

A Review on Rolling Bearing Fault Diagnosis Techniques based on Deep Learning

Tao Zhang, Yang Yu, Lianhai Wang, Shaoda Wang

College of Shipping, Shandong Jiaotong University, Weihai 264209, China

Abstract

Rolling bearings, as core transmission components in rotating machinery, directly determine the reliability of equipment operation and the safety of industrial production. They play an irreplaceable role in key sectors such as aerospace, rail transportation, petrochemical industries, and intelligent manufacturing. In complex working conditions, bearing vibration signals exhibit strong non-stationarity and significant nonlinearity. Traditional fault diagnosis methods face challenges such as a high dependency on manual feature extraction, poor robustness under complex conditions, and insufficient ability to recognize multiple faults simultaneously. Deep learning techniques, with their advantages of end-to-end automatic feature learning and complex pattern recognition, have become the core research direction in the field of rolling bearing fault diagnosis. This paper systematically reviews recent research on rolling bearing fault diagnosis based on Convolutional Neural Networks (CNN) and their fusion models. It focuses on analyzing the technical principles, model architecture design, and performance optimization mechanisms of typical methods such as COA-CNN, FFT-CNN-Transformer, MSCNN-GRU, and 2D-CNN-GRU. The paper also compares the feature enhancement effects of preprocessing techniques such as Discrete Continuous Wavelet Transform (DCWT), Fast Fourier Transform (FFT), and Variational Mode Decomposition (VMD). Additionally, it explores optimization strategies for diagnostic robustness under noise interference, varying operating conditions, and small sample scenarios. Engineering application cases, such as coal chemical circulation fans and CNC machine tool spindle boxes, are used to verify the practical effectiveness of these methods. Finally, the paper looks ahead to future developments in areas such as small sample learning, dynamic condition adaptation, model lightweighting, and multi-modal fusion, providing theoretical references and technical support for the development and industrial application of intelligent diagnostic systems for rolling bearings.

Keywords

Rolling bearings, fault diagnosis, deep learning, convolutional neural networks, feature extraction, complex working conditions, model fusion.

1. Introduction

As key supporting components of rotating machinery, rolling bearings are responsible for the positioning of rotating bodies and power transmission, and their health status directly affects the overall operational stability and service life of equipment. According to industrial fault statistics, over 40% of rotating machinery failures are caused by rolling bearing failures. Moreover, the characteristic signals in the early stages of failure are weak and can be easily obscured by high-speed operation, variable load impacts, and equipment background noise. If not diagnosed and warned in time, this may lead to equipment downtime, production interruptions, or even major safety accidents, causing huge economic losses. Traditional rolling bearing fault diagnosis methods are based on the core framework of "signal processing - feature extraction - pattern recognition," extracting fault features through time-domain analysis (mean,

root mean square, kurtosis, and other statistics), frequency-domain analysis (Fourier transform), or time-frequency analysis (wavelet transform, empirical mode decomposition (EMD))[1], and then using shallow machine learning models such as support vector machines (SVM) or BP neural networks for fault classification. However, these methods have significant limitations: first, the feature extraction process heavily relies on expert knowledge, has poor adaptability to non-stationary and nonlinear vibration signals, and struggles to capture transient fault features; second, in environments with strong noise interference or compound faults, manually designed features are easily contaminated by noise, leading to a significant decline in diagnostic accuracy; third, shallow models have limited capability to handle large-scale, multimodal industrial monitoring data, and their generalization performance may not meet the demands of complex operating conditions. With the rapid development of artificial intelligence, deep learning, thanks to its adaptive feature learning capability in deep network structures, breaks the stepwise limitations of traditional diagnostic methods and realizes a direct mapping from raw vibration signals to fault types, providing a new approach to diagnosing rolling bearing failures in complex operating conditions. Convolutional neural networks (CNNs), due to their strong local feature extraction ability and efficient parameter sharing mechanisms, have become the mainstream model for rolling bearing fault diagnosis, giving rise to optimized single-modal structures such as multi-scale CNNs and dilated convolution CNNs, as well as multi-modal fusion architectures like CNN-GRU and CNN-Transformer. At the same time, the introduction of intelligent optimization algorithms (such as the Coati Optimization Algorithm (COA) and Harris Hawk Optimization (HHO)) and advanced signal preprocessing techniques further enhances the model's adaptability to complex operating conditions. Based on six representative research papers, this article systematically reviews the research progress of deep learning in rolling bearing fault diagnosis from four dimensions: technical principles, model design, experimental validation, and engineering applications, analyzing the core advantages and applicable scenarios of various methods, and providing comprehensive references for technological innovation and engineering practice in this field.

