Research on Fault Diagnosis Algorithm of Ship Electric Propulsion Motor

Fengxin Ma, Liang Qi *, Shuxia Ye, Yuting Chen, Han Xiao and Shankai Li

School of Automation, Jiangsu University of Science and Technology, Zhenjiang 212100, China;
211210301112@stu.just.edu.cn (F.M.)
* Correspondence: alfred_02030210@just.edu.cn

Abstract: The permanent magnet synchronous motor (PMSM) has been used in electric propulsion and other fields. However, it is prone to the stator winding inter-turn short-circuit, and if no effective measures are taken, the ship's power system will be paralyzed. To realize intelligent diagnosis of inter-turn short circuits, this paper proposes an intelligent fault diagnosis method based on improved variational mode decomposition (VMD), multi-scale principal component analysis (PCA) feature extraction, and improved Bi-LSTM. Firstly, the stator current simulation dataset is obtained by using the mathematic model of the inter-turn short-circuit of PMSM, and the parameters of VMD are optimized by the grey wolf algorithm. Then, the data is coarse-grained to obtain multi-scale features, and the main features are selected as the sample data for fault classification by PCA. Subsequently, the Bi-LSTM neural network is used for training and analyzing the data of the sample set and the test set. Finally, the learning rate and the number of hidden-layer nodes of the Bi-LSTM are optimized by the whale algorithm to increase the diagnosis accuracy. Experimental results show that the accuracy of the proposed method for inter-turn short-circuited fault diagnosis is as high as 100%, which confirms the effectiveness of the method.

Keywords: fault diagnosis; permanent magnet synchronous motor; interturn short-circuited; variational mode decomposition; multi-scale; bidirectional long short-term memory neural network

1. Introduction

Ship electric propulsion technology has developed quickly in recent decades. Compared with traditional propeller propulsion, it has the advantages of high reliability, flexible operation, low pollution, low vibration and noise, and high cabin utilization. It is widely used in high-performance ships such as icebreakers, cruise ships, ferries, and engineering ships [1]. With the development of the hardware and software related to electric propulsion technology, there will be a higher proportion of ship electric propulsion technology in ship propulsion. The propulsion motor is a vital part of the ship's electric propulsion system [2]. In the future, the demand for single-unit capacity of electric propulsion ships will increase, and the DC propulsion motor will no longer meet the requirements due to the limited power. In this case, the permanent magnet synchronous motor (PMSM) becomes the ideal choice for ship propulsion motors [3]. The PMSM is widely used because of its excellent performance. However, in practical applications, various faults will occur due to poor working conditions, poor heat dissipation conditions, high environmental humidity, and frequent start and stop [4]. If the fault is not detected in time and no corresponding measures are taken, it may lead to irreversible accidents, causing serious casualties and economic losses. Therefore, fault diagnosis of PMSM has practical engineering significance.

The fault diagnosis of PMSM mainly includes four processes: data preprocessing, feature extraction, feature selection, and fault classification. To solve the problem of low diagnostic accuracy of dynamic eccentricity fault of PMSM, Xue et al. [5] proposed a dynamic eccentricity fault diagnosis method of PMSM. The fast Fourier transform is used

to analyze the stator current signal to obtain the dynamic eccentric state characteristic frequency of the stator current signal and extract the amplitude. Finally, the support vector machine with parameter optimization by a genetic algorithm is used to identify the extracted features. Most of the existing research on mechanical faults of PMSM can only detect a single fault type, and there are few studies on the application of variable speed conditions. Zhao et al. [6] proposed a mechanical fault detection method for PMSM based on improved detrended fluctuation analysis (DFA) and linear discriminant analysis (LDA). The Vold-Kalman is used to track and extract the fault feature components at the fault feature frequency, and the fault feature components are reconstructed. The improved DFA is used to extract the feature of the reconstructed signal. Finally, LDA is conducted to classify the extracted fault features. Aiming at the problem that open-circuit faults in inverters and stators are difficult to be detected, resulting in overstressing healthy switches, large torque ripples, and mechanical vibrations, Huang et al. [7] proposed a multi-break fault diagnosis method for three-phase PMSM based on symmetrical components and DC components. To evaluate the asymmetry and discriminate fault types, the magnitude ratio of the positive sequence to the negative sequence of phase currents is adopted as a fault detection index. At last, the DC components of the phase currents in different reference frames are utilized to locate faults.

