

Article

# Intelligent Fault Diagnosis of HVCB with Feature Space Optimization-Based Random Forest

Suliang Ma <sup>†</sup>, Mingxuan Chen <sup>†</sup> , Jianwen Wu <sup>\*</sup>, Yuhao Wang, Bowen Jia and Yuan Jiang

School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China; masuliang@buaa.edu.cn (S.M.); mingxuan\_chen@buaa.edu.cn (M.C.); 18810371198@163.com (Y.W.); jiabowen109@126.com (B.J.); jiangy\_luckystar@163.com (Y.J.)

<sup>\*</sup> Correspondence: wujianwen@buaa.edu.cn; Tel.: +86-10-8233-8384

<sup>†</sup> These authors contributed equally to this work.

Received: 19 March 2018; Accepted: 10 April 2018; Published: 16 April 2018



**Abstract:** Mechanical faults of high-voltage circuit breakers (HVCBs) always happen over long-term operation, so extracting the fault features and identifying the fault type have become a key issue for ensuring the security and reliability of power supply. Based on wavelet packet decomposition technology and random forest algorithm, an effective identification system was developed in this paper. First, compared with the incomplete description of Shannon entropy, the wavelet packet time-frequency energy rate (WTFER) was adopted as the input vector for the classifier model in the feature selection procedure. Then, a random forest classifier was used to diagnose the HVCB fault, assess the importance of the feature variable and optimize the feature space. Finally, the approach was verified based on actual HVCB vibration signals by considering six typical fault classes. The comparative experiment results show that the classification accuracy of the proposed method with the origin feature space reached 93.33% and reached up to 95.56% with optimized input feature vector of classifier. This indicates that feature optimization procedure is successful, and the proposed diagnosis algorithm has higher efficiency and robustness than traditional methods.

**Keywords:** high-voltage circuit breakers; mechanical fault diagnosis; wavelet packet decomposition; random forest algorithm; ensemble learning; feature space optimization

## 1. Introduction

As a piece of control and protection equipment, high-voltage circuit breakers (HVCB) play an important role in the power system. Various research has reported that the mechanical faults of HVCB will give rise to significant harm and outage costs [1]. Therefore, a handful of researchers have made efforts to analyze failure cause and optimize the structure of HVCB. With the development of machine learning development, intelligent diagnosis technology has been successfully applied in various fields. On this basis, establishing an intelligent identification model for HVCB mechanical fault has become an irresistible trend.

The contact displacement and electromagnet coil current of a closing or opening process have been used as typical information for distinguishing the operation condition of HVCB. For example, Ali Forootani et al. [2] established a dynamic model using Lagrange's method, and analyzed the resulting faults in the travel curve. An image matching-based algorithm integrating diamond search strategy was proposed to avoid the deficiencies of conventional travel-time curve measurement. This algorithm has been successfully applied in detection of mechanical properties with extreme learning machines [3]. On-line hybrid fault diagnosis model based on electromagnet coil current information was also developed [4,5]. Recently, a considerable amount of literature has been reported

on the vibration characteristics of HVCB [6–8]. All this research indicates that the vibration signal of HVCB during operation is non-stationary and non-linear, which is more difficult to capture.

Time-frequency analysis tools have been widely utilized in extracting the fault feature vector of the vibration signal. Developed by Huang et al. in 1998, empirical mode decomposition (EMD) has become one of the most potent time-frequency analysis technologies. Based on the local characteristic time scale of a signal, it can decompose complicated signals into a number of intrinsic mode functions (IMFs) containing the characteristic information of the original signal [9,10]. However, the conventional EMD method is subjected to disadvantages such as end effect, high computational complexity and inconformity of decomposition among different signals, spectrum overlapping, and chaotic boundary frequency among different intrinsic modes. Therefore, numerous experts have put forward a series of improved methods, such as local mean decomposition (LMD) [6], improved ensemble empirical mode decomposition (EEMD) [11], vibration mode decomposition (VMD) [12] and so on. These improved approaches can successfully address spectrum confusion problems but are subjected to some other weaknesses. In contrast to the EMD approach, wavelet transform (WT) was proposed as an adjustable windowed Fourier transform with limited energy leakage [13]. Moreover, WT was extended to wavelet packet transform (WPT) to effectively split the high-frequency band containing rich fault modulation information [14]. Once wavelet basis function and decomposition scale are determined, the results of wavelet package transform will be the signal under a certain scale, whose frequency components are only related to the sample frequency rather than the signal itself. Obviously, wavelet analysis is not a self-adaptive signal processing method in nature [9,12]. Owing to excellent adaptive and multi-resolution capability, WPT is still regarded as a powerful mathematical tool in handling vibration signal in various fields of engineering [15,16].

After extracting the feature vector of vibration signal using time-frequency analysis tools, building a high-accuracy and robust classifier becomes a key issue. Hussain et al. designed an expert system consisting of a signal processor module and a factual database module to diagnose the faults of power circuit breakers [17]. For the case of small sample, support vector machine (SVM) and its improved method have been applied for identifying mechanical faults of HVCBs, and satisfactory diagnostic performance can be achieved [18–20]. Along with artificial intelligence technology and meta-heuristic approaches widely applied to control field [21–24], a novel method based on particle swarm optimization-support vector domain description and kernel-based fuzzy c-means was proposed for adaptive fault diagnosis of HVCBs under real-time updating of model classification and lack of knowledge acquisition [25]. In addition, there are masses of categorizers practiced in various scenarios. For instance, decision tree classifiers can diagnose motor fault tasks or detect breast cancer based on medical data [26–28]; naive Bayes algorithm has been applied as a fault classifier to investigate the status of a monoblock centrifugal pump or an engine [29–31].

Generally speaking, the above literature discusses powerful detectors for diagnosis. Obviously, the training process of a single classifier is affected by global error rate, so the models may be biased towards the majority class and ignore the minority class. Additionally, some abnormal collected samples or the dispersion of the dataset will potentially cause over-fitting [32]. Ensemble learning, which is common in machine learning, can enhance the performance of identification systems by constructing a set of individuals and weak classifiers and combining them to classify or predict new data with weighted or unweighted voting method [33]. A novel intelligent method with multivariable ensemble-based SVM was verified by detecting complex faults of different severity levels in the locomotive roller bearings [34]. Proposed by Breiman [35] and studied by Biau et al. [36], the random forest is an improved machine learning algorithm based on random subspace and ensemble learning [37]. In real industrial environments, it is difficult to acquire a large number of samples for classifier training and estimate the significance of feature candidates for fault diagnosis. Under this circumstance, RF-based models have better performance than other classical classifiers [38–40].

In this paper, a new feature indicator integrating wavelet package transform (WPT) was proposed, and the diagnosis model was designed and optimized based on random forest for diagnosing HVCB

mechanical faults. Compared with the traditional feature indicators such as wavelet time-frequency entropy as well as general classification methods such as decision tree (DT) [28], naive Bayes classifier (NB) [29], one-class support vector machine (OCSVM) [20] and bagging tree (BT) [33], the contributions of this paper can be summarized as follows.

- (1) HVCB mechanical fault is diagnosed by a holistic approach which integrates vibration signal acquisition system, feature extraction based on WPT and optimal classifier with random forest;
- (2) Through comparing the depicted level of vibration information to that of time-frequency entropy, this paper presents time-frequency wavelet energy rate (WTFER) based on WPT and realizes full-scale description of vibration signal;
- (3) The importance of feature variable is evaluated using data out of bagging based on ensemble learning. An optimal classifier based on random forest algorithm is designed to diagnose HVCB faults.

The remainder of this paper is organized as follows: Section 2 presents the procedure of the identification system based on the vibration of HVCB and experiment platform; Section 3 discusses vibration signal time-frequency feature extraction using wavelet package transform. Section 4 discusses the establishment of HVCB fault classifier and the optimization of feature space. Section 5 compares the optimal RF identification model with other traditional diagnosis methods to highlight the advantageous accuracy for optimal RF identification model. Finally, Section 6 concludes the contributions of this paper.

## 2. Identification System and Experiment Platform

### 2.1. The Identification System

In this study, the fault diagnosis process of HVCB based on vibration signal includes three parts: data acquisition under variable faults, feature extraction using wavelet package transform (WPT) for time-frequency analysis, and optimal identification model based on random forest (RF), as shown in Figure 1. Table 1 shows the general procedure of HVCB fault detection by the proposed method.

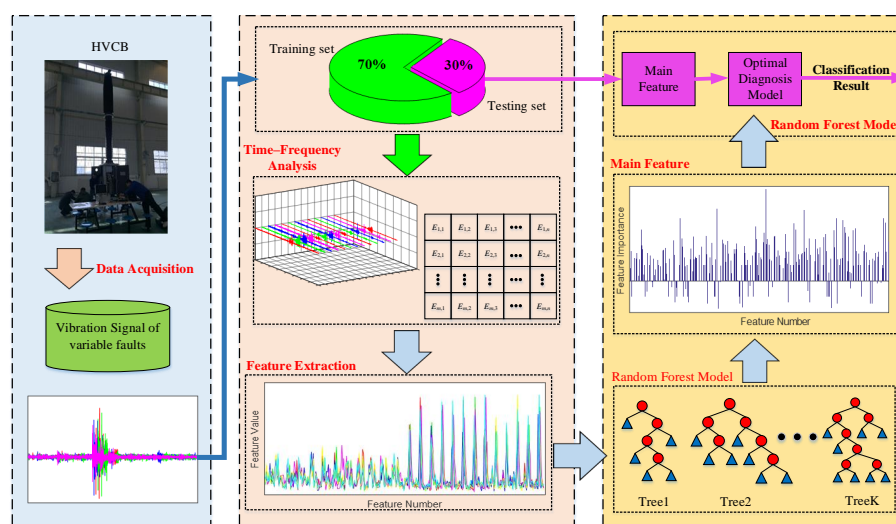


Figure 1. Structure flow of identification process.

