Mechanical Fault Diagnosis of High Voltage Circuit Breakers Based on Wavelet Time-Frequency Entropy and One-Class Support Vector Machine

Nantian Huang [1]**, Huaijin Chen** [1,†]**, Shuxin Zhang** [1,*,†]**, Guowei Cai** [1]**, Weiguo Li** [1]**, Dianguo Xu** [2] **and Lihua Fang** [1]

[1] School of Electrical Engineering, Northeast Dianli University, Jilin 132012, China;
huangnantian@126.com (N.H.); chjin1990@126.com (H.C.); caigw@mail.nedu.edu.cn (G.C.);
lionwg@163.com (W.L.); zang_fang0412@163.com (L.F.)

[2] Department of Electrical Engineering, Harbin Institute of Technology, Harbin 150001, China;
xudiang@hit.edu.cn

* Correspondence: zhang_shu_xin@126.com; Tel.: +86-159-8111-8906; Fax: +86-432-6480-6691

† These authors contributed equally to this work.

Abstract: Mechanical faults of high voltage circuit breakers (HVCBs) are one of the most important factors that affect the reliability of power system operation. Because of the limitation of a lack of samples of each fault type; some fault conditions can be recognized as a normal condition. The fault diagnosis results of HVCBs seriously affect the operation reliability of the entire power system. In order to improve the fault diagnosis accuracy of HVCBs; a method for mechanical fault diagnosis of HVCBs based on wavelet time-frequency entropy (WTFE) and one-class support vector machine (OCSVM) is proposed. In this method; the S-transform (ST) is proposed to analyze the energy time-frequency distribution of HVCBs' vibration signals. Then; WTFE is selected as the feature vector that reflects the information characteristics of vibration signals in the time and frequency domains. OCSVM is used for judging whether a mechanical fault of HVCBs has occurred or not. In order to improve the fault detection accuracy; a particle swarm optimization (PSO) algorithm is employed to optimize the parameters of OCSVM; including the window width of the kernel function and error limit. If the mechanical fault is confirmed; a support vector machine (SVM)-based classifier will be used to recognize the fault type. The experiments carried on a real SF_6 HVCB demonstrated the improved effectiveness of the new approach.

Keywords: high voltage circuit breakers; mechanical fault diagnosis; S-transform; wavelet time-frequency entropy; one-class support vector machine

1. Introduction

High voltage circuit breakers (HVCBs) play an important role in the protection and control of power systems. The faults of HVCBs will directly affect the running of the power system. Mechanical faults of HVCBs' mechanical operation mechanism are the major reasons for HVCBs faults. Therefore, research on fault diagnosis methods for HVCBs is very important for the stable operation of electric power systems. The traditional scheduled maintenance scheme will result in frequent operations and excessive overhauls. It may lead to needless intervention, and even cause HVCB faults during the maintenance [1–4]. The International Council on Large Electric Systems (CIGRE) made an investigation on the causes of failure of HVCBs. They found that 44% of main faults and 39% of secondary faults are mechanical faults [5]. An extensive diagnostic testing of circuit breakers in [4] shows that vibration

analysis is a reliable and appropriate method for non-invasive diagnostic testing. Vibration analysis is an effective signal-based approach of fault diagnosis [6]. Over the past decade, the vibration signatures generated during the operation of mechanical structure have been used for condition monitoring and fault diagnosis with good effects [7–12].

A HVCB's vibration signal is a typical non-stationary signal with strong transients. The existing vibration signal processing methods such as short-time energy [7], dynamic time warping (DTW) [8], wavelet packet transform (WPT) [9] and empirical mode decomposition (EMD) [10,11] have all achieved good results in this area. However, these methods also have some disadvantages. Dynamic time warping and short-time energy are based on the original signal. They are only sensitive to the signal changes over time. When wavelet packet decomposition is used, an appropriate wavelet basis is difficult to select. The EMD method has the indication of end effect and high computational complexity. S-transform (ST) is an effective method for time-frequency analysis. It has a multiple time-frequency resolution with the frequency by using a Gaussian window with variable window width inversely with the frequency [13]. Therefore, ST can satisfy the time-frequency analysis resolution requirements of vibration signals in different frequency domains. Besides, ST can be derived by the fast Fourier transform (FFT). Thus it is easy to realize in engineering applications. The output of ST is a two-dimensional time-frequency matrix. The characteristics of vibration signals in both the time domain and frequency domain can be fully extracted from the matrix.

As a description of the randomness status of a chaotic system, Shannon entropy contains the information characteristics of complex signals. It is suitable for feature extraction in non-stationary signals analysis. Since Shannon introduced the concept of entropy in 1948 [14], many types of information entropy have been widely used in many areas. Wavelet entropy (WE) combines the advantages of Shannon entropy and WT. It has been widely used for non-stationary signals analysis in diverse fields such as power quality transient analysis [15–17], biomedicine [18,19], fault detection and fault diagnosis [20,21]. In this paper, WTFE is used to describe the unique time-frequency characters of different HVCB mechanical statuses by vibration signal analysis.

Neural networks (NNs) [9] and SVM [10,11] have made a significant contribution to fault recognition of HVCBs. Because HVCBs generally operate infrequently, it is quite difficult to get enough vibration samples of different types of HVCB mechanical faults for training multi-class classifiers. Obviously, the fault recognition of HVCBs is a classification problem with small samples. Therefore, multi-class classification methods such as NNs which rely on lots of training samples are not appropriate for analyzing the mechanical status and identifying HVCB faults. SVM on the other hand is suitable for classification problems with small training sample sets. In order to analyze the status of HVCBs, all types of fault samples should be included in the SVM training set. However, not all types of mechanical fault samples of HVCBs are accessible to obtain. Some types of fault samples cannot be acquired in large quantities, so we are unable to cover the complete range of fault characteristics. Classification boundary deviation will be caused by the limited fault types with unbalanced fault samples Thus, some fault samples are easily mistakenly recognized as normal samples. The extreme learning machine (ELM) [22] and pairwise-coupled relevance vector machine (PCRVM) [23] have shown good effects in fault diagnosis in gas turbine generator systems, but applications in mechanical fault diagnosis of HVCBs have not been reported.

One-class classifier is a kind of pattern recognition method which can be trained by normal samples. It is suitable for classifying a small sample set. One-class support vector machine (OCSVM) [24] has great potential in the field of fault detection [25,26]. It can effectively determine whether the equipment is working in a fault condition or not. The parameter optimization is an important step that affects the classification performance of OCSVM [27]. Particle swarm optimization (PSO) is a widely used optimal method that optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality [28]. It can be used to calculate the factor such as the width factor of a kernel function to improve the classification ability of OCSVM.

This paper presents a new ST and OCSVM-based approach for HVCBs mechanical fault diagnosis. Firstly, the ST is used to process vibration signals to analyze the energy distribution in the time-frequency area. Secondly, the WTFE of a ST matrix (STM) is calculated to construct the feature vectors for describing the energy distribution of HVCB vibration signals in the time-frequency area. Then, a PSO-based OCSVM (PSO-OCSVM) which is just trained by the normal training samples is used to separate the normal and fault conditions of HVCB's mechanical operation structure. A PSO algorithm is used to optimize the parameters and improve the classification ability of traditional OCSVM. Finally, if the conditions of HVCBs are judged as a fault condition by PSO-OCSVM, the type of mechanical fault is recognized by a SVM-based classifier. Three different types of faults are simulated in a field experiment on a real HVCB to verify the validity of the new method.

