International Journal of
Automotive Manufacturing and Materials（IJAMM）

Scilight

Article

Rolling Bearing Fault Diagnosis Considering Long-Term Dependence and Time-Frequency Feature Fusion

Zhifang Chen [1,*] and Penglin He [2]

[1] Zhejiang Haikang Science&Technology Co., Ltd., Hangzhou 310059, China
[2] Shenzhen Smart Car Technology Co., Ltd., Shenzhen 518100, China
* Correspondence: 63511678@qq.com
Received: 4 November 2024; Revised: 9 April 2025; Accepted: 13 May 2025; Published: 29 July 2025

Abstract: Most of the current research models on rolling bearing fault diagnosis can effectively improve the reliability of bearings, but there are still some shortcomings. To address the issues of limited long-term dependency and sparse feature representation in existing rolling bearing fault diagnosis models, this paper proposes a novel method that incorporates long-term dependency and time-frequency feature fusion. The proposed method is based on a cross-deployed basic block with identity mapping and down-sampling, enabling the capture of long-term dependencies in bearing vibration time series signals. Furthermore, the use of skip connections facilitates effective information flow, allowing the network to integrate local dependencies, long-term dependencies, and time-frequency features at multiple scales. Experimental results on the Case Western Reserve University bearing dataset, which includes ten fault types, demonstrate that the proposed model achieves a detection accuracy of 86.8%. This performance surpasses that of the One-Dimensional Convolutional Neural Network (CNN1D) by 9% and the Multi-Layer Perceptron (MLP) by 16.3%. Moreover, the accuracy improvement is attributed to the incorporation of long-term dependency and time-frequency feature fusion. This research offers valuable insights for the intelligent diagnosis and predictive maintenance of bearings.

Keywords: mechanical engineering; rolling bearing fault diagnosis; time-frequency feature fusion; rolling bearing vibration signals; long-term dependency

1. Introduction

Rolling Bearings are deployed in diverse mechanical systems, ranging from traditional radial flux motors to hybrid motor types, wind turbines, pumps, and other industrial machines. Typically, in motors, rolling bearings play a critical role in supporting the rotor and maintaining the rotor-stator gap [1,2]. A precise rotor-stator air gap is essential for a machine's torque output, and overall efficiency [3,4]. Consequently, the maintenance of bearings significantly impacts the long-term stable operation of machines [5,6]. By implementing real-time monitoring of bearing conditions and providing early fault warnings, targeted maintenance can be conducted to extend the service life of both the bearings and the entire machine [7–9]. This approach helps reduce operating costs and optimize preventive maintenance strategies. Furthermore, it offers technical support for intelligent diagnosis and predictive maintenance of mechanical systems.

Among these, rolling bearings are the most widely used, due to their simple structure and low cost [10–14]. Currently, fault diagnosis for rolling bearings primarily relies on Data-driven Intelligent Fault Diagnosis (DIFD) methods [15]. These methods typically use accelerometer sensors to collect bearing vibration signals, which are then fed into specifically designed feature extractors for time-domain, frequency-domain, or composite feature extraction. Common techniques include Fourier transform [16], Winger-Ville distribution [17], empirical mode decomposition [18], and wavelet transform theory [19]. Once the features are extracted, they are embedded into classifiers for feature classification and fault identification. Common classifiers [20,21] include Logistic Regression

(LR) [22,23] and Support Vector Machines (SVMs) [24,25]. This technical solution constitutes the traditional fault diagnosis system, capable of completing simple fault diagnosis tasks in fixed application scenarios. However, the feature extractors in this approach often require customization by experts for different fault types, making it challenging to generalize to multiple unfamiliar fault types. Consequently, traditional fault diagnosis solutions have limited application scenarios and depend heavily on specialized manual feature customization. To ensure detection accuracy, additional complex data preprocessing or post-processing may also be necessary.

To address the challenges associated with traditional fault diagnosis technology, many scholars have explored to integration of Deep Learning (DL) into the fault diagnosis domain. This approach analyzes signal features in an end-to-end manner to obtain high-level semantic representations. For instance, Janssens et al. [26] used Convolutional Neural Networks (CNNs) to achieve four-class fault recognition in rotating machinery, preprocessing vibration data with Discrete Fourier Transform (DFT) before classification. The accuracy of the final model significantly outperforms the classical feature-engineering-based approach, which uses manually engineered features and a random forest classifier. Abdeljaber et al. [27] introduced a one-dimensional Convolutional Neural Network (CNN1D) model based on normalized vibration signals, directly performing damage detection and structural localization using raw vibration signals. Moreover, Lee et al. [28] applied CNNs for fault classification and diagnosis tasks in semiconductor manufacturing processes, utilizing sensor variables with time axis information, and convolution kernels sliding exclusively along the time axis. Furthermore, Guo et al. [29] proposed a hierarchical learning rate adaptive CNN structure for bearing fault diagnosis and damage degree investigation. The proposed model is superior to other models in terms of training speed and accuracy.

By analyzing existing fault diagnosis models based on DL (refer to Section 1), the causes of their limited fault diagnosis accuracy and generalization ability can be analyzed, including:

(1) Lack of long-term dependency: Existing fault diagnosis models are not able to capture extensive temporal dependencies due to their limited receptive fields.
(2) Poor representation ability: Detection models built by simple stacking ignore the synergy and fusion between temporal and frequency features, as well as local and long-term dependencies. This results in an overall impoverished feature space.

The model's field of view pertains to its ability to extract global semantic features, while its representation ability is crucial for the richness of the constructed feature space. To achieve high-precision and robust fault diagnosis, both aspects must be prioritized during model construction.

