Article

Fault Diagnosis of Vehicle Gearboxes Based on Adaptive Wavelet Threshold and LT-PCA-NGO-SVM

Qingyong Zhang *, Changhuan Song and Yiqing Yuan

School of Mechanical and Automotive Engineering, Fujian University of Technology, Fuzhou 350118, China; 2210101001@smail.fjut.edu.cn (C.S.); yujie@fjut.edu.cn (Y.Y.)
* Correspondence: zhqy@fjut.edu.cn; Tel.: +86-189-5031-2790

Abstract: Vehicle gearboxes are subject to strong noise interference during operation, and the noise in the signal affects the accuracy of fault identification. Signal denoising and fault diagnosis processes are often conducted independently, overlooking their synergistic potential in practical applications. This article proposes a gearbox fault identification method that integrates improved adaptive modified wavelet function noise reduction, logarithmic transformation on principal component analysis (LT-PCA), and support vector machines (SVMs) to mitigate the influence of noise and feature outliers on fault signal recognition. Initially, to address the issue of interfering signals within the original signal, an innovative adaptive wavelet function optimized by the simulated annealing (SA) algorithm is employed for noise reduction of the main intrinsic mode function (IMF) components decomposed by Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN). Subsequently, due to the persistence of high-dimension feature vectors containing numerous outliers that interfere with recognition, the LT-PCA compression and dimensionality reduction method is proposed. Experimental analyses on vehicle gearboxes demonstrate an average fault recognition rate of 96.65% using the newly proposed wavelet noise reduction function and the integrated method. This allows for quick and efficient identification of fault types and provides crucial technical support for related industrial applications.

Keywords: CEEMDAN; wavelet threshold denoising; LT-PCA; SVM; fault diagnosis

1. Introduction

The gearbox is a crucial component in mechanical equipment, and is susceptible to various factors that can lead to failures, ultimately affecting its performance. Ensuring the proper operation of the gearbox necessitates timely detection of the equipment's working status. Presently, vibration signals in gearboxes are employed to detect the condition of bearings and gearbox gears, as outlined in [1]. In the event of internal components' failure within a gearbox, abnormal vibration changes manifest during operation. These changes serve as detectable signals for fault diagnosis. Nevertheless, random noise from the surrounding environment interferes with signal collection and transmission. The accurate and effective removal of noise components while retaining the original useful signal remains a critical research focus.

The wavelet threshold denoising algorithm and empirical mode decomposition denoising algorithm have found extensive application in noise processing for fault diagnosis. Xu et al. [2] proposed a denoising method for ECG signals, combining CEEMDAN and wavelet thresholding to effectively suppress random noise. Chen et al. [3] employed the wavelet thresholding method for initial vibration signal noise reduction. Subsequently, the denoised signal underwent decomposition and reconstruction via CEEMDAN, enhancing the recognition accuracy of rolling bearings with nonstationary vibration signals. Sun et al. [4] utilized CEEMDAN to decompose the cable signal and calculate the correlation. Components with high correlation were reconstructed, and the resulting signal

underwent final wavelet noise reduction for effective noise filtration. Li et al. [5] conducted CEEMDAN decomposition of the original signal, breaking down the intrinsic mode function (IMF) into noise components, noise-dominant components, true signal-dominant components, and true signal components. Discarding the noise-dominant and real signal-dominant components, they reconstructed a noise-canceled signal with the genuine signal component, employing wavelet soft thresholding for noise reduction. Zhou et al. [6] applied CEEMDAN to decompose the blast vibration signal into IMF components, further categorizing them into major and minor components based on correlation coefficients. The denoising process involved fixed and soft thresholding techniques applied to the two groups of IMF. Abdelkader et al. [7] employed CEEMDAN to decompose the rolling bearing vibration signal, screening each IMF energy value, reducing wavelet noise in abnormal IMFs, and subsequently reconstructing the signal.

The results indicate that employing CEEMDAN with wavelet threshold noise reduction significantly reduces signal noise. However, CEEMDAN decomposition yields various intrinsic mode function (IMF) components, each containing distinct types of noise. During the wavelet noise reduction process, the efficacy varies significantly based on influencing factors such as the choice of threshold functions. Therefore, it is crucial to optimize the noise reduction process.

The gearbox is affected by many factors during operation, which generate complex nonlinear vibration signals; therefore, determining how to obtain accurate fault information is particularly important. Deng et al. [8] applied principal component analysis (PCA) to energy and kurtosis values from different intrinsic mode functions (IMFs) for automatic bearing fault identification, demonstrating PCA's effectiveness in early fault detection. Zhu et al. [9] utilized PCA to reduce the dimensionality of vibration signals, extracting fault features for training with the DBN algorithm for identification and classification. Pule et al. [10] combined PCA and SVM with cross-validation to classify bearing faults, proving its suitability for engineering applications. To address residual noise influence and effectively identify subtle fault content within limited vibration signals, this article diverges from traditional PCA direct dimensionality reduction. Instead, it adopts the LT-PCA feature dimensionality reduction method, incorporating data compression and dimensionality reduction in the fault feature classification process as the subsequent step in fault identification.

Fault diagnosis methods encompass model-based approaches [11], signal-based methods [12], and data-driven methods [13]. Model-based methods detect faults via the construction of accurate models; however, challenges arise due to parameter uncertainty factors within the system, impacting the overall model's accuracy and, consequently, fault diagnosis. Traditional signal processing methods involve time-domain, frequency-domain, and time–frequency-domain analyses. Time-domain and frequency-domain analyses may face limitations in identifying non-stationary signals, requiring further analysis. Alternatively, time–frequency-domain analyses are tailored for complex systems, facilitating the identification of vibration signal characteristics and achieving effective vibration data classification and fault detection [14]. Signal classification methods, such as wavelet transform, empirical modal decomposition (EMD), and Fourier transform (FT), can be employed for signal analysis and classification [15,16]. Data-driven fault diagnosis methods, which are not dependent on precise physical models, conduct fault monitoring by analyzing input and output data, making them well-suited for complex, non-linear, and time-varying systems. Common approaches in this category include support vector machines and neural networks [17,18].

Ahmed et al. [19] introduced a fault detection and diagnosis method based on support vector machines (SVMs). This approach involves fault classification through the generation of sampling data from sampled signals and subsequent identification of these sampled signals. While exhibiting superior performance compared to other algorithms, its effectiveness relies heavily on the extracted feature information. Han et al. [20] devised a method that combines EMD and PSO-SVM for fault feature identification. Experimental evidence

supports the algorithm's effectiveness in identifying gear failure types under varying loads. Qin et al. [21] proposed a PSO-SVM hybrid model based on sensitive dimensionless parameters, exhibiting good classification accuracy and efficiency by optimizing kernel parameter settings in SVM training. Dong et al. [22] introduced an intelligent fault diagnosis method for bearings, employing the Grey Wolf Optimization support vector machine. This method outperforms the SVM model in diagnostic effectiveness, although there is room for improvement in the optimization accuracy of GWO.

Therefore, this article introduces a signal-denoising algorithm that optimizes and enhances the wavelet threshold function through simulated annealing (SA), combined with CEEMDAN. This innovative approach addresses data loss issues associated with the over-flatness of traditional soft threshold functions after noise reduction and mitigates pseudo-interference from local oscillations in hard threshold functions during the noise reduction process. Following denoising, the LT-PCA method is implemented to compress and downscale the reconstructed signal's fault features. This strategy effectively minimizes the impact of discrete points within signal features in comparison to conventional downscaling methods. Finally, the Northern Goshawk Optimization (NGO) algorithm is employed to optimize support vector machine (SVM) parameters for the fault detection and classification of gearboxes.

2. Principle of the Method

To effectively reduce the gearbox signal noise and improve the accuracy of fault recognition, this section proposes a gearbox fault recognition method integrating improved wavelet noise reduction, LT-PCA, and NGO-SVM. The method comprises three main components: signal noise reduction processing, signal fault feature extraction processing, and fault recognition. The following presents the detailed procedure of the integrated fault noise reduction and identification method.

2.1. Principle of Wavelet Threshold Denoising

Wavelet threshold denoising is a technique that involves applying noise reduction steps to the collected signal, resulting in a corresponding low-noise signal. By carefully selecting an appropriate threshold value and threshold function, various wavelet bases are employed to adjust the wavelet coefficients. This adjustment ensures the retention of high wavelet coefficients while eliminating the low coefficients. The resultant wavelet coefficients are then reconstructed to obtain low-noise signals, facilitating subsequent analysis. The operational equation for this process is as follows:

$$f(t) = x(t) + e(t) \tag{1}$$

where $x(t)$ is the original signal and $e(t)$ is the noise signal. The steps of wavelet thresholding denoising are as follows:

1. Wavelet decomposition. Conduct wavelet decomposition of the noise-containing signal by choosing an appropriate wavelet base and determining the number of decomposition layers (i layers). This process yields the wavelet coefficients (ca, cd) for the corresponding layers, where ca and cd represent the approximation coefficients and detail coefficients after wavelet transform decomposition.
2. Thresholding. Process the wavelet coefficients by applying appropriate thresholds and threshold functions.
3. Signal reconstruction. Reconstruct the signal by utilizing the processed coefficients.

2.1.1. Optimization and Improvement of Threshold Function in Wavelet Denoising

Conventional threshold functions are of two types, namely soft threshold and hard threshold functions. Hard and soft thresholds are described as follows:

$$w_T = \begin{cases} |w|, & |w| \geq T \\ 0, & |w| < T \end{cases} \tag{2}$$

$$w_T = \begin{cases} sgn(w)(|w|-T), & |w| \geq T \\ 0, & |w| < T \end{cases} \tag{3}$$

where w is the wavelet coefficient of the original signal and w_T is the wavelet coefficient after threshold processing.

The calculation formula of threshold T is as follows [23]:

$$T_i = \frac{\sigma\sqrt{2\ln(n)}}{\ln(i+1)} \tag{4}$$

Here, n is the signal length, σ is the standard deviation of the noise, T_i is the threshold value of the i-th wavelet coefficient, and i is the number of decomposition layers. When the number of decomposition layers i is larger, then threshold T is smaller, which ensures the noise reduction process of the signal.

$$\sigma = \frac{median(|w|)}{0.6745} \tag{5}$$

where w represents the high-frequency coefficient after the i-th wavelet transform, and 0.6745 is the adjustment coefficient of noise standard deviation.

Both threshold functions aim to eliminate wavelet coefficients with small energy values while retaining or reducing those with large energy, offering a straightforward calculation approach. The hard threshold function efficiently filters out considerable weak information, preserving entirely the original information with large energy. However, the utilization of a discontinuous function for setting the threshold leads to pronounced vibrations and jump points in the signal after noise reduction, resulting in inadequate smooth connections between signal points. In contrast, the signal processed by the soft threshold function avoids interruptions and overcomes the shortcomings of the hard threshold. Nevertheless, it introduces a certain deviation in the signal energy across the entire signal scale [24]. Some useful signals are lost in this process, directly impacting the accuracy of the reconstructed signal.

To harness the benefits of both hard and soft threshold functions, researchers have proposed an improved threshold function. Experimental verification has shown that this enhanced threshold function effectively removes noise while retaining useful components in the signal, optimizing the denoising outcome.

