



# Article A Hybrid Gearbox Fault Diagnosis Method Based on GWO-VMD and DE-KELM

Gang Yao <sup>1,\*</sup>, Yunce Wang <sup>2</sup>, Mohamed Benbouzid <sup>3</sup> and Mourad Ait-Ahmed <sup>4</sup>

- Sino-Dutch Mechatronics Engineering Department, Shanghai Maritime University, Shanghai 201306, China
- <sup>2</sup> Electrical Engineering Department, Shanghai Maritime University, Shanghai 201306, China; wangyunce9032@foxmail.com
- <sup>3</sup> Institut de Recherche Dupuy de Lôme (UMR CNRS 6027 IRDL), University of Brest, 29238 Brest, France; Mohamed.Benbouzid@univ-brest.fr
- <sup>4</sup> IREENA, l'Université de Nantes/Polytech Nantes, 44602 Nantes, France; ait-ahmed-m@univ-nantes.fr
- \* Correspondence: gangyao@shmtu.edu.cn

Featured Application: The proposed GWO-VMD- and DE-KELM-based hybrid fault diagnosis method can be applied to offline and high-precision fault diagnosis of gearboxes in wind power generation systems, rail transports, and other industrial sectors including gearboxes.

Abstract: In this paper, a vibration signal-based hybrid diagnostic method, including vibration signal adaptive decomposition, vibration signal reconstruction, fault feature extraction, and gearbox fault classification, is proposed to realize fault diagnosis of general gearboxes. The main contribution of the proposed method is the combining of signal processing, machine learning, and optimization techniques to effectively eliminate noise contained in vibration signals and to achieve high diagnostic accuracy. Firstly, in the study of vibration signal preprocessing and fault feature extraction, to reduce the impact of noise and mode mixing problems on the accuracy of fault classification, Variational Mode Decomposition (VMD) was adopted to realize adaptive signal decomposition and Wolf Grey Optimizer (GWO) was applied to optimize parameters of VMD. The correlation coefficient was subsequently used to select highly correlated Intrinsic Mode Functions (IMFs) to reconstruct the vibration signals. With these re-constructed signals, fault features were extracted by calculating their time domain parameters, energies, and permutation entropies. Secondly, in the study of fault classification, Kernel Extreme Learning Machine (KELM) was adopted and Differential Evolutionary (DE) was applied to search its regularization coefficient and kernel parameter to further improve classification accuracy. Finally, gearbox vibration signals in healthy and faulty conditions were obtained and contrast experiences were conducted to validate the effectiveness of the proposed hybrid fault diagnosis method.

**Keywords:** gearboxes; fault diagnosis; vibration signals; variational mode decomposition; correlation coefficient; kernel extreme learning machine; parameter optimization

# 1. Introduction

# 1.1. Motivations

Being a crucial transmission device of drive trains, gearboxes are widely used in modern rotating machinery in the industry sectors of wind power generation, rail transport, helicopters, automobiles, mining, agriculture, etc. because of their large transmission ratio, high efficiency, and compactness [1]. However, these types of machinery and equipment are generally operating in harsh environments, and the gearboxes are often operating in high-load states. These make the components of the gearboxes prone to failure, and the failures affect the safe operation of the entire system and may even lead to catastrophic consequences and huge economic losses. Therefore, the research of gearbox fault diagnosis and the application reference values are of great theoretical significance.



Citation: Yao, G.; Wang, Y.; Benbouzid, M.; Ait-Ahmed, M. A Hybrid Gearbox Fault Diagnosis Method Based on GWO-VMD and DE-KELM. *Appl. Sci.* **2021**, *11*, 4996. https://doi.org/10.3390/ app11114996

Academic Editor: Antonio Fernández-Caballero

Received: 4 May 2021 Accepted: 26 May 2021 Published: 28 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). Vibration is one of the most widely used signals for gearbox fault diagnosis [2]. When faults happen in the components of gearboxes, abnormal non-stationary and nonlinear vibrations would be introduced in the defective locations. These vibration signals contain a wealth of fault signatures, which can be used to identify faults. However, noise contained in gearboxes makes it difficult to distinguish between useful feature signals and interference in the fault feature extraction process. This problem could lead to misdiagnosis in the fault classification stage. Therefore, finding effective methods to pre-process signals and extract fault features from non-stationary vibration signals containing wideband noise, and then classify the features accurately has become a key issue in gearbox fault diagnosis research.

In this research background, data-driven diagnostic methods are the focus, and the main purpose of the research was to develop vibration signal preprocessing, feature extraction, and fault classification techniques for accurate gearbox fault diagnosis.

#### 1.2. State of the Art of Gearbox Fault Diagnosis

#### 1.2.1. Feature Signal Adaptive Processing Aspect

For non-stationary vibration signals, their spectrum changes over time. Therefore, single frequency- or time-domain representations fail to reflect all the characteristics of the signals. In this case, Time-Frequency Analysis (TFA) becomes a very effective method for processing the gearbox vibration signals for fault feature extraction [3], and many TFA techniques, such as short-time Fourier transform (STFT), Wigner–Ville distribution (WVD), Discrete Wavelet Transform (DWT), empirical mode decomposition (EMD), VMD, etc. have been proposed and applied.

For example, Zhong et al. [4] adopted STFT to decompose the vibration signal from the time domain to the time-frequency domain for bearing fault diagnosis with changing rotational speed and load. Yao et al. [5] used discrete STFT with long- and short-time windows to obtain time-frequency spectra with different resolutions. A new signal representation approach was put forward by Yang et al. [6]; this method represented the vibration signals with a union of redundant dictionary and applied WVD to express the obtained atoms to get their time-frequency distribution. Heydarzadeh et al. [7] utilized DWT to decompose the measured vibration, acoustic, and torque signals to obtain fault features. Zhao et al. [8] proposed a Deep Residual Network (DRN) based on multiple wavelet coefficient fusion to obtain fault features. A new signal processing method combining EMD with adaptive multiscale morphological analysis was proposed by Li et al. [9]. Abdelkader et al. [10] presented an improved EMD to decompose gearbox vibration signals, and the IMFs were selected according to their energies for feature extraction. Wang et al. [11] used VMD to decompose the vibration signals, and Particle Swarm Optimization (PSO) was applied to optimize the parameters of VMD. Guo et al. [12] also adopted VMD to decompose the bearing vibration signals; the multi-scale permutation entropy and cuckoo search algorithm were further applied for VMD parameter optimization.

The application of the above-mentioned TFA techniques in the literature has achieved very interesting experimental results. However, the following deficiencies may occur when these methods are applied. (1) STFT uses fixed-size window functions, and this makes temporal and frequency resolution constant; (2) when WVD is used to analyze multi-component signals, it has a cross interference problem; (3) the predefined and invariant wavelet basis function results in DWT's lack of adaptation, and its decomposition layer number needs to be manually set; (4) for EMD, there are problems of mode mixing, and over- and under-envelopes exist; and (5) for VMD, the decomposition effect relies on its parameter setting [13].

Among the common adaptive signal processing methods, VMD is very attractive because this approach can alleviate the mode mixing problem of EMD and its varieties, such as local mean decomposition and Ensemble Empirical Mode Decomposition (EEMD), and this method also has high decomposition efficiency [13,14]. Accordingly, the VMD method was selected in this paper for gearbox vibration signal decomposition.

### 1.2.2. Fault Identification Aspect

After vibration signals are pre-processed and fault features are extracted, fault classification methods then need to be used to identify faults. In the last decade, emerging machine-learning-based intelligent fault classification methods, such as Convolution Neural Network (CNN), DRN, Extreme Learning Machine (ELM), Transfer Learning (TL), Support Vector Machine (SVM), etc., drew much attention of researchers, and a lot of interesting research work can be found in a survey of the literature.

For instance, Gao et al. [15] proposed an adaptive CNN based on Nesterov momentum and the adaptive learning rate rule for bearing fault diagnosis. Jiang et al. [16] put forward a multiscale CNN method to realize feature extraction and fault classification for wind turbine gearbox diagnosis. An anti-noise Neural Network (NN) was presented by Jin et al. [17] to diagnose bearing faults with changing loads and noise in the vibration signals. Chen et al. [18] proposed a multi-task learning-based DRN for bearing fault diagnosis. To deal with the limited sample problem, an improved Deep Learning (DL) model was developed by Saufi et al. [19] to realize gearbox fault diagnosis. Zhong et al. [20] proposed a Bayesian ELM to identify concurrent wind turbine gearbox faults based on Huang transforms and correlation techniques. Udmale et al. [21] adopted the TL technique to transfer knowledge from vibration datasets, including massive data to situations without enough sample data. A hybrid DL method involving CNN, SVM, and PSO was designed by Zhang et al. [22] to realize feature extraction and classification of wind turbine gearbox bearing faults. Zhu et al. [23] combined DL and TL to deal with the unlabeled data in the application. Yuan et al. [24] proposed a rolling bearing fault diagnosis method based on CNN and SVM to reduce manual intervention in the feature extraction process. Another two illuminating research studies include References [25,26]. Glowacz et al. [25] combined fast Fourier transform, a frequency amplitude selection method, and the mean of vector sum method to process vibration signals and to extract feature vectors for rotor fault diagnosis of electrical motors. Then, nearest neighbor, linear discriminant analysis, and linear SVM were adopted to classify faults. Sun et al. [26] presented another hybrid method, which combined EEMD, the Levy flight-based moth-flame optimization, and the naive Bayes, to realize fast bearing fault diagnosis.