2. S-Transform

The S-transform was proposed by Stockwell in 1996 [13]. The ST result $S(\tau, f)$ of an input signal $h(t)$ is defined by:

$$S(\tau, f) = \int_{-\infty}^{\infty} h(t)w(\tau - t, f)e^{-i2\pi ft}dt \tag{1}$$

$$w(t, f) = \frac{|f|}{\sqrt{2\pi}}e^{-t^2 f^2/2} \tag{2}$$

where $w(t, f)$ is the Gaussian window function. The parameter τ is a displacement factor and controls the location of the Gaussian window in the time axis. f is a parameter related to the width of Gaussian window.

As the inheritance and development of the continuous wavelet transform (CWT), ST can be derived by CWT. The one-dimensional CWT $W(\tau, d)$ of a signal $h(t)$ is defined as:

$$W(\tau, d) = \int_{-\infty}^{\infty} h(t)\psi(t - \tau, d)dt \tag{3}$$

where $\psi(t - \tau, d)$ is a mother wavelet; τ is a displacement factor; d is a scale factor. The scale factor d determines the width of the mother wavelet, while the scale factor τ determines the time location where the signal $h(t)$ is analyzed.

Let the dilation factor d as the inverse of the frequency f, *i.e.*, $d = 1/f$. Along with Equations (1) and (3), ST can be considered as a CWT with a special mother wavelet multiplied by the phase factor:

$$S(\tau, f) = e^{-i2\pi f\tau}W(\tau, f) \tag{4}$$

where the special mother wavelet is defined as the product of the Gaussian window and a complex vector:

$$\psi(t, f) = \frac{|f|}{\sqrt{2\pi}}e^{-t^2 f^2/2}e^{-i2\pi ft} \tag{5}$$

The result of ST can be calculated based on Equations (3)–(5):

$$S(\tau, f) = \int_{-\infty}^{\infty} h(t)\frac{|f|}{\sqrt{2\pi}}e^{-(\tau-t)^2 f^2/2}e^{-i2\pi ft}dt \tag{6}$$

It is obvious that Equation (6) is equal to Equations (1) and (2). Note that Equation (6) is not a strict CWT because the wavelet in Equation (5) does not satisfy the condition of zero mean for an admissible wavelet.

The frequency spectrum of ST is as follows:

$$H(f) = \int_{-\infty}^{\infty} S(\tau, f)d\tau = \int_{-\infty}^{\infty} h(t)e^{-i2\pi ft}dt \tag{7}$$

Likewise, ST result of a signal $h(t)$ can be derived by the Fourier transform, that is:

$$S(\tau, f) = \int_{-\infty}^{+\infty} H(\beta + f) \quad e^{-2\pi^2\beta^2/f^2} e^{i2\pi\beta\tau} d\beta, \, (f \neq 0) \tag{8}$$

where $H(f)$ is the spectrum of the signal $h(t)$, β is the frequency which controls the movement of Gaussian window on the frequency axis.

Let $f \to n/NT$ and $\tau \to jT$, the discrete ST can be denoted as:

$$\begin{cases} S\left[jT, \dfrac{n}{NT}\right] = \displaystyle\sum_{m=0}^{N-1} H\left[\dfrac{m+n}{NT}\right] \quad e^{-2\pi^2 m^2/n^2} \, e^{i2\pi mj/N}, \, n \neq 0 \\ S[jT, 0] = \dfrac{1}{N}\displaystyle\sum_{m=0}^{N-1} h\left(\dfrac{m}{NT}\right), \qquad\qquad\qquad\quad n = 0 \end{cases} \tag{9}$$

ST has variable time-frequency resolution. The result of ST is a two-dimensional complex matrix, called the S-matrix. With a modulus operation, we can get the module matrix of ST (STMM). The column vectors of STMM reflect the amplitude-frequency characteristics and the row vectors describe the time domain distribution of signals at a certain frequency. ST can describe the characteristics of the signal in both the time and frequency domains. Compared with WT, the decomposition of ST in the high frequency part is more detailed. The frequency resolution and anti-noise performance of ST are better than those of WT [29], therefore, ST is suitable for vibration signal processing.

3. Feature Extraction from STMM Based on Wavelet Time-Frequency Entropy

3.1. Wavelet Time-Frequency Entropy

Shannon entropy is an important part of information theory, which describes the degree of confusion of a system. The more orderly the system is, the smaller the entropy is. Shannon entropy H is defined as:

$$H = -\sum_{i=1}^{N} p_i \log p_i \tag{10}$$

where p_i is the probability of random event $Y = y_i$ and $\sum_{i=1}^{N} p_i = 1$. When $p_i = 0$, there is a convention that $p_i \log p_i = 0$.

As a powerful tool for analyzing the transient features of non-stationary signals, wavelet entropy is the combination of Shannon entropy and wavelet transform. This combination not only retains the localized features in time-frequency domains of wavelet analysis, but also embodies the representational capacity of Shannon entropy. The distributions of different kinds of fault signals in wavelet phase space are different. Several types of wavelet entropy are defined based on different principles or processing methods, such as wavelet energy entropy (WEE), wavelet time entropy (WTE), wavelet singular entropy (WSE) and wavelet time-frequency entropy (WTFE) [15]. WEE and WTE indicate the information characteristics of a signal in the time domain and fail to indicate the characteristics in the frequency domain. Regarding the fault types related to the frequency such as lack of mechanical lubrication, the two methods will appear powerless. WSE can map the correlative wavelet space into the independent linearity space and indicate the uncertainty of the energy distribution of a signal in the time-frequency domain. It is highly sensitive to the transients. Since at any moment the vibration signals of HVCBs have strong transients, the WSE results of the same type of vibration signals may present a distinct difference, thus WSE will not be suitable for extracting features of vibration signals in this study.

WTFE is composed of two vectors. The first vector stretches over the whole time space and describes the characteristics of the signal in the time domain. The second vector stretches over the whole frequency space and describes the characteristics of the signal in the frequency domain. In other

words, WTFE can measure the information features of the signal at any given instant and frequency. Therefore, the WTFE is often used in the field of fault diagnosis and detection. The definition of WTFE is as follows: let $D_j(k)$ $(j = 1, 2, \cdots, m; k = 1, 2, \cdots, n)$ be the discrete wavelet presentation, and $E_j(k) = |D_j(k)|^2$ denotes the wavelet energy at scale j and instant k. WTFE is denoted as:

$$W_{TFE} = [W_{TFEt}, W_{TFEf}] \tag{11}$$

where:

$$\begin{cases} W_{TFEt} = -\sum_{j=1}^{m} P_t \ln P_t \\ W_{TFEf} = -\sum_{k=1}^{n} P_f \ln P_f \end{cases} \tag{12}$$

where the probability P_t and P_f are defined as follows:

$$\begin{cases} P_t = E_j(k)/\sum_{j=1}^{m} E_j(k) \\ P_f = E_j(k)/\sum_{k=1}^{n} E_j(k) \end{cases} \tag{13}$$