In summary, this paper addresses the limitations of existing fault diagnosis models, especially the lack of long-term dependencies and impoverished feature spaces. Therefore, a rolling bearing fault diagnosis method that incorporates long-term dependencies and time-frequency feature fusion is considered. This approach utilizes cross-deployed basic blocks with identity mapping and down-sampling to capture long-term dependencies in bearing vibration time series signals. Additionally, skip connections facilitate effective information flow, enabling the fusion of local dependencies, long-term dependencies, and temporal and frequency features at multiple scales. The accuracy of this method was verified by comparing it with CNN1D and the Multi-Layer Perceptron Network (MLPNet) on the Case Western Reserve University bearing dataset.

2. Analysis of Existing Representative Rolling Bearing Fault Diagnosis Models

This section will analyze existing representative bearing fault diagnosis models to uncover the causes of their limited fault diagnosis accuracy and generalization ability.

2.1. Multilayer Perceptron Network

MLPNet [30] is a feedforward neural network model built on fully connected layers and is widely used for fault diagnosis tasks using bearing vibration data. Its architecture, shown in Figure 1, consists of three components: the input layer, the hidden layer, and the output layer.

Through embedding, the dataset's features are captured by the hidden layer, followed by feature integration and prediction in the output layer. Specifically, let matrix $X \in \mathbb{R}^{n \times d}$ represent a mini-batch of n samples, where each has d input features. For a single-hidden-layer MLP with h hidden units, $H \in \mathbb{R}^{n \times h}$ denotes the hidden layer output. Due to MLPNet's fully-connected topology, its output is calculated as follows:

$$\begin{cases} \mathrm{H} = \sigma\left(\mathrm{X}\cdot\mathrm{W}^{[1]} + \mathrm{b}^{[1]}\right) \\ \mathrm{O} = \mathrm{H}\cdot\mathrm{W}^{[2]} + \mathrm{b}^{[2]} \end{cases} \tag{1}$$

where $\mathrm{W}^{[1]} \in \mathbb{R}^{d\times h}$ and $\mathrm{b}^{[1]} \in \mathbb{R}^{1\times h}$, represent the hidden layer weights and biases, whereas $\mathrm{W}^{[2]} \in \mathbb{R}^{h\times q}$ and $\mathrm{b}^{[2]} \in \mathbb{R}^{1\times q}$, denote the output layer weights and biases. Moreover, the non-linear activation function $\sigma(\cdot)$ gives the MLPNet the ability to fit nonlinear features, resulting in the final network output $\mathrm{O} \in \mathbb{R}^{n\times q}$. Therefore, the MLPNet is characterized by a simple structure and easy deployment; added to that, it can process high-dimensional input data without complex feature engineering.

However, MLPNet, as a feedforward neural network based on fully connected layers, has notable limitations in handling noise within the input data. Its architecture, characterized by a large number of parameters, increases the risk of overfitting, particularly when the available bearing vibration data is limited. This susceptibility to overfitting not only affects the model's ability to generalize but also results in poor performance when confronted with noise and outliers.

Moreover, MLPNet treats input data as independent feature vectors, making it challenging to effectively model temporal dependencies in input series data, such as bearing vibration signals. Consequently, this lack of consideration for the sequential nature of data restricts its performance in fault diagnosis tasks.

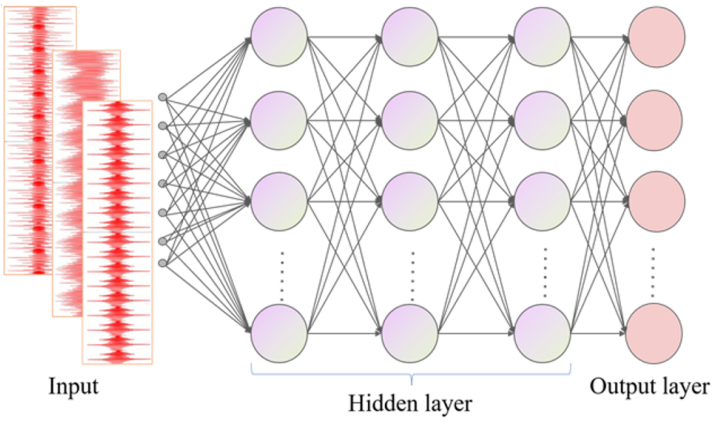


Figure 1. MLPNet model structure.

2.2. One-Dimensional Convolutional Neural Network Model

CNN1D [31] is a special CNN mainly used to process 1D sequence data involving time series and audio signals. As shown in Figure 2, the CNN1D architecture captures local dependencies in temporal signals, such as bearing vibration signals, through repeated stacking of convolutional and pooling layers. This design allows CNN1D to excel in applications like time series classification and anomaly detection.

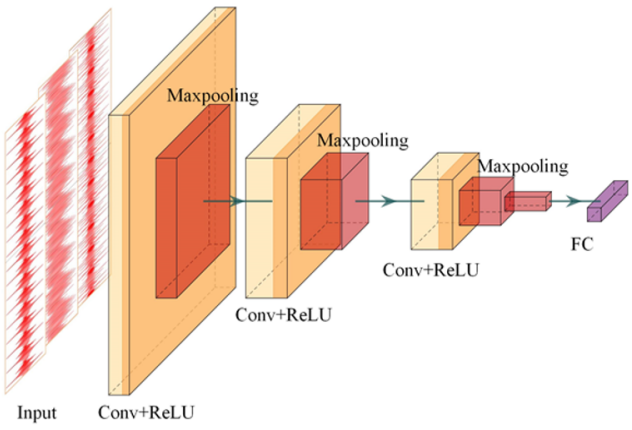


Figure 2. CNN1D model structure.

Unlike MLPNet, CCN1D can handle input sequences of variable length without necessitating fixed-length segmentation or padding. The network employs a sliding window mechanism, which enables local connections between inter-layer features using shared parameters. It results in high parameter efficiency and excellent fitting ability in few-shot learning.