In the above literature review, it can be seen that many interesting research findings based on intelligent diagnosis methods have been achieved. However, these methods also have their limitations and suitability as follows. (1) When NN-based methods are adopted, a large amount of fault samples are normally needed, which leads to a long training time in which local minimums could be trapped; and (2) when SVM is applied, it is not intuitive to determine its regularization coefficient, and the training time is long with a large-scale training set. Considering the above-mentioned methods, if no knowledge transfer is explicitly needed, if only limited samples are available, and if fast training speed is required, ELM is a suitable fault classification method due to its simplicity, fast training speed, and good generalization ability [27]. Thus, ELM was selected as the gearbox fault classification method in this paper.

#### 1.3. Organization of This Article

The rest of this paper is organized as follows. In Section 2, common faults of gearboxes are summarized, and their causes are briefly analyzed. Then the proposed vibration signal preprocessing and fault feature extraction methods are presented in Section 3, and the improved fault classification method is presented in Section 4. Afterward, in Section 5, contrast experiments are carried out on gearbox faulty data to demonstrate the effectiveness of the proposed fault diagnosis method. Finally, discussion and conclusions are presented in Section 6.

### 2. Common Faults of Gearboxes and Causes Analysis

The internal elements and structures of gearboxes are various to meet the transmission performance requirements of different applications. For example, planetary gear trains,

widely adopted in wind generation transmissions, consist of either spur, single-helical, or herringbone gears to increase the rotational speed of the rotor to match that required by the generator [28,29]. Despite the diversity of the internal structure and components, most gearboxes include various kinds of gears, gear shafts, bearings, enclosures, etc., and all these elements are prone to failure because of the following possible reasons [30]: (1) wearing or damage due to long-term operation with high alternating loads; (2) insufficient component lubrication under gearbox intensive operation; and (3) high instantaneous impacts at emergent loaded starting- or stopping-operation. According to Henriquez et al. [31], the elements with high failure rates in gearboxes are bearings and gears, a total of up to more than 60% of failures occur on these two components. Therefore, the faults of bearings and gears were taken as research objects in this paper for the following diagnosis research, which is a general study of gearboxes without focusing on a particular gearbox type.

According to Ming et al. [32], the common faults and their causes of gears and bearings are summarized in Table 1.

Elements	Common Faults	Causes
Gears	Broken Teeth	<ul> <li>Strong shocks in operation from heavy load.</li> <li>Periodically intense bending stress.</li> </ul>
	Pitting	- The shear stress on tooth flanks makes the metal on tooth surface deform and peel off to form uneven pits.
	Cracks	<ul> <li>Constant impacts during long-term operation.</li> <li>A defect of temperature control in the gear production process.</li> </ul>
Bearings	Inner Ring Faults	<ul> <li>Mainly include two types: wearing and pitting.</li> <li>Small particles in grease rubbing against the surface of the inner ring, resulting in wearing on the surface.</li> <li>Corrosion on the inner ring surface makes the metal on the surface peel off and further causes pitting.</li> </ul>
	Outer Ring Faults	<ul> <li>Mainly include three types: wear, pitting and fracture.</li> <li>Same causes as the wearing and pitting faults of the inner ring.</li> <li>Continuous impact due to overloading induces fracture faults.</li> </ul>
	Roller Faults	<ul> <li>All inner and outer ring faults can cause roller faults.</li> <li>Extra friction due to foreign particles or poor manufacturing causes damage to the rollers.</li> </ul>

Table 1. Common fau	llts of gears ar	nd bearings	[32]
---------------------	------------------	-------------	------

To identify the above common faults of gears and bearings, vibration signals of gearboxes are normally taken as featured signals [33]. Based on these vibration data, signal preprocessing, fault feature extraction, and classification methods are studied in the following sections.

# 3. Vibration Signal Preprocessing and Fault Feature Extraction

To extract effective fault features from the non-stationary gearbox vibration signals containing noise for fault classification, a three-step signal processing procedure is presented in this paper as follows.

Step 1. GWO-VMD-based vibration signal decomposition.

To reduce the effect of mode mixing during signal decomposition on the final fault classification accuracy, VMD is adopted to decompose the vibration signal, and the GWO algorithm is used to optimize the parameters of VMD for a better noise-elimination effect.

Step 2. Correlation-based vibration signal reconstruction.

After the vibration signal is decomposed by GWO-VMD, correlation coefficients of IMFs are calculated, and only highly correlated IMFs are retained for vibration signal reconstruction.

Step 3. Reconstructed signals-based fault feature extraction.

With the reconstructed vibration signals, 18 parameters, including 16 selected timedomain parameters, energies, and permutation entropies, are calculated to form fault feature vectors for the subsequent fault classification.

The above vibration signal preprocessing and feature extraction process is shown in Figure 1.



Figure 1. Workflow of vibration signal preprocessing and feature extraction.

#### 3.1. GWO-VMD-Based Vibration Signal Decomposition

#### 3.1.1. Basic Principle of VMD

VMD is a popular TFA method proposed by K. Dragomiretskiy et al. in 2014 [34]. For a non-stationary gearbox vibration signal v(t), VMD decomposes v(t) into a series of Intrinsic Mode Functions (IMFs), denoted as  $u_k(t)$ , k = 1, ..., K, where K is the number of IMFs. Each  $u_k(t)$  has limited bandwidth around its frequency center. VMD iterates to update  $u_k(t)$  in the frequency domain and extracts the spectrum gravity center as its center frequency  $\omega_k$ . According to References [34–36], VMD essentially solves a constrained variational model expressed with

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^K \| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \|_2^2 \right\} \text{s.t.} \sum_{k=1}^K u_k(t) = v(t)$$
(1)

where  $\{u_k\} = \{u_1, \ldots, u_k, \ldots, u_K\}$  is the set of the decomposed IMFs,  $\{\omega_k\} = \{\omega_1, \ldots, \omega_k, \ldots, \omega_K\}$  is the set of frequency centers of IMFs,  $\partial_t$  represents a partial derivative of time t,  $\delta(t)$  is Dirac distribution, the symbol \* stands for convolution operator, and  $\|\cdot\|_2$  represents Euclidian norm.

To solve the above constrained variational model, a quadratic penalty factor  $\alpha$  and Lagrange multiplier  $\lambda$  are introduced to transform the constrained optimization problem expressed with Equation (1) to an unconstrained one as follows [34–36]:

$$L(\{u_k\},\{\omega_k\},\lambda) = \alpha \sum_{k=1}^{K} \|\partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \|_2^2 + \|v(t) - \sum_{k=1}^{K} u_k(t) \|_2^2 + \left\langle \lambda(t), v(t) - \sum_{k=1}^{K} u_k(t) \right\rangle$$
(2)

With Equation (2), the alternating direction method of multipliers is applied to update  $u_k$ ,  $\omega_k$ , and  $\lambda$  iteratively. The updated equations are as follows [35].

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{v}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\lambda(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2}$$
(3)

$$\hat{\omega}_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega}$$
(4)

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^{n}(\omega) + \tau \left( \hat{v}(\omega) - \sum_{k=1}^{K} \hat{u}_{k}^{n+1}(\omega) \right)$$
(5)

where the symbol  $\land$  above the variables represents the updated values of  $u_k$ ,  $\omega_k$ ,  $\lambda$ , and v. Additionally,  $\tau$  is the updating parameter of Lagrange multiplier.

More details concerning the implementation process of VMD can be found in References [34–36].

# 3.1.2. Deficiencies of VMD

Compared with EMD, VMD has better decomposition capabilities in processing intermittent and noise-containing signals. When VMD parameters are properly set, the mode mixing phenomenon could be better suppressed than with EMD [37].

However, VMD also has deficiencies—its two important parameters, the mode number *K* for decomposition and the penalty factor  $\alpha$  for frequency bandwidth adjustment, need to be manually determined with experience, and the decomposition result would be affected when the two parameters are not properly set. According to the principle of VMD, in cases in which the value of  $\alpha$  is fixed, if *K* is too large, the signal will be over-decomposed; this causes a same frequency band to be decomposed into multi-mode. While if *K* is too small, the signal is under-decomposed, and the central frequencies of different bands appear on the same mode. In another case, when the value of *K* is fixed, a too big  $\alpha$  results in a bandwidth between modes that is very small, resulting in the loss of useful feature information. If the value of  $\alpha$  is too small, the bandwidth of each mode becomes wide, resulting in more noise being retained. This will further lead to low fault classification accuracy. Therefore, it is important to optimize the two parameters to get more reasonable decomposition results.

