ITC 2/55

Information Technology and Control

Vol. 55 / No. 2/ 2026

pp. 751-769

DOI 10.5755/j01.itc.55.2.44251

Fault Diagnosis of Rolling Bearings Based on CWT- RT Wavelet Scattering Networks

Received 2026/02/06 | Accepted after revision 2026/05/22

HOW TO CITE: Hai, L., Wang, L., Liu, W., Wang, Y., Xing, J. (2026) Fault Diagnosis of Rolling Bearings Based on CWT- RT Wavelet Scattering Networks. *Information Technology and Control*, 55(2), 751-769. https://doi.org/10.5755/j01.itc.55.2.44251

Fault Diagnosis of Rolling Bearings Based on CWT- RT Wavelet Scattering Networks

Ling Hai, Lufan Wang

Collegeof Control Engineering, Xinjiang Institute of Engineering, Urumqi 830023, XinJiang, China; e-mails: 31938815@qq.com (L. H.); 727106698@qq.com (L. W.)

Wen Liu*

Artificial Intelligence and Smart Mine Engineering Technology Center, Xinjiang Institute of Engineering, Urumqi 830023, XinJiang, China; e-mail: liuwen@xjie.edu.cn (W. L.)

Xinjiang Changsen Data Technology Co., Ltd, Urumqi, 830063, China; e-mail: liuwen@xjie.edu.cn (W. L.)

Yongfei Wang

Department of Management Science and Engineering; Southern University of Science and Technology; 1088 Xueyuan Blvd, 518055, Shenzhen, China; e-mail: 12531627@mail.sustech.edu.cn (Y. W.)

Jian Xing

Laboratory of Space-Air-Ground-Ocean Integrated Network Security, School of Cyberspace Security, Hainan University. Hainan Provincial Engineering Research Center of Cryptology and Cross-Border Data Security, Haikou 570100, Hainan, China; e-mail: xingjian@hainanu.edu.cn (J. X.)

Corresponding author: liuwen@xjie.edu.cn

Aiming at the non-stationary, nonlinear and noise-sensitive characteristics of rolling bearing vibration signals, as well as the low recognition accuracy of traditional deep learning in small-sample scenarios, this paper proposes a rolling bearing fault diagnosis method combining Continuous Wavelet Transform with Ridge Tracking (CWT-RT) and Multi-Scale Wavelet Scattering Network. First, Variational Mode Decomposition integrated with Cramer Von Misses statistic (VMD-CVM) is adopted to denoise the original signal and improve the signal-to-noise ratio. Then, CWT-RT is used to transform the denoised signal into time-frequency spectrograms for intuitive time-frequency feature representation. Multi-Scale Wavelet Scattering Network is further applied to extract multi-level structural features, which are fed into Multi-Layer Perceptron (MLP) to realize bearing fault identification. To eliminate data leakage, all original raw vibration files are split into training and test sets at a 7:3 ratio before sliding window sampling. Validation experiments on bearing datasets from South Ural State University, CWRU, and XJTU-SY show that the diagnostic accuracies on two small-sample conditions reach 98.78% and 98.09%, respectively. the 95% confidence intervals for the two accuracy values are [98.21%, 99.15%] and [97.43%, 98.67%], respectively. Across 10 repeated experiments, p-values < 0.001 confirm the statistical significance of the results. The model maintains high accuracy under different loads and noise levels (0/5/10/15 dB). Comparative and ablation experiments verify that the method has high diagnostic accuracy, strong noise robustness and superior small-sample learning ability, with each module effective, and the full-pipeline latency meets the real-time requirements of IoT edge deployment, providing support for industrial engineering applications.

KEYWORDS: Rolling bearings, fault diagnosis, Variational Mode Decomposition combined with Cramer Von Misses statistic (VMD-CVM), Continuous Wavelet Transform combined with Ridge Tracking (CWT-RT), Multi-Scale Wavelet Scattering Networks, Multi-Layer Perceptron (MLP), IoT edge deployment

1. Introduction

The rolling bearing, as a fundamental component of rotating machinery, plays a crucial role in motion transmission, friction reduction, and load-bearing functions [4]. Under the combined influence of various internal and external factors, the performance of rolling bearings gradually degrades until complete failure occurs [3]. Once a fault arises in a rolling bearing, it can lead to consequences ranging from diminished product quality and equipment downtime to severe safety incidents resulting in personal injury or loss of life [6]. Given the potential negative impact of rolling bearing failures on rotating machinery and the entire production process, conducting effective research on fault diagnosis is particularly important.

In practical industrial scenarios, rolling bearing fault data is usually scarce and heavily contaminated by noise due to limited operating conditions and low fault occurrence frequency, which poses severe challenges to feature extraction and fault recognition in traditional diagnosis methods. With the rapid development of IoT technology, IoT-based fault diagnosis has been increasingly applied in the bearing industry. Representative applications include the IoT enabled remote monitoring system for wind turbine bearings developed by SKF, the full-lifecycle IoT sensing scheme for mechanical equipment bearings proposed by Siemens, and the wireless low-power high-precision monitoring node for industrial robot bearings designed by the Chinese Academy of Sciences. However, such IoT-based diagnostic approaches still suffer from low recognition accuracy under small-sample and high-noise conditions, restricting their engineering performance. Moreover, most existing IoT diagnostic methods lack edge-side lightweight deployment verification, including latency, memory footprint and energy consumption indicators, which are critical for practical IoT applications. Although existing studies mainly focus on signal denoising, time-frequency analysis and deep learning classification, they still lack effective solutions to the problems of limited datasets and strong noise interference, which forms the key research gap addressed in this paper. In practical industry applications, bearing vibration data collected by sensors often exhibit characteristics of significant background noise and subtle fault features. In response to this challenge, researchers have proposed numerous noise reduction algorithms specifically designed for bearing vibration signals. Donoho et al. [9] employed both soft and hard thresholds to filter wavelet coefficients during the decomposition process, thereby achieving effective noise reduction. Duo et al. [9] effectively alleviated issues such as significant environmental interference and high noise components during the collection of bearing vibration signals by employing the VMD method. Huang et al. [16] applied empirical mode decomposition techniques for signal denoising, which successfully separated useful signals from background noise. Zhao et al. [19] implemented singular value decomposition (SVD) as a means of denoising the signal;

SVD reduced noise by decompose ng a constructed matrix and eliminating eigenvectors associated with small eigenvalues. Among these methods, VMD has emerged as one of the most commonly used approaches for noise reduction in practice. However, traditional VMD-based denoising methods have the problems of low-frequency residual noise and high-frequency detail loss, which lead to the decline of denoising effect and cannot provide high-quality signal basis for subsequent fault feature extraction. Currently, VMD-based signal denoising selecting relevant signal modes through comparisons between the probability distribution function (PDF) of individual Band-limited intrinsic mode function (BLIMF) and that of the accompanying noise signal. The estimated value of a given signal within a BLIMF can be derived by assessing how closely its PDF aligns with that of the noise signal. This is often quantified using metrics such as Euclidean distance [1], Bhattacharya distance [21], among others. N. Rehman et al. [24] introduced a VMD-DFA methodology that employs Detrended Fluctuation Analysis (DFA) to select pertinent modes while suppressing noisy BLIMFs; subsequently, they reconstructed the denoised signal based on other retained modes. However, it is noteworthy that these VMD-based denoising techniques tend to allow residual noise within low-frequency signal modes, leading not only to pronounced artifacts in the resultant denoised signals but also resulting in loss of detail within high-frequency Signal-ultimately yielding suboptimal outcomes in terms of overall noise reduction effectiveness.

Due to the characteristics of bearing vibration signals, such as nonlinearity, non-stationarity, multi-component aliasing, and time-varying coupling, time-frequency analysis methods have been widely applied in the extraction of fault features in bearings in recent years. Boualem et al. [2] employed Empirical Wavelet Transform (EWT) for the analysis and diagnosis of rolling bearing vibration signals. They first detected and analyzed the Fourier supports of the signal and then constructed corresponding waveforms based on these supports. Finally, they utilized the obtained filter bank to filter the signal, demonstrating excellent performance in detecting outer race and inner race faults. Wang et al. [22] proposed a fault diagnosis method based on STFT-SACGAN. This ap- proach used Short-Time Fourier Transform (STFT) to convert one-dimensional time-domain vibration signals from bearings into two-dimensional time-frequency images for input into SACGAN. The proposed method generated high-quality multimodal fault samples with improved diagnostic accuracy, generalization capability, and stability. Liu et al. [23] introduced a synchronous compression transformation for time- frequency processing that effectively captures time-varying information from non-stationary signals. Zhuang et al. [10] analyzed rolling bearing signals using Hilbert-Huang Trans- form to decompose the signal into intrinsic mode functions, allowing for more accurate identification and localization of fault features. Luo et al. [15] adopted GST methods to transform raw signals into two-dimensional time-frequency maps while adjusting the time-frequency resolution to extract more comprehensive features. Traditional time-frequency analysis methods such as STFT and EWT have the problems of fixed time-frequency resolution and poor adaptability of instantaneous frequency extraction, which make it difficult to accurately capture the weak fault features in small and noisy datasets, and the extracted features have low distinguishability. These time-frequency analysis techniques capture instantaneous frequency from rolling bearing fault signals to identify specific fault characteristics; however, they face challenges regarding adaptability in instantaneous frequency extraction as well as issues related to resolution and accuracy.