Wu Y et al. [25] introduced an improvement to the threshold function by incorporating an exponential function based on the soft threshold function. This modification aims to enhance the calculation speed. However, challenges persist in terms of signal compression and deviation caused by the soft threshold function, leading to the direct discarding of wavelet coefficients smaller than the threshold value. Consequently, the retention of low-frequency signals becomes limited. The expression of the threshold function is given as follows:

$$F_T = \begin{cases} sgn(w)\left(|w| - Tb^{a(T-|w|)}\right), & |w| > T \\ 0, |w| \leq T \end{cases} \tag{6}$$

Wu F et al. [26] proposed a relatively smooth threshold function, aiming to integrate the features of both soft and hard threshold functions comprehensively. While this approach leverages the advantages of both functions, it comes with a limitation in flexibility when adjusting the signal. The expression of the threshold function is given as follows:

$$F_T = \begin{cases} w - \frac{T^2}{3w} - \frac{sgn(w)}{3e^{(|w|-T)}}, & |w| \geq T \\ \frac{w}{3}e^{(|w|-T)}, & |w| < T \end{cases} \tag{7}$$

Combining the strengths gleaned from various threshold functions, this study introduces a novel and improved threshold function designed to enhance noise reduction effectiveness. The expression of the improved threshold function is as follows:

$$F_T = \begin{cases} sgn(w)\left(|w| - \dfrac{aT}{e^{(|w|-T)(1-a)}}\right), & |w| \geq T \\ sgn(w)\dfrac{1}{2}(1-a)|w|e^{(|w|-T)^3}, & |w| < T \end{cases} \tag{8}$$

where a is the adjustment parameter, w is the wavelet coefficient of the original signal, T is the set threshold, and F_T is the wavelet coefficient of the noise reduction signal.

Equation (8) approximates the hard threshold function as a tends to 0 and the soft threshold function as a tends to 1. Therefore, the adjustment parameter a can be specifically taken in the range of the (0, 1) interval. The properties of the threshold function are examined in terms of parity, continuity, and asymptote, as demonstrated below:

(1) Proving the parity of the function

When $0 < a < 1$, $F(-w) = -F(w)$, indicating that the function is odd.

(2) Proving the continuity of the function

When w approaches T^+:

$$\lim_{w \to T^+} w = \lim_{w \to T^+} sgn(w)\left(|w| - \frac{aT}{e^{(|w|-T)(1-a)}}\right) = \lim_{w \to T^+}\left(|w| - \frac{aT}{e^{(|w|-T)(1-a)}}\right) = (1-a)T \tag{9}$$

When w approaches T^-:

$$\lim_{w \to T^-} w = \lim_{w \to T^-} sgn(w)(1-a)|w|e^{(|w|-T)} = \lim_{w \to T^-}(1-a)|w| = (1-a)T \tag{10}$$

From Equations (9) and (10), $\lim\limits_{w \to T^+} w = \lim\limits_{w \to T^-} w = (1-a)T$, proving that the function is continuous at T. Similarly, the function is also continuous at $-T$. This threshold function addresses the issue where the hard threshold function causes oscillations in the reconstructed signal due to sudden signal changes at a fixed threshold.

(3) Proving the asymptote of the function

When w approaches $+\infty$:

$$\lim_{w \to +\infty} \frac{F}{w} = \frac{sgn(w)\left(|w| - \frac{aT}{e^{(|w|-T)(1-a)}}\right)}{w} = 1 - \frac{aT}{we^{(|w|-T)(1-a)}} = 1 \tag{11}$$

$$\lim_{w \to +\infty}(F - w) = \lim_{w \to \infty}\frac{aT}{e^{(|w|-T)(1-a)}} = 0 \tag{12}$$

Similarly, when w approaches $-\infty$:

$$\lim_{w \to -\infty}\frac{F}{w} = 1 \tag{13}$$

$$\lim_{w \to -\infty}(F - w) = 0 \tag{14}$$

The asymptote of F is w. As w tends to ∞, F tends to w, which effectively addresses the deficiency of deviation observed in the soft threshold function.

In summary, the demonstrated proof establishes that the improved threshold function maintains continuity across the entire real number line $(-\infty, +\infty)$. This adherence to the general rule of wavelet threshold function optimization signifies excellent flexibility and usability. The improved threshold function enhances wavelet processing at critical thresholds compared to Equation (6) and introduces the adjustment parameter a for increased flexibility in threshold processing compared to Equation (7). Figure 1 displays the traditional soft and hard threshold functions and three improved threshold function images.
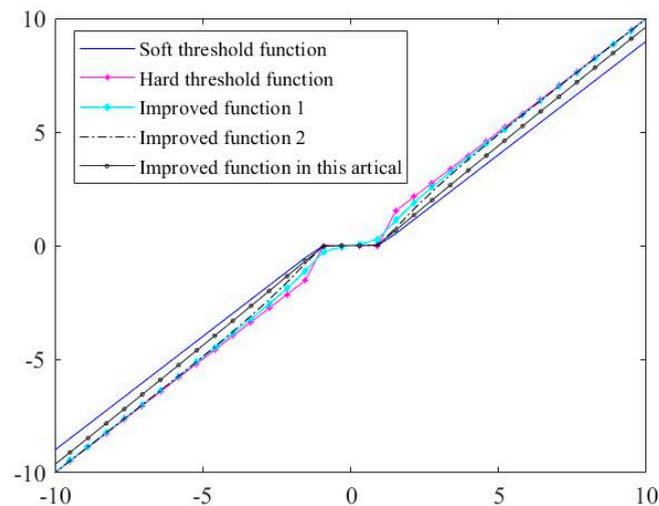
Figure 1. Threshold function curve.

Before optimizing the wavelet coefficient threshold, it is essential to adjust parameter *a* to achieve the optimal threshold function. Following the principles of wavelet noise reduction, a larger threshold than the fixed threshold results in obtaining more useful signals. In this case, parameter *a* needs to be adjusted to bring the function closer to the hard threshold function, thereby minimizing energy loss. Conversely, when the threshold is close to the fixed threshold, noise signals become more apparent, and the soft threshold function exhibits superior denoising and smoothness effects. Therefore, adjusting parameter *a* becomes necessary to align the function closer to the soft threshold function. To determine the optimal *a* for the proposed improved threshold function, the simulated annealing (SA) method can be employed.

2.1.2. Noise Reduction Method

To accurately extract fault features, a noise reduction method leveraging CEEMDAN and wavelet techniques has been proposed. This method effectively suppresses residual noise and false components, thereby accentuating fault characteristic frequencies. The specific steps are as follows:

Step 1: Perform CEEMDAN decomposition on the acquired signal to obtain multiple intrinsic mode function (IMF) components ordered from high to low. Calculate the correlation coefficients for each component.

Step 2: Determine the optimal value of the improved threshold function using SA.

Step 3: Apply wavelet noise reduction to the IMF components with larger correlation coefficients, and reconstruct the signal after noise reduction.

Step 4: Verify the signal after noise reduction using evaluation metrics. Alternatively, conduct energy spectrum analysis to extract the fault characteristic frequency.

Figure 2 presents a visual representation of the noise reduction flowchart.
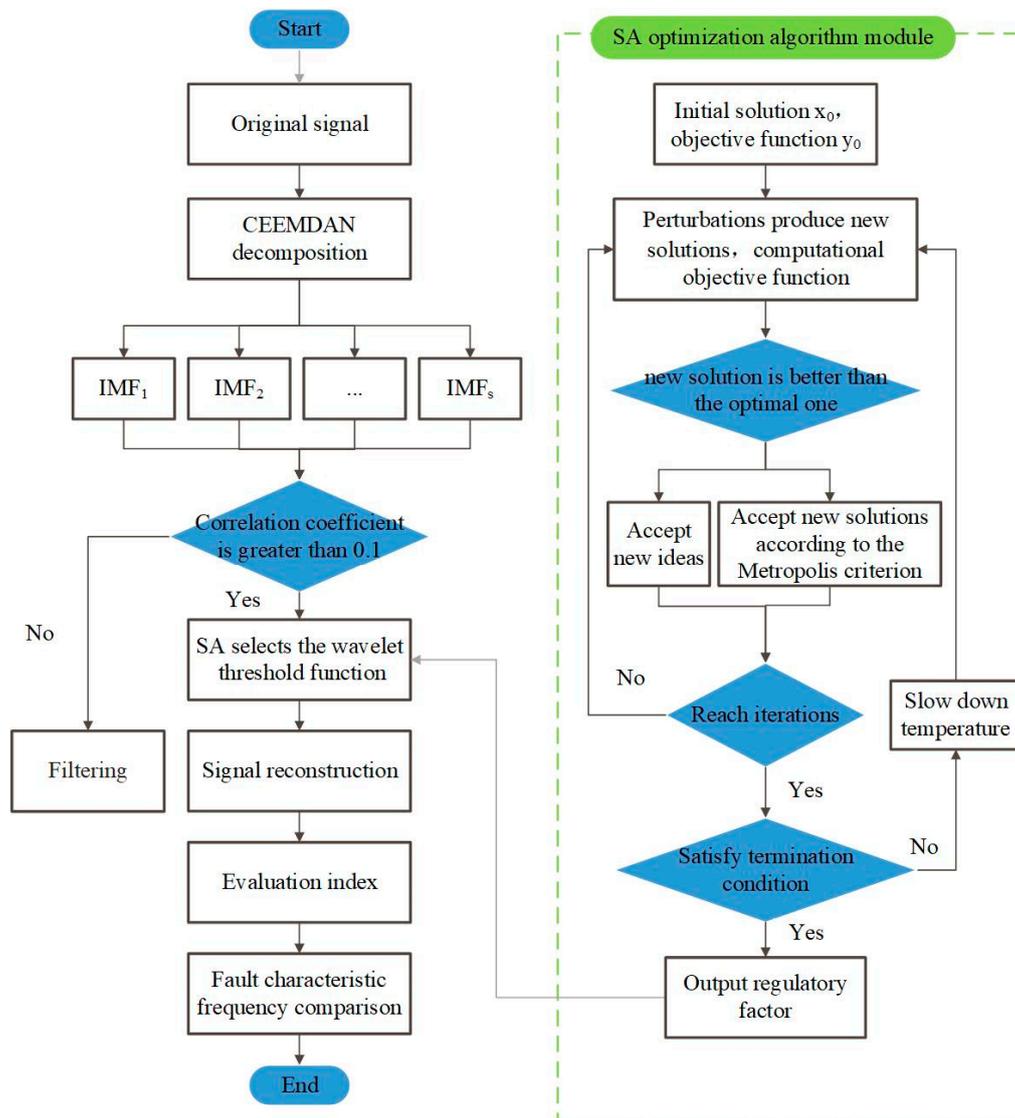
Figure 2. Noise reduction flowchart.

2.1.3. Simulation Experiment and Result Analysis

After the noise reduction process, signals containing noise can be denoised, and commonly selected evaluation indexes for denoising effect include the signal-to-noise ratio (SNR) and root mean square error (RMSE). The denoising effect is considered superior when SNR is larger and RMSE is smaller. SNR and RMSE are defined as follows:

$$SNR = 10\log\left(\sum_{i=1}^{N}\frac{S_i^2}{(Z_i - S_i)^2}\right) \tag{15}$$

$$RMSE = \frac{1}{N}\sum_{i=1}^{N}(Z_i - S_i) \tag{16}$$

Here, N is the signal length, S_i is the original signal, and Z_i is the denoised signal.

The simulated annealing experiment with an improved threshold function is conducted to obtain the optimal value. The simulation noise reduction experiments involve adding Gaussian white noise to the bump signals using MATLAB R2023b, employing a sym6 wavelet base, and setting five decomposition layers for wavelet decomposition. Test-

ing involves the use of the soft threshold function, hard threshold function, two improved thresholds from the literature [25,26], as well as the improved threshold function.

The variable *a* in the threshold function is initially explored with three values ($a = 0.1$, $a = 0.5$, $a = 0.9$) in Equation (8) for noise reduction. Given the randomness in noise addition, the subsequent results represent the average of five noise reduction attempts to mitigate errors. Table 1 summarizes the obtained signal-to-noise ratio (SNR) and root mean square error (RMSE) values. It is observed that different values lead to varied results, lacking clear regularity. To refine the selection of *a*, the simulated annealing algorithm is introduced, determining the optimal value as 0.7124 for improved threshold noise reduction. Figure 3 displays the simulated annealing iteration curve.

