Fault Diagnosis for Rolling Bearings Based on Fine-Sorted Dispersion Entropy and SVM Optimized with Mutation SCA-PSO

Wenlong Fu 1,2,*, Jiawen Tan 1,2, Yanhe Xu 3, Kai Wang 1,2 and Tie Chen 1,2

1 College of Electrical Engineering & New Energy, China Three Gorges University, Yichang 443002, China; 201708520721060@ctgu.edu.cn (J.T.); 201808520721083@ctgu.edu.cn (K.W.); chent@ctgu.edu.cn (T.C.)
2 Hubei Provincial Key Laboratory for Operation and Control of Cascaded Hydropower Station, China Three Gorges University, Yichang 443002, China
3 School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China; yh_xu@hust.edu.cn
* Correspondence: fuwenlong@ctgu.edu.cn; Tel.: +86-136-1868-0447

Received: 22 March 2019; Accepted: 14 April 2019; Published: 16 April 2019

Abstract: Rolling bearings are a vital and widely used component in modern industry, relating to the production efficiency and remaining life of a device. An effective and robust fault diagnosis method for rolling bearings can reduce the downtime caused by unexpected failures. Thus, a novel fault diagnosis method for rolling bearings by fine-sorted dispersion entropy and mutation sine cosine algorithm and particle swarm optimization (SCA-PSO) optimized support vector machine (SVM) is presented to diagnose a fault of various sizes, locations and motor loads. Vibration signals collected from different types of faults are firstly decomposed by variational mode decomposition (VMD) into sets of intrinsic mode functions (IMFs), where the decomposing mode number $K$ is determined by the central frequency observation method, thus, to weaken the non-stationarity of original signals. Later, the improved fine-sorted dispersion entropy (FSDE) is proposed to enhance the perception for relationship information between neighboring elements and then employed to construct the feature vectors of different fault samples. Afterward, a hybrid optimization strategy combining advantages of mutation operator, sine cosine algorithm and particle swarm optimization (MSCAPSO) is proposed to optimize the SVM model. The optimal SVM model is subsequently applied to realize the pattern recognition for different fault samples. The superiority of the proposed method is assessed through multiple contrastive experiments. Result analysis indicates that the proposed method achieves better precision and stability over some relevant methods, whereupon it is promising in the field of fault diagnosis for rolling bearings.

Keywords: rolling bearings; fault diagnosis; variational mode decomposition; fine-sorted dispersion entropy; mutation sine cosine algorithm-particle swarm optimization; support vector machine

1. Introduction

Rolling bearings are a crucial part in modern industrial manufacture, which can be found in linear guides, precision machine tools, and engine parts, etc., whose failure may result in serious safety accidents and economic loss. Therefore, one of the major topics to be investigated in preventing failures of mechanical systems is recognizing and diagnosing rolling bearing faults [1,2]. Unfortunately, rolling bearings are prone to failure to some extent because of complex operating conditions and structural characteristics [3]. Given this fact, effective and feasible fault diagnosis methods need to be developed for rolling bearings, thus, to improve the reliability of mechanical systems [4].

Entropy 2019, 21, 404; doi:10.3390/e21040404 www.mdpi.com/journal/entropy
Due to the rich information carried by vibration signal, most of fault diagnosis methods for rolling bearings rely on analyzing vibration signal [5]. However, the vibration signal is commonly non-stationary in actual operation, which restricts the extraction efficiency for fault features and affects the accuracy of fault diagnosis. Addressing this issue, many time-frequency signal analysis methods have been proposed to analyze non-stationary signals, including wavelet transform (WT), empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD), and variational mode decomposition (VMD). WT is a self-adaptive signal decomposition method, which has a good frequency resolution and time resolution for low-frequency components and provides a good frequency resolution and time resolution for high-frequency components [6], but the selection of an appropriate wavelet filter bank remains a challenge. EMD is adept in dealing with non-stationary signals and does not need to pre-set any basis function [7], but its performance is affected by end effects and mode mixing. To overcome these defects, as an improved version of EMD, EEMD was proposed by introducing noise-assisted analysis method [8], but it also increases the computational cost and cannot completely neutralize the added noise. Unlike the methods mentioned above, VMD is a quasi-orthogonal signal decomposition method, the main structure of which is to construct and solve a constrained variational problem [9]. Hence it realizes the separation of the frequency of each signal component, avoiding the problem of mode mixing in EMD and EEMD or filter selection in WT. Its advancement and effectiveness have been demonstrated in many previous studies. For instance, Fu et al. [10] applied VMD to decompose the wind speed time series into a set of sub-series with various frequency scales and then calculated the final predicted value of the raw wind speed by accumulating the predicted results of all components. Li et al. [11] established the approximate entropy feature vectors for each component of VMD and obtained a good fault diagnosis effect for rolling bearings. Fu et al. [12] achieved the feature extraction of signals by conducting multivariate autoregressive modelling for components of VMD.

As the non-stationarity of signals weakened by signal analysis methods, fault features can be smoothly extracted. Entropy, as a feature extraction method based on non-linear dynamic parameters, is widely applied in medical and mechanical fields, which can reflect the complexity of time series and is an effective tool for analysis of non-stationary time series. Approximate entropy (ApEn) [13] is the pioneer of introducing a feature extraction algorithm from the medical field into mechanical fault diagnosis, while it is susceptible to outliers. Permutation entropy (PE) [14] has strong anti-noise ability and is simple in calculation, but it does not take into account the difference between the average amplitude and actual amplitude. Fuzzy entropy (FE) [15] introduces the concept of uncertainty reasoning to solve the defect of sample entropy [16] which adopts a hard threshold as a discriminant criterion which may result in unstable results. This approach, however, also lowers the computational efficiency. Recently, a novel irregularity indicator termed dispersion entropy (DE), was proposed by Mostafa and Hamed [17]. DE has the advantages of fast calculation speed and less influence from outliers, which overcomes the defects of ApEn and PE to some extent, but the relationship information between neighboring amplitudes is not considered enough. In this paper, a concept of fine-sorted dispersion entropy (FSDE) is proposed to measure the features of vibration data and then construct fault feature vectors. By introducing an additional factor, FSDE is more perceptive to element values and may have better feature extraction ability than DE.

The essence of fault diagnosis for rolling bearings is actually a pattern recognition problem, that is, fault type is determined by the fault features of samples. In an engineering application, there are many alternative methods available for solving this problem. For instance, Bayesian decision has strong recognition ability by considering prior probability as well as class conditional probability, and its good accuracy requires the assumption of a suitable prior model [18]. K-nearest neighbor (KNN) based on Euclidean or Manhattan distances is easy to realize and susceptible to the distribution of samples [19]. Artificial neural network (ANN) has good ability in processing recognition problem with the support of a large number of samples [20]. Based on statistical learning and structural risk minimization theory, support vector machine (SVM) is an excellent machine learning method, which finds an optimal hyperplane that meets the classification requirements. The hyperplane maximizes the interval between
two classes while promoting classification performance. Currently, SVM combined with the feature extraction method has been successfully employed in the field of pattern recognition [21,22].

Although SVM possesses superior ability in pattern recognition, its performance is affected by parameters. In view of this, various optimization algorithms have been developed and applied to search the best parameters, such as particle swarm optimization (PSO) [23], bacterial foraging algorithm (BFA) [24], artificial sheep algorithm (ASA) [25] and sine cosine algorithm (SCA) [26]. As a novel optimization approach, the effectiveness of SCA in parameter optimization has been proved in many previous studies [27,28]. To achieve better performance in convergence precision, a hybrid optimization strategy combining mutation operator, sine cosine algorithm, and particle swarm optimization (MSCAPSO) is developed in this paper.

VMD was firstly applied to decompose the vibration signal into a set of intrinsic mode functions (IMFs), and the decomposing mode number $K$ was pre-set with central frequency observation method. Then a modified dispersion entropy containing the additional factor (FSDE) is proposed to construct the feature vectors of IMFs for different types of faults. Afterward, a hybrid optimization strategy combining mutation operator, SCA and PSO called MSCAPSO was proposed for parameter selection of SVM. Finally, the optimal SVM model was employed to classify different fault samples. The engineering application and contrastive analysis reveal the availability and superiority of the proposed method.

The paper is organized as follows: Section 2 is dedicated to the basic knowledge of VMD and SVM. Section 3 introduces the proposed fault diagnosis method based on fine-sorted dispersion entropy and MSCA-PSO optimization approach. Section 4 presents the engineering application, the result analysis of which demonstrates the superiority of the proposed method. Some discussion about the method presented in this paper is in Section 5. The conclusion is summarized in Section 6.