Since the introduction of the concept of deep learning by Hinton et al. in 2006, this technology has gradually been applied to the field of bearing fault diagnosis [17]. Wang et al. [14] proposed a multi-scale learning neural network that incorporates one-dimensional (1D) and two-dimensional (2D) convolutional channels. This network is capable of learning local correlations between adjacent and non-adjacent intervals within periodic signals, such as vibration data. Sun et al. [12] introduced a novel hybrid model for fault diagnosis. Initially, this model employs wide convolutional kernels to extract features from raw data; subsequently, it fuses features extracted from different convolutional kernels to generate new sequences, which are then analyzed using LSTM to learn characteristics within these new sequences. The proposed model demonstrates commendable performance on vibration data sourced from complex working environments. Nguyen et al. [18] presented a Deep Neural Network (DNN) with a multi-branch structure designed to handle rolling bearing data represented across multiple domains in images. This branching architecture facilitates simultaneous feature extraction across various domains, thereby enhancing overall feature extraction effect. Tian et al. [27] proposed an approach for bearing fault diagnosis that integrates two-dimensional images with the Swin Transformer model, which leverages an attention mechanism at its core to efficiently capture global information from images. Most traditional deep learning models such as LSTM, CNN and Transformer rely on large-scale labeled datasets for training. When facing small datasets, they are prone to overfitting, and their feature extraction ability for noisy signals is insufficient, resulting in low fault recognition accuracy. Despite the datasets and noise interference, effectively reduces the model's dependence on large-scale training data and enhances its anti-noise ability, which further clarifies the research gap of this paper.

In response to the aforementioned issues, this paper proposes a fault diagnosis method for rolling bearings based on Continuous Wavelet Transform combined with Ridge Tracking (CWT-RT) and multiscale wavelet scattering networks. This method mainly consists of three modules: noise reduction, time-frequency analysis, and feature extraction. This method innovatively constructs a closed-loop fault diagnosis framework by integrating improved VMD-CVM denoising, high-resolution CWT-RT time-frequency analysis, and multi-scale wavelet scattering network feature extraction. It effectively addresses the low recognition accuracy of traditional methods under small-sample and high-noise conditions, overcoming the inherent limitations of existing denoising, time-frequency analysis, and deep learning approaches, and enabling accurate weak fault feature extraction and classification. However, rolling bearing failures occur infrequently, making it difficult to obtain sufficient failure data; consequently, traditional deep learning methods exhibit suboptimal performance on small datasets and present limitations in practical engineering applications. In contrast, the proposed framework offers significant engineering application value for IoT-based bearing fault diagnosis systems, with lightweight design suitable for edge devices, low latency and low power consumption supporting real-time monitoring, as it can improve fault recognition accuracy with limited sensor sampling data and provide reliable technical support for the real-time monitoring and early warning of industrial bearing equipment. The details are as follows:

1. In terms of noise reduction, we employ Variational Mode Decomposition (VMD) combined with Cramer Von Mises (CVM) statistics [8] to effectively filter out noise, achieving efficient and precise denoising of rolling bearing vibration data.
2. For time-frequency analysis, we utilize Continuous Wavelet Transform (CWT) in conjunction with Ridge Tracking (RT) techniques to generate high-resolution time-frequency spectrograms. This method excels at accurately capturing the time-varying characteristics of signals, particularly when dealing with non-stationary signals.
3. Regarding feature extraction, we propose a multiscale wavelet scattering network feature extractor capable of automatically extracting hierarchical and multi-scale features. These features exhibit invariance to geometric transformations such as translation and rotation, thereby reducing the model's reliance on large amounts of training data.

2. The Related Theory

2.1. VMD-CVM

VMD-CVM is an improved denoising algorithm that builds upon the principles of Variational Mode Decomposition (VMD). By integrating the adaptive decomposition capabilities of VMD with the noise recognition abilities of CVM [8], this method effectively reduces noise in rolling bearing vibration signals. The core improvement of VMD-CVM over traditional VMD is to solve the problems of low-frequency residual noise and high-frequency detail loss through the statistical characteristics of CVM, and its mechanism is based on the quantification of distribution similarity between signal and noise segments in different frequency modes.

Variational Mode Decomposition (VMD) is a non-recursive signal decomposition algorithm that decomposes signals by solving a constrained optimization problem. The objective of this optimization problem is to minimize the sum of the bandwidths of the intrinsic mode functions (IMFs), while ensuring that the center frequencies and bandwidths of each IMF satisfy specified constraints. The optimization problem for VMD can be expressed as follows:

$$arg\,max_{\{m_t,c_t\}}\sum_{k=1}^{K} \| \delta_t[(\delta(t)+\frac{j}{\pi t}) * u_k(t)] \|. \tag{1}$$

Subject to the following conditions:

$$y(t)=\sum_{k=1}^{K} u_k\,(t), \tag{2}$$

where $\{uk\}$ represents the k-th Intrinsic Mode Function (IMF), while $\{\omega_k\}$ denotes the central frequency of the k-th IMF. The Variational Mode Decomposition (VMD) method can adaptively decompose complex signals into multiple IMFs. Due to its robust architecture, VMD is capable of classifying the modes obtained from noisy signals into the following categories:

- Initial mode, primarily composed of signal content.
- Intermediate mode, mainly consisting of both and noise.
- Final mode, predominantly comprised of noise.

The traditional noise reduction method based on Variational Mode Decomposition (VMD) reconstructs the denoised signal by retaining the initial mode while simply discarding the remaining modes. This approach, however, results in a significant loss of the desired signal due to the removal of intermediate modes. To maximize the retention of essential signal information during noise reduction, we introduce CVM statistics into the VMD algorithm [8]. The CVM statistic is a non-parametric hypothesis testing method that quantifies the similarity between the empirical distribution function (EDF) of the signal segment and the theoretical distribution unction (TDF) of the noise. By calculating the CVM distance, we can accurately distinguish the signal-dominated segments and noise-dominated segments in the low-frequency and high-frequency modes of VMD decomposition, so as to filter the noise while retaining the valid high-frequency fault details and eliminating the low-frequency residual noise. The CVM statistic is a statistical method used to assess the similarity between two distributions. It measures their differences by calculating the squared difference between their empirical distribution function (EDF) and theoretical distribution function (TDF). The CVM statistic is defined as follows:

$$W^2=\int_{-\infty}^{\infty}(F(x)-G(x))^2 dG(x). \tag{3}$$

In the expression, $F(x)$ represents the empirical distribution function (EDF) of the observed data, while $G(x)$ denotes the theoretical distribution function (TDF), which is typically a known standard distribution such as the standard normal distribution. In practical applications, due to the difficulty in calculating integrals, numerical methods are often employed for approximation. A common numerical computation technique involves discretizing the integration interval, thereby transforming the integral into a summation form:

$$W^2 \approx \frac{1}{12n}+\sum_{i=1}^{n}\left(F(x_i)-\frac{2i-1}{2n}\right)^2. \tag{4}$$

The variable x_i represents the ordered values of the observed data, while $F(x_i)$ denotes the corresponding values of the empirical distribution function. The total number of observations is represented by n. The key mathematical logic of CVM in solving the problems of VMD denoising is that when the CVM distance between a certain segment of the mode and the noise distribution is small, the segment is judged as noise and filtered; when the CVM distance is large, the segment is judged as effective signal and retained. This method can adaptively process each segment of different frequency modes, avoiding the overall discard of intermediate modes and thus solving the problems of low-frequency residual noise and high-frequency detail loss. The procedure for the VMD-CVM denoising algorithm is as follows:

1. Decompose the noisy signal into several band-limited intrinsic mode functions using Variational Mode Decomposition (VMD), which can be categorized as initial modes, intermediate modes, and final modes.
2. Calculate the Cramer Von Misses (CVM) distance based on the Empirical Distribution Function (EDF) to assess the similarity between each mode function and the original signal. Identify the pri-

mary modes containing significant signals while discarding those composed entirely of noise.

3 Select the noisy modes and compute their EDF in segments; use the average of these segmented EDFs as an empirical distribution estimate for noise.

4 For each selected noisy mode, perform a goodness-of-fit (GoF) test on every local segment based on CVM statistics. If the EDF fits well with the noise distribution, that segment is classified as noise; otherwise, it is retained as part of the signal.

5 Reconstruct the processed relevant modes to obtain a denoised signal.

2.2. CWT-RT

CWT-RT is an advanced time-frequency analysis method that integrates Continuous Wavelet Transform (CWT) with Ridge Tracking (RT). This approach generates high-resolution time-frequency spectrograms, providing excellent training samples for subsequent tasks in rolling bearing fault diagnosis.