As seen from Figure 1, 70% of measured data was applied for designing the random forest diagnosis model, while the remaining 30% of data was used for measurement of fault diagnosis accuracy. In this study, we tested the vibration signals of HVCB under 6 conditions (i.e., Normal case, Fault I—closing spring fatigue, Fault II—opening spring fatigue, Fault III—damping increasing for

transmission shaft, Fault IV—oil damper leaking, Fault V—looseness of base screw), with 50 samples per each condition, and 300 samples for total number.

**Table 1.** General procedure of HVCB fault detection by the proposed method.

Step	Description
Step 1	HVCB vibration signals acquisition
Step 2	Seven-layer WPT is used to decompose vibration signals into 128 components with different frequency bands
Step 3	Based on time-frequency description of WPT, the feature space is established using time-frequency energy ratio
Step 4	The RFs diagnosis model is established using feature space
Step 5	Based on RFs model, the importance of vibration signal is sorted, and the feature space is optimized;
Step 6	A new RF diagnosis model is constructed for optimizing feature space;
Step 7	The performance of the optimized RF diagnosis model is validated based on testing samples

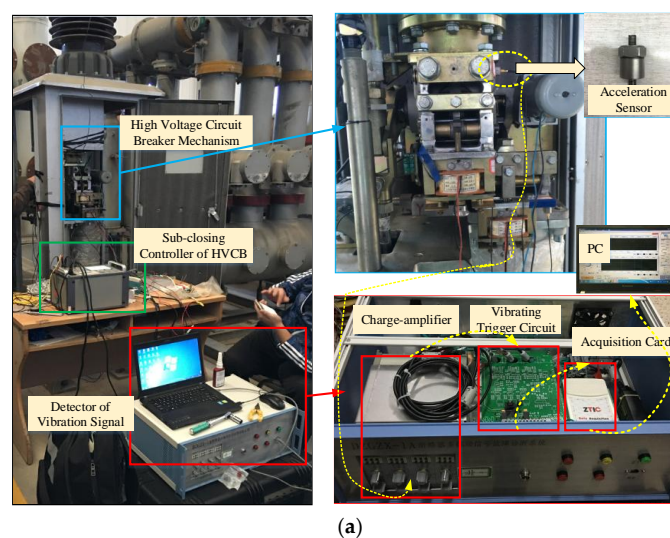
From Table 1, we can see that multiple random forest design is one of key discussed contents of this study. In Step 4, RFs model is established to evaluate the significance of vibration feature. Then, the optimal RFs model is designed for distinguishing the type of HVCB faults.

## 2.2. Vibration Signal Acquisition System with HVCB

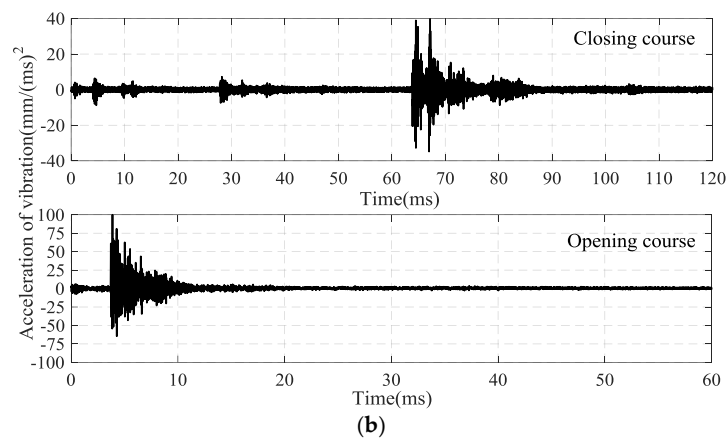
LW30-252 typed SF<sub>6</sub> HVCB (Shandong Taikai High Voltage Switchgear Co, Ltd, Shandong, China) is the object of this study. The HVCB is equipped with CT26 spring-operated mechanism of which the divide-shut brake is controlled by alternating electromagnet, the clearance between open contacts is 159 mm, and over-travel is 31 mm. As shown in Figure 2a, the vibration signal acquisition system consists of an acceleration sensor YD-111T, a charge-amplifier TS5863, vibrating trigger circuit and data acquisition card EM9118B. The yellow dotted line represents the direction of vibration signal data flow, which is finally stored in a PC. The YD-111T piezoelectric acceleration sensor has measuring range of  $\pm 10,000$  g ( $g = 9.8$  m/s<sup>2</sup>), sensitivity of 0.5 mV/g, natural frequency of 45 kHz, frequency response of 15 kHz, a maximum output voltage of  $\pm 5$  V, and a weight of 10 g.

Previous research has reported that the accuracy of vibration feature-based fault diagnosis method is affected by installation site and installation way of acceleration sensor. In this paper, the acceleration sensor was installed by threaded fastening method. Through measuring the vibration signals at multiple sites, the installation site at Figure 2a is finally selected as the optimal installation site.

When the circuit breaker receives a closing (opening) command, the acquisition system starts sampling. The sampling rate is 300 kHz and the sampling period is 120 (60) ms. In the closing (open) course, a normal vibration signal is shown in Figure 2b.



**Figure 2.** Cont.



**Figure 2.** (a) Acquisition system with HVCB; (b) The measured vibration signal at normal condition.

As shown in Figure 2, the closing process of HVCB involves multi-stage impact and attenuation process, with the maximum vibration acceleration of about 4000 g. In contrast, the opening process is only represented by trumpet-type envelope, with maximum vibration acceleration almost reaching 10,000 g. According to the operation mechanism of HVCB, it can be known that some components of HVCB, such as closing spring and oil damper, impose influence only during closing process. According to the representation form of vibration signal during opening and closing process, the signal during closing process is relatively more abundant. Therefore, the vibration signal during closing process is adopted for HVCB fault diagnosis in this study.

### 3. Feature Extraction

#### 3.1. Time and Frequency Analysis Based WPT

From the perspective of time domain, people can easily realize the description of a signal, but apparently this description is not complete. Jean Baptiste Joseph Fourier (1768–1830), a French mathematician, developed a method of expressing any function in terms of a weighted sum of cosines and sines functions [13]. Therefore, we analyzed real-world signals in both time and frequency domain. Based on Fourier transform, any finite energy signal can be represented as Equation (1).

$$\text{FT}\{x(t)\} = X(\Omega) = \int_{-\infty}^{\infty} x(t)e^{-j\Omega t} dt, \Omega = 2\pi f \quad (1)$$

where  $f$  is the frequency in Hertz, and  $\Omega t$  is the phase in radian.

We distinctly discover that Fourier transform theory could not effectively analyze the signal's spectral variations during this interval of time, which was further confirmed by the vibration signal of HVCB in Figure 1. To avoid this lack of locality property, wavelet transform can be used as an effective method to preserve signal characteristics. However, WT does not split the high frequency bands where the modulation information of vibration signal always exists. A better representation of signal is to extend WT to WPT.

WPT can be implemented by a pair of low-pass and high-pass wavelet filters, denoted as  $h(k)$  and  $g(k) = (-1)^k h(1 - k)$ , respectively. These filters, also known as Quadrature Mirror Filters (QMF) [14,15], are constructed from the selected wavelet function  $\psi(t)$ . The corresponding scaling function  $\varphi(t)$  is described as follow:

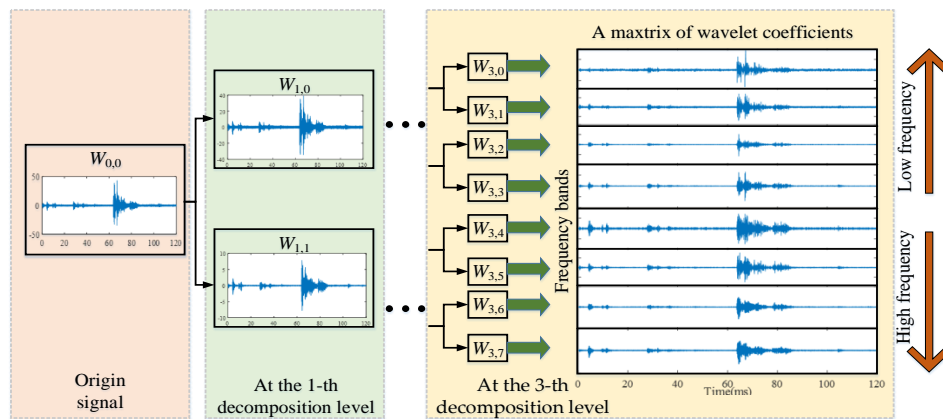
$$\begin{cases} u_{2n}(t) = \sqrt{2} \sum_k h(k) u_n(2t - k) \\ u_{2n+1}(t) = \sqrt{2} \sum_k g(k) u_n(2t - k) \end{cases} \quad (2)$$

where  $u_0(t) = \phi(t)$  and  $u_1(t) = \psi(t)$ . Correspondingly, the signal is decomposed as

$$\begin{cases} d_{j+1,2n} = \sum_m h(m-2k)d_{j,n} \\ d_{j+1,2n+1} = \sum_m g(m-2k)d_{j,n} \end{cases} \quad (3)$$

where  $d_{j,n}$  denotes the wavelet coefficients at the  $j$ -th level and the  $n$ -th sub-band;  $d_{j+1,2n}$  and  $d_{j+1,2n+1}$  denote the wavelet coefficients at the  $(j+1)$ -th level and  $2n$ -th sub-band, and the wavelet coefficients at the  $(j+1)$ -th level and the  $(2n+1)$ -th sub-bands, respectively;  $m$  is the number of the wavelet coefficients.

As illustrated in Figure 3, an origin vibration signal can be decomposed by a 3-level binary tree wavelet packet into eight sub-bands. Therefore, the reconstructed signal of wavelet coefficients  $W_j^n$  of each sub-band covers one eighth of the frequency information.