In recent years, with the rapid development of deep learning, its powerful feature extraction ability and end-to-end characteristics make it possible to directly extract fault features from the original signal for fault diagnosis tasks, avoiding the influence of artificial experience on feature extraction and effectively overcoming the disadvantages of traditional diagnosis methods [8]. Aiming at the problems of difficulty in capturing inverter faults and poor fault diagnosis accuracy in PMSM, Feng et al. [9] proposed a fusion stacked denoising autoencoder and feedforward neural network (SDA-FFNN) model. Through normalization, six frequency-domain features are integrated into a high-dimensional feature data set, and then the proposed method based on SDA-FFNN is used to complete fault diagnosis. To address the issue that motor-bearing fault diagnosis requires a complex feature extraction process and suffers from a low fault diagnosis recognition rate under various working conditions, Tang et al. [8] proposed a fault diagnosis method based on attention and a multi-scale convolutional neural network. In this method, convolution kernels of different sizes are used to extract multi-scale features of fault signals so that the model can obtain richer and complementary fault feature representations. However, multi-scale feature extraction will lead to unstable time-frequency domain feature results due to mutations at breakpoints. To solve the problem that the frequency-domain characteristics of the stator current change with the switching of motor operating conditions, which leads to the difficulty of diagnosis under multiple operating conditions, Chen et al. [10] proposed a fault diagnosis method with a convolutional neural network (CNN)-based on a phase tracker. The CNN-based phase tracker is used to convert the original signal from the time domain to the angle domain. After the angle-domain resampling, the fundamental frequency of the stator current signal under any working condition will be adjusted to the same frequency so that the fixed frequency-domain fault characteristics can be used for motor fault diagnosis. It is mentioned in this paper that CNN is usually used as a feature extraction model for time series in motor fault diagnosis [11,12].

The stator winding inter-turn short circuit fault is the most common fault of PMSM. If the fault is not detected in time and no measures are taken, the operation with a fault will increase the motor current, which will produce more heat, increase the temperature of the motor, and further destroy the insulation material and permanent magnet material, resulting in a more serious short circuit and loss of excitation [4]. A lot of research has been conducted on the stator winding inter-turn short circuit fault. Mao et al. [13] proposed an initial inter-turn short circuit fault detection method for PMSM based on VMD and logarithmic spectrum analysis. The zero-sequence voltage is taken as the fault feature, and the noise and the related harmonic components in the zero-sequence voltage component are removed by VMD to highlight the fault component. However, in practical applications, it is

difficult to select the decomposition layer and penalty factor of VMD, and it is challenging to obtain the optimal solution of the two parameters. Aiming at the problems of small fault data, unbalanced data distribution, and low data quality of inter-turn short circuit fault data in industrial production environments, Wang et al. [14] proposed an inter-turn short circuit fault diagnosis algorithm based on the stacked sparse autoencoder network. The negative sequence current and electromagnetic torque are used to construct the inter-turn short circuit fault diagnosis data set, and the stacked sparse autoencoder is employed to extract the hidden abstract features in the fault data. Finally, the SoftMax classifier is used to classify the feature signals. Zhang et al. [4] presented a permanent magnet synchronous motor fault diagnosis method based on mixup-LSTM. The fault features in the stator current signal are extracted by wavelet packet decomposition as the sample set, and the sample expansion is realized by mixup. Finally, the LSTM neural network is exploited to realize fault diagnosis (classification). As a variant of the RNN neural network, LSTM can solve the problem of gradient disappearance of the traditional RNN. Although wavelet packet decomposition has good analysis ability, it cannot suppress errors, and more accurate diagnosis results may not be obtained in the subsequent diagnosis process. Moreover, the number of layers and nodes of the hidden layer depends too much on artificial experience, and the optimal solution cannot be obtained.

Given the above problems, this paper proposes a fault diagnosis model of PMSM inter-turn short circuits based on multi-scale space. The original data is denoised by VMD. The maximum product of correlation and kurtosis is taken as the objective function, and the grey wolf (GWO) algorithm [15] is used to optimize the modal decomposition layer and penalty factor of VMD. The decomposition signal is reconstructed by kurtosis, and the noise in the signal is filtered out. The PCA [16] is adopted to select the main features to remove redundant features and reduce dimensionality. The improved coarse-grained multi-scale feature extraction is employed to reduce the instability of the time-frequency domain feature results caused by the mutation at the breakpoint. Moreover, by exploiting the advantages of the whale algorithm (WOA) [17] in optimization, the number of hidden-layer nodes and the learning rate of the Bi-LSTM network are optimized, which effectively reduces the influence of the model parameters on the diagnostic accuracy and reduces the diagnostic error. In practical engineering applications, a new fault diagnosis method is proposed to make fault diagnosis more intelligent. The diagnostic process of this method is shown in Figure 1.

