图书馆

文献检索报告

兰州理工大学图书馆

报告编号：**R2025-1219**

机构：兰州理工大学 电气工程与信息工程学院

姓名：**Abduelhadi, Alaeldden [186081101001]**

著者要求对其在国内外学术出版物所发表的科技论著被以下数据库收录情况进行查证。

检索范围：

- 科学引文索引（Science Citation Index Expanded）： 1900年-2025年

检索结果：

检索类型	数据库	年份范围	总篇数	第一作者篇数
SCI-E 收录	SCI-EXPANDED	2024 - 2025	2	2

委托人声明：

本人委托兰州理工大学图书馆查询论著被指定检索工具收录情况，经核对检索结果，附件中所列文献均为本人论著，特此声明。

作　　者（签字）：Alaeldden

完 成 人（签字）：刘　婧

完　成　日　期：2025年10月28日

完成单位（盖章）：兰州理工大学图书馆信息咨询与学科服务部

（本检索报告仅限校内使用）

图书馆

文献检索报告
SCI-E 收录

兰州理工大学图书馆

报告编号：**R2025-1219 SCI-E 收录**

数据库：科学引文索引 **(Science Citation Index Expanded)** 时间范围：**2024年至2025年**	作者姓名：**Abduelhadi, Alaeldden** 作者单位：兰州理工大学 电气工程与信息工程学院	检索人员：刘　婧 检索日期：**2025年10月28日**

检索结果：被 **SCI-E** 收录文献 **2** 篇

#	作者	地址	标题	来源出版物	文献类型	入藏号
1	Abduelhadi, A; Liang, HP; Cao, J; Chen, P	[Abduelhadi, Alaeldden; Cao, Jie] Lanzhou Univ Technol, Sch Elect & Informat Engn, Lanzhou 730050, Peoples R China.; [Liang, Haopeng] Lanzhou Univ Technol, Sch Comp & Commun, Lanzhou 730050, Peoples R China.; [Chen, Peng] Lanzhou Petrochem Univ Vocat Technol, Sch Elect & Elect Engn, Lanzhou 730060, Peoples R China.	HF-MSCN: a high frequency-multiscale cascade network for bearing fault diagnosis	***MEASUREMENT SCIENCE AND TECHNOLOGY*** 2024, 35 (11): 116120.	J Article	WOS:001289751700001
2	Abduelhadi, A; Cao, J; Liang, HP; Yuan, SQ	[Abduelhadi, Alaeldden; Cao, Jie; Yuan, Shanqin] Lanzhou Univ Technol, Sch Elect & Informat Engn, Lanzhou 730050, Peoples R China.; [Liang, Haopeng] Lanzhou Univ Technol, Sch Comp & Commun, Lanzhou 730050, Peoples R China.	Efficient fault diagnosis method based on dynamic recalibration mechanism with improved inception network for rolling bearings under limited samples	***JOURNAL OF THE BRAZILIAN SOCIETY OF MECHANICAL SCIENCES AND ENGINEERING*** 2025, 47 (11): 554.	J Article	WOS:001565595200026
					合计	2

Your manuscript MST-121271.R1 has been accepted for publication

1 message

Measurement Science and Technology <onbehalfof@manuscriptcentral.com> 29 July 2024 at 19:02
Reply-To: mst@ioppublishing.org
To: alaeldden9090@gmail.com, 928782706@qq.com, haop1115@163.com, 1341987756@qq.com

Dear Dr Liang,

Re: "HF-MSCN: A High Frequency-Multiscale Cascade Network for bearing fault diagnosis"

Manuscript reference: MST-121271.R1

We are pleased to tell you that we are accepting your Paper. We now have everything we need to publish it in Measurement Science and Technology.

You can find more information about your accepted manuscript on our Publishing Support website.

You will receive a confirmation email which will include the Digital Object Identifier. This should be used for citing your work.

DATA AVAILABILITY
The proof of your article will contain an automatically-generated data availability statement based on your answer to the data availability question on the submission form, and will replace any other data availability statement(s) already present within the manuscript. Please ensure you check the proof for the correct data citation and consistency with any other mention(s) of data availability within the manuscript itself to avoid contradiction.

If you have opted to use IOP Publishing's Figshare repository to make your data openly available this has now been verified and the DOI will be automatically inserted into the proof of your article.

Next steps

If you have chosen to publish your work as Open Access, we will send an email in the next few days with details of how to pay the article charge for your Paper.

We will email again when article proofs are ready for your final approval. Please return them by the date given in that email so that we can produce the final published version without delay.

Further information:

The Track My Article section of our website is a useful resource.

You can contact our Customer Services department at this email address.

For advice on complying with US funder requirements visit IOP Science.

All our published articles are available to readers at IOP Science.

Thank you for choosing to work with Measurement Science and Technology. We look forward to publishing your Paper.

Yours sincerely,

Poppy Clark

On behalf of:
Measurement Science and Technology
Editor-in-Chief: Andrew Yacoot
iopscience.iop.org/mst | mst@ioppublishing.org
Impact Factor: 2.7

Want to find out what is happening to your submission?
Track your article on:
Publishing support https://bit.ly/39t9yPz
WeChat https://bit.ly/2L0M9uz

Letter reference: DEC:ForAcc:REV

Gmail Alaeldden Mohammed <alaeldden9090@gmail.com>

Your Submission BMSE-D-25-00460R2 - [EMID:72b3ed17b55cddb4]

1 message

Journal of the Brazilian Society of Mechanical Sciences and Engineering "(BMSE)" 29 July 2025 at 05:54
<em@editorialmanager.com>
Reply-To: "Journal of the Brazilian Society of Mechanical Sciences and Engineering \"(BMSE)\""
<mageshwar.sureshkumar@springer.com>
To: Alaeldden Abduelhadi <alaeldden9090@gmail.com>

CC: "Jie Cao" haop1115@163.com, "Haopeng Liang" 928782706@qq.com, "Shanqin Yuan" 1076435177@qq.com

Dear Dr Abduelhadi,

We are pleased to inform you that your manuscript, "Efficient Fault Diagnosis Method Based On Dynamic Recalibration Mechanism With Improved Inception Network For Rolling Bearings Under Limited Samples", has been accepted for publication in
Journal of the Brazilian Society of Mechanical Sciences and Engineering.

You will be contacted by Springer Author Services in due course with a link to complete the grant of rights. Please note that you will receive your proofs after the publishing agreement has been received through the online system.

Please remember to quote the manuscript number, BMSE-D-25-00460R2, whenever inquiring about your manuscript.

With best regards,

Marcelo A. Savi
Editor in Chief

Reviewer #5: The revised article makes it more detailed, and I agree to accept it.

Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to Plan S principles (https://www.springernature.com/gp/open-science/plan-s-compliance) or the NIH public access policy (https://www.springernature.com/gp/open-science/us-federal-agency-compliance)) then you should select the gold OA route, and we will direct you to the compliant route where possible. Because authors warrant under our subscription licensing terms that they haven't committed to licensing any version of their article under a licence inconsistent with the terms of our agreement - including the applicable embargo period - publication under the subscription model isn't suitable for authors whose funders require no embargo.

This letter contains confidential information, is for your own use, and should not be forwarded to third parties.

Recipients of this email are registered users within the Editorial Manager database for this journal. We will keep your information on file to use in the process of submitting, evaluating and publishing a manuscript. For more information on how we use your personal details please see our privacy policy at https://www.springernature.com/production-privacy-policy. If you no longer wish to receive messages from this journal or you have questions regarding database management, please contact the Publication Office at the link below.

In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Use the following URL: https://www.editorialmanager.com/bmse/login.asp?a=r). Please contact the publication office if you have any questions.

Gmail Alaeldden Mohammed <alaeldden9090@gmail.com>

Manuscript ID: ASIM-2025-000013 Registration completed

1 message

conference <mstracking@academicenter.com> 24 June 2025 at 15:40
To: alaeldden9090@gmail.com

Dear Alaeldden:

Your manuscript has been registered successfully and is on the process of proofreading. Please pay attention to your email.

Manuscript No: ASIM-2025-000013
Title: Spiking Multi Scale Graph Convolutional Neural Networks for Fault Diagnosis of Rotating Machinery
Conference: The 4th International Conference on Advanced Sensing and Intelligent Manufacturing

Kind regards,
Yours Sincerely,
AC

IOP Publishing Measurement Science and Technology
Meas. Sci. Technol. **35** (2024) 116120 (14pp) https://doi.org/10.1088/1361-6501/ad6893

HF-MSCN: a high frequency-multiscale cascade network for bearing fault diagnosis

Alaeldden Abduelhadi[1], Haopeng Liang[2,*], Jie Cao[1] and Peng Chen[3]

[1] School of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, People's Republic of China
[2] School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, People's Republic of China
[3] School of Electronic and Electrical Engineering, Lanzhou Petrochemical University of Vocational Technology, Lanzhou 730030, People's Republic of China

E-mail: haop1115@163.com

Received 11 May 2024, revised 10 July 2024
Accepted for publication 29 July 2024
Published 13 August 2024

Abstract

In the field of data-driven fault diagnosis (FD), deep learning methods have proven their excellent performance, especially when dealing with complex signals from rotating equipment such as bearings. However, fault features in vibration signals are often mixed with noise features and distributed at different frequency scales, posing challenges for effective feature extraction. In order to solve this problem, this paper proposes a high frequency-multiscale cascade network (HF-MSCN), which enhances the noise suppression and feature learning capability of the model by combining a high-frequency convolutional block (HFCB) with a multi-scale cascade block (MSCB). HFCB effectively suppresses high-frequency noise through wide convolutional layers and self-attention mechanisms while still retaining essential high-frequency fault signals. MSCB enhances the interaction between convolutional layers at different scales by cascading the layers at different scales and strengthens the model's ability to capture subtle fault features, especially when processing periodic fault pulse signals. Finally, we investigate the internal functioning of the network using time–frequency analysis methods in signal processing to improve the interpretability of deep learning methods in FD applications and further verify the enhanced effect of HFCB and MSCB on feature extraction. We validate the effectiveness of HF-MSCN on the case western reserve university dataset as well as a self-constructed bearing composite fault dataset, and the experimental results demonstrate that the network exceeds the performance of six state-of-the-art fault diagnostic methods in high-noise environments.

Keywords: rolling bearing, fault diagnosis, interpretability of deep learning, high frequency-multiscale cascade network

1. Introduction

Rotating machinery plays a vital role in various modern industrial systems, encompassing a wide array of equipment such as bearings, compressors, gears, aircraft engines, and steam turbines. Rolling bearings and gearboxes are essential components in rotating machinery; any malfunction in the bearing or gear system might result in avoidable shutdowns, causing substantial financial losses and potentially even casualties. Research on rolling bearing fault diagnosis (FD) technology has always received widespread attention due to its extensive utilization and significance in production processes [1–3]. With advancements in artificial intelligence and data processing technologies, most current rolling bearing

* Author to whom any correspondence should be addressed.

FD methods are based on vibration data analysis. Data-driven FD methods can extract and detect useful fault information from a large amount of monitoring data and do not need to establish an accurate system model. They are suitable for complex systems where selecting an explicit model is difficult. Therefore, more and more research on data-driven FD methods of rolling bearing has been paid attention [4].

The intelligent diagnosis of rolling bearing equipment is greatly facilitated by advancements in modern signal processing methods and deep learning, which offer robust technical assistance [5, 6]. In recent years, deep learning-based FD methods have gained significant popularity due to their remarkable success in pattern recognition. These methods can process raw data directly through neural networks, which automatically extract features from the data. This capability eliminates the need for extensive preliminary data extraction, reducing the reliance on manual efforts and expert knowledge. convolutional neural networks (CNNs) have shown extraordinary success in various domains, such as image classification and speech recognition [7], driven by advances in deep learning technologies. In light of these developments, the application of CNNs in the field of bearing FD has garnered significant interest and has been increasingly adopted [8–10]. For example, Zhang *et al* [11] proposed a novel deep-learning model for intelligent FD of rolling element bearings using raw vibration signals, using wide kernels in the first convolutional layer to extract features and suppress noise and small kernels in the following layers to perform multilayer nonlinear mapping. Chen *et al* [12] proposed a new method that combines CNN and extreme learning machine (ELM) for FD. This approach emphasizes data-driven methodologies for fast, reliable mechanical FD. It comprises three steps: refining raw vibration signals using continuous wavelet transform (CWT), extracting features using a CNN with a square pooling architecture, and classification using ELM. Liu *et al* [13] introduced a multitask one-dimensional (1D) CNN that effectively handles FD tasks alongside speed and load identification tasks. This approach uses shared features to enhance diagnostic accuracy. However, there may be very subtle differences between different fault types of bearings, especially in noisy and disturbed data environments, and CNNs may have difficulty in distinguishing these subtle differences. This high inter-class similarity and intra-class variability poses a challenge to the accuracy of FD.

The attention mechanism has recently been integrated into deep learning models, which has dramatically improved interpretability in a variety of disciplines [14, 15]. By using weights to specify various input data segments, the attention mechanism allows the model to focus on vibration pattern anomalies, thereby enhancing diagnostic capabilities [16]. Furthermore, it focuses on different segments of the input sequence to produce each segment of the output, enabling it to detect long-term dependencies and complex patterns within the data [17]. Considering this benefit, numerous deep learning architectures have been proposed to solve complex FD problems. Zhao *et al* [18] proposed the deep multi-scale adversarial network with spatial attention for domain adaptation in FD. The utilization of an attention mechanism helps in directing the model's attention towards transferable elements across various operating situations. This mechanism also facilitates the extraction of rich features at different scales and enables the model to adapt its performance dynamically. A lightweight attention-based multi-scale residual CNN was developed by Xu *et al* [19] The system is designed to diagnose faults in mechanical equipment in noisy environments. Initially, a noise reduction module is used to remove undesirable data. Following this, a feature enhancement module is employed to expand the receptive field, aiding CNN in extracting more relevant features. In addition, the implementation of a joint attention module is utilized to efficiently combine these features, enhancing the utilization of discriminative features. Wang *et al* [20] introduced a multi-scale structure by splitting the original vibration signal into separate scales using an averaging technique. Attention mechanisms are then used to strengthen the ability to distinguish between different scales and channels, thereby enhancing the discriminative capability of multi-scale features.

The combination of attention and neural networks has good research prospects in the field of FD, but there is still no one standard design about the structure of attention, and most of the attention does not establish a direct connection with the fault mechanism.

In summary, deep learning methods have made great achievements in bearing FD tasks, but still face the following challenges:

(1) Although deep neural networks represented by CNNs are able to automatically learn features from vibration data, they are often unable to effectively distinguish which features are critical fault information and which are environmental interference information.
(2) Failures in bearings often occur in a complex manner. As a result, simultaneous damage at multiple locations in a bearing may lead to overlap between features, which makes the model tend to recognize composite failures as more frequently occurring single failures when dealing with them.
(3) As a 'black box' model, deep neural networks are not transparent to the user in terms of the decision-making process and internal operation mechanism, which limits their application in real industries with strict auditing.

In response to the above problems, this paper proposes a FD method based on a high-frequency multi-scale cascade network (HF-MSCN). Firstly, a high-frequency convolutional block (HFCB) is built by utilizing wide convolutional layers and a self-attention mechanism to filter high-frequency noise information. Then, a multi-scale cascade block (MSCB) is designed to extract the relationship between different fault information through the attention cascade layer. Finally, an efficient FD model is constructed using the two blocks mentioned above, and the CWT is employed to analyze the signal changes within the model and explain its diagnosis mechanism. The primary contributions of this article are summarized as follows:

(1) A HF-MSCN model is proposed, which includes two blocks: HFCB and MSCB. HFCB forces the model to distinguish between high-frequency noise features and high-frequency fault features while filtering irrelevant noise information. MSCB identifies the correlation of sophisticated fault characteristics, enabling the model to differentiate between composite faults and single faults.
(2) A rolling bearing FD framework based on HF-MSCN is designed, which can effectively handle diagnostic tasks under strong noise and compound fault conditions. Simultaneously, the framework shows excellent FD results, indicating its ability to adjust to variable industrial conditions.
(3) This article analyzes the fluctuations in non-periodic signals within the deep learning model from a time–frequency domain perspective, thereby improving its interpretability in the field of FD.

This paper is organized as follows: section 2 introduces the proposed HF-MSCN in detail, including HFCB and MSCB. Section 3 validates the effectiveness of the HF-MSCN method using two bearing datasets. Additionally, section 3 discusses the model interpretability. Finally, Section 4 provides a summary of the research conducted in this study.

2. The proposed method

2.1. Method overview

In real industrial scenarios, rolling bearings usually exhibit different types of faults, some of which may have similar vibration responses, leading to FD methods that are prone to misidentify complex fault types [21]. In addition, environmental noise is also a non-negligible interference, which causes the network to be insufficiently sensitive to fault information due to the masking of high-frequency noise components [22]. In response to these problems, this paper proposes an end-to-end HF-MSCN, which has excellent feature learning capability and anti-noise performance. The structure of HF-MSCN is shown in figure 1. It can be seen that the bearing acquisition system captures raw vibration data from rotating machinery and feeds it into the HF-MSCN FD model. The HF-MSCN model consists of two primary feature learning blocks: The HFCB and the MSCB. HFCB employs wide convolutional layers and attention to filter high-frequency noise effectively, and MSCB utilizes the interaction of adjacent layers to deal with the similarity challenge between fault classes. Moreover, the number of MSCBs can be flexibly adjusted to accommodate data of various sizes. Then, the global average pooling layer (GAP) and softmax classification layer are used to aggregate features and output diagnostic results.

HF-MSCN has two advantages: (1) HF-MSCN can focus on the high-frequency components of the vibration signal to improve the model's ability to understand and differentiate between high-frequency noise information and high-frequency fault information. (2) A cascaded hierarchical structure is built in HF-MSCN to study the correlation and difference between fault classes. This structure facilitates the connection between information in different channel domains and enhances the correlation of pulse information captured by the model. The following sections provide comprehensive details of the proposed HFCB and MSCB.

2.2. HFCB

During the operation of bearings, the friction of the parts themselves and the vibration of the surrounding machines generate environmental noise, which may mask fault information in the vibration signals. In the field of deep learning-driven FD, the use of a wide convolutional layer with a large convolutional kernel is a common strategy to resist noise interference. The wide convolutional layer is usually a convolutional layer with large convolutional kernels. For example, in a conventional convolutional layer, the size of the convolutional kernel is between 3×1 and 7×1, while in a wide convolutional layer, the size of the convolutional kernel is usually 32×1, 64×1 or even larger. The larger convolution kernel covers a longer time span, which means it has a lower frequency response to the input signal. As a result, the wide convolutional layer can act as a low-pass filter to filter out high-frequency noise from the vibration signal. However, bearing damage often produces a specific high-frequency response in the vibration signal, making it difficult for the wide convolutional layer to distinguish between noise and fault response, so relying exclusively on the wide convolutional layer to process the vibration signals may lead to reduced sensitivity in fault detection.

To address this problem, this paper designs a HFCB, and its structure is shown in figure 2. In the HFCB, the attention mechanism is used to calculate the interrelationships between the features extracted by wide convolutional layers. Then, the focused information is dynamically assigned to determine which information is strengthened or suppressed. The detailed step-by-step process is described below:

First, a wide convolutional layer is used to perform a convolutional operation independently on the input signal x, the output of which is:

$$y = F(x,\ W) \tag{1}$$

where $F(\cdot)$ represents the convolution operation and represents the output of the wide convolutional layer. When the convolutional layer processes the input data, the sub-features of each channel can participate in the convolution operation independently, thus $y = [y_1, y_2, \ldots, y_C]$, y_i represents the feature of the ith channel.

Then, for each feature map y_i, we use three independent sets of parameters W_Q, W_K, W_V to generate query (Q), key (K), and value (V), which can be achieved through 1×1 convolution operations. The specific formula is as follows:

$$Q_i = W_Q * y_i \tag{2}$$

$$K_i = W_K * y_i \tag{3}$$

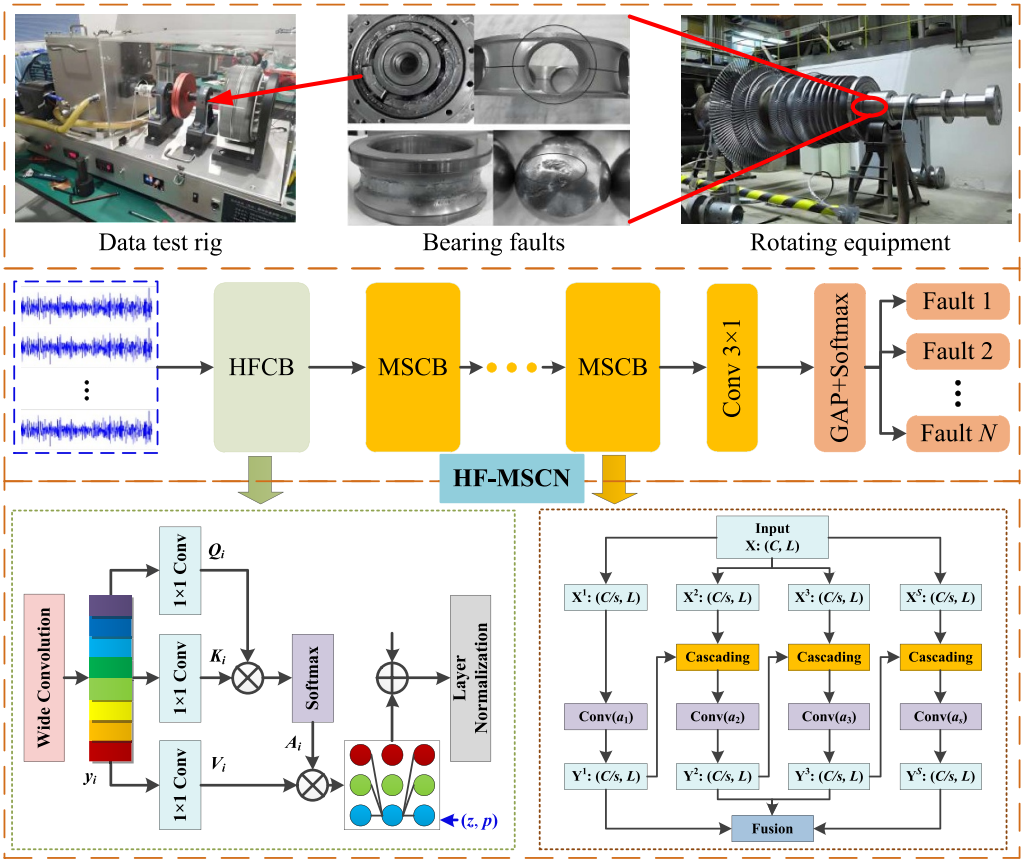


Figure 1. The structure of HF-MSCN.

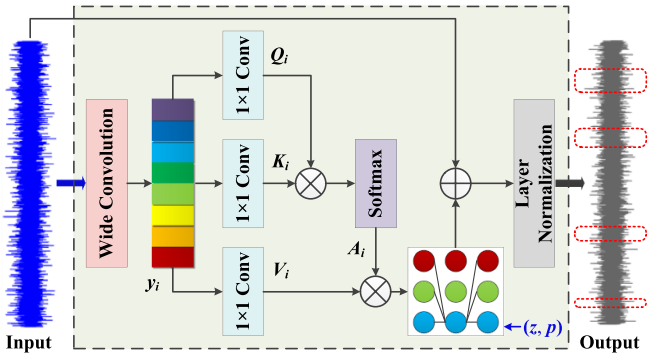


Figure 2. The structure of HFCB.

$$V_i = W_V * y_i \tag{4}$$

where Q_i, K_i, V_i represent the query, key, and value output by each channel, which will be used to calculate the attention score and analyze the interrelationships. The attention score is obtained by calculating the similarity between the query (Q) and the key (K). The calculation formula is as follows:

$$A_i = \text{softmax}\left(\frac{Q_i K_i^{\mathrm{T}}}{\sqrt{d_k}}\right) \tag{5}$$

where K_i^{T} represents the transpose of K_i, which is used for the dot product operation with the query Q_i, and d_k represents the dimension of the key vector for scaling the dot product result to stabilize the gradient. The softmax function processes the computed scores to obtain a probability from the score A_i. Furthermore, the sum of all scores is 1, reflecting the relative importance of each value (V) corresponding to the output. Finally, the attention score A_i is weighted by the corresponding value V_i to obtain the weighted output z. The formula is as follows:

$$z = \sum_i A_i V_i. \tag{6}$$

The attention mechanism dynamically determines the contribution of each feature of z above.

After the above processing, the high-frequency features of the noise in the input are dynamically suppressed, and currently, only low-frequency features and high-frequency fault features remain. Since the wide convolutional layer is more responsive to low-frequency features, in order to avoid the wide convolutional layer from being overly dependent on low-frequency features, we randomly set some output features to zero, forcing the model to learn both low-frequency features and high-frequency features. Ability is more balanced. The formula is as follows:

$$z_{\text{Dropout}} = \text{Dropout}\,(z, p) \tag{7}$$

where p represents the probability of maintaining feature activation. Finally, the residual connection between the original

input and the current output is constructed to enhance the learning objective of the network and alleviate the gradient vanishing problem. The formula is as follows:

$$o = \mathrm{LN}\left(z_{\mathrm{Dropout}} + x\right) \tag{8}$$

where + represents element-wise addition, LN represents layer normalization, which helps the network achieve faster convergence, and o represents the final output feature.

In summary, the HFCB consists of the wide convolutional layer and attention, where attention determines the importance of each feature by calculating the interrelationships between features, which is particularly important for distinguishing between noise and faults. This is because faulty signals are generally associated with periodicity, while noise is more random. Specifically, attention first transforms each channel's features into query (Q), key (K), and value (V), which creates a unique representation for each feature vector. After that, attention calculates the attention score and output weights to reflect the similarity between different features, where features with high scores are given more weight in the final output. As a result, the signals related to faults are strengthened while the noise is suppressed, and the HFCB is also able to judge the difference between high-frequency noise and high-frequency fault signals. In addition, random deactivation and residual connection operations improve network training and balance low-frequency and high-frequency feature learning.

2.3. MSCB

Bearing faults produce distinct periodic pulse signals in the vibration signal with strong signal correlation between different pulse signals. By capturing and analyzing this correlation information, the failure type and severity can be effectively identified. The proposed MSCB aims to make the network pay more attention to the correlation of pulse information within the vibration signals. The structure of MSCB is shown in figure 3, which mainly consists of cascade layers and multi-scale convolution layers. The feature extraction process is as follows. Suppose the output of the previous layer of the MSCB is $\boldsymbol{U} \in \mathrm{R}^{P\times L}$, where P represents the number of channels, and L represents the feature size of the previous layer. We use a 1×1 convolutional layer to change its channel number to C and get $\boldsymbol{X} \in \mathrm{R}^{C\times L}$, and then use $\boldsymbol{X}$ as the input for MSCB. This approach helps to improve the availability of the MSCB in the network and ensures that the MSCB is able to adapt to changes in the number of channels. In MSCB, $\boldsymbol{X}$ is first divided into s feature subsets in the channel dimension, which can be represented by $\boldsymbol{X}^i$, $1 \leqslant i \leqslant s$, and each feature subset contains C/s channel features. Therefore, $\boldsymbol{X}^i$ can be expressed as follows:

$$\boldsymbol{X}^i = \left[\boldsymbol{X}^i_1, \boldsymbol{X}^i_2, \ldots, \boldsymbol{X}^i_{C/s}\right]^{\mathrm{T}} \tag{9}$$

where $\boldsymbol{X}^i_{C/s}$ represents the C/s channel feature in the ith feature subset.

Then, the multi-scale convolution layer Conv (a_i) is assigned to the corresponding feature subset to extract multi-scale features, where the size of the convolution kernel a_1 is 3×1, $a_1 < a_2 < a_3, \ldots, < a_s$, and the size of convolution kernel a_s is $(3 + 2 \times n) \times 1$, $0 \leqslant n \leqslant s - 1$. After that, we use cascaded layers to establish interactions between different convolutional layers. It can be seen that the output feature $\boldsymbol{Y}^{s-1}$ of each convolution layer interacts with the input feature $\boldsymbol{X}^s$ of the next layer in the cascade layer, e.g. the output feature$\boldsymbol{Y}^1$ of Conv(a_1) interacts with the input feature $\boldsymbol{X}^2$ of Conv(a_2) in the cascade layer.