2. Fundamental Theories

2.1. Variational Mode Decomposition

Variational mode decomposition (VMD) is a new adaptive signal preprocessing method [29]. By setting a mode number $K$ in advance, the given signal can be decomposed into $K$ band-limited intrinsic mode functions (IMFs). The cyclic iterative method is used to obtain the optimal solution of the constrained variational problem, through which the frequency center and bandwidth of each mode are determined. The constrained variation problem can be described as follows:

$$
\min_{m_k, \omega_k} \left\{ \sum_k \| \partial_t \left[ (\delta(t) + \frac{j \pi}{m}) \ast m_k(t) \right] e^{-j \omega_k t} \|_2^2 \right\}
\quad \text{s.t.} \quad \sum_k m_k(t) = f(t), \quad k = 1, 2, \ldots, K
$$

(1)

where $m_k = [m_1, m_2, \ldots, m_k]$ represents the set of $K$ mode functions and $\omega_k = [\omega_1, \omega_2, \ldots, \omega_k]$ represents the set of central frequencies, while $\partial_t$ and $\delta_t$ are the partial derivative of time $t$ for the function and unit pulse function, respectively. $f(t)$ is the given real valued input signal.

For the above variational problem, quadratic penalty function term and Lagrange multiplier are employed to transform it into an unconstrained problem. Then Problem (1) can be specified as follows:

$$
L(m_k, \omega_k, \beta) = \alpha \sum_k \| \partial_t \left[ (\delta(t) + \frac{j \pi}{m}) \ast m_k(t) \right] e^{-j \omega_k t} \|_2^2 + \| f(t) - \sum_k m_k(t) \|^2_2 + \beta(t) \left( f(t) - \sum_k m_k(t) \right)
$$

(2)

where $\alpha$ and $\beta(t)$ represent the penalty factor and Lagrange multiplier, respectively.
Then alternate direction method is utilized to obtain the saddle point of Lagrange multiplier, that is, solving (2) by optimizing \( m_k, \omega_k \), and \( \beta \) alternately. The optimization problems of \( m_k \) and \( \omega_k \) are formulated as (3) and (4), respectively.

\[
m_k^{n+1} = \min \left\{ \alpha \| \partial \left[ \left( \delta(t) + \frac{j}{\pi t} \right) m_k(t) \right] e^{-j\omega t} \|^2_2 + \| f(t) - \sum_i m_i(t) + \frac{\beta(t)}{2} \|^2_2 \right\} \tag{3}
\]

\[
\omega_k^{n+1} = \min \left\{ \| \partial \left[ \left( \delta(t) + \frac{j}{\pi t} \right) m_k(t) \right] e^{-j\omega t} \|^2_2 \right\} \tag{4}
\]

Solving problems (3) and (4), the iterative equations are deduced as follows:

\[
\hat{m}_k^{n+1}(\omega) = \frac{f(\omega) - \sum_{i \neq k} \hat{m}_i(\omega)}{1 + 2\alpha(\omega - \omega_k)^2} \tag{5}
\]

\[
\alpha_k^{n+1} = \frac{\int_0^\infty \omega \| \hat{m}_k(\omega) \|^2 d\omega}{\int_0^\infty \| \hat{m}_k(\omega) \|^2 d\omega} \tag{6}
\]

The Lagrange multipliers can be iterated with Equation (7).

\[
\beta_k^{n+1}(\omega) = \beta_k^n(\omega) + \gamma(f(\omega) - \sum_k \hat{m}_k^{n+1}(\omega)) \tag{7}
\]

where \( \gamma \) is an updating parameter.

The main steps of VMD can be summarized as follows:

Step 1: Initialize \( m_k^0, \omega_k^0, \beta^0, n = 1 \);
Step 2: Execute loop, \( n = n + 1 \);
Step 3: Update \( m_k \) and \( \omega_k \) based on Equations (5) and (6);
Step 4: Update \( \beta \) based on Equation (7);
Step 5: If \( \sum_k \| \hat{m}_k^{n+1} - \hat{m}_k^n \|^2_2 / \| \hat{m}_k^n \|^2_2 < \epsilon \) loop end, else turn to step 2 for next iteration.

2.2. Support Vector Machine

Support vector machine (SVM) is a machine learning model developed by Vapnik [30], which can deduce the optimal solution between model complexity and learning ability based on limited information, thus, to obtain the best classification accuracy, showing unique advantages in solving learning problems with limited samples, non-linear, and high dimensional data. For a given sample set \( \{(x_i, y_i) | i = 1, 2, \ldots, n\} \), the main idea of SVM is to map samples into a high-dimensional feature space and find a hyper-plane to solve the corresponding non-linear problem in low-dimensional space. The hyper-plane function can be constructed as:

\[
w \cdot x + b = 0 \tag{8}
\]

where \( w \) is a weight vector, and \( b \) is a bias parameter, while \( w \cdot x \) represents the inner product of \( w \) and \( x \).

Taking binary classification issue as an example, to classify samples correctly with enough into classification intervals, all samples are required to meet the following constraint:

\[
w \cdot x_i + b \begin{cases} > 1 & \text{for } y_i = 1 \\ < -1 & \text{for } y_i = -1 \end{cases} \tag{9}
\]

In the sequel, the classification interval can be calculated as \( 2/\| w \|^2 \), and maximizing the interval is equivalent to minimizing \( \| w \|^2 \). To solve the linear indivisibility problem, the slack term \( \xi \) and penalty
factor $C$ are brought into Equation (9), hence the construction of the optimal hyper-plane is transformed into a constrained optimization problem:

$$\begin{align*}
\min f &= \frac{1}{2}||w||^2 + C \sum_{i=1}^{n} \xi_i \\
\text{s.t. } y_i(w^T \cdot x_i + b) &\geq 1 - \xi_i, \quad i = 1, 2, \ldots, n
\end{align*}$$ (10)

When mapping samples to high-dimensional space, different inner product kernel functions will form different algorithms. Among them, radial basis function is widely employed in the application of pattern recognition, which can be described as:

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) = \exp\left(-g||x_i - x_j||^2\right)$$ (11)

where $g$ is a kernel parameter.

The solution of the constrained optimization problem (10) is determined by a saddle point of Lagrange function, and this quadratic programming problem is transformed into the corresponding dual problem, that is:

$$\begin{align*}
\max L &= \sum_{i=1}^{n} \mu_i - \frac{1}{2} \sum_{i,j=1}^{n} \mu_i \mu_j y_i y_j K(x_i, x_j) \\
\text{s.t. } \sum_{i=1}^{n} \mu_i y_i &= 0, \quad \mu_i \geq 0, \quad i = 1, 2, \ldots, n
\end{align*}$$ (12)

where $\mu_i$ is Lagrange multiplier.

With the optimal value of $\mu_i$ obtained from the above dual problem, the optimal classification discriminant function can be ascertained as:

$$f(x) = \text{sgn}\left(\sum_{i=1}^{n} \mu_i K(x_i, x) + b\right)$$ (13)

3. Fault Diagnosis for Rolling Bearings by Fine-sorted Dispersion Entropy and Mutation SCA-PSO Optimized SVM

3.1. Fine-Sorted Dispersion Entropy

3.1.1. Dispersion Entropy

For a time series $x = [x_1, x_2, \ldots, x_n]$ of length $n$, $x_i$ is firstly normalized by a mapping function, where the standard normal cumulative distribution function is generally used [17].

$$y_i = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{x_i} e^{-\frac{(s-\mu)^2}{2\sigma^2}} ds$$ (14)

where $i = 1, 2, \ldots, n$, $\sigma$, and $\mu$ are the variance and expectation of the normal distribution, respectively. The time series $x$ is then normalized to $y = [y_1, y_2, \ldots, y_n]$, $y_i \in (0, 1)$. Further, the phase space is reconstructed for $y$:

$$y^m_j = [y_j, y_{j+\tau}, \ldots, y_{j+(m-1)\tau}]$$ (15)

where $j = 1, 2, \ldots, n - (m-1)\tau$, $m$ and $\tau$ are the embedding dimension and time delay, respectively. Then $y^m_j$ is mapped to $c$ class (1 to $c$):

$$z^c_i = \text{round}(c \cdot y_i + 0.5)$$ (16)

$$z^{mc}_{ij} = [z^c_j, z^c_{j+\tau}, \ldots, z^c_{j+(m-1)\tau}]$$ (17)
where $z_i^c$ is the $i$-th element of class sequence $z^c$, while round means rounding. Each $z_i^m$ corresponds to a dispersion pattern $\pi_{v_0 v_1 \ldots v_{m-1}}$ with elements $z_i = v_0$, $z_i = v_1$, \ldots, $z_i = v_{m-1}$. Since $\pi_{v_0 v_1 \ldots v_{m-1}}$ contains $m$ elements, each element contains $c$ values, so the number of all possible dispersion patterns are $c^m$.

The relative frequency of $\pi_{v_0 v_1 \ldots v_{m-1}}$ can be deduced as:

$$p = \frac{\text{Number}\{j \mid j \leq n - (m - 1)\tau, \pi_{v_0 v_1 \ldots v_{m-1}}\}}{n - (m - 1)\tau} \quad (18)$$

where $\text{Number}\{j \mid j \leq n - (m - 1)\tau, \pi_{v_0 v_1 \ldots v_{m-1}}\}$ is the number of occurrences of each $\pi_{v_0 v_1 \ldots v_{m-1}}$ corresponding to $z_i^m$.

Dispersion entropy can be defined according to information entropy theory:

$$DE(x, m, c, \tau) = -\sum_{n=1}^{c^m} p \cdot \ln(p) \quad (19)$$

A large DE value means that the time series is more irregular; on the contrary, the smaller the DE value, the lower the irregularity.

