEE Journal of Biomedical and Health Informatics*. <https://doi.org/10.1109/JBHI.2020.3001216>
- Jia, F., Lei, Y., Lin, J., Zhou, X., & Lu, N. (2016). Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing*, 72–73, 303–315. <https://doi.org/10.1016/j.ymssp.2015.10.025>
- Jiang, H., Li, X., Shao, H., & Zhao, K. (2018). Intelligent fault diagnosis of rolling bearings using an improved deep recurrent neural network. *Measurement Science and Technology*, 29(6), 065107. <https://doi.org/10.1088/1361-6501/aab945>
- Jiang, H., Shao, H., Chen, X., & Huang, J. (2018). A feature fusion deep belief network method for intelligent fault diagnosis of rotating machinery. *Journal of Intelligent & Fuzzy Systems*, 34(6), 3513–3521. <https://doi.org/10.3233/JIFS-169530>
- Jiang, M., & Xiang, Y. (2023). Prediction on the Remaining Useful Life of Rolling Bearings Using Ensemble DLSTM. *Shock and Vibration*, 2023, 1–16. <https://doi.org/10.1155/2023/3742912>
- Jiao, J., & Zheng, X. (2022). Fault Diagnosis Method for Industrial Robots Based on DBN Joint Information Fusion Technology. *Computational Intelligence and Neuroscience*, 2022, 4340817. <https://doi.org/10.1155/2022/4340817>

- Juodelyte, D., Cheplygina, V., Graversen, T., & Bonnet, P. (2022). Predicting Bearings' Degradation Stages for Predictive Maintenance in the Pharmaceutical Industry. *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 3107–3115. <https://doi.org/10.1145/3534678.3539057>
- Kancharla, C. R., Vankeirsbilck, J., Vanoost, D., Boydens, J., & Hallez, H. (2022). Latent Dimensions of Auto-Encoder as Robust Features for Inter-Conditional Bearing Fault Diagnosis. *Applied Sciences*, 12(3), 965. <https://doi.org/10.3390/app12030965>
- Khanam, S., Tandon, N., & Dutt, J. K. (2014). Fault Size Estimation in the Outer Race of Ball Bearing Using Discrete Wavelet Transform of the Vibration Signal. *Procedia Technology*, 14, 12–19. <https://doi.org/10.1016/j.protcy.2014.08.003>
- Kou, L., Chu, B., Chen, Y., & Qin, Y. (2022). An Automatic Partition Time-Varying Markov Model for Reliability Evaluation. *Applied Sciences*, 12(12), Article 12. <https://doi.org/10.3390/app12125933>
- Kruiger, J. F., Rauber, P. E., Martins, R. M., Kerren, A., Kobourov, S., & Telea, A. C. (2017). Graph Layouts by t-SNE. *Computer Graphics Forum*, 36(3), 283–294. <https://doi.org/10.1111/cgf.13187>
- Leite, V. C. M. N., Borges da Silva, J. G., Veloso, G. F. C., Borges da Silva, L. E., Lambert-Torres, G., Bonaldi, E. L., & de Lacerda de Oliveira, L. E. (2015). Detection of Localized Bearing Faults in Induction Machines by Spectral Kurtosis and Envelope Analysis of Stator Current. *IEEE Transactions on Industrial Electronics*, 62(3), 1855–1865. <https://doi.org/10.1109/TIE.2014.2345330>

- Li, F., Cheng, Y., Tang, B., Zhou, X., & Tian, D. (2021). Multi-layer gated temporal convolution network for residual useful life prediction of rotating machinery. *Mechanical Systems and Signal Processing*, 155, 107600. <https://doi.org/10.1016/j.ymssp.2020.107600>
- Li, H.-F., Wei, J.-L., Li, S.-H., Liu, Y.-Q., Gu, X.-H., Liu, Z.-C., & Yang, S.-P. (2023). Fatigue life prediction of high-speed train bearings based on the generalized linear cumulative damage theory. *Fatigue & Fracture of Engineering Materials & Structures*, 46(6), 2112–2120. <https://doi.org/10.1111/ffe.13984>
- Li, J., Huang, R., & Li, W. (2020). Intelligent Fault Diagnosis for Bearing Dataset Using Adversarial Transfer Learning based on Stacked Auto-Encoder. *Procedia Manufacturing*, 49, 75–80. <https://doi.org/10.1016/j.promfg.2020.06.014>
- Li, N., Zhou, R., & Zhao, X. Z. (2011). Mechanical faulty signal denoising using a redundant non-linear second-generation wavelet transform. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 225(4), 799–808. <https://doi.org/10.1243/09544062JMES2410>
- Li, X., Yuan, A., & Lu, X. (2018). Multi-modal gated recurrent units for image description. *Multimedia Tools and Applications*, 77(22), 29847–29869. <https://doi.org/10.1007/s11042-018-5856-1>
- Li, X., Zhang, W., & Ding, Q. (2019). Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. *Reliability Engineering & System Safety*, 182, 208–218. <https://doi.org/10.1016/j.res.2018.11.011>

