

Article

# Compound Fault Diagnosis of Wind Turbine Gearbox via Modified Signal Quality Coefficient and Versatile Residual Shrinkage Network

Weixiong Jiang <sup>1</sup> , Guanhui Zhao <sup>2</sup>, Zhan Gao <sup>1</sup>, Yuanhang Wang <sup>3,\*</sup> and Jun Wu <sup>1,\*</sup> 

<sup>1</sup> School of Naval Architecture and Ocean Engineering, Huazhong University of Science and Technology, Wuhan 430074, China; jiangweixiong@hust.edu.cn (W.J.); z\_gao@hust.edu.cn (Z.G.)

<sup>2</sup> College of Computer Science and Technology, Zhejiang University, Hangzhou 310058, China; 12321305@zju.edu.cn

<sup>3</sup> Sino-German College of Intelligent Manufacturing, Shenzhen Technology University, Shenzhen 518118, China

\* Correspondence: wangyuanhang@sztu.edu.cn (Y.W.); wuj@hust.edu.cn (J.W.)

**Abstract:** Wind turbine gearbox fault diagnosis is critical to guarantee working efficiency and operational safety. However, the current diagnostic methods face enormous restrictions in handling nonlinear noise signals and intricate compound fault patterns. Herein, a compound fault diagnosis method based on modified signal quality coefficient (MSQC) and versatile residual shrinkage network (VRSN) is proposed to resolve these issues. In detail, the MSQC is designed to remove the noise components irrelevant to wind turbine operation status, and it has the ability to balance the denoised effect and signal fidelity. The VRSN is constructed for compound fault diagnosis, and it consists of two heterogeneous residual shrinkage networks. The former is designed to count the number of faults, and the latter is adopted to identify the single or compound fault pattern. Finally, a self-built wind turbine gearbox compound fault test rig is adopted to verify the proposed method's effectiveness. The results demonstrate that the proposed method is competitive in terms of compound fault diagnosis accuracy.

**Keywords:** compound fault diagnosis; versatile residual shrinkage network; modified signal quality coefficient; wind turbine gearbox



Academic Editor: Ruben Puche-Panadero

Received: 29 December 2024

Revised: 25 January 2025

Accepted: 29 January 2025

Published: 3 February 2025

**Citation:** Jiang, W.; Zhao, G.; Gao, Z.; Wang, Y.; Wu, J. Compound Fault Diagnosis of Wind Turbine Gearbox via Modified Signal Quality Coefficient and Versatile Residual Shrinkage Network. *Sensors* **2025**, *25*, 913. <https://doi.org/10.3390/s25030913>

**Copyright:** © 2025 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/>).

## 1. Introduction

Due to high energy transmission efficiency and strong power output, wind turbines are of great importance when used in distributed generator infrastructures [1–3]. However, due to the intricate internal structure and harsh working environment, the malfunctions are prone to occur in the key component of a wind turbine. When 6312 wind turbines are surveyed, the distributed reliability working group of the institute of electrical and electronics engineering (DRWG-IEEE) reports that half of all faults derive from the wind turbine gearbox, such as lacking teeth, broken teeth, and bearing inner ring wear. Different faults may even happen at once [4–6]. Thus, it is an urge task to develop a fault diagnosis method for wind turbine gearboxes, especially for compound faults.

Many academics have made great efforts to develop a binary combination strategy and a probabilistic-based network for compound fault diagnosis for decades. In the binary combination strategy, multiple binary classifiers are integrated under the 1-versus-1 or 1-versus-all tactics.  $m(m - 1)/2$  groups of binary classifiers produced by the 1-versus-1 tactic or  $m$  groups of binary classifiers produced by the 1-versus-all tactic are adopted for the

$m$ -label classification problem [7]. However, there are several glaring limitations preventing the implementation of the combination strategy, such as the huge computation cost and complex fine-tuning process. In a probabilistic-based network, the Bayesian network is used to compute the fault probability distribution for the observed machine. However, the network decision thresholds and standard training data must be prepared with human intervention in advance, thus they are impractical in real-world applications [8].

Recently, algorithm adaptation methods have been explored to address the above issues. Clare et al. proposed the multilabel decision tree method based on a multilabel entropy and decision tree to realize multilabel compound fault diagnosis [9]. Zhang et al. constructed the multiclass  $k$  nearest neighbor model based on the  $k$  nearest algorithm and maximum posterior theory [10]. Tahir et al. established the rank support vector machine based on the maximum margin theory to update a set of linear classifiers, which can handle the multiclass nonlinear problem when the empirical rank error is at a minimum [11]. Liu et al. constructed a classifier chain to explore the label's interior relationship, but this fails to implement multithread operation because of its chain structure [12]. Wang et al. refined the label power-set by use of random  $k$  labelsets to propose the random  $k$  labelsets (RAKEL) for classification efficiency improvement. However, the coupling of a homogeneous-component multilabel classifier in RAKEL may impact the classification performance [13]. Thus, the unreasonable model structure and computational resource configuration will lead to inferior diagnostic accuracy and efficiency.

On the other hand, the original collected signals are usually nonstationary, nonlinear, and prone to be disturbed by the environment background noises. It is crucial for the preprocessing technique to remove noise components to reveal fault-related features. The current signal preprocessing methods mainly focus on multimodal signal fusion [14–16], high-resolution signal decomposition [17–19], and end-to-end feature extraction techniques [20–22]. However, it is still intractable to process the nonlinear noisy data by use of the traditional denoising technique. In addition, diverse fault types and compound fault patterns further intensify the challenges of signal preprocessing.

To resolve the limitations of the above methods, the compound fault diagnosis method is proposed based on the modified signal quality coefficient (MSQC) and versatile residual shrinkage network (VRSN) for a wind turbine gearbox. The MSQC is designed to detect and remove the noise components irrelevant to the wind turbine's operation status. Then, the VRSN is established for compound fault diagnosis, and it consists of two heterogeneous residual shrinkage networks used for the fault count and fault probability distribution calculation. The main contributions of the paper are as follows:

- (1) The VRSN is proposed to diagnose compound faults in a wind turbine gearbox. Different from the probabilistic-based method, the proposed network is self-adaptive, and can identify single or simultaneous faults without manual intervention for empirical threshold setting;
- (2) The multithread network structure is constructed to optimize the computation resources' configuration. Two parallel residual shrinkage networks can be implemented simultaneously to count the fault numbers and determine the fault probability distribution in responding to the real-time fault diagnosis task;
- (3) The denoised algorithm is designed to remove the noise components irrelevant to wind turbine operation status. The modified signal quality coefficient has the ability to balance the denoised effect and signal fidelity, and fault-sensitive features hidden in the originally collected signals can be captured precisely.

The remaining of this paper is organized as follows. The basic principle is introduced in Section 2. The proposed method is elaborated in Section 3. The self-built wind turbine

gearbox compound fault simulation test rig is described in Section 4. The denoised and diagnostic results are discussed in Section 5. Eventually, conclusions are drawn in Section 6.

## 2. Theoretical Background

### 2.1. Deep Residual Shrinkage Network

The residual network (ResNet) has an excellent classification ability because it can avoid the vanishing gradient and overfitting phenomena produced by model error back-propagation in the identical path. Based on this, the deep residual shrinkage network (DRSN) was proposed, introducing the soft threshold function and attention mechanism into the ResNet. It can adaptively eliminate the noise-related features to improve the model classification's performance. The soft threshold function is added to eliminate the noise-related data further, as follows:

$$f(x_{de}) = \begin{cases} x_{de} - \tau, & x_{de} > \tau \\ 0, & -\tau \leq x_{de} \leq \tau \\ x_{de} + \tau, & x_{de} < -\tau \end{cases} \quad (1)$$

where  $x_{de}$  is the denoised data,  $f(x_{de})$  is the processed feature, and  $\tau$  is the soft threshold, which can keep the prominent data stable and transform noise-related data to zero. The soft threshold is adaptively calculated by the residual shrinkage building unit (RSBU) [23].