After the repeated stacking of convolutional and pooling layers, along with the application of nonlinear activation functions, the model generates the bearing state feature vector *V*. Feeding this feature vector into a fully connected layer results in the bearing state label expression $Q = \left[q^1, q^2, q^3, ..., q^j, ..., q^n \right]$, as proposed in Equation (2):

$$Q = V \cdot W + b \tag{2}$$

where *W* and *b* represent the weight matrix and bias vector of the fully connected layer, respectively. Then, it is necessary to get the one-hot encoding label $P = \left[p^1, p^2, p^3, ..., p^j, ..., p^n \right]$ of the bearing state to generate the final fault type and train the overall network, as expressed in Equation (3):

$$\begin{aligned} p^j &= \frac{exp\left(q^j\right)}{\sum_{i=1}^{n} exp\left(q^i\right)} \\ j &\in (1, 2, ..., n), \sum_{i=1}^{n} p^j = 1 \end{aligned} \tag{3}$$

The final fault type *y* is determined using Equation (4) as follows:

$$y = argmax(P) \tag{4}$$

During network training, the cross-entropy loss function is often applied to describe the difference between predicted and true labels, as expressed below:

$$L = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{h} \hat{p}_{i,j} \cdot log\left(p_{i,j}\right) \tag{5}$$

where *N* represents the number of samples in the current batch, $p_{i,j}$ denotes the probability that the *i*-th sample is predicted to be the *j*-th type of fault, and $\hat{p}_{i,j}$ implies the corresponding true value.

The CNN1D model, while effective at capturing low overall dimension and limited receptive field, struggles with local dependencies in bearing vibration signals. Although its ability to model time dependencies surpasses MLPNet, CNN1D tends to lose long-distance temporal dependency information when considering long sequence data. This is crucial for fault diagnosis tasks based on bearing vibration signals. Although increasing the model's depth can enhance its receptive field and allow for better capture of frequency domain features, it often results in the degradation of time domain features in the final representation. Moreover, the increased number of parameters results in risks such as gradient explosion or vanishing during the training process, making it challenging to optimize the overall model parameters end-to-end. Therefore, the way to balance time-frequency information and long-distance temporal dependencies constitutes a trade-off that should be considered when creating efficient fault diagnosis models.

3. Fault Diagnosis Method for Rolling Bearings Considering Long-Term Dependencies and Time-Frequency Feature Fusion

This section will introduce a novel rolling bearing fault diagnosis method that emphasizes long-term dependencies and time-frequency feature fusion.

3.1. Modeling Motivation

Analysis of representative bearing fault diagnosis models, such MLPNet and CCN1D, reveals that existing DL-based fault diagnosis still has significant room for improvement in terms of detection accuracy and generalization. The reasons are listed as follows:

(1) *Lack of long-term dependency:* Existing fault diagnosis models are not able to capture extensive temporal dependencies due to their limited receptive fields. In the context of our work, "long-term dependency" refers to the model's ability to capture and represent temporal relationships and dependencies that span extended

periods within the bearing vibration signals. These long-term dependencies are critical for accurately diagnosing bearing faults, as bearing conditions evolve gradually over time, with subtle changes that manifest over longer time intervals.

(2) *Poor representation ability:* Detection models built by simple stacking ignore the synergy and fusion among temporal and frequency features, as well as local and long-term dependencies. This results in an overall impoverished feature space.

The model's field of view pertains to its capability to extract global semantic features, while its representation ability determines the richness of the constructed feature space. To ensure high precision and robustness in fault diagnosis, it is essential to emphasize both aspects during model development. To tackle the identified challenges, this work proposes a fault diagnosis method for rolling bearings that integrates long-term dependencies with time-frequency feature fusion.

3.2. Overall Model Structure

The proposed fault diagnosis method for rolling bearings, which incorporates long-term dependencies and time-frequency feature fusion, is inspired by the Residual Network (ResNet) [32]. Originally designed for image classification, ResNet employs two-dimensional (2D) convolution and pooling operations to effectively capture spatial features. Given the 1D nature of bearing vibration data, this paper constructs a fault diagnosis model using 1D convolution and pooling modules as fundamental components. The overall structure, shown in Figure 3, features a combination of basic blocks with identity mapping and those with down-sampling to enhance the network's expressive capability.

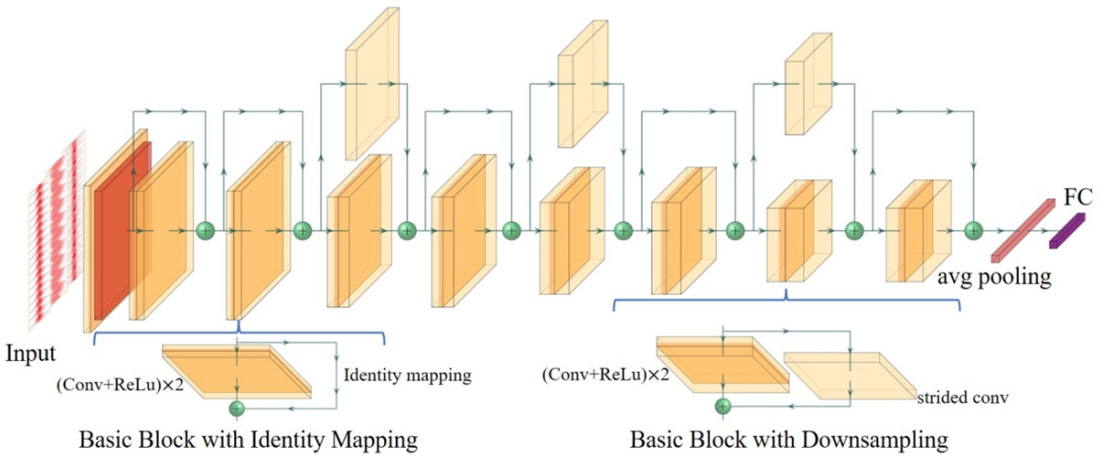


Figure 3. Rolling bearing fault diagnosis model structure considering long-term dependence and time-frequency feature fusion: ResNet1D.