- Liao, D., Yin, M., Luo, H., Li, J., & Wu, N. (2022). Machine vision system based on a coupled image segmentation algorithm for surface-defect detection of a Si_3N_4 bearing roller. *JOSA A*, 39(4), 571–579. <https://doi.org/10.1364/JOSAA.449088>
- Liao, J.-X., Dong, H.-C., Sun, Z.-Q., Sun, J., Zhang, S., & Fan, F.-L. (2022). *Attention-embedded Quadratic Network (Qtention) for Effective and Interpretable Bearing Fault Diagnosis* (arXiv:2206.00390). arXiv. <http://arxiv.org/abs/2206.00390>
- Lin, P., Fu, S.-W., Wang, S.-S., Lai, Y.-H., & Tsao, Y. (2016). Maximum Entropy Learning with Deep Belief Networks. *Entropy*, 18(7), Article 7. <https://doi.org/10.3390/e18070251>
- Liu, H., Zhou, J., Zheng, Y., Jiang, W., & Zhang, Y. (2018). Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA Transactions*, 77, 167–178. <https://doi.org/10.1016/j.isatra.2018.04.005>
- Liu, W., Guo, P., & Ye, L. (2019). A Low-Delay Lightweight Recurrent Neural Network (LLRNN) for Rotating Machinery Fault Diagnosis. *Sensors*, 19(14), Article 14. <https://doi.org/10.3390/s19143109>
- Lu, C., Wang, Z.-Y., Qin, W.-L., & Ma, J. (2017). Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Processing*, 130, 377–388. <https://doi.org/10.1016/j.sigpro.2016.07.028>
- Luo, C., Jia, M., & Wen, Y. (2017). The diagnosis approach for rolling bearing fault based on Kurtosis criterion EMD and Hilbert envelope spectrum. *2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC)*, 692–696. <https://doi.org/10.1109/ITOEC.2017.8122439>

- Lv, D., Wang, H., & Che, C. (2021). Fault diagnosis of rolling bearing based on multimodal data fusion and deep belief network. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*. <https://doi.org/10.1177/09544062211008464>
- Ma, J., Wu, J., Fan, Y., & Wang, X. (2015). The Rolling Bearing Fault Feature Extraction Based on the LMD and Envelope Demodulation. *Mathematical Problems in Engineering*, 2015, e429185. <https://doi.org/10.1155/2015/429185>
- Mao, W., Feng, W., Liu, Y., Zhang, D., & Liang, X. (2021). A new deep auto-encoder method with fusing discriminant information for bearing fault diagnosis. *Mechanical Systems and Signal Processing*, 150, 107233. <https://doi.org/10.1016/j.ymssp.2020.107233>
- Martinez, M., & Stiefelhagen, R. (2018). *Taming the Cross Entropy Loss* (arXiv:1810.05075). arXiv. <https://doi.org/10.48550/arXiv.1810.05075>
- Masci, J., Meier, U., Cireşan, D., & Schmidhuber, J. (2011). Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction. In T. Honkela, W. Duch, M. Girolami, & S. Kaski (Eds.), *Artificial Neural Networks and Machine Learning – ICANN 2011* (pp. 52–59). Springer. https://doi.org/10.1007/978-3-642-21735-7_7
- Mikolov, T. (n.d.). *Recurrent neural network based language model*.
- Mohanty, S., Gupta, K. K., & Raju, K. S. (2014). Comparative study between VMD and EMD in bearing fault diagnosis. *2014 9th International Conference on Industrial and Information Systems (ICIIS)*, 1–6. <https://doi.org/10.1109/ICIINFS.2014.7036515>