**Figure 3.** Schematic diagram of wavelet packet time-frequency analysis of vibration signal.

According to Equations (2) and (3) and the wavelet packet time-frequency analysis principle of vibration signal in Figure 3, the time-frequency processing of vibration signal of HVCB was carried out. Using db3 wavelet packet, each vibration signal was subjected to 7-level binary wavelet packet decomposition and wavelet coefficient reconstruction. The system sampling frequency was 300 kHz. The width of each of generated  $2^7$  frequency bands was 1.17 kHz. Considering the vibrating sensor YD-111T had a frequency response range from 0 to 15 kHz, the No. 0~12 wavelet coefficients at the 7th level wavelet packet were employed to characterize the time-frequency characteristics of vibration signal.

### 3.2. Feature Extraction

According to the wavelet packet time-frequency analysis results in Section 3.1 and regarding  $\Delta f = 1.17$  kHz,  $\Delta T = 10$  ms as criteria, the vibration signal during closing process was split into  $13 \times 12$  signal segments, and the energy of each signal segment was calculated using  $E = \sqrt{\sum_{t=t_0}^{t_N} x^2(t)}$ , as shown in Figure 4. In addition, considering the dispersity of vibration signal energy of HVCB and corresponding normalization processing, the proportion of each signal segment in time-frequency direction was calculated using Equation (4).  $P_{i,j}$  and  $Q_{i,j}$  represent the proportion of the  $j$ -th signal in frequency direction within the  $i$ -th period of time as well as the proportion of the  $i$ -th signal in time direction within the  $j$ -th frequency band, respectively.

$$\begin{cases} P_{i,j} = E_{i,j} / \sum_{k=1}^{13} E_{k,j} \\ Q_{i,j} = E_{i,j} / \sum_{k=1}^{12} E_{i,k} \end{cases} \quad (4)$$

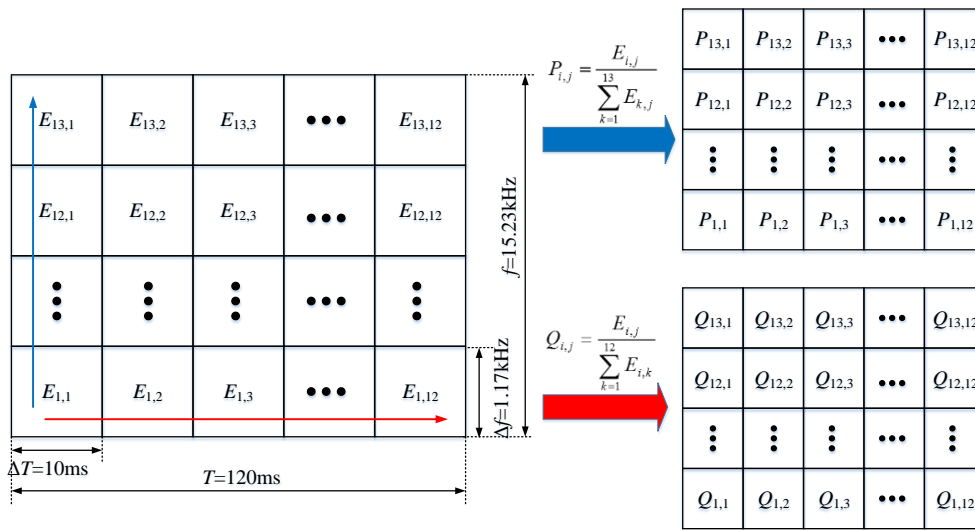


Figure 4. Calculation of wavelet packet time-frequency feature of vibration signal.

In the information theory, Shannon entropy is the most common measure for evaluating the confusion degree of a signal. In fact, it is usually described as an essential feature. Shannon entropy  $\Theta$  is defined as:

$$\Theta = - \sum_{i=1}^N \zeta_i \log \zeta_i \tag{5}$$

where  $\zeta_i$  is the probability of random event  $Y = y_i$  and  $\sum_{i=1}^N \zeta_i = 1$ . When  $\zeta_i = 0$ , there is a convention that  $\zeta_i \log \zeta_i = 0$  [20]. According to Equations (4) and (5), the Shannon entropy feature space  $WTFE = [WTFE_{f_t}, WTFE_{f_f}]$  of wavelet packet analysis matrix can be obtained, as shown in Equation (6).

$$\begin{cases} WTFE_{f_t} = [- \sum_{j=1}^{12} (P_{1,j} \times \ln(P_{1,j})), - \sum_{j=1}^{12} (P_{2,j} \times \ln(P_{2,j})), \dots, - \sum_{j=1}^{12} (P_{13,j} \times \ln(P_{13,j}))] \\ WTFE_{f_f} = [- \sum_{i=1}^{13} (Q_{i,1} \times \ln(Q_{i,1})), - \sum_{i=1}^{13} (Q_{i,2} \times \ln(Q_{i,2})), \dots, - \sum_{i=1}^{13} (Q_{i,12} \times \ln(Q_{i,12}))] \end{cases} \tag{6}$$

Shannon entropy is a type of characteristic description of signal. However, this description of signal is not complete. Taking the three sequences  $\{\zeta_1, \zeta_2, \zeta_3\}$  in Figure 5 as example, from the analysis perspective of Shannon entropy, only the difference between red sequence and other two sequences can be distinguished, i.e., the entropy value is different. Although the green curve and blue curve share identical entropy value (indicating identical dispersity of the sequence), the difference in sequence order still cannot be characterized. Therefore, in the process of feature extracting, the difference in distribution order of sequence was considered and wavelet packet analysis-based time-frequency energy was regarded as HVCB vibration signal feature to form feature space  $WTFER$ , as shown in Equation (7).

$$WTFER = [WTFER_t, WTFER_f] = [P_{1,1}, P_{1,2}, \dots, P_{1,12}, P_{2,1}, \dots, P_{13,12}, Q_{1,1}, \dots, Q_{13,12}] \tag{7}$$

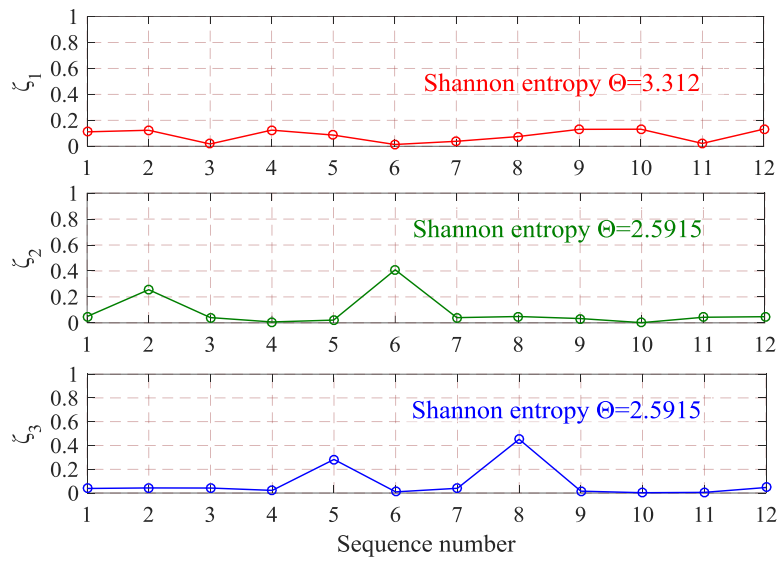
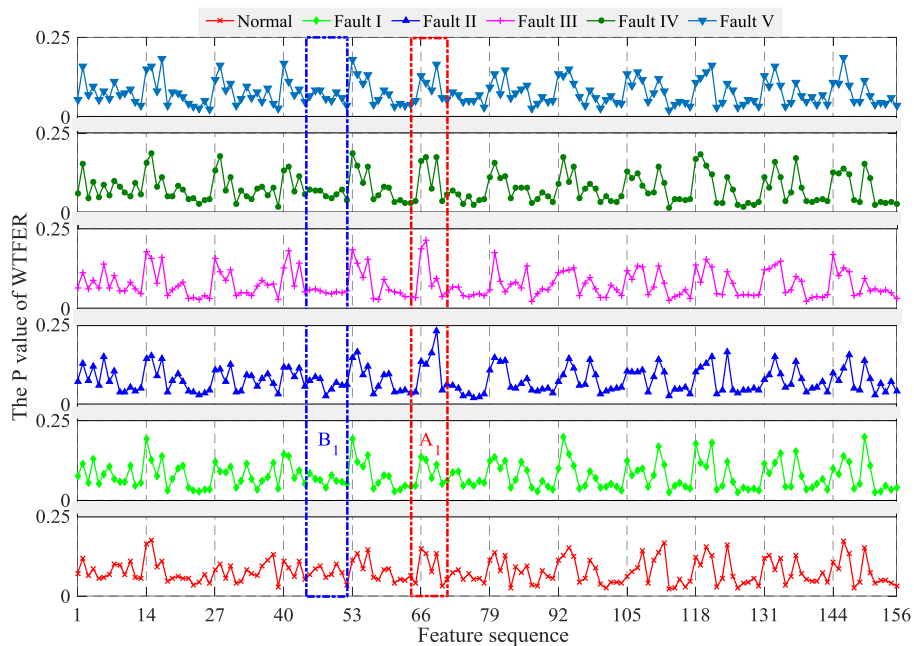


Figure 5. Incomplete description of Shannon entropy.

