

Fault Diagnosis for Key Components of Metro Vehicles based on Wavelet Threshold Denoising and EEMD

Xichun Luo, Haoran Hu*

Yunnan Jingjian Rail Transit Investment and Construction Co., Ltd., Kunming 650000, Yunnan, China

**Author to whom correspondence should be addressed.*

Copyright: © 2025 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: With the increasing adoption of intelligent operation and maintenance technologies in urban rail transit, most maintenance systems have been equipped with fault diagnosis modules targeting key components of metro vehicles. However, the integration between engineering-level diagnostic algorithms and advanced academic research remains limited. Two major challenges hinder vibration-based fault diagnosis under real-world operating conditions: the complex noise and interference caused by wheel–rail coupling and the typically weak expression of fault features. Considering the widespread application of wavelet transform in noise reduction and the maturity of ensemble empirical mode decomposition (EEMD) in handling nonlinear and non-stationary signals without parameter tuning, this study proposes a diagnostic method that combines wavelet threshold denoising with EEMD. The method was applied to bearing vibration signals collected from an operational subway line. The diagnostic results were consistent with actual disassembly findings, demonstrating the effectiveness and practical value of the proposed approach.

Keywords: Metro vehicles; Fault diagnosis; Wavelet threshold de-noising; Ensemble empirical mode decomposition

Online publication: May 15, 2025

1. Introduction

The primary concept of intelligent operation and maintenance in urban rail transit involves introducing mature technologies, such as the Internet of Things, big data, and fault diagnosis, originally developed from other industries, into metro equipment management to achieve intelligent upgrades of existing maintenance practices. With the widespread deployment of intelligent maintenance technologies and system-level products, most operational platforms in the rail transit sector are now equipped with fault diagnosis modules targeting critical components of metro vehicles. Gearboxes and bearings have emerged as primary targets for condition monitoring. Wheelset bearings represent the most crucial elements of the vehicle's running gear. A commonly employed monitoring method involves acquiring vibration signals from the surface of the axlebox using

composite vibration-impact sensors, followed by diagnostic analysis based on signal processing techniques.

Current online monitoring systems deployed in China primarily rely on techniques such as band-pass filtering, resonance demodulation, wavelet packet decomposition, acceleration envelope analysis, and time-frequency analysis. A significant gap remains between these engineering applications and the methods commonly employed in recent fault diagnosis research, such as advanced signal decomposition techniques ^[1,2], as well as machine learning and deep learning approaches ^[3-5]. While machine learning and deep learning methods require large volumes of labeled data, the main challenges stem from the nature of vibration-impact signal-based diagnostics in rail transit. These challenges include the presence of strong noise and interference resulting from wheel-rail coupling and the typically weak expression of fault signals. The proposed methodology employs noise reduction techniques to suppress interference and utilizes signal decomposition to amplify weak fault features to address these constraints. Considering the practical constraints of engineering applications, the adopted methods must be both simple and computationally efficient. Accordingly, the computational framework was constructed based on the classical wavelet thresholding denoising technique ^[6], known for its low computational overhead, and the ensemble empirical mode decomposition (EEMD) method ^[7], which operates without the need for parameter tuning.

2. Object of study

2.1. Key component

The object of study in this work is the wheelset bearing of metro vehicles currently in operation. The bogie adopts a two-axle configuration with a secondary suspension system. The bearing is mounted at the end of the wheelset, with the axlebox connected to the bogie frame via a torque arm and a locating rubber joint. The primary suspension consists of a helical spring and a vertical damper, while the bogie frame is supported by a secondary suspension comprising steel springs and hydraulic dampers. As one of the most structurally complex and critical load-bearing components, the wheelset bearing presents high research value. The bearing structure and its installation are illustrated in **Figure 1**. A sealed double-row tapered roller bearing is employed, and its key specifications are provided in **Table 1**.