# 3.1.3. GWO-Based [K, $\alpha$ ] Optimization for VMD

Grey Wolf Optimizer is a swarm intelligent technique developed by Mirjalili et al. in 2014 [38]. The algorithm mimics the leadership hierarchy of wolves, which are well known for their group hunting social behavior. When compared with the Genetic Algorithm (GA) and other swarm algorithms, such as PSO, GWO is believed to be more effective [39]. Therefore, GWO was chosen in this paper to optimize the two parameters of VMD.

In the social hierarchy of grey wolves, the fittest is named *alpha*, the second and third fittest are named *beta* and *delta*, respectively, and the rest of the wolves are called *omega*. The optimization process of GWO is led by *alpha*, *beta*, and *delta*. With the leadership in the hierarchy, the GWO hunting procedure mainly consists of three steps known as encircling prey, hunting, and attacking prey. The mathematical models of these three behaviors were formalized by Mirjalili et al. in Reference [38].

When GWO is applied to simultaneously hunt the optimal values of [K,  $\alpha$ ], the procedure is as follows.

Step 1. Parameter initialization.

Initialize the size of the wolf group, the maximum iteration steps, the upper and lower bounds of *K* and  $\alpha$ , and the values of *alpha*, *beta*, *delta*, and *omega*.

Step 2. Define and calculate the fitness function.

In order to ensure that the reconstructed vibration signals after decomposition by GWO-VMD can retain as much fault feature as possible while removing as much noise as possible, a fitness function is constructed for GWO to minimize the reconstructed signal by using permutation entropy [40] and Kurtosis [13] as follows:

$$fn = -kt + H_p \tag{6}$$

where *fn* stands for the fitness function, *kt* and *Hp* represent Kurtosis and permutation entropy, respectively. The mathematical expressions of *kt* and *Hp* are as follows:

$$kt = \frac{E(x-\mu)^4}{\sigma^4} \tag{7}$$

$$H_p = -\frac{1}{ln(m_1!)} \sum_{i=1}^{q} p_i ln p_i$$
(8)

where *x* is an arbitrary signal to calculate Kurtosis,  $\mu$  and  $\sigma$  are the mean value and standard deviation of *x*, respectively,  $p_i$  is the probability of each row in the  $m_1 \times q$  reconstruction matrix of the signal, and  $m_1$  is the embedded dimension.

Step 3. Update the wolf group.

After calculating the fitness value of each wolf, *alpha*, *beta*, *delta*, and *omega* are updated with the following rules: if fn < alpha, then alpha = fn; else if alpha < fn < beta, then beta = fn; else if *beta* < fn < omega, then omega = fn.

Step 4. Recalculate fitness values of the wolf group.

Recalculate the fitness value of each wolf, compare the new value with the previous one—if the updated value is smaller than the old one, keep the latest position of the wolf and replace the previous value with the new one; otherwise retain the previous fitness.

Step 5. Terminating condition judgement.

Determine whether the maximum iteration is reached; if not, go back to Step 2; otherwise, stop the iteration and return the optimal values of  $[K, \alpha]$ .

# 3.2. Correlation-Based Gearbox Vibration Signal Reconstruction

With the optimal values of [K,  $\alpha$ ] found by GWO, the gearbox vibration signals can be decomposed to a series of IMFs by VMD. The relation between the original vibration signal and the mode functions is the constraint condition in Equation (1), which can be re-written as follows:

$$v(t) = \sum_{k=1}^{K} IMF_k = \sum_{k=1}^{K} u_k(t)$$
(9)

To remove noise contained in v(t), the correlation coefficient, denoted as *Corr*, is used in this paper to evaluate the relevance of each IMF with v(t). The range of *Corr* is [–1, 1], and a value close to 1 indicates high correlation between the two signals. Thereby, a correlation threshold, denoted as *Corr*<sub>th</sub>, is set, and only IMFs having a higher correlation coefficient than *Corr*<sub>th</sub> are selected for signal reconstruction.

The expression of *Corr* is [41]

$$Corr_k = \frac{cov(v, u_k)}{\sqrt{D(v)}\sqrt{D(u_k)}}$$
(10)

where  $cov(v, u_k)$  represents the covariance between gearbox vibration signal v(t) and the  $k^{th}$  IMF, while D(v) and  $D(u_k)$  stand for variances of v(t) and  $u_k(t)$ , respectively.

Set  $Corr_{th} = 0.25$  in this paper for the subsequent research work, and select all IMFs with  $Corr_i \ge Corr_{th}$ ; thus, the vibration signal is reconstructed with

$$v'(t) = u_1(t) + \ldots + u_i(t), \ i \le K$$
 (11)

where v'(t) is the reconstructed vibration signal.

# 3.3. Fault Feature Extraction

Extracting distinguishing fault features is crucial for the accuracy of fault classification. From the reconstructed vibration signal v'(t), 18 dimensional vectors, including 16 time-domain parameters, signal energies, and permutation entropies, are calculated. The 18 parameters and their calculation equations are listed in Table 2.

No.	Parameter Notations Names		Mathematical Expressions
1	Maximum amplitude	$v_{max}$	$v_{max} = \max(v'(t))$
2	Minimum amplitude	$v_{min}$	$v_{\min} = \min(v'(t))$
3	Peakedness <sup>1</sup>	$v_p$	$v_p = \frac{1}{N_r} \sum v'(t)^4$
4	Peak-to-peak value	$v_{pp}$	$v_{pp} = \max(v'(t)) - \min(v'(t))$
5	Mean value	$v_m$	$v_m = rac{1}{N_n} \sum v \prime(t)$
6	Slant	$v_s$	$v_s = rac{1}{N_v} \sum v \prime(t)^3$
7	Root mean square value	$v_{rms}$	$v_{rms} = \sqrt{rac{1}{N_v}\sum v\prime(t)^2}$
8	Root square amplitude	$v_r$	$v_r = \left(rac{1}{N_v}\sum \sqrt{ert v'(t) ert} ight)^2$
9	Mean of absolute value	$v_{ma}$	$v_{ma} = \frac{1}{N_n} \sum  v'(t) '$
10	Variance	$v_{\sigma}$	$v_{\sigma} = \frac{1}{N} \sum (v'(t) - v_m)^2$
11	Shape factor $v_{sf}$		$v_{\rm sf} = rac{v_{ m rms}}{v_{ m ma}}$
12	Skewness $v_{sn}$ $v_{sn}$		$v_{sn} = rac{1}{N_v} \sum (v'(t) - v_m)^3}{v_m^{-\frac{3}{2}}}$
13	Impulse factor	$v_{if}$	$v_{if} = \frac{\tilde{v}_{max}}{v_{max}}$
14	Crest factor	$v_{cf}$	$v_{cf} = \frac{v_{max}}{v_{rms}}$
15	Clearance factor $v_{clf}$		$v_{clf} = \frac{v_{max}^{max}}{v_r}$
16	Kurtosis factor	$v_{kf}$	$v_{kf} = \frac{\frac{1}{N_v} \sum (v'(t) - v_m)^4}{v_m^2}$
17	Energy $v_E$		$v_E = \int_{-\infty}^{+\infty}  v'(t) ^2 dt$
18	Permutation entropy	$v_{pe}$	$v_{pe} = -\frac{1}{\ln(m!)} \sum_{i=1}^{q} p_i ln p_i$

Table 2. List of the 18 feature parameters.

In the calculation equation of Peakedness,  $N_v$  indicates the length of v'(t).

After calculating the parameters listed in Table 2 from v'(t), an 18-dimensional fault feature vector, denoted as *T*, for each sample datum is obtained in the following form for the subsequent fault classification:

$$T = [v_{max}, v_{min}, v_p, v_{pp}, v_m, v_s, v_{rms}, v_r, v_{ma}, v_\sigma, v_{sf}, v_{sn}, v_{if}, v_{cf}, v_{clf}, v_{kf}, v_E, v_{pe}]$$
(12)

#### 3.4. Method Verification Experiments

To demonstrate that the GWO-VMD method proposed in Section 3.1 can help signal reconstruction in noise elimination, contrast experiments were conducted and results are reported in this section. GWO-VMD and VMD were both applied to the same experimental signal for decomposition, then the correlation-based signal reconstruction proposed in Section 3.2 was applied to the two decomposition results to rebuild the signals; finally, four performance indicators—Kurtosis, SNR, RMSE, and permutation entropy—of the two reconstructed signals were compared to verify the signal reconstruction effects. The experimental procedure is shown in Figure 2.

