



ARTICLE

Research on Gearbox Fault Diagnosis Method Based on Multi-Dimensional Feature Extraction and Random Forest

Yu Zhang^{1,2,#}, Shihan Tan^{1,#}, Guangyao Lian², Congying Dun³, Qiwei Hu^{1,*} and Chiming Guo^{1,*}

¹Shijiazhuang Campus, Army Engineering University of PLA, Shijiazhuang, China

²No. 32181 Unit of PLA, Xian, China

³Army Command Academy, Nanjing, China

*Corresponding Authors: Qiwei Hu. Email: hu_q_w@163.com; Chiming Guo. Email: guochiming@nudt.edu.cn

#These authors contributed equally to this work as the first author

Received: 12 March 2026; Accepted: 22 April 2026; Published: 15 June 2026

ABSTRACT: Gearboxes are critical components in the transmission systems of various mechanical equipment. Subjected to complex and harsh operating conditions for a long time, they suffer from a high failure rate and potentially severe consequences. Traditional fault diagnosis methods are limited by problems such as noise interference, and can hardly meet the requirements in terms of diagnostic accuracy, generalization ability, and reliability. To tackle the deficiencies of traditional gearbox fault diagnosis methods, including insufficient utilization of features, poor generalization under small-sample conditions, and weak model interpretability, this paper proposes a fault diagnosis method based on multi-dimensional feature extraction and Random Forest (RF). This method integrates intelligent computing, data-driven approaches, and mechanical structural health monitoring. First, fault feature analysis is conducted from multiple dimensions including time domain, frequency domain, and envelope domain, and visualization verification is implemented using Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE). Then, the Random Forest (RF) algorithm is used for dataset training and testing, obtaining a stable diagnostic model with strong generalization ability. Finally, experimental analyses verify the effectiveness and superiority of the proposed method. The research results possess high application potential and practical value in improving the performance of gearbox fault diagnosis.

KEYWORDS: Fault diagnosis; feature extraction; principal component analysis; gearbox; random forest

1 Introduction

As a core power transmission component in the drive systems of mechanical equipment, the gearbox operates for long periods under complex and harsh working conditions such as high rotational speeds, alternating loads, and intense impact vibrations [1], making it prone to failures like tooth surface wear, tooth breakage, and gearbox damage. With a high failure rate and far-reaching impacts, it is the primary source of faults in drive system equipment [2]. Once a gearbox malfunctions, it may cause minor issues such as abnormal power transmission of the equipment and delays, or even lead to serious accidents like drive system failure in severe cases. At present, gearbox condition monitoring mainly relies on characteristic indicators constructed based on expert experience or traditional signal processing technologies [3]. However, with the intelligent upgrading of equipment, existing methods have been difficult to meet the requirements of accurate diagnosis in complex scenarios in terms of diagnostic accuracy, generalization ability, and real-time performance.

In recent years, the vigorous development of big data and artificial intelligence technologies has brought new opportunities to the field of gearbox fault diagnosis. Data-driven intelligent fault diagnosis methods build fault diagnosis models based on massive operational data [4], which can achieve higher-precision fault identification, stronger working condition adaptability and more efficient real-time diagnosis capability. There are relatively many studies on gearbox fault signal processing and feature extraction. For example, Li et al. [4] proposed a coarse-grained lattice feature method. After adaptive filtering, frequency-domain segmentation and other processes, combined with Swin Transformer modeling, the diagnostic accuracy of public datasets and experimental data exceeds 98%. Xiao et al. [5] proposed a combined method of NMD and wavelet threshold denoising. After preprocessing the signal, decomposition and envelope spectrum are performed. Experimental comparison shows that the accuracy of extracting fault characteristic frequencies is better than that of the EMD method. Bie et al. [6] proposed an improved method combining ESMD and SVM, which decomposes vibration signals to select high kurtosis envelope spectrum exponential components and constructs feature vectors using various entropies, thus effectively extracting the impact fault features of gearboxes. Cao et al. [7] proposed a method combining quadratic wavelet packet energy entropy and t-SNE, which decomposes signals to extract entropy values and fuse features, and realizes fault state identification effectively with the help of support vector machine recognition. Zhang and Wang [8] proposed a method combining wavelet packet decomposition and tree-structured pipeline optimization tool, which extracts feature vectors and then uses genetic programming to generate the optimal machine learning pipeline. Experiments confirm that this method has significant advantages. Zhao et al. [9] combined the multi-point optimized minimum entropy deconvolution correction method with improved adaptive noise complete ensemble empirical mode decomposition, and through envelope demodulation, can accurately extract the characteristic frequencies of weak faults. Zhu et al. [10] proposed a transmission path elimination enhanced variational mode decomposition method, combined with spectral editing and whale optimization algorithm for noise reduction, which can accurately extract fault features under complex transmission paths. Zheng et al. [11] proposed a spectral whitening demodulation method for bogie axle box gearboxes, which is adapted to the complex working conditions of high-speed trains, can effectively extract fault features and ensure diagnostic accuracy. Li et al. [12] pointed out that existing signal decomposition algorithms suffer from mode mixing, unstable decomposition, and poor anti-noise performance, which hinder subsequent feature extraction and fault diagnosis. To address these issues, they proposed a Spectral Distribution Decomposition (SDD) algorithm based on the Spectral Probability Density Function (SPDF) for gearbox fault diagnosis. The effectiveness and superiority of the proposed method are verified through comparative simulations with existing decomposition techniques, as well as experimental tests on fault diagnosis for laboratory gearboxes and actual wind turbine gearboxes. Chen et al. [13] aimed to solve the problem of scarce fault data for gearboxes and bearings, as well as the difficulty of effective modeling based on single health state data. They proposed a Convex Optimization Differential Analysis model (CODA model). Within a dual-spectrum framework based on natural frequency and harmonic frequency demodulation, a differential analysis mechanism is introduced to discretize complex modulation features and realize the decoupling of modulation information among multiple components.

There have also been many relevant studies on the construction of gearbox fault diagnosis models. Li et al. [14] proposed a domain-adaptive LSTM-DNN model embedded with the MMD loss function to reduce the distribution discrepancy under variable working conditions. Compared with the single LSTM model, the mean absolute error (MAE) of the proposed model is reduced by 35%, which is suitable for the cross-working condition life prediction of planetary gearboxes. Hogeia et al. [15] constructed a LogicLSTM model integrating XAI and logical tensor networks, and realized logical-neural training by virtue of the ApME loss function, achieving accurate and interpretable diagnosis of 9 types of states on the data from

the DDS platform. Dou et al. [16] converted one-dimensional signals into two-dimensional graphs via Gramian Angular Field (GAF), built a lightweight network with coordinate attention, and combined it with transfer learning fine-tuning to realize cross-component fault diagnosis of gearboxes under variable working conditions with few samples. Dong et al. [17] proposed a framework based on the nonlinear Wiener process, processed vibration signals using kernel principal component analysis (KPCA), and updated parameters through Bayesian inference. A 954-h experiment verified that the framework can dynamically optimize the maintenance time of gearboxes. Yuan et al. [18] constructed a model integrating improved multi-scale convolutional neural network (CNN) and lightweight convolutional attention, and introduced the parametric rectified linear unit. The diagnostic accuracy of the model exceeds 98.9% under complex working conditions, balancing efficiency and precision. Cheng et al. [19] designed a lightweight channel attention mechanism, obtained the time-frequency distribution of signals through wavelet transform, and combined it with transfer learning to fine-tune the network, efficiently solving the problem of gearbox fault classification under multiple working conditions with few samples. Nguyen et al. [20] combined adaptive noise control with stacked sparse autoencoder to eliminate noise in vibration signals, completed feature extraction and classification in an integrated manner, and improved the diagnostic sensitivity for multi-stage tooth breakage faults under variable rotational speeds. Desai et al. [21] fused SCADA time-series data and physical modeling data as model inputs to predict faults in wind turbine gearboxes, reducing the false alarm rate by 50%, improving the accuracy by 33%, and realizing fault early warning one month in advance. Zhu et al. [22] proposed a novel federated learning method based on an incremental generalized federated learning system to address the challenges of long global model training time and model forgetting in deep network federated learning approaches. This method employs random mapping as a bridge between broad learning and federated learning, constructing a federated learning framework built on the broad learning system. Kan et al. [23] developed a fault diagnosis method based on dual-layer adaptive personalized federated learning to tackle the issues of privacy leakage and statistical heterogeneity that often restrict the performance of fault diagnosis in industrial processes. With the aid of federated learning, multiple clients train their models locally, which effectively resolves the problem of privacy leakage.

However, in practical application scenarios, the raw vibration signals collected by sensors are affected by complex operating environments and variable working conditions, and are easily mixed with components such as background noise and interference responses. This greatly weakens the saliency of fault features. Although existing studies have made progress in feature extraction and model construction, limitations still exist: first, most methods rely only on single-dimensional features in the time or frequency domain, and cannot fully characterize fault information from non-stationary and strong-noise signals; second, traditional models have insufficient generalization ability for small-sample and high-dimensional features; third, there is a lack of systematic demonstration of feature dimensions, model adaptability, and hyperparameter selection. To fill the above gaps, this paper proposes a diagnostic framework based on multi-domain feature fusion combined with random forest, to achieve feature complementarity and improved model robustness.