Similarly, the definition of WSE is as follows: let **D** be a $m \times n$ matrix constituted by $D_j(k)$. According to singular value decomposition theory, for any $m \times n$ matrix, there exist a $m \times r$ matrix **U**, a $r \times n$ matrix **V** and a $r \times r$ diagonal matrix $\mathbf{\Lambda}$, which make:

$$\mathbf{D} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^{\mathrm{T}} \tag{14}$$

where the diagonal elements λ_l ($l = 1, 2, \cdots, r$) of $\mathbf{\Lambda}$ are called singular values of matrix **A**. The singular values are all non-negative and arranged in a non-increasing order (*i.e.*, $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_r \geqslant 0$). Then the WSE is defined as:

$$W_{SE} = -\sum_{l=1}^{r} p_l \ln p_l \tag{15}$$

where the probability p_l associated with λ_l is defined as:

$$p_l = \lambda_l / \sum_{l=1}^{r} \lambda_l \tag{16}$$

3.2. Feature Vector Extraction

ST can be considered as a special WT. Thus wavelet entropy theory is also applicable to the feature extraction of signals based on ST. A partition method for the time-frequency plane of the S-matrix is proposed to extract vibration signal features in the time-frequency area. After statistical analysis, we found that the amplitude of vibration signals of HVCBs in the frequency area higher than 10 kHz is very small. Therefore, this paper mainly analyzes the frequency area from 0 Hz to 10 kHz. Firstly, a time-frequency plane which frequency area ranges from 0 Hz to 10 kHz and time area from 0 to 150 ms ("0" is the moment the system receives the operating signal) is constructed by ST. Then, the time-frequency plane is divided into 300 congruent time-frequency blocks and the band-width and time-width of the time-frequency blocks are 1 kHz and 5 ms. The partition method is shown in Figure 1.

Let E_{ij} be the energy of time-frequency block $S_{ij}(i = 1, 2, \ldots, 10; j = 1, 2, \ldots, 30)$. Then E_{ij} is the sum of all elements in the block. Let E be the total energy of the whole time-frequency plane. A normalization processing for E_{ij} is given as:

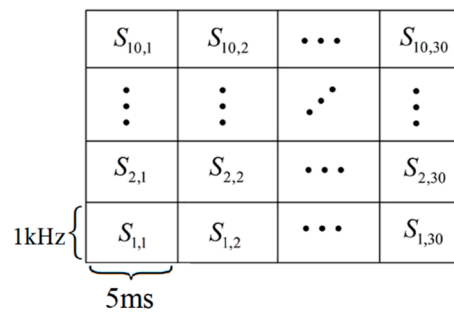$$\hat{E}_{ij} = E_{ij}/E \tag{17}$$

Figure 1. The partition of time-frequency plane.

The time component of WTFE of vibration signals is calculated by:

$$T_j = -\sum_{i=1}^{10} \hat{E}_{ij} \cdot \ln\hat{E}_{ij}, j = 1, 2, \ldots, 30 \tag{18}$$

Similarly, the frequency component of WTFE of vibration signals is calculated by:

$$F_i = -\sum_{j=1}^{30} \hat{E}_{ij} \cdot \ln\hat{E}_{ij}, i = 1, 2, \ldots, 10 \tag{19}$$

The feature vector of vibration signals is denoted as $\mathbf{Z} = [\mathbf{T}\mathbf{F}]$, where $\mathbf{T} = [T_1, \ldots, T_{30}]$ and $\mathbf{F} = [F_1, \ldots, F_{12}]$. Then $\mathbf{Z}$ is used as the input vector of OCSVM and SVM classifier.

4. Condition and Fault Classifier Based on OCSVM and SVM

SVM has good classification ability for classification problems involving small samples and high dimension data. However, because some fault training samples are difficult to obtain, there is a risk that the SVM will easily recognize fault samples as normal samples. In order to avoid this defect, the new method firstly utilizes OCSVM to accurately determine whether a HVCB mechanical failure has happened or not. When the fault is confirmed, the fault type is then identified by the SVM.

4.1. One-Class Support Vector Machine

One-class classification is an important pattern recognition methodology. It can be applied to the fields where negative samples are hard to obtain, such as fault detection, fault diagnosis, intrusion detection and disease analysis, *etc*. Compared to traditional classifiers which aim to obtain the highest recognition accuracy, the target of a one-class classifier is to identify the abnormal samples as far as possible. The latter is able to reduce the possibility that fault states will be mistaken for normal states. Therefore, a one-class classifier is appropriate for the mechanical fault diagnosis of HVCBs with a high reliability.

OCSVM is a mature and effective one-class classifier which was presented by Schölkopf *et al.* [24]. It has good fault analysis performance features, including faster training and decision speeds, lower dependence on the number of training samples and better anti-noise performance. The basic idea of OCSVM is to look for a decision hyperplane denoted by the support vector and maximize the distance from the hyperplane to the origin. Most of object samples locate on one side of the hyperplane and most of the no-object samples locate on the other side. The principle of OCSVM is shown in Figure 2.

Let $\mathbf{X} = [x_1; x_2; \cdots; x_n] \in \mathbf{R}^{n \times m}$ be the training data set, then $\mathbf{X}$ contains n m-dimensional feature vectors extracted from normal vibration signal samples. The decision hyperplane of OCSVM is given by:

$$F(x) = \langle \mathbf{w}, \mathbf{x} \rangle - \rho = 0 \tag{20}$$
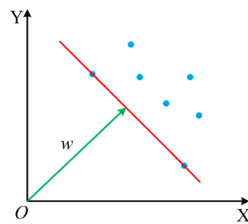
Figure 2. The principle of OCSVM.

The classification method of OCSVM can be described by the following quadratic programming problem:

$$\begin{cases} \min \dfrac{1}{2} \lVert w \rVert^2 , \\ s.t. \;\; \langle w, x_i \rangle \geqslant \rho, \;\; i = 1, \cdots, n \end{cases} \tag{21}$$

In order to improve the performance of OCSVM, the kernel theory is used for solving the linear inseparable problem. It supposes that the nonlinear mapping $\varphi : x \to \varphi(x)$ maps data from the original input space to the linear feature space. There are a slack variable ξ_i and a margin of error v in this linear feature space. ξ_i is introduced to penalize the points deviated from the hyperplane. The classifier realizes the soft interval between normal samples and fault samples with ξ_i. v is used to control the upper limit of the outliers number in the training set. Its value range is (0, 1). The expression of improved OCSVM is as follows:

$$\begin{cases} \min \dfrac{1}{2} \lVert w \rVert^2 + \dfrac{1}{vn} \sum\limits_{i=1}^{n} \xi_i - \rho, \\ s.t. \;\; \langle w, \varphi(x_i) \rangle \geqslant \rho - \xi_i, \; \xi_i \geqslant 0, \;\; i = 1, \cdots, n \end{cases} \tag{22}$$

To solve the above optimization problem, a Lagrangian function is constructed as follows:

$$(w, \xi, \rho, \alpha, \beta) = \dfrac{1}{2} \lVert w \rVert^2 + \dfrac{1}{vn} \sum\limits_{i=1}^{n} \xi_i - \rho - \sum\limits_{i=1}^{n} \alpha_i \left(\langle w, \varphi(x_i) \rangle - \rho + \xi_i \right) - \sum\limits_{i=1}^{n} \beta_i \xi_i \tag{23}$$