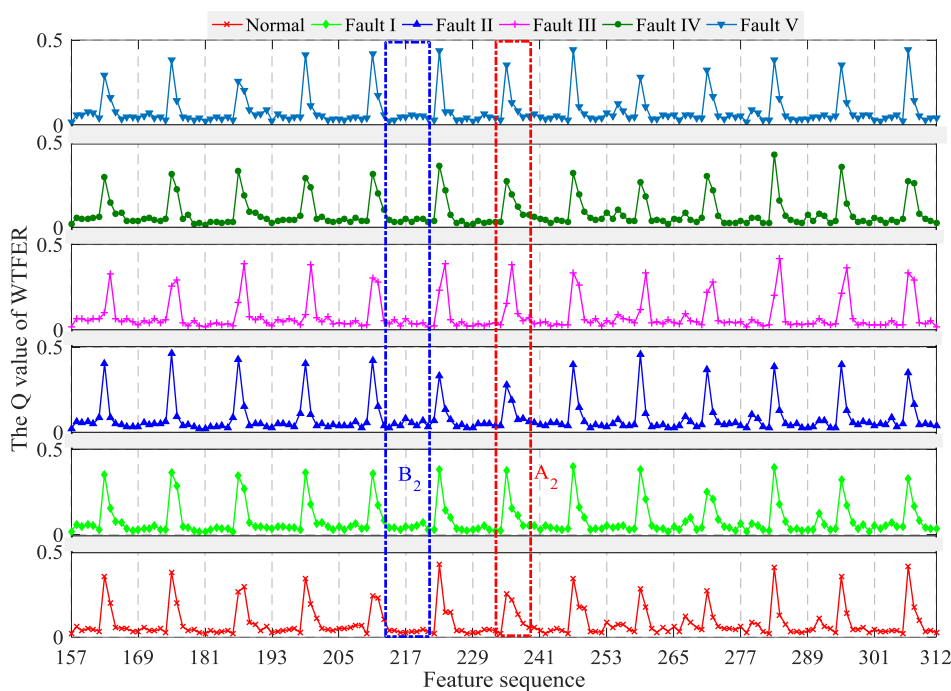
According to Equations (4) and (7), the feature sequence of HVCB vibration signal under different faults was calculated, as shown in Figure 6. It can be seen from the figure that the respective dispersities of feature sequences are identical, and Shannon entropy values are similar to each other. By comparing the feature sequences of different fault samples, it can be seen that there is a large amount of feature variables with no difference or insignificant differences, as shown in  $A_1$  and  $A_2$  of Figure 6. The classifier can make a good classification according to the differences among these features. Meanwhile, there is a large number of uncorrelated features in feature vector, as shown in  $B_1$  and  $B_2$  of Figure 6, which may affect subsequent classifier performance. Therefore, it is very important to reduce feature space dimension or to select a diagnosis algorithm with high accuracy and robustness.



(a) Feature sequence from 1 to 156 for 6 fault types

Figure 6. Cont.





(b) Feature sequence from 157 to 312 for 6 fault types

**Figure 6.** Feature description. A1, A2, B1, B2 represents a length of feature sequence (65~71th, 233~239th, 44~52th, 214~222th feature sequence, respectively).

#### 4. Fault Diagnosis Based on Random Forest

##### 4.1. Classification and Regression Tree

For classification or regression matters, a decision tree is a highly interpretable method with graph description [28] based on partition of the feature space. In general, the growth of a decision tree typically starts at the root node. When the data sample set is not pure, a series of binary tree splitting will happen until the data sample set is pure. The purity of sample data set refers to the proportion of similar samples within a node phase. Based on the value of random one of the explanatory feature variables, two child nodes are generated. The splitting process is persisted endlessly. Ultimately, a terminal node or leaf is reached, and a decision tree model is given, as illustrated in Figure 7.

It is the most important core for the whole algorithm to define the impurity of each node and minimize this impurity with reasonable partitioning. In principle, *Gini* index is a better criterion by which node impurity is minimized in a classification problem [38]. Suppose that there is a random variable  $x$  with a total of  $K$  states, each indexed by  $k$ ; and  $p_k$  is the probability of  $x$  belonging to state  $k$ . Then, the *Gini* index can be expressed as

$$Gini(x) = \sum_{k=1}^K p_k(1 - p_k) = 1 - \sum_{k=1}^K p_k^2 \tag{8}$$

In a node of classification and regression tree (CART), if a partition point  $\Delta_i$  and the  $i$ -th attribute  $C_i$  in feature space are selected, the training data set  $S$  is divided into two parts, namely  $S_1$  and  $S_2$ , then the *Gini* index of CART at a node can be calculated as

$$Gini(S, C_i) = \frac{|S_1|}{|S|} Gini(S_1) + \frac{|S_2|}{|S|} Gini(S_2) \tag{9}$$

$Gini(S, C_i)$  represents the non-determinacy or impurity of training data set  $S$  by the partition point of  $C_i = \Delta_i$ . As the *Gini* index increases, the purity of data set reduces, and the result of splitting becomes

increasingly worse. Therefore, this iterative procedure of binary partition searches the least *Gini* index throughout all possible values within the definition domain of all attribute variables, and finally a whole CART is formed, as shown in Figure 7.

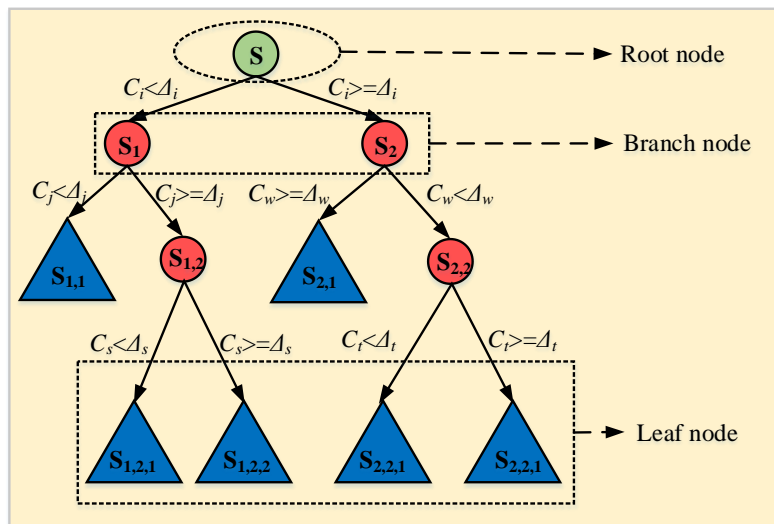


Figure 7. Model representation as a binary tree.

#### 4.2. Random Forest

The vibration information description of HVCB under the same fault is of large dispersity, i.e., the noise ratio is large. The decision tree-based techniques are limited by high variance and instability, so the classification accuracy may be greatly impacted by the noise. When uncorrelated feature variables are too high, the performance of an individual decision tree becomes worse. It is well known that individual classifiers' performance can be improved by an ensemble method that integrates a set of classifiers [32]. The most popular ensemble technique is bagging tree (BT), which proceeds by generating bootstrap samples from the original data set, constructing a CART from each bootstrap sample, and voting to combine. The majority voting rule is given by [40]

$$L = \arg \max_L \sum_{j=1}^{N_{tree}} I(H_j(C_i) == L_k) \quad (10)$$

where  $I(C_i)$  is an indication function;  $\arg()$  is a value function representing the number of trees that classify the test sample as the classification vector  $L$ .

The main principle behind bagging's performance is that the averaging reduces the variance of individual classifiers without increasing their bias. Bagging generally performs well with unstable base learners for whom small changes in the training data can lead to large changes in the learned model [32]. As an improved bagging tree method, random forest (RF) [35] randomly selects large amounts of attributes at each node of CART. Hence, RF has faster training speed, higher prediction accuracy and stronger robustness than CART and BT. Beyond that, based on bagging concept, RF provides an out of bagging (OOB) data-based unbiased estimate of test set error. OOB data are also used to estimate the importance of variables. These two estimates (test set error estimate and variable importance) are the most useful byproducts of RF. The structure of RF model is shown in Figure 8.

To increase classification accuracy, low bias of CART and low correlation of CARTs in forest are essential. To achieve low bias, each CART in forest grows to maximum depth. To obtain low correlation, randomization is applied in two aspects [37]: (1) CART is grown on a bootstrap sample of the training feature set; (2)  $m$  characteristic attributes are randomly selected out of  $d$  available. In general, the parameter  $m$  is set to  $\sqrt{d}$  for classification problem. Table 2 describes the procedure of the RF algorithm.

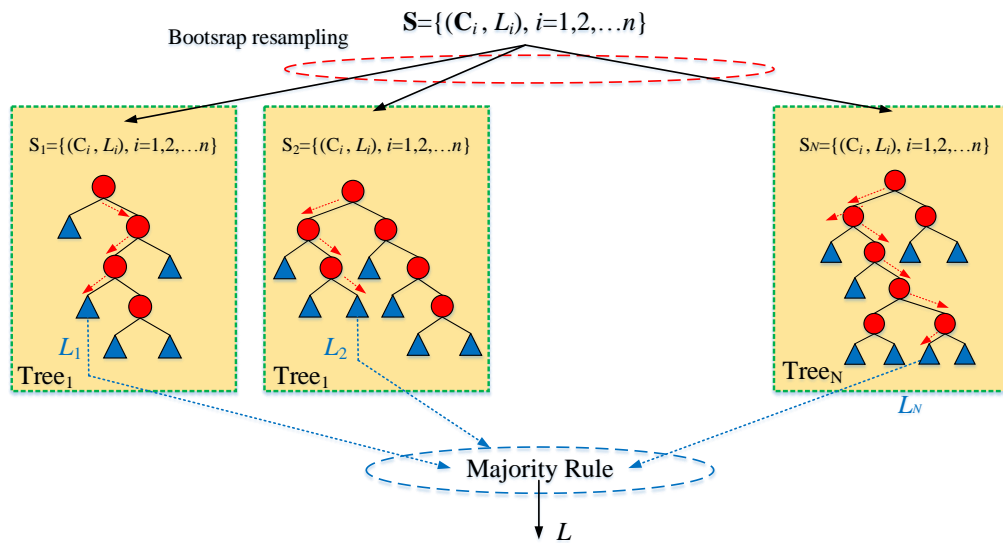


Figure 8. Random forest structure.

Table 2. Random forest algorithm.