In summary, existing fault diagnosis methods still have obvious shortcomings in the collaborative utilization of multi-domain features, adaptability to small-sample working conditions, model interpretability, and engineering practicability. Therefore, this paper proposes a fault diagnosis method based on multi-dimensional feature extraction and random forest. First, fault feature analysis is performed from multiple dimensions including the time domain, frequency domain, and envelope domain, and visualization verification is carried out using PCA and t-SNE. Then, the RF method is used for dataset training and testing to obtain a stable diagnostic model with strong generalization ability. Finally, experimental comparative analysis verifies the effectiveness and superiority of the proposed method. By systematically fusing time-domain, frequency-domain, and envelope-domain features, and combining an optimized random forest to construct

an end-to-end diagnostic framework suitable for small-sample working conditions, multi-dimensional information complementarity and high-precision fault identification are realized.

2 Fault Feature Extraction Based on Multi-Dimensions

A multi-dimensional feature extraction model is constructed and the optimal feature set parameters are obtained. The diagnostic accuracy is improved through consistency verification of features across different dimensions. This method is suitable for complex non-stationary fault signals of gearboxes, can significantly enhance feature distinguishability and diagnostic robustness, and provides a more comprehensive and reliable basis for subsequent fault identification and classification.

This paper constructs a comprehensive, complementary and highly discriminative feature set from three physical domains: time domain, frequency domain and envelope domain, rather than relying solely on a single domain or a small number of statistics. The extracted features cover four categories: amplitude statistics, distribution morphology, energy distribution and modulation characteristics, which together form a genuine multi-domain feature space. This space can reflect the laws of fault evolution from different perspectives, ensuring the completeness and discriminability of the feature space.

2.1 Vibration Signal Preprocessing

Raw vibration signals are susceptible to environmental noise, electromagnetic interference, and baseline drift during acquisition. Direct feature extraction will degrade diagnostic accuracy. Therefore, standardized preprocessing is performed on the signals prior to multi-domain feature extraction, with the procedure as follows:

(1) Wavelet threshold denoising

The db4 wavelet is used for 3-level decomposition, and soft threshold filtering is adopted to remove high-frequency random noise while retaining fault impact features.

(2) Zero-mean normalization

Zero-mean processing is applied to the denoised signal to eliminate the effects of sensor zero offset and amplitude dimensions, with the formula as follows:

$$x'_i = x_i - \mu \quad (1)$$

where μ denotes the mean value of the signal.

(3) Fixed-length segmentation and resampling

Each 6-s signal is segmented into a fixed length of 2048 points to ensure consistent input feature dimensions and avoid errors caused by inconsistent sample lengths.

After the above steps, feature extraction in the time domain, frequency domain, and envelope domain is performed.

2.2 Time-Domain Feature Extraction

Time-domain features directly reflect the statistical characteristics of vibration signals. By extracting the statistical and morphological features of signals, they reflect the amplitude distribution, fluctuation law and mutation characteristics of signals, and are sensitive to the transient impact of early faults in gearboxes. The time-domain feature vector is defined to include the following indicators:

(1) Root Mean Square (RMS)

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (2)$$

where N is the signal length and x_i is the i -th sampling point. The RMS is used to measure the overall energy level of the signal. Gearbox faults will lead to an increase in vibration energy, and the RMS value will rise accordingly.

(2) Kurtosis (K)

$$K = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^4}{\sigma^4} \quad (3)$$

where μ is the mean value of the signal and σ is the standard deviation. Kurtosis reflects the sharpness of the signal distribution. When the impact caused by faults thickens the tail of the distribution, the kurtosis value increases.

2.3 Frequency-Domain Feature Extraction

Frequency-domain features reveal the periodic components of signals through Fourier transform. As a core mathematical tool in the field of signal processing, Fourier transform can convert signals in the time domain to the frequency domain, revealing the composition of signal frequency components, which breaks the limitation that time-domain analysis is difficult to intuitively present the frequency characteristics of signals.

$$F(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-j\omega t} dt \quad (4)$$

where ω is the angular frequency and j is the imaginary unit. Through integral operation, the time-domain signal is decomposed into a superposition form of sine/cosine components with different frequencies.

The spectral centroid is an indicator reflecting the concentration trend of signal energy distribution in the frequency domain. The more the energy shifts to high frequencies, the larger the value of the spectral centroid.

$$C = \frac{\sum_{n=0}^{N-1} f(n) \cdot |X(n)|^2}{\sum_{n=0}^{N-1} |X(n)|^2} \quad (5)$$

where C is the spectral centroid, $f(n)$ is the frequency at the n -th point, $|X(n)|^2$ is the power spectrum corresponding to the frequency, and the denominator is the total power.

2.4 Envelope-Domain Feature Extraction

Envelope analysis is an important method for extracting amplitude variation features in the field of signal processing [24]. Its core lies in separating the slowly varying amplitude envelope that reflects key information from complex signals modulated by high-frequency carriers. This method can effectively eliminate redundant high-frequency components [25] and highlight the core variation trend of signals.

$$A(t) = |z(t)| = \sqrt{x^2(t) + \hat{x}^2(t)} \quad (6)$$

where the original signal is denoted as $x(t)$, the signal obtained via Hilbert transform is denoted as $\hat{x}(t)$, and the analytic signal is denoted as $z(t) = x(t) + j\hat{x}(t)$.

The envelope root mean square is an indicator reflecting the concentration degree of signal envelope energy, which can effectively highlight the amplitude fluctuations caused by faults and is often used in equipment fault diagnosis.

$$RMS_{env} = \sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2} \quad (7)$$

RMS_{env} is the envelope root mean square value, N is the number of sampling points of the envelope signal, and B is the sampling value of the i -th envelope signal.

In this paper, 12 effective features are extracted from three dimensions: time domain, frequency domain, and envelope domain, and a comprehensive feature matrix is constructed for model training and testing. The dimension of the feature matrix finally input into the random forest model is 32×12 , where 32 is the total number of samples and 12 is the dimension of the feature vector. The detailed composition is shown in Table 1.

Table 1: Composition of feature vector dimensions.

Domain	Feature Name	Physical Meaning
Time Domain	Root Mean Square, Kurtosis, Crest Factor, Root Amplitude, Waveform Index, Impulse Index	Reflects vibration amplitude, impact strength and distribution characteristics
Frequency Domain	Spectral Centroid, Mean Square Frequency, Frequency Variance	Reflects the distribution law of frequency energy
Envelope Domain	Envelope RMS, Envelope Kurtosis, Envelope Crest Factor	Reflects fault modulation and impact characteristics

3 Fault Diagnosis Model Based on Random Forest

According to the feature extraction results of the fault data mentioned above, the Random Forest (RF) method is adopted to construct an intelligent fault diagnosis model for gearboxes.

3.1 RF Model Construction

In this paper, multi-dimensional features are constructed from the time domain, frequency domain, and envelope domain, with weak correlation among these features. Random Forest (RF) is insensitive to high-dimensional features and does not require complex feature selection, enabling direct fusion of multi-domain information. Meanwhile, given the limited sample size in the gearbox experiment, RF reduces the risk of overfitting through ensemble and randomization strategies, and exhibits strong robustness to environmental noise in vibration signals, which conforms to actual industrial field conditions. In contrast, SVM relies on appropriate kernel function selection, and MLP requires a large amount of data and careful parameter tuning. RF imposes no strict assumptions on data distribution, has fewer parameters and is easier to optimize, making it more suitable for engineering applications.

Random Forest (RF) is a powerful ensemble learning algorithm composed of M decision trees [26]. Each decision tree is constructed based on a different training subset, which is obtained from the original

training set through random sampling with replacement [27]. This sampling method enables each decision tree to be trained on relatively independent and differentiated data, thereby increasing the diversity of the model. The workflow diagram of the RF model is shown in Fig. 1.

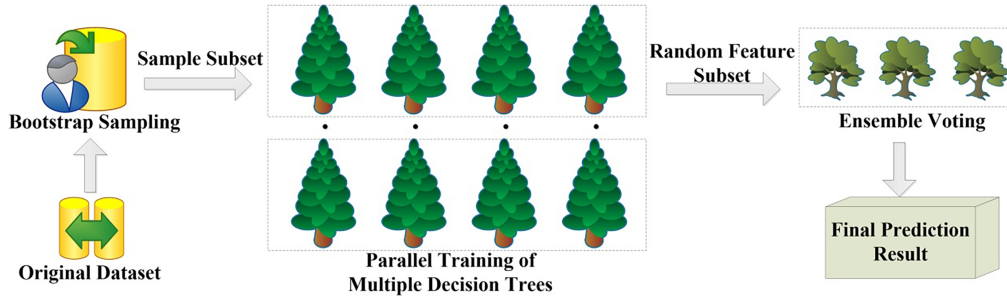


Figure 1: Workflow of the random forest model.

For classification problems, each decision tree T_m independently performs classification prediction based on the input gearbox feature vector x [28]. The final prediction result is determined by a voting mechanism, and the specific formula is as follows:

$$\hat{y} = \arg \max_c \sum_{m=1}^M I(T_m(x) = c) \quad (8)$$

where \hat{y} is the final predicted class of the RF model; c denotes the fault class; $I(\cdot)$ is the indicator function.