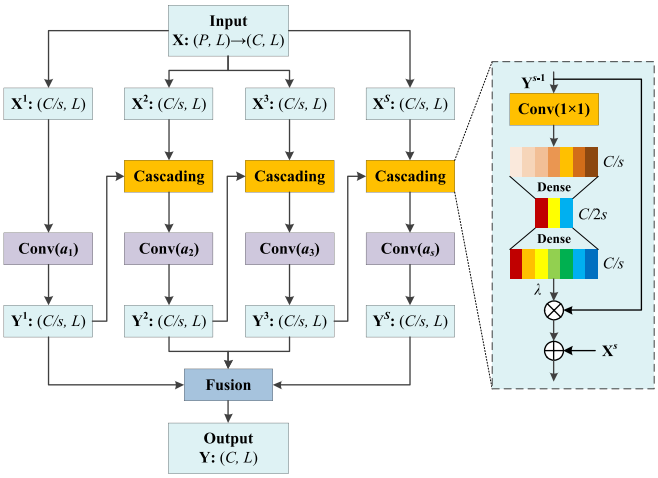


Figure 3. The structure of MSCB.

In the cascade layer, we use a 1×1 convolutional layer to aggregate the global information in the output feature $\boldsymbol{Y}^{s-1}$ and generate a feature vector $q \in \mathrm{R}^{1\times L}$ in the channel domain. Two dense layers are used to encode the importance of pulse information in the output features. Then, an activation function λ is applied to map the feature vector $q$to a fixed weight range and multiply this weight range with $\boldsymbol{Y}^{s-1}$ to enhance the meaningful signal segment features. After enhancing the pulse information in the output feature $\boldsymbol{Y}^{s-1}$, we fuse them with the input $\boldsymbol{X}^s$ of the next convolution and feed the fused features into a larger convolutional layer as a way to enable interaction between adjacent layers in multi-scale convolutional layers. The interaction between adjacent layers also helps the network capture important time-domain signal correlations more effectively.

Finally, the output features $\boldsymbol{Y}^i$ of all convolutional layers are fused, which will pass through channel fusion to obtain the output feature $\boldsymbol{Y} \in \mathrm{R}^{C\times L}$. For the entire MSCB, the relationship between the input $\boldsymbol{X}$ and the output $\boldsymbol{Y}$ can be described as follows:

$$Y = \begin{cases} F\left(\boldsymbol{X}^i, W_i\right), & i = 1 \\ F\left(\left(\boldsymbol{X}^i + \boldsymbol{Y}^i\right), W_i\right), & 2 \leqslant i \leqslant s \end{cases} \tag{10}$$

where $F(\cdot)$ represents the convolution operation, and W_i represents the weight of the convolutional layer. In summary, MSCB not only enhances the feature interaction but also makes the network more sensitive and accurate in dealing with transiently signals.

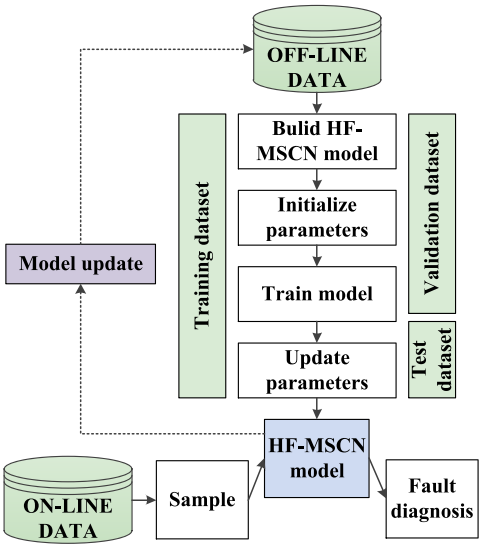


Figure 4. The fault diagnosis process of the proposed method.

2.4. FD process of the method

Figure 4 illustrates the FD process of the proposed method. This process includes two phases: offline training and online testing. In the offline training phase, offline data is used to train and build the HF-MSCN model. This involves designing the network architecture and defining the model's layers and components. Next, the model's initial parameters, including weights and biases, are set. Finally, training data is used to update the model parameters, a separate validation dataset is used to validate the model's performance, and a test dataset is used to evaluate the performance of the trained model. In the online testing phase, sampled online data is input into the HF-MSCN model for real-time FD. The well-trained and validated model can accurately classify fault conditions based on the input data. The model is periodically updated using real-time FD results and new online data. This involves retraining or fine-tuning the model with new data to ensure it remains accurate and up-to-date.

2.5. Interpretability of the method

In order to enhance the interpretability of HF-MSCN in FD, the CWT is used in this paper to analyze the time–frequency characteristics of each layers' weights in the model. First, the weight matrices are extracted from each feature layer of the HF-MSCN model. Where each weight matrix includes the weights of multiple convolutional kernels and corresponds to multiple specific output feature mappings. To perform the time–frequency analysis, the average of the weights of each convolution kernel is calculated and transformed into a 1D weight sequence. Then, CWT is applied to the 1D weight sequence, while Mexican Hat wavelet, which has a good response to transient characteristics, is chosen as the mother wavelet. Finally, the wavelet coefficients of the 1D weight sequence are obtained by stretching and translating the mother wavelet. The amplitude of these wavelet coefficients reflects the activity of the weights in both frequency and time, which can be used to identify the sensitivity and responsiveness of the model to specific frequency components. For the transformed results, we visualize them by using heat maps, where the shades of the map's colors reflect the signal strengths at different locations and frequencies.

3. Experimental results and discussion

In this section, we provide the parameters of the proposed HF-MSCN in detail and conduct fault analysis experiments on the bearing test bench of case western reserve university (CWRU) and our self-built bearing test bench.

3.1. Hyper-parameters of HF-MSCN

This paper proposes a general and flexible bearing failure method, HF-MSCN. The architecture of HF-MSCN mainly consists of stacks of HFCB, MSCB, and pooling layers. Their hyper-parameters are shown in table 1. In table 1, Kernel size represents the size of the convolutional kernel in the convolutional layer and the size of the pooling window in the pooling layer. Stride represents the reduction parameter of the feature size, e.g. the feature size is reduced by a factor of 1 when the stride is 2. Channel represents the number of channels of the layer's output feature. Output size represents the output feature tensor of the corresponding layer, which is usually expressed as 'feature size $\times$ number of channels'. '$-$' represents that the corresponding layer does not contain such parameters. In order to ensure that the input signal sample contains the complete signal period, the shape of the input signal is set to 1024×1, and the convolution kernel used in HFCB is 32×1, which can retain more valuable high-frequency information in the vibration signal. Four convolution kernels are used in the MSCB, namely, 3×1, 5×1, 7×1, and 9×1, which involve convolution kernels to help the network to extract multi-scale information comprehensively. The Stride of each pooling layer is kept to 2 to halve the signal size for each channel. As the network goes deeper, the size of the signal continues to decrease, so we use a regular convolutional layer (Conv) as the last feature layer of the network. The number of channels gradually increases from 16 to 64, along with the pattern of change of channels in deep neural networks. In the classification stage, a GAP layer and Softmax function are used to output the final diagnosis result. Multiple sets of experiments are conducted to determine the above optimal parameters. In particular, the parameters of the proposed HFCB and MSCB are discussed in more detail in the subsequent experimental section.

3.2. Comparison methods

In order to verify the superiority of the proposed method, several advanced methods are selected as comparison methods in this paper, they are MCAMDN [23], RESCNN [24], MSDARN [25], MBSDCN [26], MA1DCNN [27], GTFE-Net [28]. MCAMDN is a multi-scale network with selectable branching that resists noise interference while further

Table 1. Hyper-parameters of the HF-MSCN model.

Layer	Type	Kernel size	Stride	Channel	Output size
1	HFCB	32×1	4	16	$256 \times 1 \times 16$
2	Pooling	2×1	2	—	$128 \times 1 \times 16$
3	MSCB	$3 \times 1/5 \times 1/7 \times 1/9 \times 1$	1	32	$128 \times 1 \times 32$
4	Pooling	2×1	2	—	$64 \times 1 \times 32$
5	MSCB	$3 \times 1/5 \times 1/7 \times 1/9 \times 1$	1	64	$64 \times 1 \times 64$
6	Pooling	2×1	2	—	$32 \times 1 \times 64$
7	Conv	3×1	1	64	$32 \times 1 \times 64$
8	Pooling	2×1	2	—	$16 \times 1 \times 64$
9	GAP	—	—	—	1×64
10	Softmax	—	—	—	10/4

extracting relevant information. RESCNN is a residual network structure that consists of a residual module and a wide convolutional anti-noise module. MBSDCN adopts a multi-feature segmentation technique to enhance the richness of the feature information and improve the model's learning efficiency. MA1DCNN incorporates a joint multi-attention module, which achieves better results in bearing FD. GTFE-Net enhances information in both the time and frequency domains and utilizes a Gramian strategy to increase the model's resistance to noise.

3.3. Implementation details and evaluation metrics

In this paper, all experiments are conducted on the deep learning framework Tensorflow 2.5, and the programming language is Python 3.7. The computer used for the experiments is configured with AMD 5600H CPU, NVIDIA RTX 3050 GPU, and 16 GB of RAM. We set the model training batch for each experiment to 50, the batch size to 32, and the learning rate to 0.001. Adam optimizer is employed to minimize fluctuations throughout the model training process and to compute the average accuracy after 5 iterations as the final experimental results. Typically, rolling bearing test benches are unable to replicate the environmental noise in actual engineering. For this reason, we add Gaussian white noise to the vibration data to simulate the noise environment. The complexity of the noise environment is reflected by the signal-to-noise ratio (SNR), which is defined as follows:

$$\mathrm{SNR} = 10\log_{10}\frac{\|\boldsymbol{X}\|_2^2}{\|\boldsymbol{X} - \Phi\|_2^2} \tag{11}$$

where $\boldsymbol{X}$ represents the input signal, Φ represents the noise-corrupted signal, and $\|\cdot\|_2^2$ represents the square of the ℓ_2 - norm. First, the power of the noise is calculated by using the power of the original vibration signal and the desired SNR. After that, the noise signal is generated according to the power, in the paper Gaussian white noise signal is generated because of it has uniform power distribution at all frequencies. Finally, these noise signals are directly superimposed on the raw signal to obtain the noise affected signal. Figures 5(a) and (b) are the visualizations of vibration signal's time domain and frequency domain, which show relatively smooth vibration signals with clear periodicity characteristics and distinct frequency components. When Gaussian white noise is introduced, it can be seen from figures 5(c), (e) and (g) that the periodicity of the signal becomes less obvious and the amplitude fluctuation of the signal gradually increases. From figures 5(d), (f) and (h), it can be seen that the noise progressively drowns out the specific frequency components of the signal and the critical frequency peaks are difficult to be recognized.

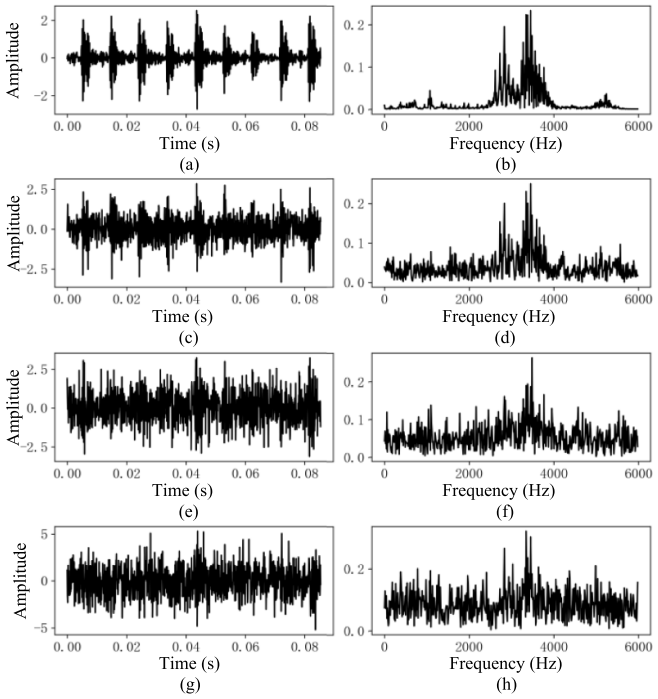


Figure 5. The visualizations of vibration signal's time domain and frequency domain under different noises. (a) and (b) the time domain and frequency domain without noise; (c) and (d) the time domain and frequency domain under 0 dB noise; (e) and (f) the time domain and frequency domain under −4 dB noise; (g) and (h) the time domain and frequency domain under −6 dB noise.

In terms of evaluation metrics for FD results, we use four indicators: accuracy, precision, recall, and F1-score, and their formulas are as follows:

Table 2. Details of the CWRU dataset.

Fault location	Fault size (mil)	Label
Normal	0	0
Inner race	7	1
Inner race	14	2
Inner race	21	3
Outer race	7	4
Outer race	14	5
Outer race	21	6
Ball	7	7
Ball	14	8
Ball	21	9

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP}+\text{TN}+\text{FN}+\text{FP}} \times 100\% \tag{12}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \times 100\% \tag{13}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \times 100\% \tag{14}$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \tag{15}$$

where TP, TN, FN, FP represent the number of true-positive samples, the number of true-negative samples, the number of false-negative samples, and the number of false-positive samples, respectively.

Case 1: Discussion of the FD of motor bearings

(1) Data description

The CWRU bearing data center designs a dataset for motor bearing failure analysis that covers multiple failure types and varying degrees of severity. The CWRU dataset consists of four main failure categories: health, outer ring fault, inner ring fault, and ball fault. There are three levels of damage size for each type of failure state, ranging from 7 mil to 21 mil. The motor maintained a specific operating speed during the experiment and operated under loads ranging from 0 to 3 hp, resulting in a total of 10 different state categories, as shown in table 2. We implement the data enhancement using the sliding window technique, which employs a window length of 1024 points to ensure that each sample covers at least one complete bearing rotation cycle and a window step size of 200 points. In fact, the CWRU dataset has 48 000 Hz and 12 000 Hz sampling rates at the drive end. At the sampling rate of 12 000 Hz, a total of 6000 data samples are generated by the sliding window technique to ensure the accuracy and robustness of the evaluation. All the samples are equally divided into six subsets; four subsets are selected as the training set, and the remaining two subsets are used as the validation set and the test set, respectively.

(2) Experimental analysis of noise environment

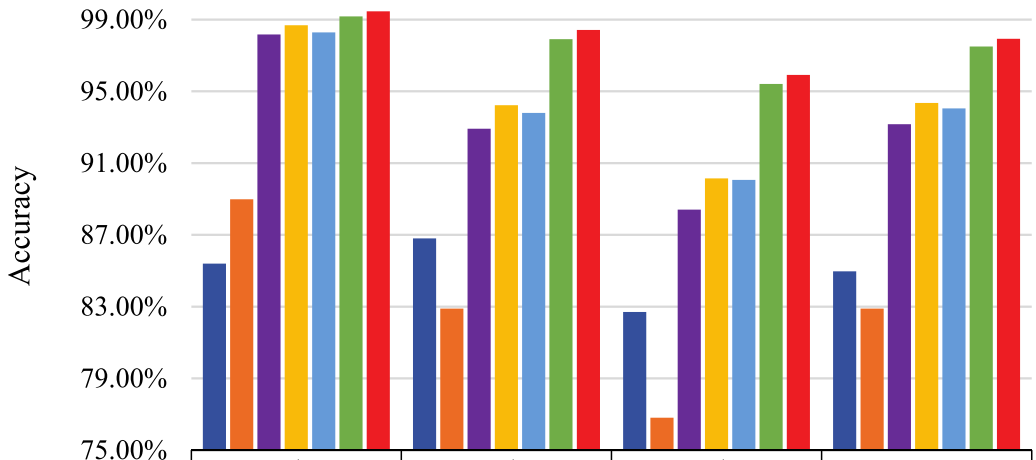
Figure 6 shows the test results of various FD methods in different noise environments. The dB in the constant coordinate represents different levels of the noise environment, and lower negative values (e.g. −4 dB, −6 dB) indicate an increase in the noise level and higher noise interference. It can be seen that the accuracy of almost all methods decreases as the noise level increases, particularly from 0 dB to −6 dB. This outcome is expected because more substantial noise interferes with the model's processing fine fault features. The proposed HF-MSCN maintains the highest accuracy at all noise levels, which indicates that it is very robust to noise. And the average accuracy of HF-MSCN is 97.93%, which is the highest among all listed methods, which means that it can maintain stable performance in different noise environments. In addition, by observing the changes in accuracy under different noise levels, HF-MSCN has the slightest decrease from 0 dB to −6 dB, from 99.46% to 95.91%, which indicates its better robustness. Overall, benefiting from the noise reduction performance of the HFCB designed in section 2 and the multi-scale feature learning capability of MSCB, the proposed HF-MSCN shows strong stability and high accuracy against different levels of noise interference.

To evaluate the performance of the proposed HF-MSCN under different noise environments (0 dB, −4 dB, −6 dB), we use confusion matrices to show the recognition accuracy across ten fault categories. In the 0 dB noise environment figure 7(a), the confusion matrix shows that HF-MSCN achieves near-perfect classification accuracy for all fault categories, with diagonal values mostly at 1.0, indicating that almost all samples are correctly classified. In the −4 dB noise environment figure 7(b), high accuracy is maintained for most categories, such as classes 0, 3, 4, 5, and 9, with diagonal values at 1.0. In the −6 dB noise environment figure 7(c), the confusion matrix shows reduced accuracy only in class 8 (0.80) and class 9 (0.96), while overall results still reflect high classification accuracy. The excellent performance of HF-MSCN in noisy environments is attributed to the inclusion of the HFCB. The HFCB uses an attention mechanism to identify important information and noise in the high-frequency data, helping the model to focus on more useful high-frequency features and enhancing its noise robustness. In summary, HF-MSCN demonstrates strong noise resistance.

Case 2: Discussion of compound FD of bearings

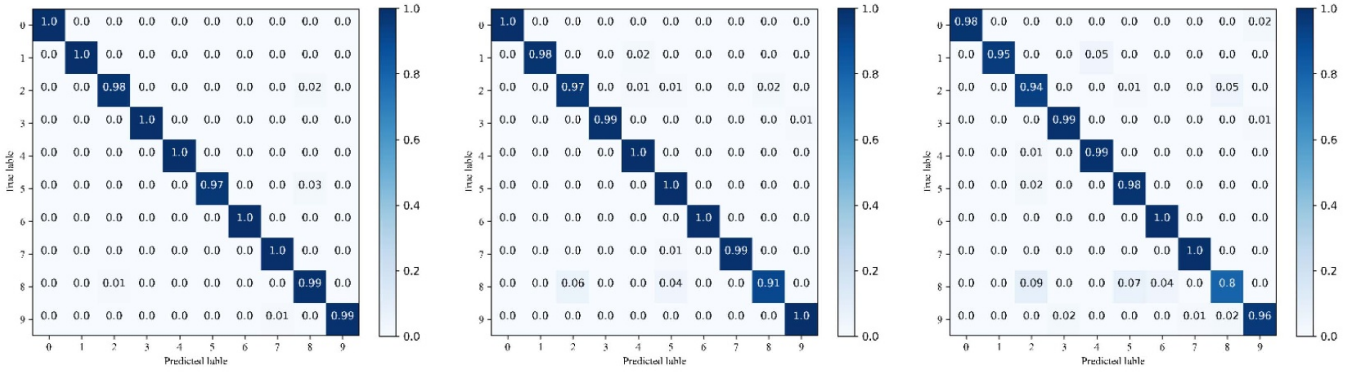
(1) Data description

We employ the machinery fault simulator (MFS) test rig manufactured by Spectrum Quest Incorporated to simulate both single-type and compound faults in bearings. The MFS test bench is shown in figure 8, including the drive end, accelerometer, and signal collector. The motor is connected to the drive shaft through a flexible coupling. We use laser technology to simulate ball faults, inner ring faults, outer ring faults and composite inner and outer ring faults on the bearings. Additionally, we collect vibration signals at a sampling frequency of 15.6 kHz. The fault locations and detailed data are presented in figure 9 and table 3, respectively. The bearings rotate at a speed of 1130 r min^{-1}, and the vibration data is transferred to a computer via a single-channel data cable and USB interface. To overcome the limitations of data collected by the sensor collection, a window length of 1024 points is used to gather data samples, with 200 samples for each



	0dB	-4dB	-6dB	Average
MCAMDN	85.40%	86.80%	82.70%	84.97%
RESCNN	88.98%	82.89%	76.81%	82.89%
MSDARN	98.17%	92.91%	88.41%	93.16%
MBSDCN	98.68%	94.23%	90.15%	94.35%
MA1DCNN	98.28%	93.80%	90.06%	94.05%
GTFE-Net	99.17%	97.91%	95.41%	97.50%
HF-MSCN	99.46%	98.42%	95.91%	97.93%

Figure 6. Diagnostic results of different methods on CWRU dataset in noisy environment.



(a) 0 dB (b) -4 dB (c) -6 dB

Figure 7. Diagnostic results on the CWRU dataset under different noise levels: (a) 0 dB, (b) −4 dB, and (c) −6 dB.

fault type, resulting in a total of 800 samples. Out of the total samples, 500 are allocated for training, 150 for verification, and 150 for testing.

(2) Experimental analysis of composite FD in noisy environments

Compound faults of rolling bearings are often seen in practical applications, and their diagnosis is more challenging compared to single faults. To assess the practical significance of the research in this paper for real industrial applications, we set three different noise environments on compound faults data. Figure 10 displays the FD results of different methods. The proposed HF-MSCN shows significant advantages throughout all levels of noise, achieving accuracies varying from 98.75% for 0 dB noise to 90.00% for −6 dB. These results indicate that HF-MSCN has strong noise robustness and complex FD capability. Furthermore, the average accuracy of HF-MSCN is 94.30%, which is much higher than other methods, verifying its excellent performance in dealing with compound faults in bearings. In addition, it is worth pointing out that HF-MSCN can achieve high accuracy in a small sample scale (500 training samples), which is particularly valuable for diagnostic situations where the cost of sample acquisition is high.

We analyze the confusion matrices for HF-MSCN in bearing compound FD under different noise environments (0 dB, −4 dB, −6 dB). Each confusion matrix shows the

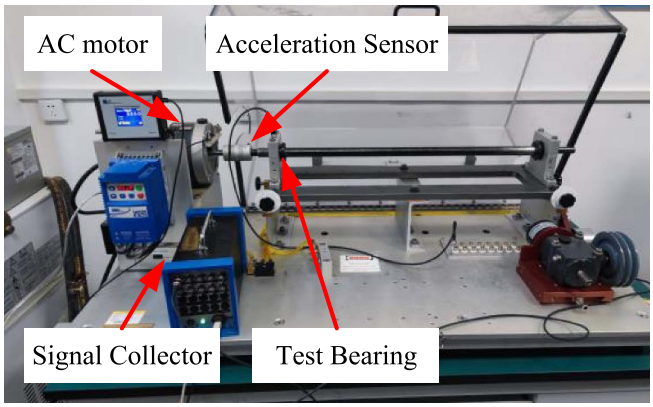


Figure 8. The MFS test bench.

Ball fault | Outer race fault

Inner race fault | Compound fault (Inner race and outer race)

Figure 9. Different types of bearing failures.

classification accuracy across four fault categories. In the 0 dB noise environment figure 11(a), the classification accuracy for classes 0 and 3 reaches 100%. There are slight misclassifications in class 1 (0.03%) and class 2 (0.02%). In the −4 dB noise environment figure 11(b), most diagonal values remain high, but the misclassification rates for some categories increase. For example, class 2 has a significant misclassification rate of 0.21. In the −6 dB noise environment figure 11(c), there are notable misclassifications in class 2 (0.14%) and class 3 (0.22%). These results demonstrate that HF-MSCN, by incorporating MSCB, exhibits excellent performance in bearing compound FD under different noise conditions. MSCB enhances inter-layer interaction and multi-scale feature capture, playing a crucial role in maintaining HF-MSCN's performance in high-noise and compound fault conditions, thus covering a wider range of FD scenarios.

Table 3. Details of the compound fault dataset.

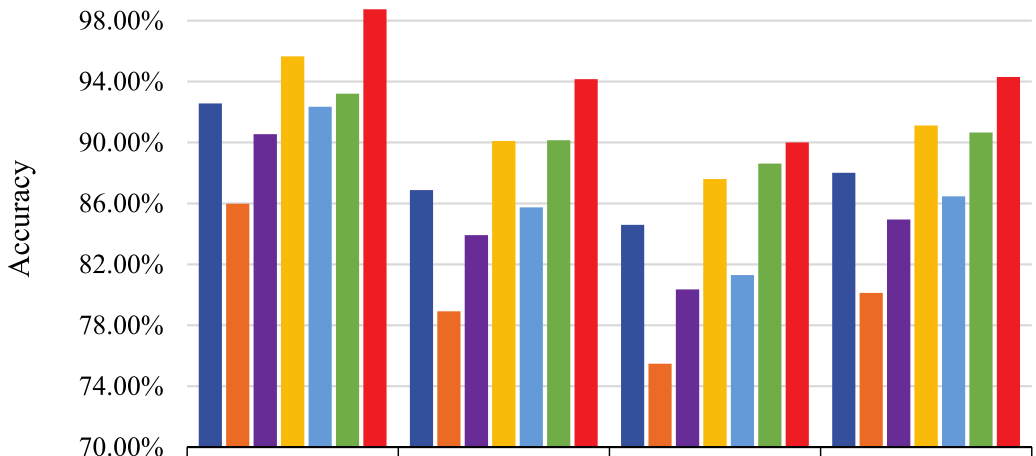
Fault location	Fault size (mil)	Label
Ball	6	0
Inner race	24	1
Outer race	6	2
Compound (Inner race and Outer race)	Inner 6&Outer 9	3

3.4. Ablation experiments

(1) Impact of the proposed HFCB and MSCB

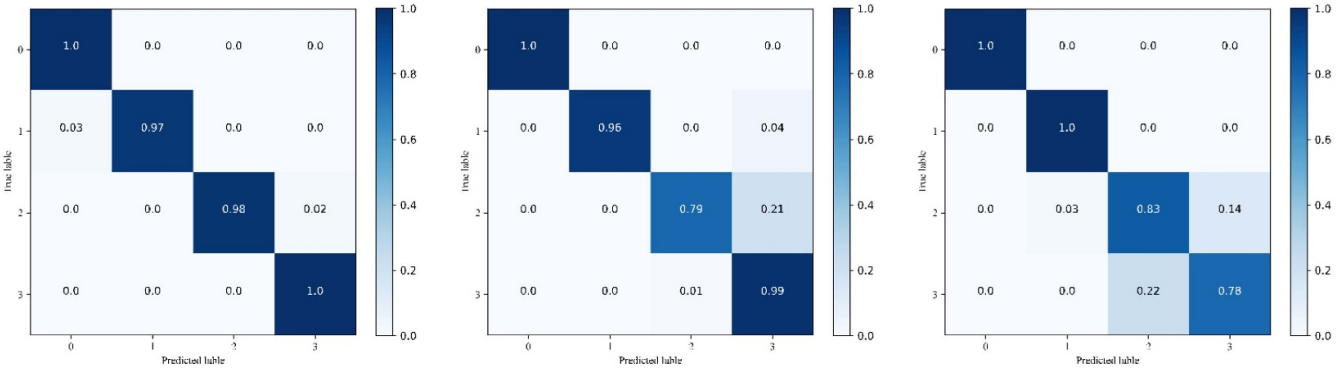
This paper focuses on HFCB for filtering high-frequency noise and MSCB for capturing fault pulse information. Four methods are designed to verify their contribution to the model performance: a baseline excluding HFCB and MSCB (Method 1), a method containing only MSCB (Method 2), a method containing only HFCB (Method 3), and a method with HFCB and MSCB (Method 4). The experiments are conducted on the CWRU dataset in a −6 dB noise environment, and the results are shown in table 4. Method 1 serves as the baseline and has the lowest of all metrics. When MSCB is included, all performance metrics are significantly improved compared to the baseline. The accuracy improves from 86.08% to 89.25%, while the F1 score increases from 86.07% to 89.6%. This enhancement demonstrates the effectiveness of the HFCB module in boosting the performance of the model. The performance improvement is even more significant when HFCB is used independently, with the accuracy increasing to 92.41% and the F1 score rising to 91.99%. This substantial improvement underscores the pivotal role of HFCB in the model, likely due to its enhanced noise immunity of the model. Furthermore, when HFCB is combined with MSCB, the model performance is optimized with both accuracy and F1 score approaching or exceeding 95%. This indicates that the combination of HFCB and MSCB can have a synergistic impact, leading to a significant improvement in the performance of FD.