3.3. Basic Block with Identity Mapping

The structure of the basic block with identity mapping is depicted in Figure 4. This block employs convolutional layers with a stride of 1 and zero padding (padding = 1), ensuring that the dimensions of the input feature map are preserved. This design maintains the integrity of the information flow while also increasing the network's receptive field. In the identity mapping branch, skipping connections directly transmit shallow features to deeper layers, effectively mitigating the network degradation issues that deep networks often encounter in deep architectures. During Back-Propagation (BP), these skip connections facilitate smoother gradient flow from deep to shallow layers, enhancing training convergence and improving overall model performance.

Crucially, the shallow features predominantly represent time-domain characteristics of the bearing vibration signals, while deeper features, which benefit from larger receptive fields and enhanced nonlinear representation capabilities, capture more complex frequency-domain and time-frequency information. The incorporation of skip

connections enables multi-scale time-frequency feature fusion, enriching the network's feature space and addressing the challenges faced by CNN1D networks in balancing time-frequency information.

By focusing on learning residual mapping rather than the entire mapping function, the model reduces the risk of overfitting, which is particularly vital in scenarios where bearing vibration data is limited.

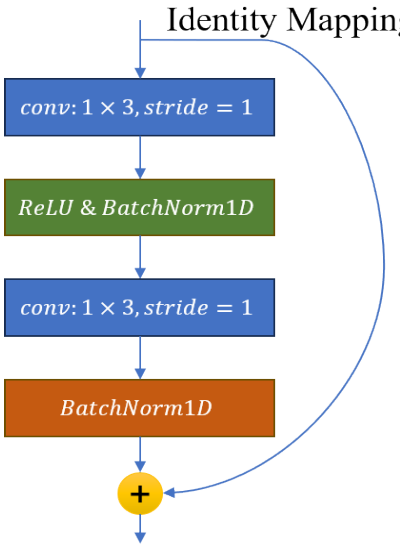


Figure 4. Basic block structure with identity mapping.

3.4. Basic Block with Down-Sampling

The structure of the basic block with down-sampling is illustrated in Figure 5. This block incudes both a convolutional branch and a down-sampling branch, utilizing stride convolution (Stride = 2) to achieve down-sampling by a factor of 2. This operation reduces the size of feature maps while preserving essential semantic information, which reducing computational cost and memory usage—an advantage for deploying the model in resource-constrained conditions. Similar to the identity mapping block, the down-sampling module also employs skip connections to directly transmit shallow features to deeper layers. This mechanism helps avoiding the network degradation issues that can arise in deeper layers. During BP, the presence of skip connections allows gradients to flow more smoothly from the deeper layers back to the shallow layers, facilitating improved convergence during training.

A key benefit of this module is that it performs down-sampling across two parallel branches, which significantly increases the model's receptive field compared to the direct identity mapping approach. When these down-sampling blocks are arranged repeatedly, as shown in Figure 3, they enable the network to model long-term dependencies in bearing time series signals effectively. The integration of skip connections further allows for the fusion of long-term and local dependencies, addressing a limitation commonly found in classic MLPNet networks.

As shown in Figure 3, the alternate deployment of basic blocks with identity mapping and those with down-sampling reflects a strategic trade-off between model performance and deployment complexity. In the identity mapping blocks, the dimensions of the input and output feature maps remain unchanged, ensuring that the integrity of the information flow is maintained. This design is crucial for achieving high-precision detection performance. However, if these blocks are excessively stacked, they can lead to a sharp increase in computational cost at each layer, complicating model training and deployment.

Conversely, the down-sampling blocks enhance the model's receptive field by reducing the dimensions of the feature maps. This capability is beneficial for capturing long-term dependencies in bearing time series signals and contributes to reducing the overall computational cost, which aids in efficient model deployment and training. Nonetheless, excessive down-sampling will severely compromise the integrity of information flow, making it challenging to fuse time and frequency information in the deeper layers of the model.

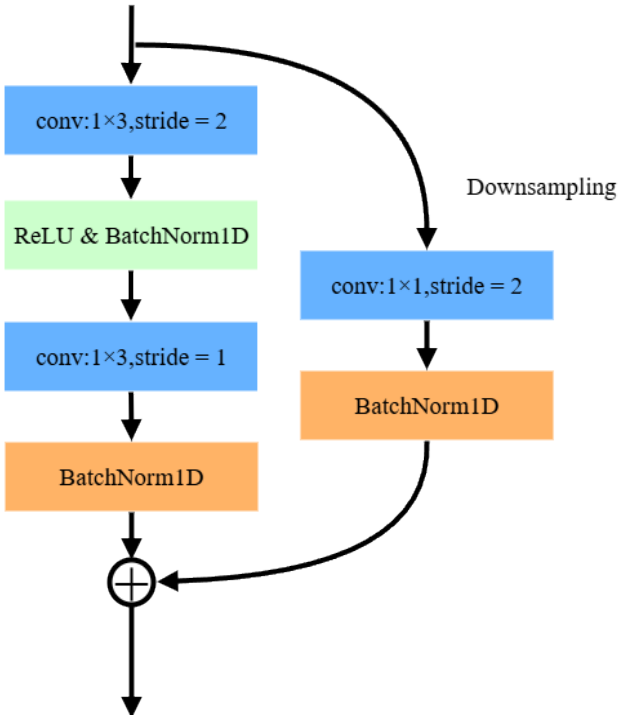


Figure 5. Basic block structure with down-sampling.

4. Experiments and Results Analysis

4.1. Experimental Setup

This section presents comparative experiments among the proposed bearing fault diagnosis model considering long-term dependencies and time-frequency feature fusion, MLPNet, and CNN1D. The experiments are implemented using Pytorch 2.0.1 (Meta, Menlo Park, CA, United States), CUDA 11.7 (NVIDIA, Santa Clara, CA, United States), and Python 3.9 (Python Software Foundation, Wilmington, DE, United States).