The Sigmoid function is embedded at the end of RSBU as the output layer of DRSN, as

$$\alpha_C = \frac{1}{1 + e^{-z_C}} \quad (2)$$

where  $Z_c$  is the output of the fully connected layer, and  $\alpha_C$  is the channel scaling value. The channel soft threshold is determined as

$$\tau_C = \alpha_C \cdot \underset{w,h}{\text{average}} |x_{W,H,C}| \quad (3)$$

where  $W$ ,  $H$ , and  $C$  denote the width, height, and channel indexes of the feature map  $x$ .

### 2.2. ICEEMDAN

The improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN), which is the extended version, can address the residual noise and spurious mode problems that arise with the CEEMDAN algorithm [24]. The principle of the ICEEMDAN algorithm is elaborated in the following.

The Gaussian white noise  $w^{(i)}$  is decomposed by the empirical mode decomposition (EMD), and it is added into the raw signal  $x^{\text{raw}}$  to construct the series.

$$x^{(i)} = x + \beta_{(i-1)} E(w^{(i)}), \quad i = 1, 2, \dots, I \quad (4)$$

where  $x$  is the raw signal,  $x^{(i)}$  is the constructed series,  $E_k(\cdot)$  is the  $k$ th order intrinsic mode function (IMF) decomposed by the EMD algorithm, and  $\beta_{(i-1)} > 0$ .

The first residual component is calculated, and the first mode is given as

$$d_1 = x - R_1 = x - \left( N(x^{(i)}) \right) \quad (5)$$

where  $d_1$  is the first mode,  $R_1$  is the first residual component, and  $N(\cdot)$  is the local mean.

The Gaussian white noise is added again. The second residual component is calculated, and the second mode is determined as

$$d_2 = R_1 - R_2 = R_1 - \left( N \left( R_1 + \beta_1 E \left( w^{(i)} \right) \right) \right) \quad (6)$$

where  $d_2$  is the second mode and  $R_2$  is the second residual component.

The  $k$ th residual component is calculated, and the  $k$ th mode is expressed as

$$d_k = R_{k-1} - R_k = R_{k-1} - \left( N \left( R_{k-1} + \beta_{k-1} E \left( w^{(i)} \right) \right) \right) \quad (7)$$

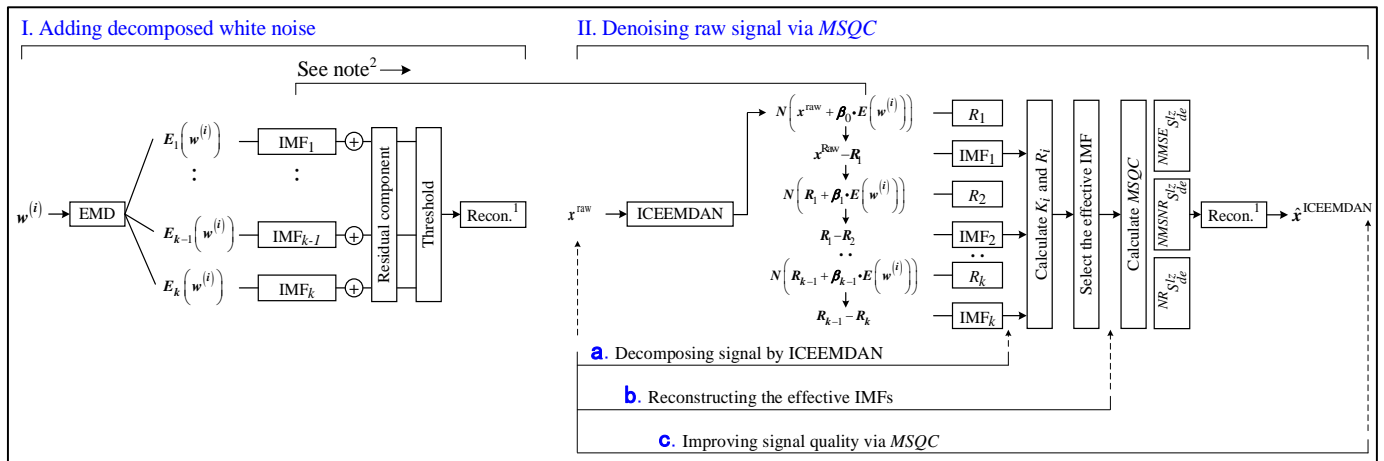
Finally, all modes and residual components are aggregated into the reconstructed signal as

$$x_a(t) = \sum_{i=1}^k d_i + R_k, i = 1, 2, \dots, k \quad (8)$$

### 3. Methodology

#### 3.1. Modified Signal Quality Coefficient

The modified signal quality coefficient (MSQC) is designed to balance the signal fidelity and noise reduction, as exhibited in Figure 1. The specific processes are elaborated as follows.



**Figure 1.** Methodology of modified signal quality coefficient. (Notes: <sup>1</sup> reconstruction; <sup>2</sup> decomposing Gaussian white noise by EMD, and adding it into the raw signal).

Step 1: The ICEEMDAN is adopted to calculate the IMF and residual components of the raw signal.

Step 2: The effective IMF component number  $NE$  is determined from 1 to  $k/2$ , and it depends on the raw signal complexity [25].

Step 3: The Pearson correlation coefficient  $R_i$  and kurtosis index  $K_i$  between the effective IMF components and raw signal are calculated. The impulse signal contained in heavy background noise can be detected by the kurtosis index, and the relevance between the raw signal and the effective IMF components can be reflected by the Pearson correlation coefficient. The above indicators are adopted to remove the irrelevant components.

$$R_i = \frac{\sum_{t=1}^T \left( d_k(t) - \overline{d_k(t)} \right) \left( x^{(i)}(t) - \overline{x^{(i)}(t)} \right)}{\sqrt{\sum_{t=1}^T \left( d_k(t) - \overline{d_k(t)} \right)^2 \left( x^{(i)}(t) - \overline{x^{(i)}(t)} \right)^2}} \quad (9)$$

$$K_i = \frac{T \sum_{t=1}^T \left( \text{IMF}_i(t) - \overline{\text{IMF}_i(t)} \right)^4}{\left( \sum_{t=1}^T \left( \text{IMF}_i(t) - \overline{\text{IMF}_i(t)} \right)^2 \right)^2} \quad (10)$$

where  $\overline{d_k(t)}$  and  $\overline{x^{(i)}(t)}$  are the averages of  $d_k(t)$  and  $x^{(i)}(t)$ .

Step 4: After selecting the effective IMF components, the reconstructed signal  $x_{bg}^{lz,(i)}(t)$  and the IMF number record matrix  $xm_{de}^{lz,(i)}(t)$  are determined, where  $l$  is the IMF number in  $x_{de}^{lz,(i)}(t)$ ,  $z$  is the reconstructed signal number,  $b$  is the first  $b$  IMFs selected by the amplitude of  $R_i$ ,  $g$  is the first  $g$  IMFs selected by the amplitude of  $K_i$ , the IMF represented by  $b$  cannot overlap with the IMF represented by  $g$ , and  $l = b + g$ . The Pearson correlation coefficients represent a prioritized index. For example, when  $M = 1, 2$ , and  $3$ , the IMF selection rule is as shown in Figure 2.

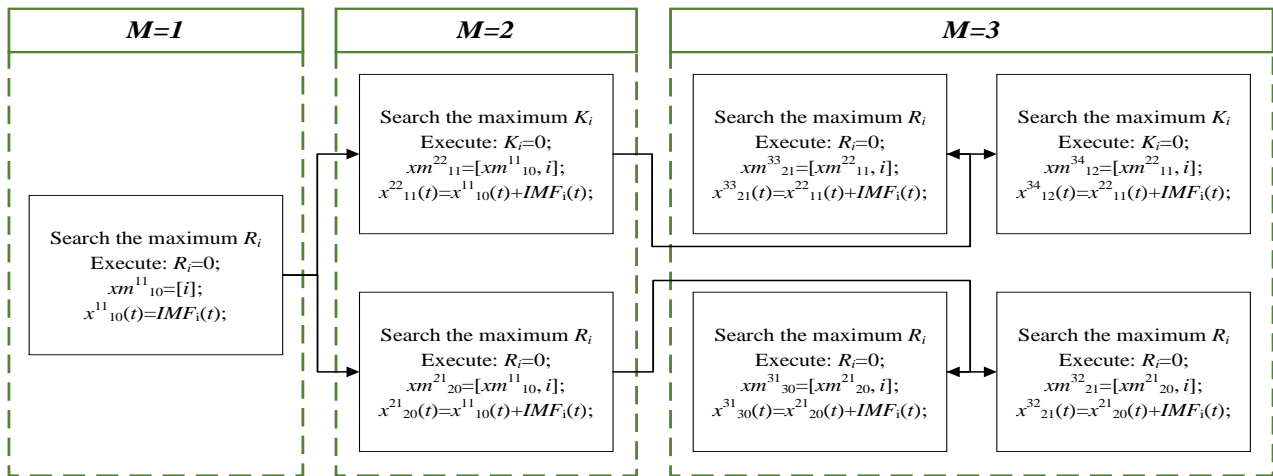


Figure 2. The selection rule of the effective IMFs.