In order to better compare the advantages and disadvantages of different methods in ablation experiments, the t-SNE technique is used to visualize the classification results. In figure 12, different colors represent different fault classes, each point represents a sample in the data, and clustered areas of color indicate that these samples are similar in some features. Method 1 does not provide a clear separation between the different fault classes, shows more mixing and overlapping, because it does not have a specialized mechanism to deal with noise and does not enhance the interaction of multiscale features. Since MSCB helps to capture fault features at different scales, method 2 shows better clustering in some fault classes. However, method 2 still has a lot of sample overlap, which shows that its ability is limited for noise suppression. Method 3 shows less overlap of samples from different fault classes, but samples from the same fault classes are not well clustered. The reason for this is that method 3 lacks MSCB and is not capable of interacting between multiple layers of features, which limits its ability to capture subtle fault features.



	0dB	-4dB	-6dB	Average
MCAMDN	92.56%	86.88%	84.60%	88.01%
RESCNN	85.98%	78.92%	75.48%	80.13%
MSDARN	90.55%	83.92%	80.35%	84.94%
MBSDCN	95.66%	90.10%	87.60%	91.12%
MA1DCNN	92.35%	85.74%	81.30%	86.46%
GTFE-Net	93.20%	90.15%	88.62%	90.66%
HF-MSCN	98.75%	94.16%	90.00%	94.30%

Figure 10. Diagnostic results of different methods on the compound fault dataset in noisy environment.



(a) 0 dB (b) -4 dB (c) -6 dB

Figure 11. Diagnostic results on the compound fault dataset under different noise levels: (a) 0 dB, (b) −4 dB, and (c) −6 dB.

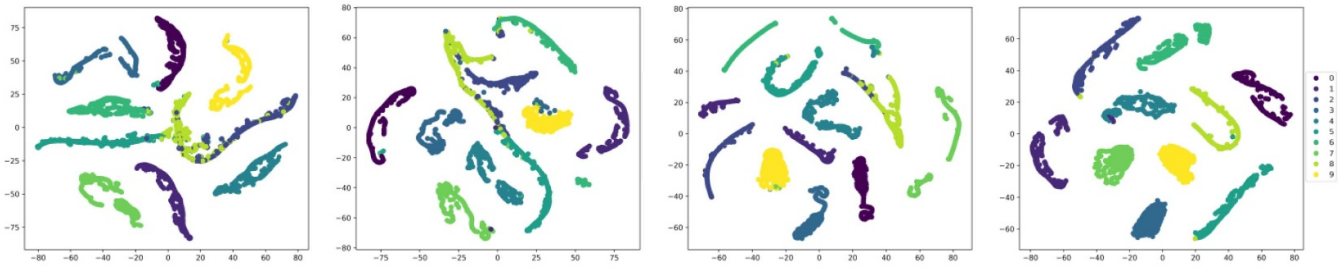
Table 4. Results of ablation experiments with different methods.

Method	HFCB	MSCB	Accuracy	Precision	Recall	F1-score
1	—	—	86.08	87.41	86.62	86.07
2	—	√	89.25	90.70	89.56	89.60
3	√	—	92.41	94.00	92.08	91.99
4	√	√	95.91	95.82	95.90	95.78

Compared to the above three methods, method 4 shows the tightest and clearest clustering effect, including intra-class aggregation, and inter-class separation. This indicates that the method combining HFCB and MSCB has the best FD performance and validates the effectiveness of the innovative part of the manuscript.

(2) The impact of activation ratio in HFCB on model performance

In the HFCB introduced in section 2, a fixed ratio is employed to deactivate certain features selectively; this approach is designed to enable the model to learn both low-frequency and high-frequency features from the vibration signal in a balanced



(a) Method 1 (b) Method 2 (c) Method 3 (d) Method 4

Figure 12. Visualization results of ablation experiments with different methods: (a) Method 1, (b) Method 2, (c) Method 3, and (d) Method 4.

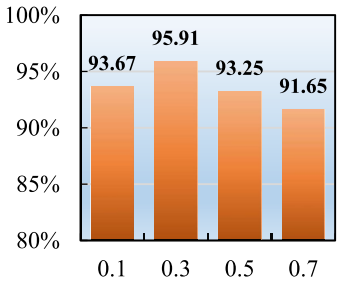


Figure 13. Results of the effect of different ratios on model performance.

manner. In order to determine the optimal ratio value, we set four ratio values in the experiment with −6 dB noise environment, and the results are shown in figure 13. It is clear that when the ratio value is 0.3, the diagnostic accuracy of the model is 95.91%, and the results of the other three ratio values are not higher than 94%, so in this article, we set the ratio value parameter in HFCB to 0.3.

(3) The impact of subset number and kernel size in MSCB on model performance

In the proposed MSCB, we divide the input features into s subsets and feed them into multi-scale convolution to achieve multi-level feature interaction, which involves the value s and the size parameter of the multi-scale convolution kernel. To ascertain these parameters, we design multiple sets of comparison experiments to achieve the best FD performance. The comparison results are shown in figure 14, where the value of s should conform to a power of 2 to ensure that the number of channels can be equally divided. The x-axis represents the values of s, and the number of channels should be divisible by specific values ($s = 2, 4, 8$). After the input features are divided, each subset corresponds to three different sets of kernel sizes, e.g. for $s = 2$, the kernel sizes are $(3 \times 1/5 \times 1)$, $(5 \times 1/7 \times 1)$, $(7 \times 1/9 \times 1)$, and for $s = 4$, the kernel sizes are $(3 \times 1/5 \times 1/7 \times 1/9 \times 1)$, $(5 \times 1/7 \times 1/9 \times 1/11 \times 1)$, $(7 \times 1/9 \times 1/11 \times 1/13 \times 1)$. Three sets of kernel sizes are represented by $k3$, $k2$, and $k1$ on the y-axis. The z-axis represents the accuracy of FD. It can be seen that among the nine sets of experimental results, the highest FD accuracy is attained with kernel size $k3$ at $s = 4$. Although the model has good accuracy

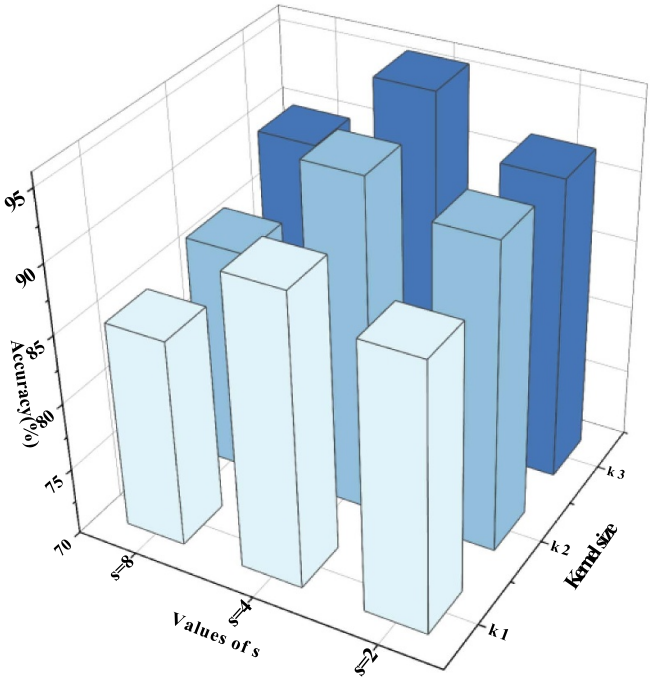
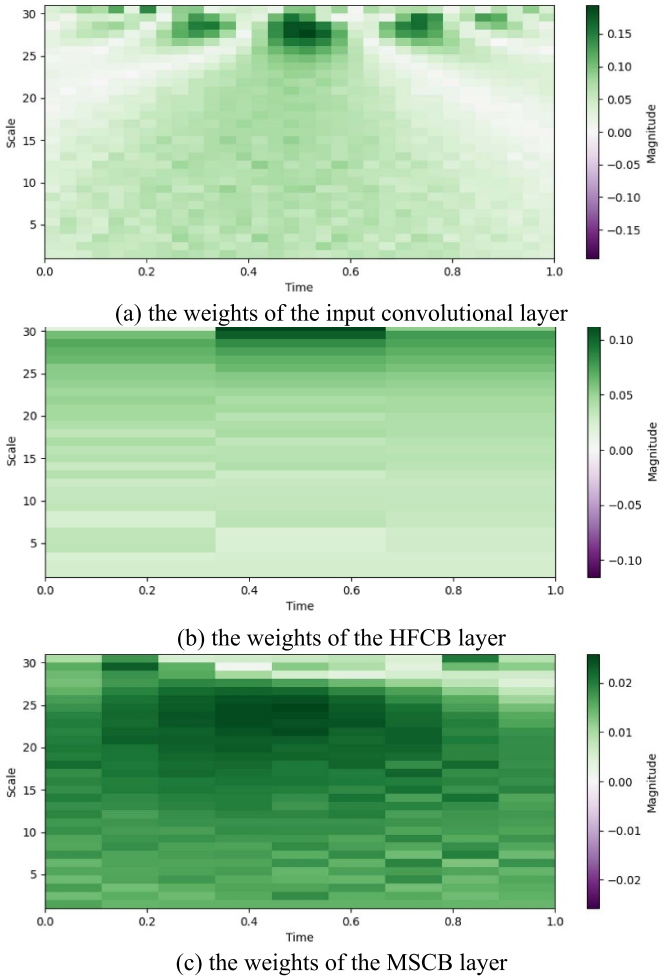


Figure 14. Comparison results for different subset numbers and core sizes.

with kernel size $k2$, the large size of the convolutional kernel leads to a higher number of parameters given to the model.

3.5. Analysis of the interpretability of HF-MSCN

This section examines the interpretability of HF-MSCN in FD. In the field of signal processing, we introduce the time–frequency transform technique to analyze the variation of weights in HF-MSCN over time and scale. The detailed procedure is as follows: first, the weights learned at each layer of the HF-MSCN model are stored, and the average of these weights is computed so that each convolution kernel is represented as a single time-domain signal. Then, the CWT is used to analyze the changes of weights at different frequencies, which is used to reveal non-periodic signal change information. Finally, the output of the CWT is visualized, where each row of the graph corresponds to a scale, each column corresponds to a time point, and the values in the matrix indicate



(a) the weights of the input convolutional layer

(b) the weights of the HFCB layer

(c) the weights of the MSCB layer

Figure 15. Weight changes over time and frequency in the network.

the signal strength of that scale at that moment. In the above way, each figure shows the weights learned at each layer over time and scale, thus helping researchers understand how the neural network model captures different frequency features in the input data. Implementing CWT to explain the model feature learning process is feasible due to the following reasons:

(1) CWT provides more detailed time–frequency resolution than other methods, such as short-time fourier transform, which is particularly suitable for revealing signals with transient features.
(2) The ability of the CWT to recognize rapid changes that occur on small time scales is helpful for understanding how the convolution kernel responds to specific features of the input signal.
(3) In neural network models, weights directly determine how the network responds to specific patterns and features in the input data. A time–frequency plot displays the intensity of activity for each weight at various frequencies. This analysis can provide an understanding of how the feature layer reacts to high or low-frequency features in the data.

In figure 15, each figure shows how weight changes over time and scale (frequency). The horizontal axis represents time, the vertical axis represents scale, and the color shades indicate the magnitude of the waveform amplitude. A decrease in scale corresponds to an increase in frequency, while an increase in scale corresponds to a decrease in frequency. Figure 15(a) illustrates the weights of the input convolutional layer, figure 13(b) indicates the output weights of the HFCB layer, and figure 15(c) indicates the output weights of the MSCB layer. Comparing figures 15(a) and (b), it is evident that there is a decrease in activity within the high-frequency region, which shows that HFCB can effectively respond to high-frequency information. Upon comparing figures 15(b) with (c), a more uniform and concentrated frequency response can be seen, suggesting that the MSCB layer effectively extracts multi-level features while suppressing unnecessary frequency components. In summary, we analyze variations in scale (frequency) and time responses in different plots and show how each layer processes and changes the transmitted signals. Furthermore, observing the effect of each layer highlights the effectiveness of the proposed HFCB layer and MSCB layer.

4. Conclusion

In this study, a novel HF-MSCN is developed to face the key challenges in machinery FD, i.e. strong noise interference and high similarity between fault classes. The framework of HF-MSCN combines HFCB and MSCB, thus helping the network to identify fault features and suppress noise in vibration signals. Extensive experiments on CWRU and self-constructed composite bearing datasets demonstrate that HF-MSCN exhibits excellent FD performance in high-noise environments, significantly outperforming six mainstream comparison algorithms. More importantly, we conduct an in-depth analysis of the feature learning mechanism in HF-MSCN and clarify how the feature layer promotes the identification and separation of fault features at different levels. This analysis verifies the effectiveness of the innovative components and enhances the interpretability of the deep learning model in the diagnostic process. In the future, we plan to continue to explore the theoretical foundation and practical applications of HF-MSCN intensely, significantly enhancing the interpretability of the network and applying this technology in a broader range of machinery FD scenarios.

Data availability statement

The data cannot be made publicly available upon publication because no suitable repository exists for hosting data in this field of study. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

The present work was funded by the National Key Research and Development Plan of China (Grant No.2020YFB1713600), the National Natural Science Foundation of China (Grant Nos. 62163023, 62063020), the Youth Science and Technology Fund of Gansu Province

(22JR5RA808), Long yuan Young Innovation Talent Team Program in Gansu Province (310100296012).

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical statement

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

ORCID iDs

Alaeldden Abduelhadi https://orcid.org/0009-0009-1407-2199
Haopeng Liang https://orcid.org/0000-0002-4342-9005
Jie Cao https://orcid.org/0009-0009-0442-8104
Peng Chen https://orcid.org/0000-0002-9928-5346

References

[1] Zhao K, Jiang H, Wang K and Pei Z 2021 Joint distribution adaptation network with adversarial learning for rolling bearing fault diagnosis *Knowl.-Based Syst.* **222** 106974
[2] Wei Z, He D, Jin Z, Shan S, Zou X, Miao J and Liu C 2023 Intelligent fault diagnosis and health stage division of bearing based on tensor clustering and feature space denoising *Appl. Intell.* **53** 24671–88
[3] Li Y, Jia Z, Liu Z, Shao H, Zhao W, Liu Z and Wang B 2024 Interpretable intelligent fault diagnosis strategy for fixed-wing UAV elevator fault diagnosis based on improved cross entropy loss *Meas. Sci. Technol.* **35** 076110
[4] Peng J, Shao H, Xiao Y, Cai B and Liu B 2024 Industrial surface defect detection and localization using multi-scale information focusing and enhancement GANomaly *Expert Syst. Appl.* **238** 122361
[5] Wei Z, He D, Jin Z, Liu B, Shan S, Chen Y and Miao J 2023 Density-based affinity propagation tensor clustering for intelligent fault diagnosis of train bogie bearing *IEEE Trans. Intell. Transp. Syst.* **24** 6053–64
[6] Zhang S, Zhang S, Wang B and Habetler T G 2020 Deep learning algorithms for bearing fault diagnostics—a comprehensive review *IEEE Access* **8** 29857–81
[7] Hou S, Lian A and Chu Y 2023 Bearing fault diagnosis method using the joint feature extraction of transformer and ResNet *Meas. Sci. Technol.* **34** 075108
[8] Bhatt D, Patel C, Talsania H, Patel J, Vaghela R, Pandya S, Modi K and Ghayvat H 2021 CNN variants for computer vision: history, architecture, application, challenges and future scope *Electronics* **10** 2470
[9] Chen X, Zhang B and Gao D 2021 Bearing fault diagnosis base on multi-scale CNN and LSTM model *J. Intell. Manuf.* **32** 971–87
[10] Gao Q, Huang T, Zhao K, Shao H and Jin B 2024 Multi-source weighted source-free domain transfer method for rotating machinery fault diagnosis *Expert Syst. Appl.* **237** 121585
[11] Zhang W, Peng G, Li C, Chen Y and Zhang Z 2017 A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals *Sensors* **17** 425
[12] Chen Z, Gryllias K and Li W 2019 Mechanical fault diagnosis using convolutional neural networks and extreme learning machine *Mech. Syst. Signal Process.* **133** 106272
[13] Liu Z, Wang H, Liu J, Qin Y and Peng D 2020 Multitask learning based on lightweight 1DCNN for fault diagnosis of wheelset bearings *IEEE Trans. Instrum. Meas.* **70** 1–11
[14] Wei Z, Wang Y, He S and Bao J 2017 A novel intelligent method for bearing fault diagnosis based on affinity propagation clustering and adaptive feature selection *Knowl.-Based Syst.* **116** 1–12
[15] Wang J, Shao H, Yan S and Liu B 2023 C-ECAFormer: a new lightweight fault diagnosis framework towards heavy noise and small samples *Eng. Appl. Artif. Intell.* **126** 107031
[16] Lv H, Chen J, Pan T, Zhang T, Feng Y and Liu S 2022 Attention mechanism in intelligent fault diagnosis of machinery: a review of technique and application *Measurement* **199** 111594
[17] Pei X, Zheng X and Wu J 2021 Rotating machinery fault diagnosis through a transformer convolution network subjected to transfer learning *IEEE Trans. Instrum. Meas.* **70** 1–11
[18] Zhao B, Zhang X, Zhan Z and Wu Q 2021 Deep multi-scale adversarial network with attention: a novel domain adaptation method for intelligent fault diagnosis *J. Manuf. Syst.* **59** 565–76
[19] Xu Y, Yan X, Feng K, Sheng X, Sun B and Liu Z 2022 Attention-based multiscale denoising residual convolutional neural networks for fault diagnosis of rotating machinery *Reliab. Eng. Syst. Saf.* **226** 108714
[20] Wang X, Zhang H and Du Z 2023 Multi-scale noise reduction attention network for aero-engine bearing fault diagnosis *IEEE Trans. Instrum. Meas.* **72** 3513810
[21] Zhang W, Li C, Peng G, Chen Y and Zhang Z 2018 A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load *Mech. Syst. Signal Process.* **100** 439–53
[22] Pancaldi F, Dibiase L and Cocconcelli M 2023 Impact of noise model on the performance of algorithms for fault diagnosis in rolling bearings *Mech. Syst. Signal Process.* **188** 109975
[23] Li A, Yao D, Yang J, Chang M and Zhou T 2024 Bearing diagnosis using an anti-noise neural network based on selectable branch multi-scale modules and attention mechanisms *IEEE Sens. J.* **24** 5830–40
[24] Zhang W, Li X and Ding Q 2019 Deep residual learning-based fault diagnosis method for rotating machinery *ISA Trans.* **95** 295–305
[25] Liang H, Cao J and Zhao X 2022 Multi-scale dynamic adaptive residual network for fault diagnosis *Measurement* **188** 110397
[26] Liang H, Cao J and Zhao X 2023 Multibranch and multiscale dynamic convolutional network for small sample fault diagnosis of rotating machinery *IEEE Sens. J.* **23** 8973–88
[27] Wang H, Liu Z, Peng D and Qin Y 2019 Understanding and learning discriminant features based on multiattention 1DCNN for wheelset bearing fault diagnosis *IEEE Trans. Ind. Inform.* **16** 5735–45
[28] Jia L, Chow T W S and Yuan Y 2023 GTFE-Net: a gramian time frequency enhancement CNN for bearing fault diagnosis *Eng. Appl. Artif. Intell.* **119** 105794

Journal of the Brazilian Society of Mechanical Sciences and Engineering (2025) 47:554
https://doi.org/10.1007/s40430-025-05880-8

TECHNICAL PAPER

Efficient fault diagnosis method based on dynamic recalibration mechanism with improved inception network for rolling bearings under limited samples

Alaeldden Abduelhadi[1] · Jie Cao[1] · Haopeng Liang[2] · Shanqin Yuan[1]

Received: 17 February 2025 / Accepted: 28 July 2025

Abstract
Small sample sizes and robust noise continue to limit the accuracy of deep learning (DL)–based bearing fault diagnosis (FD). To overcome these challenges, we propose two novel modules: Cross-layer Calibration Inception (CLC-Inception)—which combines multiscale 1D convolutions with a Squeeze-and-Excitation attention block—and Channel-Spatial Dynamic Recalibration Block (CSDRB)—comprising a Channel Recalibration Block (CRB) and a Spatial Recalibration Block (SRB) that adaptively reweights spatial and temporal features. In our method, raw vibration signals first pass through a wide convolutional layer for enhanced noise reduction; next, the Cross-layer Calibration Inception extracts deep multiscale features, which are then modified using residual convolutional layers; finally, the CSDRB prioritizes critical diagnostic information. The effectiveness of the proposed method is validated on two publicly available bearing datasets: the Case Western Reserve University (CWRU) dataset and the Lanzhou University of Technology (LUT) dataset. Comparative analyses and ablation experiments confirm that the proposed approach achieves 99.34% accuracy under small sample conditions and an average of 96.16% accuracy under strong noise, surpassing four advanced deep learning methods. These results highlight the robustness and potential applicability of the developed method for real-world bearing fault diagnosis scenarios.

Keywords Rolling bearing · Channel-spatial dynamic recalibration · Inception network · Residual CNN · Small sample fault diagnosis

1 Introduction

Rolling bearings are crucial components extensively utilized in various mechanical systems across modern industries, significantly influencing machinery reliability and operational safety [1]. Statistics indicate that over 40% of failures in rotating machinery occur due to bearing failures, primarily because rolling bearings are susceptible to damage and degradation under harsh operating conditions [2]. If these failures are not quickly recognized, they could result in the shutdown of mechanical operations, potentially leading to significant economic losses and even casualties [3, 4]. Therefore, effective Fault Diagnosis (FD) of rolling bearings is critically important to ensure industrial machinery's continued reliability, productivity, and safety. However, conventional diagnostic approaches often face difficulties under complex operational scenarios, which highlights the need for advanced diagnostic techniques capable of rapidly and accurately identifying bearing faults to avoid downtime and minimize maintenance costs.

Traditional signal processing methods such as Ensemble Empirical Mode Decomposition (EMD), Backpropagation Neural Networks (BPNN) [5], and Variable Mode Decomposition (VMD) [6] exhibit limitations in accuracy, efficiency and dependence on expert knowledge in the context of the dynamic and complicated industrial system. In recent years, the rapid development of Deep Learning (DL) has led to increased attention from researchers toward intelligent FD methods, which can lower a significant amount of labor by

Technical Editor: Jarir Mahfoud.

✉ Alaeldden Abduelhadi
alaeldden9090@gmail.com

[1] School of Electrical and Information Engineering, Lanzhou University of Technology, 730050 Lanzhou, People's Republic of China

[2] School of Computer and Communication, Lanzhou University of Technology, 730050 Lanzhou, People's Republic of China

Published online: 02 September 2025

automatically extracting features from raw vibration data and learning valuable information to help identify fault types [7]. Several DL models have been applied extensively for FD, particularly Convolutional Neural Networks (CNN) [8], Recurrent Neural Networks (RNN) [9], Generative Adversarial Networks (GAN) [10], and Deep Belief Networks (DBN) [11]. CNNs, among other DL algorithms, have clearly shifted the scene of intelligent FD owing to their powerful capture of feature capabilities and capacity to adapt to complicated signal processing tasks. For instance, a deep neural model was developed by Ye and Yu [12], which employs multiple branches with varying kernel sizes to extract multiscale features from vibration data. The method involves an attentive selection of kernels that eliminates unnecessary feature maps while highlighting helpful ones. In their cross-domain intelligent fault diagnosis research, Tian and Gu [12] proposed an approach combining feature transfer with an enhanced inception ResNet to boost the poor accuracy of rolling bearings in varying operating conditions. Sun et al. [13] proposed an approach called MA-MS1DCNN for fault diagnosis in hydraulic systems, using a multiscale one-dimensional CNN that incorporates a multi-attention mechanism with the aim of extracting deep features and strengthening the representation of related features from time-series data. These methods relied on substantial data support, and they obtained good diagnostic results without requiring professional expertise. The performance of DL models in fault identification diminishes with limited data and unbalanced data distribution. Acquiring sufficient fault samples is frequently impractical due to the complex operational conditions of rolling bearings; for example, bearings are vital components of turbomachinery that can deteriorate over months or years, so gathering enough data would take several years. Moreover, In the functioning of bearings in high-speed trains, manual data labeling is challenging due to the rapid fluctuations in load conditions [14]. Therefore, developing an efficient FD model can be challenging when sample numbers are limited.

A small sample indicates a scenario where there is an insufficient quantity of data for training the network model. DL models frequently require extensive datasets to achieve high accuracy, and the precise identification of complex fault patterns from normal operating conditions poses major obstacles when sample sizes are limited for the development of robust diagnostic models. The main goal of small sample fault diagnosis is to establish an effective method using a limited number of samples. For example, Yang et al. [15] proposed a fusion diagnosis model called CGAN-2-D-CNN for bearing fault diagnosis with small samples, using a conditional generative adversarial network to create new samples with data distributions similar to real samples and a two-dimensional CNN to extract features and classify bearing fault types. Bai et al. [16] developed a strategy utilizing multi-channel CNN and multiscale clipping fusion for data augmentation to amplify fault signals, which are subsequently transformed into time–frequency representations via the short-time Fourier transform. These studies address the issue of limited samples through data augmentation techniques, which involve increasing sample size via data cropping and generation [17]. The quantity and quality of the original data typically constrain the efficacy of augmenting data, so when the amount of original data is limited, the overlapping sampling approach may produce redundant samples, as well as when the quality of the original data is insufficient, the network model struggles to create new data samples that match the original dataset. Some studies have applied transfer learning and few-shot learning methods to deal with the difficulties of small sample FD. Transfer learning relies on information learned from source domain samples to improve diagnostic tasks in target domain samples [18]. Few-shot learning involves predicting new classes by acquiring knowledge from provided labeled classes [19]. For instance, Dong et al. [20] introduced a framework that integrates a dynamic bearings model with a transfer learning approach to produce comprehensive simulation data, where CNN and parameters of transfer learning are used to apply the diagnosis information obtained from the simulation data to actual scenarios. Su et al. [21] presented an approach consisting of two phases: data reconstruction and meta-learning. During the data reconstruction phase, noise reduction techniques are applied to extract valuable information concealed within the raw data, and a recurrent meta-learning that incorporates one-shot learning is used for boosting the FD model during the meta-learning phase. Wu et al. [22] proposed a few-shot transfer learning approach utilizing a unified 1D CNN, which addresses data dependency, transferability, and task plasticity across different methods in the few-shot context. Although the previously mentioned deep learning-based small sample FD methods have shown promising results, several obstacles still prevent their adoption. The reliance on a large number of labeled training samples from the source domains, which can be expensive and time-consuming to gather, is a major drawback of the majority of methods for transfer learning. Furthermore, few-shot learning methods frequently have complicated structures that lead to slower training optimization and diminished efficiency.

The latest advancements in studies have incorporated attention mechanisms into deep learning models, proving effective in capturing complex dependencies within sequences [23]. These mechanisms enhance the model's capacity by focusing selectively on different segments of the input sequence to generate each corresponding output segment, allowing the models to prioritize specific areas of the input data that are most relevant for producing accurate outputs, thereby improving the efficiency and effectiveness of the FD task. For example, Huang et al. [24] introduced

a multiscale CNN with a channel attention mechanism that uses maximum and average pooling layers to extract multiscale information from bearing signals, enhancing feature learning by adaptively scoring and assigning weights to the extracted features. Mekruksavanich et al. [25] proposed a deep residual network based on channel attention for recognizing complex activities using wrist-worn wearable sensors. Han et al. [26] introduced an architecture that integrates dynamic dropout, residual connections, and the multihead attention mechanism to eliminate irrelevant information and capture long-term interdependence in the data. Some studies primarily tend to the method of transforming 1-D input signals into 2-D data formats for enhanced processing, such as Liu et al. [27] developed a multiscale residual anti-noise network with a dynamic recalibration mechanism, which utilizes a short-time Fourier transform to produce a 2-D time–frequency image from 1-D vibration data. These methods above have gained satisfactory outcomes in fault identification with a few samples, either by altering the network structure or applying data augmentation techniques. However, converting 1-D signals into 2-D images presents several issues; temporal information inherent in 1-D vibration time series might be lost during the conversion process. Moreover, 2-D convolutions demand more computing power and training time than 1-D convolutions, which may not be viable for efficient industrial diagnostic operations. Additionally, the reliance on specific parameters within data-driven FD models increases the risk of overfitting, potentially compromising the effectiveness of the model.