- Mylonas, C., & Chatzi, E. (2020). *Remaining Useful Life Estimation Under Uncertainty with Causal GraphNets* (arXiv:2011.11740; Version 1). arXiv. <http://arxiv.org/abs/2011.11740>
- Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th International Conference on International Conference on Machine Learning*, 807–814.
- Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1), 1. <https://doi.org/10.1186/s40537-014-0007-7>
- Nie, X., & Xie, G. (2021). A novel normalized recurrent neural network for fault diagnosis with noisy labels. *Journal of Intelligent Manufacturing*, 32(5), 1271–1288. <https://doi.org/10.1007/s10845-020-01608-8>
- Nishat Toma, R., & Kim, J.-M. (2020). Bearing Fault Classification of Induction Motors Using Discrete Wavelet Transform and Ensemble Machine Learning Algorithms. *Applied Sciences*, 10(15), Article 15. <https://doi.org/10.3390/app10155251>
- Ozcan, I. H., Devecioglu, O. C., Ince, T., Eren, L., & Askar, M. (2022). Enhanced bearing fault detection using multichannel, multilevel 1D CNN classifier. *Electrical Engineering*, 104(2), 435–447. <https://doi.org/10.1007/s00202-021-01309-2>
- Patel, R., & Giri, V. (2015). Application of DWT and PDD for Bearing Fault Diagnosis Using Vibration Signal. *Journal of Electrical Engineering*, 15, 139–144.
- Peng, K., Jiao, R., Dong, J., & Pi, Y. (2019). A deep belief network based health indicator construction and remaining useful life prediction using improved particle filter. *Neurocomputing*, 361, 19–28. <https://doi.org/10.1016/j.neucom.2019.07.075>

- Peng, R., Zhang, X., & Shi, P. (2022). Bearing Fault Diagnosis of Hot-Rolling Mill Utilizing Intelligent Optimized Self-Adaptive Deep Belief Network with Limited Samples. *Sensors (Basel, Switzerland)*, 22(20), 7815. <https://doi.org/10.3390/s22207815>
- Pouyet, E., Rohani, N., Katsaggelos, A. K., Cossairt, O., & Walton, M. (2018). Innovative data reduction and visualization strategy for hyperspectral imaging datasets using t-SNE approach. *Pure and Applied Chemistry*, 90(3), 493–506. <https://doi.org/10.1515/pac-2017-0907>
- Puiu, A.-B., & Brabete, A.-O. (2019). *Towards NLP with Deep Learning: Convolutional Neural Networks and Recurrent Neural Networks for Offensive Language Identification in Social Media* (arXiv:1903.00665). arXiv. <https://doi.org/10.48550/arXiv.1903.00665>
- Qiu, J. (2016). *Convolutional Neural Network based Age Estimation from Facial Image and Depth Prediction from Single Image*.
- Qiu, S., Cui, X., Ping, Z., Shan, N., Li, Z., Bao, X., & Xu, X. (2023). Deep Learning Techniques in Intelligent Fault Diagnosis and Prognosis for Industrial Systems: A Review. *Sensors*, 23(3), Article 3. <https://doi.org/10.3390/s23031305>
- Rai, A., & Upadhyay, S. H. (2016). A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. *Tribology International*, 96, 289–306. <https://doi.org/10.1016/j.triboint.2015.12.037>
- Randall, R. B. (2017). A history of cepstrum analysis and its application to mechanical problems. *Mechanical Systems and Signal Processing*, 97, 3–19. <https://doi.org/10.1016/j.ymssp.2016.12.026>

- Ren, L., Sun, Y., Cui, J., & Zhang, L. (2018). Bearing remaining useful life prediction based on deep autoencoder and deep neural networks. *Journal of Manufacturing Systems*, 48, 71–77. <https://doi.org/10.1016/j.jmsy.2018.04.008>
- Rioul, O., & Vetterli, M. (1991). Wavelets and signal processing. *IEEE Signal Processing Magazine*, 8(4), 14–38. IEEE Signal Processing Magazine. <https://doi.org/10.1109/79.91217>
- Robust bearing degradation assessment method based on improved CVA - Wang—2017—*
IET Science, Measurement & Technology—Wiley Online Library. (n.d.). Retrieved 9 July 2023, from <https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/iet-smt.2016.0391>
- Rongjie, W., Yiju, Z., Meiqian, C., & Haifeng, Z. (2011). Fault diagnosis technology based on Wigner–Ville distribution in power electronics circuit. *International Journal of Electronics*, 98(9), 1247–1257. <https://doi.org/10.1080/00207217.2011.589738>
- Ruan, D., Han, J., Yan, J., & Gühmann, C. (2023). Light convolutional neural network by neural architecture search and model pruning for bearing fault diagnosis and remaining useful life prediction. *Scientific Reports*, 13(1), Article 1. <https://doi.org/10.1038/s41598-023-31532-9>
- Sanakkayala, D. C., Varadarajan, V., Kumar, N., Karan, Soni, G., Kamat, P., Kumar, S., Patil, S., & Kotecha, K. (2022). Explainable AI for Bearing Fault Prognosis Using Deep Learning Techniques. *Micromachines*, 13(9), Article 9. <https://doi.org/10.3390/mi13091471>