We can obtain the following relations by taking the partial derivatives of each variable in the Equation (23) and making them equal to zero:

$$w = \sum\limits_{i=1}^{n} \alpha_i \varphi(x_i) \tag{24}$$

$$\alpha_i = \dfrac{1}{vn} - \beta_i \leqslant \dfrac{1}{vn} \tag{25}$$

$$\sum\limits_{i=1}^{n} \alpha_i = 1 \tag{26}$$

According to the kernel function theory, the inner product of two vectors in the feature space can be represented by the kernel function in the original input space by using the nonlinear mapping, that is:

$$\langle \varphi(x_i), \varphi(x_j) \rangle = K(x_i, x_j) \tag{27}$$

Combined with Equation (27), we can get the dual form of this optimization problem:

$$\begin{cases} \min \dfrac{1}{2} \sum\limits_{i,j=1}^{n} \alpha_i \alpha_j K(x_i, x_j) \\ s.t. \sum\limits_{i} \alpha_i = 1, \;\; 0 \leqslant \alpha_i \leqslant \dfrac{1}{vn} \end{cases} \tag{28}$$

A RBF Gaussian kernel function is adopted, and its form is given as:

$$K\left(x_i, x_j\right) = \exp\left\{-\frac{||x_i - x_j||^2}{2\sigma^2}\right\} \tag{29}$$

where σ is the width parameter of RBF Gaussian kernel function.

Equation (28) describes a standard quadratic programming problem. By solving the parameters α_i and ρ, we can get the decision hyperplane in the feature space. The decision equation is written as:

$$f\left(x\right) = \sum_{i=1}^{n} \alpha_i K\left(x_i, x\right) - \rho \tag{30}$$

where the calculation formula of ρ is as follows:

$$\rho = \sum_{i=1}^{n} \alpha_i K\left(x_i, x_j\right) \tag{31}$$

4.2. Advantages of OCSVM for Condition Diagnosis

OCSVM is able to overcome the problem that fault samples are difficult to get in HVCB condition monitoring. Compared with the traditional SVM, its decision mode is more inclined to reduce the error-accept-rate, thus OCSVM is more favorable for improving equipment reliability and more suitable for mechanical condition diagnosis of HVCBs. Figure 3 compares the rationale of OCSVM and SVM, and illustrates the advantages of OCSVM in condition monitoring.
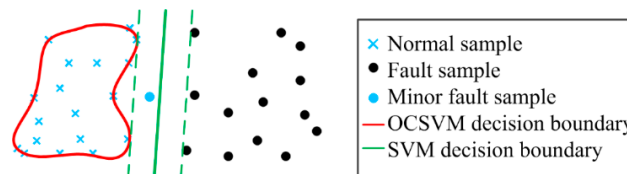

Figure 3. Comparison between the principles of OCSVM and SVM.

Figure 3 describes the results of two types of linearly separable two-class classification approaches on a 2-dimensional plane. The samples on the left side are normal samples, and the samples on the right side are fault samples. The aim of SVM is finding the lines that support the gap between the two kinds of samples, which are the two green dotted lines in Figure 3. The decision line of SVM is located in the middle of two green dotted lines. OCSVM maps the 2-dimensional input space to the high-dimensional feature space, and then seeks the decision hyperplane which supports all the target samples in this feature space. After the high-dimensional feature space is remapped to the 2-dimensional space, the hyperplane becomes a closed curve that contains all the target samples, such as the red curve in the figure.

If there is a minor fault represented by the blue dot in Figure 3, although the fault is slight, it should be identified as a fault state. According to the situation depicted in the figure, SVM identifies it as a normal condition. Conversely, because OCSVM has a more compact support region, it can correctly recognize the minor fault. The classification result of OCSVM is much more reliable than that of SVM.

4.3. An Improved PSO-Based OCSVM

The main constant parameters affecting the classification performance of OCSVM are the margin of error *v* and the width parameter of RBF kernel function σ. By adjusting the parameters *v* and σ, the distance between the hyperplane and the origin will be maximized and the classification performance of

OCSVM will be improved. In fact, these two parameters influence OCSVM's classification performance together. As the relationship between these two parameters and fitness value cannot be decided directly through the function, an intelligence optimization algorithm is used to optimize the values of v and σ. This paper adopts PSO to realize the related optimization calculation.

PSO is an intelligence algorithm for global optimization which imitates birds' flying foraging behavior [28,30]. It has a few parameters and a simple concept. The basic principle of PSO is as follows: a swarm consisting m particles is flying at a certain speed in a D-dimensional search space (in this paper $D = 2$). Each particle is considered as individually without volume. The flight speed of the particle is adjusted dynamically according to its own and its companions' flight experiences. The positions of particles are changed constantly in flight. The position and speed of the ith particle are expressed as $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \cdots, x_{iD})$ and $\boldsymbol{v}_i = (v_{i1}, v_{i2}, \cdots, v_{iD})$ respectively, where $1 \leqslant i \leqslant m$. Besides, the best position of the ith particle in history is denoted as $\boldsymbol{p}_i = (p_{i1}, p_{i2}, \cdots, p_{iD})$ and the best location that all past particles is denoted as $\boldsymbol{p}_g = (p_{g1}, p_{g2}, \cdots, p_{gD})$. For each generation, the position and speed of the d-dimension ($1 \leqslant d \leqslant D$) are changed according to the following equation:

$$v_{id}^k = wv_{id}^{k-1} + c_1 rand_1(p_{id} - x_{id}^{k-1}) + c_2 rand_2(p_{gd} - x_{id}^{k-1}) \tag{32}$$

$$v_{id}^k = \begin{cases} v_{\max}, v_{id}^k > v_{\max} \\ -v_{\max}, v_{id}^k < -v_{\max} \end{cases} \tag{33}$$

$$x_{id}^k = x_{id}^{k-1} + v_{id}^k \tag{34}$$

where w is inertia weight, c_1 and c_2 are acceleration coefficients, $rand_1$ and $rand_2$ are two uniformly distributed pseudo-random numbers in interval [0,1].

For the basic PSO algorithm, we generally define $w = 1$ and $c_1 = c_2 = 2$. The speed of a particle is limited to a maximum of $v_{\max}$. Along with Equations (32) and (34), the swarm constantly moves toward the better fitness according to the information of the particles' own experience and the shared historical information of the swarm in every iteration step.

Like other intelligent optimization algorithms, PSO also has the limitation of premature convergence that makes the algorithm get into local optima and degrades the classification performance of OCSVM. To overcome this defect and improve the convergence speed of the algorithm, this paper proposes an improved PSO algorithm with a linearly varying inertia weight and acceleration coefficients.

(1)   Adjustment of the inertia weight ω

By decreasing the inertia weight in a linear way, the algorithm can search for better solutions from the global scope. It will have a better local search capability with the increase of the number of iterations. The algorithm not only maintains a good search ability but also avoids the premature convergence phenomenon. Let $[\omega_{\min}, \omega_{\max}]$ be the value range of inertia weights, generally $\omega_{\min} = 0.4$ and $\omega_{\max} = 0.9$. Let *Iter_ max* be the maximum number of iterations. Then the inertia weight of the ith iteration is given as:

$$\omega_i = \omega_{\max} - \frac{\omega_{\max} - \omega_{\min}}{Iter\_max} \times i \tag{35}$$

(2)   Adjustment of the acceleration coefficients c_1 and c_2

Acceleration coefficients reflect the degree of information exchange between particle swarms. On the other hand, the in-flight behavior of a particle depends on its own experience with a larger c_1. This results in that the particles easily wander in their own local scope. On the other hand, particles will have a higher speed moving toward the optimal individual with a larger c_2, but this may cause a premature convergence to a local optimum. In order to solve this contradiction, researchers usually assign the same constant value to c_1 and c_2, but sometimes this method can't meet the needs of the actual situation.