Like the method adopted by Li et al. [42], experimental analytic signals were manually constructed to emulate fault vibration signals in a gearbox. The experimental signal, denoted as x(t), consisted of a periodic impulsive signal, denoted as s(t), and a Gaussian white noise with a Signal–Noise Ratio (SNR) of 5 dB, denoted as  $x_n(t)$ . The signals lasted 1 s, the sampling frequency was  $f_s = 2048$  Hz, and the sampling points were  $n_p = 2048$ . The

emulated feature frequency of a fault was  $f_m = 20$  Hz, and the frequency of carrier signal was  $f_1 = 180$  Hz. The expressions of the experimental signals are as follows:

$$\begin{cases} s(t) = 0.8 \exp(-100t_0) \cos(2\pi f_1 t) \\ x(t) = s(t) + x_n(t) \end{cases}$$
(13)

where  $t_0 = \text{mod}(n_p/f_s, 1/f_m)$ , and mod represents module operation.



Figure 2. Contrast experiments of GWO-VMD-based and VMD-based signal decomposition.

The waveforms of the experimental signals and their spectra are shown in Figure 3; the envelope spectra of s(t) and x(t) are shown in Figure 4.



Figure 3. Waveforms of the experimental signals and their spectra.



**Figure 4.** Envelope spectra of experimental signals *s*(*t*) and *x*(*t*).

From Figure 3, it can be seen that the white noise introduced many wide frequency band disturbances to s(t); and from Figure 4, the emulated fault feature frequency 20 Hz and its multi-fold frequencies (40 Hz, 60 Hz, ...) are the main components of envelope spectra. It is easy to find that the noise had obvious effects on the amplitude of these frequency components. For example, in the envelope spectrum of s(t), the emulated fault feature frequency component (20 Hz) has an amplitude of 0.2051, but after the white noise was imposed, the amplitude of the same feature frequency in x(t) became 0.1624; the amplitude decreased by 20.1%. Similar effects can also be found in other frequency components tagged in the figures.

GWO was applied to search optimal values of [K,  $\alpha$ ] for VMD to minimize the objective function shown in Equation (6). The parameter configuration of GWO is shown in Table 3. The evolutionary process of objective function value, K, and  $\alpha$  during the iterations are shown in Figure 5. After the iterations, the obtained optimization results were K=8 and  $\alpha$ =650. With these values, the decomposition results by VMD are shown in Figure 6.

Table 3. Parameter configuration of GWO.

Parameters	Parameter Values or Ranges	
Population number	15	
Maximum iterations number	20	
Range of K	[2, 8]	
Range of $\alpha$	[100, 3000]	



**Figure 5.** Evolutionary process of objective function values, *K*, and  $\alpha$  during the GWO iteration.

From Figure 6, it can be seen that the 8 IMFs had differentiable center frequencies, which were 140 Hz, 180 Hz, 220 Hz, 360 Hz, 480 Hz, 640 Hz, 771 Hz, and 938 Hz, respectively; the mode mixing problem is not obvious in the GWO-VMD-based signal decomposition.

Correlation coefficients, *CORR*, of the above 8 IMFs were calculated, and the results are shown in Table 4. IMFs with correlations higher than 0.25 were chosen; thus, IMF1–IMF4 were picked up for signal reconstruction with Equation (11). The reconstructed experimental signal and its spectrum are shown in Figure 7, and the envelop spectrum of the reconstructed signal is shown in Figure 8.

Table 4. Correlation coefficients of the IMFs.

	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5	IMF 6	IMF 7	IMF 8
CORR	0.550	0.744	0.326	0.266	0.249	0.243	0.240	0.229

Compare Figure 7 with Figures 3 and 8 with Figure 4, respectively; it can be seen that although the reconstructed signal has quite similar frequency component amplitudes as x(t), the amplitudes of the emulated fault feature frequency 20 Hz and its 2-fold and 3-fold frequencies (40 Hz and 60 Hz) in the envelope spectrum were increased in the direction of approaching their original values without noise.

For comparison reasons, VMD was also applied to decompose the experimental signal x(t), and the "central frequency observation method" [43] was adopted to determine the value of K; thus, K = 4 and  $\alpha = 2000$  was used in VMD. After the decomposition, the same procedure was followed to calculate the correlation coefficients for the obtained IMFs, and highly related components were selected to reconstruct the signal. The rebuilt experimental signal and its spectrum are shown in Figure 9, and the envelope spectrum of the reconstructed signal is shown in Figure 10.



Figure 6. IMFs and their spectra decomposed by GWO-VMD.



Figure 7. Reconstructed experimental signal and its spectrum through GWO-VMD.



Figure 8. Envelope spectrum of reconstructed experimental signals through GWO-VMD.

To quantitatively compare the reconstructed signals shown in Figures 8 and 10, four performance indicators—Kurtosis, SNR, RMSE and permutation entropy—were calculated for the two reconstructed signals and are shown in Table 5.

**Table 5.** Characteristic parameter comparison of the experimental signals.

Parameters	s(t)	<i>x</i> ( <i>t</i> )	Reconstructed x(t) through GWO-VMD	Reconstructed x(t) through VMD
Kurtosis	7.7708	5.3719	5.4677	4.7667
Signal–noise ratio	/	5.0021	8.7756	8.1462
RMSE	/	0.1019	0.0660	0.0710
Permutation entropy	0.4545	0.9320	0.6145	0.8036



Figure 9. Reconstructed experimental signal and its spectrum through VMD.



Figure 10. Envelope spectrum of reconstructed experimental signals through VMD.

From Table 5, it can be found that the Kurtosis of the emulated fault feature signal s(t) was 7.7708, but the imposed Gaussian white noise made the Kurtosis value of x(t) decrease to 5.3719, indicating the noise covered the fault and made its feature less sensible. After applying the GWO-VMD-based decomposition and the correlation coefficient-based reconstruction methods, part of the noise was removed and the Kurtosis value increased to 5.4677, higher than that of x(t). With the basic VMD having no parameter optimization, the Kurtosis value of the reconstructed x(t) was 4.7667, lower than 5.4677 and even lower than the value of x(t).

For signal–noise ratio comparison, the SNR of x(t) was 5.0021; because the GWO-VMD-based method eliminated part of the noise from x(t), the SNR of the reconstructed signal increased to 8.7756, while the basic VMD-based method only increased the SNR value of the reconstructed signal to 8.1462, smaller than 8.7756.

For RMSE comparison, smaller root mean squared error value indicates less diffused signals. The RMSE value of x(t) with noise was 0.1019; after noise elimination by GWO-VMD, the RMSE decreased to 0.0660; the VMD-based method only decreased the value to 0.0710, still higher than 0.0660.

For Permutation Entropy (PE), a smaller value means more regular signals. The PE value of the emulated periodic fault impulsive signal s(t) was 0.4545, and due to the noise, the PE of x(t) increased to 0.9320. After the noise elimination by GWO-VMD and signal reconstruction, the PE was reduced to 0.6145, lower than that by VMD-based method.

From the above comparison of the four signal characteristic parameters, it can be seen that GWO-VMD-based signal decomposition and correlation-based signal reconstruction can more effectively eliminate noise from the vibration signal than VMD does.

Furthermore, compare the envelope spectra in Figures 4, 8 and 10; it can be found that the reconstructed signal through GWO-VMD had the closest amplitudes to s(t) at the emulated fault feature frequency and its 2-fold and 3-fold frequencies (20 Hz, 40 Hz, and 60 Hz). This indicates that the GWO-VMD-based method can recover fault feature frequencies from noise better than basic VMD can.

#### 4. DE-KELM-Based Gearbox Fault Classification

With the fault feature vectors obtained by Equation (12), fault classification can be subsequently conducted, and its method is studied in this section.

#### 4.1. Basic Principle of KELM

Kernel extreme learning machines are single-hidden layer feedforward networks developed by Guang-Bin Huang [44]. When compared with other common supervised learning algorithms, such as backpropagation neural network and SVM, KELM has faster training speed and better generalization ability [45]. Therefore, KELM was adopted as a fault classification method in this paper.

Assume an ELM network has *n* inputs (representing *n*-dimensional fault feature vector inputs), *m* outputs (representing *m*-dimensional fault type vector outputs), and *L* nodes in the hidden layer. Suppose *N* distinct training sample pairs ( $x_i$ ,  $t_i$ ) are available, where i = 1, 2, ..., *N*,  $x_i = (x_{i1}, x_{i2}, ..., x_{in})^T \in \mathbb{R}^n$  is the input vector of the *i*<sup>th</sup> sample, and  $t_i = (t_{i1}, t_{i2}, ..., t_{im})^T \in \mathbb{R}^m$  is the target output of  $x_i$ . According to Reference [44], the output vector of KELM is

$$Y = H(\omega, b, x) \cdot \beta \tag{14}$$

where *Y* is the *N* × *m* output matrix of KELM,  $H(\omega, b, x)$  is the *N* × *L* output matrix of the hidden layer and is a function of input weight  $\omega$ , bias vector *b* of the hidden layer, and sample input matrix *x*, while  $\beta$  is the *L* × *m* output weight matrix between hidden and output layers.