2. Core Technologies for Rolling Bearing Fault Diagnosis

2.1. Fault Signal Characteristics and Diagnostic Challenges

Common types of rolling bearing faults include inner ring faults, outer ring faults, rolling element faults, and multi-component compound faults. The vibration signal characteristics are significantly related to the fault mechanisms: under normal operating conditions, bearing vibration intensity is low (typically <0.2 mm/s), with frequency components concentrated in the low-frequency range (0~1000 Hz); inner ring faults, caused by periodic impacts between rolling elements and the inner raceway, result in a significant increase in vibration intensity (can exceed 0.7 mm/s), with frequency distribution showing a dual-band concentration at 0~1000 Hz and 1500~2000 Hz; outer ring fault vibration signals are prominent in the low-frequency range and exhibit characteristic differences depending on the damage location (e.g., 3 o'clock, 6 o'clock, 12 o'clock positions); rolling element faults manifest as group peak characteristics around the 2000 Hz frequency range. However, in industrial scenarios, processing rolling bearing vibration signals faces three core challenges: non-stationary and nonlinear characteristics: during high-speed rotation, dynamic changes in bearing stiffness, load fluctuations, and component wear cause time-varying signal frequency components. Traditional Fourier transforms struggle to capture transient fault features, and the resolution of time-frequency analysis methods is also limited by signal complexity; strong noise interference: background noise from equipment operation, sensor measurement noise, and electromagnetic interference can easily mask weak early-stage fault features. Especially when

fault damage diameters are less than 0.178 mm, the signal-to-noise ratio can drop to -6 dB, further complicating feature extraction; variability in operating conditions: real-time changes in industrial equipment speed and load (0~3 HP) cause shifts in fault characteristic frequencies, resulting in insufficient generalization of diagnostic models trained under a single operating condition and significantly reduced cross-condition diagnostic accuracy.

2.2. Deep Learning Diagnostic Framework

The deep learning-based [2] rolling bearing fault diagnosis framework typically includes four core modules: data preprocessing, feature extraction, model training, and fault classification. Its technical advantages are mainly reflected in the following three aspects: Automatic feature learning: High-level abstract features are adaptively extracted from raw vibration signals or time-frequency images through network layers such as CNN and GRU, without the need for manually designed feature engineering, thereby avoiding subjective errors caused by dependence on expert experience; End-to-end diagnosis process: One-dimensional vibration signals or two-dimensional time-frequency images are directly input into the network, which outputs fault types after computation, simplifying the multi-step traditional diagnosis process and improving diagnostic efficiency; Robustness optimization mechanisms: Techniques such as data augmentation (overlapping sampling, random noise addition), regularization (Dropout, batch normalization), and attention mechanisms enhance the model's adaptability to noise interference and operating condition changes, improving diagnostic stability. In recent years, research on rolling bearing fault diagnosis has gradually shifted from the application of single deep learning models to an integrated architecture of 'signal preprocessing - deep learning model - intelligent optimization algorithm,' forming a multi-technology collaborative diagnostic system.