**Algorithm 1.** Random Forest**Input:**

Training set of variable fault's vibration signal,  $S = \{(C_i, L_i), I = 1, 2, \dots, n\}$ ,  $(C_i, L_i) \in R^d \times R$ ,  $(C_i, L_i)$  represent feature set and label of the  $i$ -th sampling;

Test sample  $X_i \in R^d$ ;

**Output:**

Collection of trees  $T = \{H_j(C_i), j = 1, 2, \dots, N_{\text{tree}}\}$ ;

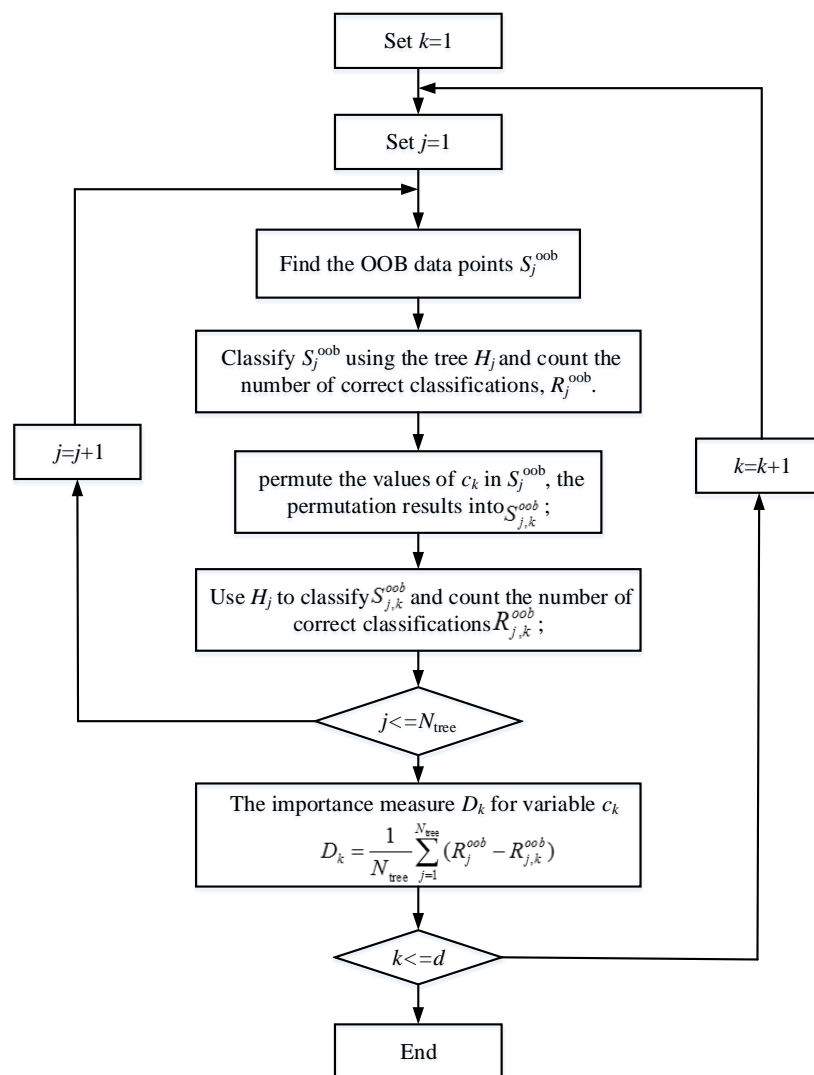
**Process:**

For  $j = 1, 2, \dots, N_{\text{tree}}$

- (1) The training set  $S$  is generated from the original training set with the bootstrap resampling method;
- (2) The decision tree is generated based on CART algorithm.  $m$  characteristic attribute values are randomly selected from  $d$  characteristic attribute values at each node of decision tree;
- (3) The best characteristic value is selected from  $m$  characteristic attribute values at each node according to the *Gini* index;
- (4) A non-pruned tree  $H_j$  is generated;
- (5) Splitting until the tree grows to the maximum, and then random forest is generated
- (6) End.

Test sample  $X_i$  is introduced into random forest and diagnosis result is recorded.

As indicated in various literature, a feature space dimension higher is by no means better during classification process. If there is too much non-essential features during feature extracting process, the main feature may be overwhelmed, which may lead to diagnosis bias. Therefore, much research on accurate estimation of variable importance has been conducted [37]. With the aid of OOB and random forest algorithm, the importance of variables in feature space was estimated. The detailed process of analyzing the importance of the  $k$ -th feature in feature space is shown in Figure 9.



**Figure 9.** The process of analyzing the importance of feature based on random forest algorithm.

As shown in Figure 9, the analysis of feature importance can effectively reduce the dimension of feature space and decrease the impact of non-essential features on diagnosis. In addition, by reducing the dimensions of feature space, the training efficiency and diagnosis accuracy of the diagnostic model can be significantly improved. This paper proposes a strategy for optimizing feature space strategy based on the above estimation process, of which the pseudo-code is shown in Table 3. The RF<sub>1</sub> was established based on original feature space and the error rate of OOB sample was recorded. According to the analysis of feature importance rank in Figure 9, top  $d \times (1 - \lambda)$  important features were adopted to constitute a new feature space, and the second-generation RF model was constructed based on this new feature space to record the error rate. After that, through analyzing the feature importance rank in such feature space, top  $d \times (1 - \lambda)$  important features were selected again to constitute another new feature space. This process was repeatedly performed until the feature space dimension was less than  $\delta$ . The feature space with minimum OOB error rate was adopted for the description and the RF model was employed as the optimal diagnosis model.

**Table 3.** Pseudo-code for optimizing feature space.

---

**Input:**  $S = \{(C_i, L_i), i = 1, 2, \dots, n\}, (C_i, L_i) \in R^d \times R, (C_i, L_i)$  RF parameters, minimum feature space dimension  $\delta$ , reduce rate  $\lambda$ .

1. **Function**
2.  $m = 1$ , design  $RF_m$  with  $S$ , calculate and sort  $D_{k,m}$  based on  $RF_m$ ;
3. **While**  $d > \delta$
4.  $m = m + 1$
5.  $S_m \leftarrow$  select the top  $d \times (1 - \lambda)$  important features to form new feature space  $S_m$  and  $d = d \times (1 - \lambda)$
6.  $RF_m \leftarrow$   $RF_m$  based on  $S_m$ ;
7.  $\{D_{k,m}, E_{\text{Oob},m}\} \leftarrow$  calculate  $D_{k,m}$  and error rate of OOB( $E_{\text{Oob},m}$ );
8. **End**
9. **Return**  $\min(\{E_{\text{Oob},m}\}) \rightarrow RF_m$
10. **End function**

---

## 5. Experimental Result and Analysis

Based on the description of HVCB vibration data samples (including training and test sample) and fault type in Section 2, feature extracting process based on vibration signal time-frequency analysis in Section 3, and the design flow of robust random forest algorithm and feature space optimization based on ensemble learning in Section 4, this section verifies the superiority of the proposed method by comparing the diagnostic performance and optimization results of different classification models with original feature space and optimized feature space.

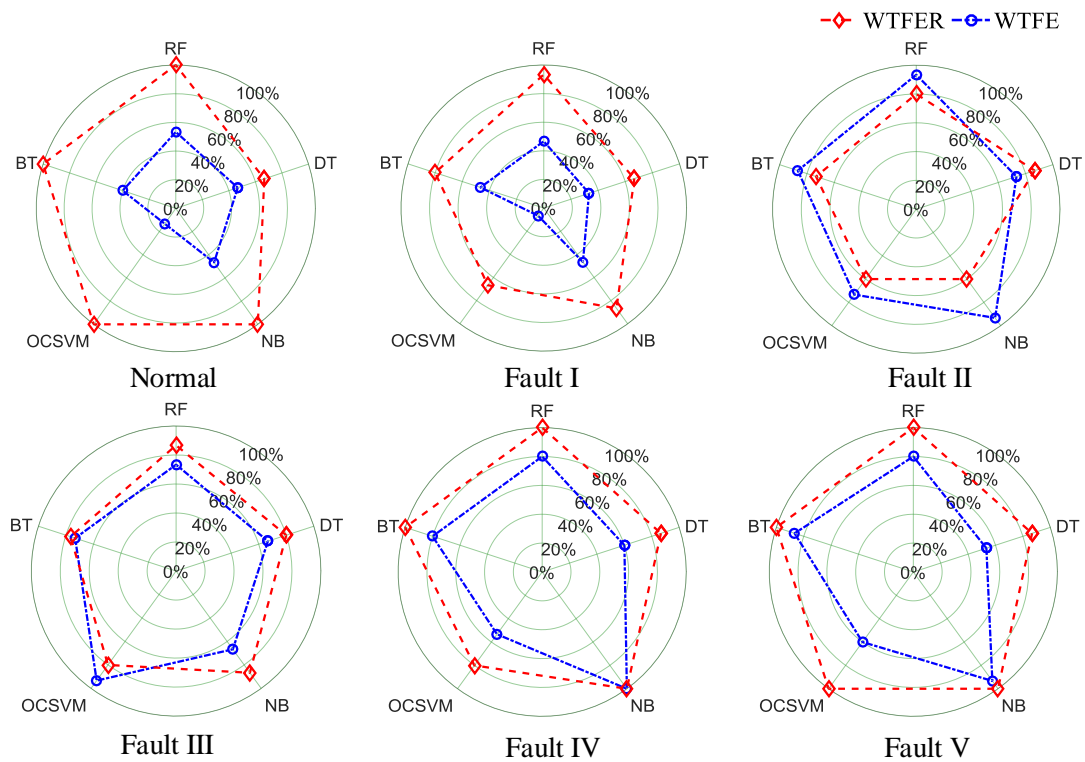
### 5.1. Diagnosis Results with the Origin Feature Space

For each fault condition, there were 50 vibration signal samples. The 50 samples for each fault condition were randomly divided into two parts: training sample set consisting of 35 samples and test sample set consisting of 15 samples. After that, based on training sample set, WTFE and WTFER were selected as the input feature vectors of diagnostic model to train the parameters of different classifier models (DT, NB, OCSVM, BT and RF) Finally, based on test sample set, we compared the diagnosis accuracy rate  $TP_i / (TP_i + FP_i)$  for single fault with the averaged diagnosis accuracy rate  $\sum_i^{n_c} TP_i / \sum_i^{n_c} (TP_i + FP)$  for all faults, as shown in Figure 10, Tables 4 and 5. Where  $TP_i$  ( $FP_i$ ) denotes the number of true (false) positive instances in the  $i$ -th fault and  $n_c$  is the number of the fault categories.