When  $M = 1$ , the reconstructed signal  $x_{10}^{11}$  and the record matrix of IMF order  $xm_{10}^{11}$  are acquired. When  $M = 2$ , the reconstructed signal  $x_{20}^{21}$ ,  $x_{11}^{22}$  and the record matrix of IMF order  $xm_{20}^{21}$ ,  $xm_{11}^{22}$  are acquired. When  $M = 3$ , the reconstructed signals  $x_{30}^{31}$ ,  $x_{21}^{32}$ ,  $x_{21}^{33}$  and  $x_{12}^{34}$  and the record matrices of IMF orders  $xm_{30}^{31}$ ,  $xm_{21}^{32}$ ,  $xm_{21}^{33}$  and  $xm_{12}^{34}$  are acquired.

Step 5: The Pearson correlation coefficient  $R_{x_{bg}^{lz,(i)}}$ , mean square error  $MSE_{x_{bg}^{lz,(i)}}$ , and redefined signal-to-noise ratio  $MSNR_{x_{bg}^{lz,(i)}}$  between the reconstructed signal and the raw signal are calculated.

$$R_{x_{bg}^{lz,(i)}} = \frac{\sum_{t=1}^T \left( x_{bg}^{lz,(i)}(t) - \overline{x_{bg}^{lz,(i)}(t)} \right) \left( x_a(t) - \overline{x_a(t)} \right)}{\sqrt{\sum_{t=1}^T \left( x_{bg}^{lz,(i)}(t) - \overline{x_{bg}^{lz,(i)}(t)} \right)^2 \left( x_a(t) - \overline{x_a(t)} \right)^2}} \quad (11)$$

$$MSE_{x_{bg}^{lz,(i)}} = \frac{1}{T} \sum_{t=1}^T \left( x_a(t) - x_{bg}^{lz,(i)}(t) \right)^2 \quad (12)$$

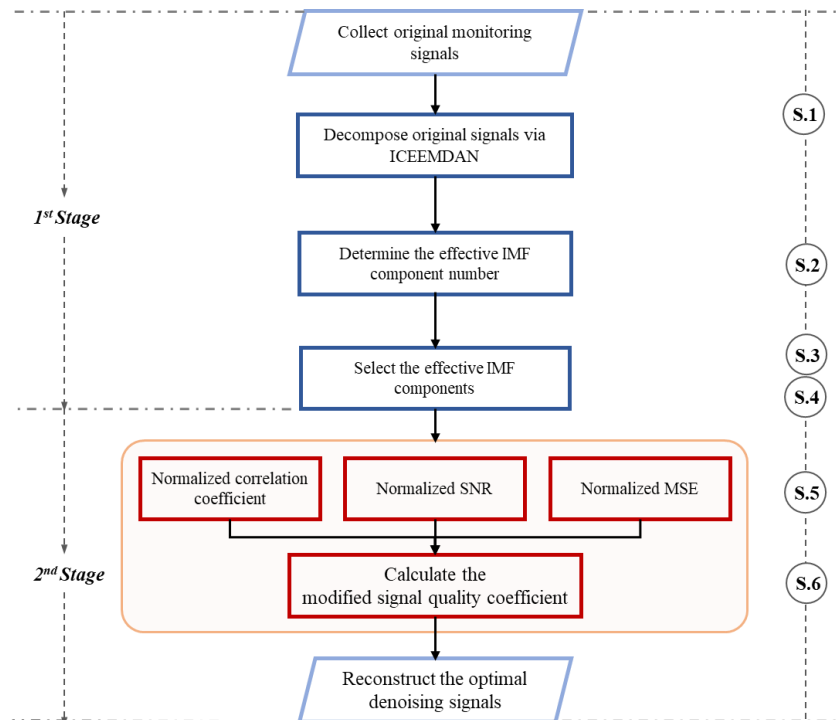
$$MSNR_{x_{bg}^{lz,(i)}} = 10 \times \log_{10} \left( \frac{\sum_{t=1}^T \left( x_a(t) \right)^2}{\sum_{t=1}^T \left( x_a(t) - x_{bg}^{lz,(i)}(t) \right)^2} \right) \quad (13)$$

The  $R_{x_{bg}^{Iz,(i)}}$  can guarantee the signal fidelity, and the larger value indicates that more fault information is contained in the reconstructed signal. The  $MSE_{x_{bg}^{Iz,(i)}}$  can represent the approximation of the reconstructed signal to the raw signal. The lower its value, the better the denoised performance. The  $MSNR_{x_{bg}^{Iz,(i)}}$  can reflect the energy proportion of the used signal to the noise signal. The larger its value, the lower the noise proportion in raw signals.

Step 6: The modified signal quality coefficient is defined as follows.

$$MSQC_{x_{bg}^{Iz,(i)}} = NR_{x_{bg}^{Iz,(i)}} + NMSNR_{x_{bg}^{Iz,(i)}} + 1/NMSE_{x_{bg}^{Iz,(i)}} \quad (14)$$

where  $NR_{x_{bg}^{Iz,(i)}}$ ,  $NMSE_{x_{bg}^{Iz,(i)}}$ , and  $NMSNR_{x_{bg}^{Iz,(i)}}$  are the normalization values of the Pearson correlation coefficient, mean square error and modified signal-to-noise ratio. The optimal denoised signal  $x_{de}$  can be determined by reconstructing  $NE$  effective IMF components with the maximal  $MSQC_{x_{bg}^{Iz,(i)}}$ . The flowchart of overall signal denoising processes based on MSQC can be seen in Figure 3.



**Figure 3.** The flowchart of overall signal denoising processes based on MSQC.

### 3.2. Versatile Residual Shrinkage Network

The versatile residual shrinkage network (VRSN) consists of Counter-DRSN and Locator-DRSN, and the architecture of VRSN is exhibited in Figure 4.

#### 3.2.1. Counter-DRSN

This network aims to count the fault number. Firstly, the original signals  $x_s$  are processed by the MSQC, and the denoised signals  $x_{de}$  with corresponding fault number labels  $M_s \in (1, \dots, W)$  are fed into Counter-DRSN. The predicted fault number is output as follows.

$$\hat{M} = \text{softmax} \left( \sum_{p,q=1}^{P,Q} W_{pq}^{l,C} f(x_{de}) + b_q^{l,C} \right) \quad (15)$$

where  $\hat{M}$  is the predicted fault number,  $s$  is the sample number,  $W_{pq}^{l,C}$  is the weight vector of Counter-DRSN between the  $p$ th neural at the  $l$ th layer and the  $q$ th neural at the  $(l + 1)$ th layer,  $x_{de}$  is the denoised signal, and  $b_q^{l,C}$  is the Counter-DRSN biases of all the  $l$ th layer's neurons for the  $(l + 1)$ th neural.