- Shao, H., Jiang, H., Lin, Y., & Li, X. (2018). A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders. *Mechanical Systems and Signal Processing*, 102, 278–297. <https://doi.org/10.1016/j.ymssp.2017.09.026>
- Shao, H., Jiang, H., Wang, F., & Wang, Y. (2017). Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet. *ISA Transactions*, 69, 187–201. <https://doi.org/10.1016/j.isatra.2017.03.017>
- Shao, H., Jiang, H., Zhang, X., & Niu, M. (2015). Rolling bearing fault diagnosis using an optimization deep belief network. *Measurement Science and Technology*, 26(11), 115002. <https://doi.org/10.1088/0957-0233/26/11/115002>
- Shao, H., Jiang, H., Zhao, H., & Wang, F. (2017). A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. *Mechanical Systems and Signal Processing*, 95, 187–204. <https://doi.org/10.1016/j.ymssp.2017.03.034>
- She, D., Jia, M., & Pecht, M. G. (2020). Sparse auto-encoder with regularization method for health indicator construction and remaining useful life prediction of rolling bearing. *Measurement Science and Technology*, 31(10), 105005. <https://doi.org/10.1088/1361-6501/ab8c0f>
- Shenfield, A., & Howarth, M. (2020). A Novel Deep Learning Model for the Detection and Identification of Rolling Element-Bearing Faults. *Sensors*, 20(18), Article 18. <https://doi.org/10.3390/s20185112>
- Shi, H., & Shang, Y. (2022). Initial Fault Diagnosis of Rolling Bearing Based on Second-Order Cyclic Autocorrelation and DCAE Combined With Transfer Learning. *IEEE Transactions on Instrumentation and Measurement*, 71, 1–18. *IEEE Transactions on Instrumentation and Measurement*. <https://doi.org/10.1109/TIM.2021.3132065>

- Simoens, J., & Vandewalle, S. (2003). A Stabilized Lifting Construction of Wavelets on Irregular Meshes on the Interval. *SIAM Journal on Scientific Computing*, 24(4), 1356–1378. <https://doi.org/10.1137/S1064827502376571>
- Smith, J. S. (2005). The local mean decomposition and its application to EEG perception data. *Journal of The Royal Society Interface*, 2(5), 443–454. <https://doi.org/10.1098/rsif.2005.0058>
- Sousa, R., Antunes, J., Coutinho, F., Silva, E., Santos, J., & Ferreira, H. (2019). Robust cepstral-based features for anomaly detection in ball bearings. *The International Journal of Advanced Manufacturing Technology*, 103(5), 2377–2390. <https://doi.org/10.1007/s00170-019-03597-2>
- Sutrisno, E., Oh, H., Vasan, A. S. S., & Pecht, M. (2012). Estimation of remaining useful life of ball bearings using data driven methodologies. *2012 IEEE Conference on Prognostics and Health Management*, 1–7. <https://doi.org/10.1109/ICPHM.2012.6299548>
- Sweke, R., Wilde, F., Meyer, J., Schuld, M., Faehrmann, P. K., Meynard-Piganeau, B., & Eisert, J. (2020). Stochastic gradient descent for hybrid quantum-classical optimization. *Quantum*, 4, 314. <https://doi.org/10.22331/q-2020-08-31-314>
- Tang, R., Yang, G., Wei, H., Mao, Y., Ture, F., & Lin, J. (2018). *Streaming Voice Query Recognition using Causal Convolutional Recurrent Neural Networks* (arXiv:1812.07754). arXiv. <https://doi.org/10.48550/arXiv.1812.07754>
- Tang, Y., Liu, Q., & Zhu, Q. (2021). Fault simulation and forecast of helical cylindrical gear of reducer based on ADAMS. *Journal of Physics: Conference Series*, 1983(1), 012019. <https://doi.org/10.1088/1742-6596/1983/1/012019>