The training process of KELM is to minimize the training error and the norm of  $\beta$  at the same time, and it can be formalized as the following optimization problem.

$$\min \frac{1}{2} \|\beta\|^2 + \frac{1}{2} C \sum_{i=1}^{N} \xi_i^2$$
(15)

s.t. 
$$h(x_i)\beta = y_i - \xi_i$$
 (16)

where *C* is the regularization factor, and  $\xi_i$  is the training error.

The weight matrix  $\beta$  can then be calculated with

$$\beta = H^T (H^T H + \frac{I}{C})^{-1} Y \tag{17}$$

If an arbitrary kernel function  $k(x_1, x_2)$  is applied to the above ELM, the inputs can be mapped from a low dimensional space to a higher one to find a linear separation solution. In this case, the output matrix, denoted as f, of KELM is calculated with

$$f = \begin{pmatrix} k(x, x_1) \\ k(x, x_2) \\ \vdots \\ k(x, x_N) \end{pmatrix}^{I} (H^{T}H + \frac{I}{C})^{-1} Y$$
(18)

where  $k(x_1, x_2)$  represents a kernel function. Common kernel functions include linear kernel, polynomial kernel, wavelet kernel, Radial Basis Function (RBF) kernel, etc. In this paper, RBF kernel was adopted with the following expression

$$k(x_1, x_2) = \exp(-\frac{\|x_1 - x_2\|^2}{g})$$
(19)

where *g* is an adjustable kernel parameter.

# 4.2. DE-KELM-Based Fault Classification

The kernel parameter *g* and regularization factor *C* are two important parameters determining training effects and classification accuracy of KELM; thus, their values need to be well determined.

In this section, a Differential Evolution (DE) algorithm, conceptualized by Storn in 1996 [46], was adopted to optimize the values of these two parameters because of its faster convergence speed and better robustness than GA [47].

Like GA, the algorithm of DE consists of two phases: population initialization and evolution. In the latter phase, mutation, crossover, and selection operations are performed. During the selection operation, all vectors in the population are evaluated by a fitness function, and only vectors with high fitness values would be selected to the next generation. These three operations are repeated until a predefined termination condition is met. More detailed iterative process of DE can be found in References [46–48].

When applying DE to optimize the kernel parameter g and regularization factor C of KELM, the workflow is shown in Figure 11.



Figure 11. Workflow of DE-based KELM parameter optimization.

## 4.3. Method Verification Experiments

To demonstrate that the proposed DE-KELM has higher fault classification accuracy than basic KELM without parameter optimization, contrast experiments were conducted on the same experimental data. The experimental procedure is shown in Figure 12.



Figure 12. Contrast experiment of DE-KELM-based and KELM-based fault feature classifications.

The experimental data included bearing vibration signals under a normal condition and three faulty conditions, i.e., inner ring fault, outer ring fault, and roller fault. The data were measured from an experimental platform, illustrated in Figure 13, of the Case Western Reserve University Bearing Data Center [49]. The approximate motor speed range of measurement was 1730–1797 rpm, and the load of motor was from 0 hp to 3 hp. Practical details of this experimental platform can be referred to in [49,50]. One hundred samples were taken randomly from each operating condition to form an experimental dataset of 400 samples. Among this dataset, 90% data of each condition were selected randomly for training and the remaining 10% were used for testing. Accordingly, the training set included 360 samples and the testing set had 40 samples. The 400 samples and their corresponding classification labels are shown in Figure 14.



Figure 13. Experimental platform of Case Western Reserve University Bearing Data Center [49].



Figure 14. 400 samples and their corresponding classification labels.

Extract a feature vector for each sample of the four working conditions according to the method presented in Section 3.3; thereof,  $18 \times 360$  and  $18 \times 40$  feature matrices were obtained to train and test the fault classifier, respectively.

Firstly, DE-KELM was used to realize the classification on the training and testing sets. The parameter configuration of DE is shown Table 6. The optimal regularization coefficient and kernel parameter found by DE were C=12.1929 and g=0.01. With these values, the feature vector classification results by DE-KELM on the training and testing sets are shown in Figure 15. The statistics of the classification results are summarized in Table 7.

Table 6. Parameter configuration of DE.

Parameters	Parameters Values or Ranges	
Population size	9	
Maximum iteration number	30	
Mutation operator	0.7	
Crossover operator	0.6	
Regularization coefficient C	[0.01, 100]	
Kernel parameter g	[0.01, 10]	



Figure 15. Fault classification results by DE-KELM on the training and testing sets.

KELM was then applied to realize fault classification on the same training and testing sets. The regularization coefficient and kernel parameter were set with empirical values of C = 0.2 and g = 1. With these values, the classification results by KELM on the same training and testing sets are shown in Figure 16. The statistics of the classification results are summarized in Table 8.

Recognized as $ ightarrow$	Healthy	Inner Ring Fault	Outer Ring Fault	<b>Roller Fault</b>	Accuracy
Healthy	10	0	0	0	100%
Inner ring fault	0	10	0	0	100%
Outer ring fault	0	0	10	0	100%
Roller fault	0	0	1	9	90%

Table 7. Classification results by DE-KELM of each condition.



Figure 16. Fault classification results by KELM on both training and testing sets.

Table 8. Classification results by KELM of	f each condition.
--	-------------------

Recognized as $ ightarrow$	Healthy	Inner Ring Fault	<b>Outer Ring Fault</b>	<b>Roller Fault</b>	Accuracy
Healthy	10	0	0	0	100%
Inner ring fault	0	10	0	0	100%
Outer ring fault	0	0	10	0	100%
Roller fault	0	1	3	6	60%

From Figures 15 and 16, it can be seen that on the training set, the classification accuracy by DE-KELM was 100% and that by KELM without parameter optimization was only 88.6%. On the testing set, the classification accuracy by DE-KELM was 97.5%, higher than the 90% that was achieved by KELM. From Tables 7 and 8, misclassified sample numbers can be seen. Only one roller fault sample in the testing set was wrongly recognized by DE-KELM, while 4 in the same fault type were misclassified by the basic KELM.

The above contrast experiments verify that KELM with parameter optimized by DE can improve classification accuracy.

# 5. Experimental Validation and Result Analysis

The techniques introduced in Sections 3 and 4 were synthesized, and the complete workflow of the proposed hybrid gearbox fault diagnosis method is as follows.

Step 1: GWO-VMD-based vibration signal decomposition.

Step 2: Correlation-based signal reconstruction.

Step 3: Fault feature calculation.

Step 4: DE-KELM-based fault classification.

The above workflow is illustrated in Figure 17.

In this section, these steps are connected to realize gear and bearing fault diagnosis on experimental datasets.

#### 5.1. Gearbox Fault Diagnosis Experimental Data

The gearbox fault diagnosis experimental datasets were also acquired from the Case Western Reserve University [49]. The data were measured in the same operating conditions as presented in Section 4.3. The datasets included vibration signals of a gearbox in eight operating conditions (healthy and faulty), which are listed in Table 9. The measured vibration signals of these conditions all lasted 75 s, and the sampling frequency was 5120 Hz; thus, in each working condition, the vibration signal included 384,000 sampling points. One datum was taken every 0.5 s as a sample; thereby, 150 samples were obtained in each situation. The sampled vibration signals are plotted and shown in Figure 18.



Figure 17. Complete workflow of the proposed hybrid gearbox fault diagnosis method.

Gearbox Elements	Classification Labels	<b>Operating</b> <b>Conditions</b>	Sample Numbers
	1	healthy bearing	150
Boaring	2	inner ring fault	150
Dearing	3	outer ring fault	150
	4	roller fault	150
	5	healthy gear	150
Gear	6	pitting fault	150
	7	broken teeth	150
	8	tooth root cracks	150

Table 9. Eight working conditions of a gearbox included in the experimental datasets [43].



**Figure 18.** Sampled vibration signals in eight operating conditions. (**a**) Healthy bearing, (**b**) bearing inner ring fault, (**c**) bearing outer ring fault, (**d**) bearing roller fault, (**e**) healthy gear, (**f**) gear pitting fault, (**g**) gear broken tooth fault, (**h**) gear tooth root crack faults.

# 5.2. Contrast Experiment I—Gearbox Fault Diagnosis with Contrasting Vibration Signal Decomposition Methods: GWO-VMD and VMD

To verify that using the GWO-VMD method proposed in this paper to decompose signals can improve the final fault classification accuracy, in this section, GWO-VMD and VMD were both applied to process the same vibration signals. Then, the same signal reconstruction, feature vector calculation, and fault classification methods were followed to diagnose bearing and gear faults. The experimental process is illustrated in Figure 19.