Table 1. SNR and RMSE values for various *a* parameters.

a	SNR/dB	RMSE
$a = 0.1$	22.61	0.133
$a = 0.5$	24.52	0.106
$a = 0.7124$	25.15	0.099
$a = 0.9$	23.78	0.116
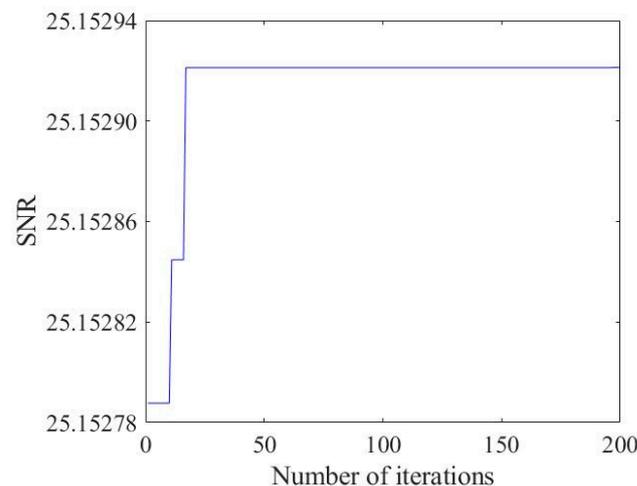
Figure 3. Simulated annealing iteration curves.

In Table 1, the *a* value obtained by simulated annealing exhibits the largest SNR and the most effective noise reduction compared to other *a* values. Noise reduction analysis is performed using soft thresholding, hard thresholding, and improved thresholding from the literature [25]. Figure 4 displays improved thresholding from the literature [26], the improved thresholding function, and the signal after noise reduction. Table 2 presents the evaluation index calculation results of various threshold functions for processing bump signals.

The signal graph in Figure 4 highlights that, when compared with the conventional soft threshold, hard threshold, and improved thresholds from the literature [25,26], the proposed improved threshold function preserves better continuity in the signal after noise reduction and retains local features. The computational results in Table 2 show that results achieved by the improved method surpass those of other threshold functions, while the RMSE is smaller. Consequently, the improved method exhibits superior noise reduction effectiveness.
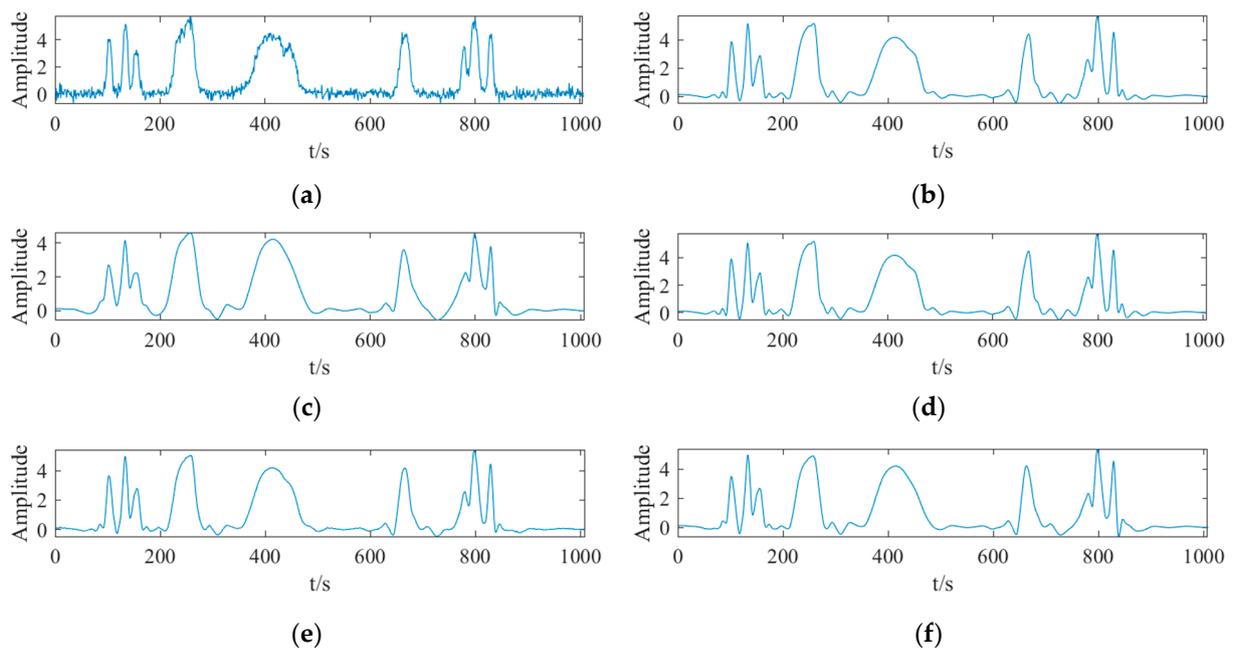
Figure 4. Comparison of noise reduction with various threshold functions. (**a**) Noisy signal. (**b**) Signal after improved threshold denoising. (**c**) Signal after soft threshold denoising. (**d**) Signal after hard threshold denoising. (**e**) Signal after improved threshold denoising in the literature [25]. (**f**) Signal after improved threshold denoising in the literature [26].

Table 2. SNR and RMSE values after noise reduction.

Classifications	SNR/dB	RMSE
Soft threshold denoising	23.07	0.126
Hard threshold denoising	23.25	0.123
Thresholding denoising in literature [25]	23.84	0.115
Thresholding denoising in literature [26]	24.32	0.109
Threshold denoising in this article	25.15	0.099

Having established the effectiveness of simulated annealing in the aforementioned analyses, practical tests were conducted using real extracted signals. When obtaining the actual signal in engineering training, the measurement sensor detects not only its own vibration signal but also considerable interference noise. Therefore, in simulated signal noise reduction tests, it becomes crucial to introduce irregular noise signals into the constructed multi-harmonic vibration simulation signal $f(t)$.

$$f(t) = \cos(2\pi f_1 t) + \frac{1}{2}\cos(2\pi f_2 t) + e(t) \tag{17}$$

where f_1 and f_2 represent the primary frequencies of the simulated signals, $f_1 = 2$ Hz, $f_2 = 5$ Hz, and $e(t)$ is the irregular noise signal.

To simulate testing conditions, a random noise of 18 dB is added to the signal $f(t)$, resulting in the simulated signal for testing. Subsequently, the signal undergoes CEEMDAN decomposition, and the decomposed IMF components are illustrated in Figure 5.
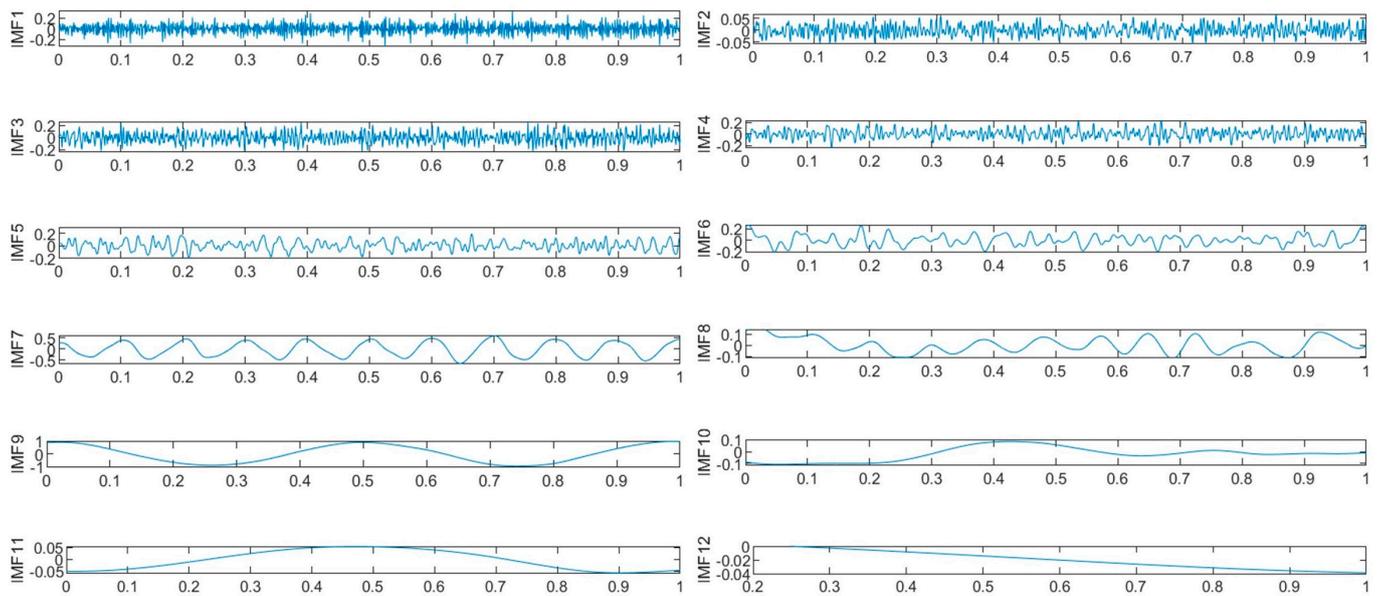
Figure 5. Decomposition result of the simulated signal using the CEEMDAN method.

Figure 5 illustrates the outcome of decomposing the simulated signal using the CEEMDAN method. In this representation, the intrinsic mode function (IMF) components are arranged in order from high frequency to low frequency, representing the frequency components inherent in the original signal. The IMF components derived from the analog signal are subjected to correlation coefficient analysis and organized based on Equation (18). Table 3 presents the correlation coefficient values, aiding in the selection of IMF components that retain more original information for subsequent noise reduction and reconstruction.

Table 3. IMF correlation coefficient value of the simulated signal.

IMF Component	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6
Correlation coefficient	0.105	0.083	0.018	0.019	0.019	0.305
IMF component	IMF7	IMF8	IMF9	IMF10	IMF11	IMF12
Correlation coefficient	0.434	0.564	0.875	0.573	0.081	0.002

The correlation coefficient formula is as follows:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X \sigma_Y} = \frac{E(XY)-E(X)E(Y)}{\sqrt{E(X^2)-E^2(X)}\sqrt{E(Y^2)-E^2(Y)}} \tag{18}$$

where X and Y represent two signals, $E(\cdot)$ is the mathematical expectation of the signal, *cov* is the covariance, and σ is the variance.

Based on the given equation, the IMF components with correlation coefficients greater than 0.1 are selected for the sym8 wavelet basis and five-layer wavelet decomposition. Other correlation coefficients of IMF are disregarded, and the remaining components are reconstructed to yield the final noise-reduced signal. Figure 6 displays the noise reduction signal.
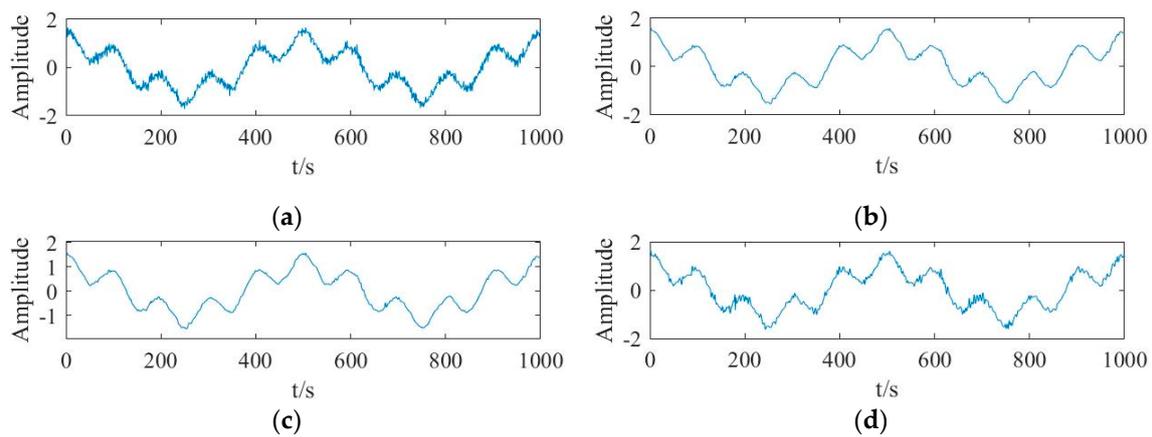
Figure 6. Comparison of the original noisy signal with the denoised signal. (**a**) Noisy signal. (**b**) Signal after CEEMDAN improved threshold denoising. (**c**) Signal after CEEMDAN soft threshold denoising. (**d**) Signal after CEEMDAN hard threshold denoising.