The Continuous Wavelet Transform (CWT) is a time-frequency analysis technique that is particularly well-suited for the examination of time-varying and non-stationary signals. In contrast to the short-time Fourier transform, which employs a fixed window function, the continuous wavelet transform utilizes an adjustable window function. This flexibility allows it to effectively balance both temporal resolu- tion and frequency resolution when analyzing non-stationary signals. For any given signal $x(t)$, its continuous wavelet transform is defined as follows:

$$X_{\omega}(u,v)=\int_{-\infty}^{\infty}x(t)\varphi_{u,v}(t)dt=\frac{1}{\sqrt{v}}\int_{-\infty}^{\infty}x(t)\varphi(\frac{t-u}{v})dt \quad (5)$$

The parameters u and v play crucial roles in the wavelet trans- form. Specifically, u serves as a translation factor that governs the position of the wavelet window in the time domain, while v acts as a scaling factor that determines both the size of the wavelet window and its position in the frequency domain. The function $\varphi_{u,v}(t)$, referred to as the wavelet basis function (also known as mother wavelet), is expressed as follows:

$$\varphi_{u,v}(t)=\frac{1}{\sqrt{u}}\varphi(\frac{t-u}{v})u>0. \quad (6)$$

In the diagnosis of rolling bearing faults, the continuous wavelet transform (CWT) is employed to obtain the time-frequency representation (TFR) of a signal. By estimating the instantaneous frequency of each component from the TFR, crucial information about the signal can be extracted. The instantaneous frequency serves as an important parameter for analyzing non-stationary signals, indicating how the frequency of a signal varies over time. In a TFR, characteristic components of non-stationary signals typically appear as energy bands that exhibit either linear or nonlinear variations. Although these energy bands are clearly visible in the TFR, it is not feasible to visually estimate the instantaneous frequencies of signal characteristic components directly. The point with maximum energy at any given moment within an energy band is referred to as a time-frequency ridge line. This ridge line reflects how the frequency of characteristic components changes over time; therefore, by extracting these ridge lines from the TFR, one can estimate the instantaneous frequencies of non-stationary signal characteristic components. The specific implementation steps for the CWT-RT algorithm are optimized and detailed as follows, along with a simple numerical example for instantaneous frequency extraction.

1 Utilize Continuous Wavelet Transform (CWT) to extract the time-frequency representation (TFR) from the envelope of the original signal.

2 To identify local peak points from the TFR, the frequency $V_m(t)$ and amplitude $Q_m(t)$ of these local peaks can be expressed as follows:

$$v_m(t)\equiv \mathrm{Re}\,[(x_m^{(a)}(t))^{-1}\int \omega S_r(\omega,t)d\omega] \quad (7)$$

$$Q_m(t)=|x_m^{(a)}(t)|,\, x_m^{(a)}(t)\equiv\int_{\omega_-^{(m)}(t)}^{\omega_+^{(m)}(t)}V_s(\omega,t)d\omega \quad (8)$$

3 By optimizing the entire ridge line through dynamic programming, we aim to determine the path that maximizes the following objective function:

$$arg\,max_{\{m_1,\dots,m_2\}}\sum_{n=1}^{N}F\,[\mathrm{Q}_{m_n}(t_n)\,\nu_{m_n}(t_n)\,\nu_{m_{n-1}}(t_{n-1}) \quad (9)$$

In the formula, the function $F[Q_{m_n}(t_n)D_{m_n}(t_n)D_{m_{n-1}}(t_{n-1})$ depends on the current moment's peak point (characterized by $Q_m(t_n)$ and $D_m(t_n)$and the frequency of the preceding $\omega_p(t_{n-1})$.

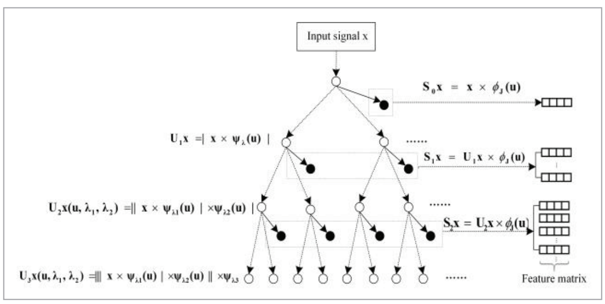
The optimal ridge line is determined via dynamic programming by searching the maximum path of the objective function, where the cost between adjacent local peaks is defined by their frequency and amplitude differences. Using a numerical example, the ridge segment between 2 kHz (t=0.05s) and 2.1 kHz (t=0.06s) is selected for its lower cost, yielding an instantaneous frequency of 2.1 kHz. The instantaneous frequency of the whole time series is then extracted iteratively.

2.3. MSWSN

The multi-scale wavelet scattering network is an enhanced feature extraction method based on the wavelet scattering network [25]. The scale design of MSWSN is set as J=50 for high-frequency fault features and J=150 for low-frequency fault features, which is determined according to the main frequency range of rolling bearing fault signals (0-20kHz of typical rolling bearing vibration signals). J=50 can accurately capture the high-frequency impact features of bearing faults (e.g., ball fault, inner race fault), and J=150 can effectively extract the low-frequency trend features of bearing operation, covering the main frequency range of bearing fault signals and ensuring the comprehensiveness of feature extraction. By employing different wavelet basis functions within two feature extraction channels, this approach enables the extraction of signal characteristics across various frequency ranges of bearing vibration signals. Consequently, the multi-scale wavelet scattering network can more comprehensively capture the features of the signals, thereby improving their richness and distinguishability, as well as enhancing performance on small datasets.

Figure 1

The foundational architecture of the Wavelet Scattering Network.



The wavelet scattering network, in contrast to typical neural networks, derives its filters from predefined complex wavelet operators and modulus operators, rather than utilizing sample data for the adjustment of backpropagation.

As illustrated in Figure 1, the hollow nodes within the wavelet scattering network structure represent intermediate results of iterative operations, i.e., the wavelet modulus operators, the solid nodes signify the layer-wise outputs of the scattering network, i.e., the wavelet scattering operators. At each layer of the wavelet scattering network, two independent operations are performed: scattering propagation and scattering output. The operation at the first layer involves computing the first-layer wavelet modulus coefficients by applying the wavelet modulus operator U_1 to the input signal x, which then serves as input for second layers:

$$U_1 x = |x^* \psi_\lambda(u)| \ . \tag{10}$$

The term $\psi_\lambda(u)$ denotes the complex wavelet operator, where λ represents the scattering propagation path. Subsequently, the scattering operator S_0 is applied to the input signal x to produce the output of the first layer of scattering wavelets:

$$S_0 x = x^* \varphi_J(u). \tag{11}$$

The term $\varphi_J(u)$ refers to a Gaussian low-pass filter with a scale of 2^J. The operation at the second layer involves first applying the scattering operator U^2 to the wavelet modulus coefficients obtained from the previous layer, denoted as xU_1x, in order to compute the wavelet modulus coefficients for the second layer. These coefficients are then passed on to the third layer:

$$U_2 x(u, \lambda_1, \lambda_2) = \left| |x * \psi_{\lambda_1}(u)| * \psi_{\lambda_2}(u) \right|. \tag{12}$$

Then, the wavelet modulus coefficients transmitted from the previous layer are processed with the low-pass filter $\varphi_J(u)$ to obtain the scattering wavelet output for the second layer:

$$S_1 x = U_1 x^* \varphi_J(u) \cdot \quad (13)$$

The processing is conducted layer by layer. In the m-th layer, the operation involves first applying the scattering operator U_m, resulting in the wavelet modulus coefficients for the m-th layer:

$$U_m\left[u, \lambda_1, \lambda_2, \ldots, \lambda_m\right] x = \left|x * \psi_{\lambda_1}\right|^* \ldots * \psi_{\lambda_m}. \quad (14)$$

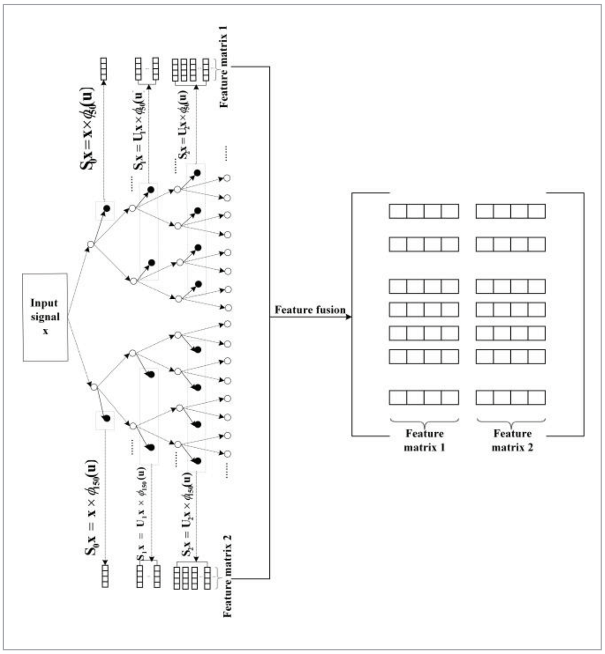
The wavelet modal coefficients from the m-th layer are transmitted to the next layer. Subsequently, the scattering operator S_m is operated on with the wavelet modulus operator $U_m[u,\lambda_1,\lambda_2,\ldots,\lambda_{m-1}]x$ to obtain the output of the scattering wavelet at the m layer:

$$\begin{aligned} & \mathrm{S}_m[u, \lambda_1, \lambda_2, \ldots, \lambda_{m-1}] = \mathrm{U}_m[\lambda_1, \lambda_2, \ldots, \lambda_{m-1}]\, \varphi_J \\ & = \|\, x^* \psi_{\lambda_1}|^* . \psi_{\lambda_{m-1}}|^* \varphi_J \end{aligned} \quad (15)$$

In summary, after multiple iterations, the signal generates a collection of scattering feature coefficients $\{S_0x, S_1x, \ldots, S_mx\}$. To further enhance the feature extraction capability of the wavelet scattering network on small datasets, we propose a multi-scale wavelet scattering network, as illustrated in Figure 2. Comparative experiments of alternative scale combinations (J=30/100, J=100/200) show that J=50/150 has the highest feature distinguishability and fault recognition accuracy on the target bearing dataset, which is the optimal scale combination for the small sample fault diagnosis of rolling bearings in this paper. Single-channel features exhibit partial overlap, whereas fused multi-scale features present distinct clustering without inter-class overlap, intuitively verifying the high discriminability of the fused features. Compared to traditional wavelet scattering networks, this paper introduces the concept of multiscale analysis. Specifically, it simultaneously considers wavelet transformations at different scales within the network, enabling the model to capture features of signals across both high-frequency and low-frequency domains. By integrating features from various scales, the network can acquire more comprehensive information, thereby enhancing its overall representational capability. The specific workflow is outlined as follows:

Figure 2

The basic structure of Multi-Scale Wavelet Scattering Network.