3.1.2. Fine-Sorted Dispersion Entropy

Dispersion entropy, as an information entropy feature, can effectively monitor the dynamic changes in vibration signal and measure its irregularity, so it can be effectively applied to the state monitoring of rolling bearings. However, its ability to feature extraction for the complex non-stationary signal can be further improved in some respects. One possible improvement would be to enhance the perception of neighboring element values. In the theory of dispersion entropy, dispersion entropy focuses on the class sequence that mapping the elements of time series into positive integers. The relationship between elements, as comparatively critical information, has not been considered enough, which suggests that the degree of difference between neighboring elements is not concerned. According to the mapping rule of dispersion entropy, the same dispersion pattern may come from multiple forms of sample vectors. Figure 1 is an example of possible vectors corresponding to the same dispersion pattern (dispersion pattern [2 2 3 3]).

![Figure 1. Example of possible vectors corresponding to the same dispersion pattern.](image-url)
To enhance the feature extraction ability of dispersion entropy, an improved dispersion entropy called fine-sorted dispersion entropy (FSDE) is proposed. FSDE introduces a factor $f$ to measure the difference between neighboring elements in the vector $y^m$. The factor $f$ is then added in class sequence $z^{m,c}_j$ as an additional element. Therefore, the relationship information between elements is brought into dispersion entropy. This operation not only inherits the features within the result of original DE but also improves the case that vectors with a distinct trend are assigned to the same class sequence, which may provide a powerful aid to feature extraction process of FSDE [31,32]. The calculation equation for factor $f$ is as follows:

$$f = \frac{\max(|d_x|)}{\rho \cdot \text{std}(|d_x|)}$$  \hspace{1cm} (20)

where $\lfloor \cdot \rfloor$ returns the largest integer that is less than or equal to the value in brackets, $d_x = \{x_{i+1} - x_i | i = 1, 2, \ldots, n - 1\}$, $d_y = \{y_{i+1} - y_i | i = 1, 2, \ldots, m - 1\}$, $\text{std}$ represent the standard deviation and $\rho$ is a precision parameter. Obviously, if $\rho > \frac{\max(|d_y|)}{\text{std}(|d_y|)}$, the value of $f$ can only be 0, and this additional element will not affect the results of dispersion pattern, which means that the value of FSDE will be exactly the same as original DE. If $\rho \in (0, \frac{\max(|d_y|)}{\text{std}(|d_y|)})$, $f$ will have more possible values, the closer $\rho$ gets to 0, the original dispersion pattern will be subdivided into more new dispersion patterns. As an example, time series [0.5114, 0.5131, 0.5134, 0.5130], [0.5012, 0.5085, 0.5152, 0.5212] and [0.5326, 0.5419, 0.5429, 0.5434, 0.5438] are di

$$z^{m,c}_j$$

$$z^{c}_j$$

$$\ldots$$

$$z^{c}_{j+(m-1)}$$

$$\Rightarrow$$

$$z^{m,c}_j$$

$$z^{c}_j$$

$$\ldots$$

$$z^{c}_{j+(m-1)}$$

$$0$$

$$1$$

$$\ldots$$

$$f$$

Figure 2. $f$ is added in class vector of dispersion entropy as an additional element.

After obtaining the new class sequence $z^{m,c}_j$, $\pi_{\nu\nu\nu\nu\nu\nu\nu}$ denotes a dispersion pattern corresponding to a $z^{m,c}_j$. The relative frequency of new dispersion pattern is calculated as:

$$p' = \frac{\text{Number}\{j | j \leq n - (m - 1)\tau, \pi_{\nu\nu\nu\nu\nu\nu\nu}\}}{n - (m - 1)\tau}$$ \hspace{1cm} (21)

Then, similar to the definition of dispersion entropy, the proposed fine-sorted dispersion entropy can be defined as:

$$FSDE(x, m, c, \tau, f) = -\sum_{n=1}^{p'} p' \cdot \ln(p')$$  \hspace{1cm} (22)
3.2. Mutation SCA-PSO Optimization

3.2.1. Sine Cosine Algorithm

Sine cosine algorithm (SCA) is a new swarm optimization algorithm, which is simple in structure and easy to realize. Using only sine and cosine functions, SCA achieves a good optimization effect. Exploration and exploitation are the two key phases when SCA processes optimization a problem (see [26] for more details). With an elaborate mechanism of exploration and exploitation, SCA is able to search for feasible solutions quickly in the searching space. Let \( Z_i = (Z_{i1}, Z_{i2}, \ldots, Z_{id})^T \) be the position of \( i \)-th individual, whereupon each solution of the optimization problem corresponds to a position of a corresponding individual in the searching space, where \( d \) is the dimension of individuals. The core updating equation of SCA is as follows:

\[
Z_i(l + 1) = \begin{cases} 
Z_i(l) + r_1 \times \sin(r_2) \times |r_3 P_i(l) - Z_i(l)|, & r_4 < 0.5 \\
Z_i(l) + r_1 \times \cos(r_2) \times |r_3 P_i(l) - Z_i(l)|, & r_4 \geq 0.5
\end{cases}
\]  

(23)

where \( P_l = (P_{l1}, P_{l2}, \ldots, P_{ld})^T \) is the best position of individual \( i \). The parameter \( r_1 = a - l(a/l_{\text{max}}) \) determine the region of individual \( i \) at next iteration, where \( l_{\text{max}}, l, \) and \( a \) are the maximum number of iterations, the current number of iterations, and a constant, respectively. The parameter \( r_2 \) is a random number in the scope of \([0, 2\pi]\), which defines the distance that the next movement of individuals should be towards. To stochastically emphasize \((r_3 > 1)\) or deemphasize \((r_3 < 1)\) the effect of the best position searched so far, a random weight \( r_3 \) within \([0, 2]\) is brought into the equation. Finally, the parameter \( r_4 \) is a random number in the range of \([0, 1]\) to switch fairly between sine and cosine components.

3.2.2. Particle Swarm Optimization

Particle swarm optimization (PSO) is a classical stochastic optimization algorithm based on swarm intelligence. In the PSO algorithm, each particle corresponds to a potential solution of the problem to be optimized [33]. Particle \( P_i \) has its own velocity and position, which are expressed as \( d \)-dimensional vectors \( v_i = (v_{i1}, v_{i2}, \ldots, v_{id})^T \) and \( s_i = (s_{i1}, s_{i2}, \ldots, s_{id})^T \), respectively. Particle \( P_i \) keeps track of the optimal solution it has found so far, namely, individual extreme value \( P_i^{\text{best}} \). The optimal solution found by the whole population is called the global extreme value \( P_g^{\text{best}} \). Particles of the swarm update themselves by tracking the above two extremes values. PSO generates a group of particles randomly to initiate the optimization and then solves the problem by iterations. In the \( l \)-th iteration, particle \( P_i \) updates its velocity and position according to the following two equations:

\[
v_i(l + 1) = w \cdot v_i(l) + c_1 \cdot \text{rand}(P_i^{\text{best}} - P_i(l)) + c_2 \cdot \text{rand}(P_g^{\text{best}} - P_i(l))
\]

(24)

\[
s_i(l + 1) = s_i(l) + v_i(l)
\]

(25)

where \( w \) is the inertia weight which is employed to balance the global and local search ability of the algorithm. \( c_1 \) and \( c_2 \) are learning factors, \( \text{rand} \) is a random number between \((0, 1)\), \( l \) is the current number of iterations.

3.2.3. Mutation SCA-PSO Optimization

In the updating strategy of SCA, sine and cosine functions in Equation (23) possess the outstanding ability of exploration but also have some drawbacks, such as slow convergence and the likelihood of being trapped in a local optimum. The reason for the above problems is that parameters of SCA are difficult to set properly, so a poor combination of parameters will result in weak exploitation, but its exploration phase will not be affected [34]. On the other hand, the PSO algorithm has the advantages of information exchange among particles and good robustness, which lead to the high probability of exploiting local optimal solution. However, similar to SCA, PSO also has the problem of local optimum in the later stage of iterations.
Since SCA and PSO have their own strengths, the combination of capabilities between SCA and PSO is expected to improve the convergence performance, which may make the hybrid algorithm powerful in both exploring different regions and exploiting promising regions. As shown in Figure 3, the SCA-PSO approach is constructed in a hierarchical form. The top layer contains \( M \) particles which are searching solutions according to the strategy of PSO. The bottom layer contains \( M \cdot N \) individuals which are evenly distributed to each top layer particle (total \( M \) groups) and their positions update by SCA strategy.