The noise reduction signal images in Figure 6 appear relatively close overall, which is attributed to the simplicity of the initial signal. However, when dealing with noise reduction in complex signal images, comparing the noise reduction effect requires computing signal-to-noise ratio (SNR) and root mean square error (RMSE) for several signals after noise reduction. Table 4 provides the SNR and RMSE values of several signals after noise reduction.

Table 4. SNR and RMSE values after noise reduction using CEEMDAN and various improved threshold functions.

Classifications	SNR/dB	RMSE
Soft threshold denoising	17.0117	0.1119
Hard threshold denoising	18.7860	0.1252
Threshold denoising in this article	19.6544	0.0912

According to the comprehensive analysis of Figure 6 and Table 4, SNR after adopting a hard threshold lies between that of a soft threshold and that of an improved threshold, indicating a general noise reduction effect. The soft threshold demonstrates superior smoothness. However, the analysis of SNR for various noise reduction signals reveals that soft threshold noise reduction exhibits a certain deviation, leading to more loss of effective information in the signal after noise reduction. Compared with the soft threshold function, the SNR of the improved wavelet is increased by 2.65 and the RMSE is reduced by 0.02. Compared with the hard threshold function, the SNR is increased by 0.87 and the RMSE is reduced by 0.03. These results indicate that the improved wavelet method is closer to the pure signal, and the noise reduction effect is superior. The noise reduction effect of this approach surpasses that of the conventional threshold function algorithm.

2.2. Dimensionality Reduction of Fault Feature Construction

Evaluating the gear fault state directly from noise-reduced vibration signals is challenging. In this context, time-domain features and frequency-domain feature parameters are initially selected. However, the abundance of feature parameters can potentially prolong subsequent identification and classification processes. To address this issue, a newly proposed dimensionality reduction method will be implemented to streamline and reduce the dimensionality of the fault information.

2.2.1. Fault Feature Vector Extraction

(1) Mixed domain feature

Time-domain statistical metrics are classified into dimensional and dimensionless types. Dimensional characteristic metrics usually include specific physical quantities, such as the amplitude and units of the signal. Dimensionless characteristic metrics provide a more abstract and generic way to describe the nature of the signal. Frequency domain analysis reveals the frequency characteristics of a signal by decomposing it into components of different frequencies. These metrics are used not only to describe the frequency distribution of the signal but also to assist in identifying significant frequency components and trends in the frequency domain. Therefore, in the time domain, dimensional characteristic indices such as mean value, absolute mean value, peak-to-peak value, and kurtosis are applied, along with dimensionless characteristic indices such as pulse factor and waveform factor. In the frequency domain, statistical indices such as the mean value of the spectrum amplitude sample, spectrum amplitude sample variance, and spectrum amplitude sample kurtosis are utilized.

(2) Wavelet domain features

To extract wavelet domain features, a multi-layer wavelet packet decomposition is performed on the signal. The resulting decomposition produces $L = 2^l$ frequency bands, where l represents the number of decomposition layers. The calculations of energy for each frequency band and the subsequent calculation of total energy (E) are conducted as follows:

$$E = \sum_{i=1}^{L} E_i \tag{19}$$

The energy entropy of the wavelet packet is as follows.

$$Energy = -\sum_{i=1}^{L} \left(\frac{E_i}{E} * \log \left| \frac{E_i}{E} \right| \right) \tag{20}$$

Wavelet packet singular spectrum entropy (Qiyi) is employed as an index to gauge the singularity or irregularity of the signal. It is calculated as follows:

$$Qiyi = -\sum_{i=1}^{L} \left(\frac{\lambda_i}{\sum_{i=1}^{L} \lambda_i} * \log \left| \frac{\lambda_i}{\sum_{i=1}^{L} \lambda_i} \right| \right) \tag{21}$$

where, λ_i represents the singular values obtained through singular value decomposition after processing the original signal into a matrix.

The singularity of a signal is assessed through wavelet packet singular entropy, which analyzes the distribution of coefficients in the wavelet packet transform. A lower value of singular entropy signifies a more uniform distribution of coefficients, indicating that the signal is regular and smooth. Conversely, a higher singular entropy value suggests an uneven distribution of coefficients, implying the presence of more irregular or singular components in the signal. Consequently, wavelet packet singular entropy is commonly employed to identify bursts, noise, or abnormal behavior in a signal.

2.2.2. Log-Transformed Principal Component Analysis Dimensionality Reduction

Principal component analysis (PCA) is a widely used technique for dimensionality reduction [27]. Its core idea involves mapping high-dimensional data to a low-dimensional space through a linear transformation while retaining as much variance in the data as possible. Dimensionality reduction is achieved by finding the main components of the data and projecting them into a new coordinate system. PCA effectively mitigates information loss by downscaling the data using the feature matrix [28]. However, outliers with large variance values can impact the data's variance during downscaling, influencing the direction of each principal component of PCA. Simply removing outliers may result in the loss of valuable

sample information. To address this, we propose a new dimensionality reduction method in this section—logarithmic transformation principal component analysis (LT-PCA).

To stabilize fluctuations in the dataset and maintain as constant a variance as possible, logarithmic transformation is chosen for data processing. This transformation compresses the data, reduces fluctuations, and eliminates large differences in orders of magnitude.

We define the sample as Z_t, the sample mean $E(Z_t) = \mu_t$, and set the distribution standard deviation to be linearly correlated with the mean; then, $\sqrt{Var(Z_t)} = \mu_t \times \sigma$, where σ is the correlation coefficient.

When x is sufficiently small, $\log(1 + x) \approx x$, leading to the following formula:

$$\log(Z_t) \approx \log(\mu_t) + \frac{Z_t - \mu_t}{\mu_t} \tag{22}$$

Furthermore, $E\left(\frac{Z_t - \mu_t}{\mu_t}\right) = 0$, $E(\log(Z_t)) \approx \log(\mu_t)$, as demonstrated in the following derivation process:

$$
\begin{aligned}
Var[\log(Z_t)] \quad &= E[\log(Z_t) - E(\log(Z_t))]^2 = E[\log(Z_t) - \log(\mu_t)]^2 \\
&\approx E\left(\frac{Z_t - \mu_t}{\mu_t}\right)^2 = \frac{1}{(\mu_t)^2} E(Z_t - E(Z_t))^2 \\
&= \frac{1}{(\mu_t)^2} Var(Z_t) \approx \sigma^2
\end{aligned} \tag{23}
$$

To summarize, $E(\log(Z_t)) \approx \log(\mu_t)$, $Var[\log(Z_t)] \approx \sigma^2$. Logarithmic transformations can effectively convert a model prone to significant fluctuations with increasing independent variables into a more stable model. Assuming D is a dataset with m samples and n variables, the transformation matrix is defined by the formula:

$$P = \log_a D \tag{24}$$

In the formula, P represents the logarithmic transformation form of D. Compared with D, P offers greater stability and is used to replace D in PCA modeling.

Taking the logarithm of sample data has the effect of decreasing the difference between large and small values on a logarithmic scale, thereby reducing data dispersion. When the variance of the data is associated with the mean, the logarithmic transformation stabilizes the data, creating a more homogeneous relationship between variance and mean. This proves beneficial for specific data analysis and modeling tasks, aiding in the reduction of heteroskedasticity and facilitating easier data processing and analysis. It is essential to note that the logarithmic transformation may not be suitable for all datasets, especially those containing negative numbers or zeros.

The differences between the LT-PCA method and the PCA method are exemplified by the coordinate images below.

From the left panel of Figure 7, it is evident that outliers with large difference values significantly affect the principal component direction. Although these outliers can be addressed during dimensionality reduction, their noticeable deviation from the main data direction contradicts the requirement for effective data dimensionality reduction. However, with LT-PCA, the first principal component direction can encapsulate most of the data information, markedly mitigating the impact of outlier points. Therefore, the LT-PCA method proposed in this section proves to be more realistic than traditional PCA dimensionality reduction processing, particularly in handling discrete points.
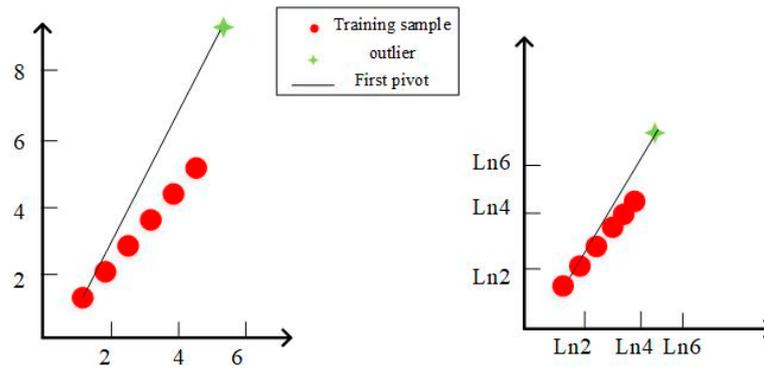
Figure 7. Schematic diagram of LT-PCA.

2.3. Fault Classification and Recognition Research

Support vector machine (SVM) serves as a linear classifier rooted in a supervised learning approach. It transforms the problem into solving convex quadratic programming. The core concept involves mapping sample data into a high-dimensional feature space through a kernel function. The objective is to identify an optimal hyperplane in this space, maximizing the classification interval between different categories and achieving effective linear classification [29]. The key is to find a hyperplane that can better distinguish between different categories and ensure its superior performance in the high-dimensional space, to improve the accuracy of classification. The principle of SVM is shown in Figure 8, where the red circles and green asterisks indicate two categories of samples. L is the optimal classification surface, and L_1 L_2 represents the hyperplane of the nearest two categories of samples from the optimal surface, parallel to it. These hyperplanes serve as the positive and negative boundaries of SVM. The distance between L_1 and L_2 constitutes the classification interval, and the points located on the positive and negative boundaries are referred to as support vectors.
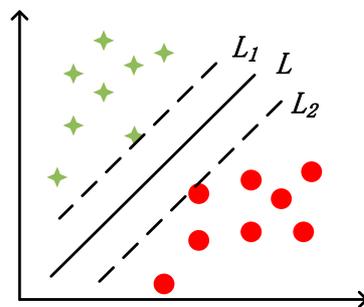
Figure 8. Schematic diagram of SVM.

Utilizing SVM allows training to identify and classify feature vectors after dimensionality reduction, and the machine can be optimized using intelligent algorithms to enhance identification accuracy. One such optimization approach is Northern Goshawk Optimization (NGO), a heuristic optimization algorithm inspired by the hunting behavior of hawks in nature [30]. NGO serves as a meta-heuristic optimization algorithm applicable to solving various optimization problems. Consequently, the parameters of SVM can be effectively optimized using this algorithm.

3. Construction and Analysis of the Gearbox Vibration Test Bench

In this study, the gearbox of a new energy vehicle is used for engineering practical experiments. Figure 9 displays the gearbox vibration test bench, where the vibration acceleration signal during gearbox operation is captured by a piezoelectric acceleration sensor installed on the gearbox housing. The installation position of the acceleration sensor [31] on the gearbox housing is displayed in Figure 10.
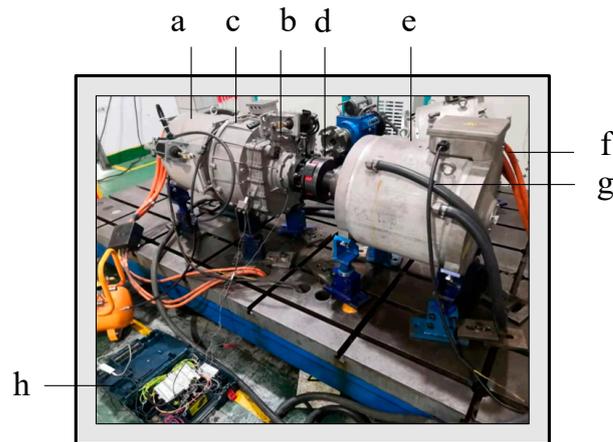
Figure 9. Gearbox vibration test bench device for a vehicle. (a—Drive motor; b—Sensor; c—transmission; d—coupling; e—load motor; f—Three-phase line; g—Cooling water pipe; h—Collection toolbox).