3.2 Parameter Optimization

The selection of parameters has a crucial impact on the performance of the RF model. To improve the accuracy and stability of the model, this paper adopts the method of grid search combined with cross-validation to optimize its key parameters.

Number of decision trees: The number of decision trees determines the scale of the ensemble model. A larger number of decision trees can enhance the stability and generalization ability of the model, but it will also increase the computational cost [29]. The value range considered in this paper is $\{50, 100, 200, 300\}$.

Maximum depth: The maximum depth limits the growth of decision trees. A smaller maximum depth can prevent overfitting, but may lead to underfitting; a larger maximum depth can enable the model to better fit the training data, but may increase the risk of overfitting. The value range is set as $\{5, 10, 15, 20, \text{None}\}$, where None means no depth limit, the decision tree will grow until all leaf nodes are pure or other stopping conditions are met.

Number of split features: When splitting at each node, a certain number of features are randomly selected for evaluation. This increases the diversity among decision trees and helps improve the overall performance of the model. The value range is $\{\sqrt{n}, \log_2(n), n\}$, where n is the total number of features.

Minimum number of samples at leaf nodes: This parameter limits the minimum number of samples in leaf nodes and prevents decision trees from being overly complex. The value range is $\{1, 2, 4\}$.

The objective function adopts the average accuracy of 5-fold cross-validation, and the calculation formula is as follows:

$$\text{Score} = \frac{1}{5} \sum_{i=1}^5 \text{Acc}_i \quad (9)$$

By traversing all possible hyperparameter combinations and calculating the objective function value corresponding to each combination, the optimal hyperparameter combination of the random forest model is finally determined in this paper as follows:

$$\left\{ \begin{array}{l} n_estimators = 100 \\ max_depth = 10 \\ max_features = \sqrt{n} \\ min_samples_leaf = 1 \end{array} \right\} \quad (10)$$

This set of parameters ensures the classification accuracy of the model while effectively avoiding overfitting, enabling the model to achieve better stability and generalization ability on small-sample datasets.

3.3 Model Evaluation Metrics

To comprehensively evaluate the performance of the model on the four-classification problem, the following evaluation metrics are adopted in this paper:

Accuracy: As the most commonly used evaluation metric, accuracy refers to the proportion of correctly predicted samples to the total number of samples. The calculation formula is as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

Macro-Precision: Macro-average precision takes into account the precision of each category and avoids the impact of class imbalance on the evaluation results. The calculation formula is as follows:

$$P_{macro} = \frac{1}{k} \sum_{i=1}^k P_i \quad (12)$$

Macro-Recall: Macro-average recall measures the model's ability to identify faults of each category. The calculation formula is as follows:

$$R_{macro} = \frac{1}{k} \sum_{i=1}^k R_i \quad (13)$$

F1-Score: F1-score is the harmonic mean of precision and recall, which comprehensively considers the precision and recall of the model. The calculation formula of macro-average F1-score is as follows:

$$F1 = \frac{2 \times P_{macro} \times R_{macro}}{P_{macro} + R_{macro}} \quad (14)$$

where TP , TN , FP , and FN denote true positive, true negative, false positive, and false negative, respectively; k is the number of categories; P_i and R_i represent the precision and recall of the i -th fault category, respectively.

4 Experimental Verification

To verify the effectiveness and superiority of the proposed gearbox fault diagnosis method, vibration signals of the gearbox were collected using a test bench. Time-domain, frequency-domain, and envelope-domain features were extracted and input into the constructed RF model for experimental analysis and comparative verification, thereby providing experimental support for the diagnostic application of gearbox faults.

4.1 Experimental Platform

The experimental platform consists of two main parts: a gear reducer and a signal acquisition system, whose structure is shown in Fig. 2. The core components include a base, a gear reducer, a three-phase asynchronous motor, an electromagnetic speed-regulating motor, a magnetic powder brake, a speed regulator, and a signal acquisition device. In this experiment, the JZQ250 type two-stage parallel shaft gearbox is taken as the research object. This gear reducer is widely used in various mechanical equipment and has important research significance.

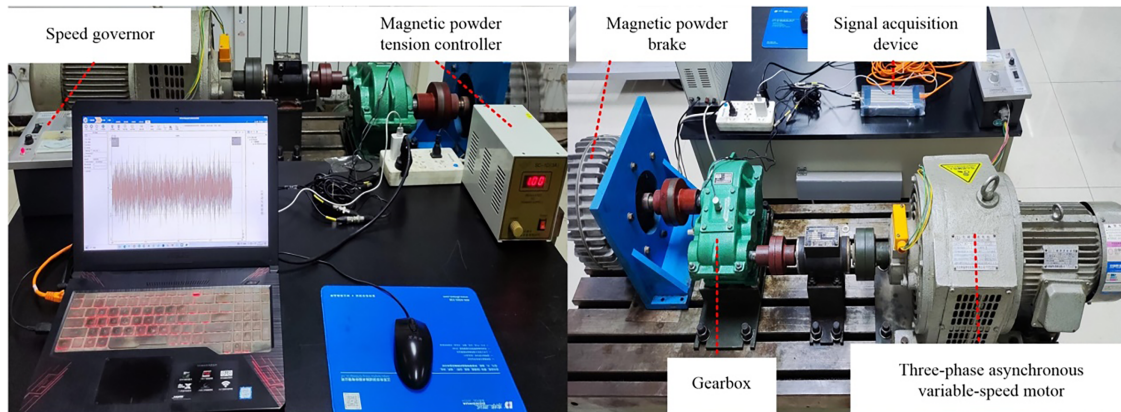


Figure 2: Gearbox experimental platform.

For the vibration signal acquisition process, the Donghua DH5981 dynamic signal acquisition instrument was selected, paired with the IEPE low-impedance voltage-output general-purpose vibration sensor (Model 14100), with the sampling frequency set to 20 kHz.

An industrial-standard vibration sensor arrangement is adopted. The sensors are installed vertically on the bearing housing of the gearbox, close to the shortest transmission path of vibration energy, which can effectively capture gear meshing vibration and fault impact signals, ensuring that the acquired vibration data have high signal-to-noise ratio and high representativeness. This installation position follows the general specification layout for rotating machinery fault diagnosis, which can stably obtain fault-sensitive features and meet the requirements of fault feature extraction and diagnostic identification.

The specific connection method is as follows: after installing the vibration sensor at the designated position of the reducer, it is connected to the Donghua DH5981 dynamic signal acquisition instrument via a dedicated connecting cable. The acquisition instrument then transmits the vibration signals to the supporting signal acquisition software of Donghua. This software can implement core functions such as signal storage, format conversion and preliminary analysis, providing fundamental support for subsequent data processing.

4.2 Vibration Signal Acquisition Scheme

Common fault modes of gearboxes include gear wear, gear cracks, gearbox tooth breakage, and tooth surface scuffing. This experiment mainly takes gear cracks as an example for research and analysis. First, to explore how to identify gear crack faults of varying degrees, four working conditions were set for the intermediate gear under the same operating conditions, namely normal state, 2 mm crack, 5 mm crack, and 8 mm crack.

Vibration signal acquisition should be initiated after the experimental platform operates stably, so as to avoid the interference of factors such as insufficient equipment preheating and voltage fluctuations in the

initial stage of startup on signal quality. The experimental parameters are set as follows: the motor input speed is 1200 r/min, the sampling frequency of the vibration sensor is 20 kHz, the data acquisition duration of each group is 6 s, and the next group of acquisition is carried out after an interval of 5 s between groups, with a total of 8 groups of data obtained. The specific signal acquisition scheme is shown in [Table 2](#).

Table 2: Fault signal acquisition scheme.

Working Condition Label	Working Condition Type	Sampling Frequency	Sampling Time	Sampling Interval	Number of Samples	Input Rotational Speed
S1	Normal State	20,000 Hz	6 s	5 s	8	1200 r/min
S2	2 mm Crack	20,000 Hz	6 s	5 s	8	1200 r/min
S3	5 mm Crack	20,000 Hz	6 s	5 s	8	1200 r/min
S4	8 mm Crack	20,000 Hz	6 s	5 s	8	1200 r/min

To fully verify the diagnostic performance and stability of the proposed method, experiments are carried out in this paper under the premise of consistent controlled variables and strictly identical working conditions. Eight groups of high-quality, high-signal-to-noise ratio valid samples are collected under each working condition to ensure good representativeness and consistency of the samples. A stratified 5-fold cross-validation strategy is adopted in the experiment to comprehensively evaluate the model. Through multiple random divisions of the training set and test set, the value of data is fully exploited, and the reliability and statistical significance of experimental results are effectively improved.

4.3 Fault Feature Extraction

For the collected dataset samples, the stratified 5-fold cross-validation strategy is adopted for data partitioning. According to four working condition categories, all samples are randomly divided into five mutually exclusive subsets via proportional stratified sampling. In each iteration, four subsets are used as the training set and one subset as the test set. The process is repeated five times and the average results are obtained.