In summary, DL methods have achieved significant advancements in bearing FD tasks; nonetheless, they continue to encounter the following challenges:

1) Data scarcity and noise resilience: Existing models necessitate substantial amounts of labeled data, which are frequently expensive and labor-intensive to acquire, and exhibit considerable performance decline when utilized in small sample or high-noise conditions, thereby constraining their practical applicability in real-world scenarios.
2) Model robustness: Transforming one-dimensional vibration signals into two-dimensional formats may result in the loss of crucial temporal data and increase computational demands. Meanwhile, the shift toward deeper and more parameter-intensive CNN architectures elevates the likelihood of overfitting and training inconsistencies, thereby hindering effective and dependable implementation in industrial settings.

To address these two limitations, this paper proposes a lightweight FD method based on Cross-layer Calibration Inception (CLC-Inception) and Channel-Spatial Dynamic Recalibration Block (CSDRB). The proposed method processes one-dimensional vibration signals through CLC-Inception enhanced with Squeeze-and-Excitation (SE) attention, which improves the network's adaptability and robustness through multiscale information representation achieved by employing diverse convolutional kernel sizes. Additionally, the CSDRB facilitates adaptive adjustment of the network's weights based on spatial mapping and channel-wise feature information. These innovations significantly improve the efficiency of FD, even in scenarios with limited training data and strong noise interference. The primary contributions of this research can be summarized as follows:

1) An efficient and intuitive feature extraction framework is developed, combining CSDRB with CLC-Inception enhanced by SE attention and residual CNN structures. This allows the network to effectively identify critical features and extract comprehensive diagnostic information from raw signals.
2) The CSDRB is strategically designed to capture both temporal and spatial characteristics inherent in vibration signals, thus significantly improving diagnostic accuracy across diverse operational conditions. Moreover, incorporating a wide convolutional layer and a Global Average Pooling (GAP) layer effectively drops the number of parameters, facilitating faster training.
3) Extensive experimental validation on the widely used CWRU bearing dataset confirms the model's exceptional performance, particularly in scenarios characterized by small training sample sizes. Further validation performed on a laboratory-designed LUT-bearing test rig also demonstrates the model's robustness and reliability in accurately diagnosing both composite and single faults, even under severely limited data conditions.

This paper is structured as follows: Sect. 2 presents the theoretical background, providing a foundational overview for understanding the subsequent methodologies. Section 3 describes the proposed FD method, which includes CLC-Inception and CSDRB, detailing the overall framework developed for this study. Section 4 discusses the experimental analysis, giving insights into the discussion and evaluating the effectiveness of the proposed method. Finally, Sect. 5 summarizes the research conducted in this article.

2 Theoretical backgrounds

2.1 The Inception network

Introduced by Google researchers [28], the inception network serves as a foundational building block for Convolutional Neural Networks (CNNs). The architecture 1-D inception network, as shown in Fig. 1, is designed

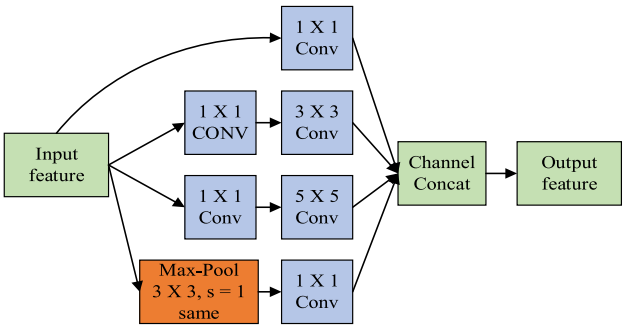


Fig. 1 Structure of the inception network

to enhance the capability of a CNN by facilitating multi-level feature selection through the incorporation of filters of various sizes within a single network layer. It is distinguished by several key features that set it apart from traditional CNN layers. The inception network simultaneously applies multiple convolutional filters of different sizes (e.g., 1×1, 3×3, 5×5) to the input, enabling the network to capture information at various scales and complexities. Additionally, 1×1 convolution is used for dimensionality reduction, which decreases the number of parameters while maintaining network depth. The network also integrates a parallel pooling branch, typically max pooling (e.g., with a pool size of 3), to provide an additional form of spatial aggregation. The outputs from all convolutional filters and the pooling layer are then concatenated along the channel dimension, ensuring that subsequent layers have access to a comprehensive set of features extracted at multiple scales. This multifaceted design significantly enhances the network's efficiency and effectiveness, particularly in tasks where capturing multiscale information is critical for accurate identification.

2.2 Squeeze-and-excitation block (SE)

The SE block proposed by Hu et al. [29] is one of the most influential works in attention mechanisms, designed to calibrate channel-wise features by considering inter-channel dependencies. The architecture of the 1D SE is depicted in Fig. 2, consisting of three operations, namely squeezing, excitation, and scaling. In the squeezing operation, the input feature maps $\mathrm{I}=[i_1, i_2, \ldots, i_c]$ of a SE block have a shape of $H\times C$, where H is the height, and C is the channels. The input I is first squeezed through a global average pooling (GAP) to create a weight vector, thus aggregating the global spatial information for the channel representation, which can also be considered as the statistics of each channel feature. The channel descriptor $d \in \mathbb{R}^C$ is generated by Eq. (1), which describes the global receptive information of each feature map I_c.

$$d_c = F_{sq}(I_c) = \frac{1}{H}\sum_{i=1}^{H} I_c^{(i)} \tag{1}$$

After executing the squeeze operation, the dimensions of the feature maps are $1\times C$. The excitation operation adaptively recalibrates channel-wise features by utilizing the information aggregated in the channel descriptor. The attention weight for each channel of the input feature maps is computed using the sigmoid function σ and the rectified linear unit (ReLU). The excitation vector s is derived as outlined in Eq. (2) by encoding the channel descriptor d into a shape of $1\times(C/r)$ and then decoding it back to a shape of $1\times C$ through two fully connected (FC) layers. These procedures introduce a nonlinear transformation and facilitate a smooth gating calculation, resulting in a channel weight vector.

$$s = \mathrm{F}_{\mathrm{ex}}(d, W) = \sigma\left(\mathrm{W}_2\ \mathrm{ReLU}\left(\mathrm{W}_1\ \mathrm{d}\right)\right) \tag{2}$$

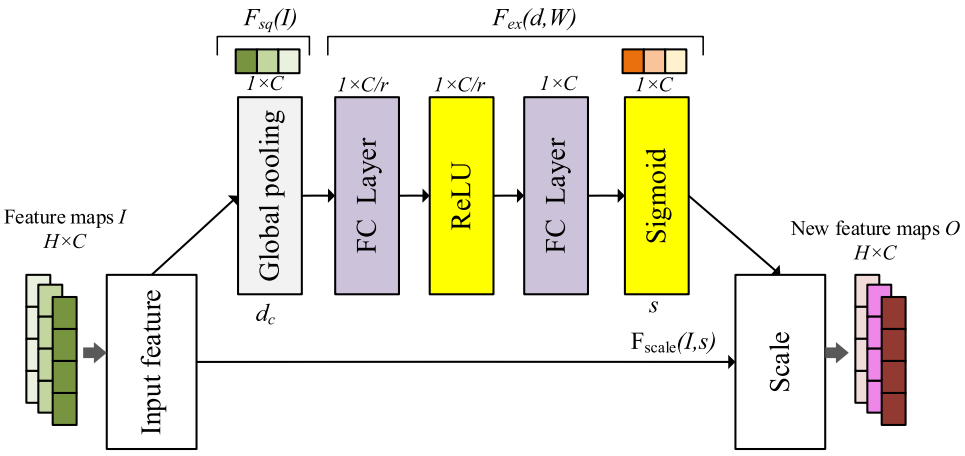


Fig. 2 Structure of the SE block

where $W_1 \in \mathbb{R}^{\frac{C}{r}\times C}$ and $W_2 \in \mathbb{R}^{C\times\frac{C}{r}}$ represent the parameters of the two FC layers, respectively, r is the reduction ratio, which helps to reduce both the capacity of the SE block.

Finally, the input feature maps $i_C \in \mathbb{R}^H$ is multiplied by the corresponding excitation vector s_c to produce the final output $O=[o_1, o_2, \ldots, o_c]$ in the scaling operation. In other words, the input feature maps are reweighted by the channel weight vector. After going through the SE block, input and output feature maps I and O have the same shape. The computation can be executed as follows:

$$o_c = F_{scale}(I_c, s_c) = s_c \times i_c \tag{3}$$

2.3 Residual CNNs

A prominent type of feedforward neural network utilized in FD is the Convolutional Neural Network (CNN) [30]. A conventional CNN primarily consists of a convolutional layer, pooling layer, fully connected layer, activation function, and batch normalization, as seen in Fig. 3. The convolutional layer (Conv) extracts features from raw input through the sliding operation of the convolution kernel. Following a sequence of convolutional operations on the input data, the convolutional layer is capable of extracting features with enhanced abstraction. To speed up neural network processing, each filter has an identical convolutional kernel to minimize parameters; this concept is referred to as weight sharing. The feature mapping is generated via the nonlinear activation function. The relationship between input and output is established as follows:

$$y_j^{l+1} = ReLU\left(\sum_i^k x_i^l . K_{ij}^l + b_j^l\right) \tag{4}$$

where x_i^l denotes the input of the l th layer, ReLU is the activation function, K_{ij}^l and b_j^l represent the weight and bias, respectively, and y_j^{l+1} is the j th feature map output by the l th layer.

The pooling layer is often placed after the convolutional layer to minimize feature dimensionality and parameters while further highlighting the extracted features. Average and maximum pooling layers are popular. The maximum pooling layer outputs the maximum pooled area value and is more common. The maximum pooling layer mathematical model is:

$$P_{(i,j)}^{l+1} = \underset{(j-1)W+1\le t\le jW}{Max}\left\{a_t^l(t)\right\} \tag{5}$$

where $a_t^l(t)$ represents the activation value of the t th neurons in the l th layer, W is the width of the pooling region, and $P_{(i,j)}^{l+1}$ is the value of the corresponding neuron after the output of the pooling layer.

The fully connected layer (FC), generally at the end of the network, integrates the features of previous convolution and pooling layers into a one-dimensional column vector, which then connects to the labels output layer and finally completes the classification by the Softmax function. The Softmax function is calculated as follows:

$$Q_j(v) = \frac{e^{v_j}}{\sum_{j=1}^n e^{v_j}} \tag{6}$$

where v_j represents the j th input feature, n donates the number of categories, and $Q_j(v)$ represents the probability that v_j is the j th category.

Batch normalization (BN) is a regularization method that standardizes the data inside the network model, hence accelerating the training process while ensuring that the original data maintains as much expressiveness as possible. The calculation can be performed as follows:

$$\mu = \frac{1}{N_{batch}}\sum_{s=1}^{N_{batch}} x_s \tag{7}$$

$$\sigma^2 = \frac{1}{N_{batch}}\sum_{s=1}^{N_{batch}} \left(x_s - \mu\right)^2 \tag{8}$$

$$\hat{x_s} = \frac{x_s-\mu}{\sqrt{\sigma^2-\varepsilon}} \tag{9}$$

$$y_s = \gamma\hat{x_s} + \beta \tag{10}$$

where x_s is the s th input, N_{batch} donates the number of small data batches, μ and σ^2 indicate the mean and variance of small data batches, respectively, ε represents a constant that is greater than 0 but close to 0, $\hat{x_s}$ is the result of normalizing the data, γ and β are the scale and shift parameters that can be learned by the network, respectively, and y_s is the s th output of the neuron after BN.

A residual neural network (ResNet) addresses the issue of gradient vanishing in deep convolutional neural networks by utilizing multiple residual blocks that are connected in sequence to learn residual features effectively. Residual connectivity, commonly known as identity mapping, functions as a technique for feature transfer in ResNet. A residual block was developed by He et al. [31], which utilizes two convolutional layers and implements

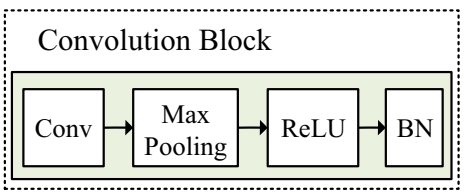


Fig. 3 Structure block of CNNs

residual connectivity. Figure 4 illustrates the structure of the residual block. In this representation, x represents the input, $H(x) = F(x) + x$ indicates the output and the residual neural network incorporates cross-layers on top of standard deep convolutional networks and exclusively learns the difference between the output and the input, expressed as $F(x) = H(x) — x$. It has been demonstrated through experiment validation that this learning approach enhances the backpropagation of the gradient. Also, it strengthens the feature learning capacity of the residual network.

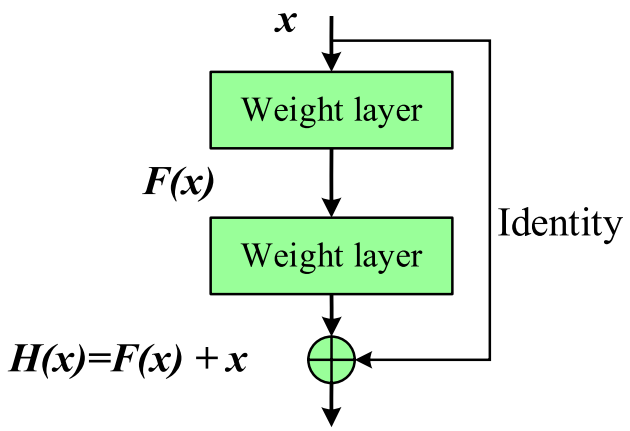


Fig. 4 Structure of the residual block

3 Proposed fault diagnosis method

3.1 Method overview

The advanced FD technique that employs CNN is widely used in the FD of rolling bearings. In cases where fault samples are limited, achieving optimal diagnostic results is difficult. As a result, we designed a novel deep learning framework with two new modules (CLC-Inception and CSDRB), as illustrated in Fig. 5, to enhance classification accuracy and robustness, particularly under limited data conditions and high noise interference. The proposed model generally involves four essential components: a wide convolutional layer, CLC-Inception, CSDRB, and CNN residual connections. A wide convolutional layer is initially used to mitigate background noise while maintaining essential signal characteristics. The CLC-Inception conducts parallel convolutions with varying kernel sizes to effectively capture multiscale features during the network's early stages. The extracted features are processed through convolutional blocks and CSDRB—an innovative component of the architecture. Each CSDRB employs channel and spatial attention. The channel recalibration block utilizes global average pooling to selectively highlight important channels, whereas the spatial recalibration block allows the model to concentrate on relevant spatial regions and reduce the influence of less relevant ones. The refined features are processed using global average pooling and Softmax classifier to determine the fault

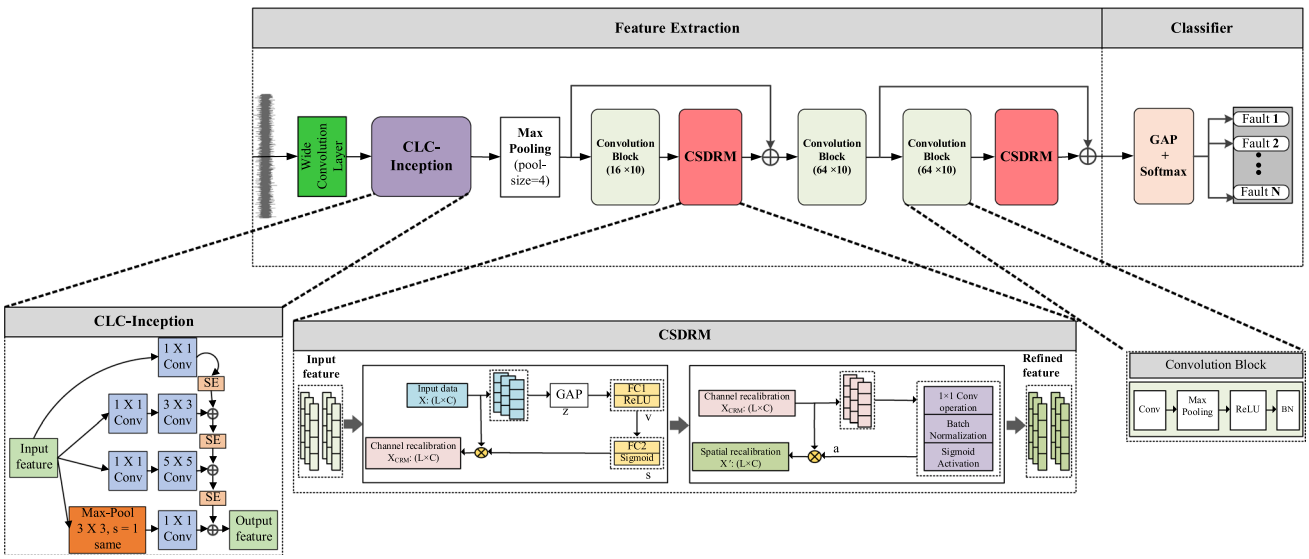


Fig. 5 Diagram of the proposed model structure. The architecture begins with a wide convolutional layer to suppress high-frequency noise while preserving essential signal features. The CLC-Inception module follows, performing parallel 1×1, 3×3, and 5×5 convolutions with SE-based channel recalibration. A Max Pooling layer (pool_size = 4) reduces feature dimensionality. Features then pass through three convolution blocks (kernel size = 10, filters = 16/64/64, ReLU activation, batch normalization) for hierarchical feature extraction. Between these blocks, CSDRB modules apply channel and spatial attention to refine representations by highlighting important channels and relevant spatial regions. Finally, Global Average Pooling and Softmax classify the fault type

category. The combination of multiscale feature representation, attention-based recalibration, and residual learning enhances the model's capacity to extract discriminative and noise-resistant features, leading to a fault diagnostic model characterized by high accuracy and generalizability. The model construction approach is described as follows:

1) This research utilizes a one-dimensional raw vibration signal as the input for the model. This method avoids the necessity for specific expertise in signal processing and enables end-to-end FD.
2) Throughout the collection process, vibration signals might get contaminated by noise. To mitigate the impact of high-frequency noise on fault detection, the first layer of the model has a wide convolutional layer, which is usually a convolutional layer with large convolutional kernels that help to alleviate high-frequency noise interference [32].
3) The temporal properties of vibration signals, which consist of one-dimensional time-series data, are crucial. Thus, the CLC-Inception design is enhanced by the integration of SE attention. This enables the model to focus on the most significant features, resulting in more comprehensive and reliable signal characterization, hence enhancing the model's performance and accuracy.
4) CSDRB functions to modify the network's attention toward varying information, including spatial mapping and global features, thereby minimizing irrelevant information or noise and improving the representation of fault features. Additionally, the residual connection facilitates the extraction of features through its robust depth feature identification capability, enhancing the network's ability to learn complex abstract information.
5) The convolution block offers substantial advantages in scenarios with limited samples by permitting efficient feature extraction and parameter sharing. It captures localized patterns, decreasing overfitting and improving generalization, and its hierarchical structure facilitates the network's ability to learn critical low-level and high-level features required for FD. GAP substitutes the fully connected layer to decrease the parameter count of the model and improve training efficiency. Then, the fault types are identified using the Softmax function. The subsequent sections outline the detailed specifications of the proposed method.

3.2 Cross-layer calibration inception (CLC-Inception)

The CLC-Inception module is an advanced architectural design that integrates the strengths of multiscale feature extraction inherent in Inception networks with the adaptive channel-wise recalibration capabilities of the Squeeze-and-Excitation (SE) mechanism. As shown in Fig. 6, the module receives an input feature tensor and processes it in parallel through various convolutional and pooling branches, each designed to capture features at distinct receptive fields. Specifically, the input feature is first split into four parallel paths:

1) 1×1 Convolution Path: This branch applies a single 1×1 convolution directly to the input feature, capturing local channel-wise interactions with minimal computational expense.
2) $1\times1\rightarrow3\times3$ Convolution Path: This branch employs a 1×1 convolution for early dimensionality reduction, succeeded by a 3×3 convolution to acquire mid-level contextual information with a modest receptive field, which facilitates quick and expressive feature transformation.
3) $1\times1\rightarrow5\times5$ Convolution Path: Similarly, this branch begins with a 1×1 convolution for dimension reduction, followed by a larger 5×5 convolution that extracts broader contextual dependencies, enabling the module to detect more global features in the input.
4) Pooling Path: This path implements a 3×3 max-pooling operation with (stride 1, padding set to "same" to preserve dimensions), so successfully collecting translation-invariant features. The output is then processed by a 1×1 convolution to adjust channel dimensions and incorporate learnable transformations.

CLC-Inception is characterized by the insertion of SE blocks at various stages. Following each convolutional branch, a SE block is used to recalibrate the output feature maps by explicitly modeling inter-channel relationships. Figure 6 illustrates that each convolutional output is independently recalibrated using its respective SE block. The adjusted outputs from all branches are then concatenated or summed, progressively passing through SE calibration

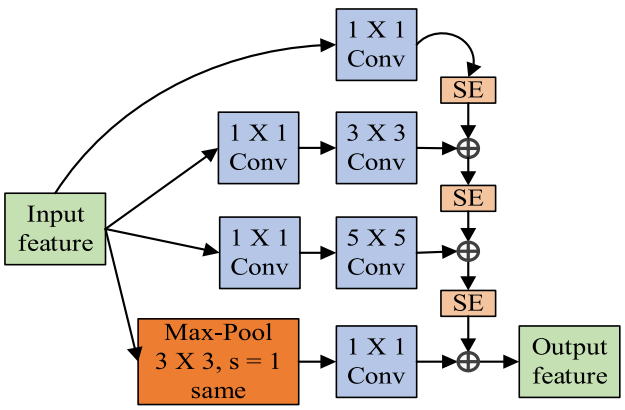


Fig. 6 Structure of CLC-Inception

layers after merging. This hierarchical application of SE blocks guarantees that both local and global characteristics are adaptively prioritized, preserving a consistent emphasis on relevant diagnostic information across scales. Finally, the adjusted feature maps are produced as the output feature, prepared for subsequent phases of the network. The CLC-Inception module effectively enhances the model's capacity to extract, recalibrate, and integrate rich, multiscale, and channel-aware representations by merging multiscale convolution paths with cross-layer SE calibration, demonstrating particular efficacy in bearing fault diagnosis under small data and noisy conditions.

3.3 Channel-spatial dynamic recalibration block (CSDRB)

To enhance feature learning capabilities, the Channel-Spatial Dynamic Recalibration Block (CSDRB) is proposed, specifically designed to adaptively recalibrate both channel-wise and temporal features. Unlike conventional squeeze-and-excitation attention or multiscale CNNs, the CSDRB explicitly integrates channel-wise and spatial-wise attention mechanisms into a cohesive recalibration framework. As illustrated in Fig. 7, CSDRB consists of two primary components: the Channel Recalibration Block (CRB) and the Spatial Recalibration Block (SRB). These two blocks work interactively, enabling the model to emphasize discriminative features more effectively. This interaction significantly improves overall feature representation in FD tasks. Furthermore, due to its streamlined structure and reduced number of training parameters, the CSDRB can be seamlessly integrated into various network architectures, making it particularly suitable for FD scenarios involving small or noisy data. Detailed explanations of the CRB and SRB components are provided below.

(1) CRB: The convolution operation functions in a local space, making it challenging to capture relational information across channels. The sensitivity of feature mapping differs across channels, with certain features potentially containing noise or lacking relevance to fault information [33]. This can mislead the network's feature learning process and decrease diagnosis efficiency. To improve the network's ability to extract global information, CRB is developed from the viewpoint of feature mapping across various channels, with the objective of adaptively amplifying channels linked to fault information while mitigating channels that have noise or incorrect data through explicit computation of inter-channel relationships.

The objective of CRB is to modify the significance of various channels within a feature map, allowing the network to emphasize channels that provide more useful information for a specific task and suppress less relevant ones. Figure 8 displays the structure of CRB, with the input feature map $X \in R^{L\times C}$ where C denotes the number of input channels, and L represents the size of the input feature (the sequence length).

Initially, GAP is utilized on the feature map to extract global information across the sequence dimension, resulting in a global descriptor, which can be stated as follows:

$$z = \frac{1}{L}\sum_{i=1}^{L} X_{i,c} \forall c \in [1, \mathrm{C}] \tag{11}$$

where $z \in R^C$ is a global descriptor that captures the average activation of the corresponding channel across over temporal positions.

Then, two fully connected layers are applied to calculate recalibration weights for each channel, starting by lowering the channel dimension to learn cross-channel dependencies and then extending it again using a sigmoid activation to get channel-wise recalibration weights. The first fully connected

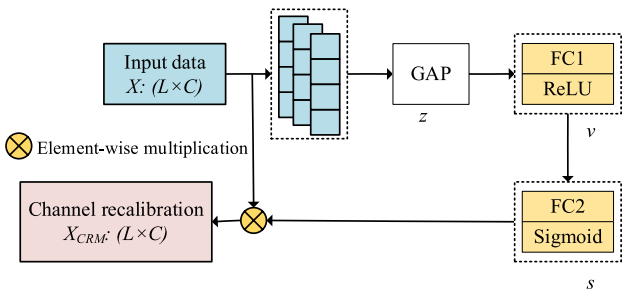


Fig. 8 Algorithm structure of CRB

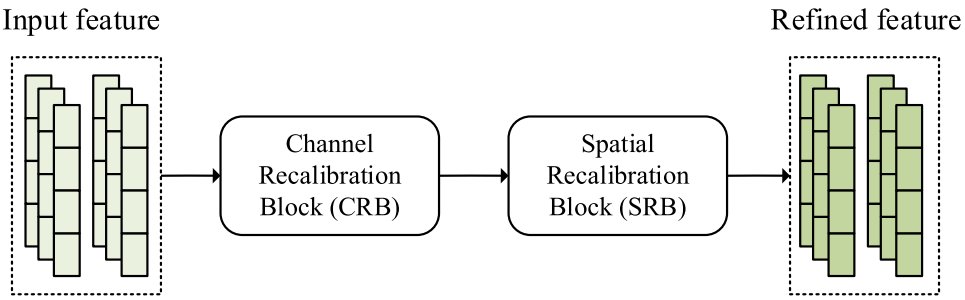


Fig. 7 Fundamental framework of the CSDRB

layer, FC_1, is the reduction layer, which reduces the number of channels by a reduction ratio r. The output from FC_1 is v, which can be computed as follows:

$$v = ReLU\left(W_1 z + b_1\right) \tag{12}$$

where $W_1 \in R^{C/r \times C}$ represents the FC_1 weights, $b_1 \in R^{C/r}$ represents the bias for the first layer, r donates as the reduction ratio and is set to 16, and $ReLU$ is an activation function. The expansion layer is the second fully connected layer FC_2. The channel representation is expanded back to the original number of channels using a sigmoid activation function σ with output values in the range [0,1].

$$s = \sigma\left(W_2 v + b_2\right) \tag{13}$$

where $W_2 \in R^{C \times C/r}$ represents the weights of the FC_2, $b_2 \in R^C$ represents the bias for the second layer, and $s \in R^C$ is the output of FC_2, which is channel-wise recalibration weights.

Finally, the original feature map $X \in R^{L \times C}$ and channel-wise recalibration weights $s \in R^C$ are element-wise multiplied. The X_{CRM} output is a recalibrated feature map obtained by multiplying each channel by its learned weight s. It can be calculated as follows:

$$X_{CRM} = X \cdot s \tag{14}$$

where $X_{CRM} \in R^{L \times C}$ represents the channel-wise recalibration.