Figure 10. The mounting position of the accelerometer.

Initially, the normal gearbox platform serves as the driving end to provide forward speed, with a single motor as the load delivering forward torque. This setup simulates the working condition of the gearbox driving in the forward direction. The gearbox is then set to first gear with a torque of 300 Nm at the load end. Vibration data are collected at three speed points: 1200 rpm, 1800 rpm, and 2400 rpm, with a minimum of 20 datasets collected at each speed point to mitigate the influence of accidental factors. The collected samples have a length of 10 s, a maximum frequency of effective analysis set at 5000 Hz, and a resolution of 0.1 Hz, and 80 groups of data are extracted for further use. Following the collection of vibration data in the normal state, samples representing the tooth surface of light, moderate, and severe failure states are gathered, as shown in Figure 11. This set of input samples is then used to conduct data feature extraction and dimensionality reduction experiments.

Figure 11. Faulty gear condition. (**a**) Minor tooth failure state. (**b**) Moderate tooth failure state. (**c**) Severe tooth failure state.

In planetary gear systems, characteristic frequency analysis becomes more intricate compared to fixed-axis gears due to the simultaneous rotation and revolution of planetary gears. Taking a single-row planetary gear as an example, the tooth ring is fixed, the sun wheel serves as the power input, the planetary frame acts as the output, and the gearbox functions as the reducer. The number of teeth for the sun wheel, planetary frame, and gear ring is denoted as Z_s, Z_p, and Z_r, respectively. The input speed is represented by n_s, and only one sun wheel, one planetary frame, and one gear ring are included. The fault characteristic frequency of the main components of the planetary gear train is shown as follows:

The fault characteristic frequency of the sun gear is as follows:

$$f_s = N * \frac{n_s}{60} * \left(\frac{Z_r}{Z_r + Z_s} \right) \tag{25}$$

The fault characteristic frequency of the planet gear is as follows:

$$f_p = \frac{n_s}{60} * \left(\frac{Z_r}{Z_r + Z_s} \right) * \left(\frac{Z_s}{Z_p} \right) \tag{26}$$

The fault characteristic frequency of the ring gear is as follows:

$$f_r = N * \frac{n_s}{60} * \left(\frac{Z_r}{Z_r + Z_s} \right) * \left(\frac{Z_s}{Z_r} \right) \tag{27}$$

The fault characteristic frequency of the planetary carrier is as follows:

$$f_p = \frac{n_s}{60} * \left(\frac{Z_s}{Z_r + Z_s} \right) \tag{28}$$

The characteristic frequency of meshing faults in the planetary gear train is as follows:

$$f_s = \frac{n_s}{60} * \left(\frac{Z_r * Z_s}{Z_r + Z_s} \right) \tag{29}$$

The parameters of the new energy vehicle gearbox used in the bench experiment are shown in Table 5.

Table 5. Planetary gear train parameter settings.

Parameter	Sun Gear	Planet Gear	Ring Gear
Number of teeth	32	31	92
Module (mm)	2.5	2.5	2.5
Pressure angle (°)	20	20	20
Helix angle (°)	0	0	0
Addendum coefficient	1	1	1
Dedendum coefficient	0.25	0.25	0.25
Profile Shift coefficient	0	0	1.112

For speeds of 1200, 1800, and 2400 rpm, the fault characteristic frequency is calculated using Formulas (25) to (29), as shown in Table 6.

Table 6. Fault characteristic frequency of each component under different input speeds.

Input Speed (RPM)	Input Frequency (Hz)	Meshing Frequency (Hz)	Sun Gear (Hz)	Planet Gear (Hz)	Ring Gear (Hz)	Planet Carrier (Hz)
1200	20	474.84	59.35	15.32	5.16	20.65
1800	30	712.26	89.03	22.98	7.74	30.97
2400	40	949.68	118.71	30.63	10.32	41.29

4. Experimental Analysis

The algorithm proposed in this article involves the LT-PCA dimensionality reduction optimization of input variables. Additionally, NGO is introduced to optimize the penalty factor 'c' and kernel parameter 'g' in the SVM model [32], aiming to obtain the optimal parameters for improved classification prediction. The method flow is shown in Figure 12.
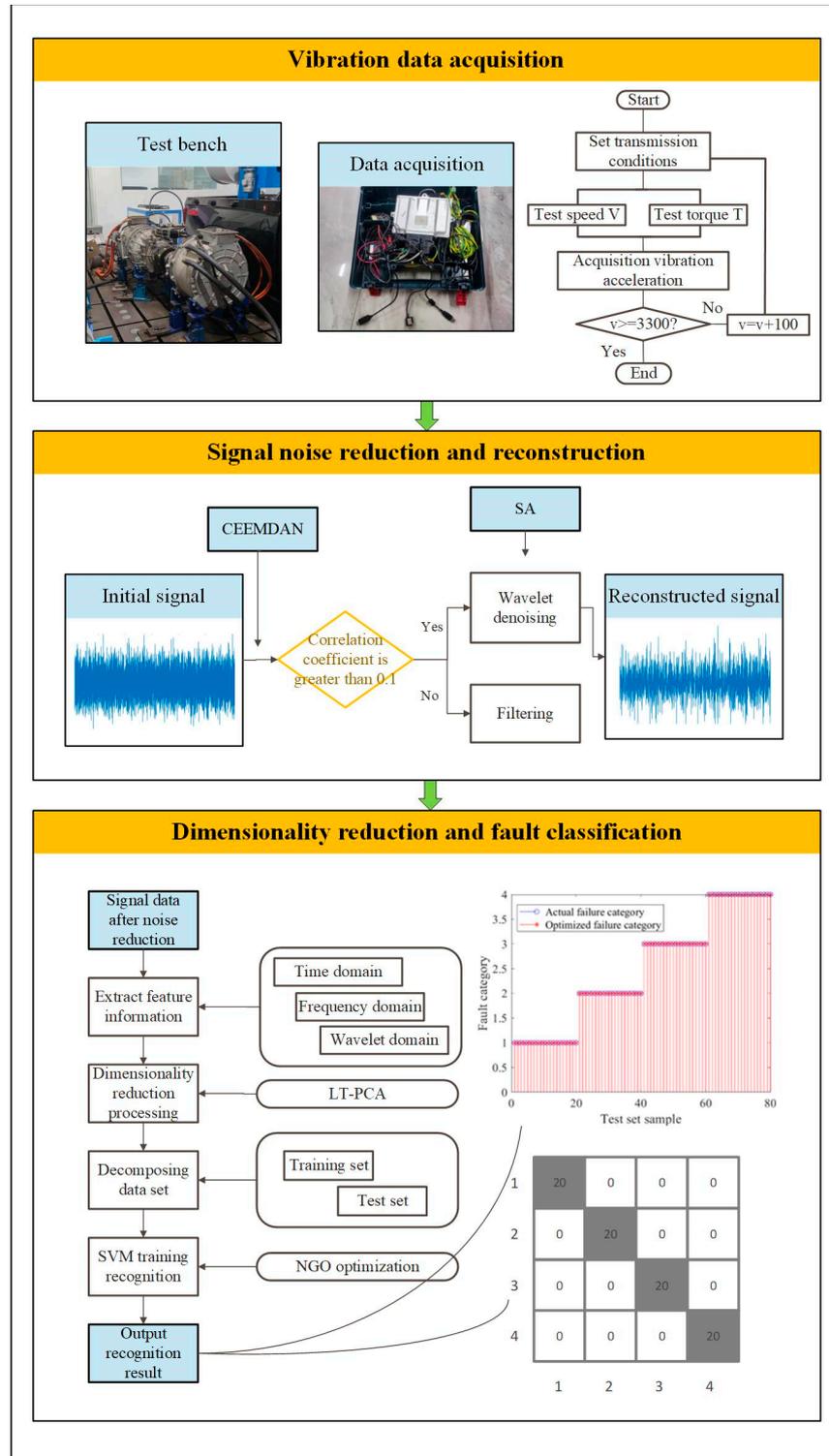
Figure 12. Troubleshooting flowchart.

4.1. Signal Denoising

Because of the influence of the experimental environment and other factors, the vibration data contain significant noise, making the calculation of the SNR challenging. The fault form can be analyzed according to the time domain and energy spectrum of the signal. The theoretical fault characteristic frequency of the planetary gear single tooth is 30.6 Hz, and its frequency is obtained through the calculation of gearbox parameters. As the motor speed is not stable, accurately determining the theoretical frequency and its similar frequencies is crucial for subsequent data processing and analysis.

The signal with CEEMDAN decomposition is denoised by wavelet threshold. The IMF component with a correlation coefficient greater than 0.1 was decomposed by sym8 wavelet and five-layer wavelet, and the IMF component after noise reduction was reconstructed. Figure 13 displays the time-domain diagram of the IMF component obtained by the CEEMDAN decomposition of the gear fault signal, and Table 7 presents the calculated correlation coefficients. Figure 14 displays the original gear fault signal and the signal with or without CEEMDAN decomposition.
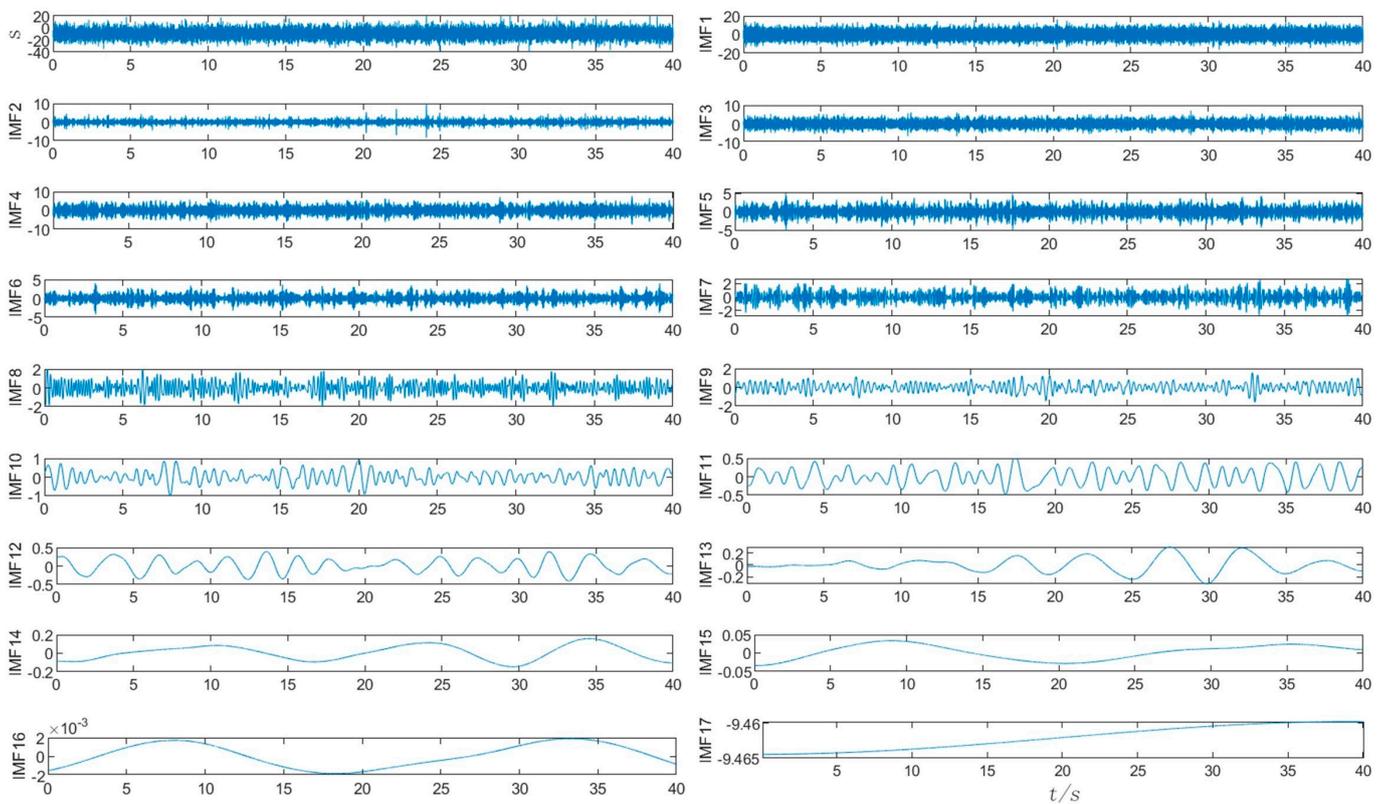
Figure 13. Decomposition result of gear failure signals using the CEEMDAN method.