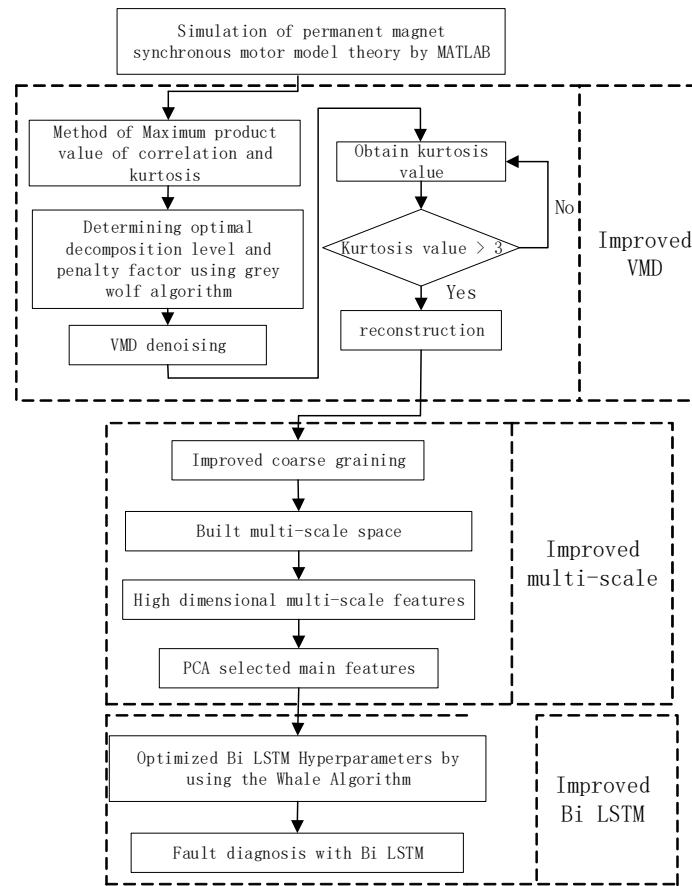



Figure 1. The diagnosis process of the proposed method.

2. Theoretical Background

2.1. Mathematical Model of PMSM under Interturn Short-Circuit Fault

The equivalent diagram of PMSM under the inter-turn short-circuit fault is shown in Figure 2. It is assumed that the inter-turn short-circuit fault occurs on the a-phase stator winding of the PMSM. In the figure, a short circuit is added to the a-phase winding of the motor. The short-circuit resistance R_f divides the a-phase winding into two parts: a_1 and a_2, where a_1 represents the normal part, a_2 represents the fault part, and the current i_f is the short-circuit current flowing through the resistance R_f.

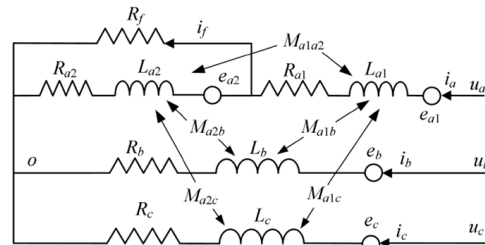



Figure 2. PMSM equivalent model of a phase inter-turn short-circuited fault.

The short-circuit turn ratio is defined as:

$$u = \frac{n}{N} \tag{1}$$

where n indicates the number of short-circuit turns for the phase stator winding, and N indicates the total number of turns of a phase stator winding.

In the case of an inter-turn short-circuit fault, the voltage equation of PMSM can be expressed as:

$$V_{sf} = R_{sf}i_{sf} + L_{sf}\frac{di_{sf}}{dt} + e_{sf} + V_0 \tag{2}$$

V_{sf} is defined as:

$$V_{sf} = \begin{bmatrix} R_{a1} + R_{a2} & 0 & 0 & -R_{a2} \\ 0 & R_b & 0 & 0 \\ 0 & 0 & R_c & 0 \\ R_{a2} & 0 & 0 & -u(R_{a1} + R_{a2}) - R_f \end{bmatrix} \tag{3}$$

where V_{sf} denotes the resistance matrix; R_f denotes the fault resistance, R_{a1}, R_{a2}, R_b, and R_c denote the resistance of stator windings a_1, a_2, b, and c, respectively; $i_{sf} = \begin{bmatrix} i_a & i_b & i_c & i_d \end{bmatrix}^T$ denotes the current matrix; i_a, i_b, and i_c denote the stator winding currents of phases a, b, and c, respectively; i_f denotes the short-circuit current.