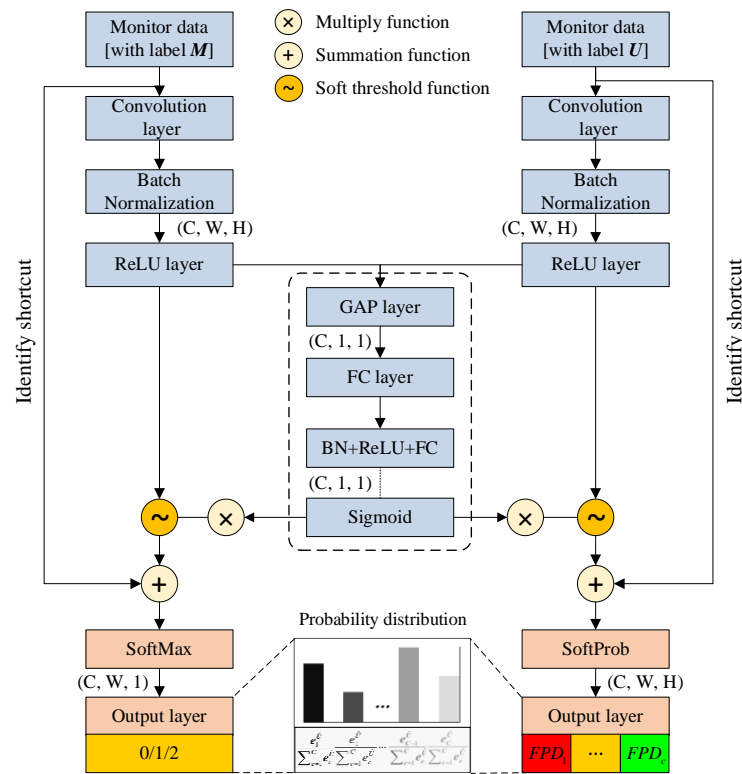


Figure 4. The architecture of versatile residual shrinkage network.

The objective function of Counter-DRSN is given by

$$Obj = \frac{1}{s} \sum_{i=1}^s (M_i - \hat{M}_i)^2 \tag{16}$$

where  $Obj$  is the objective function.

### 3.2.2. Locator-DRSN

The purpose of this network is to calculate the fault probability distribution and determine the fault pattern. Similarly, the original signals are denoised by the MSQC. The denoised signals with corresponding fault category labels  $U_s \in (1, \dots, C)$  are input into Locator-DRSN, and the soft threshold function is added to eliminate the noise-relative components. The fault probability distribution is output from the fully connected layer of the Locator-DRSN as

$$\hat{U} = SoftProb \left( \sum_{p,q=1}^{P,Q} W_{pq}^{l,L} f(x_{de}) + b_q^{l,L} \right) \tag{17}$$

where  $\hat{U}$  is the predicted fault probability distribution,  $W_{pq}^{l,L}$  is the weight vector of Locator-DRSN, and  $b_q^{l,L}$  is the network bias.



Then, the Locator-DRSN is updated by the objective function as follows

$$l_L = -\frac{1}{S} \sum_{i=1}^S \sum_{j=1}^C E_{ij} \log(\hat{U}_{ij}) \quad (18)$$

where  $E_{ij}$  is one if the sample  $i$  belongs to fault category  $j$ , and zero otherwise.  $\hat{U}_{ij}$  is the occurrence probability of the  $j$ th fault category.

Finally, the predicted fault number and probability distribution are aggregated to determine the fault pattern as

$$LABEL_i(x_{de}) = \left\{ c_i \mid \underset{\{i=1, \dots, \hat{M}\}}{\operatorname{argmax}} \{ \hat{M}, \dots, 0 \} \hat{U}(x_{de}) \right\}_{\{c=1, \dots, C\}} \quad (19)$$

where  $LABEL_i(x_{de}) \in (c_1, \dots, c_{\hat{M}})$  is the predicted label of the VRSN.

#### 4. Experimental Study

To verify the proposed diagnosis method, the self-built wind turbine gearbox compound fault test platform was adopted to place the sample under single and compound fault patterns [26]. This is composed of a multistage gearbox, a generator motor, a prime mover, an electric load simulator, a data collection device, and a laptop, as shown in Figure 5. The rotational speed of a prime mover is 1400 RPM, and two meshing gears rotate at 1184 and 840 RPM. One health condition (H1), five single fault patterns (SFP1-SFP5) and six compound fault patterns (CFP1-CFP6) are simulated as exhibited in Figure 5 and Table 1. The fault units are processed artificially. For instance, the broken tooth is manufactured by the laser cutting method to cut a part of a gear tooth.

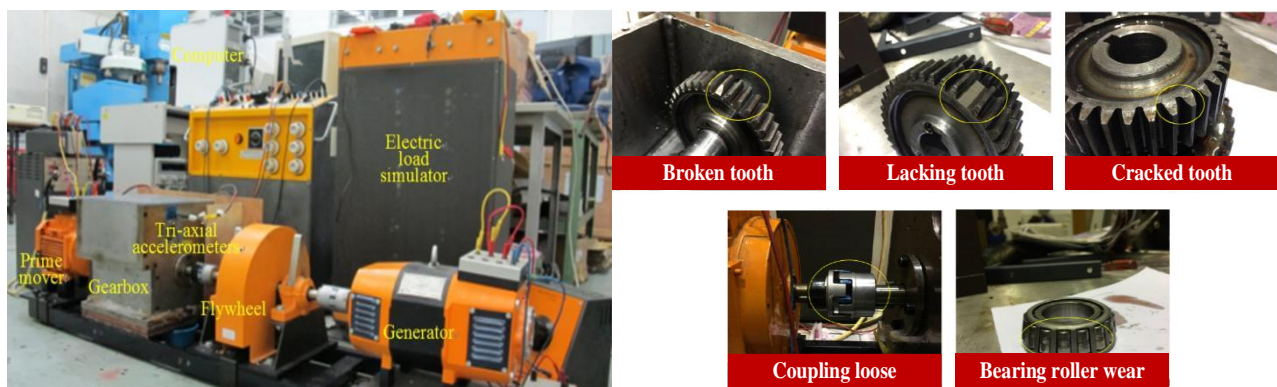


Figure 5. The self-built wind turbine gearbox compound fault test platform.

Table 1. A detailed description of fault number and category label.

Fault Pattern	Fault Description	Fault Label	Count Label	Training/Validation/Testing Number
H1	Health (H1)	0	0	1050/150/300
SFP1	Broken tooth (BT)	1	1	1050/150/300
SFP2	Lacking tooth (LT)	2	1	1050/150/300
SFP3	Cracked tooth (CT)	3	1	1050/150/300
SFP4	Coupling loose (CL)	4	1	1050/150/300
SFP5	Bearing roller wear (BRW)	5	1	1050/150/300



Table 1. Cont.

Fault Pattern	Fault Description	Fault Label	Count Label	Training/Validation/Testing Number
CFP1	BT and LT	1, 2	2	1050/150/300
CFP2	BT and CT	1, 3	2	1050/150/300
CFP3	BT and CL	1, 4	2	1050/150/300
CFP4	BT and BRW	1, 5	2	1050/150/300
CFP5	LT and CL	2, 4	2	1050/150/300
CFP6	LT and BRW	2, 5	2	1050/150/300

The raw vibration signal was acquired by use of the triaxial accelerometer (Type, NI-cDaq-9174; sensitivity, 100 mV/g). Each sample consisted of 2048 sampling points ( $2 \text{ s} \times 1024 \text{ Hz}$ ). There were 1500 samples under each fault pattern, and they were divided into 70%, 10% and 20% for the model training, validation and testing, respectively. The proposed algorithm was executed with MATLAB R2022b and Python v3.8, and conducted on a personal laptop with Intel Core i7 Processor 14900HK CPU, 32 GB memory, and a Microsoft Windows 11 enterprise operation system. The overall schematic of the experimental study design can be seen in Figure 6.