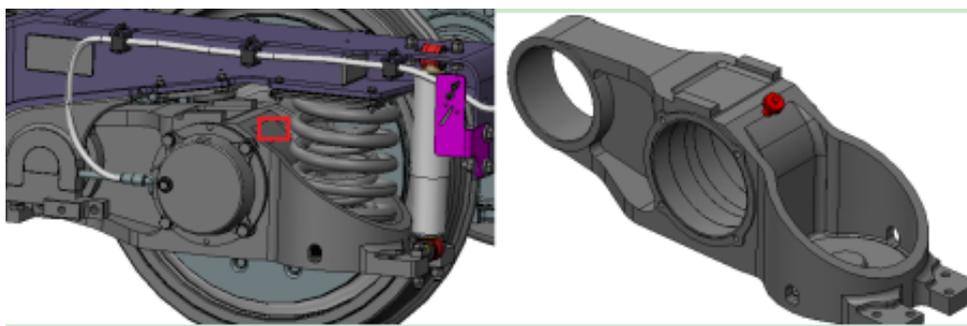


Figure 1. Schematic diagram of bearing and sensor installation

Table 1. Bearing specifications

Bearing model	Rolling element diameter (mm)	Pitch diameter (mm)	Number of rolling elements	Contact angle (°)
SKF 8670-01	24.27	179.268	20	10

2.2. Data acquisition

A composite digital sensor is mounted on the axlebox of each wheelset bearing to capture vibration and impact signals during train operation. One sensor is installed per axlebox, positioned above and slightly offset from the bearing, as illustrated in the schematic diagram. The sensor is secured using an M12 mounting interface with an effective tread depth of at least 16 mm. The angle between the sensor's central axis and the vibration direction (defined by the line connecting the sensor and the bearing center) must satisfy the condition $A \leq \pm 10^\circ$.

Following data acquisition, the onboard sensor transmits the collected signals to the ground system via the train-ground wireless transmission subsystem, which acts as a bridge between the onboard sensing unit and the ground-based integrated application platform. Data transmission is achieved through components including the onboard communication gateway, ground communication gateway, broadband combined antenna with supporting cables, and a wireless transmission network. This setup establishes a wireless communication channel for transmitting monitoring and fault data from the train to the ground subsystem. The backend system then downloads or analyzes the data to perform fault diagnosis, as illustrated in **Figure 2**.

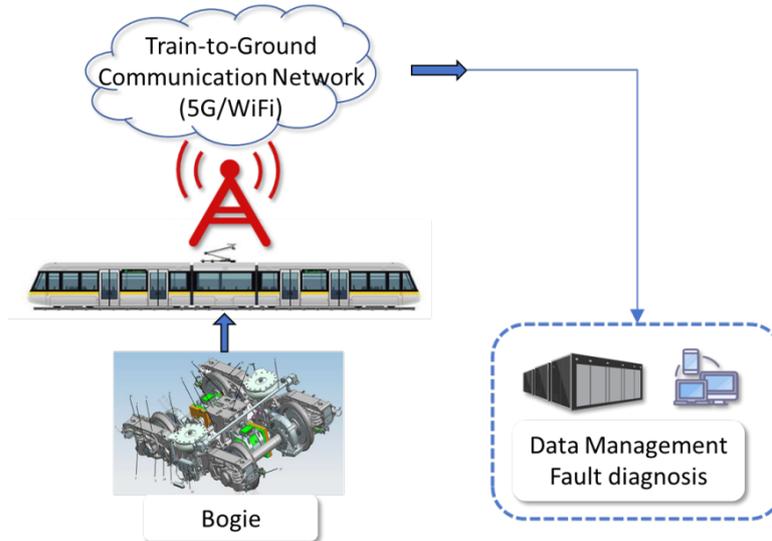


Figure 2. Schematic diagram of bearing and sensor installation

3. Theoretical framework

3.1. Wavelet denoising

The wavelet transform performs signal filtering by applying a suitable set of filters to decompose the signal into multiple frequency components. At each decomposition level, the signal is split into high-frequency detail coefficients and low-frequency approximation coefficients. The approximation component is then passed to the next level for further wavelet decomposition.