We use an appropriate method which chooses a larger c_1 and a smaller c_2 at the beginning of the algorithm, and then gradually decreases c_1 and increases c_2. By this adjustment, particles tend to fly in the entire search space in the early stages, so the region that contains the optimal solution does not get lost. Particles finally tend to fly to the globally optimal solution. Compared to the traditional method, particles learn more from the particles which have reached the historical optimal solution. In this paper, the adjustment measure of acceleration coefficients is defined as follows:

$$c_1 = (c_{1f} - c_{1i})\frac{i}{Iter\_max} + c_{1i} \tag{36}$$

$$c_2 = (c_{2f} - c_{2i})\frac{i}{Iter\_max} + c_{2i} \tag{37}$$

where c_{1i} and c_{1f} are the initial value and final value of c_1, c_{2i} and c_{2f} are the initial value and final value of c_2. There is a symmetry variation in this paper, namely c_1 linearly decreases from 2.5 to 0.5 and c_2 linearly decreases from 0.5 to 2.5.

4.4. Fault Diagnostic Process

The flow chart of the diagnosis method is shown in Figure 4. In practical engineering applications, the vibration signal that has been diagnosed as a normal signal can be added to the normal sample set, so the fault diagnosis system can constantly adapt to the change of running conditions of a circuit breaker, and its learning ability can be improved.
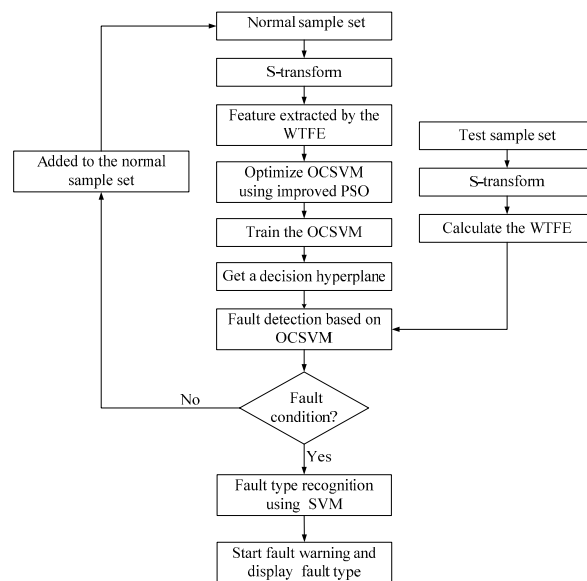

Figure 4. The flow of the fault diagnosis process.

5. Experimental Results and Analysis

5.1. Data Collection and Processing

The experiment adopts LW9-72.5 series outdoor high voltage SF_6 circuit breakers as the analysis object. The vibration signal acquisition system is built with a CA-YD-182A piezoelectric acceleration sensor made in Jiangsu United Electronic Technology Co., Ltd. (Yangzhou, China) and NI-9234 DAQ devices made by National Instruments (NI, Austin, TX, USA). The acceleration sensor is used for vibration signal acquisition. The DAQ device is used to record the data with 25.6 kS/s sampling rate for a time period of 150 ms during opening operation. The vibration signal acquisition system for a circuit breaker is shown in Figure 5.
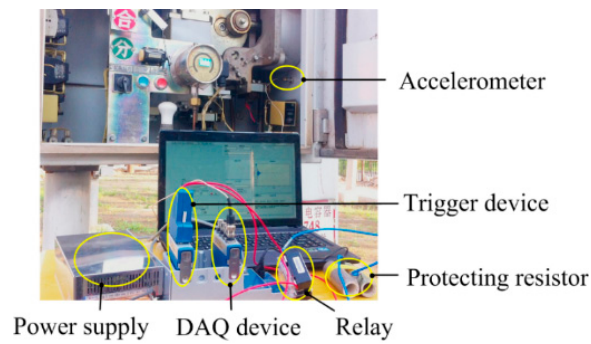
Figure 5. The vibration signal acquisition system for a circuit breaker.

Three kinds of fault types are simulated in the experiment: (1) z jam fault of the iron core (fault type I); (2) base screw looseness (fault type II); (3) lack of mechanical lubrication (fault type III). To avoid the HVCB damage from excessive operations, 40 vibration signals of the health and 40 vibration signals per fault type were collected.
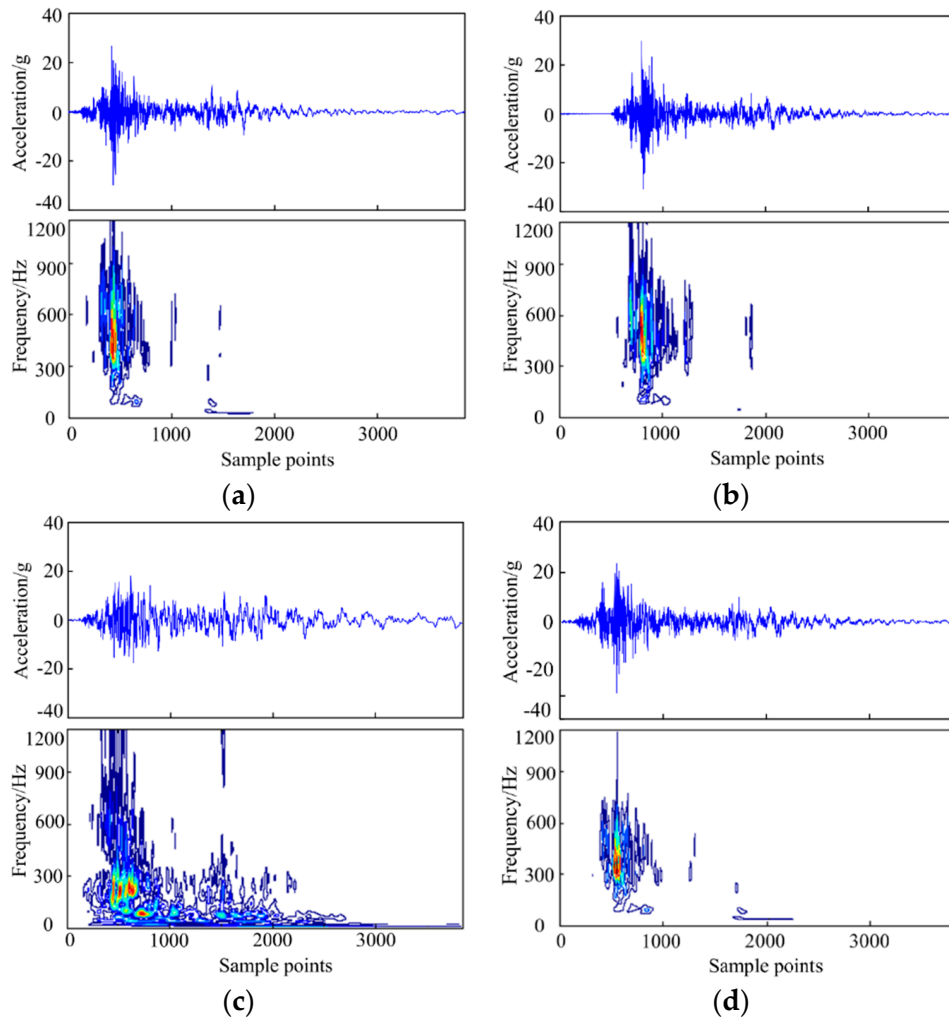

Figure 6. (**a**) The normal signal and its STMM contour plot; (**b**) The signal of fault type I and its STMM contour plot; (**c**) The signal of fault type II and its STMM contour plot; (**d**) The signal of fault type III and its STMM contour plot.