**Figure 3.** The structure of the mutation sine cosine algorithm and particle swarm optimization (SCA-PSO).

In iterations, the optimal value found by each group in the bottom layer is saved by a top layer particle. Through this operation, the top layer particles focus on exploitation by using advantages of PSO, while the bottom layer individuals focus on exploration, giving play to advantages of SCA [34]. The updating strategy of bottom layer individuals is changed to the following equation:

\[
Z_{ij}^{l+1} = \begin{cases} 
Z_{ij}^l + r_1 \times \sin(r_2) \times [r_3 s_i^l - Z_{ij}^l], & r_4 < 0.5 \\
Z_{ij}^l + r_1 \times \cos(r_2) \times [r_3 s_i^l - Z_{ij}^l], & r_4 \geq 0.5
\end{cases}
\]  

(26)

where \( Z_{ij}^l \) is the position of \( j \)-th (\( j = 1, 2, \ldots, N \)) bottom layer individual that belongs to the \( i \)-th (\( i = 1, 2, \ldots, M \)) top layer particle. The position of \( i \)-th top layer particle is described as \( s_i^l \). \( l \) is the current number of iterations.

Moreover, analysis of PSO shows that the population diversity of PSO gradually decreasing with each iteration and it is more likely to result in a local optimum in the later stage of iterations. To solve this problem, a mutation operator is introduced to enrich the diversity of PSO particles [35]. The updating rules of top layer particles containing the mutation operators are as follows:

\[
v_i^l = w \cdot v_i^l + c_1 \cdot \text{rand}(s_i^{l \text{best}} - s_i^l) + c_2 \cdot \text{rand}(s_{g \text{best}}^l - s_i^l)
\]  

(27)

\[
s_i^{l+1} = \begin{cases} 
s_i^l + v_i^l, & l \neq nT \\
[1 + G(0.5 - \text{unif})] s_i^l, & l = nT
\end{cases} \quad i = 1, 2, \ldots
\]  

(28)

where \( G \) and \( T \) (\( T < t_{\text{max}} \)) are the given mutation amplitude and mutation period, respectively, while \( \text{unif} \) represents a random number that conforms to the uniform distribution \( U(0, 1) \). It can be observed that during the updating process, the position \( s_i \) will mutate periodically, which gives PSO particles a certain ability to jump out of the local optimum even in the later stage of iterations.

The computational complexity of the proposed mutation sine cosine algorithm and particle swarm optimization (SCA-PSO) algorithm is \( O(M(N_{\text{sca}} + N_{\text{pso}})) \), where \( t_{\text{sca}} \) and \( t_{\text{pso}} \) are the computational costs of updating all the search agents in each iteration for SCA and PSO, respectively. Accordingly, the computational complexity of SCA and PSO are \( O(t_{\text{sca}}) \) and \( O(t_{\text{pso}}) \), respectively.

The main steps of proposed mutation SCA-PSO (MSCAPS0) approach are as below, and its block diagram is shown in Figure 4:
Step 1: Initialize $M$ particles in the top layer as well as $M \cdot N$ individuals in the bottom layer randomly and set mutation parameters;

Step 2: Calculate the fitness values of all searching agents;

Step 3: Update $s_i$ on the basis of the best solution found by each group in the bottom layer according to Equation (26);

Step 4: Update $s^{ib}_{l}^{gbest}$ in top layer based on Equations (27) and (28);

Step 5: If the maximum number of iterations is not meet, go to step 3;

Step 6: Return the value of $s^{ib}_{l}^{gbest}$ as the optimal solution.

In this section, a novel fault diagnosis method based on FSDE and SVM optimized by MSCAPSO is proposed to improve the fault diagnosis accuracy. The specific steps are detailed as follows:

Step 1: Determine the mode number $K$ of VMD by central frequency observation method;

Step 2: Decompose samples into sets of IMFs with VMD;

Step 3: Calculate the FSDE value of each IMF;

Step 4: Construct feature vectors of different fault samples with FSDE values;

Step 5: Optimize parameters $C$ and $g$ for SVM with the proposed MSCAPSO optimization method;

Step 6: Train the SVM model with optimal parameters $C$ and $g$;

Step 7: Apply the optimal SVM model to classify different fault samples and evaluate the performance of the model.

The flowchart of the proposed fault diagnosis model is shown in Figure 5.
4. Engineering Application

4.1. Data Collection

Vibration signals with different fault locations and sizes gathered from Bearings Data Center of Case Western Reserve University (CWRU) [36] were employed as the experiment data in this research. The experiment stand was mainly composed of a motor, an accelerometer, a torque sensor/encoder and a dynamometer. The bearing was a deep groove ball bearing with model SKF6205-2RS. Accelerometers were applied to collect vibration signal. Meanwhile, the torque sensor/encoder was for the measurement of power and speed. By using electro-discharge machining, the experiment device simulated three fault states of the rolling bearing: inner race fault, ball element fault, and outer race fault, and the depth of faults was 0.011 inches. In the experiment, vibration signals collected from the drive end were taken as the research objects, where the sample frequency was 12 kHz. To fully verify the effectiveness of the proposed fault diagnosis method, nine types samples were applied in this paper, namely inner race fault, ball fault, and outer race fault with diameters of 0.007, 0.014, and 0.021 inches. In addition, the experiments are conducted under the load of 2, 1, and 0 hp with motor speed 1750, 1772, and 1797 rpm. Further, the vibration signal of each type of faults contained 59 samples, and each sample contained 2048 sample points. The experimental data are listed in Table 1.

Table 1. Description of the experimental data.

| Motor Load (hp) | Position of Fault | Defect Size (Inches) | Label of Classes | Number of Samples |
|-----------------|------------------|----------------------|------------------|-------------------|
| 2/1/0           | Inner race       | 0.007                | L1               | 59                |
|                 | Inner race       | 0.014                | L2               | 59                |
|                 | Inner race       | 0.021                | L3               | 59                |
|                 | Ball             | 0.007                | L4               | 59                |
|                 | Ball             | 0.014                | L5               | 59                |
|                 | Ball             | 0.021                | L6               | 59                |
|                 | Outer race       | 0.007                | L7               | 59                |
|                 | Outer race       | 0.014                | L8               | 59                |
|                 | Outer race       | 0.021                | L9               | 59                |
4.2. Application to Fault Diagnosis of Rolling Bearings

The proposed FSDE-MSCAPSO method has been applied to diagnose nine types of faults and compared with other six relevant fault diagnosis methods at feature extraction as well as parameter optimization stage. More specifically, fuzzy entropy (FE) [37], permutation entropy (PE) [4], and dispersion entropy (DE) were applied at the feature extraction stage for comparison; SCA and PSO were applied at the parameter optimization stage. The parameters of the contrastive experiments were set in the same way as the proposed method.

In the proposed method, the first step is to decompose the signal of each type of fault into a set of IMFs. The mode number K of VMD needs to be determined in advance with the central frequency observation method. If K is too large, central frequencies of neighboring IMFs may be too close, then mode mixing will emerge. Normalized central frequencies under different K values are listed in Table 2, where K was ascertained by the sample of ball fault with a diameter of 0.014 inches (L5). As listed in Table 2, when K was set 5, the first two central frequencies were relatively close. Similarly, as illustrated in Figure 6, during the iterative calculation of VMD, if K was 5 or greater, there were some central frequencies of IMFs close to each other, which meant excessive decomposition had occurred. Hence, parameter K was set to 4.

The waveforms of the original signals are shown in Figure 7. It can be observed that waveforms of different fault positions or different defect sizes are different to a certain extent. The examples of decomposition results with different fault positions (L2, L5, L8) and different defect sizes (L4, L5, L6) are illustrated in Figure 8, which shows that IMFs decomposed from the same types of faults are quite different.

Table 2. Normalized central frequencies with different K values.

| Number of Modes | Normalized Central Frequencies |
|-----------------|--------------------------------|
| 2               | 0.2623 0.0438                  |
| 3               | 0.2698 0.2227 0.0418           |
| 4               | 0.2700 0.2246 0.1112 0.0367    |
| 5               | 0.2760 0.2614 0.2209 0.1107 0.0366 |
| 6               | 0.2992 0.2745 0.2606 0.2200 0.1106 0.0366 |
| 7               | 0.3301 0.2754 0.2610 0.2206 0.1146 0.0548 0.0315 |
| 8               | 0.3663 0.2847 0.2711 0.2548 0.2180 0.1144 0.0546 0.0314 |

Figure 6. The variation of central frequency with iteration under different K values.
well distinguished.

which shows that FSDE value of every IMF is di-

mension. However, FSDE may also be oversensitive to small changes and increase the computation. Considering all these factors, embedding dimension is large, there will be a high computational cost. When dealing with these factors, embedding dimension is usually set from two to seven, if $m$ is large, there will be a high computational cost. When dealing with $c$, it is usually suggested to choose from four to eight, if $c$ is small, it may not be able to recognize two values with very different amplitudes [38]. $\rho$ is the key parameter that distinguishes FSDE from DE. The closer $\rho$ to 0, an original dispersion pattern will be subdivided into more new dispersion patterns, however, FSDE may also be oversensitive to small changes and increase the computation. Considering all these factors, embedding dimension $m$ was set at three and the number of class $c$ was set at six in our research. Moreover, for the practical purpose, it was recommended to assume the time delay $\tau = 1$ [38], while precision parameter $\rho$ was set as 2.9 to balance the precision and computational complexity.