Afterwards, comparative analysis of signal waveforms is conducted, followed by multi-dimensional feature extraction using the method proposed above. Finally, distribution boxplots are applied, and dimensionality reduction and visualization are performed on the multi-dimensional feature space based on PCA and t-SNE.

As shown in [Fig. 3](#): The vibration amplitude of the healthy state S1 remains stable within ± 0.2 g, with a steady signal and no obvious impacts; the amplitude of the 2 mm crack state S2 expands to ± 0.3 g, and periodic impacts begin to appear; the amplitude of the 5 mm crack state S3 further increases to ± 0.4 g, with shortened impact intervals and enhanced energy; the maximum amplitude of the 8 mm crack state S4 exceeds ± 0.5 g, with both impact intensity and frequency reaching the highest levels. As the crack size increases, the vibration amplitude, impact characteristics, and energy level show a stepwise increasing trend, which is consistent with the gear crack propagation mechanism.

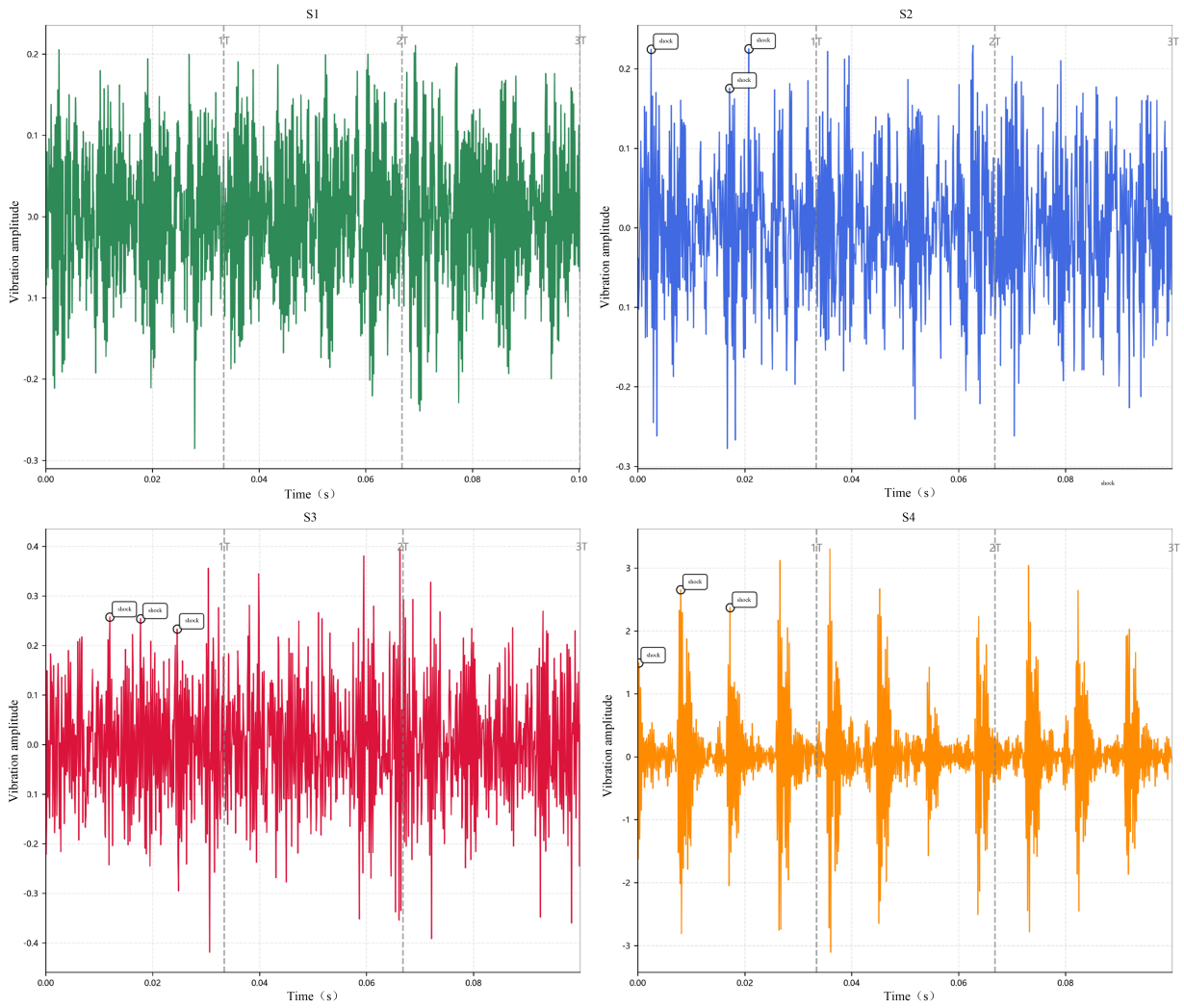


Figure 3: Comparison chart of time-domain waveforms.

As illustrated in Fig. 4, the energy distribution of the normal state S1 is uniform across the full frequency band without distinct peaks. Obvious prominent peaks emerge at the gear meshing characteristic frequencies for S2, S3 and S4, and the peak amplitudes increase by 2 to 5 times with the expansion of cracks. Meanwhile, the vibration energy gradually concentrates in the high-frequency band, which is positively correlated with the fault severity.

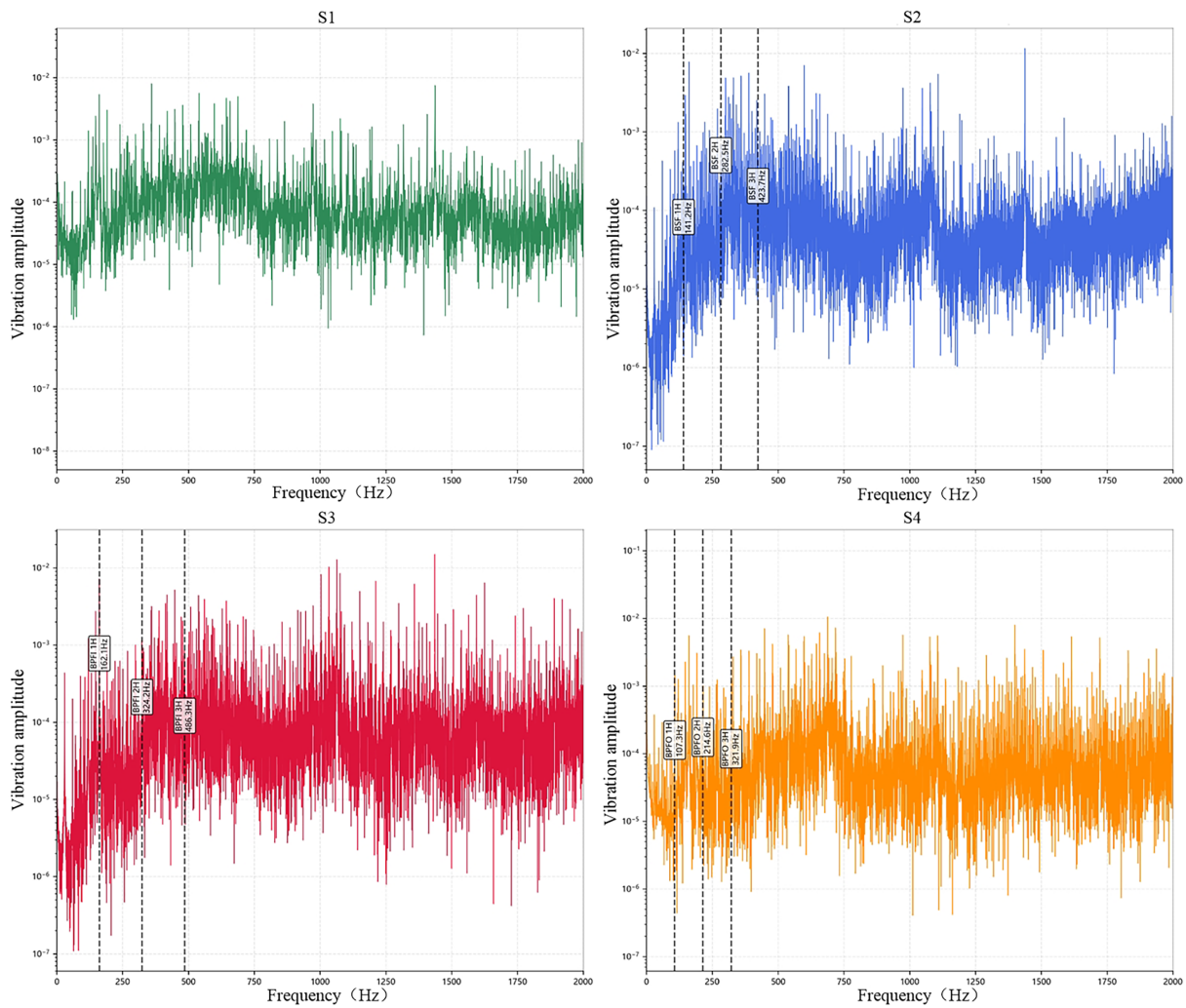


Figure 4: Frequency domain analysis diagram.

It can be observed from Fig. 5 that envelope analysis can effectively separate high-frequency carriers and highlight the slowly varying characteristics of fault impacts. The envelope curve of S1 is stable with slight amplitude fluctuations. For S2–S4, the envelope amplitude gradually increases and the fluctuations become more severe as the crack propagates, and the root mean square value of the envelope rises by more than three times, which clearly reflects the evolution law of fault severity.