- Toma, R. N., Piltan, F., Im, K., Shon, D., Yoon, T. H., Yoo, D.-S., & Kim, J.-M. (2022). A Bearing Fault Classification Framework Based on Image Encoding Techniques and a Convolutional Neural Network under Different Operating Conditions. *Sensors (Basel, Switzerland)*, 22(13), 4881. <https://doi.org/10.3390/s22134881>
- Tong, J., Luo, J., Pan, H., Zheng, J., & Zhang, Q. (2020). A Novel Cuckoo Search Optimized Deep Auto-Encoder Network-Based Fault Diagnosis Method for Rolling Bearing. *Shock and Vibration*, 2020, 1–12. <https://doi.org/10.1155/2020/8891905>
- Wan, L., Chen, Y., Li, H., & Li, C. (2020). Rolling-Element Bearing Fault Diagnosis Using Improved LeNet-5 Network. *Sensors (Basel, Switzerland)*, 20(6), 1693. <https://doi.org/10.3390/s20061693>
- Wang, C., Gan, M., & Zhu, C. (2018). Fault feature extraction of rolling element bearings based on wavelet packet transform and sparse representation theory. *Journal of Intelligent Manufacturing*, 29(4), 937–951. <https://doi.org/10.1007/s10845-015-1153-2>
- Wang, C., Jiang, W., Yang, X., & Zhang, S. (2021). RUL Prediction of Rolling Bearings Based on a DCAE and CNN. *Applied Sciences*, 11(23), Article 23. <https://doi.org/10.3390/app112311516>
- Wang, C., Sun, H., Zhao, R., & Cao, X. (2020). Research on Bearing Fault Diagnosis Method Based on an Adaptive Anti-Noise Network under Long Time Series. *Sensors (Basel, Switzerland)*, 20(24), 7031. <https://doi.org/10.3390/s20247031>
- Wang, F., Liu, X., Deng, G., Yu, X., Li, H., & Han, Q. (2019). Remaining Life Prediction Method for Rolling Bearing Based on the Long Short-Term Memory Network.

- Neural Processing Letters*, 50(3), 2437–2454. <https://doi.org/10.1007/s11063-019-10016-w>
- Wang, H., Liu, Z., Peng, D., & Qin, Y. (2020). Understanding and Learning Discriminant Features based on Multiattention 1DCNN for Wheelset Bearing Fault Diagnosis. *IEEE Transactions on Industrial Informatics*, 16(9), 5735–5745. IEEE Transactions on Industrial Informatics. <https://doi.org/10.1109/TII.2019.2955540>
- Wang, H., Liu, Z., Peng, D., Yang, M., & Qin, Y. (2022). Feature-Level Attention-Guided Multitask CNN for Fault Diagnosis and Working Conditions Identification of Rolling Bearing. *IEEE Transactions on Neural Networks and Learning Systems*, 33(9), 4757–4769. <https://doi.org/10.1109/TNNLS.2021.3060494>
- Wang, J., Qiao, L., Ye, Y., & Chen, Y. (2017). Fractional envelope analysis for rolling element bearing weak fault feature extraction. *IEEE/CAA Journal of Automatica Sinica*, 4(2), 353–360. IEEE/CAA Journal of Automatica Sinica. <https://doi.org/10.1109/JAS.2016.7510166>
- Wang, X., & He, Q. (2015). *Machinery Fault Signal Reconstruction Using Time-Frequency Manifold*. 19, 777–787. https://doi.org/10.1007/978-3-319-09507-3_68
- Wang, Y., Deng, L., Zheng, L., & Gao, R. X. (2021). Temporal convolutional network with soft thresholding and attention mechanism for machinery prognostics. *Journal of Manufacturing Systems*, 60, 512–526. <https://doi.org/10.1016/j.jmsy.2021.07.008>
- Wang, Y., Ding, X., Zeng, Q., Wang, L., & Shao, Y. (2021). Intelligent Rolling Bearing Fault Diagnosis via Vision ConvNet. *IEEE Sensors Journal*, 21(5), 6600–6609. <https://doi.org/10.1109/JSEN.2020.3042182>