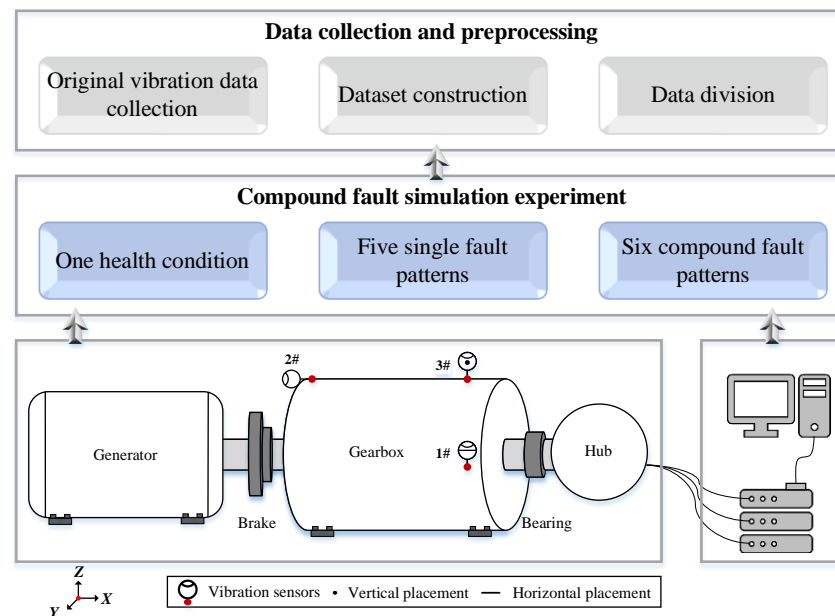
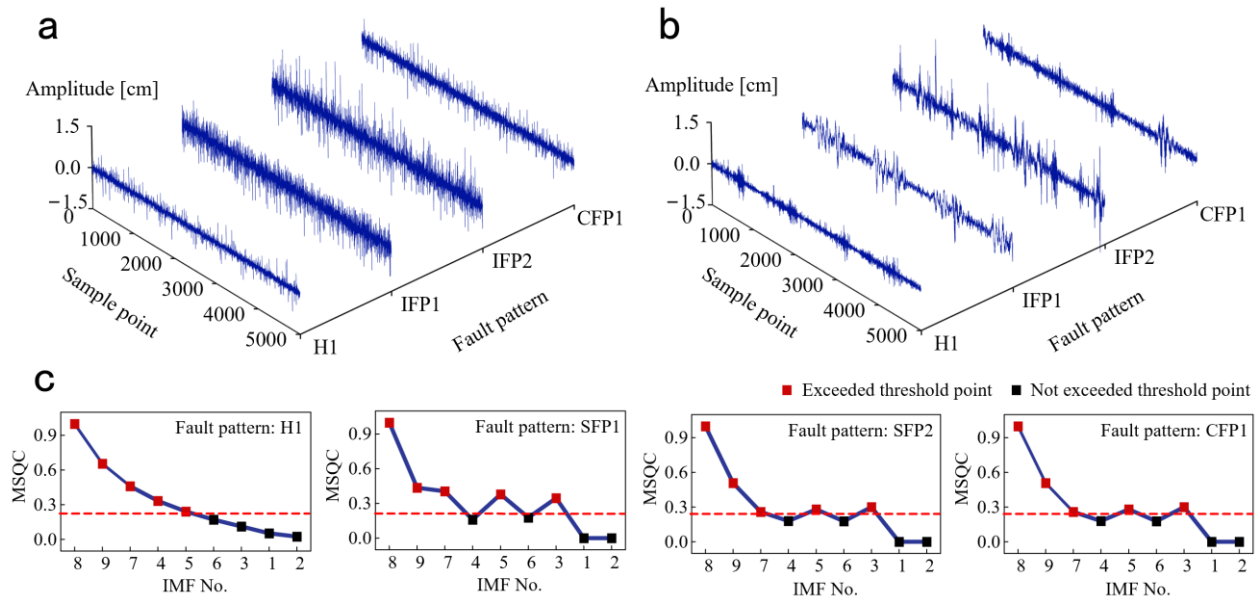


Figure 6. The overall schematic of the experimental study design.

## 5. Results and Discussion

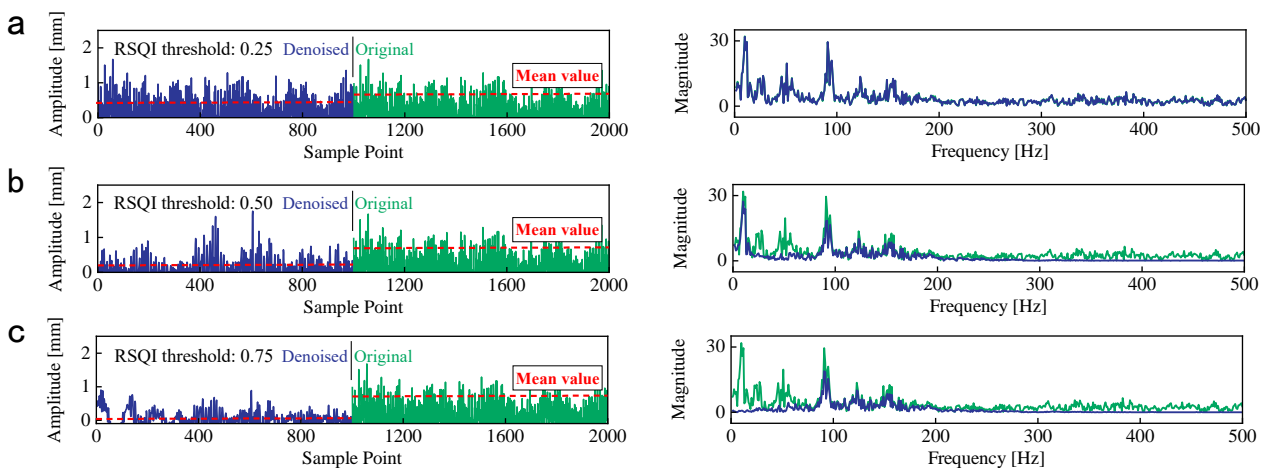
### 5.1. Signal Denoised Performance

The denoised results of raw vibration signals under H1, SFP1, SFP2, and CFP1 are shown in Figure 7. In Figure 7a,b, the amplitudes of vibration signals under SFP1, SFP2, and CFP1 are obviously larger than that under H1. This phenomenon helps us to identify different fault patterns from aspects of time-domain discrepancy. In Figure 7c, the MSQCs of each IMF component under four fault patterns are exhibited. When the MSQC threshold is 0.25, the denoised signals are obtained by reconstructing the effective IMF components exceeding the MSQC threshold. This indicates that the no. 3, no. 5, and no. 7–9 IMFs are reserved under SFP1, SFP2, and CFP1, but the no. 3 IMF is eliminated, and no. 4 IMF is reserved under H1.



**Figure 7.** The denoised results of raw vibration signals under four fault patterns. (a) Raw vibration signals. (b) Denoised signals. (c) The MSQC values.

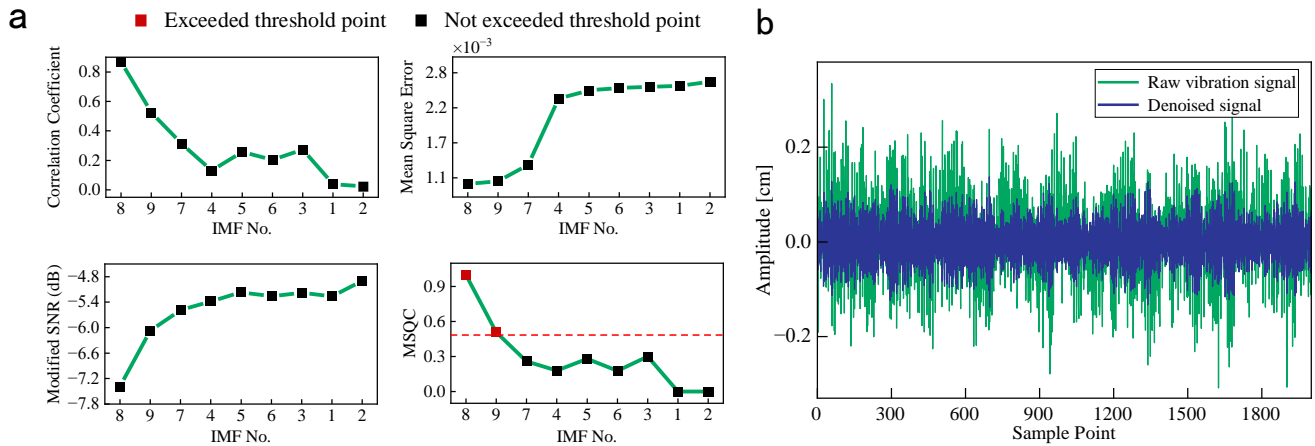
To achieve the optimal denoised effect, the denoised results of CFP1 with MSQC threshold = 0.25, 0.50, and 0.75 are exhibited in Figure 8. The waveforms of raw vibration and denoised signals show no obvious differences, and the frequency components are also similar. The influence of irrelevant noise components cannot be eliminated with the lower MSQC threshold. On the contrary, only 80–200 Hz frequency components are left, and the amplitude of denoised signals decreases significantly. The higher MSQC threshold eliminates the plentiful frequency components relevant to the wind turbine operation condition, and it will bring about a negative impact on the fault identification. Therefore, an appropriate threshold is critical for the denoised performance, and it is set to 0.5 through the repeated experimentation and signal analyses, as shown in Figure 8b.