(2) SRB: Impulse excitation happens as a rolling bearing with regional defects interacts with other components, resulting in the vibration signal data with a fault characteristic. Identifying the fault segment in the vibration signal simplifies failure mechanism diagnosis and minimizes noise. SRB is developed to strengthen the network focus on the captivating part of the original within the intricate signal, consequently simplifying the learning process of the model and making crucial information representation more efficient. As seen in Fig. 9, SRB aims to highlight or diminish specific temporal positions within the feature map, enabling the network to discern which areas in the sequence hold greater significance. In this study, the input to SRM is typically the recalibrated feature map from the CRM, but it can also be the original input feature map.

First, 1×1 convolution is performed on the input throughout the channel dimension to create spatial recalibration weights a for each temporal point along the sequence. Batch normalization (BN) is then used to stabilize and speed training by normalizing the input. The sigmoid activation σ is subsequently applied to guarantee that the recalibration weights are in the range [0,1], allowing the network to either emphasize (a close to 1) or suppress (a close to 0) certain temporal positions. The output weights $a \in R^{L \times 1}$ are computed as:

$$a = \sigma\left(W * X_{CRM} + b\right) \tag{15}$$

where $W \in R^{1 \times C \times 1}$ denotes the convolution kernel, which has a kernel size of 1 and works along the channel dimension,$X_{CRM} \in R^{L \times C}$ represents the channel-wise recalibration, $b \in R^C$ denotes the bias term, and σ represents the sigmoid activation function.

Eventually, element-wise multiplication is executed between the spatial recalibration weights a and the input feature map X_{CRM}, resulting in the spatial recalibrated output $X' \in R^{L \times C}$. The element-wise multiplication permits the SRB to adaptively modify the significance of each temporal location according to the learned weights. This can be depicted as follows:

$$X' = X_{CRM} \cdot a \tag{16}$$

3.4 Rolling bearing FD procedure

In this paper, the key hyperparameters for the proposed FD model based on CSDRB with CLC-Inception are determined through empirical analysis, ablation experiments, and references in related literature. Specifically, we conducted a series of experiments using grid search and cross-validation to fine-tune hyperparameters such as the convolution

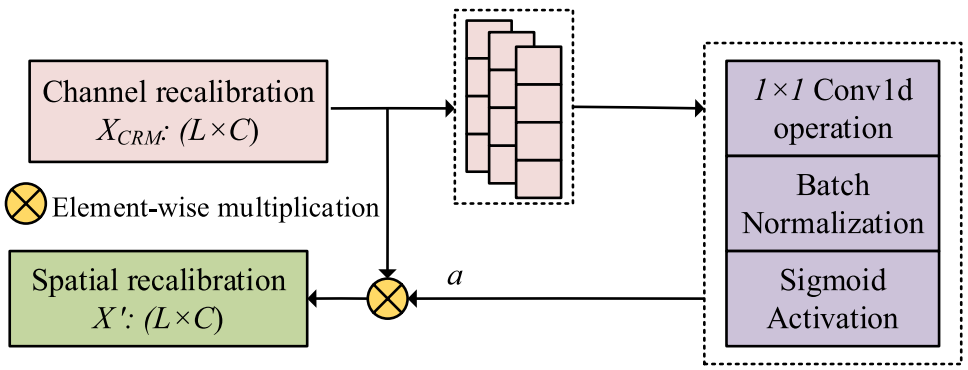


Fig. 9 Algorithm structure of SRB

kernel sizes, stride lengths, recalibration weights, and learning rates. Parameters were initially chosen based on prior work and domain knowledge and then iteratively adjusted to achieve optimal diagnostic performance on the validation set. This approach ensured both model stability and generalization across varying conditions. For transparency and reproducibility, the final parameter values are clearly listed in Table 1. The research procedure of the proposed method for diagnosing bearing faults, which is shown in Fig. 10, encompasses several stages: data preprocessing, model training, and model testing. The specific procedures are as follows:

Table 1 Summary of specific parameters for the proposed model

Stage	Hyperparameters	Value
Feature extraction	Input	Data size 1024 × 1
	Wide convolution layer	Kernel size 32 Stride 4 Filters 16 Activation ReLU
	CLC-Inception	16, 32, 32, 8, 16, 16 filters for respective branches 1 × 1, 3 × 3 reduced, 3 × 3, 5 × 5 reduced, and 5 × 5
	Convolution	Kernel size 10 Stride 1 Max Pooling pool_size = 4 Activation ReLU batch normalization () Filters 16, 64, and 64
	CSDRB	-
Classifier	Softmax	Number of classes 10/4
Training	Loss function	Cross-entropy loss
	Optimizer	Adam (learning rate of 0.0001)
	Batch size	12
	Number of epochs	30

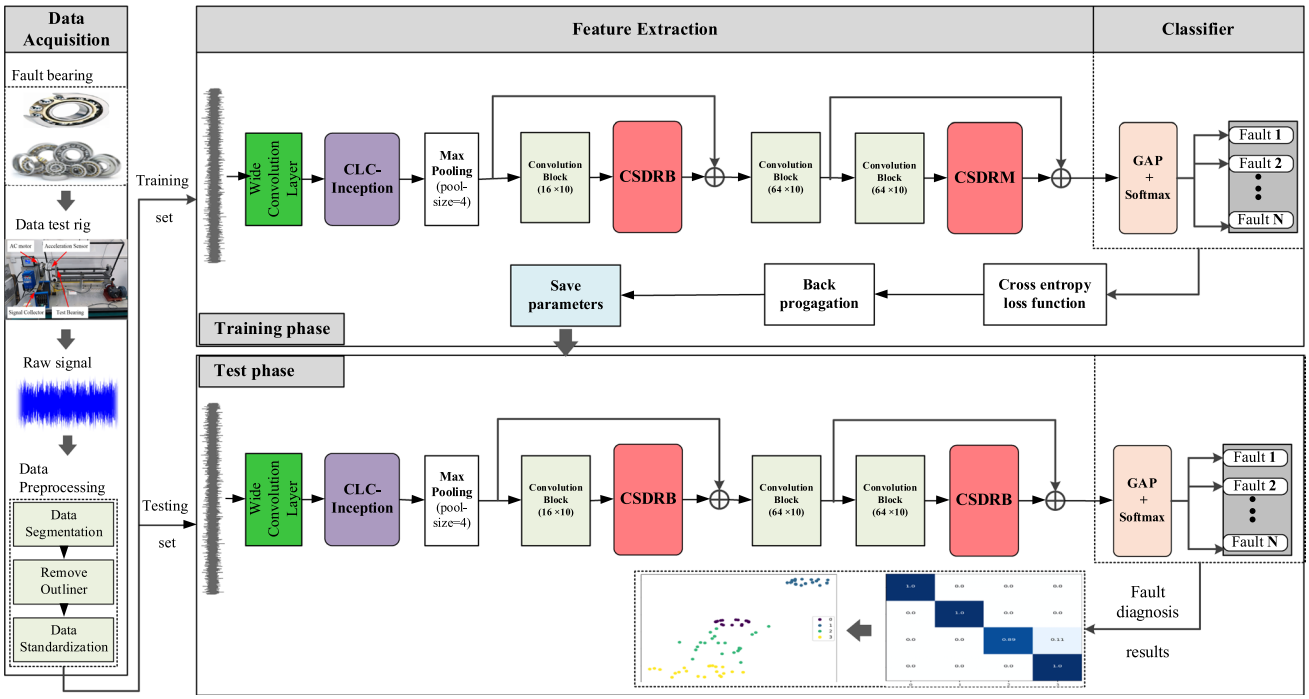


Fig. 10 Bearing fault diagnosis procedure

1) Collect the raw vibration data on bearing faults.
2) Data preprocessing: The first step is data segmentation, which involves splitting the obtained vibration data into segments of a given length. Then, outliers are removed since anomalous conditions may arise during data gathering, and the combined vibration signals may not be error-free. Before the data are transferred to the network, the samples are standardized to ensure adherence to a standard Gaussian distribution. Finally, one-hot encoding is used for various fault types, and the dataset is partitioned into training and test sets.
1) Provide the training set as the model's input. The samples that exhibit various fault types first undergo feature learning/extraction, wherein the wide convolution layer eliminates noise interference, CLC-Inception extracts deep features, and CSDRB captures both spatial and temporal fault characteristics. The channel and spatial weights are dynamically modified via CSDRB. The classifier finalizes the fault identification procedure.
2) Adam gradient optimization is employed to adjust the weights and biases of each layer during backpropagation, while the cross-entropy loss between the actual values and the predicted values is determined at each iteration.
3) To preserve the model parameters until the training loss value approaches the early stopping setting, repeat steps (3) and (4).
4) Input the test set into the trained model to evaluate its performance. The Softmax activation function is used to diagnose bearing faults.

4 Experiment analysis and discussion

In this section, the public bearing datasets from Case Western Reserve University (CWRU) and the mechanical fault simulation (MFS) datasets from Lanzhou University of Technology (LUT) are utilized to evaluate the diagnostic performance of the proposed method. Also, additional challenging tasks are done in two case studies, which are FD under small sample conditions and noisy environments. All algorithm programs are implemented in Tensorflow2.5, and the programming language is Python 3.7. The computer configuration includes AMD-R5-5600H CPU, NVIDIA RTX 3060 GPU, and 16 GB RAM.

Four evaluation metrics are employed for FD results: Accuracy, Precision, Recall, and F1-score, with their respective formulae detailed below:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FN+FP} \times 100\% \tag{17}$$

$$\text{Percision} = \frac{TP}{TP+FP} \times 100\% \tag{18}$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \tag{19}$$

$$F1\_score = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \times 100\% \tag{20}$$

where TP, TN, FN, FP refer to the quantities of true-positive samples, true-negative samples, false-negative samples, and false-positive samples, respectively.

4.1 Case 1

4.1.1 (1) Data description

The CWRU bearing data center designs a motor bearing failure analysis dataset that covers multiple failure types and varying degrees of severity [34]. The CWRU dataset comprises four main failure categories: health, ball fault, inner ring fault (IR), and outer ring fault (OR). Defect diameters for each type of failure are 0.007, 0.014, and 0.021 inches, respectively. Figure 11 shows the CWRU test bench; the motor sustained the specified working speed while functioning under loads from 0 to 3 hp, producing 10 distinct state categories. The bearing data used in this study are collected via the acceleration sensor at the drive end of the motor, with a sampling rate of 12 kHz; a total of 6000 data samples are generated by the sliding window technique to ensure the accuracy and robustness of the evaluation. The data enhancement is implemented using the sliding window technique, which employs a window length of 1024 points to ensure that each sample covers at least one complete bearing rotation cycle and a window step size of 100 points. As displayed in Table 2, there are two small sample datasets, Dataset 1 and Dataset 2, each containing ten signal data types. We divide the CWRU dataset into training and test sets according to the ratio of 2:3 (test size is 0.6). Dataset 1 includes 420 test samples and 280 training samples, with each class comprising 70 samples split into 28 for training and 42 for testing.

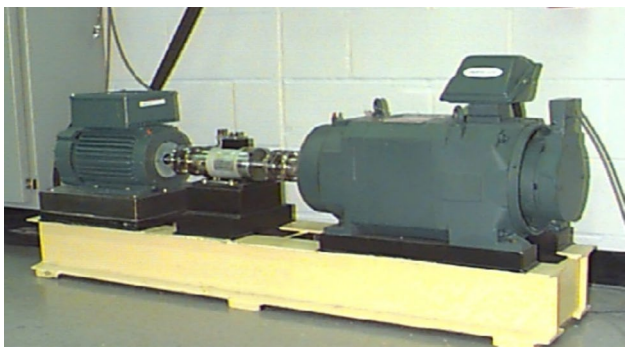


Fig. 11 CWRU test bench

Table 2 Details of the CWRU dataset

Label	Fault location	Fault size (mil)	Dataset 1 Training/Testing	Dataset 2 Training/Testing
0	Normal	–	28/42	40/60
1	IR	7	28/42	40/60
2	IR	14	28/42	40/60
3	IR	21	28/42	40/60
4	OR	7	28/42	40/60
5	OR	14	28/42	40/60
6	OR	21	28/42	40/60
7	Ball	7	28/42	40/60
8	Ball	14	28/42	40/60
9	Ball	21	28/42	40/60
	ALL		280/420	400/600

Dataset 2 consists of 600 test samples and 400 training samples, with each class containing 100 samples divided into 40 for training and 60 for testing.

4.1.2 (2) Experimental analysis under small sample conditions

The testing performance of the proposed FD method under small sample conditions is demonstrated in Table 3, which employs the CWRU dataset. Dataset 1 comprises 200 training samples, resulting in an Accuracy of Dataset 1 is 99.28%, with Precision, Recall, and F1-score values of 99.34%, 99.37%, and 99.34%, respectively. In contrast, Dataset 2, which contains 400 training samples, exhibits superior performance, achieving an Accuracy of 99.83%, as well as Precision, Recall, and F1-score values of 99.83%, 99.81%, and 99.82%, respectively. The results indicate that our model performs exceptionally well on both datasets. Dataset 2 shows slightly improved results, likely due to the increased number of training samples, presumably enhancing the model's generalization capability. This underscores the effectiveness and robustness of the proposed method for fault identification, even in the presence of restricted sample sizes, and illustrates its capacity to maintain consistent and high accuracy across a wide range of dataset sizes.

Table 3 Results on CWRU under small sample conditions

CWRU Datasets	Accuracy	Precision	Recall	F1-score
Dataset 1	99.28%	99.34%	99.37%	99.34%
Dataset 2	99.83%	99.83%	99.81%	99.82%

4.1.3 (3) Testing performance under small sample conditions and noisy environment

Bearings are frequently influenced by noise during operation, which changes the characteristics of vibration signals. Consequently, FD in conditions of limited samples and noisy environments presents a huge obstacle. This article simulates true industrial environmental noise by mixing the original vibration signal with Gaussian white noise at different signal-to-noise ratios (SNR). The SNR is defined as follows:

$$\mathrm{SNR}_{\mathrm{dB}} = 10\log\left(\frac{\mathrm{P}_{\mathrm{signal}}}{\mathrm{P}_{\mathrm{noise}}}\right) \tag{21}$$

where P_{signal} represents the original signal power and P_{noise} represents the Gaussian noise power.

The experiments in this section are conducted using the CWRU dataset as the primary experimental dataset. Four noisy environments are simulated with SNR = −5, −3, −1, and 0 dB. The DL models are trained on original data and noise-free signals, whereas their performance is assessed using noise-mixed data. This testing methodology compels the DL models to extract deeper and more robust features from the original data, thereby improving their resistance to noise. The limited sample conditions comprise Dataset 1 and Dataset 2, containing 280 and 400 training examples, respectively, while the noise settings encompass −5, −3, −1, and 0 dB values. The experimental results are detailed in Table 4, which illustrates a decrease in SNR correlates with an increase in noise interference intensity. A notable decline in diagnostic accuracy occurs as the SNR decreases from 0 dB to −5 dB. Despite this, the proposed method consistently maintains a high level of accuracy, depicting its robustness in challenging noise conditions with limited samples.

Figures 12 and 13 illustrate the confusion matrix results for fault classification in scenarios characterized by limited sample sizes and different noise levels (0 dB, −1 dB, −3 dB, and −5 dB) utilizing the CWRU dataset. In Fig. 12a, the

Table 4 Results on CWRU under small sample conditions and noisy environment

Number of samples	SNR(dB)	Accuracy	Precision	Recall	F1-score
700 (Dataset 1)	0	99.04%	99.04%	98.99%	99.00%
	-1	98.33%	98.56%	98.52%	98.49%
	-3	97.61%	97.57%	97.53%	97.53%
	-5	95.47%	95.65%	95.07%	95.17%
1000 (Dataset 2)	0	99.83%	99.83%	99.81%	99.82%
	-1	99.16%	99.21%	99.23%	99.22%
	-3	97.33%	98.03%	97.37%	97.43%
	-5	95.66%	95.62%	95.65%	95.59%

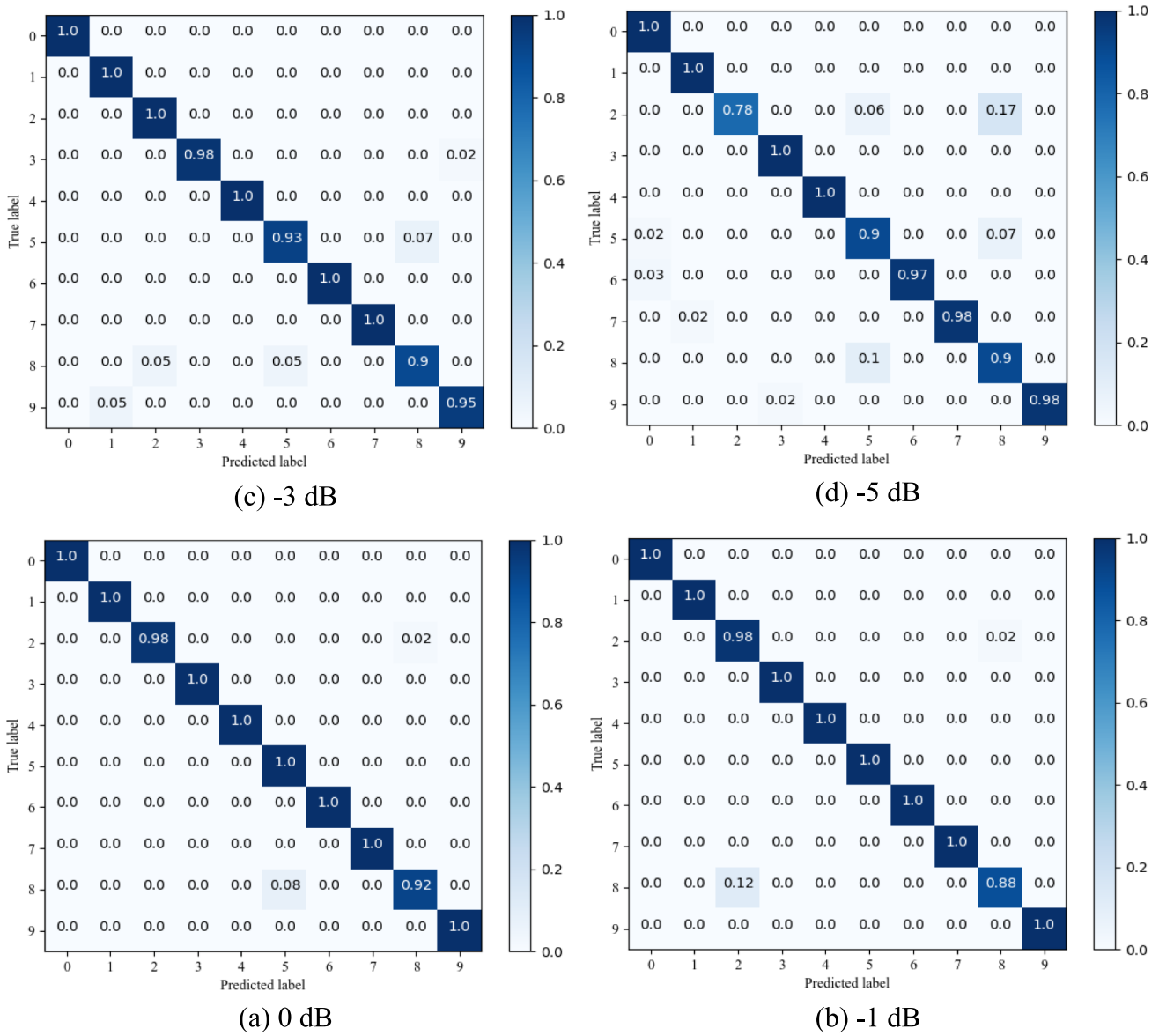


Fig. 12 Confusion matrix results on the CWRU dataset with 700 samples under different noise levels: (a) 0 dB, (b) -1 dB, (c) -3 dB, and -5 dB

model demonstrates exceptional classification performance at a 0 dB noise level, as evidenced by the high accuracy across all classes, with diagonal values predominantly at 1.0. As noise levels rise to −1 dB, −3 dB, and −5 dB, minor misclassifications appear; for example, in Fig. 12b for class 2, accuracy drops to 0.88. Figure 12c demonstrates that high accuracy is preserved across most categories, including classes 0, 1, 2, 4, 6, and 7, which exhibit diagonal values of 1.0. In the −5 dB Fig. 12d, the confusion matrix indicates a decline in accuracy specifically for class 2, class 5, and class 8, with respective values of 0.78, 0.90, and 0.90. Concurrently, the overall results continue to demonstrate high classification accuracy.

In Fig. 13, the larger training samples elevate robustness across all noise levels. At 0 dB and -1 dB, the model exhibits its excellent classification accuracy with negligible errors. Figure 13c demonstrates that the classification accuracy achieves 100%; however, class 8 reveals a significant misclassification rate of 0.23 at −3 dB. In −5 dB Fig. 13d, notable misclassifications occur in class 8 and class 5, exhibiting accuracies of 0.74 and 0.88, respectively. The overall performance is superior to that depicted in Fig. 12, with significantly lower misclassification rates. The results indicate that although the proposed method is effective with small sample sizes, increasing the dataset substantially improves FD accuracy and noise resilience.

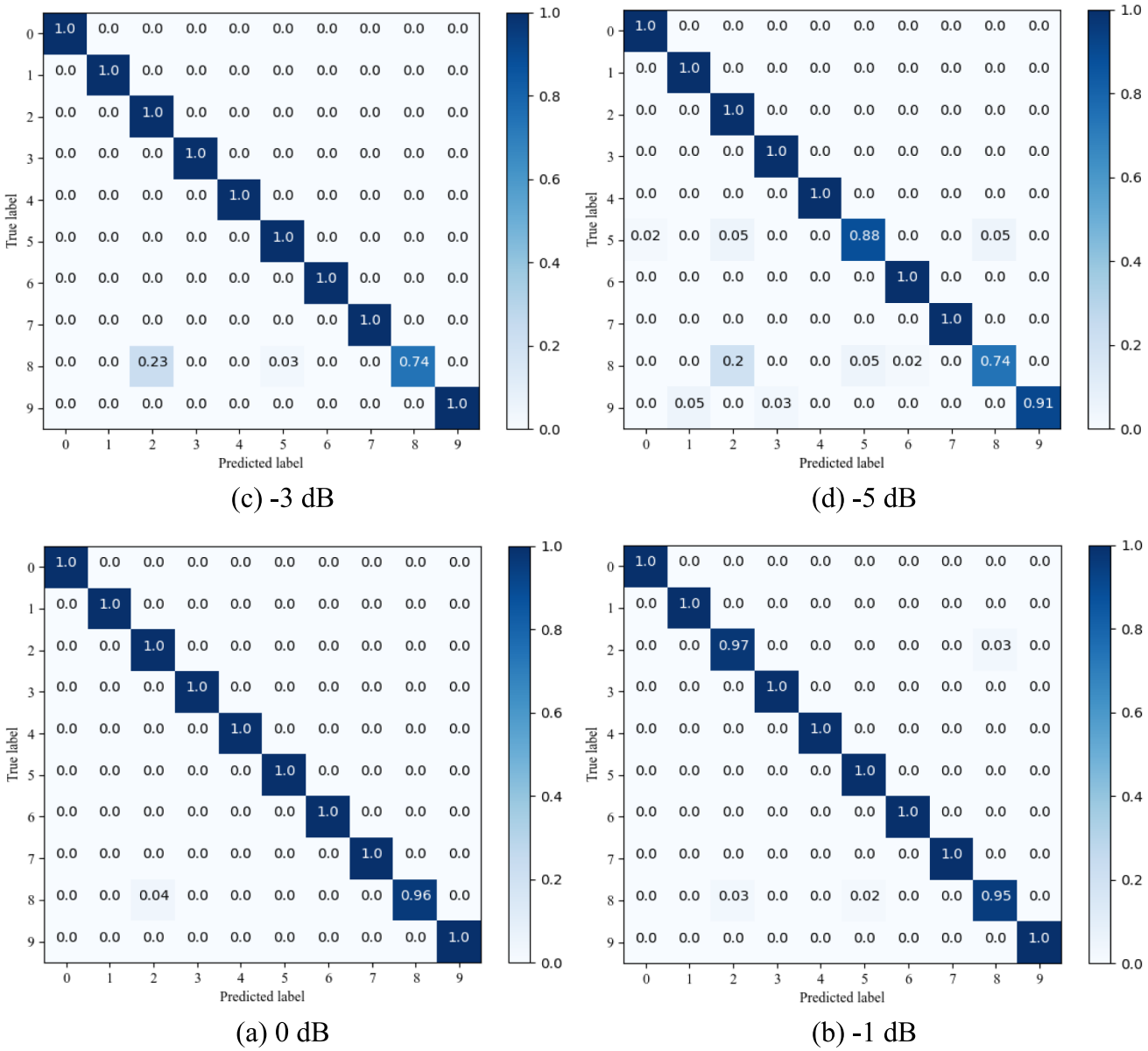
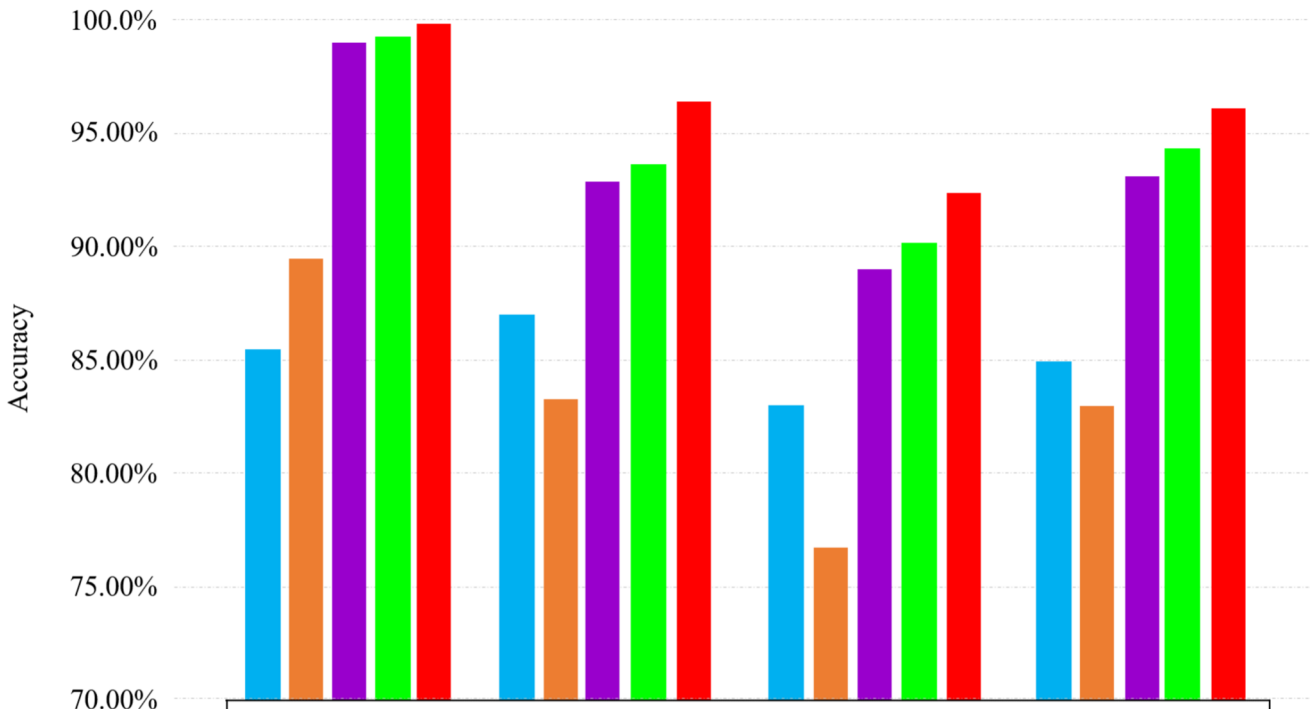


Fig. 13 Confusion matrix results on the CWRU dataset with 1000 samples under different noise levels: (a) 0 dB, (b) -1 dB, (c) -3 dB, and -5 dB

4.1.4 (4) Comparison with small sample fault diagnosis methods

To validate the superiority of the proposed method, this paper selects several advanced methods for comparison: MCAMDN [35], RESCNN [36], MSDARN [37], and MA1DCNN [38]. MCAMDN is a multiscale network featuring selectable branching, designed to mitigate noise interference while enhancing the extraction of useful data. RESCNN is a residual network architecture comprising a residual module and a wide convolutional anti-noise block. MSDARN integrates multiscale learning with channel attention mechanisms to facilitate dynamic feature selection. MA1DCNN integrates a joint multi-attention module, resulting in improved performance in bearing fault identification. Figure 14 presents the diagnosis outcomes of multiple FD methods on the CWRU dataset in noisy situations (0 dB, −4 dB, and −6 dB). The proposed method has the maximum accuracy of 99.83% at 0 dB, exceeding MCAMDN (85.40%), RESCNN (88.98%), MSDARN (98.17%), and MA1DCNN (98.28%). As the noise grows to -4 dB, all other methods decrease accuracy, but the proposed method outperforms others. At the most extreme noise level of -6 dB, the proposed method obtains 92.33% accuracy, surpassing MA1DCNN (90.06%) and MSDARN (88.41%), whereas MCAMDN and RESCNN perform worse at 82.70% and



Methods	SNR 0 dB	-4 dB	-6 dB	Average
MCAMDN	85.40%	86.80%	82.70%	84.97%
RESCNN	88.98%	82.89%	76.81%	82.89%
MSDARN	98.17%	92.91%	88.41%	93.16%
MAIDCNN	98.28%	93.80%	90.06%	94.05%
Proposed Method	99.83%	96.33%	92.33%	96.16%

Fig. 14 Diagnosis results of different methods on the CWRU dataset under noisy environments

76.81%, respectively. These results suggest that the proposed method can maintain excellent diagnostic accuracy across various noise levels, even under difficult conditions where other methods experience significant performance degradation.