- Wang, Y., Xiang, J., Markert, R., & Liang, M. (2016). Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications. *Mechanical Systems and Signal Processing*, 66–67, 679–698. <https://doi.org/10.1016/j.ymssp.2015.04.039>
- Wang, Y., Yao, H., & Zhao, S. (2016). Auto-encoder based dimensionality reduction. *Neurocomputing*, 184, 232–242. <https://doi.org/10.1016/j.neucom.2015.08.104>
- Wang, Z., Zhang, Q., Xiong, J., Xiao, M., Sun, G., & He, J. (2017). Fault Diagnosis of a Rolling Bearing Using Wavelet Packet Denoising and Random Forests. *IEEE Sensors Journal*, 17(17), 5581–5588. *IEEE Sensors Journal*. <https://doi.org/10.1109/JSEN.2017.2726011>
- Wen, L., Gao, L., & Li, X. (2019). A New Deep Transfer Learning Based on Sparse Auto-Encoder for Fault Diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(1), 136–144. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. <https://doi.org/10.1109/TSMC.2017.2754287>
- Wu, Z., & Huang, N. E. (2009). Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 01(01), 1–41. <https://doi.org/10.1142/S1793536909000047>
- Xia, T., & Xi, L. (2019). Manufacturing paradigm-oriented PHM methodologies for cyber-physical systems. *Journal of Intelligent Manufacturing*, 30(4), 1659–1672. <https://doi.org/10.1007/s10845-017-1342-2>
- Xiong, X., Hongkai, J., Li, X., & Niu, M. (2020). A Wasserstein gradient-penalty generative adversarial network with deep auto-encoder for bearing intelligent fault

- diagnosis. *Measurement Science and Technology*, 31(4), 045006.
<https://doi.org/10.1088/1361-6501/ab47df>
- Xu, F., Shu, X., Li, X., & Tang, R. (2021). Health indicator construction for roller bearing based on an unsupervised deep belief network with a novel sigmoid zero local minimum point model. *Structural Health Monitoring*, 20(4), 2110–2123.
<https://doi.org/10.1177/1475921720963951>
- Xu, G., Liu, M., Jiang, Z., Söffker, D., & Shen, W. (2019). Bearing Fault Diagnosis Method Based on Deep Convolutional Neural Network and Random Forest Ensemble Learning. *Sensors*, 19(5), 1088. <https://doi.org/10.3390/s19051088>
- Y, L. (1989). Generalization and network design strategies. *Technical Report*.
<https://cir.nii.ac.jp/crid/1570291224524363008>
- Yadavar Nikraves, S. M., Rezaie, H., Kilpatrik, M., & Taheri, H. (2019). Intelligent Fault Diagnosis of Bearings Based on Energy Levels in Frequency Bands Using Wavelet and Support Vector Machines (SVM). *Journal of Manufacturing and Materials Processing*, 3(1), Article 1. <https://doi.org/10.3390/jmmp3010011>
- Yan, C., Zhao, M., Lin, J., Liang, K., & Zhiqiang Zhang. (2020). Fault signature enhancement and skidding evaluation of rolling bearing based on estimating the phase of the impulse envelope signal. *Journal of Sound and Vibration*, 485, 115529.
<https://doi.org/10.1016/j.jsv.2020.115529>
- Yang, L., Wang, H., Tang, P., & Li, Q. (2021). CaptionNet: A Tailor-made Recurrent Neural Network for Generating Image Descriptions. *IEEE Transactions on Multimedia*, 23, 835–845. *IEEE Transactions on Multimedia*.
<https://doi.org/10.1109/TMM.2020.2990074>

- Yoder, J., & Priebe, C. E. (2017). Semi-supervised k-means++. *Journal of Statistical Computation and Simulation*, 87(13), 2597–2608.
<https://doi.org/10.1080/00949655.2017.1327588>
- Yu, K., Lin, T. R., & Tan, J. (2020). A bearing fault and severity diagnostic technique using adaptive deep belief networks and Dempster–Shafer theory. *Structural Health Monitoring*, 19(1), 240–261. <https://doi.org/10.1177/1475921719841690>
- Zhang, C., Pan, L., Wang, S., Wang, X., & Tomovic, M. (2018). An accelerated life test model for solid lubricated bearings used in space based on time-varying dependence analysis of different failure modes. *Acta Astronautica*, 152, 352–359.
<https://doi.org/10.1016/j.actaastro.2018.08.027>
- Zhang, G., Niwa, K., & Kleijn, W. B. (2019). *Rapidly Adapting Moment Estimation* (arXiv:1902.09030). arXiv. <https://doi.org/10.48550/arXiv.1902.09030>
- Zhang, W., Li, C., Peng, G., Chen, Y., & Zhang, Z. (2018). A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mechanical Systems and Signal Processing*, 100, 439–453. <https://doi.org/10.1016/j.ymssp.2017.06.022>
- Zhang, W., Peng, G., Li, C., Chen, Y., & Zhang, Z. (2017). A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals. *Sensors (Basel, Switzerland)*, 17(2), 425. PubMed.
<https://doi.org/10.3390/s17020425>
- Zhang, Y., Zhou, T., Huang, X., Cao, L., & Zhou, Q. (2021). Fault diagnosis of rotating machinery based on recurrent neural networks. *Measurement*, 171, 108774.
<https://doi.org/10.1016/j.measurement.2020.108774>