**Figure 8.** The denoised performance of CFP1 with four MSQC thresholds. (a) The denoised results with MSQC threshold = 0.25. (b) The denoised results with MSQC threshold = 0.5. (c) The denoised results with MSQC threshold = 0.75.

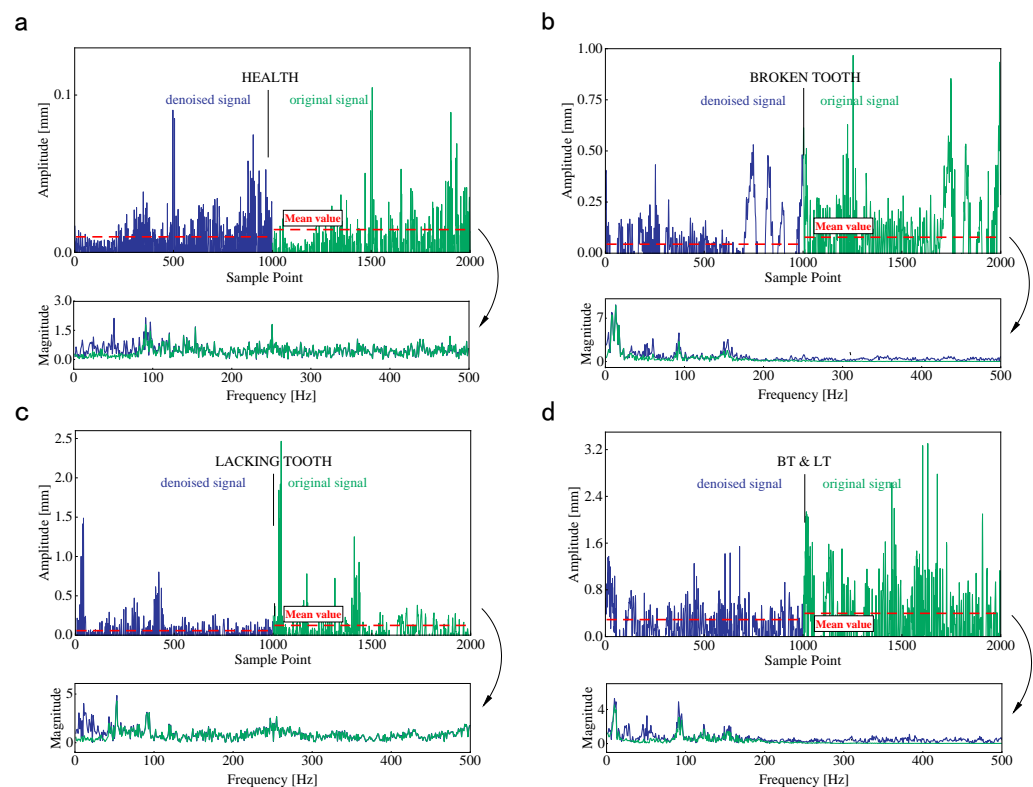
The denoised algorithm is used to process the raw vibration signals under CFP1, as shown in Figure 9. The MSQCs of no. 3, no. 5, and no. 7–9 IMFs exceed the threshold due to the lower mean square error, modified signal-to-noise, and larger correlation coefficient, as exhibited in Figure 9a. The reserved IMFs have plentiful wind turbine operation status information, and they are integrated to obtain the denoised signals. In Figure 9b, the

waveforms of denoised signals present obvious periodicity, and their amplitude fluctuation is more subtle than that of the raw vibration signal.



**Figure 9.** The denoised performance of CFP1 with the MSQC threshold = 0.5. (a) Correlation coefficient, mean square error, modified SNR and MSQC. (b) The raw vibration and denoised signals.

The redundant IMF components and residuals are removed, and the effective IMFs are left to reconstruct the denoised signals. The original and denoised signals under H1, IFP1, IFP2, and CFP1 are shown in Figure 10. The influence of fault occurrence is more prominent on the waveform and periodicity of denoised signals, and the average of denoised signals decreases obviously. In addition, the irrelevant frequency components are eliminated according to the wind turbine operation status and fault characteristic, and the distribution of denoised signal energy is more centralized. The raw vibration signals under twelve fault patterns are processed as mentioned above, and this will help to enhance the performance of the proposed fault diagnosis method.



**Figure 10.** The original and denoised signals under four fault patterns. (a) H1. (b) IFP1. (c) IFP2. (d) CFP1.

### 5.2. Diagnosis Result and Discussion

The repetition experiments were conducted to evaluate the performance stability of the compound fault diagnosis method. In Figure 11, the diagnosis accuracies of 15 repetition experiments are recorded. The overall fault diagnosis accuracies for the test sets are steady, and the mean value is up to 96.16%. Compared with the model training approach using raw vibration signals, the mean value of diagnosis accuracy increases by 4.71% (from 91.45% to 96.16%). This indicates that the denoised algorithm can eliminate irrelevant noise interference and improve diagnostic performance effectively.

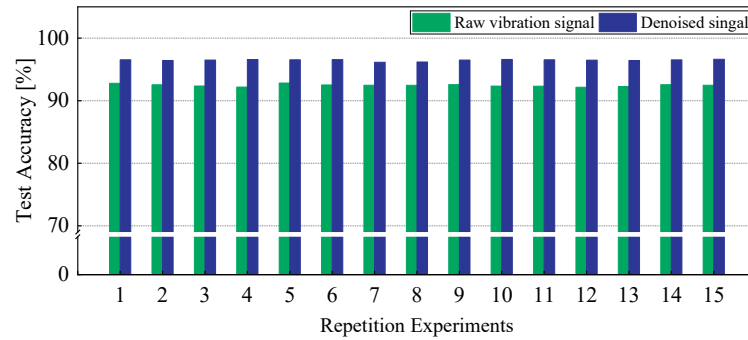


Figure 11. The test accuracy of fifteen repeated experiments before and after signal denoising.

The diagnosis results after 15 repetition experiments are shown in Figure 12. In Figure 12a, the average fault probability distribution and fault number are recorded. The predicted fault numbers show larger fluctuation under IFP3 and CFP5. Their test accuracies are relatively lower. However, the propose compound diagnosis method can still identify fault patterns precisely, as exhibited in Figure 12b, and only some of the test samples are misjudged, which were identified as BT/CT under IFP3 and CL/CT under CFP5.