To better visualize and understand the feature distributions learned by the proposed method on the CWRU dataset under noisy environments at different noise levels, the t-SNE technique [39] is applied to depict the classification outcomes. The t-SNE uses various colors to denote several fault categories; each point is a sample inside the dataset, and the clustered regions of color denote that these samples have similarities in some characteristics. In Fig. 15a at 0 dB, the 10 fault types are distinctly separated, and the clustering is clearest, suggesting that the method successfully differentiates between classes with minimal noise interference. At −4 dB Fig. 15b, the separation between certain clusters begins to diminish, with specific fault classes showing overlap, indicating a minor decline in classification performance. At the highest noise level of -6 dB in Fig. 15c, the clustering shows further degradation, with considerable overlap among several classes, especially for those that are more sensitive to noise, highlighting the difficulties introduced by severe noise conditions. Despite the increasing noise levels, the t-SNE visualization indicates that the proposed method can distinguish fault types under noisy environments degree, highlighting its robustness in extracting and preserving discriminative features.

4.2 Case 2

4.2.1 (1) Data description

The Machinery Fault Simulator (MFS) test rig, designed by Spectrum Quest Incorporated (SQI), is employed to simulate both single and compound faults. Figure 16

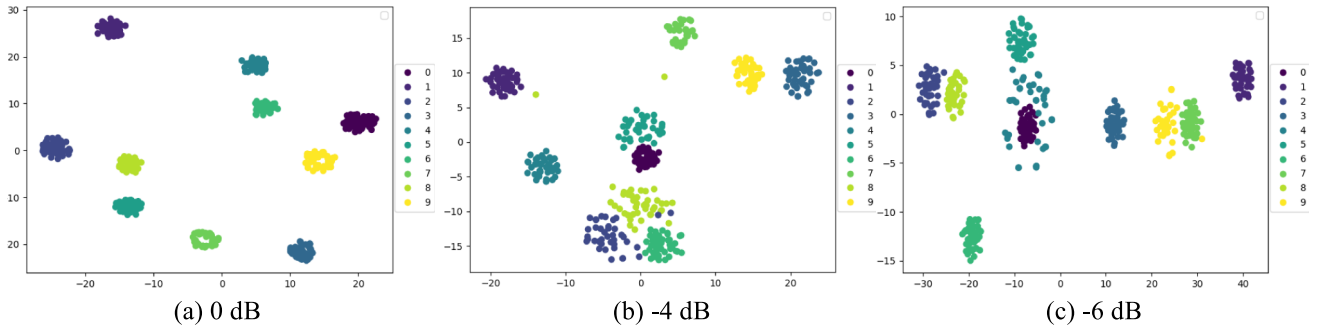


Fig. 15 Visualization results of the proposed method on the CWRU dataset under noisy environments: (a) 0 dB, (b) -4 dB, and (c) -6 dB

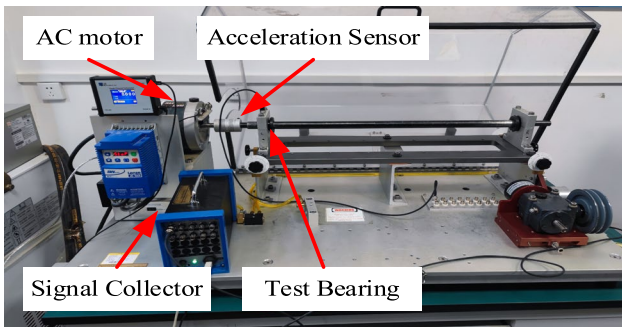


Fig. 16 MFS test bench

Table 5 Details of the compound fault dataset

Label	Fault location	Fault size (mil)
0	Ball	6
1	Inner race	24
2	Outer race	6
3	Compound (Inner race and Outer race)	Inner 6 & Outer 9

presents the MFS test setup, which includes components such as the drive end, accelerometer, and signal collector. The motor is connected to the drive shaft via a flexible coupling. Laser technology is utilized to induce various fault types in bearings, including ball faults, inner ring faults, outer ring faults, and combined inner and outer ring faults. Vibration signals are recorded at a sampling frequency of 15.6 kHz. The bearings operate at a rotational speed of 1130 r/min, with vibration data transmitted to a computer through a single-channel data cable and USB interface [40]. A window length of 1024 points is used to gather data samples, with a window step size of 100 points.

In practical applications, inner-race defects typically originate at the rolling contact zone where the loaded roller or ball induces localized stress and micro-pitting, whereas outer-race damage often arises from misalignment or housing-induced stress concentrations; when both occur simultaneously, which results in compound faults, the interaction of the two damage sites exacerbates vibration and accelerates bearing failure. Table 5 provides detailed information about the composite fault dataset with four categories. A total of 200 samples, comprising 50 samples for each fault type, with 80 designated for training and 120 for testing, according to a 2:3 ratio, as seen in Table 5.

Figure 17 illustrates the visualizations of the vibration signal in both the time domain and frequency domain for four distinct fault types. The vibration magnitudes for these fault types exhibit significant variation, correlating with their respective severities. In the time-domain plots (a, c, e, g), the 6 mil ball defect produces the smallest oscillations, approximately ± 0.04 units—indicating a minor disturbance, whereas the 24 mil inner-race fault generates the largest excursions of around ± 0.6 units, marking it as the most severe in terms of impact amplitude. The compound fault (6 mil inner and 9 mil outer) yields intermediate spikes of about ± 0.2 units. In a compound fault, every impact from an inner-race pit is superimposed on the shocks generated by outer-race damage, resulting in time-domain waveforms that blend periodic impulses of differing amplitudes and intervals (as seen in subplot (e)), rather than the clean, isolated spikes of single faults. The 6 mil outer-race defect also reaches roughly ± 0.2 units but with a more irregular waveform. In the frequency domain (b, d, f, h), the ball fault's spectrum peaks near 1.2 units at its characteristic frequency, and the inner-race fault shows a broad, lower-frequency peak around 20 units. The compound fault's dominant spectral amplitude is about 4 units; this superposition yields multiple characteristic peaks and sidebands (subplot f) that can mask or distort one another, making it difficult to attribute energy at a given frequency to one fault location versus the other. The outer-race defect exhibits the highest frequency-domain peak of roughly 7.5 units. Consequently, extracting reliable,

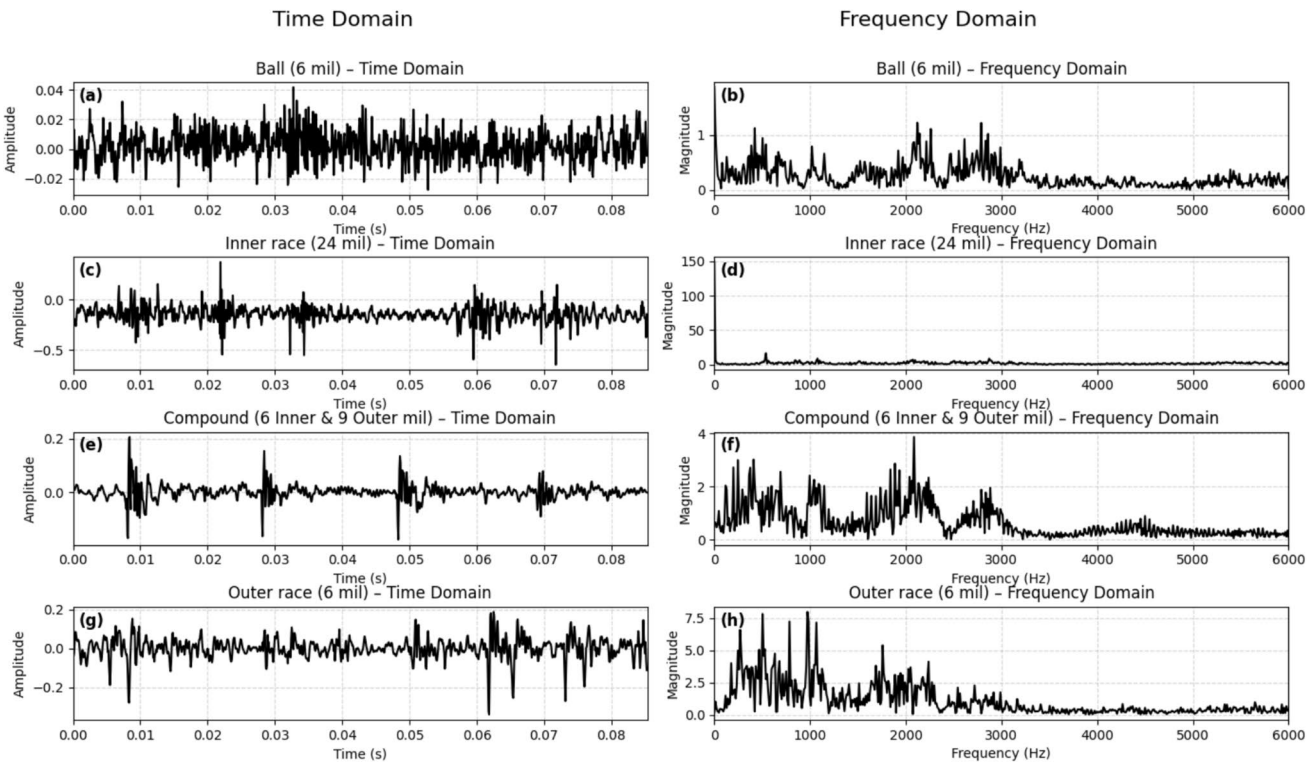


Fig. 17 Visualizations of the vibration signal's time domain and frequency domain for four fault types. (a), (b) the time domain and frequency domain of ball fault; (c), (d) the time domain and frequency domain of inner race fault; (e), (f) the time domain and frequency domain of compound fault; (g), (h) the time domain and frequency domain outer race fault

Table 6 Results on compound fault dataset under small sample conditions

Number of samples	Accuracy	Precision	Recall	F1-score
Dataset (200 Samples)	99.16%	99.19%	99.21%	99.19%
100 Samples	98.33%	98.61%	98.68%	98.61%

discriminative features from compound faults requires sophisticated multiscale and attention-based mechanisms to disentangle these intertwined patterns, which is why they are inherently more complex to diagnose than single-fault conditions.

4.2.2 (2) Experimental analysis under small sample conditions

Table 6 demonstrates the performance of the proposed method in identifying bearing defects under limited sample conditions, utilizing a dataset of 200 samples and when the dataset is reduced to 100 samples, evaluated across four metrics: Accuracy, Precision, Recall, and F1-score. The approach attains near-perfect performance with 200 samples, with an Accuracy of 99.16%, Precision of 99.19%, Recall of 99.21%, and F1-score of 99.19%, demonstrating its robust diagnosis capabilities under mostly limited data. A reduction in the sample size to 100 results in a slight decrease in performance, with accuracy, Precision, Recall, and F1-score decreasing to 98.33%, 98.61%, 98.68%, and 98.61%, respectively. Nonetheless, the method continues to exhibit superior performance, demonstrating its resilience and capacity for generalization despite the scarcity of data.

Table 7 Results on compound fault dataset under small sample conditions and noisy environment

Compound fault dataset (200 samples)	SNR(dB)			
	-5	-3	-1	0
Accuracy	96.66%	97.50%	98.33%	99.16%
Precision	96.56%	97.49%	98.48%	99.10%
Recall	96.32%	97.44%	98.14%	99.32%
F1-score	96.35%	97.45%	98.25%	99.20%

4.2.3 (3) Testing performance under small sample conditions and noisy environment

This part utilizes 120 testing samples and 80 training samples to verify the effectiveness of the proposed method. As seen in Table 7, The proposed method shows strong potential

in diagnosing bearing faults under small sample conditions with noisy environments employing Gaussian white noise according to Eq. (21) at different SNR: 0, −1, −3, and −5 dB. Among these four fault types, the compound inner and outer race vibration signal is the most challenging to detect. The overlapping impulsive patterns at two distinct defect frequencies create a highly non-stationary waveform in which individual fault signatures can mask one another, as seen in Fig. 18a. The original compound fault vibration signal, characterized by impulsive peaks resulting from the combined inner and outer race defects, is superimposed on a baseline of moderate background variability. In Fig. 18b, the original compound signal is mixed by additive white Gaussian noise at an SNR of 0 dB, leading to noise amplitudes that are, on average, similar to the fault impulses. Although the key defect-related spikes are still identifiable, In subfigure (c) (SNR = –1 dB), the noise amplitude slightly surpasses that of the signal, which further obscures the periodic fault patterns and diminishes the prominence of the impulsive events. In subfigure (d) (SNR = –3 dB), noise significantly obscures the waveform. In subfigure (e) (SNR = –5 dB), the noise energy completely obscures the signal, rendering the fault impulses nearly indistinguishable amidst the dense random fluctuations. As the noise level decreases (higher SNR), the model consistently improves performance across all parameters, including Accuracy, Precision, Recall, and F1-score. At a low SNR of −5 dB, the approach achieves Accuracy, Precision, Recall, and F1-score values of 96.66%, 96.56%, 96.32%, and 96.35%, respectively, demonstrating strong diagnosis efficacy despite significant noise interference. The

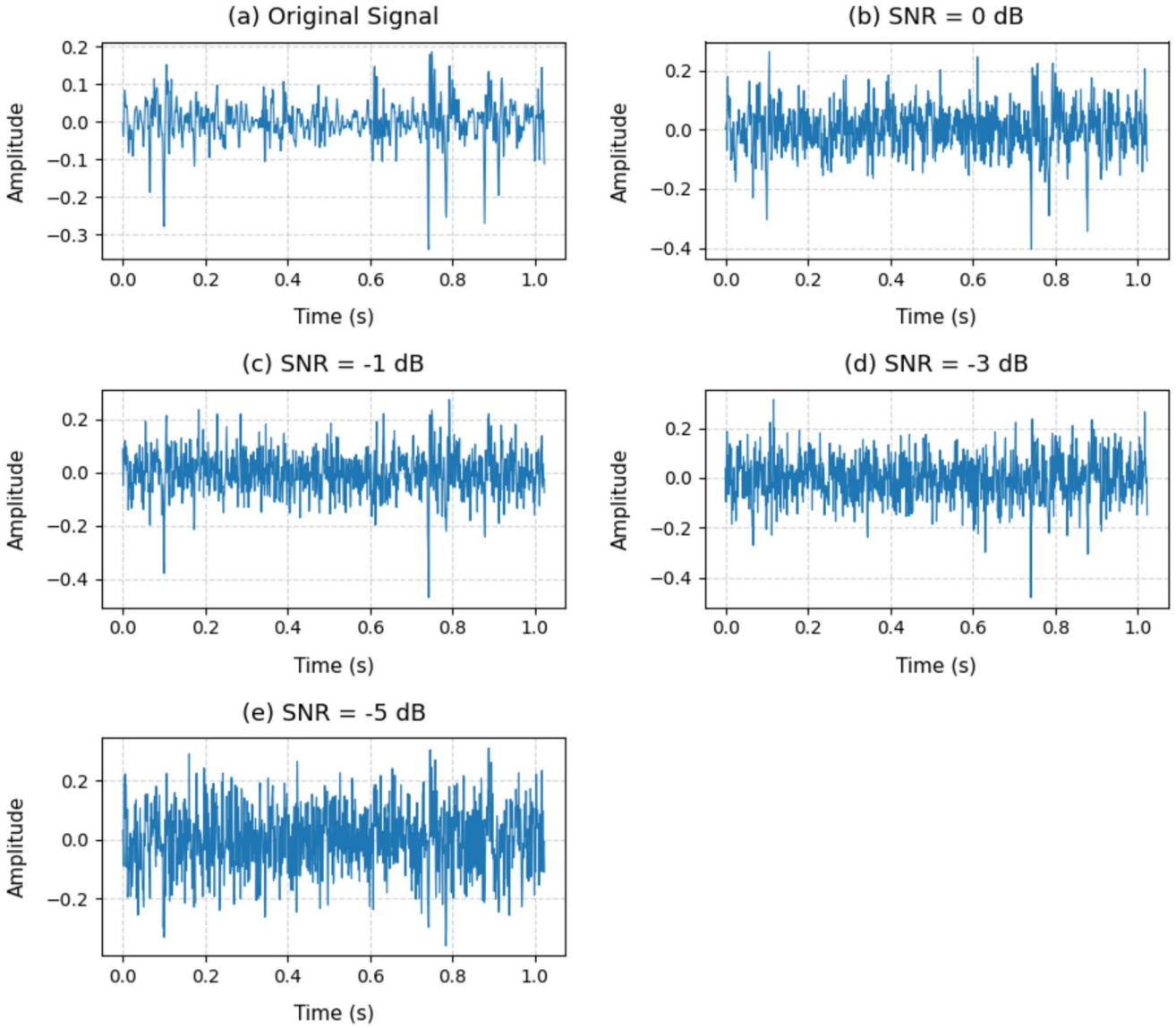


Fig. 18 Visualization of compound inner and outer race raw vibration signal with different levels of noise

proposed method achieves near-perfect results at an SNR of 0 dB, with an accuracy of 99.16% and an F1-score of 99.20%, as noise decreases.

In Fig. 19, the confusion matrix provides a detailed representation of the testing performance of the proposed method, evaluated with limited samples and different noise levels, applying the compound fault dataset with 80 training samples. Figure 19a at 0 dB shows near-perfect classification with all classes achieving almost 100% accuracy, except for a minor misclassification of 3% in class 2. In Fig. 19b, at − 1 dB, performance slightly decreases, with class 2 indicating a misclassification rate of 7%, while the other classes remain unchanged. In subfigure (c) -3 dB, class 2 retains a 3% misclassification rate, whereas class 3 experiences a 7% decline in accuracy, reflecting increased noise. At the highest noise level of − 5 dB (subfigure d), classification accuracy is further degraded, with class 2 and class 3 showing notable misclassification rates of 11% and 4%, respectively.

In summary, the above results prove the method's robustness at low noise levels, maintaining excellent classification accuracy for most classes. Misclassifications grow with noise, yet the proposed method performs well in high-noise situations.

4.2.4 (4) Impact of the proposed CSDRB and CLC-Inception

Four methods are established to assess their impact on the model performance regarding Accuracy, Precision, Recall, and F1-score. Table 8 describes the results on the implications of including or excluding the CSDRB and

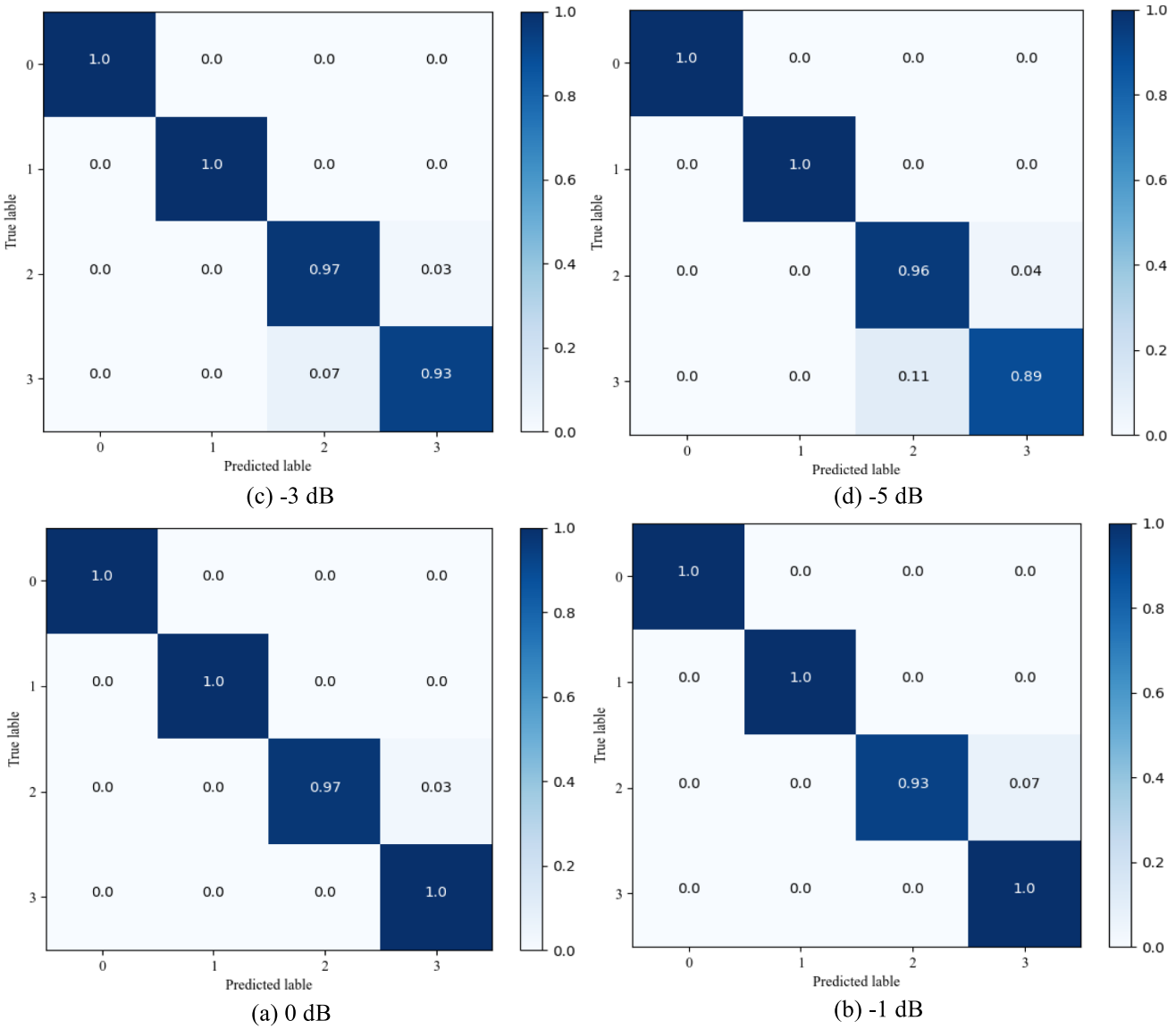


Fig. 19 Confusion matrix results of the compound fault dataset at varying noise levels: 0 dB, -1 dB, -3 dB, and -5 dB

Table 8 Results of the impact of the proposed CSDRB and CLC-Inception

Method	CSDRB	CLC-Inception	Accuracy	Precision	Recall	F1-score
1	–	–	91.66%	91.60%	90.58%	90.64%
2	–	√	94.16%	95.27%	94.53%	94.31%
3	√	–	96.66%	96.88%	96.28%	96.52%
4	√	√	99.16%	99.19%	99.21%	99.19%

CLC-Inception. Method 1, serving as a baseline without either module, achieves the lowest performance metrics, with an Accuracy, Precision, Recall, and F1-score of 89.16%, 92.07%, 88.79% and 8806%, respectively. Method 2 only contains CLC-Inception, resulting in a significant increase in performance compared to the baseline; the accuracy rises from 89.16% to 9416%, and the F1-score escalates from 88.06% to 94.31%. This improvement emphasizes the substantial impact of CLC-Inception on the ability of the model to extract more useful features. Similarly, Method 3, which only utilizes CSDRB, markedly enhances performance across all metrics, achieving, e.g., an Accuracy of 96.66% and an F1-score of 96.63%. When CSDRB is used independently, the performance improvement is even more noticeable, exhibiting the effectiveness of CSDRB in dynamically recalibrating features for fault identification. Method 4, incorporating both CSDRB and CLC-Inception, has the greatest Accuracy, Precision, Recall, and F1-score of 99.16%, 99.19%, 99.21%, and 99.19%. This suggests that integrating both modules strengthens generalization, adaptability, and the ability of the model to capture various crucial features.

The t-SNE, illustrated in Fig. 20, is utilized to visualize the classification results of all four methods and evaluate the impact of CSDRB and CLC-Inception on performance. In Fig. 20a Method 1, excluding both CSDRB and CLC-Inception, results in poor class separability, characterized by overlapping clusters and unclear class borders. Figure 20b and c illustrates Method 2 and Method 3, respectively, involving CLC-Inception while excluding CSDRB and vice versa. The classification of faults reveals notable enhancements. Eventually, Fig. 20d highlights that Method 4, which combines CSDRB with CLC-Inception, achieves the best separability with distinct feature clusters, indicating its superior performance to all the other methods.

In summary, the results confirm that implementing either CSDRB or CLC-Inception solely helps the model perform better compared to the baseline. However, the combination of both modules in Method 4 makes the biggest difference, showing their complementary roles in improving classification accuracy and the general robustness of the model. According to this study, combining advanced multiple-feature extraction and dynamic attention mechanisms is necessary for achieving state-of-the-art performance.

4.2.5 (5) Comparison with recent fault diagnosis methods

To thoroughly evaluate the effectiveness of the proposed approach, several state-of-the-art methods from recent literature are selected for comparative analysis, including MBSDCN, GTFE-Net, and HF-MSCN. Specifically, MBSDCN [41] employs a multibranch feature segmentation strategy to enhance feature representation and improve model learning efficacy. GTFE-Net [42] implements a Gramian-based noise reduction strategy that transforms vibration signals into 2D matrices, multiplies them by their Gramian to amplify low-rank periodic components, and incorporates the denoised output with raw and frequency domain inputs into a three-branch CNN for bearing fault diagnosis. HF-MSCN [40] combines frequency domain and time domain filtering methods to address fault detection challenges in high-noise industrial environments with rotating machinery.

The experimental results are in Table 9. In FD experiments under high-noise conditions with a -5 dB SNR, the proposed model achieves a significantly higher accuracy of 96.66%, demonstrating superior noise robustness and feature recognition capability. In contrast, methods such as MBSDCN, GTFE-Net, and HF-MSCN exhibit various limitations in handling high noise interference or limited data. The proposed design greatly surpasses current SE and MS-CNN algorithms by incorporating both the CLC-Inception and CSDRB modules. CLC-Inception enhances multiscale feature representation by cross-layer channel recalibration, whereas CSDRB incorporates channel and spatial attention processes to more accurately highlight discriminative features. This dual recalibration strategy markedly exceeds the representational capacity of SE attention and multiscale CNN techniques by directly and jointly augmenting both spatial and temporal feature significance. Collectively, these innovations constitute clear theoretical distinctions and practical advantages. As a result, the proposed framework indicates strong adaptability and accuracy in complex industrial scenarios.