- Zhao, S., Liang, L., Xu, G., Wang, J., & Zhang, W. (2013). Quantitative diagnosis of a spall-like fault of a rolling element bearing by empirical mode decomposition and the approximate entropy method. *Mechanical Systems and Signal Processing*, 40(1), 154–177. <https://doi.org/10.1016/j.ymssp.2013.04.006>
- Zhiwei, L. (2022). Bearing Fault Diagnosis of End-to-End Model Design Based on 1DCNN-GRU Network. *Discrete Dynamics in Nature and Society*, 2022, e7167821. <https://doi.org/10.1155/2022/7167821>
- Zhong, X., Wang, F., & Ban, H. (2022). Development of a plug-and-play anti-noise module for fault diagnosis of rotating machines in nuclear power plants. *Progress in Nuclear Energy*, 151, 104344. <https://doi.org/10.1016/j.pnucene.2022.104344>
- Zhou, Q., Shen, H., Zhao, J., Liu, X., & Xiong, X. (2019). Degradation State Recognition of Rolling Bearing Based on K-Means and CNN Algorithm. *Shock and Vibration*, 2019, e8471732. <https://doi.org/10.1155/2019/8471732>
- Zhou, S., Xiao, M., Bartos, P., Filip, M., & Geng, G. (2020). Remaining Useful Life Prediction and Fault Diagnosis of Rolling Bearings Based on Short-Time Fourier Transform and Convolutional Neural Network. *Shock and Vibration*, 2020, e8857307. <https://doi.org/10.1155/2020/8857307>
- Zhou, Y., Chen, J., Dong, G. M., Xiao, W. B., & Wang, Z. Y. (2011). Wigner–Ville distribution based on cyclic spectral density and the application in rolling element bearings diagnosis. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 225(12), 2831–2847. <https://doi.org/10.1177/0954406211413215>

- Zhu, G., Zhu, Z., Xiang, L., Hu, A., & Xu, Y. (2023). Prediction of bearing remaining useful life based on DACN-ConvLSTM model. *Measurement*, 211, 112600. <https://doi.org/10.1016/j.measurement.2023.112600>
- Zhu, J., Nostrand, T., Spiegel, C., & Morton, B. (2014). Survey of Condition Indicators for Condition Monitoring Systems. *Annual Conference of the PHM Society*, 6(1), Article 1. <https://doi.org/10.36001/phmconf.2014.v6i1.2514>
- Zhu, Z., Lei, Y., Qi, G., Chai, Y., Mazur, N., An, Y., & Huang, X. (2023). A review of the application of deep learning in intelligent fault diagnosis of rotating machinery. *Measurement*, 206, 112346. <https://doi.org/10.1016/j.measurement.2022.112346>

APPENDICES

Appendix A: Journal Publications

1. **Li Xiuyu** and S. J. Tanjong, “Classification of bearing degradation stage based on automatic label assignment and multi-scale channel-attention network.”, PERTANIKA Journal of Science and Technology. **Status: Accepted by Pertanika Journal of Science and Technology**
2. **Li Xiuyu** and S. J. Tanjong, “Robust convolutional neural network for fault diagnosis of rolling bearings in noisy industrial environments”, Jordan Journal of Mechanical and Industrial Engineering. **Status: Accepted by Edelweiss Applied Science and Technology**
3. **Li Xiuyu**, S. J. Tanjong and A. R. H. Rigit, “Current development and future directions in deep learning-based techniques for bearing fault diagnosis.”, Journal of Engineering Science and Technology Review. **Status: Manuscript improvement, pending submission.**