According to the above method, vibration signals are processed by ST. Figure 6 shows the vibration signals and their contour plot after ST analysis under different conditions, including the healthy ones and three types of faults. From Figure 6, we can find that the time-frequency energy distributions of the different types of vibration signals have obvious differences. Compared with the normal signal, the energy distribution of the iron core jam fault signal has an apparent time delay; the base screw looseness fault signal has a strong energy distribution in a lower frequency area; and the energy center of the third fault signal has slightly changed in both the time and frequency domains. The time and frequency characteristics can thus be extracted to analyze the operating condition of a HVCB's mechanical operation system.

5.2. Feature Extraction and Analysis

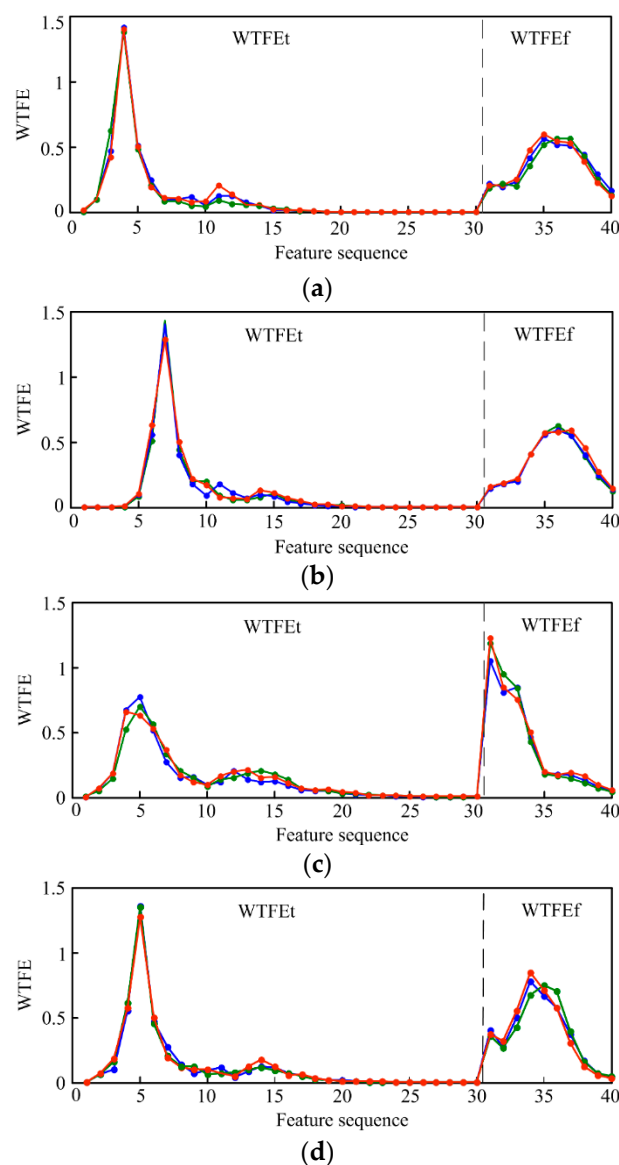We can get the WTFE feature vectors according to the feature vector extraction method mentioned before.

Figure 7. (**a**) WTFE feature distribution of the normal signals; (**b**) WTFE feature distribution of the iron core jam fault signals; (**c**) WTFE feature distribution of the base screw looseness fault signals; (**d**) WTFE feature distribution of the lack of mechanical lubrication fault signals.

The WTFE feature distributions of four kinds of vibration signals are shown in Figure 7, where the first 30 features reflect the energy distribution of the signal in the time domain (WTFEt) and the other 10 features reflect the energy distribution in the frequency domain (WTFEf). For clarity, each type only shows three data points. Figure 7 shows that different types of vibration signals have significantly different feature distributions. According to these differences, the classifier can achieves a good classification effect. In order to prove the diagonosis ability of different feature presentation methods, we present a comparison between the WTFE and WSE. Firstly, the STMM is divided into 50 submatrixes along the time axis. Then the WSE of each of submatrix is calculated based on Equations (14)–(16) to form 50 dimensional input feature vectors. The WSE feature distributions of four kinds of vibration signals are shown in Figure 8.
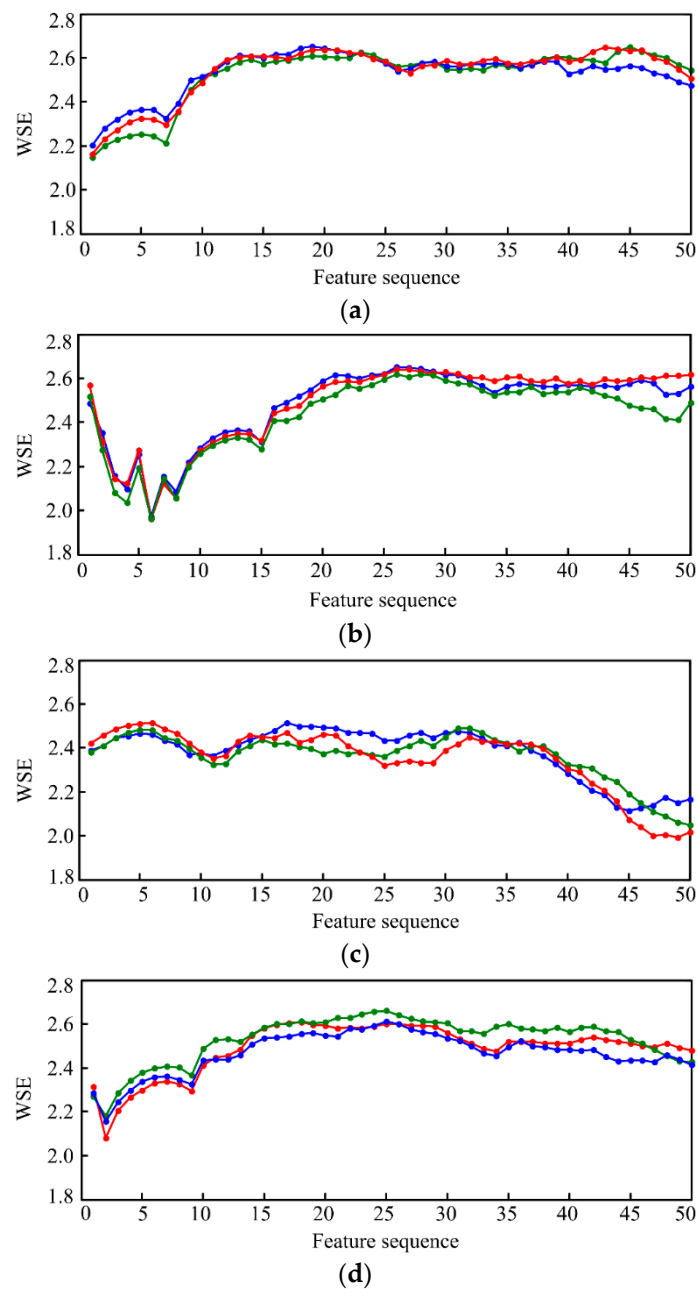

Figure 8. (**a**) WSE feature distribution of the normal signals; (**b**) WSE feature distribution of the iron core jam fault signals; (**c**) WSE feature distribution of the base screw looseness fault signals; (**d**) WSE feature distribution of the lack of mechanical lubrication fault signals.