3. Progress in Research on Typical Deep Learning Diagnostic Methods

3.1. Single-Modal Diagnostic Methods based on Optimized CNN

3.1.1. COA-CNN: CNN Parameter Global Optimization Driven by Intelligent Optimization

The core parameters of traditional CNN models (such as convolution kernel size, quantity, learning rate, etc.) often rely on empirical selection, which can easily lead the model to fall into local optima and fail to fully exploit feature extraction capabilities. To address this issue, Bie Fengfeng [3] and others proposed a rolling bearing fault diagnosis method based on CNN optimized by the Coati Algorithm (COA-CNN). Its technical innovations are mainly reflected in the following three aspects: 1. Differential Continuous Wavelet Transform (DCWT) Signal Preprocessing: High-frequency noise is filtered through first-order differentiation, and then the one-dimensional vibration signal is converted into a two-dimensional time-frequency map using continuous wavelet transform, effectively preserving the complete time-frequency information of the signal and solving the problem of insufficient time-frequency resolution in traditional wavelet transforms. 2. COA Optimization for CNN Parameter Mechanism: By simulating the foraging behavior of the coati, CNN parameters are globally optimized in two stages. The first stage updates the population positions based on a prey expulsion strategy. The second stage dynamically adjusts the search space through predator avoidance behavior, ultimately achieving a globally optimal configuration for key parameters such as the number of convolution kernels and the learning rate. 3. Experimental Verification and Engineering Application: On the SQI-MFS mechanical fault comprehensive simulation test platform (bearing model MBER-12K, sampling frequency 12800Hz, motor speed 20Hz), COA-CNN achieved a classification accuracy of 93.8% for four typical states: normal, inner race fault, outer race fault, and rolling element fault, which represents an improvement of 6.0, 4.7, and 8.4 percentage points over traditional CNN (87.8%), SVM (89.1%), and BP neural networks (85.4%),

respectively. In engineering applications of circulating fans in coal chemical enterprises (fan speed 1481 r/min, sampling frequency 12800Hz), this method further improved the fault recognition accuracy of fans No. 5, No. 6, and No. 7 to 96.3%, while the computational time (18.2s) was significantly lower than that of traditional CNN (21.3s), demonstrating its effectiveness under complex industrial conditions. The core value of the COA-CNN method lies in solving the CNN parameter tuning problem through an intelligent optimization algorithm, significantly improving the model's adaptability to strong noise and non-stationary signals. It is particularly suitable for fault diagnosis in industrial equipment under multi-condition and high-interference scenarios. However, it should be noted that this method is relatively sensitive to the COA population size (set as 30 in the experiments) and the number of iterations (160 times), which need to be dynamically adjusted according to the data complexity to balance optimization efficiency and diagnostic performance.

3.1.2. Multi-scale CNN and Dilated Convolution: Enhanced Design for Feature Extraction

Traditional CNN models tend to overlook temporal dimension information when processing vibration signals, leading to the loss of sequential features and affecting fault diagnosis accuracy. To address this issue, Chen Chao et al. proposed a fault diagnosis method that integrates a Multi-Scale Convolutional Neural Network (MSCNN) with Gated Recurrent Units (GRU), referred to as MSCNN-GRU. The structural innovations primarily include three aspects: 1. Multi-scale feature extraction module design: Using large convolution kernels of sizes 64×1 , 32×1 , 16×1 , and 8×1 with long strides (e.g., 8×1 stride), the input length to the GRU layer is reduced while capturing local features at different scales. Following each convolutional path, a GRU layer is added to leverage GRU's update and reset gate mechanisms to strengthen sequential feature learning, avoiding the traditional CNN's defect of losing temporal correlations in the signal. 2. Dilated convolution for enhanced spatial feature extraction: Three sequential dilated convolution layers with dilation rates of 1, 2, and 7 expand the receptive field from the traditional 3×3 to 21×21 , effectively capturing large-range spatial features. At the same time, the heterogeneous dilation rates weaken grid effects to prevent loss of information continuity. 3. Noise resistance and generalization optimization: A Dropout layer (dropout rate 0.5) is introduced at the model input layer, forcing the network to train based on a few key features, enhancing noise robustness. Using a global average pooling (GAP) layer instead of traditional fully connected layers reduces the model parameters from 1.233 million to 72,000, significantly lowering the overfitting risk and improving generalization. Experimental results on the Case Western Reserve University (CWRU) rolling bearing dataset (bearing model SKF6205, fault diameters 0.178–0.534 mm, sampling frequency 12 kHz) show that the MSCNN-GRU model achieves 100% diagnosis accuracy without noise interference and still maintains 86.75% accuracy under -6dB strong noise, significantly outperforming comparison models such as WDCNN (37.43%), ResNet18 (24.93%), and CNN-LSTM (31.68%). Cross-dataset validation on the Jiangnan University rolling bearing dataset shows that the model's accuracy decreases the least as noise increases, further demonstrating its generalization capability. The MSCNN-GRU method provides an effective solution for rolling bearing fault diagnosis under strong noise, small sample sizes, and variable load conditions, although large convolution kernel designs increase computation per convolution, necessitating a trade-off between feature extraction precision and computational efficiency.