The inductance matrix can be expressed as:

$$L_{sf} = \begin{bmatrix} L_{a1} + L_{a2} + 2M_{a1a2} & M_{a1b} + M_{a2b} & M_{a1c} + M_{a2c} & -L_{a2} - M_{a1a2} \\ M_{a1b} + M_{a2b} & L_b & M_{bc} & -M_{a2b} \\ M_{a1c} + M_{a2c} & M_{bc} & L_c & -M_{a2c} \\ L_{a2} + M_{a1a2} & M_{a2b} & M_{a2c} & -L_{a2} \end{bmatrix} \tag{4}$$

where L_{a1}, L_{a2}, L_b, and L_c denote the self-inductance of stator windings a_1, a_2, b, and c, respectively; $M_{j,k}$ indicates the mutual induction between stator windings j and k ($j \in \{a_1, a_2, b, c\}$ and $k \in \{a_1, a_2, b, c\}$).

The counter electromotive force matrix of the three-phase stator winding and short-circuit winding can be expressed as:

$$e_{sf} = \begin{bmatrix} e_a \\ e_b \\ e_c \\ e_d \end{bmatrix} = \frac{d}{dt} \begin{bmatrix} \lambda_{PM,a} \\ \lambda_{PM,b} \\ \lambda_{PM,c} \\ \lambda_{\lambda_{PM,f}} \end{bmatrix} \tag{5}$$

$$\begin{cases} \lambda_{PM,a} = \lambda_{PM,1}cos(\theta) + \sum_{v=2k+1}\lambda_{PM,v}cos(v\theta - \theta_v) \\ \lambda_{PM,b} = \lambda_{PM,1}cos(\theta - \frac{2\Pi}{3}) + \sum_{v=2k+1}\lambda_{PM,v}cos(v\theta - \theta_v\frac{2\Pi}{3}) \\ \lambda_{PM,b} = \lambda_{PM,1}cos(\theta + \frac{2\Pi}{3}) + \sum_{v=2k+1}\lambda_{PM,v}cos(v\theta - \theta_v + v\frac{2\Pi}{3}) \\ \lambda_{PM,b} = u\lambda_{PM,a} \end{cases} \tag{6}$$

where $k \in Z$, $\lambda_{PM,1}$ denotes the fundamental amplitude of the flux linkage; $\lambda_{PM,v}$ denotes the vth harmonic amplitude of the flux linkage; $\lambda_{PM,f}$ denotes the flux linkage of the short-circuit winding; θ denotes the rotor electrical angle; θ_v denotes the angle between the fundamental and the vth harmonic of the flux linkage; $V_0 = \begin{bmatrix} V_0 & V_0 & V_0 & 0 \end{bmatrix}^T$ denotes the zero-sequence voltage.

The three-phase winding resistance, self-inductance, and mutual inductance of the normal PMSM are equal. The relationship between the short-circuit turn ratio and the resistance, inductance, and back electromotive force in the stator winding can be expressed as:

$$\begin{cases} R_{a1} = (1-u)R_a \\ R_{a2} = uR_a \\ L_{a1} = (1-u)^2L_a \\ L_{a2} = u^2L_a \end{cases} \tag{7}$$

$$\begin{cases} M_{a1a2} = u(1-u)L_a \\ M_{a2b} = (1-u)M_{ab} \\ M_{a2c} = (1-u)M_{ac} \\ \quad M_{a2b} = uM_{ab} \\ \quad M_{a2c} = uM_{ac} \end{cases} \tag{8}$$

$$\begin{cases} \lambda_{PM,f} = u\lambda_{Pm,a} \\ \quad e_f = ue_a \end{cases} \tag{9}$$

When the PMSM suffers from an inter-turn short-circuit fault, its electromagnetic torque can be expressed as:

$$T_e = \frac{e_a i_a + e_b i_b + e_c i_c - e_f i_f}{\omega_r} \tag{10}$$

where w_r denotes the mechanical angular velocity of the rotor of the PMSM, w_e denotes the electrical angular velocity of the rotor, and n_p denotes the pole pairs of the PMSM.