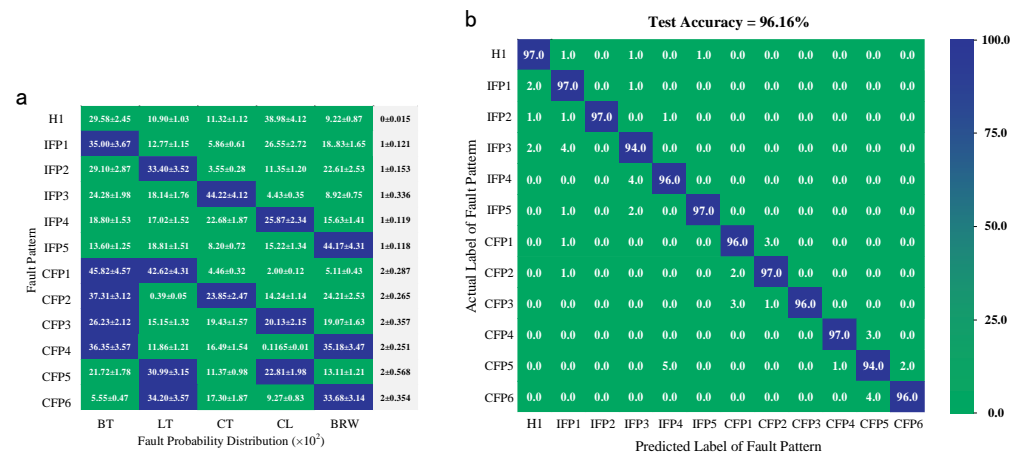
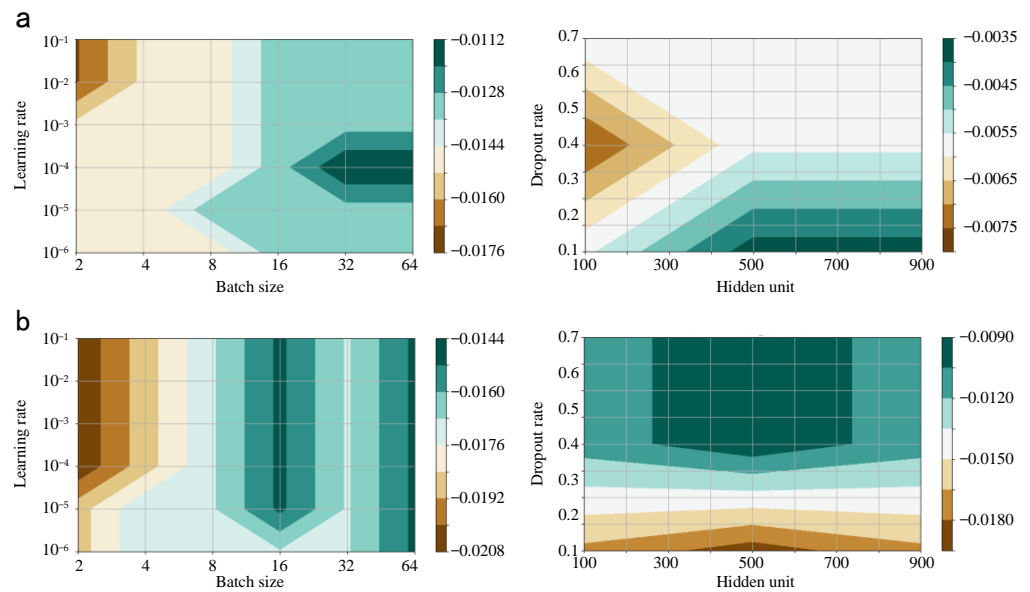


Figure 12. The diagnosis results after 15 repetition experiments. (a) The predicted fault probability distribution and fault number. (b) The multiclass confusion matrix of fault diagnosis for the test sets.

The determination of the deep neural network parameter is still a great problem. There are no mature theories to guide the process of parameter setup [27]. Herin, the hyper parameters of the proposed network such as batch size, learning rate, hidden unit, and the dropout rate are determined by means of the cross-validation and repeated experience [28].

In Figure 13, the relation between the parameter group and the testing loss is exhibited for the last trial. In detail, the candidates of batch size, learning rate, dropout rate, and hidden unit are [2, 4, 8, 16, 32, 64],  $[1 \times 10^{-6}, 1 \times 10^{-5}, 1 \times 10^{-4}, 1 \times 10^{-3}, 1 \times 10^{-2}, 1 \times 10^{-1}]$ , [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7], and [100, 200, 300, 400, 500, 600, 700, 800, 900]. The optimal batch size, learning rate, dropout rate, and hidden unit of Counter-DRSN are 64,  $1 \times 10^{-4}$ , 0.1, and 900. Those of the Locator-DRSN are 64,  $1 \times 10^{-4}$ , 0.4, and 500.



**Figure 13.** The parameters of Counter-DRSN and Locator-DRSN. (a) The cross-validation result of Counter-DRSN. (b) The cross-validation result of Locator-DRSN.

### 5.3. Comparison Analysis

To verify the effectiveness of the proposed method, there are five compound fault diagnosis algorithms used for the comparison analysis, including pairwise probabilistic multilabel classification (PPMLC) [8], random  $k$  labelsets (RAKEL) [13], dual-extreme learning machine (Dual-ELM) [6], and wavelet transform multi-label convolutional neural network (WT-MLCNN) [27]. The denoised algorithm was here applied in the comparative algorithms for further denoised effect analysis. To eliminate the contingency in the results, twenty repeated experiments were conducted, and the same samples were used for the model training, validation, and testing. The comparison analysis results are recorded in Table 2.

**Table 2.** The compound fault diagnosis performances of seven methods.

Method No.	Diagnosis Strategies	Running Time (s)	Average Accuracy $\pm$ Standard deviation (%)		
			Training Set	Validation Set	Testing Set
Method 1 [8]	PPMLC	57.95	86.98 $\pm$ 6.24	86.12 $\pm$ 6.78	86.03 $\pm$ 6.89
Method 2	PPMLC with MSQC	75.34	89.87 $\pm$ 5.12	88.54 $\pm$ 5.64	88.32 $\pm$ 5.79
Method 3 [13]	RAKEL	81.34	82.34 $\pm$ 7.81	81.83 $\pm$ 8.12	81.54 $\pm$ 8.34
Method 4	RAKEL with MSQC	102.54	86.73 $\pm$ 6.875	84.57 $\pm$ 7.51	84.53 $\pm$ 7.53
Method 5 [6]	Dual-ELM	21.75	83.64 $\pm$ 7.84	82.57 $\pm$ 7.96	82.53 $\pm$ 7.99
Method 6	Dual-ELM with MSQC	39.57	87.51 $\pm$ 6.57	86.54 $\pm$ 6.89	85.98 $\pm$ 7.23
Method 7 [27]	WT-MLCNN	207.24	93.76 $\pm$ 3.251	92.87 $\pm$ 3.427	92.72 $\pm$ 3.457
Method 8	WT-MLCNN with MSQC	231.25	98.37 $\pm$ 1.387	95.81 $\pm$ 2.101	95.83 $\pm$ 2.024
Method 9 (Ours)	The proposed method	52.34	98.54 $\pm$ 1.360	96.34 $\pm$ 1.785	96.26 $\pm$ 1.823

The average test accuracy of the proposed compound method reached up to 96.26%, followed by method 8 (95.83%), method 7 (92.72%), method 2 (88.32%), method 1 (86.03%), and so on. It is obvious that after applying the denoised algorithm, the average accuracies increased by about 4% for four comparative algorithms. However, the standard deviations of methods 3, 4, and 5 were still larger, and thus showed inferior performance stability. To improve the efficiency of diagnosis, method 5 only required 21.75 s to execute the diagnostic tasks, followed by method 6 (39.57 s), method 9 (52.34 s), method 1 (57.95 s), and so on. The execution time for the proposed method is close to those of the first two machine learning-based methods. The comparison analysis results demonstrate that the proposed method is superior in terms of diagnosis accuracy and efficiency.

## 6. Conclusions

The compound fault diagnosis method is proposed based on the modified signal quality coefficient and versatile residual shrinkage network. The main conclusions are summarized as follows:

1. The signal denoised algorithm is designed to remove the noise components irrelevant to wind turbine operation status. In this paper, the modified signal quality coefficient can balance the denoised effect and signal fidelity, and fault-sensitive features hidden in the original collected signals can be excavated precisely;
2. A versatile residual shrinkage network is constructed for the compound fault diagnosis. Unlike the probabilistic-based method, the proposed network is self-adaptive, and is used to identify single- or simultaneous-fault scenarios without the manual intervention required for setting the empirical threshold;
3. An effective multithread network structure is constructed to optimize the computation resource configuration. Two parallel residual shrinkage networks can be implemented simultaneously to count the number of faults and determine the fault probability distribution used for responding to a real-time fault diagnosis task.

In future research, physical interpretability methods such as the physics-knowledge-guided method can be employed for the construction of neural network because they show good performance in improving the stability of neural networks and enhancing the interpretability of the diagnostic processes [29]. In addition, different fault categories entail different degrees of risk regarding the health of a wind turbine, and the fault risk weight can be evaluated by use of the multi-criteria decision-making method, as an extension of the current research topic.

**Author Contributions:** Conceptualization, J.W. and G.Z.; methodology, J.W. and W.J.; software, J.W.; validation, J.W.; formal analysis, J.W. and G.Z.; investigation, J.W. and W.J.; resources, W.J. and Z.G.; data curation, J.W.; writing—original draft preparation, J.W. and W.J.; writing—review and editing, J.W. and W.J.; visualization, J.W.; supervision, W.J.; project administration, W.J. and Y.W.; funding acquisition, W.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the National Natural Science Foundation of China under the Grant No. 523B2100, in part by the Interdisciplinary Research Program of Huazhong University of Science and Technology under Grant No. 2024JCYJ028, and in part by Shenzhen Science and Technology Program under Grant No. JCYJ20240813113102003.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are available on request from the authors.