3.2. Multi-modal Diagnosis Method Based on CNN-Temporal Model Fusion

3.2.1. 2D-CNN-GRU: Spatial and Temporal Feature Co-learning Architecture

To address the low diagnostic accuracy of traditional CNN models under multiple coexisting faults (such as inner and outer ring compound faults) and variable operating conditions, Zhang Xiong [4] and others proposed a rolling bearing multi-fault diagnosis model (2D-CNN-GRU) that

integrates a two-dimensional convolutional neural network (2D-CNN) with a gated recurrent unit (GRU). Its technical features are mainly reflected in the following three aspects: Dual-branch feature extraction architecture: 2D-CNN is used as a spatial feature extractor. Through three convolutional layers (kernel sizes $20 \times 20 \times 32$, $9 \times 9 \times 64$, and $4 \times 4 \times 128$) and max-pooling layers, local fault spatial features are extracted from the two-dimensional time-frequency diagram; the GRU layer serves as a temporal feature extractor, utilizing the gating mechanism to model the time dependencies of vibration signals, compensating for the CNN model's insufficient capability to model long-sequence information. Cross-dataset performance validation: On the XJTU-SY rolling bearing dataset (bearing speed 2100 r/min, radial force 12 kN) and the QPZZ-II multi-condition dataset (including 12 states such as acceleration/deceleration, unbalanced loading, and compound faults), the diagnostic accuracy of the 2D-CNN-GRU model exceeds 95%, reaching up to 98%, significantly outperforming comparative models such as AlexNet (71.73%), MobileNet-V2 (51.92%), and LeNet-5 (66.15%). Model efficiency optimization: Through parameter pruning and structural simplification, the 2D-CNN-GRU model has only 609,000 parameters, a 73.9% reduction compared to AlexNet (2,332,000), and the training time is reduced by more than 40%, meeting the real-time diagnostic requirements of industrial scenarios. The core advantage of this model lies in the collaborative use of spatial and temporal features, effectively enhancing adaptability to compound faults and variable operating conditions, making it suitable for industrial scenarios with multiple fault types and complex working conditions. However, it should be noted that this method requires converting one-dimensional vibration signals into two-dimensional time-frequency diagrams, increasing the computational cost of preprocessing, which needs further optimization in scenarios with extremely high real-time requirements.

3.2.2. Comparative Analysis of MSCNN-GRU and 2D-CNN-GRU

Both MSCNN-GRU and 2D-CNN-GRU focus on mining temporal dimension information, achieving collaborative learning of spatial-temporal features through the integration of CNN and GRU. However, their technical approaches differ significantly: MSCNN-GRU uses a one-dimensional CNN to directly process raw vibration signals, avoiding information loss during the time-frequency transformation process and performing better under high-noise conditions (-6dB noise accuracy 86.75% vs 83.64%); 2D-CNN-GRU indirectly extracts spatial features through two-dimensional time-frequency maps, making it more suitable for complex compound fault scenarios, and demonstrating stronger generalization ability in multi-fault classification (12 categories). From an engineering application perspective, MSCNN-GRU is more suitable for scenarios with limited sensor resources and high real-time requirements, while 2D-CNN-GRU is better suited for high-precision multi-fault diagnosis of critical equipment.