The first three FSDE and DE vectors of different fault samples (L1–L9) are compared in Table 3, which shows that FSDE value of every IMF is different from the corresponding DE value, but the overall trend remains similar. 3D projection of entropy vectors for different types of faults is shown in Figure 9. It can be seen that feature vectors extracted by PE are not clearly distinguished, quite a lot of feature vectors from different types of faults are still mixed together. This situation is improved in the figures of FE and DE to a certain extent, while feature vectors extracted by FSDE are basically well distinguished.
Table 3. Fine-sorted dispersion entropy and dispersion entropy of intrinsic mode functions (IMFs).

| Fault Label | Sample Number | Fine-Sorted Dispersion Entropy | Dispersion Entropy |
|-------------|---------------|-------------------------------|--------------------|
|             |               | IMF1  | IMF2  | IMF3  | IMF4  | IMF1  | IMF2  | IMF3  | IMF4  |
| L1          | 1             | 0.4428 | 0.4126 | 0.3860 | 0.3154 | 0.4110 | 0.3740 | 0.3202 | 0.2509 |
|             | 2             | 0.4421 | 0.4143 | 0.3854 | 0.3154 | 0.4074 | 0.3734 | 0.3204 | 0.2520 |
|             | 3             | 0.4402 | 0.4116 | 0.3862 | 0.3151 | 0.4030 | 0.3700 | 0.3201 | 0.2495 |
| L2          | 1             | 0.3799 | 0.3671 | 0.3828 | 0.3003 | 0.3486 | 0.3344 | 0.3157 | 0.2370 |
|             | 2             | 0.3806 | 0.3713 | 0.3849 | 0.2959 | 0.3606 | 0.3520 | 0.3170 | 0.2380 |
|             | 3             | 0.3822 | 0.3675 | 0.3864 | 0.3069 | 0.3611 | 0.3347 | 0.3163 | 0.2452 |
| L3          | 1             | 0.4074 | 0.5448 | 0.4277 | 0.3061 | 0.3730 | 0.5225 | 0.3881 | 0.2514 |
|             | 2             | 0.4231 | 0.5133 | 0.4222 | 0.3050 | 0.3907 | 0.4924 | 0.3829 | 0.2519 |
|             | 3             | 0.3986 | 0.5113 | 0.4246 | 0.3064 | 0.3576 | 0.4859 | 0.3773 | 0.2509 |
| L4          | 1             | 0.3993 | 0.3838 | 0.3611 | 0.2623 | 0.3146 | 0.3056 | 0.2970 | 0.2031 |
|             | 2             | 0.4034 | 0.3876 | 0.3555 | 0.2452 | 0.3120 | 0.3041 | 0.2914 | 0.1789 |
|             | 3             | 0.4028 | 0.3874 | 0.3622 | 0.2891 | 0.3135 | 0.3108 | 0.2968 | 0.2288 |
| L5          | 1             | 0.3708 | 0.3707 | 0.3884 | 0.2769 | 0.3063 | 0.3112 | 0.3176 | 0.2218 |
|             | 2             | 0.3638 | 0.3725 | 0.3861 | 0.2999 | 0.3043 | 0.3129 | 0.3143 | 0.2198 |
|             | 3             | 0.3731 | 0.3856 | 0.3873 | 0.2783 | 0.3021 | 0.3203 | 0.3162 | 0.2206 |
| L6          | 1             | 0.3513 | 0.3916 | 0.3938 | 0.3026 | 0.2897 | 0.3157 | 0.3190 | 0.2397 |
|             | 2             | 0.3847 | 0.3943 | 0.3944 | 0.3031 | 0.3109 | 0.3153 | 0.3178 | 0.2415 |
|             | 3             | 0.3851 | 0.3882 | 0.4021 | 0.3127 | 0.3117 | 0.3135 | 0.3187 | 0.2499 |
| L7          | 1             | 0.4561 | 0.5815 | 0.4964 | 0.3546 | 0.4372 | 0.5582 | 0.4718 | 0.2926 |
|             | 2             | 0.4613 | 0.5820 | 0.4910 | 0.3444 | 0.4414 | 0.5569 | 0.4685 | 0.2859 |
|             | 3             | 0.4606 | 0.5864 | 0.4960 | 0.3754 | 0.4395 | 0.5632 | 0.4736 | 0.3167 |
| L8          | 1             | 0.4052 | 0.3892 | 0.3477 | 0.2739 | 0.3282 | 0.3083 | 0.2825 | 0.2159 |
|             | 2             | 0.4162 | 0.3953 | 0.3234 | 0.2770 | 0.3329 | 0.3159 | 0.2669 | 0.2165 |
|             | 3             | 0.4024 | 0.3884 | 0.3539 | 0.2979 | 0.3296 | 0.3140 | 0.2892 | 0.2320 |
| L9          | 1             | 0.4793 | 0.4115 | 0.3643 | 0.3088 | 0.4746 | 0.4041 | 0.3244 | 0.2656 |
|             | 2             | 0.4899 | 0.3863 | 0.3996 | 0.3169 | 0.4855 | 0.3758 | 0.3743 | 0.2778 |
|             | 3             | 0.4906 | 0.5040 | 0.4291 | 0.3139 | 0.4825 | 0.4966 | 0.4234 | 0.2751 |

In the optimization stage, the feature vectors from different types of faults were divided randomly into two parts with 30 and 29 vectors, where 30 were used for training while the rest, 29, for testing. The proposed MSCAPSO approach was applied to optimize the penalty factor C and the kernel parameter g of SVM. The total number of searching agents and iteration times were set 40 and 100, respectively, while searching ranges of C and g were both $[2^{-10}, 2^{10}]$. Considering the maximum number of iterations and the number of optimization parameters, the mutation period was set at five and mutation amplitude was set at one. Fitness values were calculated by five-fold cross-validation for the training samples. Hence, the optimization effect of the given C and g could be measured by the average accuracy of cross-validation, then the optimal C and g were obtained.
The convergence procedure of MSCAPSO is depicted in Figure 10a, from which it can be observed that the average fitness value of particles rises rapidly at an early stage of the iterations and then tends to be stable. That is, the particles are close to the global optimal solution. The shaded part shows the distribution of convergence curves in 10 runs. The comparison of different optimization methods is shown in Figure 10b, where all convergence experiments are based on the same FSDE feature vectors, and each convergence curve is the average of 10 runs. The figure shows that the convergence curve of PSO fluctuates greatly, and the average fitness value is difficult to be further improved after 50 iterations. SCA has a good overall convergence trend, but the convergence rate is significantly lower than the other two methods. MSCAPSO has already converged to a high level in the early stage and keeps the best convergence effect among all methods in the whole iterations.

With the optimal parameters \( C \) and \( g \), the SVM model optimized by MSCAPSO was trained and applied to recognize testing samples. For an in-depth verification of the proposed method, diagnosis results were averaged over 10 runs, and training samples were chosen at random in every run. Furthermore, the deviation of each result was calculated as a reference for result evaluation. In this application, adjusted Rand index (ARI), normalized mutual information (NMI), F-measure (F), and accuracy (ACC) as four widely used evaluation metrics are employed to evaluate the diagnosis results [39,40], which reflects the matching degree between classification result and real distribution information of samples. The scope of ARI is \([-1, 1]\) and the rest scopes are all \([0, 1]\). A higher value

**Figure 9.** 3D projection of entropy vectors for different types of faults: (a) vectors of permutation entropy, (b) fuzzy entropy, (c) dispersion entropy(d), fine-sorted dispersion entropy.

**Figure 10.** Convergence procedure of proposed optimization approach: (a) distribution of convergence curves in 10 runs (b) comparison of different optimization methods.
indicates a better classification performance. The optimal $C$ and $g$ are presented corresponding to the best accuracy among 10 runs.

Fault diagnosis results with different methods under variable load conditions are shown in Table 4 and Figure 11. It can be observed from Table 4 that four evaluation metrics of the proposed FSDE-MSCAPSO method are all the highest when the motor loads are 2 hp and 1 hp, i.e., 0.9637, 0.9646, 0.9839, 0.9839 and 0.9615, 0.9827, 0.9828. As well, the deviations of the evaluation values are also at a very small level among all methods. When the motor load is 0 hp, although the value of proposed method is slightly lower than FE-PSO in each metric, its performance is still better than DE-PSO/SCA methods than before improvements. To be specific, the comparison of FE-PSO, DE-PSO and FSDE-PSO indicates that the proposed FSDE has a better feature extraction ability than FE, PE and DE in the field of fault diagnosis. Similarly, this fact is also verified in the comparison among PE-SCA, DE-SCA, and FSDE-SCA. Moreover, it can be concluded from the contrastive experiments between FSDE-SCA/PSO and FSDE-MSCAPSO that the proposed MSCAPSO optimization strategy improves the diagnosis performance of SVM.

The boxplots of evaluation values are shown in Figure 12, illustrating the performances of different diagnostic methods. As shown in the figure, the proposed method achieves better overall performance and stability for fault diagnosis than other contrastive methods. Furthermore, several statistical metrics that have been widely used for robustness evaluation [10,41,42] are applied to demonstrate the robustness of the diagnosis models, i.e., mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). The error metrics of the outputs of different models under various loads are listed in Table 5, as well as the corresponding histograms shown in Figure 13. It can be observed that error metrics of FSDE-MSCAPSO are the lowest when the motor loads are 2 hp and 1 hp. When the motor load is 0 hp, although the metric values of the proposed method are not the lowest among all models, it is still at a low level and in the second position in each metric, indicating the proposed model has good robustness on the whole.