For any function $x(t)$ in $L^2(R)$, the continuous wavelet transform concerning a wavelet basis is given by:

$$CWT(a, b) = \frac{1}{\sqrt{a}} \int x(t) \varphi\left(\frac{t-b}{a}\right) dt \quad (1)$$

In the wavelet transform, the coefficients a and b represent the scale and translation parameters, respectively. By multiplying each coefficient by the appropriately scaled and shifted wavelet, the original signal can be reconstructed. The scale parameter controls the frequency resolution of the transform. A smaller value

of a corresponds to a compressed time window and a broader frequency bandwidth, resulting in a higher center frequency. Conversely, a larger stretch of the wavelet captures slower variations in the signal and reflects lower-frequency components. The translation parameter b adjusts the time localization of the wavelet, thereby shifting the analysis window along the time axis to reveal the signal's temporal features at different resolutions. Wavelet transform maintains strong localization capabilities in both time and frequency domains, making it highly effective for suppressing noise while preserving critical signal features. As a result, it has been widely adopted in signal-denoising applications. The wavelet-based denoising process begins with the assumption of a one-dimensional noisy signal, expressed as:

$$f(t) = s(t) + n(t), t = 1, 2, \dots \quad (2)$$

Where $f(t)$ denotes the true signal, $s(t)$ represents the observed (original) signal, and $n(t)$ corresponds to the noise component.

The principle of wavelet denoising lies in removing Gaussian white noise by filtering out non-informative components, thereby recovering the effective part of the signal. In this study, the wavelet threshold denoising method is adopted^[8]. The overall procedure is illustrated in **Figure 3**.

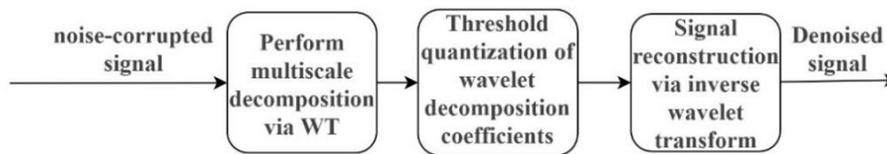


Figure 3. Flow chart of wavelet transforms threshold denoising

As shown in **Figure 3**, the denoising procedure includes three main steps: (1) multi-scale decomposition of the noisy signal using wavelet transform; (2) thresholding of the wavelet decomposition coefficients to suppress noise; (3) signal reconstruction based on the thresholded coefficients.

The most critical step in the wavelet denoising process is selecting the optimal thresholding function and the corresponding threshold value. Currently, no unified theoretical standard exists for determining the thresholding function, as it reflects different strategies for processing wavelet coefficients. In practical implementations, a heuristic threshold is typically applied to the wavelet coefficients, and soft thresholding is used to process the signal.

3.2. Ensemble empirical mode decomposition and kurtosis criterion

EEMD is an effective method for analyzing nonlinear, non-stationary, or non-white noise signals. Built upon empirical mode decomposition (EMD), EEMD introduces artificial white noise and applies multiple iterations of decomposition. The signal is ultimately decomposed into a series of frequency- or amplitude-modulated components. These components, which are non-sinusoidal and mutually distinct, are referred to as intrinsic mode functions (IMFs)^[9]. The final result is obtained by averaging the corresponding IMFs across all ensemble trials. For a given signal $y(t)$, the EEMD procedure is illustrated in **Figure 4**.

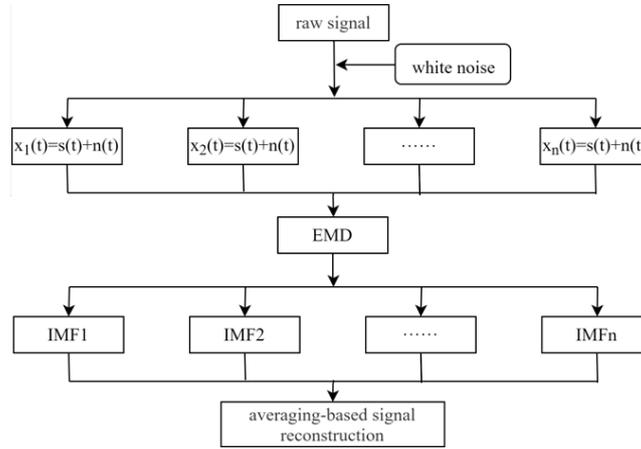


Figure 4. Flow chart of EEMD algorithm

Through EEMD, the original signal $y(t)$ is decomposed into a set of ensemble-averaged IMFs and a residual term r .