3.3. Global Feature Fusion Method based on CNN-Transformer

Although CNN models possess powerful local feature extraction capabilities, their ability to model global contextual information is insufficient, which can easily lead to the omission of dispersed fault features. To address this issue, Xu Kunbo [5] and others proposed a fusion diagnostic model based on Fast Fourier Transform (FFT), dual-channel CNN, and Transformer (FFT-CNN-Transformer). The technical innovations are mainly reflected in the following three aspects: Dual-channel feature extraction in time and frequency domains: By using FFT to convert one-dimensional time-domain signals into frequency-domain signals, a time-frequency dual-channel CNN architecture is constructed. The time-domain channel uses a 1D CNN (convolution kernel 3×1 , channels $1 \rightarrow 32 \rightarrow 64 \rightarrow 128$) to extract local temporal features; the frequency-domain channel uses a similar network structure (channels $1 \rightarrow 16 \rightarrow 32 \rightarrow 64$) to capture frequency distribution features, addressing the limitations of traditional single-domain analysis with limited information. Transformer encoder for global feature fusion: The decoder

part of the Transformer model is removed, retaining only the encoder layers. Multi-head attention is used to dynamically allocate weights to dual-channel features, enhancing the focus on key fault features (such as inner race fault feature frequency 162.23 Hz, outer race fault feature frequency 527.04 Hz) while suppressing non-key features and noise interference. Efficient training strategy: Batch normalization (BN) layers are introduced to accelerate model convergence, Dropout (probability 0.5) is used to prevent overfitting, and the Adam optimizer (learning rate 0.0002) with cross-entropy loss function is chosen to optimize model parameters, improving training efficiency and stability. Experiments on the CWRU rolling bearing dataset (sampling frequency 12 kHz, fault diameter 0.021 in) show that the FFT-CNN-Transformer model achieves an average diagnostic accuracy of 99.85%, significantly outperforming comparative models such as multi-scale graph-Transformer (98.20%), DRSN-Transformer (97.46%), and CNN-Transformer (96.66%). The training time is 15.13 s, improved by 13.8% compared to CNN-Transformer (17.55 s), validating its efficiency. The core value of this model lies in realizing global feature fusion through the Transformer, effectively overcoming the limitations of CNN in local feature extraction, making it suitable for complex scenarios with dispersed fault features and superimposed multi-frequency components. However, it should be noted that the computational complexity of the Transformer's self-attention mechanism is high, and on large-scale datasets, lightweight optimization techniques such as sparse attention and model quantization are required.

3.4. Basic Diagnostic Methods based on Traditional CNN

3.4.1. One-Dimensional CNN: Lightweight Models and Engineering Practical Design

To address the diagnostic cost and real-time requirements of equipment monitoring systems for small and medium-sized enterprises, Jiang Wei and others constructed a basic multi-layer CNN model based on the CWRU dataset, focusing on practical optimizations for engineering scenarios. Its technical features are mainly reflected in the following three aspects: standardized data preprocessing flow: wavelet transform is used to remove signal noise, data augmentation is achieved through time-axis shifting and random noise addition, and short-time Fourier transform (STFT) is applied for time-frequency conversion to improve signal quality; lightweight model structure design: a 2-layer one-dimensional CNN (3×1 convolution kernel) with 2 max-pooling layers is used, ReLU activation function and Adam optimizer are chosen, with the model containing only 120,000 parameters—over 90% fewer than complex fusion models, reducing training time to one-third of traditional models; multi-fault classification performance verification: testing on 10 fault types (including normal condition, inner ring, outer ring, rolling element faults, and various defect diameters) achieved an accuracy of 98.9%, meeting the diagnostic needs of small- to medium-scale industrial scenarios. Although this method does not introduce complex fusion architectures, the combination of standardized preprocessing and lightweight modeling provides a simple and effective diagnostic solution for industrial scenarios with low cost and high real-time requirements, offering significant practical engineering value.