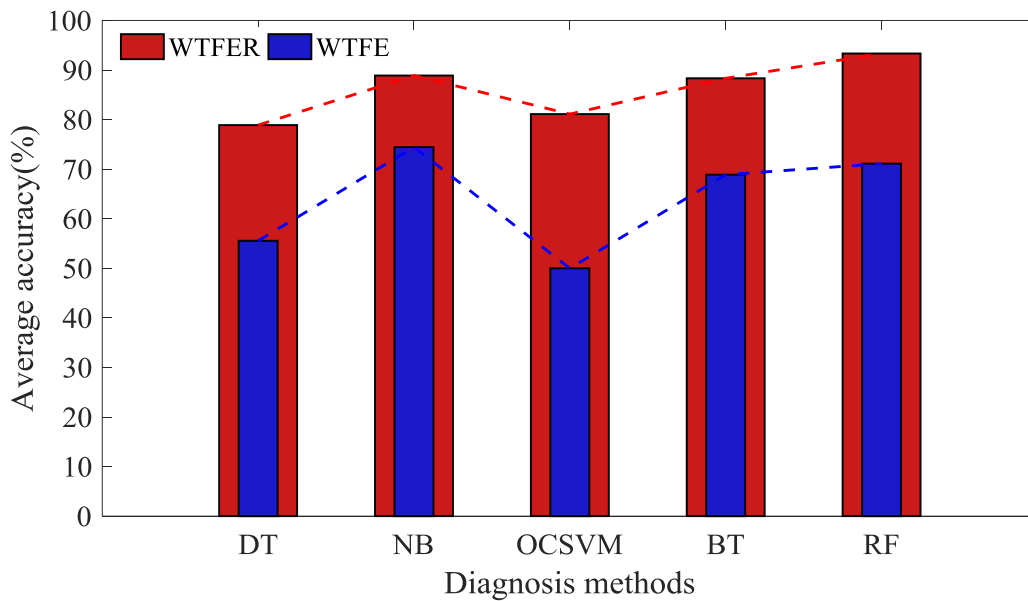
According to Figure 10 and Tables 4 and 5, it can be known that (1) despite any classifier model, the diagnosis accuracy with WTFER parameter as feature vector input is always significantly higher than that with WTFE parameter as input. This indicates the Shannon entropy of signal is an incomplete description of original signal feature, which will greatly impact on diagnosis accuracy. Selecting a proper feature space is beneficial for improving diagnosis accuracy of model; (2) In spite of analysis from diagnosis accuracy of single fault or from average diagnosis accuracy of all faults, the classifier model based on integrated method (BT and RF) always has higher diagnosis accuracy (93.33%) than that of single classifier model (DT-78.89%, NB-88.89% and OCSVM-81.11%), indicating the superiority of integrated learning method.

**Table 4.** WTFER-based diagnosis results.

Fault Type	Accuracy Rate				
	DT	NB	OCSVM	BT	RF
Normal	66.67%	100%	100%	100%	100%
Fault I	66.67%	86.67%	66.67%	80%	93.33%
Fault II	86.67%	60%	60%	73.33%	80%
Fault III	80%	86.67%	80%	86.67%	86.67%
Fault IV	86.67%	100%	80%	100%	100%
Fault V	86.67%	100%	100%	100%	100%
Total	78.89%	88.89%	81.11%	90%	93.33%



(a) Comparison accuracy of each fault type using various methods in the WTFER and WTFE feature space



(b) Comparison accuracy of all test samples using various methods in the WTFER and WTFE feature space

Figure 10. Diagnosis results based on WTFER feature space.

Table 5. WTFE-based diagnosis results.

Fault Type	Accuracy Rate				
	DT	NB	OCSVM	BT	RF
Normal	46.67%	46.67%	13.33%	40%	53.33%
Fault I	33.33%	46.67%	6.67%	46.67%	46.67%
Fault II	73.33%	93.33%	73.33%	86.67%	93.33%
Fault III	66.67%	66.67%	93.33%	73.33%	73.33%
Fault IV	60%	100%	53.33%	80%	80%
Fault V	53.33%	93.33%	60%	86.67%	80%
Total	55.56%	74.45%	50%	68.89%	71.11%

### 5.2. Diagnosis Results with the Optimal Feature Space

Using the random forest model based on WTFER feature space in Section 5.1 and the optimization flow of feature space in Table 3, we optimized the input feature vector of classifier model, realized the description of optimized feature space, and designed optimized random forest classifier model of which the OOB error curve is shown in Figure 11a. The error curves of OOB data diagnosis of random forest under different feature spaces are shown in Figure 11b. Under optimized feature space, the comparisons of feature vectors of different fault samples are shown in Figure 11c.

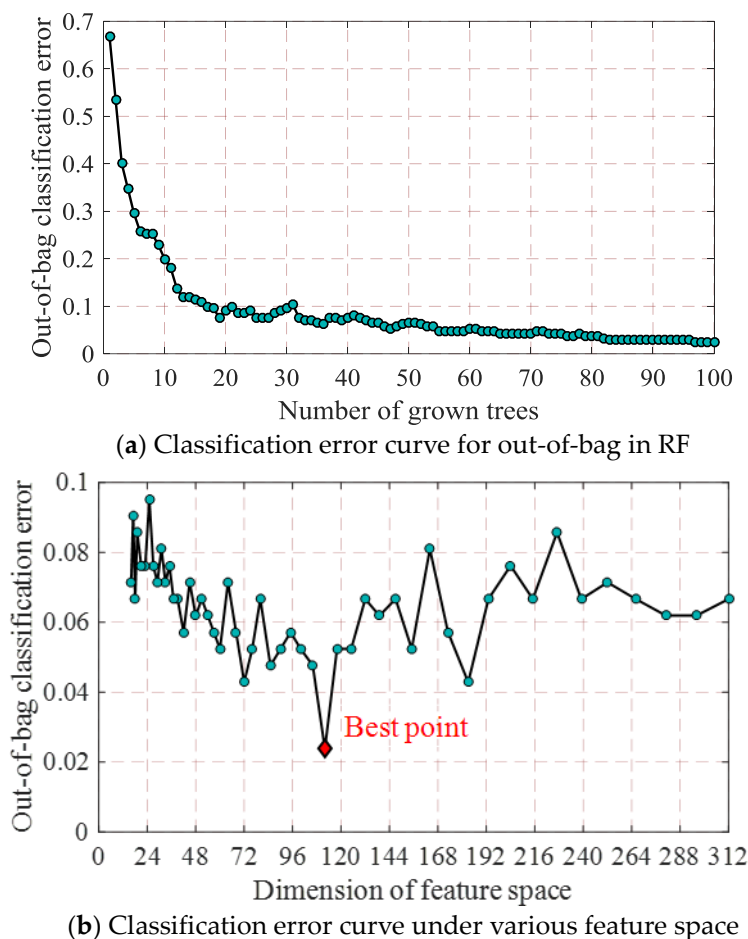
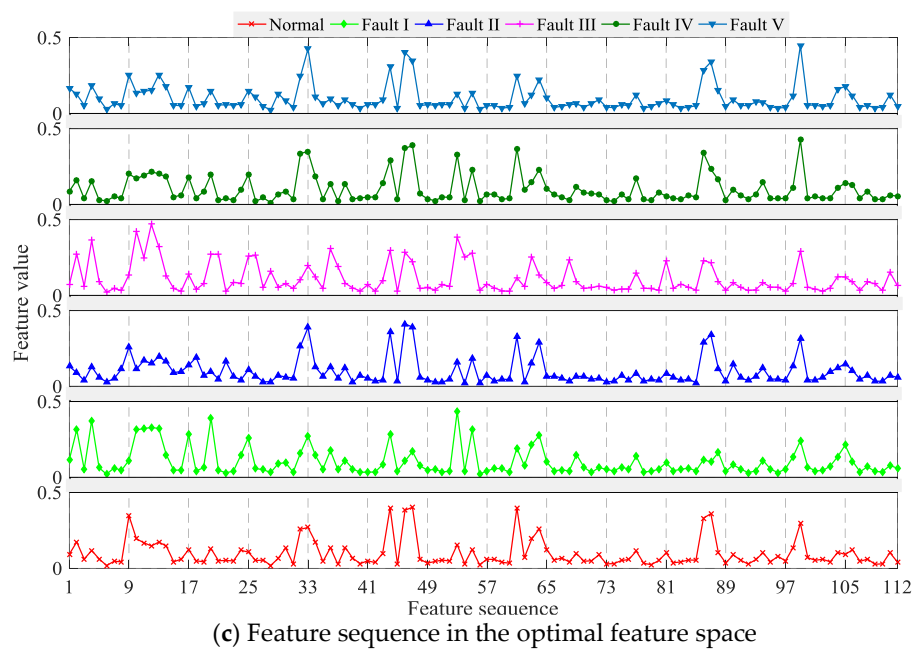


Figure 11. Cont.



**Figure 11.** Calculation results of parameters of optimized feature space.

As shown in Figure 11a, the OOB error of random forest under optimized feature space gradually decreases and tends to be stable, indicating that the model has sound diagnosis performance. As shown in Figure 11b, as the feature space dimension decreases, the OOB error of random forest does not display monotonical increasing or decreasing trend, but there exists an optimized feature space (a point where minimum OOB error locates), i.e., the feature space constituted by top 112 important features. By comparing Figure 11c with Figure 6, it can be known that the optimization process greatly reduced low-correlated variables, decreased the dimension of feature space, so that the reserved features can indicate the difference among faults more clearly.