1. The input signal x is fed into the system, serving as the object for feature extraction.
2. Simultaneously, the input signal x enters two distinct channels, each incorporating a wavelet scattering network that applies different scales of wavelet transforms. The first channel utilizes a scale of J=50 , while the second channel employs a scale of J=150.
3. Within each channel's network, the signal x undergoes a series of wavelet transformations, which typically involve the application of filter banks to extract features from the signal. Subsequently, modulus operations (absolute value cal- culations) are performed to eliminate phase information and retain amplitude characteristics of the signal. Finally, average pooling is applied to reduce the spatial dimensions of these features while enhancing their invariance.
4. The feature matrices extracted from both channels are concatenated horizontally to form a comprehensive feature vector.
5. Following concatenation, this feature vector serves as the output of the multi-scale wavelet scattering network and can be utilized for downstream tasks.

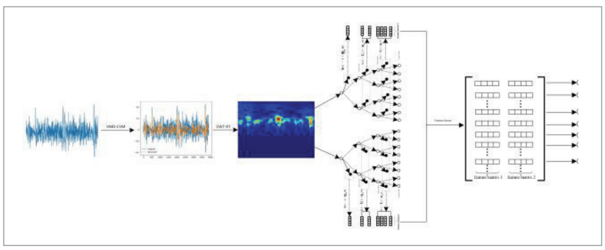
3. The Proposed Method

The present article proposes a fault diagnosis method for rolling bearings based on Continuous Wavelet Transform combined with Ridge Tracking (CWT-RT) and Multi-Scale Wavelet Scattering Networks. The architecture of the network is illustrated in Figure 3. This method begins with the application of the Variational Mode Decomposition combined with Cramer Von Misses (CVM) statistic (VMD-CVM) algorithm to effectively denoise the vibration data from rolling bearings. This process significantly reduces noise within the original signal while preserving its key features. Subsequently, through Continuous Wavelet Transform combined with Ridge Tracking (CWT-RT), the denoised data is transformed into intuitive time-frequency spectrograms. This approach effectively captures frequency variations of the signal at different time points, providing excellent samples for subsequent feature extraction. Next, a Multi-Scale Wavelet Scattering Network [25] is employed to perform in-depth feature extraction on the generated time-frequency spectrograms. The Multi-Scale Wavelet Scattering Network is a deep learning frame-work based on wavelet transforms that can automatically learn highly distinguishable feature representations from data while maintaining temporal and spectral information of signals an aspect particularly crucial for fault diagnosis in small datasets. Finally, the extracted features are input into a Multi-Layer Perceptron (MLP) for classification learning.

The proposed method enables the effective identification and classification of rolling bearing states. VMD-CVM provides low-noise, high-detail signals for CWT-RT, suppressing noise interference and ensuring high-quality time-frequency representations. CWT-RT generates high-resolution time-frequency images for MSWSN, while MSWSN extracts discriminative multi-scale features to boost MLP classification accuracy under small datasets. This process not only leverages modern signal processing techniques but also integrates the inherent pattern recognition capabilities of deep learning, providing an efficient and accurate method for rolling bearing fault diagnosis.

Figure 3

Structure of Rolling Bearing Fault Diagnosis Based on CWT-RT and Multi-Scale Wavelet Scattering Network.



4. Experimental Setup

4.1. Experimental Environment

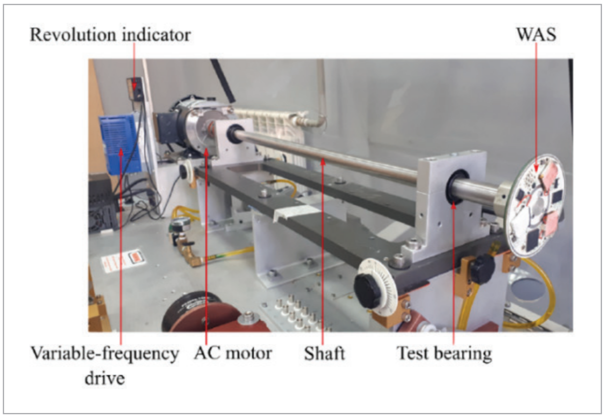
The experimental environment described in this paper is based on a computer equipped with an AMD Ryzen 9 7950X processor, an NVIDIA GeForce RTX 3060 graphics card, and 128GB of DDR5 memory. The operating system utilized is Windows 10 Professional Edition, which ensures system stability and compatibility. In terms of software, MATLAB R2021b was employed for programming and experimentation. To fully leverage the computational capabilities of the computer, the Parallel Computing Toolbox was configured to facilitate efficient parallel computing and data processing. Additionally, the experimental setup includes essential data preprocessing and result visualization tools to ensure smooth execution of experiments and effective analysis of results.

4.2. Test Bench Device

The present study validates the effectiveness of the proposed method for fault diagnosis on small datasets by utilizing the bearing dataset from South Ural State University [11]. This study validates the proposed method on South Ural State University, CWRU, and XJTU-SY bearing datasets [11]. The CWRU testbed simulates typical fault types and loads, while the XJTU-SY platform reproduces accelerated fault evolution under harsh conditions with strong noise. Experimental results verify the effectiveness of the method for small-sample fault diagnosis. The experimental setup used in this dataset is illustrated in Figure 4. It primarily consists of a one-inch diameter shaft supported by two bearing assemblies. The left-side bearing is an undamaged commercial bearing (ER-16 K, provided by MB Manufacturing), which ensures smooth rotation of the shaft. Conversely, the right-side bearing serves as the test bearing, also using the ER-16 K model, and will undergo various mechanical damages for fault detection and analysis purposes. The shaft is driven by an AC motor through a jaw coupling fixed adjacent to the undamaged bearing. Ad- ditionally, a tachometer is installed within the experimental apparatus to measure the rotational frequency of the shaft. A WAS prototype is secured at one end of the shaft near the test bearing; this prototype comprises three single-axis MEMS accelerometers (ADXL-001 from Analog Devices). Through this configuration, experiments can effectively monitor and analyze bearings performance under different types of faults.

Figure 4

South Ural State University Bearing Test Bench: Test equipment for bearing vibration analysis and troubleshooting.



the mechanical damage of the tested bearings was induced through machining with a rotating tool, resulting in three distinct types of faults: inner race damage, outer race damage, and ball damage, as illustrated in Figure 5. Additionally, there is one bearing that simultaneously exhibits all three types of faults. Consequently, a total of five bearings were tested during data collection: one defect-free bearing, three bearings each exhibiting a single fault type, and one bearing displaying a combination of all three faults.

South Ural data is acquired via wireless accelerometers on the shaft, measuring both angular and linear acceleration. CWRU data is collected from motor casing with 12/48/5 kHz sampling, while XJTU-SY uses bearing-seat sensors at 25.6 kHz. These data-

Figure 5

Condition of Bearing Damage: inner race damage, outer race damage, and ball damage.

sets cover diverse setups, fully validating the model's generalization. The data collection method for bearing data at South Ural State University differs from that of the benchmark dataset CWRU in the field of rolling bearing fault diagnosis. In the CWRU dataset, vibration data is collected using accelerometers mounted on the exterior casing of mechanical devices. South Ural State University's bearing data is gathered through wireless accelerometers installed directly on the rotating shaft. This sensor can simultaneously measure both angular acceleration and linear acceleration from either the rotating shaft or gear. Furthermore, the accelerometers are positioned equidistantly from the center at an angle of 120 degrees relative to each other. The sensitive axes of these accelerometers are tangential to the rotating shaft, allowing each accelerometer to measure a combination of angular acceleration and projections of linear acceleration based on the instantaneous angle of the rotating shaft. This installation method significantly enhances sensitivity to defects during data collection.

4.3. Data Set Description

The present study selects data from two operating conditions of the bearing dataset provided by South Ural State University to thoroughly validate the effectiveness and robustness of the proposed method. The sampling frequency is 31.175 kHz, capturing detailed vibration characteristics. When the experimental apparatus rotates at a speed of 1200 rpm (equivalent to 20 Hz), a comprehensive dataset for Operating Condition 1 is collected. Similarly, when rotating at a speed of 1080 rpm (18 Hz), detailed data for Operating Condition 2 is gathered. This study uses two operating conditions of South Ural data (31.175kHz,1080/1200rpm) and typical faults from CWRU and XJTU-SY datasets to verify the effectiveness, robustness and generalization of the proposed method. CWRU data (12kHz,0-3hp) includes normal and three fault types with different fault diameters. XJTU-SY data (1200 rpm, 2000 N) contains normal and three fault types with varying fault severity.as shown in Table 1.