2.2. Variational Mode Decomposition

VMD is an adaptable signal processing method based on Wiener filtering with VMD, the input signal $f_{(t)}$ is finally resolved into multiple modal component signals u_k; the unilateral spectrum of each mode u_k is obtained by the Hilbert transform, and each mode modulates its corresponding spectrum to the corresponding baseband by adding an exponential term. A confined variational problem is defined as follows:

$$min\left\{ \sum_k \left\| \partial_t [(\delta(t) + \frac{j}{\Pi t}) \cdot u_k(t)] e^{-jw_k t} \right\|_2^2 \right\} \tag{11}$$

$$s.t. \sum_k u_k = f \tag{12}$$

By using the quadratic punishment factor and the Lagrange multiplier operator $\lambda_{(t)}$, the confined variational problem is transformed into an unconfined variational problem. Then, it is solved by the alternating direction algorithm of multipliers and converted to the frequency domain by Fourier isometric transformation. Finally, the updated expressions of u_k^{n+1} and w_k^{n+1} are obtained:

$$\hat{u}_k^{n+1}(w) = \frac{\hat{f}(w) - \sum_{i \neq k} \hat{u}_i(w) + (\frac{\hat{\lambda}(w)}{2})}{1 + 2\alpha(w - w_k)^2} \tag{13}$$

$$w_k^{n+1} = \frac{\int_0^\infty w |\hat{u}_i(w) + |^2 dw}{\int_0^\infty |\hat{u}_k(w)|^2 dw} \tag{14}$$

2.3. Constructing Multi-Scale Features

The features extracted by traditional methods cannot efficiently render the state features of the signal, which affects the diagnostic accuracy. Considering the non-stationarity of the signal, the features at a single scale may overlap in the feature space, and it is difficult to fully reveal the signal characteristics. By referring to the multi-scale concept and considering the characteristics of the signal at multiple scales to ensure the overall integrity and local details of the signal, this paper proposes a multi-scale feature index that can reveal the performance of the signal in multi-scale space. Meanwhile, based on the traditional time-domain index, combined with the advantages of multi-scale space on feature space overlap and signal cross-scale complexity, this paper constructs a multi-scale index as the basis for fault classification. Firstly, the time-domain signal is coarse-grained, which makes it difficult to obtain its distribution at different scales. For the time-domain

sequence $x = \{x_1, x_2, \cdots, x_n\}$, a signal segmentation of length τ is performed, where the *jth* signal segment can be expressed as:

$$\left\{x_{p,(j-1)\tau+1}, x_{p,(j-1)\tau+2}, \cdots, x_{p,j\tau}, \cdots\right\}\left(1 \le j \le \frac{n}{\tau}\right) \tag{15}$$

The corresponding coarse-grained sequence $y^{(\tau)}$ can be expressed as:

$$\begin{cases} y^{(\tau)} = \left\{y_1^{(\tau)}, y_2^{(\tau)}, \cdots, y_j^{(\tau)}, \cdots\right\} \\ y_j^{\tau} = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau} x_i \end{cases}, 1 \le j \le \frac{n}{\tau} \tag{16}$$

where τ is the scale factor. By extracting the time domain features of the coarse-grained sequence $y^{(\tau)}$ of different scales, the required multi-scale features can be obtained. Since the multi-scale sequence $y^{(1)}$ is the original time-domain sequence x when $\tau = 1$, the calculated result is the traditional time-domain feature.

2.4. Principal Component Analysis

The principle of PCA is re-linearly combining the high-dimensional variable indexes in the original fault data to construct a new variable index with a lower dimension. These low-dimensional variable indicators are orthogonal and unrelated to each other. Based on this, the coordinate projection of the data sample on the vector is determined by the difference in the independent variables between the samples. The differences reflected by other vectors are reduced in turn, and these vectors are called main components.

The steps of PCA feature extraction are as follows:

Step 1: Given an input signal matrix $X = \{x_{ij} : x_{ij} \in R^{n \times m}\}$, where each row $x_i \in R^{1 \times m}, i = 1, \cdots, n$ of the matrix indicates an index and each column $x_j \in R^{n \times 1}, j = 1, \cdots, m$ of the matrix represents a sample. The average value of the sample x_j can be represented as:

$$\bar{x}_j = \frac{1}{m}\sum_{j=1}^{m} x_j \tag{17}$$

Step 2: Let C_o be the covariance matrix. The eigenvalues $\lambda_i, i = 1, \cdots, n$ and eigenvectors $d_i, i = 1, \cdots, n$ of the covariance matrix C_o are calculated.