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

## References

1. Wang, C.; Yang, J.; Jie, H.; Tian, B.; Zhao, Z.; Chang, Y. An uncertainty perception metric network for machinery fault diagnosis under limited noisy source domain and scarce noisy unknown domain. *Adv. Eng. Inform.* **2024**, *62*, 102682. [[CrossRef](#)]
2. He, F.; Ye, Q. A bearing fault diagnosis method based on wavelet packet transform and convolutional neural network optimized by simulated annealing algorithm. *Sensors* **2022**, *22*, 1410. [[CrossRef](#)] [[PubMed](#)]
3. Wang, C.; Tian, B.; Yang, J.; Jie, H.; Chang, Y.; Zhao, Z. Neural-transformer: A brain-inspired lightweight mechanical fault diagnosis method under noise. *Reliab. Eng. Syst. Safe.* **2024**, *251*, 110409. [[CrossRef](#)]
4. Mohammad, H.; Vlastic, F.; Zacek, J.; Maya, B.; Mazal, P. Using Acoustic Emission for Condition Monitoring of the Main Shaft Bearings in 4-Point Suspension Wind Turbine Drivetrains. *Nondestruct. Test. Eval.* **2023**, *39*, 2108–2131. [[CrossRef](#)]
5. Tang, S.; Ma, J.; Yan, Z.; Zhu, Y.; Khoo, B.C. Deep transfer learning strategy in intelligent fault diagnosis of rotating machinery. *Eng. Appl. Artif. Intel.* **2024**, *134*, 108678. [[CrossRef](#)]



6. Yang, Z.X.; Wang, X.; 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](#)]
7. Zhou, D.; Blaabjerg, F.; Franke, T.; Tønnes, M.; Lau, M. Comparison of Wind Power Converter Reliability with Low-Speed and Medium-Speed Permanent-Magnet Synchronous Generators. *IEEE Trans. Ind. Electron.* **2015**, *62*, 6575–6584. [[CrossRef](#)]
8. Wong, P.K.; Zhong, J.; Yang, Z.; Vong, C.M. Sparse Bayesian Extreme Learning Committee Machine for Engine Simultaneous Fault Diagnosis. *Neurocomputing* **2016**, *174*, 331–343. [[CrossRef](#)]
9. Clare, A.; King, R.D. Knowledge Discovery in Multi-Label Phenotype Data. In *Principles of Data Mining and Knowledge Discovery*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2001; Volume 2168, pp. 42–53. [[CrossRef](#)]
10. Zhang, M.L.; Zhou, Z.H. ML-KNN: A Lazy Learning Approach to Multi-Label Learning. *Pattern Recognit.* **2007**, *40*, 2038–2048. [[CrossRef](#)]
11. Tahir, M.A.; Kittler, J.; Bouridane, A. Multi-Label Classification Using Stacked Spectral Kernel Discriminant Analysis. *Neurocomputing* **2016**, *171*, 127–137. [[CrossRef](#)]
12. Liu, B.; Tsoumakas, G. Dealing with Class Imbalance in Classifier Chains via Random Undersampling. *Knowl.-Based Syst.* **2020**, *192*, 105292. [[CrossRef](#)]
13. Wang, R.; Kwong, S.; Wang, X.; Jia, Y. Active k -Labelsets Ensemble for Multi-Label Classification. *Pattern Recognit.* **2021**, *109*, 107583. [[CrossRef](#)]
14. Wang, L.; Chen, Z.; Zou, H.; Huang, D.; Pan, Y.; Cheang, C.F.; Li, J. A Deep Learning-Based High-Temperature Overtime Working Alert System for Smart Cities with Multi-Sensor Data. *Nondestruct. Test. Eval.* **2024**, *39*, 164–184. [[CrossRef](#)]
15. Wang, L.; Cao, H.; Xu, H.; Liu, H. A Gated Graph Convolutional Network with Multi-Sensor Signals for Remaining Useful Life Prediction. *Knowl.-Based Syst.* **2022**, *252*, 109340. [[CrossRef](#)]
16. Hassan, M.U.; Khan, T.; Zafar, T.; Yousuf, W.B.; Shah, A. Degradation Prognostics of Aerial Bundled Cables Based on Multi-Sensor Data Fusion. *Nondestruct. Test. Eval.* **2024**, *40*, 489–507. [[CrossRef](#)]
17. Wang, Z.; He, X.; Yang, B.; Li, N. Subdomain Adaptation Transfer Learning Network for Fault Diagnosis of Roller Bearings. *IEEE Trans. Ind. Electron.* **2022**, *69*, 8430–8439. [[CrossRef](#)]
18. Sahu, P.; Rai, R.; Patel, N. Deep learning-based fault classification of rolling bearings under noisy conditions using CEEMD-VMD-IMF with magnitude scalogram images. *J. Mech. Sci. Technol.* **2024**, *10*, 102409. [[CrossRef](#)]
19. Lv, M.; Li, H. Nonlinear Dispersive Component Decomposition: Algorithm and Applications. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3515614. [[CrossRef](#)]
20. Zhang, H.; Zhao, S.; Qiang, W.; Chen, Y.; Jing, L. Feature Extraction Framework Based on Contrastive Learning with Adaptive Positive and Negative Samples. *Neural Netw.* **2022**, *156*, 244–257. [[CrossRef](#)]
21. Ding, K.; Chen, X.; Jiang, M.; Yang, H.; Chen, X.; Zhang, J.; Gao, R.; Cui, L. Feature extraction and fault diagnosis of photovoltaic array based on current–voltage conversion. *Appl. Energy* **2024**, *353*, 122135. [[CrossRef](#)]
22. Hu, H.; Peng, G.; Wang, X.; Zhou, Z. Weld Defect Classification Using 1-D LBP Feature Extraction of Ultrasonic Signals. *Nondestruct. Test. Eval.* **2018**, *33*, 92–108. [[CrossRef](#)]
23. Colominas, M.A.; Schlotthauer, G.; Torres, M.E. Improved Complete Ensemble EMD: A Suitable Tool for Biomedical Signal Processing. *Biomed. Signal Process. Control* **2014**, *14*, 19–29. [[CrossRef](#)]
24. Miao, Y.; Zhang, B.; Li, C.; Lin, J.; Zhang, D. Feature Mode Decomposition: New Decomposition Theory for Rotating Machinery Fault Diagnosis. *IEEE Trans. Ind. Electron.* **2023**, *70*, 1949–1960. [[CrossRef](#)]
25. Zhao, M.; Zhong, S.; Fu, X.; Tang, B.; Pecht, M. Deep Residual Shrinkage Networks for Fault Diagnosis. *IEEE Trans. Ind. Inform.* **2020**, *16*, 4681–4690. [[CrossRef](#)]
26. Jiang, W.; Wu, J.; Zhu, H.; Li, X.; Gao, L. Paired Ensemble and Group Knowledge Measurement for Health Evaluation of Wind Turbine Gearbox under Compound Fault Scenarios. *J. Manuf. Syst.* **2023**, *70*, 382–394. [[CrossRef](#)]
27. Liang, P.; Deng, C.; Wu, J.; Yang, Z.; Zhu, J.; Zhang, Z. Compound Fault Diagnosis of Gearboxes via Multi-Label Convolutional Neural Network and Wavelet Transform. *Comput. Ind.* **2019**, *113*, 103132. [[CrossRef](#)]
28. Wang, S.; Shuai, H.; Hu, J.; Zhang, J.; Liu, S.; Yuan, X.; Liang, P. Few-shot fault diagnosis of axial piston pump based on prior knowledge-embedded meta learning vision transformer under variable operating conditions. *Expert Syst. Appl.* **2025**, *269*, 126452. [[CrossRef](#)]
29. Yin, C.; Li, Y.; Wang, Y.; Dong, Y. Physics-guided degradation trajectory modeling for remaining useful life prediction of rolling bearings. *Mech. Syst. Signal Pr.* **2025**, *224*, 112192. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.