For the training process, the Adam optimizer is employed with a batch size of 32 and a dropout rate of 0.5. the training is conducted over 100 epochs. A cosine annealing learning rate adjustment strategy is applied during training to optimize the learning process, as described in Equation (6):

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min})\left(1 + cos\left(\frac{T_{cur}}{T_{max}}\pi\right)\right) \quad (4)$$

where $\eta_{\min}$ is the minimum learning rate, $\eta_{\max}$ represents the maximum learning rate (initial learning rate), and T_{cur} denotes the current epoch.

To verify the effectiveness of the proposed rolling bearing fault diagnosis model that incorporates long-term dependencies and time-frequency feature fusion, this section utilizes bearing data from the Case Western Reserve University (CWRU) dataset [33]. The experimental setup is illustrated in Figure 6.

In this study, the bearing fault conditions were created using electrical discharge machining technology, focusing on single-point faults. The faults were categorized by three fault diameters: 0.007 inches, 0.014 inches, and 0.021 inches. These faults were positioned at three locations: the ball, inner ring, and outer ring of the bearing, as shown in Figure 7.

The experimental setup included four different load conditions for bearing operation: 0 hp, 1 hp, 2 hp, and 3 hp, which corresponded to motor speeds of 1797 RPM, 1772 RPM, 1750 RPM, and 1730 RPM, respectively. To monitor the bearing conditions, acceleration sensors were mounted on the motor housing using magnetic bases. Vibration signals were recorded using a 16-channel DAT recorder (TASCAM, Montebello, CA, United States) at sampling frequencies of 12 kHz and 48 kHz.

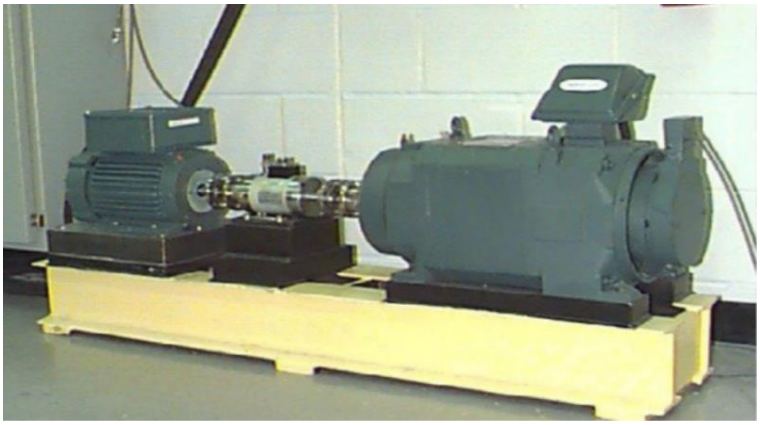


Figure 6. CWRU rolling bearing test bench.

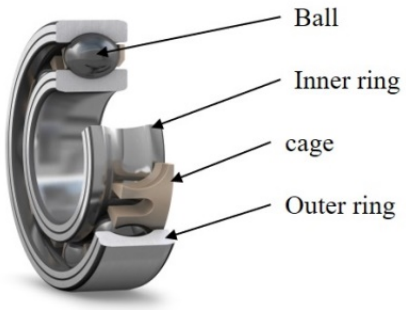


Figure 7. Structure of rolling bearing.

4.2. CWRU Data Description

This section utilizes bearing vibration data monitored at the motor drive end under different load conditions, sampled at a frequency of 48 kHz. The rolling bearing fault dataset is summarized in Table 1, including normal operating conditions as well as three types of faults: inner ring fault, outer ring fault, and ball fault. Each fault type encompasses three damage levels: 0.007 inches, 0.014 inches, and 0.021 inches, resulting in a total of 10 distinct conditions. Partial time-domain waveforms of the vibration signals corresponding to these conditions are depicted in Figure 8.

Table 1. Rolling bearing fault description and corresponding label.

Serial Number	Fault Location	Fault Size/Inch	Label
1	Normal	0	Normal
2	Ball	0.007	Ball 0.007
3		0.014	Ball 0.014
4		0.021	Ball 0.021
5	Inner ring	0.007	Inner 0.007
6		0.014	Inner 0.014
7		0.021	Inner 0.021
8	Outer ring	0.007	Outer 0.007