3.4.2. VMD-CNN: Signal Decomposition and Feature Enhancement Collaborative Method

The operating environment of the spindle box bearings in machine tools is complex, influenced by multiple sources of interference such as gear mesh noise and cutting vibrations, making fault feature extraction challenging. To address this issue, Chen Shuori [6] proposed a fusion diagnostic method based on Variational Mode Decomposition (VMD) and CNN (VMD-CNN). The technical innovations mainly lie in the following three aspects: KHA-optimized VMD parameters: Using the Krill Herd Algorithm (KHA) with envelope entropy as the fitness function to optimize the number of decomposition layers and the quadratic penalty factor of VMD, enabling adaptive decomposition of vibration signals to obtain the optimal Intrinsic Mode

Functions (IMFs);Kurtosis criterion to select effective components: Selecting IMF components with the highest kurtosis values for signal reconstruction, emphasizing fault impact features while suppressing background noise interference;CNN fault classification: Inputting the reconstructed signals into a 3-layer CNN model (convolution kernel 5×1, max pooling), achieving a diagnostic accuracy of 91.53% for five machine tool bearing conditions (normal, inner race fault, outer race fault, rolling element fault, and composite fault), outperforming comparative methods such as EMD-CNN (86.59%) and EEMD-CNN (88.20%).The core advantage of this method is that it addresses the incomplete decomposition problem of non-stationary signals through VMD and combines it with CNN for precise fault feature extraction, making it particularly suitable for industrial scenarios with strong impacts and multiple interference sources, such as machine tools and rolling mills. However, VMD has high computational complexity, requiring a balance between decomposition accuracy and real-time performance.

4. Key Technology Comparison and Performance Analysis

4.1. Comparison of Preprocessing Methods

Signal preprocessing is a key step in improving the performance of rolling bearing fault diagnosis. The feature enhancement effects and applicable scenarios of different preprocessing techniques vary significantly. A detailed comparison is shown in Table 1:

Table 1. Comparison of Different Preprocessing Techniques

Preprocessing method	Core Principle	Technical advantage	Limitations	Applicable Scenarios
Differential Continuous Wavelet Transform (DCWT)	First-order difference denoising Continuous wavelet transform	High time-frequency resolution, effectively suppresses high-frequency noise, and preserves transient features	Computationally intensive and sensitive to the choice of wavelet basis functions	Non-stationary signals and scenarios with strong noise interference (such as coal chemical circulating fans and wind power bearings)
Fast Fourier Transform(FFT)	Time-domain signal→Frequency domain signal conversion	The frequency components are clear, the computational efficiency is high, and it is convenient to extract characteristic frequencies.	Only applicable to steady signals, unable to capture transient characteristics	Scenarios with stable operating conditions and known fault characteristic frequencies (such as laboratory simulation tests or fixed-speed motors)
Variational Mode Decomposition(VMD)	Adaptive decomposition into IMF components, minimizing constrained variational problems	High decomposition accuracy, strong anti-mode aliasing capability, suitable for nonlinear signals	Parameter-sensitive, requires optimization with intelligent algorithms	Multi-component, high-impact signal scenarios (such as machine tool spindle boxes, rolling mill bearings)
Short-Time Fourier Transform(STFT)	Sliding window Fourier transform, taking into account both time-domain and frequency-domain information	Realize time-frequency localized analysis with moderate computational complexity	The window size is fixed, and there is a trade-off between time and frequency resolution.	Mid-to-low frequency signals, scenarios with relatively concentrated fault characteristics (such as low-speed rotating equipment)

4.2. Engineering Application Case Analysis

4.2.1. Fault Diagnosis of Coal Chemical Circulating Fan Bearings (COA-CNN)

Bie Fengfeng et al. took the rolling bearings of the circulating fan in a certain coal chemical enterprise as the research object (fan model: unknown, speed 1481 r/min, sampling frequency