Table 7. IMF correlation coefficient value of the noisy signal.

IMF Component	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	IMF9
Correlation coefficient	0.766	0.566	0.541	0.459	0.318	0.231	0.172	0.130	0.090
IMF component	IMF10	IMF11	IMF12	IMF13	IMF14	IMF15	IMF16	IMF17	
Correlation coefficient	0.074	0.056	0.035	0.018	0.012	0.014	0.009	0.002	
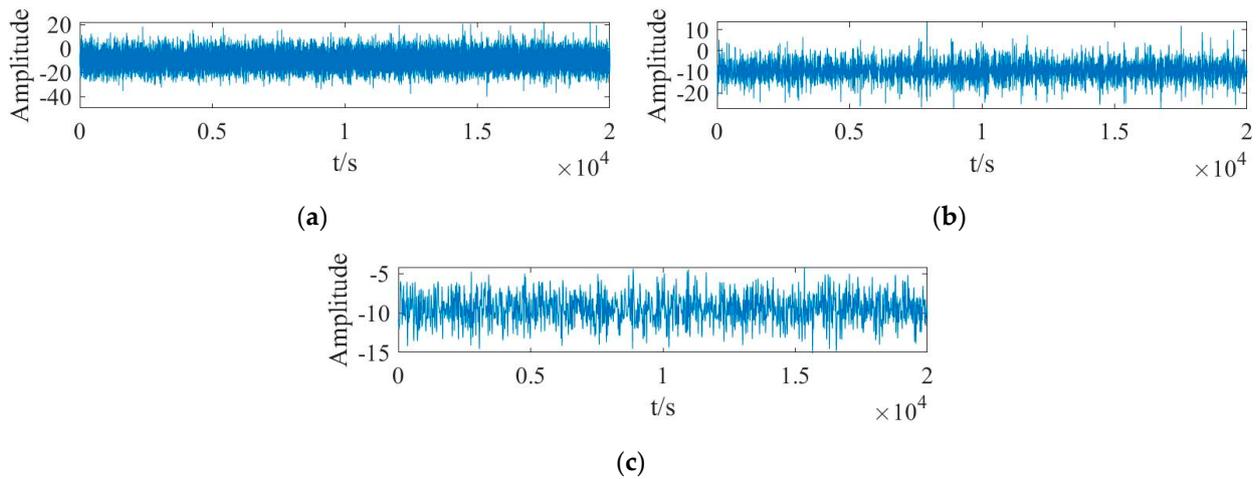
Figure 14. Gear failure signal comparison. (**a**) Noisy signal. (**b**) Denoising signal using CEEMDAN and improved thresholds. (**c**) Denoising signal with an improved threshold is used without CEEMDAN.

Figure 14 reveals numerous burr impact signals generated by gear meshing in the signal, and the fault form cannot be identified by time-domain analysis at this stage. Because the measured signal lacks an original pure signal, indicators such as SNR cannot be used to determine noise reduction results. The energy spectrum of the signal is selected as the analysis method, and the energy spectrum transformation analysis of the signal is performed [33], as displayed in Figure 15.
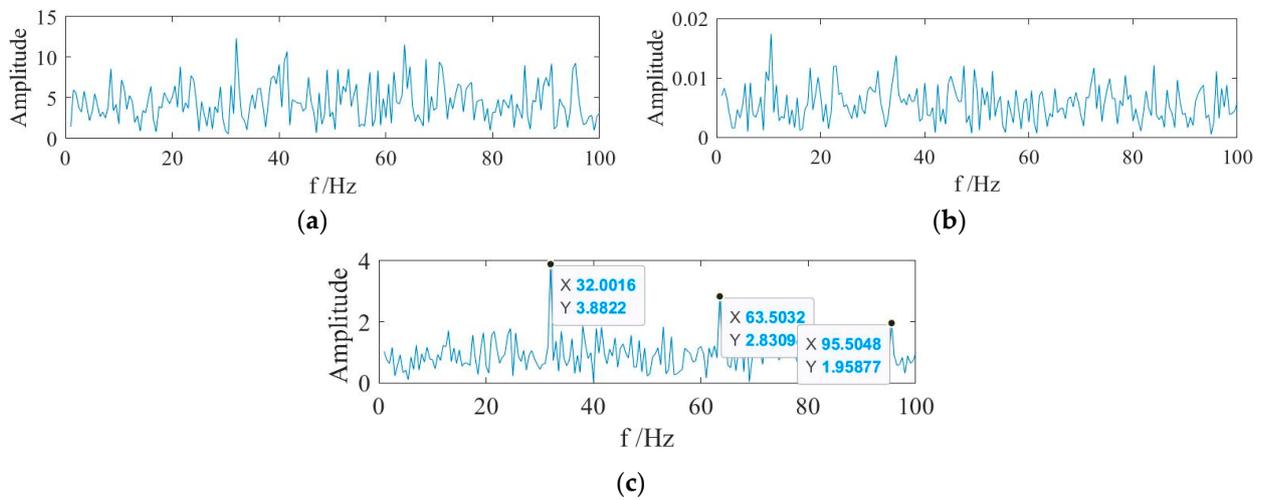
Figure 15. Gear failure signal energy spectrum. (**a**) The energy spectrum of noisy signals. (**b**) The energy spectrum after noise reduction is decomposed without using CEEMDAN. (**c**) The energy spectrum after noise reduction is decomposed using CEEMDAN.

The Teager–Kaiser energy operator, a nonlinear operator, efficiently extracts signal energy by enhancing transient characteristics to highlight shock components. The continuous-time signal $x(t)$ us defined as follows [34]:

$$\psi[x(t)] = [\dot{x}(t)]^2 - x(t)\ddot{x}(t) \tag{30}$$

where $\dot{x}(t)$ and $\ddot{x}(t)$ are the first and second order derivatives, respectively, of $x(t)$. For discrete signals, it is defined as:

$$\psi[x(n)] = x^2(n) - x(n+1)x(n-1) \tag{31}$$

According to Equation (31), the signal energy is calculated using three consecutive sampling points, which has a small computational effort. It exhibits better adaptability to the non-stationary and transient characteristics of the signal.

The energy spectrum of the signal with noise in Figure 15a,b reveals that the fault frequency is hidden in the noise, and the fault information cannot be accurately identified. However, direct wavelet denoising fails to clearly identify the characteristic frequency, leaving the actual fault frequency obscured and resulting in a less-than-ideal denoising effect. In Figure 15c, the proposed method's decomposition and noise reduction unveil the characteristic fault rotation frequency and the frequency doubling of the planetary wheel, which can be extracted as follows: 32.0, 63.5, and 95.5 Hz. The characteristic frequency is obvious, and the energy value corresponding to it is prominent, nearing the theoretical calculation frequency. Thus, the noise reduction method is effective.

4.2. Signal Feature Vector Dimensionality Reduction

To demonstrate the accurate reflection of gearbox fault characteristics by the LT-PCA extracted in this article, PCA and LT-PCA dimensionality reduction analyses are carried out on the hybrid domain features and wavelet domain features of the obtained signal, respectively. In this process, the contribution rate and cumulative contribution rate of each principal component are shown in Table 8. Notably, the first four cumulative contribution rates contain over 99% of the information, surpassing the theoretical value of 85% [35]. Consequently, the decision is made to downsize the data to a four-dimensional feature vector, facilitating improved visualization and result comparison.

Table 8. Contribution rate and cumulative contribution rate for each principal component.

Feature	Contribution Rate (%)	Cumulative Contribution Rate (%)
M1	85.55	85.55
M2	9.59	95.14
M3	2.65	97.79
M4	1.53	99.32
M5~M27	0.68	100

The features of the two principal components meeting the conditions after dimensionality reduction are projected into two-dimensional space, as shown in Figure 16.
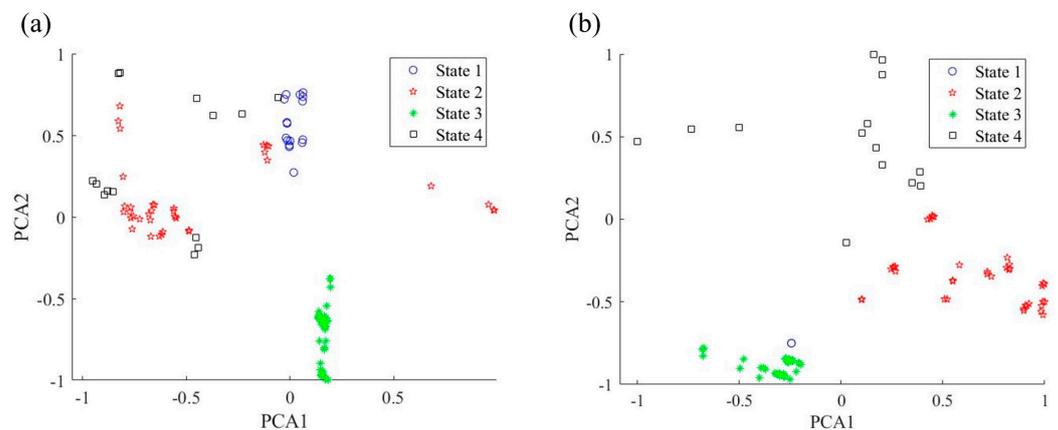
Figure 16. *Cont.*
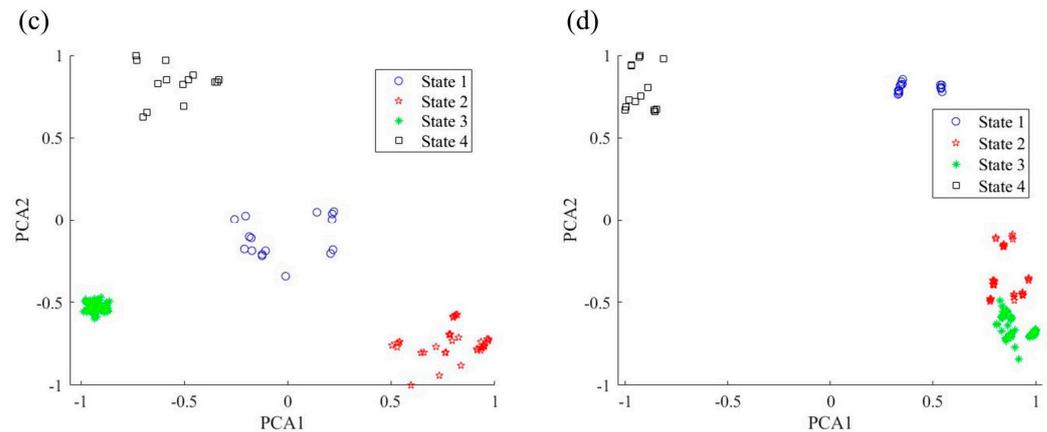
Figure 16. Two-dimensional principal component analysis. (**a**) Two-dimensional principal component analysis of time–frequency mixed domain PCA. (**b**) Two-dimensional principal component analysis of wavelet domain PCA. (**c**) LT-PCA two-dimensional principal component analysis in the time–frequency mixed domain. (**d**) LT-PCA two-dimensional principal component analysis in the wavelet domain.