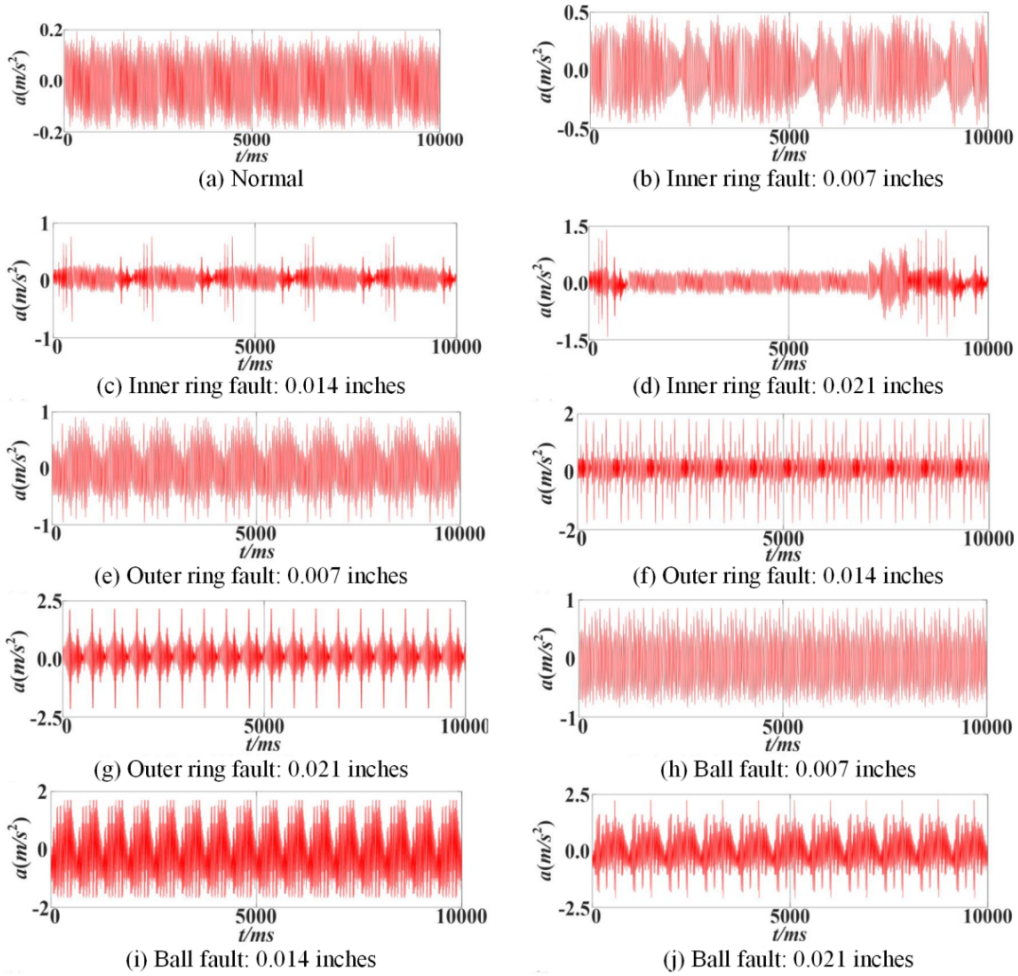
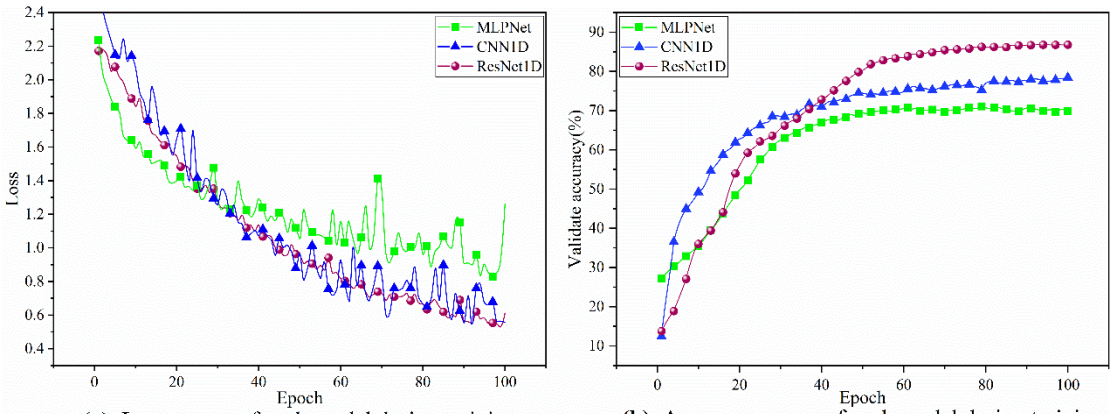


Figure 8. Partial vibration signal diagrams of 10 rolling bearing states.

4.3. Analysis of Experimental Results

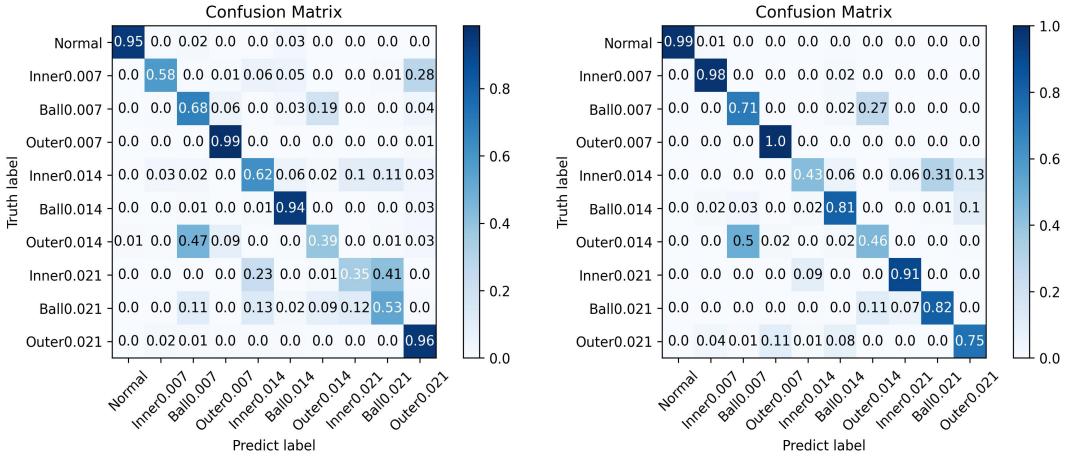
In the experiment, we consistently employed a sampling frequency of 48 kHz, utilizing vibration signals recorded under 3 HP load as the training set for each model, while the test set comprised vibration signals collected under 0 HP load. These two load-speed conditions represent the maximum operational divergence available in the dataset. That is, the 3 hp and 0 hp working conditions are at opposite ends of the mechanical and dynamical spectrum, offering the greatest challenge for model generalization. The difference in mechanical load and rotational speed between them leads to substantial variations in both the spectral content and the transient dynamics of the vibration signal. By choosing this most extreme scenario, we aimed to rigorously evaluate the model's ability to generalize across domains—a core requirement for deployment in real-world industrial settings, where machinery often operates under dynamically changing loads and environmental conditions.

The loss and accuracy curves for each model during the training process are presented in Figure 9, illustrating the models' convergence behavior and performance over training epochs. These curves clearly demonstrate that as training progresses, the accuracy steadily increases and eventually stabilizes, while the loss decreases and converges. Such behavior is a strong indication that the model is effectively learning and generalizing well to the data, without showing signs of overfitting. The confusion matrix depicting the model inference results on the test set is presented in Figure 10.