Based on the kurtosis criterion, the kurtosis value of each IMF obtained from EEMD is calculated. IMF components with kurtosis values greater than 3, which typically indicate the presence of fault-induced impulsive features, are selected for Hilbert envelope spectrum analysis. Fault characteristic frequencies are then extracted from the resulting Hilbert envelope spectrum.

The kurtosis value used in this analysis is a dimensionless parameter that reflects the distribution characteristics of the vibration signal. Its mathematical definition is given as follows ^[10]:

$$K = \frac{E(x-u)^4}{\sigma^4} \quad (3)$$

Where x denotes the vibration signal under analysis, u is the mean value of x ; σ represents the standard deviation of x ; E denotes the expectation (mean) of the signal.

The kurtosis value describes the sharpness of a signal and directly reflects the intensity of its impulsive components. It remains unaffected by bearing speed, size, or load, making it a reliable indicator for fault diagnosis. Impulsive components serve as critical signatures of bearing faults. In a healthy bearing, where the vibration amplitude follows a normal distribution, the kurtosis value is approximately 3. The onset of a fault leads to a significant increase in kurtosis. A higher kurtosis value suggests a greater proportion of impulsive content within the signal, indicating more prominent fault-induced impacts.

4. Methods

The proposed algorithm consists of the following steps: (1) Wavelet transform is applied to denoise the vibration signal, taking advantage of its time-frequency localization capability; (2) Ensemble empirical mode decomposition (EEMD) is then performed on the denoised signal; (3) Based on the kurtosis criterion, IMF components with strong relevance to bearing fault characteristics are selected and used to reconstruct the signal; (4) Hilbert transform is applied to obtain the envelope spectrum, enabling the identification of fault characteristic frequencies and thus completing the bearing fault diagnosis. The detailed algorithmic procedure is illustrated in **Figure 5**.

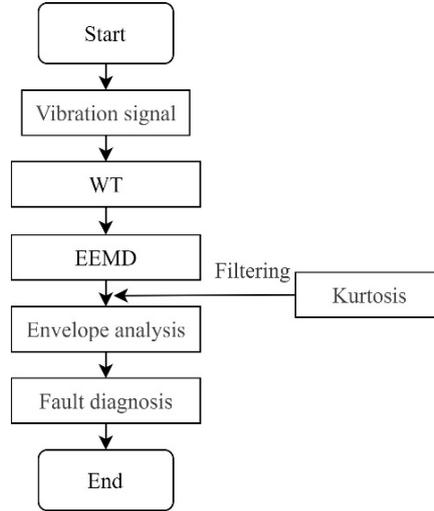


Figure 5. Algorithm flowchart

5. Experimental verification

5.1. Basic data

During data acquisition, the metro vehicle operated at a constant speed of 70 km/h, measured at the axlebox. Given a wheel diameter of 840 mm, the rotational frequency of the rolling bearing was calculated to be 8.125 Hz. The vibration signal was sampled at a frequency of 1040 Hz.

The characteristic fault frequency of the inner race of the rolling bearing can be calculated using the following formula:

$$f_r = n/60 \quad (4)$$

$$f_i = \frac{z}{2} \left(1 + \frac{d}{D} \cos \alpha\right) f_r \quad (5)$$

The characteristic fault frequency of the outer race is calculated as:

$$f_o = \frac{z}{2} \left(1 - \frac{d}{D} \cos \alpha\right) f_r \quad (6)$$

The characteristic fault frequency of the rolling element is calculated as:

$$f_b = \frac{D}{2d} \left[1 - \left(\frac{d}{D}\right)^2 \cos^2 \alpha\right] f_r \quad (7)$$

The characteristic fault frequency of the cage is calculated as:

$$f_c = \frac{1}{2} \left(1 - \frac{d}{D} \cos \alpha\right) f_r \quad (8)$$

Where z is the number of rolling elements; α represents the contact angle; f_r denotes the shaft rotational frequency (Hz); f_i is the inner race fault frequency (Hz); f_o denotes the outer race fault frequency (Hz); d is the diameter of rolling element (mm); D is the pitch diameter (mm).