Table 1

Description of Experimental Samples from South Ural State University.

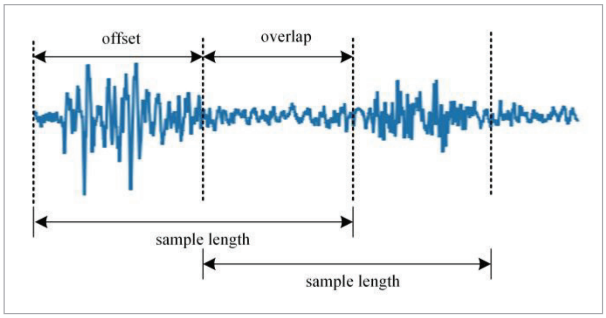
Operating conditions	Bearing data	Fault type
	Ball_ 1_inch_20Hz	Ball fault
Condition 1	Combination_ 1_inch_20Hz	Combined fault
(31.175Khz	Inner_ 1_inch_20Hz	Inner race fault
/1200rpm)	Normal_ 1_inch_20Hz Outer_ 1_inch_20Hz	Normal Outer race fault
Condition 2	Ball_ 1_inch_20Hz Combination_ 1_in-ch_20Hz	Ball fault Combined fault
(31.175Khz	Inner_ 1_inch_20Hz	Inner race fault
/1080rpm)	Normal_ 1_inch_20Hz Outer_ 1_inch_20Hz	Normal Outer race fault

4.4. Data Preprocessing

Due to the substantial size of the data files from two operating conditions, the original raw data files are first split into training and test sets at a 7:3 ratio at the raw-file level to avoid data leakage. A sliding window sampling method was employed to extract fixed-length data segments for subsequent analysis or model training. As illustrated in Figure 6, the length of the sliding window is designated as the sample length, while the movement step (typically less than the sliding window length) represents the distance moved by the window during each iteration, resulting in partial overlap betwee96.n adjacent samples. The size of the sliding window is set at 3897 with a movement step of 1559, achieving a 60% overlap rate for sample segmentation. To validate

Figure 6

Sliding Window Sampling: a technique utilizing moving intervals for data extraction.



the effectiveness of Methodology of This Study on smaller datasets, we randomly selected 383 samples from each category within Operating Condition 1 to form its corresponding small dataset. Similarly, after applying sliding window sampling to Operating Condition 2, we also selected 383 samples from each category to create its small dataset.

For the CWRU and XJTU-SY datasets, we also split raw files first and then randomly select 383 samples from each category randomly select 383 samples from each category to form small datasets for cross-dataset, cross-load and cross-noise validation experiments. After applying the sliding window sampling technique, the time series data for each category under working Condition 1 is presented in Figure 7. Subsequently, the VMD-CVM algorithm is employed to perform noise reduction on the small dataset samples from both cases, resulting in 383 noise-reduced samples for each category. The outcomes of this noise reduction process for Condition 1 are illustrated in Figure 8. Following this, the CWT-RT method is utilized to analyze the time-frequency characteristics of the noise-reduced samples, yielding time-frequency spectro- grams for each fault state; specifically, the time-frequency spectrogram corresponding to Condition 1 is displayed in Figure 9. This provides high-quality training samples essential for subsequent model training. The time-frequency spectrogram samples from all three datasets are divided into a training set and a test set at a ratio of 7:3, respectively, for model training and validation under different operating conditions and noise levels. The time-frequency spectrogram samples from both conditions are divided into a training set and a test set at a ratio of 7:3, respectively.

Figure 7

Time Series Data: its types are, from left to right, ball fault, combined fault, inner race fault, normal, outer race fault.

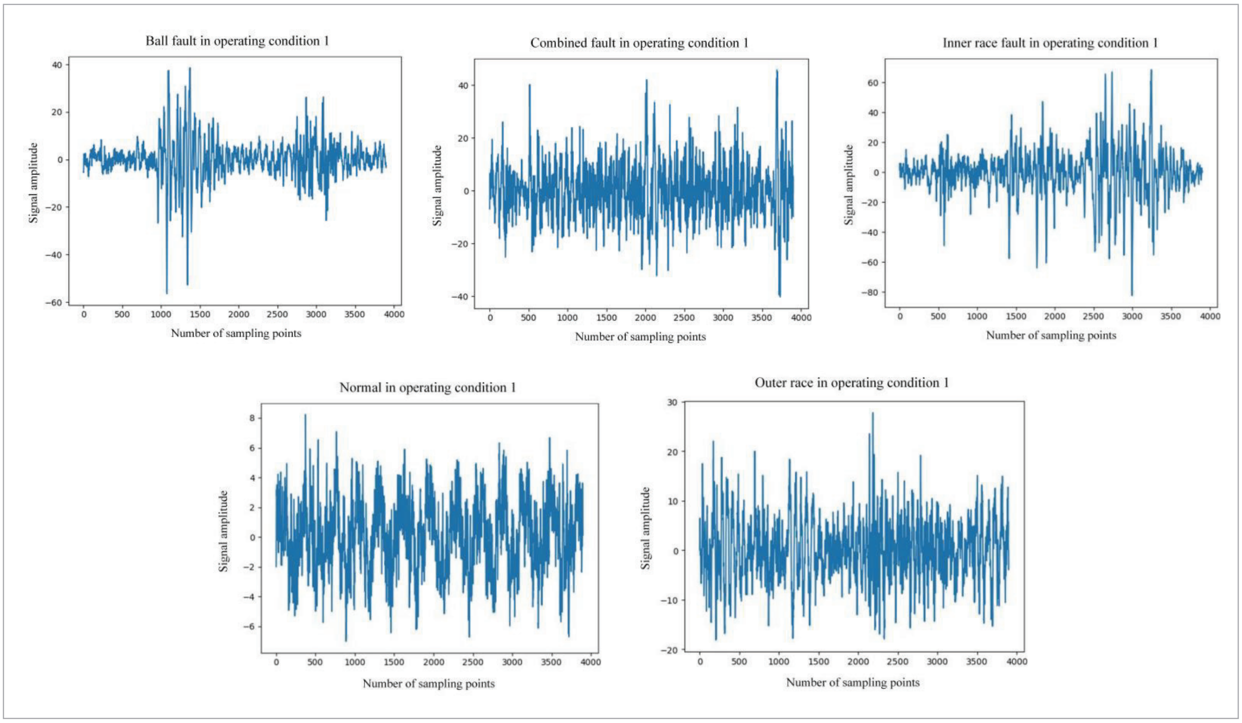


Figure 8

Data After Denoising: its types are, from left to right, ball fault, combined fault, inner race fault, normal, outer race fault.