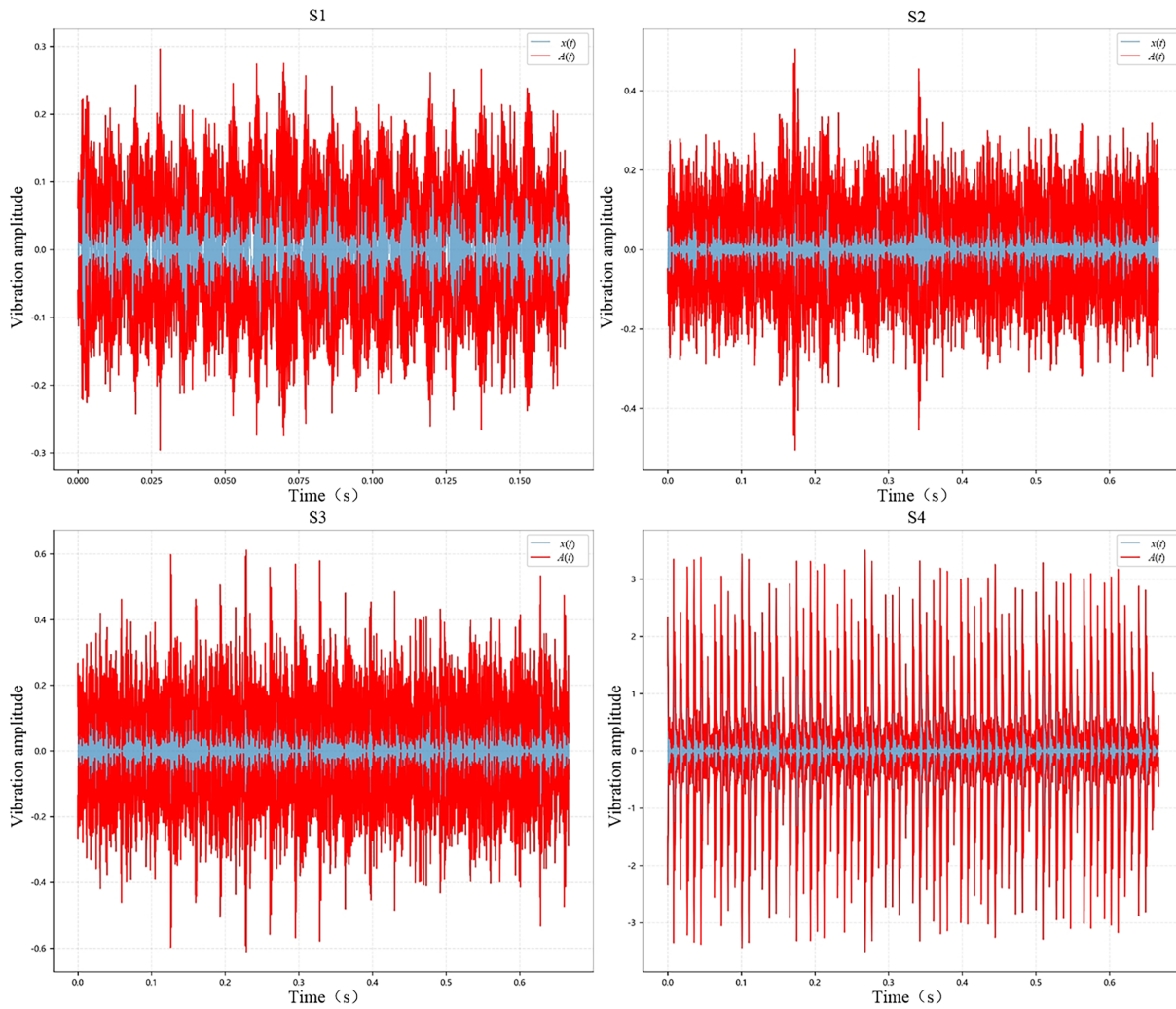


Figure 5: Comparison chart of envelope analysis.

Fig. 6 presents the box plot of feature distribution, a statistical chart for displaying data distribution that enables rapid identification of data features and anomalies. Significant differences exist in the distribution of various statistical features of vibration signals, such as root mean square (RMS), kurtosis, spectral centroid, and envelope root mean square, when the gearbox operates normally and under different fault conditions. Specifically, in the normal state, the RMS values show concentrated distribution with relatively low magnitudes; the kurtosis values are low and centralized; the spectral centroid values are high and stable; and the envelope RMS values are low and concentrated. In contrast, under fault conditions, especially with an 8 mm gear crack, the RMS values exhibit more dispersed distribution with higher magnitudes; the kurtosis values increase sharply with scattered distribution; the spectral centroid values are low with fluctuations; and the envelope RMS values are high and dispersed. These differences indicate that various statistical features can serve as effective indicators for distinguishing between the normal and fault states of the gearbox as well as for identifying different fault types.

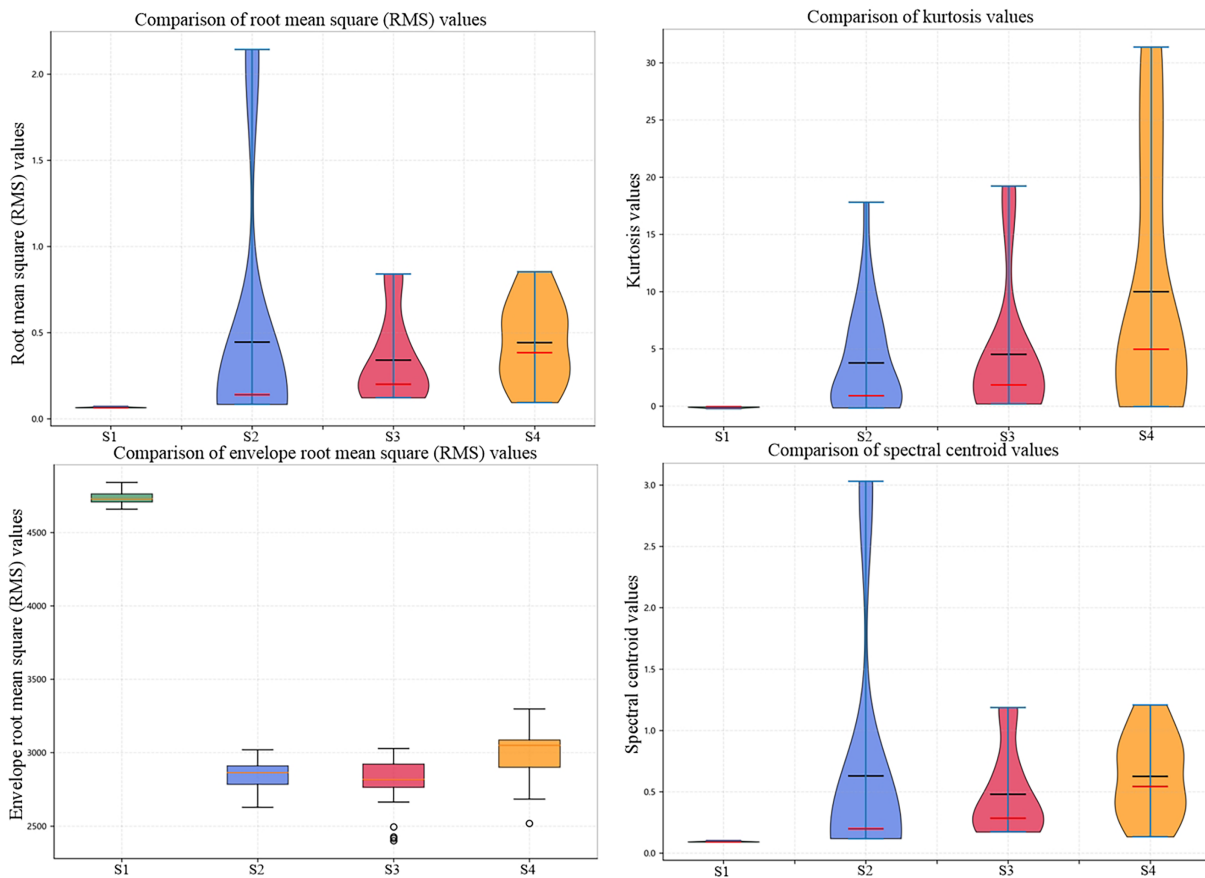


Figure 6: Box plot of feature distribution.

As can be seen from the multi-dimensional feature space analysis in Fig. 7, in the feature space distribution of Principal Component Analysis (PCA), the sample points of S1 state (green dots), S2 state (blue dots), S3 state (red dots) and S4 state (orange dots) have a certain degree of separability. The normal sample points are relatively concentrated in a specific area, while the fault sample points are distributed in different ranges, indicating that PCA can partition the feature space of gearbox samples under different states to a certain extent. The principal component contribution chart shows that the explained variance ratio of PC1 reaches 0.400, making it the most dominant principal component. With the increase in the number of principal components, the cumulative contribution degree rises gradually, which indicates that the first several principal components can explain most of the data variance. In the t-SNE feature space distribution, the clustering effect of sample points under different states is more obvious. The normal sample points and various types of fault sample points form relatively independent clusters, respectively, suggesting that t-SNE has a stronger ability to distinguish gearbox samples under different states after dimensionality reduction, which is more conducive to subsequent fault identification and classification.