4.3 Discussion

Rolling bearings often operate under highly variable loads and speeds, generating vibration signals that can be subtle, transient, and easily obscured by industrial noise,

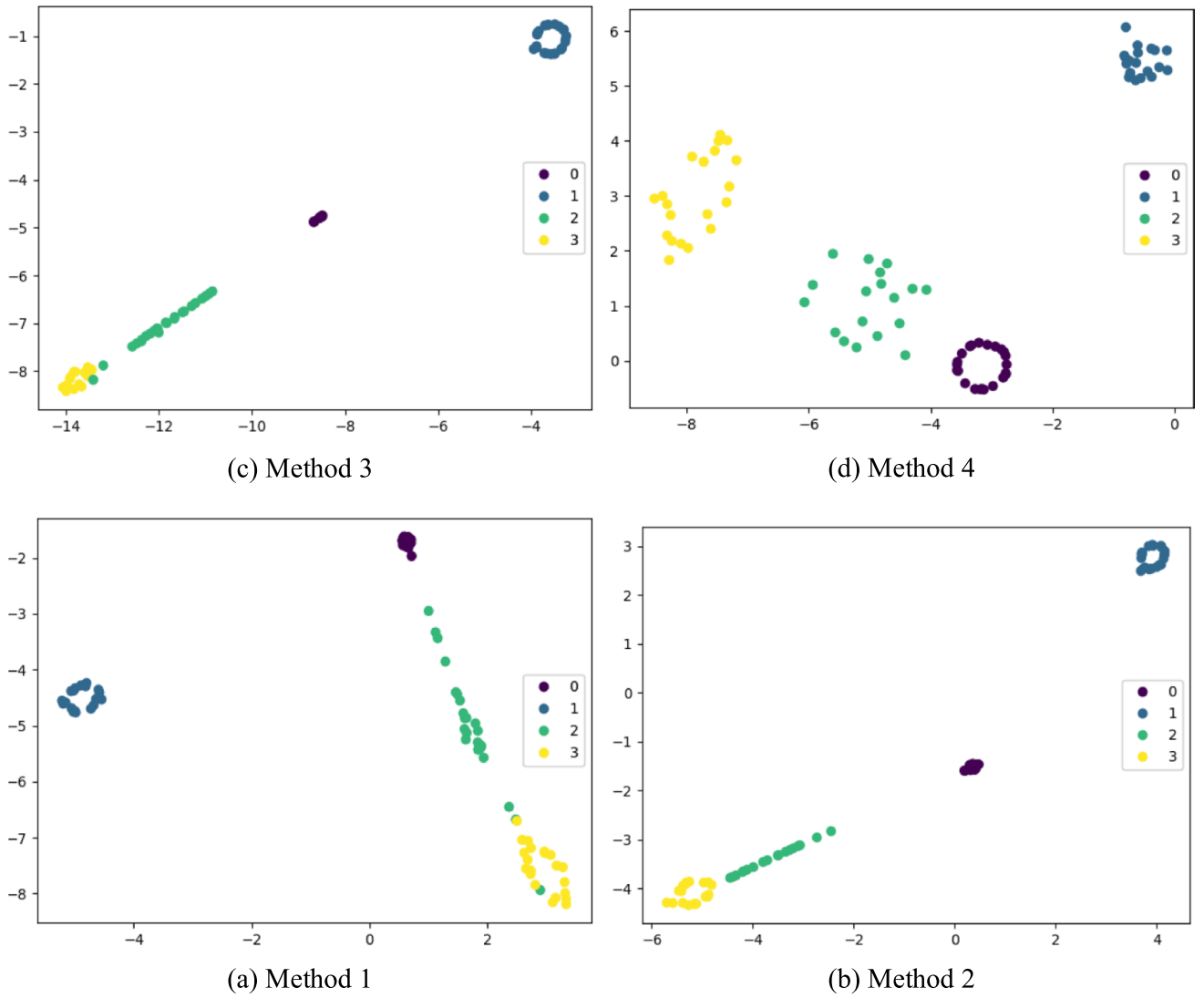


Fig. 20 Visualization results of the impact of the proposed CSDRB and CLC-Inception: (a) Method 1, (b) Method 2, (c) Method 3, and (d) Method 4

Table 9 Comparison results of different methods on the compound fault dataset under noisy environments

Methods	Accuracy
MBSDCN	88.71%
GTFE-Net	90.45%
HF-MSCN	93.01%
Proposed method	96.66%

especially when only small numbers of failure samples are available for training. Advanced CNN-based methods struggle in these conditions because fixed-scale convolutions cannot simultaneously capture both fine-grain fault and broader waveform distortions, and their feature maps can be overwhelmed by background noise. The proposed CLC-Inception directly addresses these limitations by employing parallel 1D convolutions at multiple scales to detect both short-duration impacts and longer-duration envelope variations, while the integrated SE block enhances the relevance of each scale's features. The subsequent CSDRB further modifies these representations by adaptively reweighting channels and temporal segments that correspond to incipient faults, thus isolating critical diagnostic cues even when signal-to-noise ratios are low. The proposed method is validated on both small sample and high-noise conditions across two benchmark datasets, yielding significant accuracy improvements—up to 99.34% in data-limited scenarios and 96.16% under severe noise with small data—over four state-of-the-art methods. All reported classification accuracies are averages from ten independent training runs, each using the same predefined partitioning of the training and testing samples as described in Tables 2 and 5. These results demonstrate that

our method is not only justified but essential for reliable bearing FD in challenging real-world environments.

5 Conclusion

To address the challenges of limited training data and high noise levels in rolling bearing FD, a new end-to-end framework that integrates CLC-Inception and CSDRB to efficiently capture spatial and temporal characteristics inherent in vibration signals, which significantly enhances diagnostic accuracy under complex environments. The CLC-Inception is specifically designed to extract and recalibrate multiscale, channel-wise representations, thereby addressing the problem of feature underrepresentation when sample sizes are small, while convolutional layers enhanced by residual connections strengthen the network's ability to learn discriminative fault features. The CSDRB solves the issue of irrelevant feature dominance by adaptively emphasizing the most informative spatial–temporal cues, which helps the model improve its capability to identify deep discriminative characteristics. A Softmax classifier is finally applied to the refined feature set for accurate fault classification.

Experimental results from the public CWRU dataset and a self-constructed composite bearing dataset indicate that the proposed method outperforms four advanced techniques under both data scarcity and robust noise interference. The proposed method demonstrates high diagnostic accuracy in the tested scenarios; however, it has yet to be assessed under various load and speed conditions of real-world industrial operations. Future research will incorporate advanced transfer learning techniques alongside domain adaptation, allowing the model to acquire invariant features that generalize effectively under varying operating parameters. The incorporation of these strategies aims to strengthen the framework's generalization performance under varying load and speed conditions.

Acknowledgements This work was funded by the National Natural Science Foundation of China (Grant Nos. 62163023, 62063020), the Youth Science and Technology Fund of Gansu Province (22JR5RA808), Long yuan Young Innovation Talent Team Program in Gansu Province (310100296012).

Data availability Data are available upon request to the corresponding author.

Declarations

Conflicts of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Sahu D, Dewangan RK, Matharu SPS (2024) An investigation of fault detection techniques in rolling element bearing. J Vib Eng Technol 12(4):5585–5608
2. Xie Z, Chen J, Feng Y, Zhang K, Zhou Z (2022) End to end multi-task learning with attention for multi-objective fault diagnosis under small sample. J Manuf Syst 62(October 2021):301–316. https://doi.org/10.1016/j.jmsy.2021.12.003
3. Qiang R, Zhao X (2025) Multiscale bidirectional transformer network for rolling bearing fault diagnosis. J Braz Soc Mech Sci Eng 47(5):220
4. Luo Z, Pan S, Dong X, Zhang X (2025) Interpretable quadratic convolutional residual neural network for bearing fault diagnosis. J Braz Soc Mech Sci Eng 47(4):1–17. https://doi.org/10.1007/s40430-025-05457-5
5. Cai B et al (2020) Fault detection and diagnostic method of diesel engine by combining rule-based algorithm and BNs/BPNNs. J Manuf Syst 57(June):148–157. https://doi.org/10.1016/j.jmsy.2020.09.001
6. Chen G, Yan C, Meng J, Wang H, Wu L (2021) Improved VMD-FRFT based on initial center frequency for early fault diagnosis of rolling element bearing. Meas Sci Technol 32(11):115024
7. Zhao H et al (2023) Intelligent diagnosis using continuous wavelet transform and Gauss convolutional deep belief network. IEEE Trans Reliab 72(2):692–702. https://doi.org/10.1109/TR.2022.3180273
8. Yao R, Zhao H, Zhao Z, Guo C, Deng W (2024) Parallel convolutional transfer network for bearing fault diagnosis under varying operation states. IEEE Trans Instrum Meas 73:1–13. https://doi.org/10.1109/TIM.2024.3480212
9. Yang X et al (2024) Multiple IGBT Open-circuit fault diagnosis strategy for MMC using feature reconstruction based recurrent neural network. IEEE J Emerg Sel Top Power Electron 12(6):5625–5637. https://doi.org/10.1109/JESTPE.2024.3432189
10. Chen Y, Qiang Y, Chen J, Yang J (2024) FMRGAN: feature mapping reconstruction GAN for rolling bearings fault diagnosis under limited data condition. IEEE Sens J 24(15):25116–25131. https://doi.org/10.1109/JSEN.2024.3415713
11. Liao X, Ming X, Xia M (2024) KBRDBN: an interpretable deep belief network for the fault diagnosis of the trolley mechanism in ship-to-shore cranes. IEEE Trans Instrum Meas 73:1–12. https://doi.org/10.1109/TIM.2023.3318717
12. Ye Z, Yu J (2021) Akrnet: a novel convolutional neural network with attentive kernel residual learning for feature learning of gearbox vibration signals. Neurocomputing 447:23–37. https://doi.org/10.1016/j.neucom.2021.02.055
13. Sun J, Ding H, Li N, Sun X, Dong X (2024) Intelligent fault diagnosis of hydraulic system based on multiscale one-dimensional convolutional neural networks with multiattention mechanism. Sensors (Basel). https://doi.org/10.3390/s24227267
14. Wang C, Xu Z (2021) An intelligent fault diagnosis model based on deep neural network for few-shot fault diagnosis. Neurocomputing 456:550–562. https://doi.org/10.1016/j.neucom.2020.11.070
15. Yang J, Liu J, Xie J, Wang C, Ding T (2021) Conditional GAN and 2-D CNN for bearing fault diagnosis with small samples. IEEE Trans Instrum Meas 70:1–12. https://doi.org/10.1109/TIM.2021.3119135
16. Bai R, Xu Q, Meng Z, Cao L, Xing K, Fan F (2021) Rolling bearing fault diagnosis based on multi-channel convolution neural network and multiscale clipping fusion data augmentation. Meas. J. Int. Meas. Confed. 184(February):109885. https://doi.org/10.1016/j.measurement.2021.109885

17. Yin H, Li Z, Zuo J, Liu H, Yang K, Li F (2020) Wasserstein generative adversarial network and convolutional neural network (WG-CNN) for bearing fault diagnosis. Math Probl Eng. https://doi.org/10.1155/2020/2604191
18. Kumar A, Vashishtha G, Gandhi CP, Zhou Y, Glowacz A, Xiang J (2021) Novel convolutional neural network (NCNN) for the diagnosis of bearing defects in rotary machinery. IEEE Trans Instrum Meas. https://doi.org/10.1109/TIM.2021.3055802
19. Ma L, Shi F, Wu Z, Peng K (2023) A Survey of few-shot learning-based compound fault diagnosis methods for industrial processes, IEEE 6th International conference on industrial cyber-physical systems. 1–4, https://doi.org/10.1109/ICPS58381.2023.10128105.
20. Dong Y, Li Y, Zheng H, Wang R, Xu M (2022) A new dynamic model and transfer learning based intelligent fault diagnosis framework for rolling element bearings race faults: solving the small sample problem. ISA Trans 121:327–348. https://doi.org/10.1016/j.isatra.2021.03.042
21. Su H, Xiang L, Hu A, Xu Y, Yang X (2022) A novel method based on meta-learning for bearing fault diagnosis with small sample learning under different working conditions. Mech Syst Signal Process 169(December):108765. https://doi.org/10.1016/j.ymssp.2021.108765
22. Wu J, Zhao Z, Sun C, Yan R, Chen X (2020) Few-shot transfer learning for intelligent fault diagnosis of machine. Measurement 166:108202. https://doi.org/10.1016/j.measurement.2020.108202
23. Wang Y, Yan J, Sun Q, Jiang Q, Zhou Y (2020) Bearing intelligent fault diagnosis in the industrial internet of things context: a lightweight convolutional neural network. IEEE Access 8:87329–87340
24. Huang YJ, Liao AH, Hu DY, Shi W, Bin Zheng S (2022) Multi-scale convolutional network with channel attention mechanism for rolling bearing fault diagnosis. Meas. J. Int. Meas. Confed. 203(August):111935. https://doi.org/10.1016/j.measurement.2022.111935
25. Mekruksavanich S, Jitpattanakul A, Sitthithakerngkiet K, Youplao P, Yupapin P (2022) Resnet-SE: channel attention-based deep residual network for complex activity recognition using wrist-worn wearable sensors. IEEE Access 10:51142–51154. https://doi.org/10.1109/ACCESS.2022.3174124
26. Han S, Sun S, Zhao Z, Luan Z, Niu P (2024) Deep residual multiscale convolutional neural network with attention mechanism for bearing fault diagnosis under strong noise environment. IEEE Sens J 24(6):9073–9081. https://doi.org/10.1109/JSEN.2023.3345400
27. Liu B, Yan C, Liu Y, Wang Z, Huang Y, Wu L (2023) Multiscale residual antinoise network via interpretable dynamic recalibration mechanism for rolling bearing fault diagnosis with few samples. IEEE Sens J 23(24):31425–31439. https://doi.org/10.1109/JSEN.2023.3328007
28. Szegedy C et al., (2015) Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition 07-12-June, pp. 1–9, https://doi.org/10.1109/CVPR.2015.7298594
29. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition., pp. 7132–7141
30. Wang L, Zou T, Cai K, Liu Y (2024) Rolling bearing fault diagnosis method based on improved residual shrinkage network. J Braz Soc Mech Sci Eng 46(3):1–12. https://doi.org/10.1007/s40430-024-04729-w
31. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778
32. Wei L, Peng X, Cao Y (2024) Enhanced fault diagnosis of rolling bearings using an improved inception-lstm network. Nondestruct. Test. Eval. 40(7):1–20
33. Zhang X, He C, Lu Y, Chen B, Zhu L, Zhang L (2022) Fault diagnosis for small samples based on attention mechanism. Measurement 187:110242
34. Smith WA, Randall RB (2015) Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study. Mech Syst Signal Process 64:100–131
35. Li A, Yao D, Yang J, Chang M, Zhou T (2023) Bearing diagnosis using an anti-noise neural network based on selectable branch multiscale modules and attention mechanisms (August 2023). IEEE Sens J 24(5):5830–5840
36. Zhang W, Li X, Ding Q (2019) Deep residual learning-based fault diagnosis method for rotating machinery. ISA Trans 95:295–305
37. Liang H, Cao J, Zhao X (2022) Multiscale dynamic adaptive residual network for fault diagnosis. Measurement 188:110397
38. Wang H, Liu Z, Peng D, Qin Y (2019) Understanding and learning discriminant features based on multiattention 1DCNN for wheelset bearing fault diagnosis. IEEE Trans Ind Inform 16(9):5735–5745
39. Van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9(11):2579–2605
40. Abduelhadi A, Liang H, Cao J, Chen P (2024) HF-mscn: a high frequency-multiscale cascade network for bearing fault diagnosis. Meas Sci Technol 35(11):116120
41. Liang H, Cao J, Zhao X (2023) Multibranch and multiscale dynamic convolutional network for small sample fault diagnosis of rotating machinery. IEEE Sens J 23(8):8973–8988
42. Jia L, Chow TWS, Yuan Y (2023) GTFE-net: a gramian time frequency enhancement CNN for bearing fault diagnosis. Eng Appl Artif Intell 119:105794

Spiking Multi Scale Graph Convolutional Neural Networks for Fault Diagnosis of Rotating Machinery

Alaeldden Abduelhadi*
School of Electrical and Information Engineering
Lanzhou University of Technology
Lanzhou, China
alaeldden9090@gmail.com

Jie Cao
School of Electrical and Information Engineering
Lanzhou University of Technology
Lanzhou, China
haop1115@163.com

Haopeng Liang
School of Computer and Communication
Lanzhou University of Technology
Lanzhou, China
928782706@qq.com

***Abstract*—To address the challenges faced by deep learning in processing time-series data, especially the noise interference encountered when diagnosing bearing fault signals, this paper introduces an innovative Spiking Multiscale Graph Convolutional Neural Network (SMGCNN) method. This method takes advantage of the superiority of Spiking Neural Networks (SNNs) in dynamic time processing, enhancing the model's ability to handle signals that change over time. Then, a multiscale graph convolution is designed to capture the correlation between features at different scales, which takes the output features of convolution as graph nodes to construct graph structural relationships. Through empirical verification on the bearing fault data from the University of Paderborn and the publicly available bearing data from Southeast University, the results show that SMGCNN exhibits remarkable accuracy in bearing fault diagnosis tasks, especially in low signal-to-noise ratio environments, where it demonstrates high accuracy and robustness.**

Keywords-Rolling bearing; fault diagnosis; deep learning; spiking multiscale graph convolution neural network

I. INTRODUCTION

With the advancement of industrial technology, rotating equipment has rapidly evolved, and its operational status is essential for industrial production. Rolling bearings, as critical components, directly determine whether mechanical parts function normally [1], [2]. Bearings, constantly subjected to load as they rotate, eventually fail due to fatigue after cyclic loading and unloading [3]. Bearing failures can lead to severe vibration and noise in machinery, consequently degrading equipment quality and safety [4]. Timely and accurate monitoring of rolling bearings and accurate fault identification in noisy environments are crucial for ensuring the safe and reliable operation of equipment and avoiding economic losses [5].

Considering the rapid development of technology, neural networks are increasingly applied in fault diagnosis. Their robust data processing capabilities and self-learning characteristics offer novel solutions for complex system fault diagnosis. Convolutional Neural Networks (CNN) have shown significant advantages in handling multi-sensor data, noisy labels, and model comparisons. Zhang et al. [6] proposed a deep learning method based on CNN to address the information loss in bearing vibration data processing due to non-linear and non-stationary characteristics. Verstraete et al. [7] suggested a method that uses time-frequency representations to convert raw data into images for classification and fault diagnosis using deep CNN, addressing the uncertainties and biases introduced by traditional feature extraction and selection in system analysis. Literature [8] proposed a method combining data and regression characteristics for wind turbine bearing fault diagnosis, addressing the increasing fault rates, rising maintenance costs, and deteriorating safety conditions due to equipment aging. Literature [9] introduced a method that learns effective fault features directly from raw vibration signals and performs feature extraction and classification simultaneously in an end-to-end system, addressing the separate design of feature extraction and classification in traditional fault diagnosis methods that require additional signal processing and expertise. These studies somewhat resolve traditional feature extraction issues and offer efficient feature learning and time-frequency representation conversion advantages. However, it lacks vitality when dealing with tasks involving time-series data and dynamic environments.

To address these issues, various methods for handling time-series data, including Recurrent Neural Networks (RNN), Transformer, Spiking Neural Networks (SNN), etc., have provided alternative solutions. Among them, SNN has gained popularity among researchers as an emerging neural network. Lin et al. [10] presented an improved Spiking Neural Network to enhance the diagnostic performance for inter-shaft bearing faults, thereby addressing certain limitations of traditional methods. WANG et al. [11] proposed the application of transformer neural networks to enhance the accuracy and timeliness of bearing fault diagnosis, mitigating the negative impact of bearing faults on the performance and reliability of mechanical systems. Xuan et al. [12] proposed a Wavelet Gradient Integrated Spiking Neural Network (WGI-SNN) framework for diagnosing complex correlations between multiple single-point defects. Literature [13] introduced a method that converts features extracted using Local Mean Decomposition (LMD) into probabilistic spike sequences and incorporates multi-layer learning algorithms to facilitate multi-layer training of SNNs, addressing the difficulty in expanding SNNs to deep learning due to their internal state and error function's lack of continuity and differentiability. Wang et al. [14] proposed a novel fault diagnosis method for power systems using an improved Adaptive Fuzzy Spiking Neural P system (AFSN P systems)

and Particle Swarm Optimization (PSO) algorithm to enhance the efficiency and accuracy of power system diagnostics. These studies collectively focus on utilizing advanced neural network models like SNNs and Transformer Neural Networks to improve fault detection efficiency. Despite progress, challenges remain in adapting to complex operational conditions for accuracy requirements.

In existing research, models tend to excel either in feature extraction or in time-series processing; nevertheless, when both aspects are considered, the diagnostic accuracy of the model is significantly compromised in the presence of noise in the environment. Extracting useful features from a large amount of raw data and selecting the most effective features for fault diagnosis is a significant challenge. Simultaneously, obtaining contextual information and correlations in time-series data is particularly important. CNNs succeed at extracting spatial features, but in real-world applications, fault diagnosis systems need to be able to handle different types of equipment and various operating conditions. This raises concerns about model complexity and adaptability to time-series data. SNNs, as powerful time-series neural networks, are effective in handling time-series data, yet their deficiency in extracting complex spatial features is efficiently supplemented by CNN. Considering this, our study introduces a method that combines CNNs and SNNs, aiming to fully exploit their advantages in spatial and temporal feature extraction. This approach seeks to enhance the accuracy and efficiency of bearing fault diagnosis, particularly in handling signals with complex spatial patterns and dynamic temporal characteristics. Here are the contributions made by the proposed method:

1. This article introduces an improved SMGCNN method, achieving high-quality fault diagnosis.
2. A novel Multiscale Graph Convolution (MGC) is proposed. MGC regards the output features of each convolution as graph nodes and constructs graph structural relationships for multiscale features, so as to enhance the model's ability to capture the correlations between features at different scales.
3. Experimental validation is conducted on two datasets, demonstrating the effectiveness of the method.

The subsequent chapters outline the existing methods in the second section, the third section provides a detailed description of the proposed method along with a method framework diagram, the fourth section conducts experimental validation on the proposed method, and the fifth section summarizes this work and provides prospects for future research.

II. RELATED WORKS

In recent years, deep learning technologies have been extensively applied in the field of fault diagnosis. Particularly, CNNs have become a focal point of research due to their outstanding capabilities in feature extraction and pattern recognition.

A. Application of Convolution Neural Networks in Fault Diagnosis

CNNs are a type of deep learning architecture especially suited for processing data with grid-like topologies, such as images and time-series data. CNNs achieve effective feature extraction and pattern recognition by learning the hierarchical features of data [15]. In the realm of fault diagnosis, CNNs can automatically extract useful features from raw sensor data. This process typically involves the following steps:

Convolutional Layers: These layers extract local features from the input data through convolution operations, represented as:

$$F(I,K) = \sum_{i=1}^{m}\sum_{j=1}^{n} I(i,j)\cdot K(i,j) \tag{1}$$

where F is the output of the convolution operation, I is the input data, K represents the convolution kernel, and m and n are the dimensions of the kernel.

Activation Functions: Activation functions introduce non-linearity to the network, enabling it to learn more complex features. A common activation function is the Rectified Linear Unit (ReLU) [16]:

$$\mathrm{ReLU}(x) = \max(0, x) \tag{2}$$

Pooling Layers: Pooling layers reduce the dimensionality of features, decreasing computation and preventing overfitting. A common pooling operation is max pooling:

$$P(x) = \max_{i \in \text{window}} x_i \tag{3}$$

where $P(x)$ is the output of the pooling operation, and x_i are the pixel values within the pooling window. By stacking and combining these layers, CNNs learn high-level feature representations from raw data, which is crucial for fault diagnosis. For instance, CNNs can extract fault characteristics from the vibration signals of mechanical equipment, facilitating the effective identification of various fault types [17].

In subsequent sections, we will explore the application of SNNs in fault diagnosis and compare their strengths and limitations with CNNs in addressing fault diagnosis issues.

B. Application of Convolution Neural Networks in Fault Diagnosis

SNNs are a type of neural network that mimic the working mechanism of biological neural systems [18], particularly excelling in processing time-related data. Unlike traditional deep learning models, neurons in SNNs can emit spike signals at specific time points, thereby more effectively processing dynamic and time-varying information. In fault diagnosis, SNNs can effectively process time-series data obtained from sensors, such as vibration or sound signals [19]. The basic working principle includes:

Spike Encoding: Converting analog signals into spike sequences. Rate encoding is a typical method where the frequency of spikes is proportional to the signal's intensity:

$$r(t) = \frac{1}{T}\sum_{i=1}^{N}\delta(t - t_i) \tag{4}$$

where $r(t)$ denotes the spike rate at time t, δ is the Dirac delta function, t_i represents the firing time of the $i-th$ spike, T is the observation window length, and N is the total number of spikes.

Temporal Dynamics Processing: Neurons in SNNs adjust their membrane potential dynamically based on the timing of input spikes. Neurons emit a spike and reset their state when the membrane potential exceeds a certain threshold. The dynamics of a neuron are given by the following equation:

$$\tau \frac{dV(t)}{dt} = -V(t) + I_{syn}(t) \tag{5}$$

where $V(t)$ is the membrane potential, τ donates the time constant, and $I_{syn}(t)$ is the synaptic input.

Learning and Adaptation: SNNs learn specific input patterns by adjusting synaptic weights, which is often accomplished through Spike-Timing-Dependent Plasticity (STDP) mechanisms:

$$\Delta w = \begin{cases} A_+ \exp\left(\dfrac{-\Delta t}{\tau_+}\right), & \text{if } \Delta t > 0 \\ -A_- \exp\left(\dfrac{\Delta t}{\tau_-}\right), & \text{if } \Delta t < 0 \end{cases} \tag{6}$$

where Δw represents the change in synaptic weight, Δt stands for the time difference between pre and post-neuron pulses, A_+ and A_- denote positive and negative learning rates, while τ_+ and τ_- are time windows.

One key advantage of SNNs in the field of fault diagnosis applications is their high sensitivity to temporal information, enabling them to effectively identify and classify complex fault patterns based on time-series data. Additionally, SNNs often exhibit better energy efficiency compared to traditional CNN models, which is particularly crucial in applications on embedded and mobile devices. However, the training and optimization of SNNs are more intricate compared to traditional deep learning models which to some extent limits their widespread adoption in industrial applications. Future research may focus on simplifying the training process of SNNs and enhancing their generalization capabilities.

III. METHOD OF PULSE CONVOLUTIONAL NEURAL NETWORKS FOR FAULT DIAGNOSIS

Current fault diagnosis methods often exhibit insufficient comprehensiveness in capturing time-series information, and there is a lack of effective correlation in capturing contextual information. Traditional SNNs mainly focus on simulating the dynamics of time, but their feature extraction capability is relatively limited when dealing with complex signals such as bearing fault signals. CNNs have strong feature extraction capabilities on time-series data, capable of automatically identifying and learning important features. This discovery inspired us to develop a new fault diagnosis method that excels in both dynamic behavior processing and feature extraction [20]. The overall structural framework is shown in Figure 1.

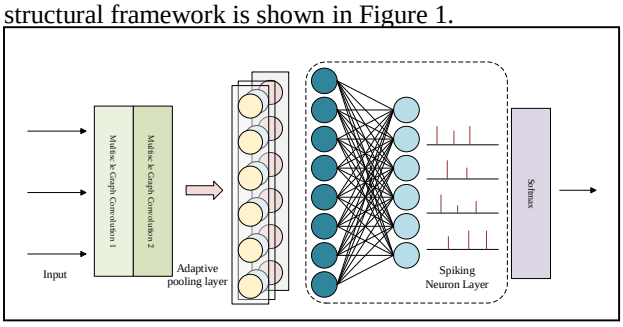


Figure 1. Fault diagnosis model.

A. Spiking Neural Unit

This research utilizes Spiking Neuron Units, which are essentially designed for processing time-series data and efficiently identifying dynamic patterns over time. In fault diagnosis, SNN can capture subtle changes in equipment states over time, which may indicate potential faults. Compared to traditional deep neural networks, SNN transmits information through pulse signals (rather than continuous numerical values) and typically requires lower computational resources. Its unique structure and dynamic properties enable it to recognize complex and nonlinear patterns, which is particularly advantageous in handling complex mechanical systems and diagnosing hidden faults.

As shown in Figure 1, the network employs two layers of SNN neural units, and these layers use the Sigmoid activation function to approximate the threshold excitation mechanism of neurons, allowing the network to process information in a more biologically consistent manner. This processing allows the network to effectively integrate input signals at various time scales, thereby decoding the transient and cumulative features in bearing fault signals.