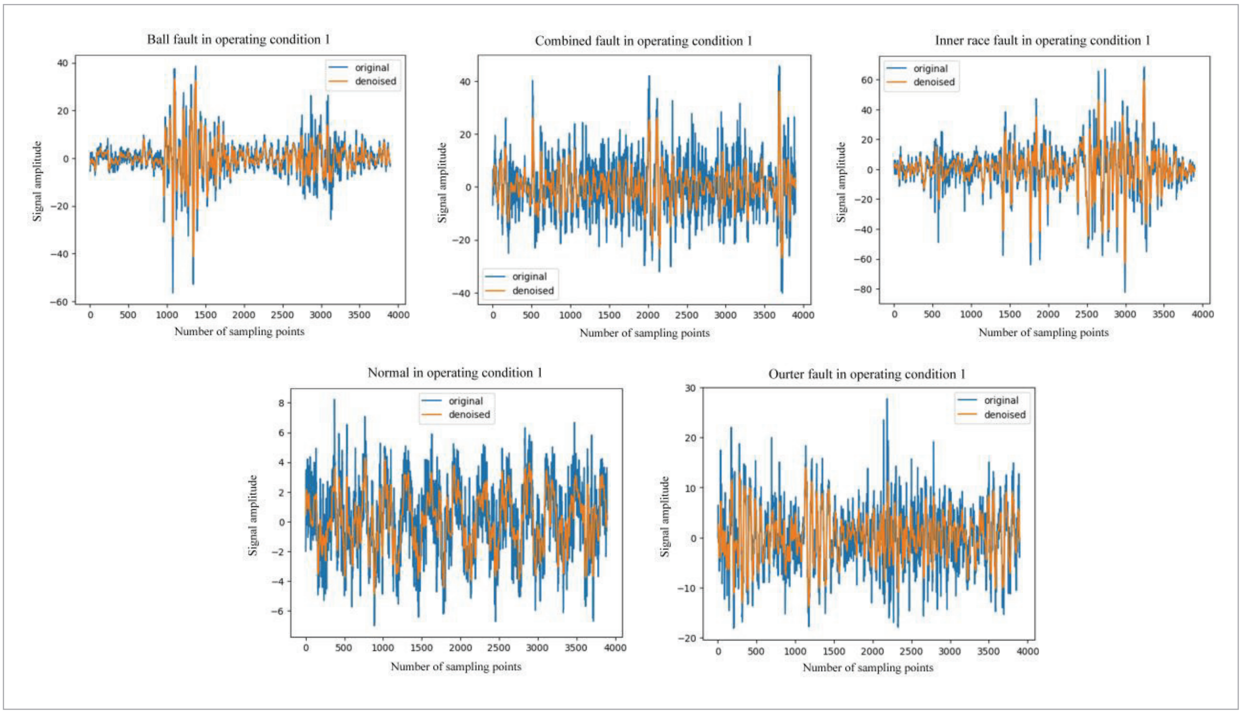
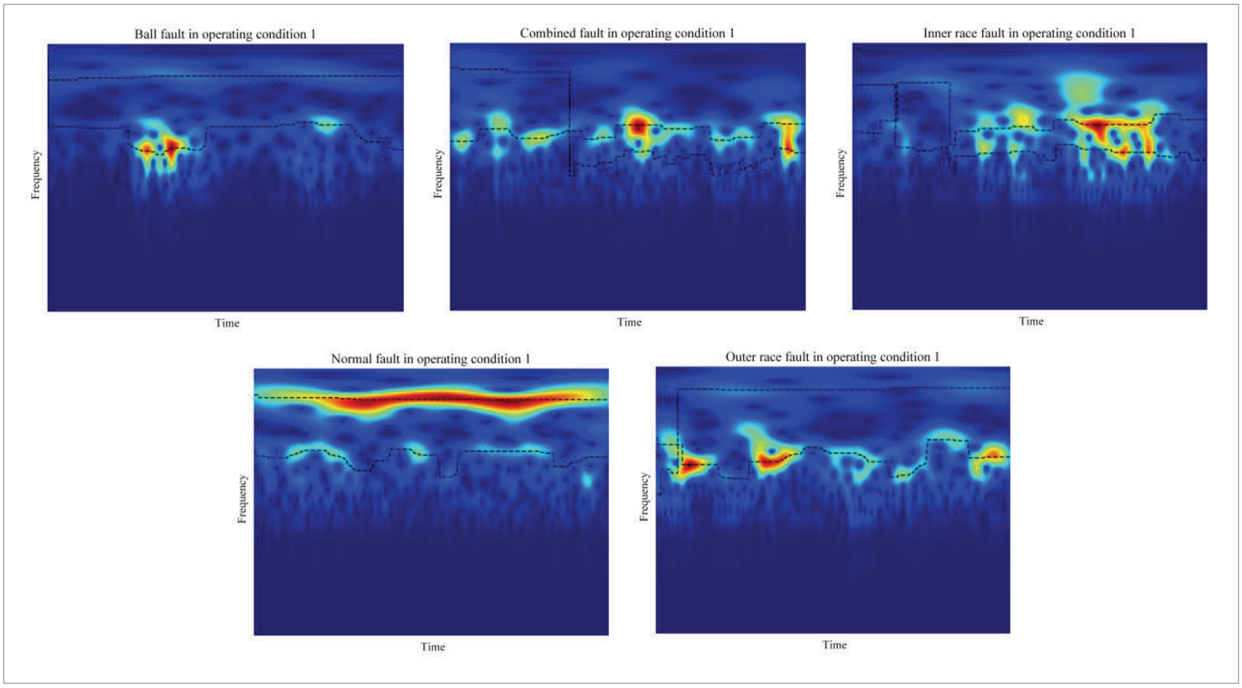


Figure 9

Time-Frequency Spectrum Diagram: its types are, from left to right, ball fault, combined fault, inner race fault, normal, outer race fault.



4.5. Parameter Setting

The present study constructs an innovative deep learning architecture, namely the Multi-scale Wavelet Scattering Network (MWSN) combined with a Multilayer Perceptron (MLP). The primary objective of this model is to enhance the performance of classification tasks on small datasets related to rolling bearings. The design of MWSN is cleverly designed; it employs a dual-channel strategy where each channel utilizes wavelet scattering networks at different scales, specifically J=50 and J=150, for feature extraction. This multi-scale analysis facilitates the capture of rich detail information from signals. Each channel's wavelet scattering network consists of three layers. After layer-by-layer calculation, the two channels extract features sized 1530×723 and 1530×1173 respectively. After feature fusion, these dimensions combine to form an input size of 1530 × 1896 for the MLP model. The MLP parameters (21 hidden neurons, tanh activation, 0.01 learning rate) are optimized using grid search and 5-fold cross-validation. The search encompasses the number of neurons, activation functions, and learning rates. The selected configuration balances fitting and generalization, alleviates over/under-fitting, and enables stable convergence for high-dimensional multi-scale features in small-sample scenarios. The MLP comprises an input layer, two hidden layers, and one output layer. Each hidden layer contains 21 neurons and employs the tanh activation function to stimulate the nonlinear representation capabilities of the deep network. For the 5-class mutually exclusive bearing fault classification task, the output layer uses the softmax activation function to output normalized class probabilities, and the cross-entropy loss function is adopted. The output layer utilizes a sigmoid function for probability prediction while employing mean squared error as its loss function. To ensure stability and convergence during training, we set a maximum iteration limit of 100 and established the target training error threshold at 1×10^{-6}. Additionally, a moderate learning rate of 0.01 was selected as the step size for gradient updates.

4.6. Evaluation Metrics

The present article employs accuracy, recall, precision, and F1 score as metrics to evaluate the performance of fault diagnosis models. Accuracy is defined as the proportion of correctly predicted samples to the total number of samples. The calculation formula is as follows:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}. \quad (16)$$

In this context, TP (True Positive) refers to the number of samples that the model correctly predicts as faults; FP (False Positive) denotes the number of samples that the model incorrectly predicts as faults; TN (True Negative) indicates the number of samples that the model accurately identifies as normal states; and FN (False Negative) represents the number of samples that are mistakenly predicted by the model as normal states. Recall is defined as the proportion of correctly predicted fault samples relative to all actual fault samples. The calculation formula is as follows:

$$Recall = \frac{TP}{TP+FN}. \quad (17)$$

The recall rate reflects the model's ability to correctly identify faults among all actual fault samples. Precision refers to the proportion of samples predicted as faults by the model that are indeed actual faults. The calculation formulas are as follows:

$$Precision = \frac{TP}{TP+FP}. \quad (18)$$

The precision rate reflects the model's ability to correctly identify faults among the samples predicted as faulty. The F1 score is the harmonic mean of precision and recall, serving to balance the relationship between these two metrics. The calculation formula is as follows:

$$F1\_Scone = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \quad (19)$$

The F1 score provides a comprehensive metric that takes into account both precision and recall, facilitating the identification of an optimal balance between the two. By utilizing these metrics, we can effectively and comprehensively assess the performance of fault diagnosis models.

5. Analysis of Results

5.1. Performance Evaluation

The fault diagnosis model based on CWT-RT and multi-scale wavelet scattering networks [25] was trained using small datasets from the two operating conditions shown in Table 1. Each small dataset contains five different fault states, with each category comprising 383 samples. The model training was completed once the target training error reached 1×10^{-6}, after which a test set was utilized to evaluate the model's performance. The testing accuracy for the small datasets under operating Condition 1 and operating Condition 2 of South Ural State University [11] reach 98.78% and 98.09%, respectively; the testing accuracy of CWRU dataset under 0/1/2/3hp loads are 99.21%, 98.85%, 98.32% and 97.89%, respectively; the testing accuracy of XJTU-SY dataset under 1200rpm/2000N load is 97.56%. as illustrated in Figure 10. The confusion matrices for the test sets of both operating conditions are presented in Figure 11. In the confusion matrix for the small dataset of operating Condition 1, all normal state samples (115) were classified correctly, indicating that this model demonstrates a high level of accuracy in distinguishing between normal and faulty states, as well as exhibiting significant sensitivity and reliability towards ball faults and Combination faults. Furthermore, while there were a few misclassifications concerning inner race faults and outer race faults, the overall performance still reflects considerable accuracy and reliability; thus, affirming the effectiveness of this model within the small dataset of operating Condition 1. In contrast, within the confusion matrix for operating Condition 2 of small dataset, rolling element faults achieved perfect classification results. However, misclassifications occurred among other four categories; notably, outer race faults exhibited the highest error rate at approximately 5.2%. Despite these errors, it is important to highlight that the overall accuracy attained by models trained on this small dataset under operating Condition 2 reached an impressive rate of 98.09%, thereby confirming its high level of accuracy and reliability as well. The confusion matrices of CWRU and XJTU-SY datasets are shown in the Supplementary Material, which further verifies the high recognition accuracy of the proposed model on different benchmark datasets.

Figure 10

Accuracy of the Test set: left side corresponds to Condition 1, right side corresponds to Condition 2.