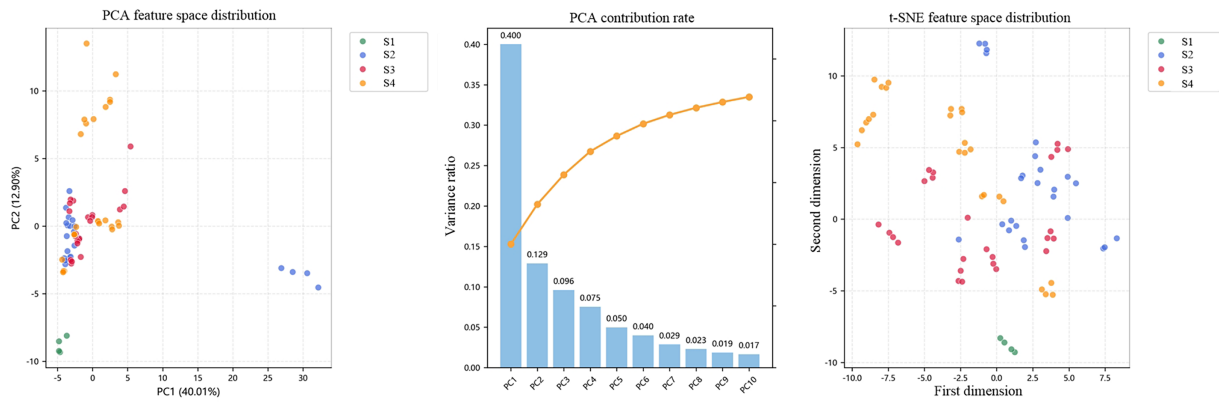


Figure 7: Multi-dimensional space analysis chart.

Finally, the multi-dimensional features including time-domain, frequency-domain and envelope analysis are integrated into a comprehensive feature matrix, forming a complete feature set that can reflect the multi-domain characteristics of gearbox faults.

4.4 Comparative Analysis of Fault Diagnosis Models

For the extracted fault feature set, the previously constructed fault diagnosis model based on Random Forest (RF) was trained, and a comprehensive comparative analysis was conducted. The models involved in the comparison included Random Forest (RF), Gradient Boosting (GB), Extremely Randomized Trees (ET), Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), Logistic Regression (LR), K-Nearest Neighbors (KNN), and Gaussian Naive Bayes (GNB). Through the training and evaluation of these models on the same dataset, quantitative analysis was carried out from multi-dimensional metrics such as accuracy, precision, recall, F1-score, and cross-validation stability. Ultimately, a comprehensive performance comparison result was formed, which demonstrated the effectiveness and superiority of the selected fault diagnosis model.

From the perspective of model characteristics, methods such as SVM and MLP usually have stronger nonlinear fitting capabilities in large-scale data scenarios, but their performance is sensitive to sample size. The multi-dimensional feature and random forest framework constructed in this paper is originally designed to adapt to industrial fault diagnosis scenarios with small samples and high-dimensional features. Through ensemble learning and random attribute selection mechanisms, random forest can maintain strong feature learning and classification ability under limited sample conditions. All comparative models in this paper adopt a unified grid search strategy to complete parameter optimization, ensuring consistent comparison conditions. The experimental results can effectively reflect the applicability and performance differences of different methods in gearbox fault diagnosis tasks with small samples.

Based on the comprehensive evaluation results in Table 3, the Random Forest (RF) algorithm exhibited the optimal performance in the fault diagnosis task, achieving an accuracy and F1-score of 95.83% each. It outperformed other algorithms significantly in both the recognition accuracy of various fault categories and overall stability, demonstrating excellent adaptability to the data characteristics of this task. The performance of Gradient Boosting (GB) was extremely close to that of RF, with its core metrics being basically on par. However, due to its serial iterative training mechanism, GB might be slightly less efficient when processing large-scale datasets, resulting in relatively poor adaptability to real-time diagnosis scenarios. Both Extremely Randomized Trees (ET) and Support Vector Machine (SVM) delivered good overall performance, with an

accuracy of 91.67% for each. Among them, ET was slightly more sensitive to noise than RF because of its more extreme random feature splitting strategy. Although SVM performed stably on medium-dimensional features, the parameter tuning difficulty and computational complexity of its kernel function would increase significantly when dealing with high-dimensional complex nonlinear relationships. Constrained by the sample size and feature dimensions, the Multi-Layer Perceptron (MLP) failed to give full play to the complex pattern fitting capability of deep learning models, achieving an accuracy of 87.50%, and its generalization performance in small-sample scenarios needs to be improved. Although the K-Nearest Neighbors (KNN) algorithm achieved an accuracy of 91.67%, its prediction speed would decrease significantly with the increase of data volume due to its mechanism relying on inter-sample distance calculation, making it difficult to meet the real-time requirements of large-scale diagnosis tasks. In contrast, Gaussian Naive Bayes (GNB) achieved an accuracy of only 66.67%, performing relatively poorly in this task, because its strict assumption of feature independence was inconsistent with the strong correlation among features in the actual fault data.

Table 3: Diagnostic performance of different algorithms.

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
RF	95.83	96.35	95.83	95.83
GB	95.83	96.30	95.83	95.80
ET	91.67	93.52	91.67	91.91
SVM	91.67	93.52	91.67	91.91
MLP	87.50	88.76	87.50	87.47
LR	79.17	84.63	91.67	79.16
KNN	91.67	91.67	91.67	91.67
GNB	66.67	68.75	66.67	62.97

Based on the multi-metric performance comparison chart of models in [Fig. 8](#), the Random Forest (RF) and Gradient Boosting (GB) models exhibited excellent performance across accuracy, precision, recall and F1-score, with their metric values remaining close and at high levels, which indicates that these two models have outstanding comprehensive performance in classification tasks. The Extremely Randomized Trees (ET) and Support Vector Machine (SVM) models also achieved relatively high values in all metrics, showing good performance. The Multi-Layer Perceptron (MLP) model performed moderately, while the Logistic Regression (LR) and Naive Bayes (NB) models yielded relatively low metric values. In particular, the NB model showed significantly lower accuracy, precision, recall and F1-score than other models, reflecting poor classification performance. In terms of the comparison of model performance radar charts, the RF and GB models stood out in all metrics, with their radar charts covering a wide area, which demonstrates their strong performance in multiple aspects. The ET model also reached a favorable level with balanced performance across all metrics. The radar charts of models such as SVM and K-Nearest Neighbors (KNN) covered a relatively smaller area, indicating slightly inferior performance. The NB model had the smallest coverage area in its radar chart, which further reflects its weak performance in these evaluation metrics. Overall, the above results are consistent with those in [Table 2](#). Ensemble learning models including RF, GB and ET achieved superior performance in this gearbox fault classification task, whereas simple models such as NB showed relatively insufficient performance.

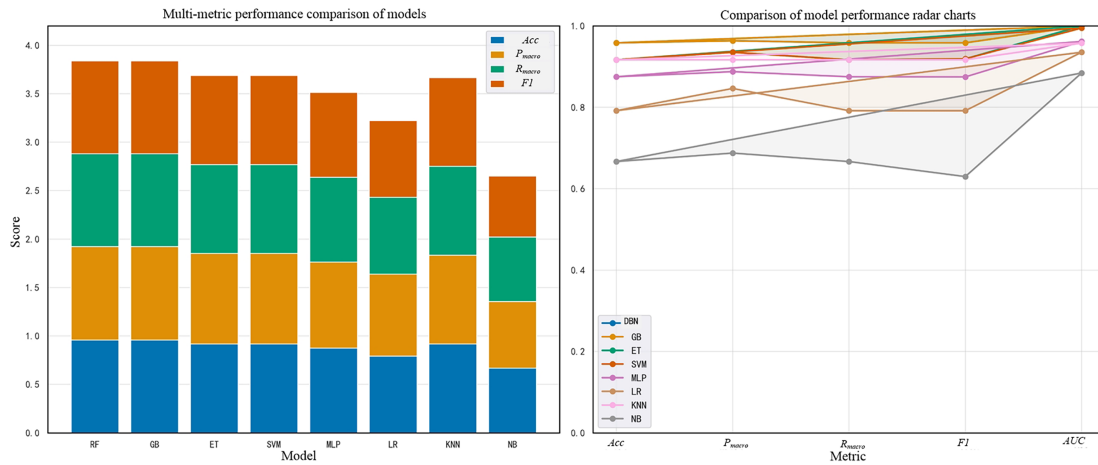


Figure 8: Multi-metric performance comparison chart of models.

As can be seen from Fig. 9, the Random Forest (RF) model demonstrated remarkable advantages. In terms of performance, its F1-score reached a relatively high level, close to 0.95, indicating that the RF model could accurately identify various types of faults and achieve excellent classification results in the gearbox fault classification task. Meanwhile, in terms of training time, the RF model required a relatively short training duration, which was much lower than that of the Gradient Boosting (GB) model. This characteristic of ensuring high performance while maintaining high training efficiency enables the RF model to not only complete the model training process rapidly in practical applications, but also provide accurate results for gearbox fault diagnosis, thus achieving a favorable balance between training efficiency and classification performance.