From Figure 8, the WSE feature distributions of different kinds of vibration signals show different characteristics. However, we can reveal some disadvantages of the WSE method by comparing Figures 7 and 8. First, the WSE method can't clearly and visually display the real change rules of vibration signals. Second, the WSE feature vectors of the same types of signals are more dispersed than the WTFE ones. Third, the difference between the high value and the low value in the WTFE feature vector is too small. These latter two characteristics of WSE method will degrade the performance of a classifier.

5.3. Classification Using OCSVM-SVM

We select the improved PSO to optimize the parameters of OCSVM in two-dimensional space. According to the abovementioned adjustment method of the inertia weight and accelerated coefficient, a program is written to realize the parameter optimization. The number of swarms is 30 and the number of iterations is 50. After running PSO, we obtain the optimal solution $v = 0.82$, $\sigma = 17.68$. Figure 9 shows the relationship between fitness and iterations. From Figure 9, we can find that the globally optimal solution has appeared in the ninth iteration, and in later iterations swarms just to get close to the particle which has the optimal fitness. Therefore, the average fitness increases gradually.
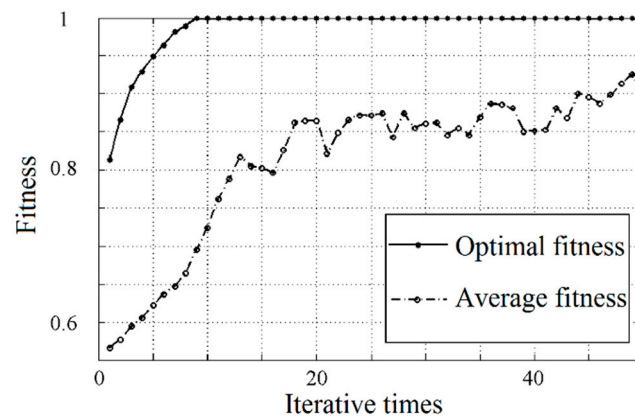

Figure 9. Fitness curve of PSO.

We select the WTFE as the input feature vector of the classifier. The classifier needs to be trained using the training samples before use. For each type of signal, there are 40 vibration samples. The 40 samples of each type are divided randomly into two groups. One is the training sample set, and the other is the test sample set. Each set contains 20 samples. The training sample set of the normal state is used to train the OCSVM. The training sample sets of each type of faults are used to train the SVM. A total of 80 samples of four test sample sets are used to test the classification effect. The test results are shown in Table 1, where, the state determination accuracy (STA) reflects the ability of the classifier to determine whether the circuit breaker is healthy or not and the classification accuracy (CA) reflects the ability of the classifier to identify the specific type of a sample.

Table 1. Diagnosis results using the new approach.

Test Sample	Diagnosis Results				STA	CA
	Normal State	Fault Type I	Fault Type II	Fault Type III		
Normal state	18	0	0	2	90%	90%
Fault type I	0	20	0	0	100%	100%
Fault type II	0	0	20	0	100%	100%
Fault type III	0	0	0	20	100%	100%

From the diagnosis results, the OCSVM-SVM failed to recognize the normal sample completely, but it can accurately classify all the fault samples into the correct fault type. In fact, for the fault diagnosis of circuit breakers, the risk of recognizing fault samples as normal samples is much higher than that of recognizing normal samples as fault samples, therefore this approach still achieves a good diagnosis effect. When the WSE is selected as the input feature vector, we can get a comparative results shown in Table 2.

Table 2. Diagnosis results with the feature vector of WSE.

Test Sample	Diagnosis Results				STA	CA
	Normal State	Fault Type I	Fault Type II	Fault Type III		
Normal state	14	0	1	5	70%	70%
Fault type I	0	20	0	0	100%	100%
Fault type II	0	0	18	2	100%	90%
Fault type III	2	0	1	17	90%	85%

From Tables 1 and 2 we can find that the diagnosis results with the WSE method are greatly inferior to those of the new approach, especially for the classification of normal state conditions. Thus the WSE method is not suitable for the feature extraction of vibration signals of HVCBs.

To explain the merit of OCSVM-SVM against other popular used classifiers, comparison experiments between SVM, ELM based classifier and the new approach are designed. The training method and test samples of SVM and ELM are the same as in the new approach. The experimental results are shown in Table 3.

Table 3. Diagnosis results by using the SVM and ELM methods.

Classifier	Test Sample	Diagnosis Results				STA	CA
		Normal State	Fault Type I	Fault Type II	Fault Type III		
SVM	Normal state	17	0	0	3	85%	85%
	Fault type I	0	20	0	0	100%	100%
	Fault type II	0	0	20	0	100%	100%
	Fault type III	2	0	0	18	90%	90%
ELM	Normal state	18	0	0	2	90%	90%
	Fault type I	0	20	0	0	100%	100%
	Fault type II	0	0	0	20	100%	100%
	Fault type III	3	0	0	17	85%	85%

Table 3 shows that the SVM and ELM methods have about the same classification ability. They all correctly identify all samples of type I and II faults. For the type III faults, the two classifier fail to identify all samples. The AC of SVM is 90% and that of ELM is 85%. Since the OCSVM-SVM can identify all samples of the fault type I, II and III, the fault recognition ability of the new approach is better than that of SVM and ELM.

In a real power system environment, there may some types of new faults that we have not recorded before. Once this happens, the multiple fault classifiers cannot identify this fault type because a lack of training samples. Therefore, it is very important that the classifier can determine it as a fault state accurately. Considering this case, we compare the STA of OCSVM-SVM, SVM and ELM. Suppose the fault type III is the unknown fault, then no samples of fault type III are involved in the training of the three types of classifiers. Twenty sets of type III fault vibration data are selected as the new test samples. The diagnosis results are shown in Table 4.

In Table 4, neither of the two methods can correctly identify the specific fault type without training samples, but the STA of OCSVM-SVM is 100% while that of SVM and ELM is 0. That means OCSVM-SVM can correctly determine the state of the fault whose type has not been recorded before. Therefore, we can conclude that the OCSVM-SVM method has better fault detection capability and it is more suitable for circuit breaker fault diagnosis which requires a higher reliability.

Table 4. Diagnosis results of the case of lack of training samples by using the OCSVM-SVM, SVM and ELM methods.

Classifier	Test Sample	Diagnosis Results			STA	CA
		Normal State	Fault Type I	Fault Type II		
OCSVM-SVM	Fault type III	0	14	6	100%	0
SVM	Fault type III	20	0	0	0	0
ELM	Fault type III	20	0	0	0	0

6. Conclusions

This paper presents a new method based on WTFE and improved OCSVM for mechanical fault diagnosis of HVCBs. ST is employed to process and analyze vibration signals. The WTFE is selected as the vibration signal feature. It characterizes the signal in the time domain and frequency domain. A new classifier based on OCSVM-SVM is built to improve the classification performance of the diagnosis system. An optimal PSO algorithm is used to optimize the OCSVM parameters. Experimental results show that the new approach has higher STA and CA than the traditional SVM and ELM methods, and the accuracy for some faults increases by more than 10%. Especially in the mechanical fault condition analysis of fault types without training samples, the new method shows a conspicuous advantage, therefore, the new method can significantly increase power system security and reliability.