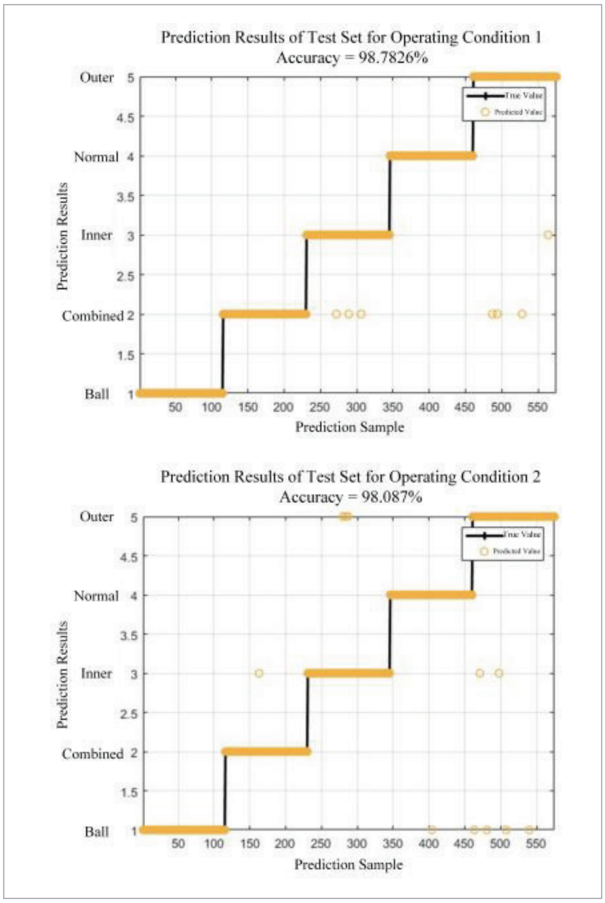
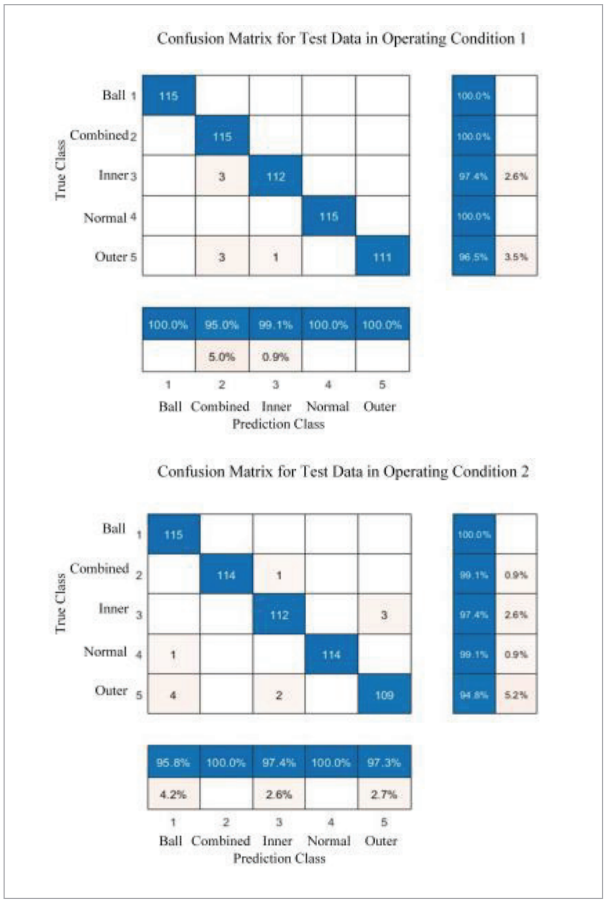


Figure 11

Confusion Matrix of the Test Set: left side corresponds to Condition 1, right side corresponds to Condition 2.



5.2. Comparative Experiment

In order to validate the superior fault diagnosis performance of the model proposed in this paper, five different deep learning models were compared. Specifically, small datasets from two operating conditions were trained and tested across these five distinct deep learning models, with the testing results presented in Table 2. The comparative experimental results indicate that the model proposed in this paper significantly outperforms five other deep learning models in fault diagnosis performance on two small datasets under different operating conditions. For the 20 Hz small dataset, the accuracy of our method reaches 98.78%, with precision, recall, and F1-score values of 98.83%, 98.78%, and 98.80% respectively. Compared to LSTM [20], DRSN [13], WDCNN [7], DCA-BiGRU [5], and WT-Swin Transformer [27], our method demonstrates superior overall performance, achieving higher accuracy, precision, recall, and F1-score; specifically, accuracy is improved by 15.36%, 13.6%, 9.02%, 2.45%, and 3.38% respectively. Compared with SVM, k-NN and Naive Bayes, the proposed method improves accuracy by 12.54%, 14.27% and 16.89%, with slightly higher but millisecond-level inference time, satisfying industrial real-time demands. For the small dataset at 18 Hz, our method also exhibits the highest performance metrics across all indicators, with accuracy improvements of 14.19%, 12.49%, 7.6%, 3.22%, and 3.37% compared to the aforementioned models. These findings suggest that our approach offers a distinct advantage over conventional deep learning methods for bearing fault diagnosis within small datasets.

Table 2
Comparative Experimental Results of the Models.

Algorithm	20 Hz Small Data				18 Hz Small Data			
	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
LSTM	83.42	84.25	82.64	83.44	83.90	84.21	83.64	83.92
DBSN	85.18	86.91	83.56	85.20	85.60	86.72	84.53	85.61
WDCNN	89.76	90.20	89.36	89.78	90.49	89.70	91.32	90.50
DCA-BiGRU	96.33	97.48	95.26	96.36	94.87	96.52	93.33	94.90
WT-SwinTransformer	95.40	96.25	94.58	95.41	94.72	95.50	94.01	94.75
Our Approach	98.78	98.83	98.78	98.80	98.09	98.11	98.09	98.10

5.3. Ablation Experiment

To validate the effectiveness of each module in the fault diagnosis model proposed in this paper, we conducted an ablation study on the suggested model. The ablation experiment quantifies the accuracy contribution rate of each module (VMD-CVM, CWT-RT, MSWSN) on the small datasets of South Ural State University (20Hz/18Hz), CWRU (3hp) and XJTU-SY, and deeply analyzes the synergy mechanism of each module. The ablation experiments include the following aspects:

1 The original bearing data from South Ural State University was used as input for the Wavelet Scattering Network (WSN). features were extracted from the raw bearing data using WSN. These features exhibit excellent invariance and stability, after which they are fed into a Multilayer Perceptron (MLP) for training. Finally, we evaluate the model's performance on a test set and compute various evaluation metrics.

2 In the feature extraction phase, an improved multiscale wavelet scattering network was employed to extract features. By incorporating multiscale analysis, this approach enhances both feature richness and robustness, effectively improving diagnostic performance on small datasets.

3 Building upon both the multiscale wavelet scattering network and multilayer perceptron, we further introduced CWT-RT time-frequency analysis methods to enhance resolution and accuracy in time-frequency analysis, thereby providing high-quality training samples for fault diagnosis tasks.

4 We directly applied our proposed method for fault diagnosis and calculated evaluation metrics based on the test set results. Table 3 presents a comparative analysis of different module combinations within our ablation study concerning their outcomes on the test set while detailing various evaluation metrics to comprehensively assess each module's impact on overall diagnostic performance.

Table 3
Results of the Model Ablation Experiment.

Algorithm	20 Hz Small Data				18 Hz Small Data			
	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
WSN-MLP	85.22	85.15	85.22	85.18	90.26	90.29	90.26	90.27
MSWSN-MLP	88.35	88.17	88.35	88.26	93.57	93.56	93.57	93.56
CWT-RT-MSWSN-MLP	97.74	97.79	97.74	97.76	97.22	97.23	97.22	97.22
Our Approach	98.78	98.83	98.78	98.80	98.09	98.11	98.09	98.10

The results of the ablation experiments indicate that replacing the wavelet scattering network with a multi-scale wavelet scattering network leads to an improvement in accuracy. The multi-scale wavelet scattering network is capable of capturing features of the data at different scales. This multi-scale analysis enables the model to better comprehend both local and global information within signals, thereby enhancing its ability to recognize complex patterns. Furthermore, after incorporating the CWT-RT time-frequency analysis method, there is a significant increase in model accuracy. The CWT-RT time-frequency analysis method combines continuous wavelet transform with ridge tracking techniques, providing high-resolution time-frequency representations and tracking important frequencies. It demonstrates exceptional performance in real-time analysis of complex and non-stationary signals. Finally, by integrating the VMD-CVM algorithm, we observe a further enhancement in accuracy, underscoring the effectiveness of using the VMD-CVM algorithm for noise reduction.

5.4. Validation of Generalization Performance

In order to further validate the effectiveness of the model and assess whether it is overfitting, we utilized the remaining datasets from Condition 1, excluding the small dataset at 20 Hz, as a test set. The trained model on the 20 Hz small dataset was then evaluated using this test set. The results are presented in Table 4. The experimental results indicate that the model trained on 20 Hz small dataset continues to demonstrate a high level of accuracy when evaluated against the remaining 20 Hz dataset in Condition 1. This finding serves as compelling evidence of the model's effectiveness and suggests that it can be successfully applied to other data from the same condition while maintaining superior performance levels.

5.5. Computation Cost Analysis

The full pipeline of the proposed method includes VMD-CVM denoising, CWT-RT time-frequency transformation, MSWSN feature extraction and MLP classification. The computation cost is quantified on the experimental platform (AMD Ryzen 9 7950X, NVIDIA RTX 3060, 128GB DDR5) and IoT edge device (NVIDIA Jetson Nano):

1. **Preprocessing overhead:** VMD-CVM denoising takes 2.3ms per sample, CWT-RT time-frequency transformation takes 3.1ms per sample.
2. **Feature extraction time:** MSWSN multi-scale feature extraction takes 4.7ms per sample.
3. **Inference latency:** MLP classification takes 0.8ms per sample.
4. **Full-pipeline latency:** 10.9ms per sample on PC, 18.3ms per sample on Jetson Nano, both meeting the millisecond-level real-time requirements of IoT industrial monitoring.
5. **Memory footprint:** The peak memory of the model is 128MB on edge devices, which is compatible with low-memory IoT sensors.
6. **Energy consumption:** The average power consumption is 2.3W during edge deployment, suitable for battery-powered wireless IoT nodes.