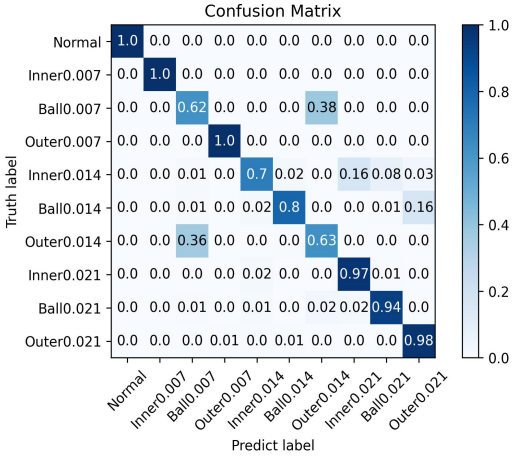


(a) Loss curves of each model during training **(b)** Accuracy curves of each model during training

Figure 9. Accuracy curve and loss curve of each model in training process.



(a) MLPNet confusion matrix. **(b)** CNN1D confusion matrix.



(c) ResNet1D confusion matrix.

Figure 10. Confusion matrix of inference results of each model on the test set.

Due to the strong temporal sequence characteristics of bearing vibration data, the CNN1D model excels in capturing the temporal correlations inherent in these signals, allowing for more effective extraction of fault-related features. In contrast, MLP is limited in its ability to model time series correlations. Additionally, CNN1D benefits from local connectivity and a weight sharing mechanism, which allow it to efficiently

extract local dependency information from the vibration signals. In comparison, MLPNet requires larger architectures and more parameters to model this information. In addition, CNN1D's parameter sharing mechanism offers higher parameter efficiency, enabling efficient training on relatively small datasets; this is particularly important for small sample data such as vibration signals. Ultimately, CNN1D achieves a detection accuracy of 77.8% on the test set, while MLPNet's accuracy is limited to 70.5%.

The rolling bearing fault diagnosis model, referred as ResNet1D, demonstrates a significantly larger network receptive field compared to both CNN1D and MLPNet, owing to its deeper network structure and the integration of basic blocks with down-sampling. This structural advantage facilitates the model's ability to capture long-term dependencies in bearing time series signals while effectively extracting frequency-domain and time-frequency features. Additionally, the implementation of skip connections within ResNet1D facilitates effective information flow, enabling the fusion of local dependencies, long-term dependencies, and multi-scale time-frequency features. These measures significantly improve the model's generalization ability, allowing it to better mitigate overfitting particularly in complex vibration signal datasets.

As a result, ResNet1D achieved the highest detection accuracy of 86.8% on the test set, surpassing CNN1D network by 9% and MLPNet by 16.3%. This notable accuracy improvement can be attributed to the successful incorporation of long-term dependencies and time-frequency feature fusion into the model.

Additionally, the computational complexity and inference time of the model are evaluated on an NVIDIA GeForce RTX 4090 GPU (NVIDIA, Santa Clara, CA, United States). The input data consists of 1D vibration signals with a length of 1024, as specified in the experimental setup. The experimental results demonstrate that the proposed method contains only 3.84 million parameters and exhibits a low computational complexity of 174.29 million FLOPs, with an inference time of 4.65 ms. Therefore, the model is suitable for deployment on edge devices or embedded systems, as it does not require substantial computational resources.

5. Conclusions

To sum up, this paper establishes a rolling bearing fault diagnosis method considering long-term dependencies and time-frequency feature fusion. The main research conclusions are summarized as follows:

(1) This paper thoroughly analyzed existing representative rolling bearing fault diagnosis models, MLPNet and CNN1D, and determines the reasons for their limited fault diagnosis accuracy and generalization. These reasons include (1) a lack of long-term dependencies where the overall model dimension is relatively low, and its restricted receptive field struggles to capture extensive temporal dependencies, and (2) the impoverished feature space where models built by layer-by-layer stacking ignore the synergy and fusion between time-frequency features and local and long-term dependencies;

(2) To address the challenge linked to the lack of long-term dependencies and impoverished feature space in current models, this paper establishes a rolling bearing fault diagnosis model that considers both long-term dependencies and time-frequency feature fusion. The proposed model captures long-term dependencies in bearing vibration time series signals while alternately deployed basic blocks with identity mapping and down-sampling. It also employs skip connections to offer effective information flow pathways for the network to fuse local dependencies, long-term dependencies, and multi-scale time-frequency features;

(3) Validation experiments were conducted using bearing data obtained from Case Western Reserve University (CWRU) dataset. The experimental results highlight that ResNet1D achieved the highest detection accuracy of 86.8% on the test set, outperforming the CNN1D network by 9% and the MLPNet network by 16.3%. Therefore, the accuracy enhancement is obtained by introducing long-term dependencies and time-frequency feature fusion.