Author Contributions: Shuxin Zhang designed the research method. Nantian Huang wrote the draft. Huaijin Chen contributed to the experimental section. Weiguo Li and Lihua Fang gave a detailed revision. Guowei Cai and Dianguo Xu provided important guidance. All authors have read and approved the final manuscript.

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

References

1. Runde, M.; Aurud, T.; Lundgaard, L.E.; Ottesen, G.E.; Faugstad, K. Acoustic diagnosis of high voltage circuit-breakers. *IEEE Trans. Power Deliv.* **1992**, *7*, 1306–1315. [CrossRef]
2. Demjanenko, V.; Valtin, R.A.; Soumekh, M.; Haidu, H.; Antur, A.; Hess, D.P.; Soom, A.; Wright, S.E.; Tangri, M.K.; Park, S.Y. A noninvasive diagnostic instrument for power circuit breakers. *IEEE Trans. Power Deliv.* **1992**, *7*, 656–663. [CrossRef]
3. Polycarpou, A.A.; Soom, A.; Swarnakar, V.; Valtin, R.A.; Acharya, R.S.; Demjanenko, V.; Soumekh, M.; Benenson, D.M.; Porter, J.W. Event timing and shape analysis of vibration bursts from power circuit breakers. *IEEE Trans. Power Deliv.* **1995**, *11*, 848–857. [CrossRef]
4. 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]
5. CIGRE Working Group. *Final Report of the Second International Enquiry on High Voltage Circuit Breaker Failures and Defects in Service*; CIGRE Report No. 83; CIGRE: Paris, France, 1994.
6. Gao, Z.W.; Cecati, S.; Ding, S.X. A survey of fault diagnosis and fault-tolerant techniques-Part I: Fault diagnosis with model-based and signal-based approaches. *IEEE Trans. Ind. Electron.* **2015**, *62*, 3757–3767. [CrossRef]
7. Meng, Y.P.; Jia, S.L.; Shi, Z.Q.; Rong, M.Z. The detection of the closing moments of a vacuum circuit breaker by vibration analysis. *IEEE Trans. Power Deliv.* **2006**, *21*, 652–658. [CrossRef]
8. Landry, M.; Leonard, F.; Landry, C.; Beauchemin, R. 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]
9. Lee, D.S.; Lithgow, B.J.; Morrison, R.E. New fault diagnosis of circuit breakers. *IEEE Trans. Power Deliv.* **2003**, *18*, 454–459. [CrossRef]

10. Huang, J.; Hu, X.G.; Yang, F. Support vector machine with genetic algorithm for machinery fault diagnosis of high voltage circuit breaker. *Measurement* **2011**, *44*, 1018–1027. [CrossRef]

11. Huang, J.; Hu, X.G.; 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]

12. Høidalen, H.K.; Runde, M. Continuous monitoring of circuit-breakers using vibration analysis. *IEEE Trans. Power Deliv.* **2005**, *20*, 2458–2465. [CrossRef]

13. Stockwell, R.G.; Mansinha, L.; Lowe, R.P. Localization of the complex spectrum: The S transform. *IEEE Trans. Signal Process.* **1996**, *44*, 998–1001. [CrossRef]

14. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]

15. He, Z.Y.; Chen, X.Q.; Luo, G.M. Wavelet entropy measure definition and its application for transmission line fault detection and identification. In Proceedings of the International Conference on Power System Technology, Chongqing, China, 22–26 October 2006.

16. Chen, J.K.; Li, G.Q. Tsallis wavelet entropy and its application in power signal analysis. *Entropy* **2014**, *16*, 3009–3025. [CrossRef]

17. He, Z.Y.; Gao, S.B.; Chen, X.Q.; Zhang, J.; Bo, Z.Q.; Qian, Q.Q. Study of a new method for power system transients classification based on wavelet entropy and neural network. *Int. J. Electr. Power Energy Syst.* **2011**, *33*, 402–410.

18. Kumar, Y.; Dewal, M.L.; Anand, R.S. Epileptic seizure detection using DWT based fuzzy approximate entropy and support vector machine. *Neurocomputing* **2014**, *133*, 271–279. [CrossRef]

19. Sharma, R.; Pachori, R.B.; Acharya, U.R. An integrated index for the identification of focal electroencephalogram signals using discrete wavelet transform and entropy measures. *Entropy* **2015**, *17*, 5218–5240. [CrossRef]

20. Boškoski, P.; Juričić, Đ. Fault detection of mechanical drives under variable operating conditions based on wavelet packet Rényi entropy signatures. *Mech. Syst. Signal Process.* **2012**, *31*, 369–381. [CrossRef]

21. Dubey, R.; Samantaray, S.R. Wavelet singular entropy-based symmetrical fault-detection and out-of-step protection during power swing. *IET Gener. Transm. Distrib.* **2013**, *7*, 1123–1134. [CrossRef]

22. Wong, P.K.; Yang, Z.X.; Vong, C.M.; Zhong, J.H. Real-time fault diagnosis for gas turbine generator systems using extreme learning machine. *Neurocomputing* **2014**, *128*, 249–257. [CrossRef]

23. Yang, Z.X.; Wong, P.K.; Vong, C.M.; Zhong, J.H.; Liang, J.J. Simultaneous-fault diagnosis of gas turbine generator systems using a pairwise-coupled probabilistic classifier. *Math. Probl. Eng.* **2013**, *2013*. [CrossRef]

24. Schölkopf, B.; Platt, J.C.; Shawe-Taylor, J.; Smola, A.J.; Williamson, R.C. Estimating the support of a high-dimensional distribution. *Neural Comput.* **2001**, *13*, 1443–1471. [CrossRef] [PubMed]

25. Mahadevan, S.; Shah, S.L. Fault detection and diagnosis in process data using one-class support vector machines. *J. Process Control* **2009**, *19*, 1627–1639. [CrossRef]

26. Shin, H.J.; Eom, D.H.; Kim, S.S. One-class support vector machines—An application in machine fault detection and classification. *Comput. Ind. Eng.* **2005**, *48*, 395–408. [CrossRef]

27. Xiao, Y.C.; Wang, H.G.; Xu, W.L. Parameter selection of Gaussian kernel for one-class SVM. *IEEE Trans. Cybern.* **2015**, *45*, 927–939. [PubMed]

28. Kennedy, J.; Eberhart, R.C. Particle swarm optimization. In Proceedings of the IEEE International Conference on Neural Networks, Piscataway, NJ, USA, 27 November–1 December 1995; pp. 1942–1948.

29. Mishra, S.; Bhende, C.N.; Panigrahi, K.B. Detection and classification of power quality disturbances using S-transform and probabilistic neural network. *IEEE Trans. Power Deliv.* **2008**, *23*, 280–287. [CrossRef]

30. Olsson, A.E. *Particle Swarm Optimization: Theory, Techniques and Applications*; Nova Science Publishers: Hauppauge, NY, USA, 2011.