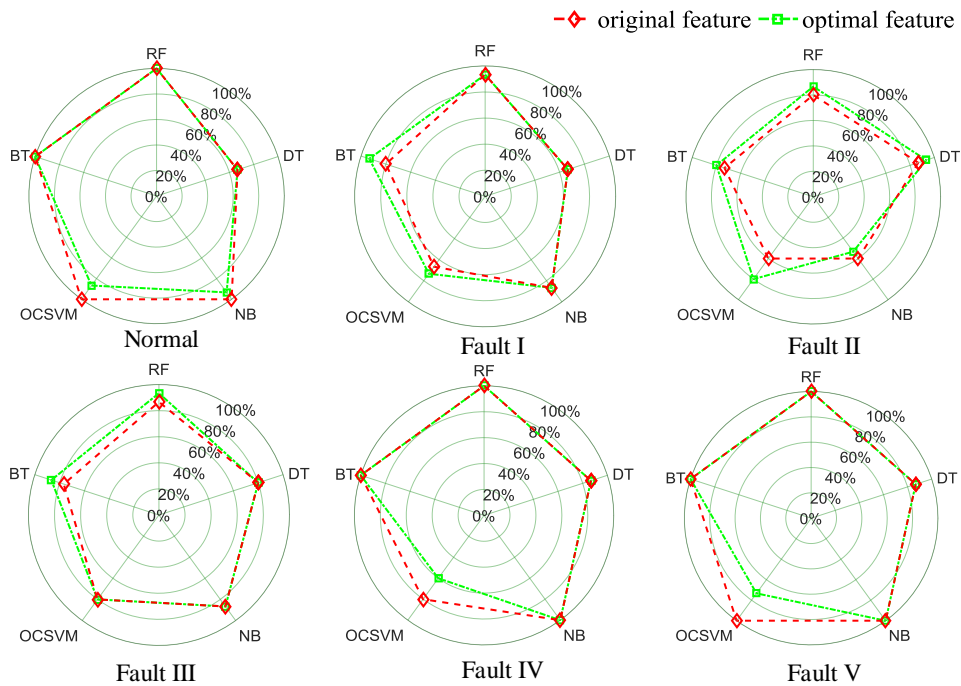
Each classifier model was reconstructed based on optimized feature space. The diagnosis results of different models after feature space optimization were compared based on vibration signal test set, as shown in Figure 12 and Table 6.

By comparing Figures 10 and 12, it can be known that (1) after feature space optimization, the diagnosis accuracy of the random forest model increases to 95.56%, which is significantly higher than that of other classifier models; (2) under optimized feature space, the diagnosis accuracies of all classifiers are increased, indicating that the proposed optimization process is effective, and the removal of low-correlated feature variables is conducive to the improvement of fault diagnosis results.

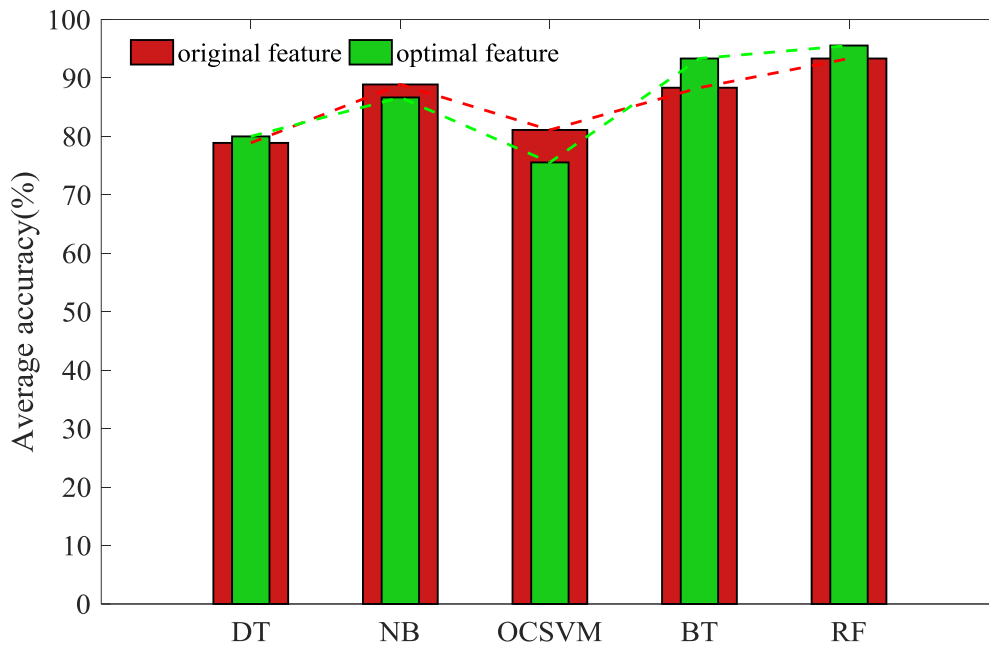
**Table 6.** Diagnosis results after feature space optimization.

Fault Type	Accuracy Rate				
	DT	NB	OCSVM	BT	RF
Normal	66.67%	93.33%	86.67%	100%	100%
Fault I	66.67%	86.67%	73.33%	93.33%	93.33%
Fault II	93.33%	53.33%	80%	80%	86.67%
Fault III	80%	86.67%	80%	86.67%	93.33%
Fault IV	86.67%	100%	60%	100%	100%
Fault V	86.67%	100%	73.33%	100%	100%
Total	80%	86.67%	75.56%	93.33%	95.56%





(a) Comparison accuracy of each fault type using various methods in original and optimal feature space



(b) Comparison accuracy of all test samples using various methods in original and optimal feature space

Figure 12. Diagnosis results after feature space optimization.

### 6. Conclusions

In this work, a novel diagnosis method for HVCB mechanical faults is proposed based on WTFER feature space and random forest model. The experimental results show the superiority of the proposed method in HVCB fault diagnosis. Conclusions can be drawn as follows:

- (1) To timely discover potential risk of HVCB, a precise diagnosis method and detection system for the mechanical fault of HVCB is designed based on the vibration signal under different conditions;
- (2) As the frequency of HVCB vibration signal changes with time, a time-frequency analysis technique is applied in the feature extraction process. Compared with incomplete descriptions of WTFE based on Shannon entropy, a WTFER feature vector describes the vibration signal characteristics more comprehensively and effectively;
- (3) In comparison to the conventional single classification, random forest with bagging technology and random split selection skill achieves better performance and higher accuracy in mechanical fault diagnosis of HVCB;
- (4) As one of the most useful byproducts of random forest algorithm, the assessment of feature importance can be used to optimize the feature space for further improving diagnosis precision of the RF model.

Currently, the vibration information-based diagnostic technique for HVCB is still in the phase of artificial feature extraction. With the development of artificial intelligence technology, automatic identification diagnosis of characteristic space description using deep learning will become the research focus in the future. In addition, the fault normally does not occur alone but occurs multiple times during the operation of HVCB; therefore, applying the single-fault diagnosis model in identification of combined faults will become a key research direction for intelligent fault diagnosis of HVCB.

**Acknowledgments:** This paper is financially supported by the National Natural Science Foundation of China (No. 51377007 and No.51677002) and the Specialized Research Fund for the Doctoral Program of Higher Education in China (No. 20131102130006).

**Author Contributions:** Suliang Ma designed the research method. Mingxuan Chen and Suliang Ma wrote the draft. Yuhao Wang and Suliang Ma contributed to the experimental section. Bowen Jia gave a detailed revision. Jianwen Wu and Yuan Jiang provided important guidance. All authors have read and approved the final manuscript.

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

## Nomenclature

HVCB	High-voltage circuit breakers
DT	Decision tree algorithm
NB	Naive bayes algorithm
OCSVM	One class support vector Machine
BT	Bagger decision tree algorithm
RF	Random forest
OOB	Out of bagging
WPT	wavelet package transform
WTFE	wavelet time-frequency entropy
WTFER	wavelet time-frequency energy rate
$E_{i,j}$	energy in the $i$ -th time and $j$ -th frequency section
$P_{i,j}$	energy rate alone frequency direction
$Q_{i,j}$	energy rate alone time direction
$S$	training sample set
$X$	test sample set
$C_i$	$i$ -th sample feature $\in \mathbf{R}^d$ ,
$d$	feature space dimension
$k$	$k$ -th feature in feature space
$L_i$	$i$ -th sample label (Fault type) $\in R$
$H_j$	$j$ -th decision tree
$T$	collection of trees represent forest
$N_{\text{tree}}$	number of tree in the forest
$S_j^{\text{oob}}$	OOB data set in $j$ -th decision tree

$S_{j,k}^{\text{oob}}$	OOB data set that permuting $k$ -th feature in $S_j^{\text{oob}}$ for $j$ -th decision tree
$R_j^{\text{oob}}$	number of correct classification in $j$ -th decision tree with OOB data
$R_{j,k}^{\text{oob}}$	number of correct classification in $j$ -th decision tree for $S_{j,k}^{\text{oob}}$
$D_k$	important measure for $k$ -th feature
$\delta$	minimum feature space dimension
$\lambda$	reduce rate
$\text{RF}_m$	$m$ -th random forest model
$E_{\text{oob},m}$	error rate of OOB data in $m$ -th random forest model