Author Contributions: Z. C.: conceptualization, methodology, software, visualization, writing—original draft preparation; P.H.: data curation, supervision, validation, writing—reviewing and editing. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability: Statement: The data are provided as requested.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chong, Y.C.; Staton, D.A.; Mueller, M.A.; Chick, J. An experimental study of rotational pressure loss in rotor-stator gap. *Propuls. Power Res.* **2017**, *6*, 147–156.
2. Hu, B.; Yao, Y.; Wang, C.; Chen, X. The effect of rotor roughness on flow and heat transfer in rotor–stator cavities with different axial gap. *Appl. Therm. Eng.* **2024**, *251*, 123535.
3. Heins, G.; Thiele, M.; Patterson, D.; Lambert, N. Increase in operating range and efficiency for variable gap axial flux motors. In Proceedings of the 2014 IEEE Energy Conversion Congress and Exposition (ECCE), Pittsburgh, PA, USA, 14–18 September 2014; pp. 5870–5876.
4. Manne, V.H.B.; Vacca, A.; Merrill, K. A numerical method for evaluating the torque efficiency of hydraulic orbit motors considering deformation effects and frictional losses. *Mech. Syst. Signal Process.* **2021**, *146*, 107051.
5. Yang, W.; Court, R. Experimental study on the optimum time for conducting bearing maintenance. *Measurement* **2013**, *46*, 2781–2791.
6. Zimroz, R.; Bartelmus, W.; Barszcz, T.; Urbanek, J. Diagnostics of bearings in presence of strong operating conditions non-stationarity—A procedure of load-dependent features processing with application to wind turbine bearings. *Mech. Syst. Signal Process.* **2014**, *46*, 16–27.
7. You, K.; Qiu, G.; Gu, Y. Optimizing prior distribution parameters for probabilistic prediction of remaining useful life using deep learning. *Reliab. Eng. Syst. Saf.* **2024**, *242*, 109793.
8. You, K.; Qiu, G.; Gu, Y. Remaining useful life prediction of lithium-ion batteries using EM-PF-SSA-SVR with gamma stochastic process. *Meas. Sci. Technol.* **2023**, *35*, 015015.
9. You, K.; Qiu, G.; Gu, Y. A 3-D attention-enhanced hybrid neural network for turbofan engine remaining life prediction using CNN and BiLSTM models. *IEEE Sens. J.* **2023**, *24*, 21893–21905.
10. You, K.; Wang, P.; Gu, Y. Towards efficient and interpretative rolling bearing fault diagnosis via quadratic neural network With Bi-LSTM. *IEEE Internet Things J.* **2024**, *11*, 23002–23019.
11. You, K.; Qiu, G.; Gu, Y. Rolling bearing fault diagnosis using hybrid neural network with principal component analysis. *Sensors* **2022**, *22*, 8906.
12. You, K.; Qiu, G.; Gu, Y. An efficient lightweight neural network using BiLSTM-SCN-CBAM with PCA-ICEEMDAN for diagnosing rolling bearing faults. *Meas. Sci. Technol.* **2023**, *34*, 094001.
13. You, K.; Wang, P.; Huang, P.; Gu, Y. A sound-vibration physical-information fusion constraint-guided deep learning method for rolling bearing fault diagnosis. *Reliab. Eng. Syst. Saf.* **2025**, 253, 110556.
14. You, K.; Lian, Z.; Chen, R.; Gu, Y. A novel rolling bearing fault diagnosis method based on time-series fusion transformer with interpretability analysis. *Nondestruct. Test. Eval.* **2024**, 1–27.
15. Zhou, Q.; Li, J.; Xu, H. Artificial Intelligence and Its Roles in the R&D of Vehicle Powertrain Products. *Int. J. Automot. Manuf. Mater.* **2022**, *1*, 6.
16. Neşe, S.V.; Kılıç, O.; Akıncı, T.Ç. Analysis of wind turbine blade deformation with STFT method. *Energy Educ. Sci. Technol. Part A-Energy Sci. Res.* **2012**, *29*, 679–686.
17. Zhou, Y.; Chen, J.; Dong, G.M.; Xiao, W.B.; Wang, Z.Y. Wigner–Ville distribution based on cyclic spectral density and the application in rolling element bearings diagnosis. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2011**, *225*, 2831–2847.
18. Lei, Y.; Lin, J.; He, Z.; Zuo, M.J. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.* **2013**, *35*, 108–126.
19. Yan, R.; Gao, R.X.; Chen, X. Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Process.* **2014**, *96*, 1–15.
20. Kudo, M.; Sklansky, J. Comparison of algorithms that select features for pattern classifiers. *Pattern Recognit.* **2000**, *33*, 25–41.
21. Murphy, K.P. Naive bayes classifiers. *Univ. Br. Columbia* **2006**, *18*, 1–8.
22. Yan, J.; Lee, J. Degradation assessment and fault modes classification using logistic regression. *J. Manuf. Sci. Eng.* **2005**, *127*, 912–914.
23. Yang, Y.; Gao, X.; You, J.; Zhang, D.; Zhang, Z.; Song, Y. A Control System Design for an Intelligent Unmanned Automotive. *Int. J. Automot. Manuf. Mater.* **2024**, *3*, 6.
24. Widodo, A.; Yang, B.S. Support vector machine in machine condition monitoring and fault diagnosis. *Mech. Syst. Signal Process.* **2007**, *21*, 2560–2574.
25. Xie, S.; Li, Z.; Arvin, F.; Ding, Z. A Review of Multi-vehicle Cooperative Control System in Intelligent Transportation. *Int. J. Automot. Manuf. Mater.* **2023**, *2*, 5.
26. Janssens, O.; Slavkovikj, V.; Vervisch, B.; Stockman, K.; Loccufier, M.; Verstockt, S. Convolutional neural network based fault detection for rotating machinery. *J. Sound Vib.* **2016**, *377*, 331–345.
27. Abdeljaber, O.; Avci, O.; Kiranyaz, S.; Inman, D.J. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *J. Sound Vib.* **2017**, *388*, 154–170.
28. Lee, K.B.; Cheon, S.; Kim, C.O. A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes. *IEEE Trans. Semicond. Manuf.* **2017**, *30*, 135–142.
29. Guo, X.; Chen, L.; Shen, C. Hierarchical adaptive deep convolution neural network and its application to bearing

fault diagnosis. *Meas* **2016**, *93*, 490–502.
30. Kim, T.; Adali, T. Fully complex multi-layer perceptron network for nonlinear signal processing. *J. VLSI Signal Process. Syst. Signal Image Video Technol.* **2002**, *32*, 29–43.
31. Malek, S.; Melgani, F.; Bazi, Y. One-dimensional convolutional neural networks for spectroscopic signal regression. *J. Chemom*. **2018**, *32*, e2977.
32. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
33. Case Western Reserve University Bearing Data Center. Available online: https://engineering.case.edu/bearingdatacenter (accessed on 22 October 2024).