Figure 16 is used to illustrate the distribution of data points for four different categories. Each point represents a data sample, with colors distinguishing the four categories (states 1, 2, 3, and 4 indicating normal state, mild, moderate, and severe faults, respectively), and the axes indicating different feature dimensions.

Figure 16, it is evident that different domains perform differently in PCA dimensionality reduction concerning classification effects under various faults. In Figure 16a, for the time-frequency hybrid domain, the clustering effect is obviously the worst after normal PCA dimensionality reduction. Except for the moderate fault and normal state, which exhibit some aggregation effects, the classification effects of other faults show overlapping categories, indicating a serious misdiagnosis phenomenon in PCA downsizing classification relying on the time–frequency hybrid domain. Conversely, under LT-PCA dimensionality reduction, as shown in Figure 16c, the aggregation effect is evident, and different faults are clearly demarcated, achieving the best classification effect.

For wavelet domain features, as shown in Figure 16b, under normal PCA dimensionality reduction, the aggregation effect of data points is obvious for normal state and moderate fault cases, and the accuracy of analysis and diagnosis is high. However, for mild and severe fault categories, the proximity of data points makes it challenging to separate them, and the wide gap between data points of a single category increases the likelihood of misjudgment when determining the fault type. In the wavelet domain features under LT-PCA dimensionality reduction, as shown in Figure 16d, the aggregation of different fault datasets is more effective than ordinary PCA aggregation. However, light and moderate faults appear mixed, and fault categories cannot be effectively distinguished.

Overall, the effect of dimensionality reduction in the time–frequency hybrid domain is more pronounced than in a single time–frequency domain. After LT-PCA dimensionality reduction, both time–frequency hybrid and wavelet domains exhibit excellent aggregation effects compared with ordinary PCA dimensionality reduction of the principal components. The results demonstrate that the proposed dimensionality reduction method in this part is superior to the traditional method of dimensionality reduction.

4.3. Fault Category Identification

Following LT-PCA dimensionality reduction, aggregated fault feature vectors are obtained. These feature vectors serve as inputs for a support vector machine (SVM) in the fault pattern recognition of gearboxes. For each fault type, 80 sets of samples are divided into two parts: 60 sets for training the SVM and 20 sets for testing its accuracy. In these

samples, labels 1, 2, 3, and 4 denote normal state, mild fault, moderate fault, and severe fault, respectively.

The following is the optimization process of the NGO-SVM. The optimization parameters 'c' and 'g' optimization process and fitness curve of the NGO algorithm are shown in Figure 17.
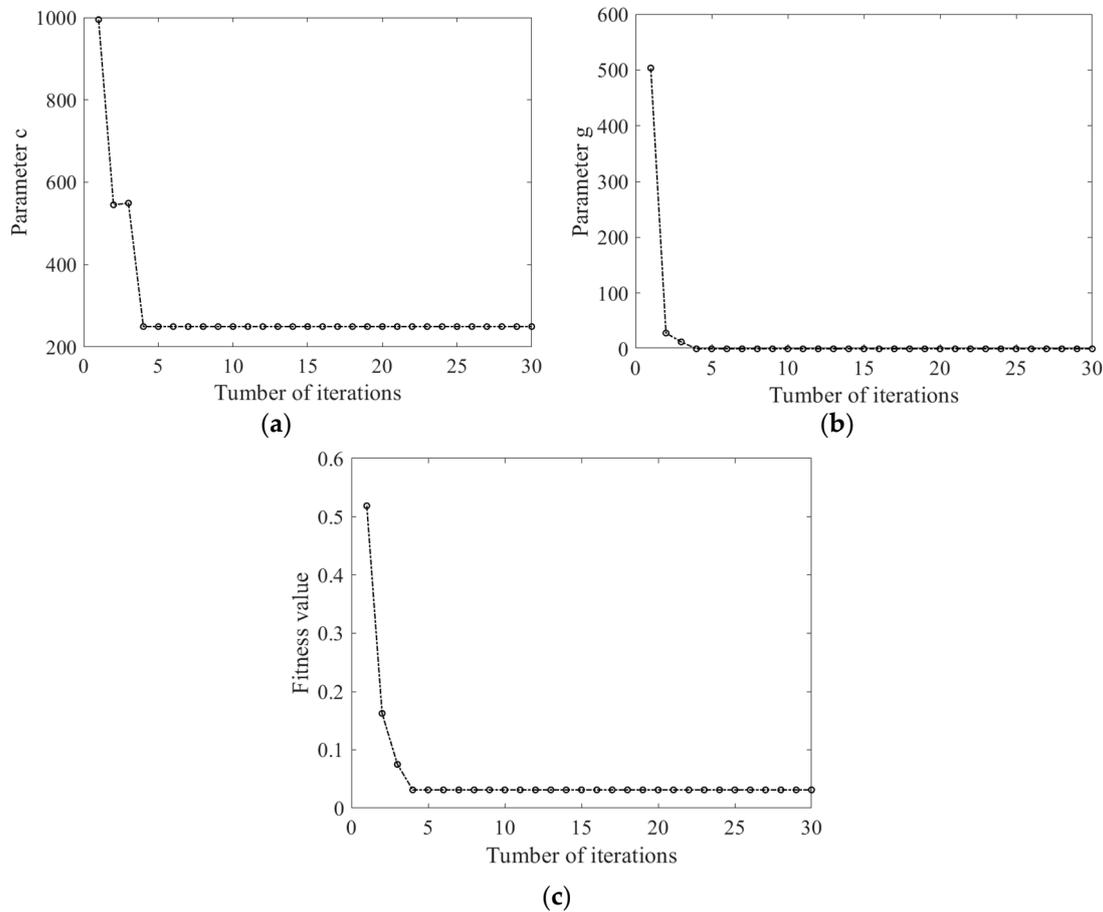
Figure 17. Optimization process and iteration curve of parameters 'c' and 'g'. (**a**) Optimization process of parameter 'c'. (**b**) Optimization process of parameter 'g'. (**c**) Iteration curve.

According to Figure 17, optimizing the SVM using the NGO algorithm finds the parameter configurations that give the optimal cross-validation accuracy after only four iterations, specifically, $c = 147$ and $g = 0.02$. This indicates that the NGO algorithm converges rapidly, finding high-performance parameter configurations with minimal computations. Such fast convergence is highly beneficial for large-scale datasets and complex models, significantly reducing the optimization search time. The results show that the NGO algorithm has a significant advantage in optimizing SVM models.

Figure 18 reveals multiple misclassified sample points in SVM classification without using the optimization algorithm directly. However, after optimization, the SVM classifies the points perfectly, demonstrating the excellent performance of the optimization algorithm.
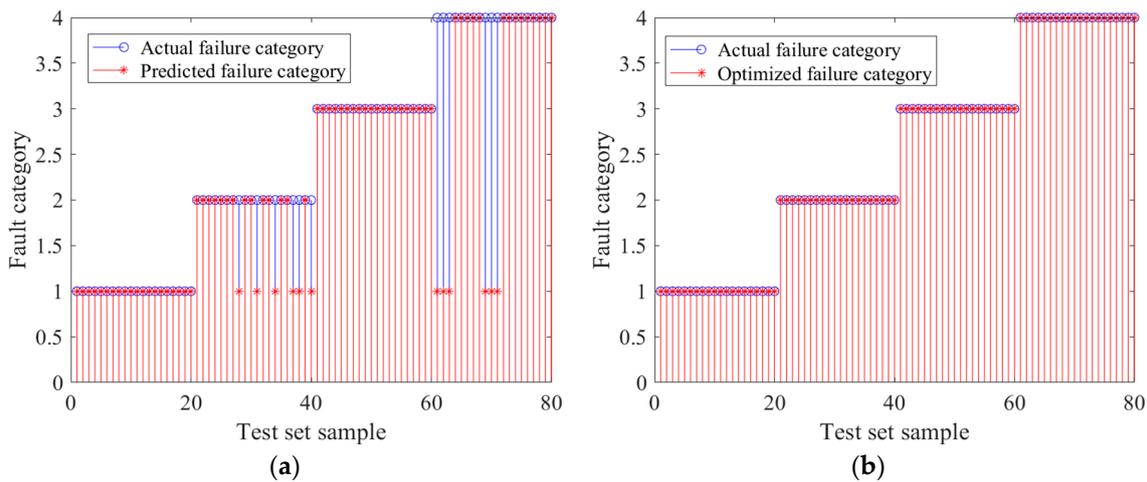
Figure 18. Original classification and classification prediction effect of NGO optimization algorithm. (**a**) Original classification prediction effect. (**b**) Classification prediction effect of NGO optimization algorithm.

The confusion matrix (CM) is a valuable tool in machine learning and statistics for evaluating the performance of a classification model, offering a detailed perspective on how the model performs across different categories. Its primary purpose is to compare the model's classification results with the actual categories, enabling the calculation of various performance metrics. When testing experimental data, fault diagnosis results are visually represented using the confusion matrix, as depicted in Figure 19. The horizontal and vertical coordinates of the figure, labeled 1, 2, 3, and 4, represent the four data categories, allowing for joint diagnosis of no fault, as well as mild, moderate, and severe faults. The rows of the graph represent true category cases, and the columns represent the predicted cases. For example, in column 2 of Figure 19a, where the horizontal coordinate is 2 (representing mild faults), 14 samples are correctly classified, and 6 samples are considered to be in the no-fault state, with an accuracy of 70%. Similarly, for the NGO-SVM, see column 2 of Figure 19b, where the classification accuracy is 100%. The classification prediction of the NGO optimization algorithm surpasses that of the original SVM algorithm.
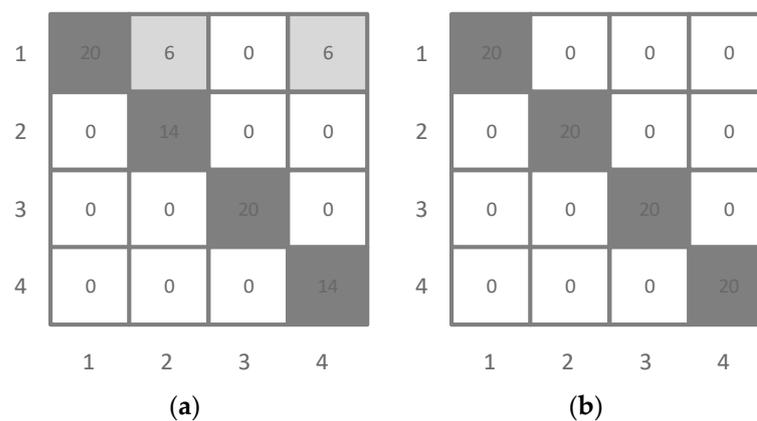
Figure 19. Confusion matrix for classification of original and NGO optimization algorithms. (**a**) Original classification test set confusion matrix. (**b**) NGO optimization algorithm test set confusion matrix.

Similarly, the processed feature dataset is analyzed using MFO-SVM and PSO-SVM algorithms to compare the classification accuracy of each state under the same operating conditions. The classification accuracies of the various algorithms are presented in Table 9, and the results are averaged over 10 tests.

Table 9. Classification accuracy of different optimization algorithms.

Algorithm	Classification Accuracy (%)
SVM	85
NGO-SVM	96.65
MFO-SVM	89.2
PSO-SVM	91.8

In Table 9, it is evident that the unoptimized SVM has a classification accuracy of 85% on this feature dataset, indicating that it is not well-adapted to the needs of the particular problem due to the lack of parameter optimization. This indicates that there is room for improvement of the unoptimized SVM on this task. Optimization of the SVM using different algorithms led to higher classification accuracy compared to the unoptimized SVM. The average classification accuracy of NGO-SVM is 96.65%, which is much higher than the classification accuracy of ordinary SVM, and the optimized SVM accuracy is improved by 11.65%. The NGO algorithm finds a more suitable parameter configuration through adaptive search, which improves the performance of SVM. It demonstrates the effectiveness of the method.