Based on the above formulas, the characteristic fault frequencies corresponding to the bearing studied in

this work are as follows: 92.08 Hz for an inner race fault, 70.42 Hz for an outer race fault, 29.47 Hz for a rolling element fault, and 3.52 Hz for a cage fault.

5.2. Fault diagnosis

Due to the extremely low failure rate of bearings in metro operation, only a limited number of early-stage outer race faults were observed in the dataset analyzed in this study. As a result, the fault-related signal components are relatively weak.

(1) Wavelet Denoising: The spectrum and envelope spectrum after wavelet-based denoising are shown in **Figure 6**. As observed in the figure, distinct impact components are present, and the signal-to-noise ratio is relatively high. A prominent frequency component at 68.81 Hz appears in both the denoised and envelope spectra, close to the calculated outer race fault frequency of 70.42 Hz. In addition, the shaft rotational frequency of 8.125 Hz is also clearly identifiable in the spectrum. However, multiple interfering components remain, and the absence of harmonic features prevents clear identification of the fault type. Therefore, further signal decomposition is required to enhance the fault-related information within the original signal.

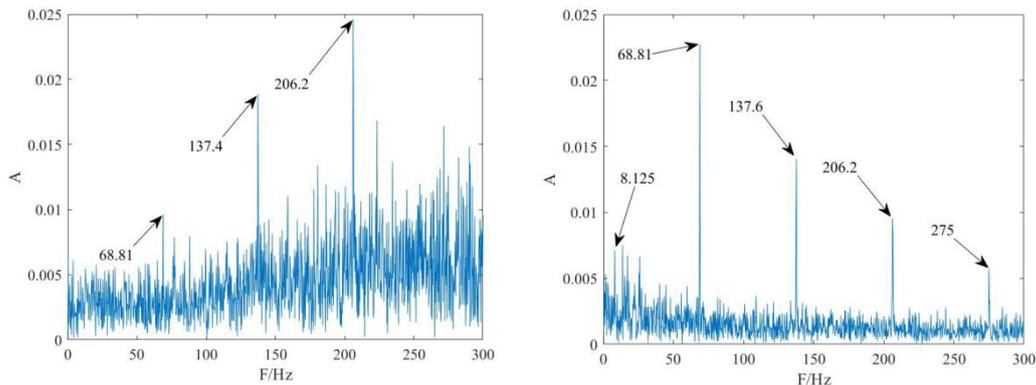


Figure 6. Denoised signal. (a) Spectrum; (b) Envelope spectrum

(2) EEMD processing: The wavelet-denoised bearing vibration signal is further decomposed using EEMD, resulting in a set of intrinsic mode functions (IMFs), denoted as IMF1, IMF2, ..., IMF9. The time-domain waveforms of each IMF component are shown in **Figure 7**, and the corresponding kurtosis values are listed in **Table 2**.