12,800 Hz). Using a 1A314E three-axis acceleration sensor, vibration signals were collected from the rear end (measuring point 1), front end (measuring point 2), and output end (measuring point 3) of the fan to carry out engineering application verification. Operating conditions and signal feature analysis: The horizontal vibration intensity at the drive end of Fan No. 5 was 3.16 mm/s (close to the alarm threshold of 4.5 mm/s), and the vibration signal showed a deterioration trend; the vibration intensity of Fan No. 7 was 16 mm/s (exceeding the shutdown threshold of 11.2 mm/s), with no obvious low-frequency fault characteristics and severe high-frequency noise interference. Diagnosis results: The COA-CNN model, through DCWT preprocessing and parameter optimization, achieved a fault identification accuracy of 96.3% for Fans No. 5, 6, and 7, successfully diagnosing the rolling element fault of Fan No. 7, which is 6.8 percentage points higher than the traditional CNN model (89.5%). This provides a precise basis for equipment shutdown maintenance and prevents production losses caused by the escalation of faults.

Chen Shuori and others focused on a certain type of CNC machine tool spindle box rolling bearing (deep groove ball bearing, fault depth 1–2 mm, width 0.2–0.4 mm), collecting vibration signals under five operating conditions (117,000 vibration points per condition) to carry out engineering application verification: Interference and signal feature analysis: During machine tool operation, there are multiple sources of interference such as gear meshing noise and cutting vibrations, resulting in a low signal-to-noise ratio (0–2 dB), which easily masks fault features. Diagnosis results: The VMD-CNN model, through KHA optimization of VMD parameters, achieves effective signal decomposition and feature enhancement. For compound faults (inner and outer rings), the diagnosis accuracy reaches 91.53%, an improvement of 5 percentage points compared to the traditional EMD-CNN method, enabling early warning of machine tool bearing faults and reducing equipment maintenance costs.

5. Challenges and Future Development Directions

5.1. Current Challenges in Research

Although deep learning-based rolling bearing fault diagnosis technology has made significant progress, it still faces the following challenges in industrial applications: Small sample and data imbalance issues: In industrial scenarios, normal operation data of rolling bearings account for over 90%, while fault data (especially early-stage and rare faults) are scarce. Existing deep learning models largely rely on large-scale labeled data, and their generalization ability is insufficient under small-sample conditions; Insufficient adaptability to dynamic operating conditions: Real-time variations in industrial equipment speed and load cause frequency shifts in fault features. Most existing models are trained under fixed conditions, resulting in a 10%~20% drop in diagnostic accuracy across different conditions, making it difficult to meet dynamic operating requirements; Conflict between model lightweight and real-time performance: Complex fusion models (such as FFT-CNN-Transformer) have parameters in the millions to tens of millions and high computational complexity, making them difficult to deploy on edge devices (e.g., sensor nodes, embedded systems), thus failing to meet industrial real-time diagnostic needs; Poor model interpretability: Deep learning models are essentially "black-box" structures, making it hard to explain the relationship between fault features and diagnostic results, which hinders validation and trust by industrial experts and limits practical engineering applications.

5.2. Future Development Directions

To address the above challenges and in consideration of industrial scenario requirements, the future development of rolling bearing fault diagnosis technology can move in the following directions: 1. Development of few-shot learning techniques: By combining transfer learning

(domain adaptation, fine-tuning of pre-trained models), generative adversarial networks (GANs), and other technologies, the problem of scarce fault data can be addressed. For instance, GANs can be used to generate simulated early fault data to expand the training dataset; pre-trained models can learn general features on large-scale public datasets (such as CWRU) and then be fine-tuned with a small amount of industrial data to improve generalization under few-shot conditions. 2. Optimization of robustness under dynamic operating conditions: Introduce adaptive feature alignment techniques (e.g., dynamic time warping (DTW), attention mechanisms) to achieve dynamic matching of fault features; develop variable-speed fault diagnosis models based on instantaneous frequency estimation to accommodate fluctuations in industrial equipment speed; integrate reinforcement learning techniques to allow models to autonomously adjust parameters with changing operating conditions, enhancing dynamic adaptability. 3. Lightweight model design and edge deployment: Employ model compression (pruning, quantization, knowledge distillation), lightweight network architectures (such as MobileNet, EfficientNet), and hardware acceleration (FPGA, GPU) techniques to reduce model parameter scale to below the million level while ensuring diagnostic accuracy, meeting edge device deployment requirements. For example, knowledge distillation can transfer the knowledge of complex fused models to lightweight models to achieve a balance between accuracy and efficiency. 4. Enhancement of model interpretability: Leverage attention visualization (heatmaps), causal inference, and model distillation to reveal the contribution mechanism of key fault features to diagnostic results. For example, heatmaps can intuitively display the time-frequency regions that CNNs focus on, explaining the basis of fault identification; causal inference can analyze the causal relationship between features and fault types, improving model reliability. 5. Construction of a multimodal fusion diagnostic system: Integrate multi-source monitoring data such as vibration, temperature, oil analysis, and acoustic emission to build multimodal deep learning models (e.g., multi-input CNN-Transformer) for more comprehensive and reliable fault diagnosis. For instance, vibration signals can be used to identify early faults, combined with temperature signals to assess fault severity, thereby enhancing diagnostic comprehensiveness.