5. Discussion and Conclusions

5.1. Discussion

This article analyzes the actual vibration signal situation by constructing a gearbox vibration test stand, simulating normal, mild, moderate, and severe damage states of gears, and determining the degree of influence of the fault state on the gearbox. Experimental verification confirms the superior performance of the new wavelet noise reduction function and LT-PCA dimensionality reduction proposed in this article compared to other methods. The noise reduction effect of the analog signal using the threshold function proposed in this article is evaluated and compared with the direct use of soft and hard thresholds and the threshold functions proposed in references [25,26]. The results show an effective improvement in the SNR, a reduction in the RMSE, and a superior noise reduction effect. Collecting the actual signal in the presence of surrounding environmental noise poses challenges in obtaining a pure signal, making analysis through quantitative indicators such as SNR difficult. Thus, the energy spectrum is employed to assess the noise reduction effect. In the energy spectrum, the characteristic frequency of the fault can be obtained, indicating the success of the noise reduction method. In the experiments, LT-PCA exhibited superior data aggregation compared to PCA in the dimensionality reduction process. It effectively mitigated the influence of discrete points by compressing the data beforehand, resulting in more accurate fault feature extraction. Analyzing the actual spectrum after noise reduction reveals a slight difference with the theoretical frequency, potentially influenced by factors such as input speed fluctuation, shaft misalignment, gear gap, and other operational aspects of the gearbox, causing interference with gear meshing.

5.2. Conclusions

To address the issue of low accuracy in classifying noisy fault signals, this article proposes a new adaptive wavelet denoising function and logarithmic principal component analysis method to effectively optimize and enhance the accuracy of the SVM algorithm in detecting noisy fault signals. The proposed algorithm, compared with methods solely focused on fault classification, is innovative and yields the following conclusions:

(1)   A new wavelet threshold function is adopted, reducing signal noise by optimizing parameters through a simulated annealing algorithm. The fault characteristics of the signal after noise reduction are clearer than those before noise reduction and other noise reduction methods, providing a better foundation for subsequent signal processing.

(2) In comparison with existing fault feature vector extraction methods, this article combines PCA and data compression to mitigate the abnormal influence caused by outliers in the feature vector. The combined algorithm, as opposed to traditional dimensionality reduction algorithms, delivers accurate feature data for data classification.

Through a series of experiments, the selected algorithm proves superior to traditional methods in denoising, dimensionality reduction, and recognition classification, achieving an average classification accuracy of 96.65%. The system may be influenced by factors such as input speed fluctuations, shaft alignment, and gear clearance, potentially causing interference with gear meshing and resulting in errors between the failure frequency and the theoretical value. This aspect is not explored in depth in this paper, and future studies will conduct a thorough analysis of the multifactorial signals affecting vibration to deepen our understanding further.

Author Contributions: Conceptualization, Q.Z.; Data curation, Q.Z. and C.S.; Funding acquisition, Q.Z.; Investigation, C.S.; Methodology, C.S.; Project administration, Q.Z.; Software, C.S.; Super-vision, Y.Y.; Writing—original draft, C.S.; Writing—review and editing, Q.Z. and C.S. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

Abbreviations

LT-PCA	Logarithmic Transformation and Principal Component Analysis	RMSE	Root Mean Square Error
SVM	Support Vector Machine	PCA	Principal Component Analysis
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise	NGO	Northern Goshawk Optimization
IMF	intrinsic mode function	MFO	Moth–Flame Optimization
SA	Simulated Annealing	PSO	Particle swarm optimization
SNR	Signal-to-Noise Ratio		

References

1. Zhu, H.; Cheng, J.; Zhang, C.; Wu, J.; Shao, X. Stacked pruning sparse denoising autoencoder based intelligent fault diagnosis of rolling bearings. *Appl. Soft Comput.* **2020**, *88*, 106060. [CrossRef]
2. Xu, Y.; Luo, M.; Li, T.; Song, G. ECG signal denoising and baseline wander correction based on CEEMDAN and wavelet thresh-old. *Sensors* **2017**, *17*, 2754. [CrossRef] [PubMed]
3. Chen, W.; Li, J.; Wang, Q.; Han, K. Fault feature extraction and diagnosis of rolling bearings based on wavelet thresholding denoising with CEEMDAN energy entropy and PSO-LSSVM. *Measurement* **2021**, *172*, 108901. [CrossRef]
4. Sun, K.; Zhang, J.; Shi, W.; Guo, J. Extraction of partial discharge pulses from the complex noisy signals of power cables based on CEEMDAN and wavelet packet. *Energies* **2019**, *12*, 3242. [CrossRef]
5. Li, G.; Guan, Q.; Yang, H. Noise reduction method of underwater acoustic signals based on CEEMDAN, effort-to-compress complexity, refined composite multiscale dispersion entropy and wavelet threshold denoising. *Entropy* **2018**, *21*, 11. [CrossRef]
6. Zhou, Z.; Zhang, J.; Cheng, R.; Rui, Y.; Cai, X.; Chen, L. Improving purity of blasting vibration signals using advanced Empirical Mode Decomposition and Wavelet packet technique. *Appl. Acoust.* **2022**, *201*, 109097. [CrossRef]
7. Abdelkader, R.; Kaddour, A.; Bendiabdellah, A.; Derouiche, Z. Rolling bearing fault diagnosis based on an improved denoising method using the complete ensemble empirical mode decomposition and the optimized thresholding operation. *IEEE Sens. J.* **2018**, *18*, 7166–7172. [CrossRef]
8. Deng, L.; Zhang, A.; Zhao, R. Intelligent identification of incipient rolling bearing faults based on VMD and PCA-SVM. *Adv. Mech. Eng.* **2022**, *14*, 16878140211072990. [CrossRef]
9. Zhu, J.; Hu, T.; Jiang, B.; Yang, X. Intelligent bearing fault diagnosis using PCA–DBN framework. *Neural Comput. Appl.* **2020**, *32*, 10773–10781. [CrossRef]

10. Pule, M.; Matsebe, O.; Samikannu, R. Application of PCA and SVM in Fault Detection and Diagnosis of Bearings with Varying Speed. *Math. Probl. Eng.* **2022**. [CrossRef]

11. Ekanayake, T.; Dewasurendra, D.; Abeyratne, S.; Ma, L.; Yarlagadda, P. Model-based fault diagnosis and prognosis of dynamic systems: A review. *Procedia Manuf.* **2019**, *30*, 435–442. [CrossRef]

12. Wang, Z.; Huang, Z.; Song, C.; Zhang, H. Multiscale adaptive fault diagnosis based on signal symmetry reconstitution preprocessing for microgrid inverter under changing load condition. *IEEE Trans. Smart Grid* **2016**, *9*, 797–806. [CrossRef]

13. Tariq, M.F.; Khan, A.Q.; Abid, M.; Mustafa, G. Data-driven robust fault detection and isolation of three-phase induction motor. *IEEE Trans. Ind. Electron.* **2018**, *66*, 4707–4715. [CrossRef]

14. Ma, Z.; Zhao, M.; Li, B.; Fan, H. A novel blind deconvolution based on sparse subspace recoding for condition monitoring of wind turbine gearbox. *Renew. Energy* **2021**, *170*, 141–162. [CrossRef]

15. Ravikumar, K.; Madhusudana, C.; Kumar, H.; Gangadharan, K. Classification of gear faults in internal combustion (IC) engine gearbox using discrete wavelet transform features and K star algorithm. *Eng. Sci. Technol. Int. J.* **2022**, *30*, 101048. [CrossRef]

16. Zhong, T.; Qu, J.; Fang, X.; Li, H.; Wang, Z. The intermittent fault diagnosis of analog circuits based on EEMD-DBN. *Neurocomputing* **2021**, *436*, 74–91. [CrossRef]

17. Tuerxun, W.; Chang, X.; Hongyu, G.; Zhijie, J.; Huajian, Z. Fault diagnosis of wind turbines based on a support vector machine optimized by the sparrow search algorithm. *IEEE Access* **2021**, *9*, 69307–69315. [CrossRef]

18. Lin, Y.; Xiao, M.; Liu, H.; Li, Z.; Zhou, S.; Xu, X.; Wang, D. Gear fault diagnosis based on CS-improved variational mode decomposition and probabilistic neural network. *Measurement* **2022**, *192*, 110913. [CrossRef]

19. Ahmed, H.O.A.; Wong, M.L.D.; Nandi, A.K. Classification of bearing faults combining compressive sampling, laplacian score, and support vector machine. In Proceedings of the IECON 2017—43rd Annual Conference of the IEEE Industrial Electronics Society, Beijing, China, 29 October–1 November 2017.

20. Han, D.; Zhao, N.; Shi, P. Gear fault feature extraction and diagnosis method under different load excitation based on EMD, PSO-SVM and fractal box dimension. *J. Mech. Sci. Technol.* **2019**, *33*, 487–494. [CrossRef]

21. Qin, A.; Hu, Q.; Zhang, Q.; Lv, Y.; Sun, G. Application of sensitive dimensionless parameters and PSO–SVM for fault classification in rotating machinery. *Assem. Autom.* **2019**, *40*, 175–187. [CrossRef]

22. Dong, Z.; Zheng, J.; Huang, S.; Pan, H.; Liu, Q. Time-shift multi-scale weighted permutation entropy and GWO-SVM based fault diagnosis approach for rolling bearing. *Entropy* **2019**, *21*, 621. [CrossRef]

23. Cao, D.; Bi, Y.; Huang, Q.; Chen, Y.; Guo, L.; Lai, M. BOTDR denoising scheme based on joint improvement of wavelet threshold. *Foreign Electron. Meas. Technol.* **2022**, *41*, 83–86.

24. Yang, C.; Nie, C.; Wang, H.; Ruan, X. Research of noise reduction algorithm and effect evaluation about EMG interference based on improved wavelet threshold. *Electron. Meas. Technol.* **2021**, *44*, 80–86.

25. Wu, Y.; Xing, H.; Li, J. Wavelet denoising algorithm with improved threshold function. *J. Electron. Meas. Instrum.* **2022**, *36*, 9–16.

26. Wu, F.; Ma, C.; Cheng, K. Study on wavelet denoising method of vibration signal based on improved threshold. *J. Hefei Univ. Technol. (Nat. Sci.)* **2022**, *45*, 873–877+900.

27. Peng, C.; Shangguan, W.; Xing, Y.; Cai, B. Fault diagnosis method for on-board equipment of CTCS based on dual-view fault feature extraction. *J. China Railw. Soc.* **2022**, *44*, 63–70.

28. Tan, L.; Wu, F.; Li, W. Image compression and reconstruction based on PCA. Journal of Physics: Conference Series. *IOP Publ.* **2021**, *1944*, 012021.

29. Zhang, A.; Yu, D.; Zhang, Z. TLSCA-SVM fault diagnosis optimization method based on transfer learning. *Processes* **2022**, *10*, 362. [CrossRef]

30. Dehghani, M.; Hubálovský, Š.; Trojovský, P. Northern goshawk optimization: A new swarm-based algorithm for solving optimization problems. *IEEE Access* **2021**, *9*, 162059–162080. [CrossRef]

31. Tan, J. *Learn NVH from Here*; Machinery Industry Press: Beijing, China, 2018.

32. Ma, L.; Wang, G.; Zhang, P.; Huo, Y. Fault Diagnosis Method of Circuit Breaker Based on CEEMDAN and PSO-GSA-SVM. *IEEJ Trans. Electr. Electron. Eng.* **2022**, *17*, 1598–1605. [CrossRef]

33. Yang, J. Bearing Fault Diagnosis Method in the Machining Process. *Mach. Des. Manuf.* **2021**, 112–116. [CrossRef]

34. Galezia, A. Teager-Kaiser energetic trajectory for machine diagnosis purposes. *J. Vibroeng.* **2017**, *19*, 1014–1025. [CrossRef]

35. Wang, Z.; Han, D.; Li, M.; Liu, H.; Cui, M. The abnormal traffic detection scheme based on PCA and SSH. *Connect. Sci.* **2022**, *34*, 1201–1220. [CrossRef]