## References

1. Heising, C.R.; Janssen, A.L.J.; Lanz, W.; Colombo, E.; Dialynas, E.N. Summary of CIGRE 13.06 Working Group worldwide reliability data and maintenance cost data on high voltage circuit breakers above 63 kV. In Proceedings of the 1994 IEEE Industry Applications Society Annual Meeting, Denver, CO, USA, 2–6 October 1994; Volume 3, pp. 2226–2234.
2. Niu, W.; Liang, G.; Yuan, H.; Baoshu, L. A Fault Diagnosis Method of High Voltage Circuit Breaker Based on Moving Contact Motion Trajectory and ELM. *Math. Probl. Eng.* **2016**, *2016*, 1–10. [[CrossRef](#)]
3. Forootani, A.; Afzalian, A.A.; Ghomsheh, A.N. Model-based fault analysis of a high-voltage circuit breaker operating mechanism. *Turk. J. Electr. Eng. Comput. Sci.* **2016**, *25*, 2349–2362. [[CrossRef](#)]
4. Fei, M.; Yi, P.; Zhu, K.; Zheng, J. On-line hybrid fault diagnosis method for high voltage circuit breaker. *J. Intell. Fuzzy Syst.* **2017**, *33*, 1–12. [[CrossRef](#)]
5. Fei, M.; Mei, J.; Zheng, J.; Wang, Y. Development and Application of Distributed Multilayer On-line Monitoring System for High Voltage Vacuum Circuit Breaker. *J. Electr. Eng. Technol.* **2013**, *8*, 813–823.
6. Huang, N.; Fang, L.; Cai, G.; Xu, D.; Chen, H.; Nie, Y. Mechanical Fault Diagnosis of High Voltage Circuit Breakers with Unknown Fault Type Using Hybrid Classifier Based on LMD and Time Segmentation Energy Entropy. *Entropy* **2016**, *18*, 322. [[CrossRef](#)]
7. Landry, M.; Leonard, F.; Landry, C.; Brikci, F. An Improved Vibration Analysis Algorithm as a Diagnostic Tool for Detecting Mechanical Anomalies on Power Circuit Breakers. *IEEE Trans. Power Deliv.* **2008**, *23*, 1986–1994. [[CrossRef](#)]
8. Runde, M.; Ottesen, G.E.; Skyberg, B.; Ohlen, M. Vibration analysis for diagnostic testing of circuit-breakers. *IEEE Trans. Power Deliv.* **1996**, *11*, 1816–1823. [[CrossRef](#)]
9. Huang, J.; Hu, X.; Geng, X. An intelligent fault diagnosis method of high voltage circuit breaker based on improved EMD energy entropy and multi-class support vector machine. *Electr. Power Syst. Res.* **2011**, *81*, 400–407. [[CrossRef](#)]
10. Liu, M.; Wang, K.; Sun, L.; Zhen, J. Applying empirical mode decomposition (EMD) and entropy to diagnose circuit breaker faults. *Optik-Int. J. Light Electron Opt.* **2015**, *126*, 2338–2342.
11. Zhang, J.; Liu, M.; Wang, K.; Sun, L. Mechanical Fault Diagnosis for HV Circuit Breakers Based on Ensemble Empirical Mode Decomposition Energy Entropy and Support Vector Machine. *Math. Probl. Eng.* **2015**, *2015*. [[CrossRef](#)]
12. Huang, N.; Chen, H.; Cai, G.; Fang, L.; Wang, Y. Mechanical Fault Diagnosis of High Voltage Circuit Breakers Based on Variational Mode Decomposition and Multi-Layer Classifier. *Sensors* **2016**, *16*, 1887. [[CrossRef](#)] [[PubMed](#)]
13. Gargour, C.; Gabrea, M.; Ramachandran, V.; Lina, J.-M. A short introduction to wavelets and their applications. *IEEE Circuits Syst. Mag.* **2009**, *9*, 57–68. [[CrossRef](#)]
14. Wang, Y.; Xu, G.; Lin, L.; Jiang, K. Detection of weak transient signals based on wavelet packet transform and manifold learning for rolling element bearing fault diagnosis. *Mech. Syst. Signal Process.* **2015**, *54–55*, 259–276. [[CrossRef](#)]
15. Yan, R.; Gao, R.X.; Chen, X. Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Process.* **2014**, *96*, 1–15. [[CrossRef](#)]
16. Kumar, A.; Kumar, R. Time-frequency analysis and support vector machine in automatic detection of defect from vibration signal of centrifugal pump. *Measurement* **2017**, *108*, 119–133. [[CrossRef](#)]
17. Hussain, A.; Lee, S.J.; Choi, M.S.; Brikci, F. An expert system for acoustic diagnosis of power circuit breakers and on-load tap changers. *Expert Syst. Appl.* **2015**, *42*, 9426–9433. [[CrossRef](#)]

18. Ni, J.; Zhang, C.; Yang, S.X. An Adaptive Approach Based on KPCA and SVM for Real-Time Fault Diagnosis of HVCBs. *IEEE Trans. Power Deliv.* **2011**, *26*, 1960–1971. [[CrossRef](#)]
19. Huang, J.; Hu, X.; Yang, F. Support vector machine with genetic algorithm for machinery fault diagnosis of high voltage circuit breaker. *Measurement* **2011**, *44*, 1018–1027. [[CrossRef](#)]
20. Huang, N.; Chen, H.; Zhang, S.; Cai, G.; Li, W.; Xu, D.; Fang, L. Mechanical Fault Diagnosis of High Voltage Circuit Breakers Based on Wavelet Time-Frequency Entropy and One-Class Support Vector Machine. *Entropy* **2016**, *18*, 7. [[CrossRef](#)]
21. Marinaki, M.; Marinakis, Y.; Stavroulakis, G.E. Vibration control of beams with piezoelectric sensors and actuators using particle swarm optimization. *Expert Syst. Appl.* **2011**, *38*, 6872–6883. [[CrossRef](#)]
22. Bououden, S.; Chadli, M.; Karimi, H.R. An ant colony optimization-based fuzzy predictive control approach for nonlinear processes. *Inform. Sci.* **2015**, *299*, 143–158. [[CrossRef](#)]
23. Yang, T.; Qiu, W.; Ma, Y.; Chadli, M.; Zhang, L. Fuzzy model-based predictive control of dissolved oxygen in activated sludge processes. *Neurocomputing* **2014**, *136*, 88–95. [[CrossRef](#)]
24. Chadli, M.; Borne, P. *Multiple Models Approach in Automation: Takagi-Sugeno Fuzzy Systems*; John Wiley & Sons: Hoboken, NJ, USA, 2012; p. 256, ISBN 978-1-848-21412-5.
25. Zhu, K.; Mei, F.; Zheng, J. Adaptive fault diagnosis of HVCBs based on P-SVDD and P-KFCM. *Neurocomputing* **2017**, *240*, 127–136. [[CrossRef](#)]
26. Seera, M.; Lim, C.P.; Chu, K.L. Motor fault detection and diagnosis using a hybrid FMM-CART model with online learning. *J. Intell. Manuf.* **2016**, *27*, 1–13. [[CrossRef](#)]
27. Seera, M.; Lim, C.P.; Tan, S.C.; Loo, C.K. A hybrid FAM-CART model and its application to medical data classification. *Neural Comput. Appl.* **2015**, *26*, 1799–1811. [[CrossRef](#)]
28. Azar, A.T.; El-Metwally, S.M. Decision tree classifiers for automated medical diagnosis. *Neural Comput. Appl.* **2013**, *23*, 2387–2403. [[CrossRef](#)]
29. Vernekar, K.; Kumar, H.; Gangadharan, K.V. Engine gearbox fault diagnosis using empirical mode decomposition method and Naïve Bayes algorithm. *Sādhanā* **2017**, *42*, 1–11.
30. Glowacz, A. Recognition of Acoustic Signals of Induction Motors with the Use of MSAF10 and Bayes Classifier. *Arch. Metall. Mater.* **2016**, *61*, 153–158. [[CrossRef](#)]
31. Muralidharan, V.; Sugumaran, V. A comparative study of Naïve Bayes classifier and Bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis. *Appl. Soft Comput.* **2012**, *12*, 2023–2029. [[CrossRef](#)]
32. Collell, G.; Prelec, D.; Patil, K.R. A simple plug-in bagging ensemble based on threshold-moving for classifying binary and multiclass imbalanced data. *Neurocomputing* **2018**, *275*, 330–340. [[CrossRef](#)] [[PubMed](#)]
33. Ahmed, M.; Rasool, A.G.; Afzal, H.; Siddiqi, I. Improving Handwriting based Gender Classification using Ensemble Classifiers. *Expert Syst. Appl.* **2017**, *85*, 158–168. [[CrossRef](#)]
34. Zhang, X.L.; Wang, B.J.; Chen, X.F. Intelligent fault diagnosis of roller bearings with multivariable ensemble-based incremental support vector machine. *Knowl.-Based Syst.* **2015**, *89*, 56–85. [[CrossRef](#)]
35. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
36. Biau, G.; Devroye, L.; Lugosi, G. Consistency of Random Forests and Other Averaging Classifiers. *J. Mach. Learn. Res.* **2009**, *9*, 2015–2033.
37. Verikas, A.; Gelzinis, A.; Bacauskiene, M. Mining data with random forests: A survey and results of new tests. *Pattern Recognit.* **2011**, *44*, 330–349. [[CrossRef](#)]
38. Cerrada, M.; Zurita, G.; Cabrera, D.; Sánchez, R.-V.; Artés, M.; Li, C. Fault diagnosis in spur gears based on genetic algorithm and random forest. *Mech. Syst. Signal Process.* **2016**, *70–71*, 87–103. [[CrossRef](#)]
39. Seera, M.; Wong, M.L.D.; Nandi, A.K. Classification of Ball Bearing Faults using a Hybrid Intelligent Model. *Appl. Soft Comput.* **2017**, *57*, 427–435. [[CrossRef](#)]
40. Wang, Z.; Zhang, Q.; Xiong, J.; Xiao, M.; Sun, G.; He, J. Fault Diagnosis of a Rolling Bearing Using Wavelet Packet Denoising and Random Forests. *IEEE Sens. J.* **2017**, *17*, 5581–5588. [[CrossRef](#)]