Figure 19. Process of contrast experiment I.

5.2.1. Vibration Signals Decomposition through GWO-VMD and VMD, Respectively

GWO-VMD was applied first to decompose the obtained gearbox vibration signals. The parameters of GWO are shown in Table 10; the optimal values of [K,  $\alpha$ ] obtained by GWO in the eight working conditions are shown in Table 11.

Table 10. Parameter configuration of GWO.

Parameters	Parameter Values or Ranges	
Population number	9	
Maximum iterations number	100	
Mode Number	[2, 15]	
Penalty factor range	[100, 6000]	

**Table 11.** Optimal values of  $[K, \alpha]$  obtained by GWO in the eight working conditions.

Gearbox Elements	Working Conditions	[Κ, α]
Bearing	healthy bearing inner ring fault	[11, 4018] [15, 139]
	roller fault	[5, 2678]
Gear	healthy gear pitting fault broken teeth tooth root cracks	[5, 2330] [15, 100] [5, 3652] [5, 3019]

For comparison, VMD was then applied to decompose the same experimental data. The value of  $\alpha$  was set to 2000, and the value of *K* was determined by the "central frequency observation method" [43]. The calculated values of *K* in the eight working conditions are listed in Table 12 with the value of  $\alpha$ .

With the parameter configurations shown in Tables 11 and 12, the vibration signals in the eight conditions were decomposed by GWO-VMD and VMD, respectively. Then the IMFs having correlation coefficients greater than 0.25 were taken to reconstruct the vibration signals.

<b>Gearbox Elements</b>	Working Conditions	[ <i>K</i> , α]
Bearing	healthy bearing inner ring fault outer ring fault roller fault	[3, 2000] [4, 2000] [4, 2000] [5, 2000]
Gear	healthy gear pitting fault broken teeth tooth root cracks	[4, 2000] [5, 2000] [4, 2000] [4, 2000]

**Table 12.** Values of  $[K \alpha]$  for VMD in the eight working conditions.

5.2.2. Fault Classification Results on the Two Sets of Feature Vectors Obtained through GWO-VMD and VMD, Respectively

After the reconstructed signals were obtained through the decomposition by GWO-VMD and VMD, respectively, 18 feature parameters were calculated on the 150 samples; thus, two  $18 \times 150$  feature matrices were obtained. Among the 150 samples in each condition, 130 were taken randomly to form a training set, and the remaining 20 were used to make up a testing set.

The DE-KELM proposed in this paper was then applied to classify the two feature matrices, respectively. The parameter settings of the DE are shown in Table 6. Fault classification results by DE-KELM on the feature matrices calculated through GWO-VMD and VMD are shown in Figures 20 and 21, respectively. The classification accuracy of the two experiments is shown in Table 13.



**Figure 20.** Fault classification results by DE-KELM on the feature matrix calculated through GWO-VMD-based decomposition method.

Table 13. Fault classification accuracy of the training and testing sets by DE-KELM.

Signal Decomposition Method	Classification Accuracy of the Training Set (by DE-KELM)	Classification Accuracy of the Testing Set (by DE-KELM)
GWO-VMD	100%	98.125%
VMD	100%	91.875%

As can be seen from Figure 20 and Table 13, when the vibration signal was processed by GWO-VMD, DE-KELM classified all the 1040 training samples correctly and the classification accuracy was 100%; for the 160 testing samples, only 3 of them were misclassified, and the classification accuracy was 98.125%.



**Figure 21.** Fault classification results by DE-KELM on the feature matrix calculated through VMDbased decomposition method.

From Figure 21 and Table 13, when the vibration signal was processed by the contrasting VMD method, for the 1040 training samples, DE-KELM also correctly classified with 100% accuracy; while for the 160 testing samples, 13 were misclassified, and the classification accuracy was 91.875%.

From the above two contrast experiments, it can be seen that when GWO-VMD was used to decompose the signal, after which the same signal reconstruction and feature extraction were conducted, the obtained fault features can improve the fault diagnosis accuracy under the same fault classification method.

# 5.3. Contrast Experiment II—Fault Classification by DE-KELM and KELM, Respectively, on the Same Fault Feature Vectors

To demonstrate that the proposed DE-KELM fault classification method can achieve higher diagnosis accuracy than KELM, in this section, the same fault features obtained in Section 5.2 through GWO-VMD decomposition were used to test DE-KELM and KELM, respectively. The contrasting experimental process is illustrated in Figure 22.





The training and testing sets were divided in the same way as in the previous experiment. The classification results by DE-KELM are presented in Figure 20, the results by KELM on the same training and testing sets are shown in Figure 23, and the accuracy of the two classification methods is summarized in Table 14.

As can be seen from Figure 23 and Table 14, for the 160 testing samples, 4 of them were misclassified by KELM, and the classification accuracy on the testing set was 97.5%, which is lower than 98.125%, the accuracy obtained by DE-KELM.

The above experimental results show that the proposed DE-KELM method can achieve higher classification accuracy on the same fault features than KELM.



**Figure 23.** Fault classification results by KELM on the feature matrix calculated through GWO-VMDbased decomposition method.

Fault Classification Methods	Classification Accuracy of the Training Set	Classification Accuracy of the Testing Set
DE-KELM	100%	98.125%
KELM	97.9808%	97.5%

# 6. Conclusions and Discussion

In this paper, we proposed a vibration signal-based hybrid method for realizing gearbox fault diagnosis and carried out experiments to verify its effectiveness. The main contribution of the proposed method is the combination of signal processing, machine learning, and optimization techniques to effectively eliminate noise contained in vibration signals and to achieve high diagnostic accuracy. This hybrid method includes GWO-VMD-based vibration signal decomposition, correlation-based signal reconstruction, fault feature extraction, and DE-KELM-based gearbox fault classification.

To eliminate as much noise as possible from the vibration signal and to alleviate the mode mixing problem during signal decomposition, VMD was adopted for vibration signal decomposition and GWO was used to optimize its [K,  $\alpha$ ] by minimizing the objective function shown in Equation (6); after IMFs were obtained, the correlation coefficient of each IMF was calculated to select highly correlated components for signal reconstruction. Method validation experiments were conducted and reported in Section 3 to demonstrate that the reconstructed signal by the proposed method eliminates more noise and has better Kurtosis, signal–noise ratio, RMSE, and permutation entropy values than those by the basic VMD method.

With the rebuilt vibration signals, 18 feature parameters were calculated for each sample to form fault feature vectors. Then, KELM was used to classify the features with its parameters optimized by the DE algorithm for higher diagnostic accuracy. Method validation experiments were conducted and reported in Section 4 to demonstrate that the proposed DE-KELM can achieve higher classification accuracy than basic KELM.

Finally, the gearbox vibration signals in eight different operating conditions were obtained and contrast experiments were conducted to verify that the proposed GWO-VMD-and DE-KELM-based gearbox fault diagnosis method can achieve higher diagnostic accuracy than basic VMD and KELM.

Compared with other available hybrid fault diagnosis methods in the literature survey, for instance, the two methods achieved interesting research results in [25,26]; the method proposed in this paper adopts the parameter-optimized time-frequency analysis technique (GWO-VMD), which can reflect more characteristics of the non-stationary vibration signal in the signal preprocessing and feature extraction stage. This method can also alleviate the

mode mixing problem included in EMD and its varieties. In the fault classification stage, the proposed DE-KELM method used KELM as a fault classifier, which can achieve faster training speed than SVM-based methods.

Although the effectiveness of the proposed method was demonstrated with the experiments in Sections 3–5, there are also limitations in the proposed hybrid diagnosis method-the introduction of optimization techniques to VMD and KELM leads to longer overall signal processing and fault classification time. In the experiments of Sections 3 and 5, it took 5–10 min for DE to search optimal values for [K,  $\alpha$ ]. This deficiency makes the proposed hybrid method not applicable to real-time diagnosis. Therefore, improving the real-time capability of the proposed hybrid method will be one of the future research goals of the authors. Concerning the practical applications of the proposed method, it can be applied to offline and high-precision fault diagnosis of gearboxes operating in scenarios with rich ambient noise, such as wind generation systems, rail transports, etc. Additionally, the authors believe that the method is also applicable to other diagnostic problems when the measured signal contains noise—rotor fault diagnosis of electrical motors, for example. This will be verified in subsequent research related to this report. Furthermore, considering the complexity and diversity of gearbox failures, combining the proposed method of this research report with transfer learning techniques and applying them to diagnose faults with limited samples will also be an important research goal to be achieved, and this hybrid method will be compared with model-based gearbox diagnosis methods to verify its effectiveness.

**Author Contributions:** Methodology and writing, G.Y.; software, Y.W.; validation, M.A.-A.; supervision, M.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Shanghai Science and Technology Committee Foundation with Grant No. 19040501700.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

**Data Availability Statement:** The experimental data in Sections 4 and 5 can be downloaded from the homepage of the Case Western Reserve University Bearing Data Center [49] with the following weblink: https://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website, accessed on 1 May 2021. These data are also available from the correspondence author of this article.