Table 4. Fault diagnosis results with different methods under variable load conditions.

| Motor Load (hp) | Methods       | $C$         | $g$          | Evaluation Metrics |
|-----------------|---------------|-------------|--------------|--------------------|
|                 |               | Adjusted Rand Index (ARI) | Normalized Mutual Information (NMI) | F-Measure (F) | Accuracy (ACC) |
| 2               | FE-PSO        | 204.766     | 7.978        | 0.9309 [−0.025, 0.043] | 0.9288 [−0.021, 0.038] | 0.9687 [−0.010, 0.020] | 0.9686 [−0.011, 0.020] |
|                 | PE-SCA        | 616.339     | 411.111      | 0.9185 [−0.043, 0.046] | 0.9207 [−0.032, 0.048] | 0.9628 [−0.020, 0.022] | 0.9628 [−0.020, 0.022] |
|                 | DE-PSO        | 13.124      | 1024         | 0.9461 [−0.023, 0.028] | 0.9463 [−0.021, 0.033] | 0.9759 [−0.010, 0.013] | 0.9759 [−0.010, 0.013] |
|                 | DE-SCA        | 12.762      | 420.762      | 0.9456 [−0.022, 0.028] | 0.9513 [−0.020, 0.019] | 0.9755 [−0.010, 0.013] | 0.9755 [−0.010, 0.013] |
|                 | FSDE-PSO      | 54.370      | 151.024      | 0.9572 [−0.016, 0.009] | 0.9603 [−0.019, 0.011] | 0.9808 [−0.008, 0.004] | 0.9808 [−0.008, 0.004] |
|                 | FSDE-SCA      | 25.206      | 200.267      | 0.9549 [−0.014, 0.019] | 0.9589 [−0.021, 0.019] | 0.9797 [−0.007, 0.009] | 0.9797 [−0.007, 0.009] |
|                 | FSDE-MSCAPSO  | 6.085       | 390.627      | 0.9637 [−0.019, 0.019] | 0.9646 [−0.016, 0.014] | 0.9839 [−0.007, 0.008] | 0.9839 [−0.007, 0.008] |
Table 4. Cont.

| Motor Load (hp) | Methods     | C     | $g$      | Evaluation Metrics |
|----------------|-------------|-------|----------|--------------------|
|                |             |       |          | Adjusted Rand Index (ARI) | Normalized Mutual Information (NMI) | F-Measure (F) | Accuracy (ACC) |
| 1              | FE-PSO      | 6.908 | 2.358    | 0.904 [-0.035, 0.035] | 0.9243 [-0.050, 0.035] | 0.9681 [-0.027, 0.018] | 0.9682 [-0.027, 0.018] |
|                | PE-SCA      | 69.617| 294.338  | 0.9019 [-0.043, 0.044] | 0.8942 [-0.021, 0.015] | 0.9537 [-0.012, 0.015] | 0.9544 [-0.015, 0.015] |
|                | DE-PSO      | 2.217 | 586.466  | 0.8814 [-0.040, 0.051] | 0.8827 [-0.046, 0.041] | 0.9458 [-0.021, 0.027] | 0.9460 [-0.027, 0.026] |
|                | DE-SCA      | 331.824| 250.205  | 0.8864 [-0.042, 0.047] | 0.8869 [-0.035, 0.034] | 0.9548 [-0.022, 0.024] | 0.9566 [-0.024, 0.023] |
|                | FSDE-PSO    | 2.754 | 1024.000 | 0.9483 [-0.023, 0.018] | 0.9468 [-0.024, 0.024] | 0.9765 [-0.011, 0.008] | 0.9766 [-0.011, 0.008] |
|                | FSDE-SCA    | 4.532 | 690.248  | 0.9386 [-0.029, 0.044] | 0.9387 [-0.025, 0.040] | 0.9971 [-0.015, 0.021] | 0.9972 [-0.015, 0.020] |
|                | FSDE-SCA    | 4.532 | 690.248  | 0.9386 [-0.029, 0.044] | 0.9387 [-0.025, 0.040] | 0.9971 [-0.015, 0.021] | 0.9972 [-0.015, 0.020] |
|                | FSDE-MSCAPSO| 5.004 | 1024.000 | 0.9615 [-0.038, 0.021] | 0.9598 [-0.025, 0.019] | 0.9827 [-0.018, 0.010] | 0.9828 [-0.017, 0.010] |
| 0              | FE-PSO      | 9.453 | 9.781    | 0.9141 [-0.034, 0.052] | 0.9144 [-0.035, 0.051] | 0.9596 [-0.018, 0.025] | 0.9598 [-0.017, 0.025] |
|                | PE-SCA      | 853.002| 32.932   | 0.8850 [-0.045, 0.031] | 0.8862 [-0.048, 0.038] | 0.9447 [-0.021, 0.018] | 0.9448 [-0.021, 0.017] |
|                | DE-PSO      | 16.319| 456.170  | 0.8942 [-0.055, 0.038] | 0.8997 [-0.054, 0.032] | 0.9512 [-0.027, 0.018] | 0.9510 [-0.028, 0.018] |
|                | DE-SCA      | 155.860| 163.498  | 0.8753 [-0.037, 0.035] | 0.8855 [-0.025, 0.028] | 0.9596 [-0.020, 0.018] | 0.9598 [-0.020, 0.018] |
|                | FSDE-PSO    | 16.138| 549.663  | 0.9036 [-0.020, 0.045] | 0.9111 [-0.034, 0.046] | 0.9954 [-0.010, 0.022] | 0.9956 [-0.009, 0.021] |
|                | FSDE-SCA    | 1024.000| 2.400    | 0.8873 [-0.034, 0.022] | 0.8948 [-0.042, 0.024] | 0.9973 [-0.016, 0.011] | 0.9975 [-0.016, 0.010] |
|                | FSDE-MSCAPSO| 216.051| 448.183  | 0.9079 [-0.031, 0.033] | 0.9086 [-0.035, 0.033] | 0.9567 [-0.015, 0.017] | 0.9571 [-0.015, 0.016] |

Figure 11. Comparison of evaluation values for different methods (2 hp).
Figure 12. Boxplots of evaluation values for different methods under variable load conditions: (a) 2hp, (b) 2hp, (c) 0hp; the x-axis tick labels correspond to: 1: Fuzzy entropy-particle swarm optimization (FE-PSO), 2: Permutation entropy-sine cosine algorithm (PE-SCA), 3: Dispersion entropy-particle swarm optimization (DE-PSO), 4: Dispersion entropy-sine cosine algorithm (DE-SCA), 5: Fine-sorted dispersion entropy-particle swarm optimization (FSDE-PSO), 6: Fine-sorted dispersion entropy-sine cosine algorithm (FSDE-SCA), 7: Fine-sorted dispersion entropy-mutation operator, sine cosine algorithm and particle swarm optimization (FSDE-MSCAPSO).
Table 5. Error metrics of results of different methods.

| Method          | MAE 2 hp | MAE 1 hp | MAE 0 hp | RMSE 2 hp | RMSE 1 hp | RMSE 0 hp | MAPE 2 hp | MAPE 1 hp | MAPE 0 hp |
|-----------------|----------|----------|----------|-----------|-----------|-----------|-----------|-----------|-----------|
| FE-PSO          | 0.0314   | 0.0318   | 0.0402   | 0.0324    | 0.0354    | 0.0420    | 0.0325    | 0.0331    | 0.0421    |
| PE-SCA          | 0.0372   | 0.0456   | 0.0552   | 0.0390    | 0.0462    | 0.0566    | 0.0388    | 0.0478    | 0.0586    |
| DE-PSO          | 0.0241   | 0.0567   | 0.0490   | 0.0253    | 0.0592    | 0.0516    | 0.0248    | 0.0605    | 0.0519    |
| DE-SCA          | 0.0245   | 0.0540   | 0.0602   | 0.0256    | 0.0558    | 0.0617    | 0.0252    | 0.0573    | 0.0642    |
| FSDE-PSO        | 0.0192   | 0.0234   | 0.0444   | 0.0195    | 0.0240    | 0.0453    | 0.0195    | 0.0240    | 0.0466    |
| FSDE-SCA        | 0.0203   | 0.0280   | 0.0525   | 0.0210    | 0.0300    | 0.0531    | 0.0208    | 0.0289    | 0.0555    |
| FSDE-MSCAPSO    | 0.0161   | 0.0172   | 0.0429   | 0.0169    | 0.0190    | 0.0443    | 0.0164    | 0.0176    | 0.0450    |

Figure 13. Comparison of mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) for diagnosis models under different load conditions: (a) MAE, (b) RMSE, (c) MAPE.