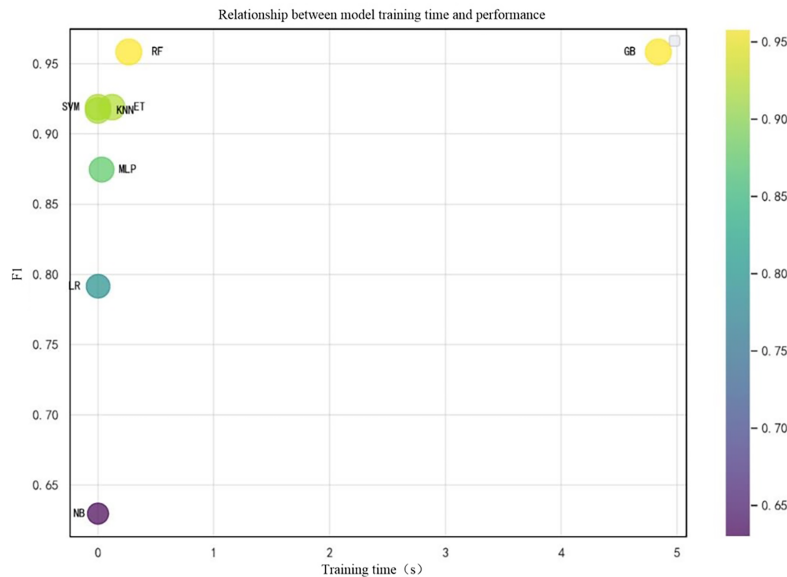


Figure 9: Relationship chart between model training time and performance.

As can be clearly observed from Fig. 10, the Random Forest (RF) model exhibited exceptionally outstanding performance. The narrow box of its box plot indicates that the RF model had minimal performance

fluctuations during the cross-validation process, demonstrating excellent stability. This means that the RF model could maintain stable and favorable classification results across different data subsets, with strong generalization ability. Meanwhile, the mean F1-score of the RF model stayed at a high level, which shows that in the gearbox fault classification task, the model could not only perform stably but also ensure high classification accuracy, enabling precise identification of various gearbox faults. Therefore, it serves as an excellent model choice that balances stability and accuracy for gearbox fault diagnosis tasks.

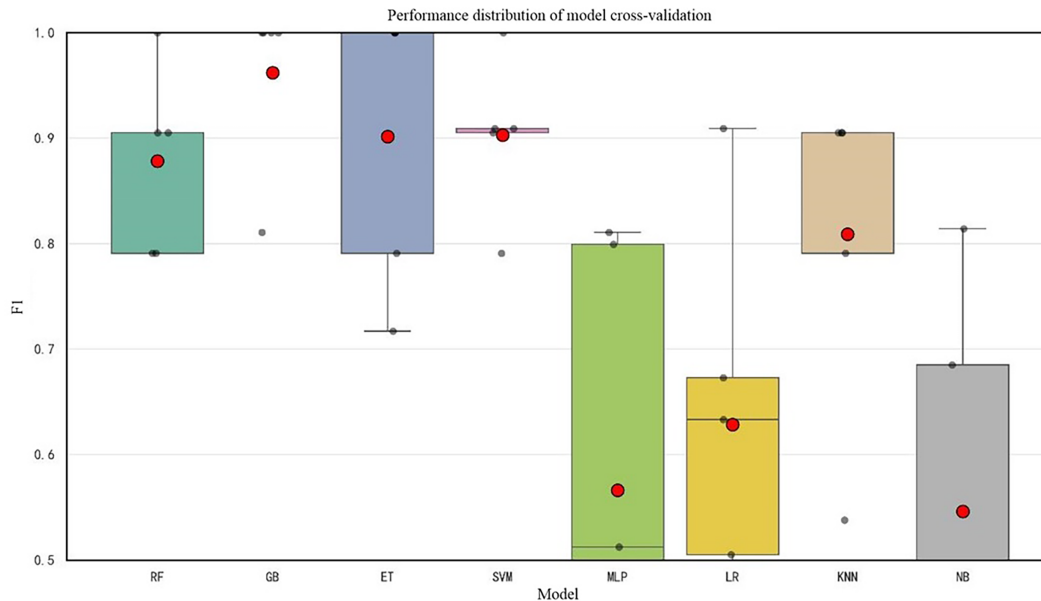


Figure 10: Distribution chart of model cross-validation performance.

According to the analysis results of the confusion matrix of the RF model in Fig. 11, the overall performance of RF is excellent, with an F1-score of 0.958, indicating good comprehensive classification performance of the model. In the raw count matrix, all 8 samples of state S1 are correctly predicted; all 8 samples of state S2 are accurately classified; and the 8 samples of state S3 are also correctly identified. However, one sample of state S4 is misclassified as S1, with the remaining seven samples correctly predicted. From the row-normalized proportion matrix, the prediction accuracies of classes S1, S2, and S3 are all 1.00, meaning all samples of these three classes are correctly classified. The prediction accuracy of class S4 is 0.88, indicating a certain degree of misclassification for this class, with 12% of S4 samples incorrectly predicted as other categories.

Based on the comprehensive evaluation results of all categories, the RF model demonstrated significant superiority in the source domain fault diagnosis task. In terms of performance, its accuracy and F1-score both reached 95.83%. In the multi-metric performance comparison, all core indicators maintained high and balanced levels, leading to remarkable classification results for gearbox faults and high overall classification accuracy. In terms of stability, the RF model showed minimal performance fluctuations in cross-validation, with a narrow box in the box plot, which indicates that it could maintain stable and favorable classification results across different data subsets and possessed strong generalization ability. In terms of efficiency, the RF model required a short training time, which was much lower than that of the GB model. It could complete training rapidly and be put into practical applications, achieving a sound balance between training efficiency and classification performance. Therefore, it is an ideal model choice for gearbox fault diagnosis tasks.

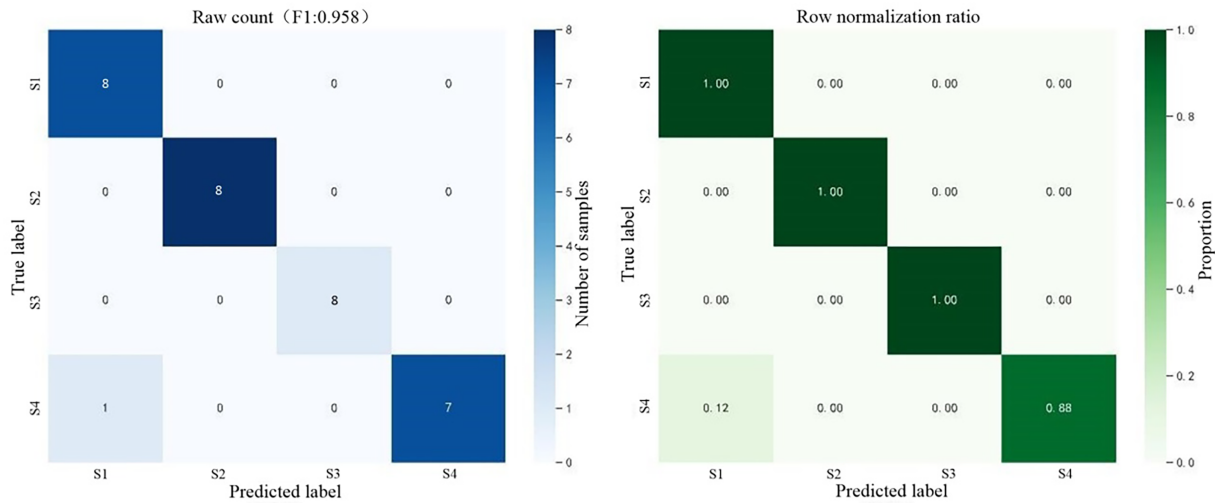


Figure 11: Confusion matrix heatmap of RF.

5 Conclusion

To address the challenges of insufficient diagnostic accuracy, weak generalization ability, and the difficulty of capturing fault information with single-dimensional features in gearboxes under complex working conditions, this study proposes a gearbox fault diagnosis method based on multi-dimensional feature extraction and Random Forest (RF). The main research conclusions are as follows:

- (1) The multi-dimensional feature extraction strategy significantly improves the completeness and distinguishability of fault information. By extracting key indicators such as root mean square values, this method captures the time-domain statistical characteristics, frequency-domain periodic components, and envelope amplitude variation features of gearbox vibration signals. It effectively overcomes the defect of incomplete characterization of fault information by single-dimensional features, providing more accurate feature support for subsequent diagnostic models.
- (2) The diagnostic model based on Random Forest achieves a good balance among diagnostic performance, training efficiency, and generalization ability. With strong modeling capability for complex data, the model exhibits excellent performance on the gearbox fault dataset. Compared with traditional models such as Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP), its accuracy and F1-score are both improved by more than 4%. Meanwhile, it features short training time, which can meet the efficiency requirements of practical diagnostic scenarios.
- (3) The effectiveness and superiority of the proposed method are verified through experimental validation and comparative analysis. The results of feature distribution boxplots and PCA, t-SNE visualization demonstrate that the extracted features can effectively distinguish different fault states. The confusion matrix and multi-metric performance comparison show that the Random Forest model achieves an accuracy of 95.83%, with minimal performance fluctuation in cross-validation and strong generalization ability. Its identification accuracy and stability for gearbox faults are significantly superior to those of other comparative algorithms.