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

#### Nomenclature

Acronyms or Variables	Connotations
VMD	Variational Mode Decomposition
GWO	Wolf Grey Optimizer
KELM	Kernel Extreme Learning Machine
DE	Differential Evolutionary
IMF	Intrinsic Mode Function
TFA	Time-Frequency Analysis
STFT	Short-Time Fourier Transform
WVD	Wigner-Ville Distribution
DWT	Discrete Wavelet Transform
EMD	Empirical Mode Decomposition
PSO	Particle Swarm Optimization
EEMD	Ensemble Empirical Mode Decomposition
CNN	Convolution Neural Network
DRN	Deep Residual Network
ELM	Extreme Learning Machine
TL	Transfer Learning

SVM	Support Vector Machine
NN	Neural Network
DL	Deep Learning
SNR	Signal-Noise Ratio
PE	Permutation Entropy
GA	Genetic Algorithm
$\overline{v}(t)$	Gearbox vibration signal
(t)	The $k^{th}$ IME of $v(t)$
$\mathcal{U}_{k}(l)$	Number of imfe
	Frequency contars of the $k^{th}$ IME (Hz)
w <sub>k</sub>	Penalty factor in VMD algorithm
a	Lagran as multiplicer in VMD
Λ fra	Eitness function of CMO
jn	Kunteria in CINO fita and function
KI	Rurtosis in GWO fitness function
нр	Permutation entropy in GWO fitness function
μ	Mean value for Kurtosis calculation
$\sigma$	Standard deviation for Kurtosis calculation
<i>m</i> <sub>1</sub> , <i>q</i>	Dimensional parameters of the reconstruction matrix for <i>Hp</i> calculation
$p_i$	Probability of each row of the reconstruction matrix
alpha, beta, delta, omega	Grey wolf names in their social hierarchy
Corr	Correlation coefficient
Corr <sub>th</sub>	Correlation coefficient threshold for signal reconstruction
v'(t)	Reconstructed gearbox vibration signal
$v_{max}$	Maximum amplitude of $v'(t)$
$v_{min}$	Minimum amplitude of $v'(t)$
$v_p$	Peakedness of $v'(t)$
$v_{pp}$	Peak-to-peak value of $v'(t)$
$v_m$	Mean value of $v'(t)$
$v_s$	Slant of $v'(t)$
v <sub>rms</sub>	Root mean square value of $v'(t)$
<i>v</i> <sub>r</sub>	Root square amplitude of $v'(t)$
$v_{ma}$	Mean of absolute value of $v'(t)$
$v_{\sigma}$	Variance of $v'(t)$
$v_{sf}$	Shape factor of $v'(t)$
v <sub>sn</sub>	Skewness of $v'(t)$
<i>v<sub>if</sub></i>	Impulse factor of $v'(t)$
v <sub>cf</sub>	Crest factor of $v'(t)$
V <sub>clf</sub>	Clearance factor of $v'(t)$
Vkf	Kurtosis factor of $v'(t)$
τη τη	Energy of $v'(t)$
Una Dina	Permutation entropy of $v'(t)$
$N_{\tau}$	Length of $v'(t)$
T	Fault feature vector
x(t)	Constructed experimental signal with noise
s(t)	Periodic impulsive signal for experiments
$r_{\rm el}(t)$	Cause white noise for experiments
<i>x<sub>n</sub>(t)</i>	KELM input node number
л I	KELM hidden laver node number
L 111	KELM nudden nyer node number
N	Number of KELM training sample pairs
Γ <b>ν</b> γ.	Input vector of the <i>i</i> <sup>th</sup> KEI M sample
$x_1$	Target output of r.
$\frac{1}{\gamma}$	$N \times m$ output matrix of KELM
і Ц	$N \times I$ output matrix of KELM hidden layer
11	$N \wedge L$ output matrix of NELW fillen layer
w h	RELM IIIPUT WEIGITI IIIdulix Bias vastar of KEI M hidden lavor
U	Dias vector of KELIVI fildden layer
x e	NELIVI sample input matrix
p	$L \times m$ output weight matrix of KELM
L	Regularization factor of KELM

$\xi_i$	Training error of KELM
$k(x_1, x_2)$	Kernel function of KELM
f	Output function of KELM
8	Kernel parameter of RBF kernel

#### References

- 1. Ruan, H.; Wang, Y.; Li, X.; Qin, Y.; Tang, B.; Wang, P. A Relation-Based Semisupervised Method for Gearbox Fault Diagnosis With Limited Labeled Samples. *IEEE Trans. Instrum. Meas.* **2021**, 70. [CrossRef]
- Jin, X.; Cheng, F.; Peng, Y.; Qiao, W.; Qu, L. Drivetrain Gearbox Fault Diagnosis VIBRATION-AND CURRENT-BASED AP-PROACHES. IEEE Ind. Appl. Mag. 2018, 24, 56–66. [CrossRef]
- Chen, X.; Feng, Z. Time-Frequency Analysis of Torsional Vibration Signals in Resonance Region for Planetary Gearbox Fault Diagnosis Under Variable Speed Conditions. *IEEE Access* 2017, 5, 21918–21926. [CrossRef]
- Zhong, D.; Guo, W.; He, D. An Intelligent Fault Diagnosis Method based on STFT and Convolutional Neural Network for Bearings Under Variable Working Conditions. In Proceedings of the 2019 Prognostics and System Health Management Conference (PHM-Qingdao), Qingdao, China, 25–27 October 2019; pp. 1–6. [CrossRef]
- 5. Yao, J.; Zhao, J.; Deng, Y.; Langari, R. Weak Fault Feature Extraction of Rotating Machinery Based on Double-Window Spectrum Fusion Enhancement. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 1029–1040. [CrossRef]
- 6. Yang, B.; Liu, R.; Chen, X. Sparse Time-Frequency Representation for Incipient Fault Diagnosis of Wind Turbine Drive Train. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 2616–2627. [CrossRef]
- Heydarzadeh, M.; Kia, S.H.; Nourani, M.; Henao, H.; Capolino, G. Gear fault diagnosis using discrete wavelet transform and deep neural networks. In Proceedings of the IECON 2016—42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy, 23–26 October 2016; pp. 1494–1500. [CrossRef]
- 8. Zhao, M.; Kang, M.; Tang, B.; Pecht, M. Multiple Wavelet Coefficients Fusion in Deep Residual Networks for Fault Diagnosis. *IEEE Trans. Ind. Electron.* **2019**, *66*, 4696–4706. [CrossRef]
- 9. Li, Y.; Xu, M.; Liang, X.; Huang, W. Application of Bandwidth EMD and Adaptive Multiscale Morphology Analysis for Incipient Fault Diagnosis of Rolling Bearings. *IEEE Trans. Ind. Electron.* **2017**, *64*, 6506–6517. [CrossRef]
- 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]
- 11. Wang, X.-B.; Yang, Z.-X.; Yan, X.-A. Novel Particle Swarm Optimization-Based Variational Mode Decomposition Method for the Fault Diagnosis of Complex Rotating Machinery. *IEEE-ASME Trans. Mechatron.* **2018**, *23*, 68–79. [CrossRef]
- Guo, Z.; Liu, M.; Wang, Y.; Qin, H. A New Fault Diagnosis Classifier for Rolling Bearing United Multi-Scale Permutation Entropy Optimize VMD and Cuckoo Search SVM. *IEEE Access* 2020, *8*, 153610–153629. [CrossRef]
- 13. Wang, C.; Li, H.; Huang, G.; Ou, J. Early Fault Diagnosis for Planetary Gearbox Based on Adaptive Parameter Optimized VMD and Singular Kurtosis Difference Spectrum. *IEEE Access* 2019, *7*, 31501–31516. [CrossRef]
- 14. Li, F.; Li, R.; Tian, L.; Chen, L.; Liu, J. Data-driven time-frequency analysis method based on variational mode decomposition and its application to gear fault diagnosis in variable working conditions. *Mech. Syst. Signal Process.* **2019**, *116*, 462–479. [CrossRef]
- 15. Gao, S.; Pei, Z.; Zhang, Y.; Li, T. Bearing Fault Diagnosis Based on Adaptive Convolutional Neural Network With Nesterov Momentum. *IEEE Sens. J.* 2021, 21, 9268–9276. [CrossRef]
- 16. Jiang, G.; He, H.; Yan, J.; Xie, P. Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox. *IEEE Trans. Ind. Electron.* **2019**, *66*, 3196–3207. [CrossRef]
- 17. Jin, G.; Zhu, T.; Akram, M.W.; Jin, Y.; Zhu, C. An Adaptive Anti-Noise Neural Network for Bearing Fault Diagnosis Under Noise and Varying Load Conditions. *IEEE Access* 2020, *8*, 74793–74807. [CrossRef]
- Chen, L.; Xu, G.; Tao, T.; Wu, Q. Deep Residual Network for Identifying Bearing Fault Location and Fault Severity Concurrently. *IEEE Access* 2020, *8*, 168026–168035. [CrossRef]
- 19. Saufi, S.R.; Bin Ahmad, Z.A.; Leong, M.S.; Lim, M.H. Gearbox Fault Diagnosis Using a Deep Learning Model With Limited Data Sample. *IEEE Trans. Ind. Inform.* 2020, *16*, 6263–6271. [CrossRef]
- 20. Zhong, J.-H.; Zhang, J.; Liang, J.; Wang, H. Multi-Fault Rapid Diagnosis for Wind Turbine Gearbox Using Sparse Bayesian Extreme Learning Machine. *IEEE Access* 2019, *7*, 773–781. [CrossRef]
- Udmale, S.S.; Singh, S.K.; Singh, R.; Sangaiah, A.K. Multi-Fault Bearing Classification Using Sensors and ConvNet-Based Transfer Learning Approach. *IEEE Sens. J.* 2020, 20, 1433–1444. [CrossRef]
- 22. Zhang, X.; Han, P.; Xu, L.; Zhang, F.; Wang, Y.; Gao, L. Research on Bearing Fault Diagnosis of Wind Turbine Gearbox Based on 1DCNN-PSO-SVM. *IEEE Access* 2020, *8*, 192248–192258. [CrossRef]
- Zhu, J.; Chen, N.; Shen, C. A New Deep Transfer Learning Method for Bearing Fault Diagnosis Under Different Working Conditions. *IEEE Sens. J.* 2020, 20, 8394–8402. [CrossRef]
- 24. Yuan, L.; Lian, D.; Kang, X.; Chen, Y.; Zhai, K. Rolling Bearing Fault Diagnosis Based on Convolutional Neural Network and Support Vector Machine. *IEEE Access* 2020, *8*, 137395–137406. [CrossRef]
- 25. Glowacz, A.; Glowacz, W.; Kozik, J.; Piech, K.; Gutten, M.; Caesarendra, W.; Liu, H.; Brumercik, F.; Irfan, M.; Khan, Z.F. Detection of Deterioration of Three-phase Induction Motor using Vibration Signals. *Meas. Sci. Rev.* **2019**, *19*, 241–249. [CrossRef]