5. Discussion

After the detailed comparative analysis above, the superiority of the proposed FSDE and MSCAPSO methods have been effectively demonstrated, as well, the accuracy of the diagnosis model has also been verified by experiments on various loads, fault sizes, and locations. It is worth noting that the proposed FSDE method only considers the relationship between neighboring elements in a given series, and the overall information of the series is not taken into account. According to the previous references [31,43], the improved method using the overall amplitude information of the series has been realized on other entropy methods, which is expected to further improve the performance of the proposed method. Furthermore, some of the parameters in the FSDE are still empirical values in the present paper. In future studies, reasonable metrics can be introduced to evaluate the influence of parameters on the approximation between FSDE and DE, so as to better determine the value of parameters. Furthermore, the proposed MSCAPSO algorithm can be optimized through better strategies to further improve the convergence speed and global search ability. In addition, the multi-objective optimization that has been widely utilized in the field of controlling [44–47] could be implemented in fault diagnosis, which is expected to improve the accuracy and reduce the variance of the outputs of the model [48]. Moreover, although the diagnosis experiments considered various load conditions, fault sizes, and locations, the signal to noise ratio of the CWRU bearing data is relatively high, and the conclusions obtained in the experiment may not be very comprehensive. In the future work, we plan to build our own bearing test device to collect many more vibration signals, and the results obtained under such conditions may be more practical.

Considering that the CWRU bearing data set has been selected by many researchers as experimental data to validate their methods and techniques, the diagnosis accuracies of some other methods based on this data set are listed in Table 6 to compare with the methods presented in this paper. Note that the
segmentation or fault types of experimental samples may not be exactly the same, the accuracies listed is for general description.

Table 6. Comparison among several other fault diagnosis model. The abbreviations are as follows: Chaos quantum sine cosine algorithm (CQSCA), hybrid gravitational search algorithm (HGSA), extreme learning machine (ELM), multivariate autoregressive (MAR), time shift multiscale fuzzy entropy (TSMFE), Laplacian support vector machine (LapSVM), improved multiscale dispersion entropy (IMDE), max-relevance min-redundancy (mRMR), back-propagation neural network (BPNN), improved multiscale permutation entropy (IMPE), quantum behaved particle swarm optimization (QPSO), least squares support vector machine (LSSVM), generalized composite multiscale permutation entropy (GCMPE), semi-supervised kernel Marginal Fisher analysis (SSKMFA).

| Reference | Feature Extraction | Classification | Accuracy |
|-----------|--------------------|----------------|---------|
| [4]       | IMFs decomposed by VMD+PE | CQSCA-SVM | 0.9789 |
| [8]       | IMFs decomposed by EEMD | HGSA-ELM | 0.9938 |
| [12]      | IMFs decomposed by VMD+MAR | HMGWOSCA-SVM | 0.9808 |
| [20]      | IMPE | QPSO-LSSVM | 0.9800 |
| [22]      | TSMFE | C1: LapSVM | 0.9800 |
|           |       | C2: SVM | 0.9844 |
|           |       | C1: mRMR+ELM | 0.9056 |
|           |       | C2: mRMR+BPNN | 0.9860 |
|           |       | C3: mRMR+SVM | 0.8840 |
| [49]      | IMDE | C1: mRMR+ELM | 0.9056 |
| [50]      | GCMPE | PSO-SVM | 0.9889 |
| [51]      | SSKMFA | KNN | 0.9850 |

6. Conclusions

A novel fault diagnosis method for rolling bearings by fine-sorted dispersion entropy and mutation SCA-PSO optimized SVM is presented to diagnose faults of various sizes, locations, and motor loads in this study. Due to the non-stationarity of original vibration signals, signals collected from different types of faults were firstly split by VMD into sets of IMFs, before which the decomposing mode number K was determined by a central frequency observation method. Then an improved dispersion entropy containing an additional factor, termed FSDE, was proposed to enhance the perception for relationship information between neighboring elements. It was employed to construct the feature vectors of different fault samples later. Afterward, a hybrid optimization strategy combining advantages of mutation operator, SCA and PSO called MSCAPSO, was proposed to optimize the SVM model. The optimal SVM model was subsequently applied to realize the classification for different fault samples. The contrastive experiments in Section 4 were implemented among the proposed FSDE-MSCAPSO method and other relevant methods, where nine types of faults with different locations and sizes were successfully diagnosed. The analysis of results indicates that the proposed method achieves the highest 0.9637, 0.9646, 0.9839, and 0.9839 in four widely used evaluation metrics, which exhibits its superior precision and stability over other methods.

Author Contributions: W.F. conceived the research, J.T. performed the experiments and was responsible for paper writing, Y.X. provided guidance for experiments and contributed to revise the paper, K.W. helped analyze results, T.C. provided recommendations for this research.

Funding: This work was funded by the National Natural Science Foundation of China (NSFC) [51741907, 51809099], Hubei Provincial Major Project for Technical Innovation [2017AA132], the Open Fund of Hubei Provincial Key Laboratory for Operation and Control of Cascaded Hydropower Station [2017KX06], Research Fund for Excellent Dissertation of China Three Gorges University [2019SSPY070].

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

| Symbols | Abbreviations |
|---------|---------------|
| K       | mode number   |
| m_k     | set of K mode functions |
| \( \omega_k \) | set of central frequencies |
| \( \partial_t \) | partial derivative of time \( t \) |
| \( \delta_t \) | unit pulse function of time \( t \) |
| \( f(t) \) | real valued input signal |
| \( a \) | penalty factor for VMD |
| \( \beta(t) \) | Lagrange multiplier for VMD |
| \( w \) | weight vector |
| \( b \) | bias parameter |
| \( \xi \) | slack term |
| \( C \) | penalty factor for SVM |
| \( g \) | kernel parameter |
| \( \mu_i \) | Lagrange multiplier for SVM |
| \( x \) | time series |
| \( n \) | length of time series |
| \( \sigma \) | variance of the normal distribution |
| \( \mu \) | expectation of the normal distribution |
| \( y_i \) | reconstruction matrix of \( x \) |
| \( m \) | embedding dimension |
| \( \tau \) | time delay |
| \( c \) | number of class |
| \( z_{m,c} \) | class sequence of DE |
| \( \pi_{z_{m,c}} \) | dispersion pattern of DE |
| \( f \) | factor of FSDE |
| \( \rho \) | precision parameter |
| \( \zeta_{m,c} \) | class sequence of FSDE |
| \( \zeta_{z_{m,c}} \) | dispersion pattern of FSDE |
| \( Z_i \) | position of individual |
| \( P_i \) | best position of individual |
| \( v_i \) | velocity of particle |
| \( s_i \) | position of particle |
| \( P_{EEMD} \) | individual extreme value |
| \( \mu_{EEMD} \) | global extreme value |
| \( c_1, c_2 \) | learning factor |
| \( Z_{b,l} \) | position of bottom layer individual |
| \( s_{b,l} \) | position of top layer particle |
| \( M \) | number of top layer particle |
| \( N \) | number of bottom layer individual |
| \( G \) | mutation amplitude |
| \( l \) | current number of iterations |

## References

1. Brkovic, A.; Gajic, D.; Gligorijevic, J.; Sagicajic, I.; Georgieva, O.; Gennaro, S.D. Early fault detection and diagnosis in bearings for more efficient operation of rotating machinery. *Energy* 2017, 136, 63–71. [CrossRef]
2. Zhao, X.; Qin, Y.; He, C.; Jia, L.; Kou, L. Rolling element bearing fault diagnosis under impulsive noise environment based on cyclic correntropy spectrum. *Entropy* 2019, 21, 50. [CrossRef]
3. Mbo’O, C.P.; Hameyer, K. Fault diagnosis of bearing damage by means of the linear discriminant analysis of stator current features from the frequency selection. *IEEE Trans. Ind. Appl.* 2016, 52, 3861–3868. [CrossRef]
4. Fu, W.; Tan, J.; Li, C.; Zou, Z.; Li, Q.; Chen, T. A hybrid fault diagnosis approach for rotating machinery with the fusion of entropy-based feature extraction and SVM optimized by a chaos quantum sine cosine algorithm. *Entropy* 2018, 20, 626. [CrossRef]
5. Yuan, R.; LV, Y.; Song, G. Fault diagnosis of rolling bearing based on a novel adaptive high-order local projection denoising method. *Complexity* 2018, 2018, 3049318. [CrossRef]
6. Gligorijevic, J.; Gajic, D.; Brkovic, A.; Sagicajic, I.; Georgieva, O.; Gennaro, S.D. Online condition monitoring of bearings to support total productive maintenance in the packaging materials industry. *Sensors* 2016, 16, 316. [CrossRef]
7. Ali, J.B.; Fnaiech, N.; Saidi, L.; Chebel-Morello, B.; Fnaiech, F. Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Appl. Acoust.* 2015, 89, 16–27.

8. Meng, L.; Li, C.; Zhang, X.; Li, R.; An, X. Compound feature selection method and parameter optimization of ELM for fault diagnosis of rolling element bearings. *ISA Trans.* 2016, 65, 556–566.