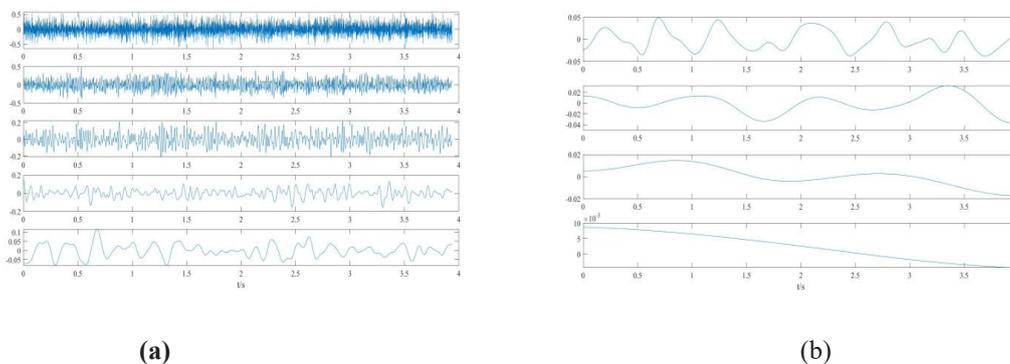


Figure 7. Time-domain waveforms of IMF components obtained from EEMD decomposition

Table 2. Kurtosis values of IMF components

IMF	IMF ₁	IMF ₂	IMF ₃	IMF ₄	IMF ₅
Kurtosis	3.01199	3.02087	2.94800	3.64594	3.17786
IMF	IMF ₆	IMF ₇	IMF ₈	IMF ₉	
Kurtosis	2.08514	2.71039	2.51784	1.60057	

(3) Reconstruction and envelope analysis: IMF components with kurtosis values greater than 3 are selected for signal reconstruction. After applying the Hilbert transform to the reconstructed signal, the resulting envelope spectrum is shown in **Figure 8**, respectively.

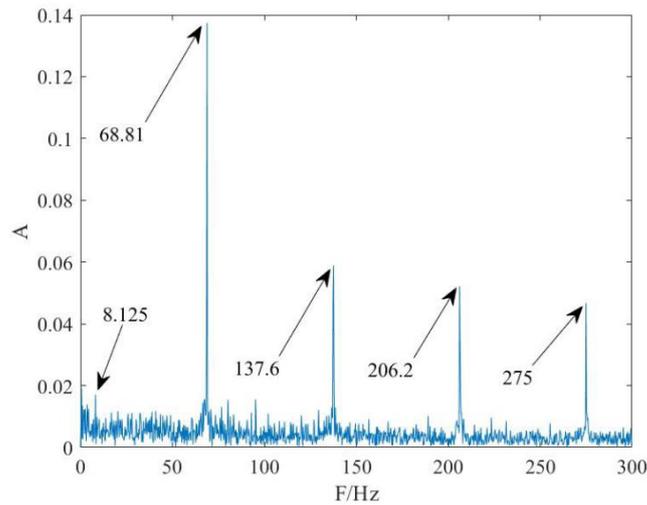


Figure 8. Envelope spectrum of the signal after wavelet denoising and EEMD

As shown in the above figures, the envelope spectrum obtained through signal decomposition and reconstruction using the proposed method shows significantly enhanced impact components. The frequency at 8.125 Hz corresponds to the rotational frequency of the bearing. Prominent peaks are also present at 68.81 Hz, as well as their second and third harmonics at 137.6 Hz and 206.2 Hz, respectively. These frequencies closely match the calculated outer race fault frequency of 70.42 Hz. Accordingly, the fault can be attributed to the bearing’s outer race, as **Figure 9**. This diagnosis is consistent with the observed damage during bearing disassembly and aligns with fault signature characteristics reported by a leading bearing manufacturer.

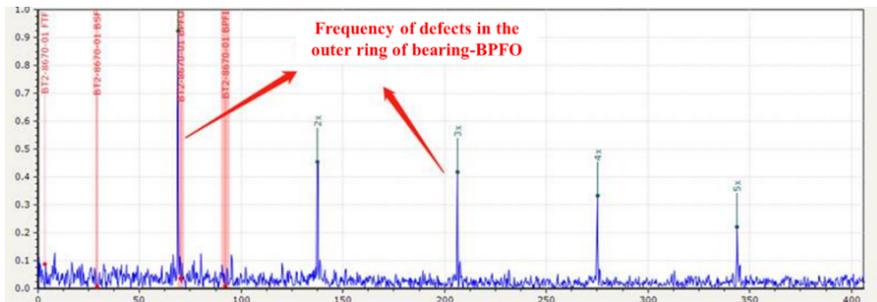


Figure 9. Fault analysis results of a renowned bearing manufacturer

6. Conclusion

To address the challenges of high noise levels and weak fault signals in vibration data collected from metro vehicles during actual operation, this study proposes a fault diagnosis method combining wavelet threshold denoising and EEMD, with a focus on engineering applicability. The method was applied to bearing fault diagnosis for metro vehicles. Experimental validation using real-world operational data confirmed that the extracted fault features are consistent with actual bearing failures.