- 26. Sun, S.; Przystupa, K.; Wei, M.; Yu, H.; Ye, Z.; Kochan, O. Fast bearing fault diagnosis of rolling element using Levy Moth-Flame optimization algorithm and Naive Bayes. *Eksploat. Niezawodn. Maint. Reliab.* **2020**, *22*, 730–740. [CrossRef]
- Yang, Z.-X.; Wang, X.-B.; Wong, P.K. Single and Simultaneous Fault Diagnosis With Application to a Multistage Gearbox: A Versatile Dual-ELM Network Approach. *IEEE Trans. Ind. Inform.* 2018, 14, 5245–5255. [CrossRef]
- 28. Ren, F.; Qin, D. Investigation of the effect of manufacturing errors on dynamic characteristics of herringbone planetary gear trains. In Proceedings of the International Gear Conference 2014, Lyon, France, 26–28 August 2014; pp. 230–239. [CrossRef]
- 29. Breeze, P. Chapter 5—Drive Trains, Gearboxes, and Generators. In *Wind Power Generation*, 1st ed.; Breeze, P., Ed.; Academic Press: Salt Lake City, UT, USA, 2016; pp. 41–48. ISBN 9780128040386. [CrossRef]
- Wang, C.; Li, H.; Ou, J.; Huang, G. Early Weak Fault Diagnosis of Gearbox Based on ELMD and Singular Value Decomposition. In Proceedings of the 2019 Prognostics and System Health Management Conference (PHM-Qingdao), Qingdao, China, 25–27 October 2019; pp. 1–5. [CrossRef]
- Henriquez, P.; Alonso, J.B.; Ferrer, M.A.; Travieso, C.M. Review of Automatic Fault Diagnosis Systems Using Audio and Vibration Signals. *IEEE Trans. Syst. Man Cybern. Syst.* 2014, 44, 642–652. [CrossRef]
- 32. Ming, Q.; Chen, L.; Li, Y.; Yan, J. Fault Diagnosis and Status Monitoring of the Bearing. In *Bearing Tribology*, 1st ed.; Ming, Q., Chen, L., Li, Y., Yan, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2017; pp. 239–306. ISBN 978-3-662-53095-5. [CrossRef]
- 33. Yu, J.; Zhou, X. One-Dimensional Residual Convolutional Autoencoder Based Feature Learning for Gearbox Fault Diagnosis. *IEEE Trans. Ind. Inform.* 2020, *16*, 6347–6358. [CrossRef]
- 34. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. IEEE Trans. Signal Process. 2014, 62, 531–544. [CrossRef]
- 35. Ma, T.; Zhang, X.; Jiang, H.; Wang, K.; Xia, L.; Guan, X. Early Fault Diagnosis of Shaft Crack Based on Double Optimization Maximum Correlated Kurtosis Deconvolution and Variational Mode Decomposition. *IEEE Access* **2021**, *9*, 14971–14982. [CrossRef]
- Liu, H.; Xiang, J. A Strategy Using Variational Mode Decomposition, L-Kurtosis and Minimum Entropy Deconvolution to Detect Mechanical Faults. *IEEE Access* 2019, 7, 70564–70573. [CrossRef]
- Lahmiri, S. Comparing Variational and Empirical Mode Decomposition in Forecasting Day-Ahead Energy Prices. *IEEE Syst. J.* 2017, 11, 1907–1910. [CrossRef]
- 38. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey Wolf Optimizer. Adv. Eng. Softw. 2014, 69, 46–61. [CrossRef]
- Alsadie, D. TSMGWO: Optimizing Task Schedule Using Multi-Objectives Grey Wolf Optimizer for Cloud Data Centers. *IEEE Access* 2021, 9, 37707–37725. [CrossRef]
- 40. Chen, X.; Yang, Y.; Cui, Z.; Shen, J. Wavelet Denoising for the Vibration Signals of Wind Turbines Based on Variational Mode Decomposition and Multiscale Permutation Entropy. *IEEE Access* **2020**, *8*, 40347–40356. [CrossRef]
- 41. Ahmed, H.; Nandi, A.K. Compressive Sampling and Feature Ranking Framework for Bearing Fault Classification With Vibration Signals. *IEEE Access* 2018, *6*, 44731–44746. [CrossRef]
- 42. Li, B.; Zhang, P.; Liu, D.; Mi, S.; Ren, G. Feature extraction for roller bearing fault diagnosis based on adaptive multi-scale morphological gradient transformation. *J. Vib. Shock* **2011**, *30*, 104–108. [CrossRef]
- Shao, S.; McAleer, S.; Yan, R.; Baldi, P. Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning. *IEEE Trans. Ind. Inform.* 2019, 15, 2446–2455. [CrossRef]
- Huang, G.-B.; Zhou, H.; Ding, X.; Zhang, R. Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Trans.* Syst. Man Cybern. Part B-Cybern. 2012, 42, 513–529. [CrossRef]
- Song, J.; Zheng, Z.; Liu, Y.; Guo, X.; Zhou, J. Wavelet Packet-based Kernel Extreme Learning Machine for Sensor Faults Diagnosis of Hypersonic Vehicle. In Proceedings of the 2018 Chinese Automation Congress (CAC), Xian, China, 30 November–2 December 2018; pp. 2017–2022. [CrossRef]
- 46. Storn, R. On the usage of differential evolution for function optimization. In Proceedings of the North American Fuzzy Information Processing, Berkeley, CA, USA, 19–22 June 1996; pp. 519–523. [CrossRef]
- Nam, P.; Malinowski, A.; Bartczak, T. Comparative Study of Derivative Free Optimization Algorithms. *IEEE Trans. Ind. Inform.* 2011, 7, 592–600. [CrossRef]
- 48. Bilal, M.; Pant, M.; Zaheer, H.; Garcia-Hernandez, L.; Abraham, A. Differential Evolution: A review of more than two decades of research. *Eng. Appl. Artif. Intell.* 2020, 90. [CrossRef]
- The Case Western Reserve University Bearing Data Center. Available online: https://csegroups.case.edu/bearingdatacenter/ pages/welcome-case-western-reserve-university-bearing-data-center-website (accessed on 28 April 2021).
- 50. Khamoudj, C.E.; Tayeb, F.B.-S.; Benatchba, K.; Benbouzid, M.; Djaafri, A. A Learning Variable Neighborhood Search Approach for Induction Machines Bearing Failures Detection and Diagnosis. *Energies* **2020**, *13*, 2953. [CrossRef]