B. Multisacle Graph Convolution

Multiscale convolution extracts feature at different scales through convolution kernels of different sizes in the same layer. In fault diagnosis, it can capture fault features in different ranges, which helps to identify complex patterns. However, some modules only aggregate and fuse features at several scales, analyze the features of each scale independently, and ignore the similarities or dependencies between features at different scales, making it difficult to meet the requirements of complex fault feature analysis. To address this problem, this paper proposes Multiscale graph convolution, which takes the output features of convolution as graph nodes and constructs graph structural relationships to enhance the capture of correlations between features at different scales. The feature extraction process is divided into three steps, and the structure is shown in Figure 2.

Step 1: Multiscale convolutional feature extraction. The input features are split into multiple sub-features in the channel dimension. For each sub-feature, convolution operations with different kernel sizes are performed to extract features at different time scales. The convolution process includes one-dimensional convolution, batch normalization, and ReLU activation function. These

multiscale features contain rich fault information, and it is necessary to establish dependencies among them for fault identification.

Step 2: Construction of Feature Graph Structure. To fully explore the interrelations among multiscale features, this study constructs a graph structure where each node represents the feature embedding extracted from a specific scale. Specifically, Global Average Pooling (GAP) is applied to each scale-specific feature map to compress the spatial dimensions and generate a one-dimensional global descriptor vector. The resulting descriptor H_i from the *i*-th scale serves as a node in the graph structure. Then, all node vectors are then stacked to form the node feature matrix $H \in \mathbb{R}^{N\times d}$, where N is the number of nodes, and d is the dimensionality of each node feature. To define the connections between nodes, the model adopts dot-product similarity as the similarity metric. The affinity between any two node features H_i and H_j is computed as:

$$A_{ij} = H_i \bullet H_j^T$$

where A_{ij} quantifies the similarity between nodes i and j. This computation yields a symmetric similarity matrix $A \in \mathbb{R}^{N\times N}$, where higher values indicate stronger correlations between multiscale features. This graph construction not only retains individual scale-specific representations but also captures their structural correlations, providing a solid foundation for subsequent graph convolution operations.

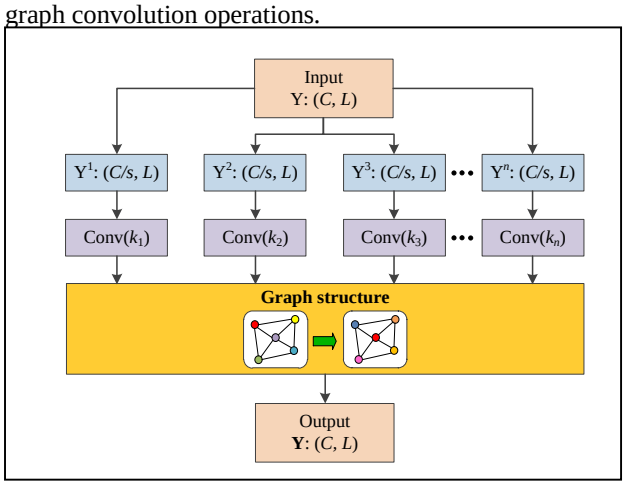


Figure 2. Multiscale graph convolution.

Step 3: Multiscale Feature Weighted Fusion. The Softmax function is applied to the adjacency matrix to calculate the attention weight matrix for normalization. The attention weights are used to perform a weighted sum of node features at different scales to update the node features. The weighted features of each node incorporate supplementary information from other scales. Finally, all the updated node features are concatenated along the channel dimension to obtain the final multiscale fusion features. These features include the convolution features at different scales and the results of adaptive fusion through the graph structure.

Multiscale graph convolution utilizes the "message-passing" mechanism in the graph network, allowing each scale feature to retain its own characteristics while capturing relevant information from other scale features, learning the dependencies between different scale features, and enhancing the ability to handle complex fault data.

C. SMGCNN Method for Fault Diagnosis

This research proposes the SMGCNN method, where the introduction of SNN aims to handle dynamic features in bearing signals more efficiently. CNNs, as a common deep learning tool, not only perform feature extraction but also provide rich and fine-grained inputs to the SNN layers. Through this combination, the network extracts spatial features while better utilizing the time dynamic processing advantages of SNN. This combination improves the ability to capture complex signal features, making the network more accurate in handling bearing fault diagnosis tasks.

Firstly, the network employs two multiscale graph convolution layers (MGC 1 and MGC 2) to process the bearing input signals. MGC 1 contains four multiscale convolution layers. The kernel sizes of different convolution layers are 3×1, 5×1, 7×1, and 9×1 respectively. Each convolution layer has 4 filters with the stride of 1. MGC 2 maps the features to 8 channels, and using the same kernel size and stride. This configuration helps extract more advanced feature representations.

$$\begin{aligned} &\text{First multiscale graph convolution: } \mathrm{c}_1 = \mathrm{MGC}(x) \\ &\text{Second multiscale graph convolution: } \mathrm{c}_2 = \mathrm{MGC}(\mathrm{c}_1) \end{aligned} \tag{7}$$

To reduce the number of parameters and improve the model's generalization ability, we introduced an adaptive average pooling layer that standardizes the length of each feature map to 64.

$$\text{Adaptive Average Poolying layer: } p = \mathrm{Pool}(c_2) \tag{8}$$

This adaptive mechanism permits the model to deal with input signals of different lengths while maintaining consistent output sizes. Subsequent to convolution and pooling, the signal is fed into two pulse neuron layers (layer1 and layer2). Layer1 converts the pooled feature maps into a hidden layer with a size defined by the user as a hidden size. Layer2 transforms the output of the hidden layer into the final class predictions with a size of number of classes.

$$\begin{aligned} &\text{First Pulse Neuron Layer(layer1): } h_1 = \sigma(W_h h_0 + b_h) \\ &\text{Second Pulse Neuron Layer(layer2): } h_2 = \sigma(W_0 h_1 + b_0) \end{aligned} \tag{9}$$

where σ is the Sigmoid activation function, W_h, W_o are weight matrices, b_h, b_o are bias terms, h_0 is the input feature, and h_1 and h_2 are the outputs of the two pulse neuron layers. In these two layers, we use the Sigmoid function as the activation function to simulate the excitation and inhibition processes of neurons.

$$\text{Time Dependency: } h_t = \sigma(W \cdot h_{t-1} + U \cdot a_t + b) \tag{10}$$

where h_t is the output of neurons at time t, h_{t-1} is the output at the previous time step, a_t is the input at the current

time step, W and U are weight matrices, and b is the bias term. This expression highlights the benefits of SNN in managing dynamic changes in bearing signals. Finally, within our network, the pulse neuron layers can be represented in the following mathematical form:

$$h_{\text{final}} = \sigma\left(W_{\text{snn}} \cdot \text{Pool}\left(\text{ReLU}\left(W_2 * c_2 + b_2\right)\right) + b_{\text{snn}}\right) \quad (11)$$

This innovation makes the network more suitable for handling bearing signals with complex patterns and dynamic changes.

IV. EXPERIMENT

To evaluate the diagnostic performance of the proposed method, the experimental data in this study consists of two parts, namely, the bearing dataset from the University of Paderborn and the bearing dataset from Southeast University.

Experiment 1 utilizes a set of bearing data provided by the University of Paderborn. The experimental setup, involves the collection of data, including vibration signals recorded during the operation of a gearbox. Experiment 2 employs bearing data from Southeast University, encompassing fault information related to vibration signals during bearing operation. These experiments are conducted to validate and demonstrate the effectiveness and accuracy of the proposed methodology, emphasizing its reliability across different datasets in the domain of deep learning-based bearing fault diagnosis.

A. Experiment 1

1) Experimental Dataset: The dataset used in Experiment 1, is sourced from the benchmark dataset for rolling bearing condition monitoring at Purdue University. This dataset has a sampling frequency of 64 kHz and a sampling duration of 4 seconds. It includes a variety of rolling bearing current signals and vibration signals recorded under different rotational speeds and loads. It is utilized for evaluating the performance of various machine learning algorithms in the context of rolling bearing fault diagnosis. In this study, we specifically use vibration signals from the same operating conditions, which encompass normal operation and four distinct fault states. These four fault states are: inner race fault 1, inner race fault 2, healthy, outer race fault 1, and outer race fault 2. These signals are employed to validate the performance of the proposed SMGCNN fault diagnosis model.

2) Experimental Results: After training the model on an assigned training dataset, it is evaluated on a testing dataset. Additionally, it is compared with several recent deep learning methods in the field of bearing fault diagnosis. The results are shown in Table I From the experimental results in Table I, various deep learning models are evaluated, including WDCNN, ADSAE, FFT-DNN, ResNet-Tanh, DBN, and SMGCNN. In comparison to existing methods, SMGCNN exhibits the highest accuracy. It has lower FLOPs of 5.3G compared to ADSAE and a relatively smaller parameter count of 0.003M. Its computational complexity falls within a moderately suitable range compared to other models. From Table I, it can be observed that the proposed method achieves state-of-the-art accuracy in deep learning-based bearing fault diagnosis. This is possibly due to the effective integration of features in its model architecture, enabling higher accuracy at a relatively lower computational complexity. ResNet-Tanh and DBN perform well, likely due to their deep structures, facilitating the learning of more complex features. The lower accuracy of WDCNN and FFT-DNN might be related to their architectures or training strategies, or they may require further tuning. The relatively poor performance of ADSAE could be attributed to excessive model complexity leading to overfitting or training instability.

TABLE I. TABLE OF DIAGNOSTIC ACCURACY RESULTS

Method	Accuracy	Flops(G)	Parameters(M)
WDCNN	95.3%	0.26	0.002
ADSAE	91.8%	12.37	0.018
FFT-DNN	95.6%	0.65	0.007
ResNet-Tanh	97.1%	0.64	0.002
DBN	97.0%	0.25	3.1
SMGCNN	99.1%	5.3	0.003

Figure 3 shows the confusion matrix. This confusion matrix consists of 5 categories, where each row represents the actual category, and each column represents the predicted category. For the "inner1" category, the model correctly predicted 171 instances of "inner1," with a correct classification rate of 98.28%. There are 3 instances of "inner1" wrongly predicted as "health," accounting for 1.72% of the total. For the "inner2" category, the model correctly predicted 177 instances of "inner2," achieving a 100% correct classification rate. In the "health" category, the model correctly predicted all 177 instances, achieving a 100% correct classification rate. For the "inner" category, the model correctly predicted 174 instances of "inner," with a 100% correct classification rate. In the "Outer2" category, the model correctly predicted 174 instances, with a correct classification rate of 98.28%. There are 1.72% of "Outer2" instances wrongly predicted as "Outer1." From Figure 3, it is evident that the model performs well in most categories, but there are some misclassifications in certain categories.

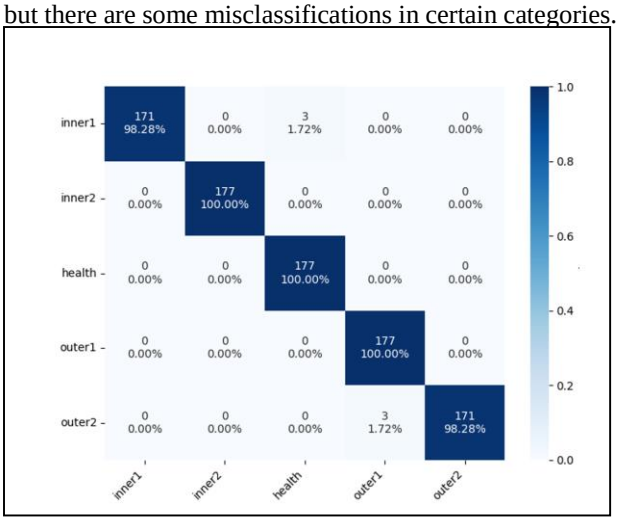


Figure 3. Confusion matrix of fault classification results in experiment 1

In real-world scenarios, the ability to attain stable diagnostics under certain sensor/industrial environment noise is crucial. Therefore, an evaluation of diagnostic performance is performed under Signal-to-Noise Ratio (SNR) ranging from -4 to 6 dB, as seen in Figure 4. In Figure 4, the experimental results comprise two sets of values: SNN_values and NOSNN_values. Throughout the experiments, the SNN_values gradually rise from 0.884 to 0.9832. This indicates that the model's performance progressively enhances with the incorporation of SNN, perhaps attributable to the advantages of SNN in learning and capturing data features. NOSNN_values are relatively lower but also increase gradually from 0.771 to 0.9666 as the experiments progress. In the absence of SNN, the model's performance also improves, possibly because of other components or the nature of the data itself.

Both SNN_values and NOSNN_values demonstrate a consistent upward trend, likely due to the model's learning process incrementally optimizing and adapting to the task's features. The SNN_values are relatively higher than NOSNN_values, suggesting that the utilization of SNN has a positive impact on performance. In certain experiments, SNN may have better adaptability to specific data features or task characteristics, resulting in better performance when using SNN. Figure 4 illustrates that the proposed method achieves diagnostic accuracy of over 95% even at a SNR ratio of 0, highlighting that the SNN approach designed in this paper effectively extracts signal features and exhibits high robustness when dealing with samples affected by noise. To validate the effectiveness of the proposed method, further ablation experiments are conducted to confirm the function of this model.

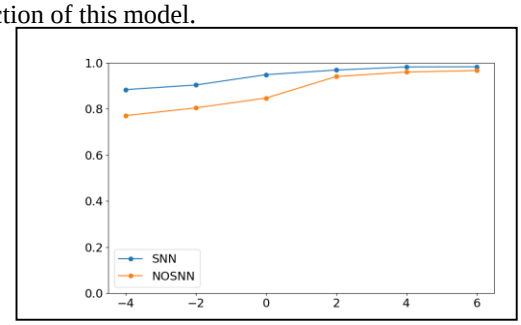


Figure 4. Diagnostic accuracy under different SNR

B. *Experiment 2*

1) Experimental Dataset: The experiment utilizes an openly available bearing dataset from Southeast University, which was collected through a dynamic simulation system within a power transmission chain. This test rig primarily consists of components such as a variable-speed motor, motor controller, planetary gearbox, parallel-axis gearbox, load, and load controller. It serves as a reliable source of real-world data for mechanical fault diagnosis in the context of bearing failures.

In this context, the dataset consists of two different operating conditions with rotational speeds of 30-2 and 20-2, corresponding to two types of faults: ball fault, comb fault, Health, inner fault, and outer fault. The data is collected at a sampling frequency of 10KHz with a sampling duration of 10 seconds. During the data processing, 22 time-domain and frequency-domain features are extracted for each sample, including mean, standard deviation, slope, energy, spectral peak frequency, and more. These feature extractions aim to fully exploit the information within the data and enhance the diagnostic performance of fault detection. The same diagnostic model is experimented with four different sample sizes, each consisting of 1050 individual samples. There are a total of 10 categories, comprising 9 fault types and 1 normal condition. Therefore, in each experimental group, the total number of samples is 10500. This division is carried out randomly in a 4:1 ratio between training and testing samples. This division is performed to evaluate the accuracy of the proposed fault diagnostic models under varying sample sizes.

2) Experimental Results: A training dataset is used to train the model, which is then evaluated on a test dataset. Furthermore, comparisons are done with some contemporary deep learning algorithms in the field of bearing fault diagnosis, and the results are shown in Table III.

The experimental results in Table II involve the performance comparison of different deep learning models, including WDCNN, ADSAE, FFT-DNN, ResNet-Tanh, DBN, and SMGCNN. Here is a comprehensive analysis of these results and possible reasons: SMGCNN exhibits the best performance, achieving an accuracy of 98.41%, which is the highest among all models. ResNet-Tanh, DBN, and FFT-DNN also perform relatively well, with accuracies of 95.13%, 95.24%, and 94.61%, respectively.

TABLE II. TABLE OF DIAGNOSTIC ACCURACY RESULTS

Method	Accuracy	Flops(G)	Parameters(M)
WDCNN	93.23%	0.36	0.003
ADSAE	91.38%	13.37	0.028
FFT-DNN	94.61%	0.85	0.017
ResNet-Tanh	95.13%	0.84	0.022
DBN	95.24%	0.45	4.1
SMGCNN	98.41%	7.3	0.037

The confusion matrix is an essential tool for evaluating the performance of a classification model. The confusion matrix, as shown in Figure 5, is a 10-class matrix, with each row representing the actual category and each column denoting the predicted category. For the "comb_20" category: The model correctly classifies 207 instances of "comb_20," with an accuracy of 98.57%. Three instances of "comb_20" are misclassified as "inner_20," representing 1.43% of the total. For the "outer_20" category: The model correctly classifies 201 instances of "outer_20," with an accuracy of 95.71%. However, there are 9 instances of "outer_20" that are incorrectly classified as "heath_20," making up 4.29% of the total. For the "ball_30" category: The model correctly identifies 198 instances of "ball_30," resulting in an accuracy of 95.65%. Three instances of "ball_30" are mistakenly identified as "inner_30", while 9 instances are misclassified as "outer_30". For the "inner_30" category: The model accurately identifies 201 instances of "inner_30," with an accuracy of 95.71%. There

are 3 instances of "inner_30" that are misclassified as "ball_30," and 6 instances are incorrectly classified as "health_30." In the "Outer_30" category, the model correctly classifies 207 instances as "outer," attaining a 98.57% accuracy rate in this classification. Nevertheless, there is a 1.43% misclassification rate where some "outer" instances are mistakenly categorized as "inner_30."

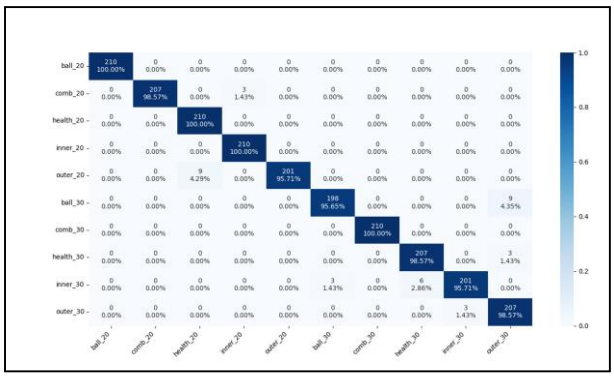


Figure 5. Confusion matrix of fault classification results in experiment 2

In Figure 6, it can be observed that in this set of experiments, SNN_values and NOSNN_values are provided within the SNR range of -2 to 6. As SNR gradually increases from -2 to 6, SNN_values show a corresponding incremental trend, rising from 0.9429 to 0.9732. It is noteworthy that SNN exhibits a more pronounced improvement in performance under low SNR ratio conditions. Similarly, when the SNR progressively rises from -2 to 6, NOSNN_values also reveal a gradual upward trend, rising from 0.8474 to 0.9566. Under low SNR ratio conditions, the performance of the model without SNN improves slightly, although not considerably. SNN_values perform better overall compared to NOSNN_values, with a more pronounced improvement, particularly in low SNR ratio conditions. This suggests that SNN has a certain advantage in handling noise within signals. As the SNR ratio increases, the model's performance significantly improves, primarily because higher SNR ratio conditions make it easier for the model to learn distinct signal features. The pulse neural network structure of SNN exhibits a distinct advantage in dealing with complex signals, including noise, which is more pronounced under low SNR ratio conditions. The proposed method can achieve diagnostic accuracy of over 95% even in 0 SNR conditions, illustrating the effectiveness of the SNN approach developed in this study in successfully extracting signal features and exhibiting substantial robustness to noise.

To comprehensively evaluate the effectiveness of the proposed approach, several state-of-the-art methods published in recent literature are chosen for comparative analysis, including MCAMDN, GTFE-Net, MSDARN, MBSDCN, and HF-MSCN. Specifically, MSDARN [21] utilizes a dynamic attention mechanism combined with multi-scale convolutional neural networks to achieve robust performance. MBSDCN [22] implements a multi-branch feature segmentation strategy to enrich feature representation and optimize model learning capability. Meanwhile, HF-MSCN [23] combines frequency-domain and time-domain filtering techniques to address fault detection challenges in high-noise industrial environments involving rotating machinery. The experimental results in Table III. In the fault diagnosis experiments under high-noise conditions, the SMGCNN model achieved a significantly higher accuracy of 94.29%, demonstrating superior noise robustness and feature recognition capability. In contrast, methods such as MSDARN, MBSDCN, and HF-MSCN exhibited various limitations in handling multiscale features or high-frequency noise. By integrating the temporal modeling capability of spiking neural networks with the structural awareness of multiscale graph convolution, SMGCNN effectively suppresses low-relevance noise connections through a graph enhancement mechanism and captures critical fault features. As a result, SMGCNN exhibits strong adaptability and accuracy in complex industrial scenarios.

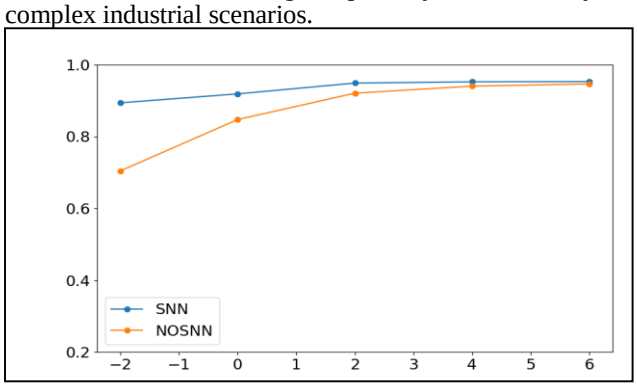


Figure 6. Diagnostic accuracy at different SNR

TABLE III. COMPARISON RESULTS WITH ADVANCED METHODS

Method	Accuracy
MSDARN	85.50 %
MBSDCN	90.89 %
HF-MSCN	91.30 %
SMGCNN	94.29 %

C. Discussion

The results of this experiment demonstrate that SNN outperforms other advanced deep learning models. This achievement is primarily attributed to the SMGCNN's integration of the strengths of both CNN and SNN, effectively combining spatial and temporal sequence features. It is worth noting that in low signal-to-noise ratio environments, SMGCNN exhibits significant robustness, highlighting its superior performance in handling noisy and interfered data. These findings explicitly indicate that merging different types of network layers, such as CNN and SNN, is a key strategy for enhancing model performance under complex conditions, especially when dealing with data noise and uncertainty. The experimental results also reveal the contributions and limitations of CNN and SNN layers when applied separately to model performance. The CNN layer plays an essential role in capturing spatial features, while the SNN layer significantly enhances the model's resistance to noise. This further underscores the importance of combining different types of network layers when designing deep learning models, especially in noisy environments.

V. CONCLUSION

Traditional deep learning methods often have difficulty accurately capturing the overall temporal dynamics and contextual information in time series data, especially when there is noise interference as encountered in rolling bearings. This study proposes the SMGCNN method to address this issue. This method combines the advantages of SNN in processing time series data and the multiscale feature extraction ability of MGC. Experimental results show that SMGCNN performs excellently in terms of accuracy. Especially in low signal-to-noise ratio environments, it shows higher accuracy and robustness compared to other recent deep learning models. In addition, the analysis of the confusion matrix further confirms the excellent performance of SMGCNN in diagnosing different categories. Future work can explore more efficient network structures or optimization strategies to reduce the computational burden of the model. Moreover, future research on SMGCNN can involve testing and optimizing the model in a wider range of datasets and real-world application scenarios.

ACKNOWLEDGMENT

This work was funded by the National Natural Science Foundation of China (Grant Nos. 62163023, 62063020), the Youth Science and Technology Fund of Gansu Province (22JR5RA808), Long yuan Young Innovation Talent Team Program in Gansu Province (310100296012).

REFERENCES

[1] Y. Sun and W. Wang, “Role of image feature enhancement in intelligent fault diagnosis for mechanical equipment: A review,” Eng. Fail. Anal., vol. 156, p. 107815, 2024.

[2] D. V Patel, V. M. Bhojawala, and K. M. Patel, “Size estimation of spall-like fault on rolling element bearing using an integrated approach incorporating total variation regularization and singular spectrum analysis,” Meas. Sci. Technol., vol. 34, no. 3, p. 35601, 2022.

[3] Liu, Bin, et al. "ISEANet: an interpretable subdomain enhanced adaptive network for unsupervised cross-domain fault diagnosis of rolling bearing." Advanced Engineering Informatics 62 (2024): 102610.

[4] T. Han, L. Ding, D. Qi, C. Li, Z. Fu, and W. Chen, “Compound faults diagnosis method for wind turbine mainshaft bearing with Teager and second-order stochastic resonance,” Measurement, vol. 202, p. 111931, 2022.

[5] Zhang, Y, Zhao, X, Peng, Z, et al. WD-KANTF: An interpretable intelligent fault diagnosis framework for rotating machinery under noise environments and small sample conditions. Advanced Engineering Informatics, 66, 103452.

[6] C. Lu, Z. Wang, and B. Zhou, “Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification,” Adv. Eng. Informatics, vol. 32, pp. 139–151, 2017.

[7] D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane, and M. Modarres, “Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings,” Shock Vib., vol. 2017, no. 1, p. 5067651, 2017.

[8] B. Cui, Y. Weng, and N. Zhang, “A feature extraction and machine learning framework for bearing fault diagnosis,” Renew. Energy, vol. 191, pp. 987–997, 2022.

[9] G. Jiang, H. He, J. Yan, and P. Xie, “Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox,” IEEE Trans. Ind. Electron., vol. 66, no. 4, pp. 3196–3207, 2018.

[10] J. Wang, T. Li, C. Sun, R. Yan, and X. Chen, “Improved spiking neural network for intershaft bearing fault diagnosis,” J. Manuf. Syst., vol. 65, pp. 208–219, 2022.

[11] Z. Yang, J. Cen, X. Liu, J. Xiong, and H. Chen, “Research on bearing fault diagnosis method based on transformer neural network,” Meas. Sci. Technol., vol. 33, no. 8, p. 85111, 2022.

[12] J. Xuan, Z. Wang, S. Li, A. Gao, C. Wang, and T. Shi, “Measuring compound defect of bearing by wavelet gradient integrated spiking neural network,” Measurement, vol. 223, p. 113796, 2023.

[13] L. Zuo, F. Xu, C. Zhang, T. Xiahou, and Y. Liu, “A multi-layer spiking neural network-based approach to bearing fault diagnosis,” Reliab. Eng. Syst. Saf., vol. 225, p. 108561, 2022.

[14] J. Wang, H. Peng, M. Tu, J. M. Pérez-Jiménez, and P. Shi, “A fault diagnosis method of power systems based on an improved adaptive fuzzy spiking neural P systems and PSO algorithms,” Chinese J. Electron., vol. 25, no. 2, pp. 320–327, 2016.

[15] Q. Yin, E. L. Johnson, and N. Ofen, “Neurophysiological mechanisms of cognition in the developing brain: Insights from intracranial EEG studies,” Dev. Cogn. Neurosci., vol. 64, p. 101312, 2023.

[16] K. L. dos Santos and M. P. dos Santos Silva, “Deep cross-training: An approach to improve deep neural network classification on mammographic images,” Expert Syst. Appl., vol. 238, p. 122142, 2024.

[17] L. Jing, M. Zhao, P. Li, and X. Xu, “A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox,” Measurement, vol. 111, pp. 1–10, 2017.

[18] M. Pfeiffer and T. Pfeil, “Deep learning with spiking neurons: Opportunities and challenges,” Front. Neurosci., vol. 12, p. 409662, 2018.

[19] S. B. Furber et al., “Overview of the SpiNNaker system architecture,” IEEE Trans. Comput., vol. 62, no. 12, pp. 2454–2467, 2012.

[20] C. Xiao, J. Chen, and L. Wang, “Optimal mapping of spiking neural network to neuromorphic hardware for edge-AI,” Sensors, vol. 22, no. 19, p. 7248, 2022.

[21] Liang H, Cao J, Zhao X. Multi-scale dynamic adaptive residual network for fault diagnosis[J]. Measurement, 2022, 188: 110397.

[22] Liang H, Cao J, Zhao X. Multibranch and multiscale dynamic convolutional network for small sample fault diagnosis of rotating machinery[J]. IEEE Sensors Journal, 2023, 23(8): 8973-8988.

[23] Abduelhadi A, Liang H, Cao J, et al. HF-MSCN: a high frequency-multiscale cascade network for bearing fault diagnosis[J]. Measurement Science and Technology, 2024, 35(11): 116120.