9. Liu, H.; Jing, J.; Jian, M. Fault diagnosis of electromechanical actuator based on VMD multifractal detrended fluctuation analysis and PNN. *Complexity* 2018, 2018, 9154682. [CrossRef]

10. Fu, W.; Wang, K.; Zhou, J.; Xu, Y.; Tan, J.; Chen, T. A hybrid approach for multi-step wind speed forecasting based on multi-scale dominant ingredient chaotic analysis, KELM and synchronous optimization strategy. *Sustainability* 2019, 11, 1804. [CrossRef]

11. Li, K.; Su, L.; Wu, J.; Wang, H.; Chen, P. A rolling bearing fault diagnosis method based on variational mode decomposition and an improved kernel extreme learning machine. *Appl. Sci.* 2017, 7, 1004. [CrossRef]

12. Fu, W.; Tan, J.; Zhang, X.; Chen, T.; Wang, K. Blind parameter identification of MAR model and mutation hybrid GWO-SCA optimized SVM for fault diagnosis of rotating machinery. *Complexity* 2019, 2019, 3264969.

13. Pincus, S. Approximate entropy (ApEn) as a complexity measure. *Chaos* 1995, 5, 110–117. [CrossRef] [PubMed]

14. Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* 2002, 88, 174102. [CrossRef] [PubMed]

15. Chen, W.; Wang, Z.; Xie, H.; Yu, W. Characterization of surface EMG signal nased on fuzzy entropy. *IEEE Trans. Neural Syst. Rehabil. Eng.* 2007, 15, 266–272. [CrossRef] [PubMed]

16. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol.-Heart Circ. Physiol.* 2000, 278, H2039–H2049. [CrossRef] [PubMed]

17. Rostaghi, M.; Azami, H. Dispersion entropy: A measure for time-series analysis. *IEEE Signal Process. Lett.* 2016, 23, 610–614. [CrossRef]

18. Cai, B.; Huang, L.; Xie, M. Bayesian networks in fault diagnosis. *IEEE Trans. Ind. Inform.* 2017, 13, 2227–2240. [CrossRef]

19. Wang, D. K-nearest neighbors based methods for identification of different gear crack levels under different motor speeds and loads: Revisited. *Mech. Syst. Signal Process.* 2016, 70–71, 201–208. [CrossRef]

20. Raj, N.; Jagadanand, G.; George, S. Fault detection and diagnosis in asymmetric multilevel inverter using artificial neural network. *Int. J. Electron.* 2017, 105, 559–571. [CrossRef]

21. Shi, P.; Liang, K.; Han, D.; Zhang, Y. A novel intelligent fault diagnosis method of rotating machinery based on deep learning and PSO-SVM. *J. Vibroeng.* 2017, 19, 5932–5946. [CrossRef]

22. Zhu, X.; Zheng, J.; Pan, H.; Bao, J.; Zhang, Y. Time-shift multiscale fuzzy entropy and laplacian support vector machine based rolling bearing fault diagnosis. *Entropy* 2018, 20, 602. [CrossRef]

23. Jiang, M.; Jiang, L.; Jiang, D.; Li, F.; Song, H. A sensor dynamic measurement error prediction model based on NAPSO-SVM. *Sensors* 2018, 18, 233. [CrossRef]

24. Xu, Y.; Zheng, Y.; Du, Y.; Yang, W.; Peng, X.; Li, C. Adaptive condition predictive-fuzzy PID optimal control of start-up process for pumped storage unit at low head area. *Energy Convers. Manag.* 2018, 177, 592–604. [CrossRef]

25. Xu, Y.; Li, C.; Wang, Z.; Zhang, N.; Peng, B. Load frequency control of a novel renewable energy integrated micro-grid containing pumped hydropower energy storage. *IEEE Access* 2018, 6, 29067–29077. [CrossRef]

26. Mirjalili, S. SCA: A sine cosine algorithm for solving optimization problems. *Knowl-Based Syst.* 2016, 96, 120–133. [CrossRef]

27. Fu, W.; Wang, K.; Li, C.; Li, X.; Li, Y.; Zhong, H. Vibration trend measurement for a hydropower generator based on optimal variational mode decomposition and an LSSVM improved with chaotic sine cosine algorithm optimization. *Meas. Sci. Technol.* 2019, 30, 015012. [CrossRef]

28. Fu, W.; Wang, K.; Li, C.; Tan, J. Multi-step short-term wind speed forecasting approach based on multi-scale dominant ingredient chaotic analysis, improved hybrid GWO-SCA optimization and ELM. *Energy Convers. Manag.* 2019, 187, 356–377. [CrossRef]

29. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* 2014, 62, 531–544. [CrossRef]

30. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995.
31. Fadlallah, B.H.; Chen, B.; Keil, A.; Principe, J.C. Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information. Phys. Rev. E 2013, 87, 022911. [CrossRef]

32. Xiaofeng, L.; Yue, W. Fine-grained permutation entropy as a measure of natural complexity for time series. Chin. Phys. B 2009, 18, 2690–2695. [CrossRef]

33. Kennedy, J.; Eberhart, R.C. Particle swarm optimization. In Proceedings of the International Symposium on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995; pp. 1942–1948.

34. Issa, M.; Hassani, A.E.; Oliva, D.; Helmi, A.; Ziedan, I.; Alzohairy, A.M. ASCA-PSO: Adaptive sine cosine optimization algorithm integrated with particle swarm for pairwise local sequence alignment. Expert Syst. Appl. 2018, 99, 56–70. [CrossRef]

35. Pehlivanoglu, Y.V. A new particle swarm optimization method enhanced with a periodic mutation strategy and neural networks. IEEE Trans. Evol. Comput. 2013, 17, 436–452. [CrossRef]

36. Bearing Data Center of the Case Western Reserve University. Available online: http://csegroups.case.edu/bearingdatacenter/pages/download-data-file (accessed on 28 December 2018).

37. Zhao, H.; Sun, M.; Deng, W.; Yang, X. A new feature extraction method based on EEMD and multi-scale fuzzy entropy for motor bearing. Entropy 2016, 19, 14. [CrossRef]

38. Rostaghi, M.; Ashory, M.R.; Azami, H. Application of dispersion entropy to status characterization of rotary machines. J. Sound Vib. 2019, 438, 291–308. [CrossRef]

39. Nguyen, T.P.Q.; Kuo, R.J. Partition-and-merge based fuzzy genetic clustering algorithm for categorical data. Appl. Soft Comput. 2019, 75, 254–264. [CrossRef]

40. Fahad, A.; Alshatri, N.; Tari, Z.; Alamri, A.; Khalil, I.; Zomaya, A.Y.; Foufou, S.; Bouras, A. A survey of clustering algorithms for big data: Taxonomy and empirical analysis. IEEE Trans. Emerg. Top. Comput. 2014, 2, 267–279. [CrossRef]

41. Zhang, Q.; Gao, J.; Dong, H.; Mao, Y. WPD and DE/BBO-RBFNN for solution of rolling bearing fault diagnosis. Neurocomputing 2018, 312, 27–33. [CrossRef]

42. Yang, Z.; Wang, J. A combination forecasting approach applied in multistep wind speed forecasting based on a data processing strategy and an optimized artificial intelligence algorithm. Appl. Energy 2018, 230, 1108–1125. [CrossRef]

43. Azami, H.; Escudero, J. Amplitude-aware permutation entropy: Illustration in spike detection and signal segmentation. Comput. Methods Programs Biomed. 2016, 128, 40–51. [CrossRef]

44. Lai, X.; Li, C.; Zhou, J.; Zhang, N. Multi-objective optimization for guide vane shutting based on MOASA. Renewable Energy 2019, 139, 302–312. [CrossRef]

45. Li, C.; Wang, W.; Chen, D. Multi-objective complementary scheduling of hydro-thermal-RE power system via a multi-objective hybrid grey wolf optimizer. Energy 2019, 171, 241–255. [CrossRef]

46. Liu, Y.; Qin, H.; Mo, L.; Wang, Y.; Chen, D.; Pang, S.; Yin, X. Hierarchical flood operation rules optimization using multi-objective cultured evolutionary algorithm based on decomposition. Water Resour. Manag. 2019, 33, 337–354. [CrossRef]

47. Zhang, C.; Peng, T.; Li, C.; Fu, W.; Xia, X.; Xue, X. Multiobjective optimization of a fractional-order PID controller for pumped turbine governing system using an improved NSGA-III algorithm under multiworking conditions. Complexity 2019, 2019, 5826873. [CrossRef]

48. Wang, W.; Tahir, H.; Ren, Z.; Zhou, K. Robust fault detection and diagnosis for uncertain nonlinear systems. J. Syst. Eng. Electron. 2009, 20, 1031–1036.

49. Yan, J.; Jia, M. Intelligent fault diagnosis of rotating machinery using improved multiscale dispersion entropy and mRMR feature selection. Knowl. Based Syst. 2019, 163, 450–471. [CrossRef]

50. Zheng, J.; Pan, H.; Yang, S.; Cheng, J. Generalized composite multiscale permutation entropy and laplacian score based rolling bearing fault diagnosis. Mech. Syst. Signal Process. 2018, 99, 229–243. [CrossRef]

51. Jiang, L.; Xuan, J.; Shi, T. Feature extraction based on semi-supervised kernel marginal fisher analysis and its application in bearing fault diagnosis. Mech. Syst. Signal Process. 2013, 41, 113–126. [CrossRef]