6. Conclusion

The present study addresses the characteristics of rolling bearing vibration signals, which are non-stationary, nonlinear, and susceptible to noise interference. Additionally, it tackles the issue of low recognition rates in deep learning when applied to small datasets. We propose a fault diagnosis model for rolling bearings based on CWT-RT and Multi-Scale Wavelet Scattering Networks. Strict raw-file level train-test splitting is adopted to eliminate data leakage, and softmax+cross-entropy is used for standard multi-class classification. Experimental results demonstrate that this diagnostic model achieves outstanding performance on two small datasets under different operating conditions from South Ural State University. The model's accuracy reached 98.78% and 98.09%, respectively. In comparative

Table 4

Test Results of the Remaining Dataset at 20Hz.

Test Data	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
20Hz Residual Dataset	20Hz Small Data Model	98.73	98.74	98.73	98.73

experiments, our proposed fault diagnosis model outperformed common deep learning methods in terms of accuracy, precision, recall, and F1-score. In ablation studies, we systematically decompose each functional module to observe how these changes impact overall performance, thereby validating the effectiveness and significance of each module. The full-pipeline latency, memory and energy consumption meet the IoT edge deployment requirements. Despite the significant progress made by our method in diagnosing bearing faults within small datasets, several challenges remain. Future work will focus on enhancing the model's generalization capability-specifically improving its ability to recognize faults across varying operational conditions-to further increase its practical applicability.

Acknowledgement

This work was supported by the Xinjiang Uygur Autonomous Region Strategic Talent Program Distinguished Engineer (XJRC-2025-GX-PY-GCS-045) and the Xinjiang Urumqi Hongshan Science and Technology Innovation Young Talents Program (B241013006).

References

1. Abramovich, F., Benjamini, Y. Adaptive Thresholding of Wavelet Coefficients. Computational Statistics and Data Analysis, 1996, 22(4), 351-361. https://doi.org/10.1016/0167-9473(96)00003-5
2. Benazzouz, D., Merainani, B., Ould-Bouamama, B., Rahmoune, C. Rolling Bearing Fault Diagnosis Based Empirical Wavelet Transform Using Vibration Signal. In 2016 8th International Conference on Modelling, Identification and Control (ICMIC), 2016: 526-531. https://doi.org/10.1109/ICMIC.2016.7804169
3. Bi, Y., Peng, B., Wan, S., Xue, B., Zhang, M. A Survey on Fault Diagnosis of Rolling Bearings. Algorithms, 2022, 15(10). https://doi.org/10.3390/a15100347
4. Cen, J., Chen, Z., Xiong, J. Rolling Bearing Fault Diagnosis Using Time-Frequency Analysis and Deep Transfer Convolutional Neural Network. IEEE Access, 2020, 8, 150248-150261. https://doi.org/10.1109/ACCESS.2020.3013885
5. Chen, B., He, C., Lu, Y., Zhang, L., Zhang, X., Zhu, L. Fault Diagnosis for Small Samples Based on Attention Mechanism. Measurement, 2022, 187, 110242. https://doi.org/10.1016/j.measurement.2021.110242
6. Chen, X., Gao, D., Zhang, B. Bearing Fault Diagnosis Base on Multi-Scale CNN and LSTM Model. Journal of Intelligent Manufacturing, 2021, 32(4), 971-987. https://doi.org/10.1007/s10845-020-01600-2
7. Chen, Y., Li, C., Peng, G., Zhang, W., Zhang, Z. A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals. Sensors, 2017, 17(2). https://doi.org/10.3390/s17020425
8. Cramér, H. On the Composition of Elementary Errors: First Paper: Mathematical Deductions. Scandinavian Actuarial Journal, 1928, 1928(1), 13-74. https://doi.org/10.1080/03461238.1928.10416862
9. Donoho, D. De-Noising by Soft-Thresholding. IEEE Transactions on Information Theory, 1995, 41(3), 613-627. https://doi.org/10.1109/18.382009
10. Duan, L., Li, T., Yuan, Z., Zhang, L. Intelligent Fault Diagnosis of Rolling Element Bearings Based on HHT and CNN. In 2018 Prognostics and System Health Management Conference (PHM-Chongqing), 2018: 292-296. https://doi.org/10.1109/PHM-Chongqing.2018.00056
11. Eremeeva, V., Ibryaeva, O., Sakovskaya, V., Sinitsin, V. Intelligent Bearing Fault Diagnosis Method Combining Mixed Input and Hybrid CNN-MLP Model. Mechanical Systems and Signal Processing, 2022, 180, 109454. https://doi.org/10.1016/j.ymssp.2022.109454
12. Fan, Y., Sun, H. Fault Diagnosis of Rolling Bearings Based on CNN and LSTM Networks Under Mixed Load and Noise. Multimedia Tools and Applications, 2023, 82(28), 43543-43567. https://doi.org/10.1007/s11042-023-15325-w
13. Fu, X., Pecht, M., Tang, B., Zhao, M., Zhong, S. Deep Residual Shrinkage Networks for Fault Diagnosis. IEEE Transactions on Industrial Informatics, 2020, 16(7), 4681-4690. https://doi.org/10.1109/TII.2019.2943898
14. Gao, S., Guo, Q., Li, Y., Song, Y., Wang, D. Application of Multiscale Learning Neural Network Based on CNN in Bearing Fault Diagnosis. Journal of Signal Processing Systems, 2019, 91(10), 1205-1217. https://doi.org/10.1007/s11265-019-01461-w

15. Huang, G., Luo, J. Check Valve Fault Diagnosis Based on Generalized S-Transform and Deep Belief Network. J. Electron. Meas. Instrum, 2019, 33, 192-198. http://jemi.etmchina.com/jemi/ch/reader/view_abstract.aspx?file_no=B1900245&flag=1

16. Huang, N. E., Liu, H. H., Long, S. R., Shen, Z., Shih, H. H., Tung, C. C., Wu, M. C., Yen, N. C., Zheng, Q. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 1998, 454(1971), 903-995. https://doi.org/10.1098/rspa.1998.0193

17. Huang, W., Li, Y., Zeng, Y., Wei, Y., Xu, M. A New Rolling Bearing Fault Diagnosis Method Based on Multiscale Permutation Entropy and Improved Support Vector Machine Based Binary Tree. Measurement, 2016, 77, 80-94. https://doi.org/10.1016/j.measurement.2015.11.043

18. Hoang, D. T., Kang, H. J., Nguyen, V. C., Tran, X. T., Van, M. A Bearing Fault Diagnosis Method Using Multi-Branch Deep Neural Network. Machines, 2021, 9(12). https://doi.org/10.3390/machines9120345

19. Jia, X., Zhao, M. A Novel Strategy for Signal Denoising Using Reweighted SVD and Its Applications to Weak Fault Feature Enhancement of Rotating Machinery. Mechanical Systems and Signal Processing, 2017, 94, 129-147. https://doi.org/10.1016/j.ymssp.2017.02.036

20. Jin, B., Sun, S., Zhao, H. Sequential Fault Diagnosis Based on LSTM Neural Network. IEEE Access, 2018, 6, 12929-12939. https://doi.org/10.1109/ACCESS.2018.2794765

21. Jung, Y. Y., Lavrik, I., Ruggeri, F., Vidakovic, B. Bayesian False Discovery Rate Wavelet Shrinkage: Theory and Applications. Communications in Statistics Simulation and Computation, 2008, 37(6), 1086-1100. https://doi.org/10.1080/03610910802049649

22. Li, H., Wang, H., Zhu, H. Multi-Mode Data Generation and Fault Diagnosis of Bearings Based on STFT-SACGAN. Electronics, 2023, 12(8). https://doi.org/10.3390/electronics12081910

23. Liu, Q., Zeng, Y., Xu, Y. Synchrosqueezing Extracting Transform and Its Application in Bearing Fault Diagnosis Under Non-Stationary Conditions. Measurement, 2021, 173, 108569. https://doi.org/10.1016/j.measurement.2020.108569

24. Mandic, D. P., Rehman, N. Multivariate Empirical Mode Decomposition. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2010, 466(2117), 1291-1302. https://doi.org/10.1098/rspa.2009.0502

25. Mallat, S. Group Invariant Scattering. Communications on Pure and Applied Mathematics, 2012, 65(10), 1331-1398. https://doi.org/10.1002/cpa.21424

26. Mushe, D. Bearing Fault Diagnosis Based on VMD Noise Reduction and CNN. Noise and Vibration Control, 2021, 41(5), 155. https://doi.org/10.6598/j01.itc.54.3.421

27. Tian, Z., Wang, J., Yan, W., Zhang, X. A Bearing Fault Diagnosis Method Based on 2D Images and Transformer. In 2023 13th International Conference on Information Technology in Medicine and Education (ITME), 2023: 277-282. https://doi.org/10.1109/ITME60234.2023.00064