This paper takes gear crack faults as the typical research object, and focuses on verifying the fault diagnosis performance under different crack severities. The proposed multi-dimensional feature extraction and random forest diagnosis framework has good feature generalization ability and model adaptability. It can not only effectively distinguish the severity of cracks, but is also theoretically applicable to the

classification and identification of various typical gearbox fault types such as tooth surface wear, tooth root fracture, and tooth surface scuffing. Future research will further expand the fault types, increase the number of samples and working conditions, carry out diagnosis experiments under multiple fault modes and various operating conditions, and comprehensively verify the universality and engineering applicability of the proposed method.

Acknowledgement: Author Chiming Guo sincerely acknowledges the financial support from the National Natural Science Foundation of China under contract number 71871220 (Titled: Dynamic Maintenance Optimization of Complex Systems Considering Correlation and Task Constraints).

Funding Statement: Not applicable.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Yu Zhang, Shihan Tan; methodology, Qiwei Hu; software, Chiming Guo; validation, Guangyao Lian; formal analysis, Congying Dun; investigation, Shihan Tan, Qiwei Hu; resources, Chiming Guo; data curation, Congying Dun; writing—original draft preparation, Yu Zhang; writing—review and editing, Shihan Tan, Guangyao Lian, Congying Dun; visualization, Qiwei Hu; supervision, Congying Dun; project administration, Chiming Guo. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Data available on request from the authors. The data that support the findings of this study are available from the Corresponding Author, [Chiming Guo], upon reasonable request.

Ethics Approval: Not applicable.

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

References

1. Kumar S, Kumar V, Sarangi S, Singh OP. Gearbox fault diagnosis: a higher order moments approach. *Measurement*. 2023;210:112489. doi:10.1016/j.measurement.2023.112489.
2. Damou A, Ratni A, Benazzouz D. Intelligent multi-fault identification and classification of defective bearings in gearbox. *Adv Mech Eng*. 2024;16(4):16878132241246673. doi:10.1177/16878132241246673.
3. Sun S, Zhang S, Wang W. A new monitoring technology for bearing fault detection in high-speed trains. *Sensors*. 2023;23(14):6392. doi:10.3390/s23146392.
4. Li X, Jia B, Liao Z, Wang X. Bearing-fault-feature enhancement and diagnosis based on coarse-grained lattice features. *Sensors*. 2024;24(11):3540. doi:10.3390/s24113540.
5. Xiao M, Wen K, Zhang C, Zhao X, Wei W, Wu D. Research on fault feature extraction method of rolling bearing based on NMD and wavelet threshold denoising. *Shock Vib*. 2018;2018:9495265. doi:10.1155/2018/9495265.
6. Bie F, Gu S, Guo Y, Yang G, Peng J. Research on gearbox impact feature extraction method based on improved ESMD. *Insight*. 2022;64(1):20–7. doi:10.1784/insi.2022.64.1.20.
7. Cao J, Zhang X, Yin R, Ma Z. The feature extraction method based on quadratic wavelet packet energy entropy and t-SNE for bearing fault diagnosis. *Proc Inst Mech Eng Part C J Mech Eng Sci*. 2025;239(2):520–31. doi:10.1177/09544062241283331.
8. Zhang D, Wang Y. The fault diagnosis of rolling bearing based on WPD and TPOT. In: *Proceedings of the 2019 Chinese Automation Congress (CAC); 2019 Nov 22–24; Hangzhou, China*. p. 1029–34. doi:10.1109/CAC48633.2019.8996312.
9. Zhao L, Zhang Y, Zhu D. Feature extraction for rolling element bearing weak fault based on MOMEDA and ICEEMDAN. *J Vibroeng*. 2018;20(6):2352–62. doi:10.21595/jve.2018.19309.
10. Zhu D, Chen J, Yin B. Fault feature extraction of rolling element bearing based on TPE-EVMD. *Measurement*. 2021;183:109880. doi:10.1016/j.measurement.2021.109880.

11. Zheng Z, Song D, Xu X, Lei L. A fault diagnosis method of bogie axle box bearing based on spectrum whitening demodulation. *Sensors*. 2020;20(24):7155. doi:10.3390/s20247155.
12. Li D, Wang G, Zhang M, Dai W. Spectral distribution decomposition and its application in gearbox fault diagnosis. *Mech Mach Theory*. 2026;220:106359. doi:10.1016/j.mechmachtheory.2026.106359.
13. Chen J, Xu Y, Liu T, Yang M, Zhang K, Zhang H. A convex optimization difference analysis model for intelligent fault detection and diagnosis of gearboxes. *Mech Syst Signal Process*. 2026;251:114207. doi:10.1016/j.ymsp.2026.114207.
14. Li H, Wang C, Zhang Y. A study of a domain-adaptive LSTM-DNN-based method for RUL prediction of planetary gearbox. *Processes*. 2023;11(8):2245–56. doi:10.3390/pr11072002.
15. Hogeia E, Onchiş DM, Yan R, Zhou Z. LogicLSTM: logically-driven long short-term memory model for fault diagnosis in gearboxes. *J Manuf Syst*. 2024;77:892–902. doi:10.1016/j.jmsy.2024.10.003.
16. Dou S, Cheng X, Du Y, Wang Z, Liu Y. Gearbox fault diagnosis based on Gramian angular field and TLCA-MobileNetV3 with limited samples. *Int J Metrol Qual Eng*. 2024;15:15. doi:10.1051/ijmqe/2024004.
17. Dong E, Zhang Y, Zhan X, Bai Y, Cheng Z. A novel dynamic predictive maintenance framework for gearboxes utilizing nonlinear Wiener process. *Meas Sci Technol*. 2024;35(12):126210. doi:10.1088/1361-6501/ad762e.
18. Yuan B, Li Y, Chen S. Efficient gearbox fault diagnosis based on improved multi-scale CNN with lightweight convolutional attention. *Sensors*. 2025;25(9):2636. doi:10.3390/s25092636.
19. Cheng X, Dou S, Du Y, Wang Z. Gearbox fault diagnosis method based on lightweight channel attention mechanism and transfer learning. *Sci Rep*. 2024;14(1):743. doi:10.1038/s41598-023-50826-6.
20. Nguyen CD, Prosvirin AE, Kim CH, Kim JM. Construction of a sensitive and speed invariant gearbox fault diagnosis model using an incorporated utilizing adaptive noise control and a stacked sparse autoencoder-based deep neural network. *Sensors*. 2021;21(1):18. doi:10.3390/s21010018.
21. Desai A, Guo Y, Sheng S, Phillips C, Williams L. Prognosis of wind turbine gearbox bearing failures using SCADA and modeled data. *Annu Conf PHM Soc*. 2020;12(1):10. doi:10.36001/phmconf.2020.v12i1.1292.
22. Zhu H, Wang H, Wang Y, Qian X, Wang Q. A bearing fault diagnosis method based on an incremental broad federated learning system. *Eng Res Express*. 2026;8(3):035234. doi:10.1088/2631-8695/ae39a5.
23. Kan X, Chen X, Li L, Peng X, Zhong M. A fault diagnosis method based on dual-layer adaptive personalized federated learning in wind turbines. *IFAC Pap*. 2025;59(20):1848–53. doi:10.1016/j.ifacol.2025.11.427.
24. Yang T, Guo Y, Wu X, Na J, Fung RF. Fault feature extraction based on combination of envelope order tracking and cICA for rolling element bearings. *Mech Syst Signal Process*. 2018;113:131–44. doi:10.1016/j.ymsp.2017.03.050.
25. Wang J, Qiao L, Ye Y, Chen Y. Fractional envelope analysis for rolling element bearing weak fault feature extraction. *IEEE/CAA J Autom Sin*. 2017;4(2):353–60. doi:10.1109/JAS.2016.7510166.
26. Liu K, Gu Y, Tang L, Du Y, Zhang C, Zhu J. Random forest grid fault prediction based on genetic algorithm optimization. *Front Phys*. 2025;13:1480749. doi:10.3389/fphy.2025.1480749.
27. Cerrada M, Zurita G, Cabrera D, Sánchez RV, Artés M, Li C. Fault diagnosis in spur gears based on genetic algorithm and random forest. *Mech Syst Signal Process*. 2016;70:87–103. doi:10.1016/j.ymsp.2015.08.030.
28. Quiroz JC, Mariun N, Mehrjou MR, Izadi M, Misron N, Mohd Radzi MA. Fault detection of broken rotor bar in LS-PMSM using random forests. *Measurement*. 2018;116:273–80. doi:10.1016/j.measurement.2017.11.004.
29. Liu J, Cai B, Yan S, Sun P. Transformer fault diagnosis based on the improved QPSO and random forest. *Meas Sci Technol*. 2024;35(9):096206. doi:10.1088/1361-6501/ad574c.