Wavelet threshold denoising effectively filters out non-informative noise components and preserves the essential signal content, thereby enhancing the signal-to-noise ratio of vibration data. EEMD further decomposes the denoised signal into a set of IMFs that isolate impact-related components across different frequency bands. By applying the kurtosis criterion to select relevant IMFs and reconstructing the signal, Hilbert envelope analysis enables the extraction of fault characteristic frequencies. The proposed method demonstrates strong robustness in handling non-stationary signals.

Future research in fault diagnosis of key metro vehicle components should aim to enhance the adaptability of signal decomposition techniques under low signal-to-noise ratio conditions. Moreover, expanding the dataset and incorporating deep learning-based models may offer further improvements in diagnostic accuracy and reliability.

Acknowledgments

Thank you to Yunnan Jingjian Rail Transit Investment and Construction Co., Ltd. for providing the testing site.

Disclosure statement

The authors declare no conflict of interest.

References

- [1] Miao B, Zhang W, Liu J, et al., 2021, Review on Frontier Technical Issues of Intelligent Railways Under Industry 4.0. *Journal of Traffic and Transportation Engineering*, 21(1): 115–131. <https://doi.org/10.19818/j.cnki.1671-1637.2021.01.005>
- [2] Dan Y, Liu T, Chu W, et al., 2024, Application of Genetic Algorithm to Optimize Variational Mode Decomposition in Bearing Fault Feature Extraction. *Noise and Vibration Control*, 44(1): 148.
- [3] Wang Z, Yao L, Qi X, et al., 2021, Fault Diagnosis of Planetary Gearbox based on Parameter Optimized VMD and Multi-domain Manifold Learning. *Journal of Vibration and Shock*, 40(1): 110–118.
- [4] Kumar A, Vashishtha G, Gandhi CP, et al., 2021, Novel Convolutional Neural Network (NCNN) for the Diagnosis of Bearing Defects in Rotary Machinery. *IEEE Transactions on Instrumentation and Measurement*, 70: 1–10. <https://doi.org/10.1109/TIM.2021.3055802>
- [5] Lv J, Xiao Q, Zhai X, et al., 2024, A High-performance Rolling Bearing Fault Diagnosis Method based on Adaptive Feature Mode Decomposition and Transformer. *Applied Acoustics*, 224: 110156. <https://doi.org/10.1016/j.apacoust.2024.110156>
- [6] Patil AR, Buchaiah S, Shakya P, 2024, Combined VMD-morlet Wavelet Filter-based Signal De-noising Approach and Its Applications in Bearing Fault Diagnosis. *Journal of Vibration Engineering & Technologies*, 12: 7929–7953.

- [7] Yi C, Lin J, Zhang W, et al., 2015, Fault Diagnostics of Railway Axle Bearings based on IMF's Confidence Index Algorithm for Ensemble EMD. *Sensors*, 15(5): 10991–11011. <https://doi.org/10.3390/s150510991>
- [8] Donoho DL, 1995, De-noising by Soft Thresholding. *IEEE Transactions on Information Theory*, 41(3): 613–627. <https://doi.org/10.1109/18.382009>
- [9] Niu Y, Fei J, Li Y, et al., 2020, A Novel Fault Diagnosis Method based on EMD, Cyclostationary, SK and TPTSR. *Journal of Mechanical Science and Technology*, 34: 1925–1935.
- [10] Yu Y, Zhao X, Yu C, 2022, TMSST-CK Fault Feature Extraction Method for Flexible Thin-wall Bearing based on the Gini Index Principle. *Measurement Science and Technology*, 34: 025017.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.