6. Conclusion

A review of six research articles on rolling bearing fault diagnosis shows that deep learning-based intelligent diagnostic methods, particularly hybrid models that integrate convolutional neural networks (CNN) with recurrent neural networks (such as GRU and LSTM) or attention mechanisms (such as Transformers), have become the current mainstream trend. Compared with traditional methods that rely on manual feature extraction, deep learning methods have stronger adaptive feature extraction capabilities and higher diagnostic accuracy. Models such as COA-CNN, FFT-CNN-Transformer, MSCNN-GRU, and 2D-CNN-GRU have demonstrated excellent diagnostic performance under various experimental conditions, with accuracy rates generally exceeding 95%, and some models achieving over 99% on public datasets. These models effectively enhance robustness and generalization under complex conditions such as noise interference, variable operating conditions, and small sample sizes by incorporating optimization algorithms, multi-scale feature extraction, time-frequency fusion, and attention mechanisms. Moreover, studies generally use the Case Western Reserve University (CWRU) dataset for validation, with some also utilizing datasets from Jiangnan University, XJTU-SY, and others to enhance the engineering applicability of the models. Overall, the future development of intelligent fault diagnosis technology for rolling bearings will focus on integrating spatiotemporal features, optimizing network structures, and improving model lightweight performance and real-time capabilities.

References

- [1] Chen Chao, Guo Jing, Qu Huawei, et al. Fault Diagnosis of Rolling Bearings Based on Multi-Scale Convolutional Neural Networks and Gated Recurrent Units [J]. Bearings, 2025, (09): 109-119. DOI:10.19533/j.issn1000-3762.202308052.
- [2] Jiang Wei, Wang Yongxing. Research on Rolling Bearing Fault Diagnosis Method Based on Convolutional Neural Network [J]. Marine Electrical Technology, 2025, 45(07): 7-10. DOI:10.13632/j.meee.2025.07.002.
- [3] Bie Fengfeng, Zhou Zhaolong, Li Qianqian, et al. Research on Rolling Bearing Fault Diagnosis Method Based on COA-CNN [J]. Noise and Vibration Control, 2025, 45(04):136-142. DOI: CNKI:SUN:ZSZK.0.2025-04-022.
- [4] Zhang Xiong, Qu Weiyang, Wang Wenqiang, et al. Rolling Bearing Multi-Fault Diagnosis Model Based on Improved CNN-GRU Model [J/OL]. Mechanical and Electrical Engineering, 1-10 [2025-10-15]. <https://link.cnki.net/urlid/33.1088.TH.20250521.1526.010>.
- [5] Xu Kunbo, Liu Dongjun, Zong Zekai, et al. Rolling Bearing Fault Diagnosis Based on FFT-CNN-Transformer [J/OL]. Journal of Naval Aviation University, 1-10 [2025-10-15]. <https://link.cnki.net/urlid/37.1537.V.20250429.1122.002>.
- [6] Chen Shuori, Han Xianfeng. Application Research on Machine Tool Bearing Fault Diagnosis Based on Convolutional Neural Networks [J]. Mechanical Management Development, 2025, 40(09):87-89. DOI:10.16525/j.cnki.cn14-1134/th.2025.09.030.