Step 3: The eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$ are arranged in decreasing order, and the eigenvectors corresponding to the eigenvalues are $d_i, i = 1, \cdots, n$. The accumulative contribution rate ς of the first r principal components can be expressed as:

$$\varsigma = \sum_{i=1}^{r} \lambda_i / \sum_{i=1}^{n} \lambda_i \tag{18}$$

Step 4: If $\varsigma \ge 0.80$, construct a matrix $E \in R^{n \times r}$ composed of the corresponding feature vectors. Then, the new sample matrix X' can use $E = (d_1, d_2, \cdots, d_r)$ to map the original bearing vibration signal matrix to a new space

$$X' = E^T X \tag{19}$$

where $X' \in R^{r \times m}$.

2.5. Long Short-Term Memory Neural Network

The neurons of the LSTM model have a cell state and three gate mechanisms. As shown in Figure 3, the LSTM model has three gates: the forgetting gate, the updating gate, and the output gate, through which the cell state is protected and controlled.

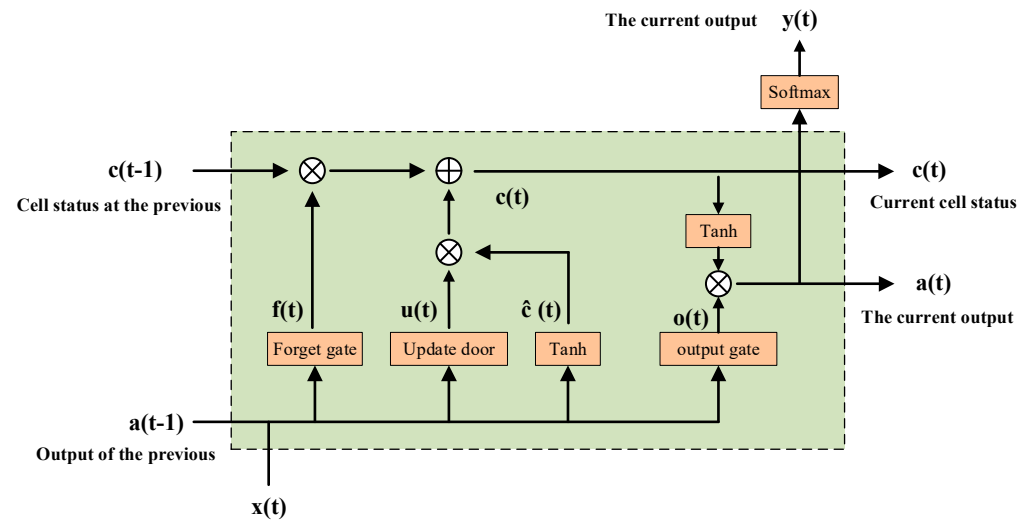



Figure 3. The structure of the LSTM model.

Forgetting door: The sigmoid function determines which information message is discarded, and it can be expressed as:

$$f(t) = \sigma\left(W_f[a(t-1), x(t)] + b_f\right) \tag{20}$$

where $f(t)$ is the quantity outgoing to each cell in state $c(t-1)$; $a(t-1)$ is the outgoing at time $t-1$; $x(t)$ is the intromission at time t; σ is a sigmoid function, $\sigma(x) = (1 + e^{-x})^{-1}$; W_f is the weight of each variable, and b_f is the bias term.

Update door: Update storage information message. First, the sigmoid function calculates the result $u(t)$ to determine which information message to update. Then, a new candidate value vector $\hat{c}(t)$ is generated according to the Tanh function and added to the cell state. By multiplying the old cell state with the forgetting gate $f(t)$, part of the old information message is forgotten, $u(t) * \hat{c}(t)$ is added, and the current cell state is updated. It can be expressed as:

$$u(t) = \sigma(W_u[a(t-1), x(t)] + b_u) \tag{21}$$

$$\hat{c}(t) = tanh(W_c[a(t-1), x(t)] + b_c) \tag{22}$$

$$c(t) = u(t) \odot \hat{c}(t) + f(t) \odot c(t-1) \tag{23}$$

where $u(t) \in [0, 1]$; $tanh$ outputs a value between -1 and 1; $c(t-1)$ is the cell status value at time $t-1$; $\hat{c}(t)$ indicates the information message to be extracted from the intromission information at time t, and $c(t)$ indicates the updated cell status value.

Output gate: Output information message based on the cell status. The sigmoid function determines the quantity of output information. The $tanh$ function processes $c(t)$ to obtain a value between -1 and 1. $o(t)$ and $c(t)$ are multiplied to obtain the output value at time t, which can be expressed as:

$$o(t) = \sigma(W_o[a(t-1), x(t)] + b_o) \tag{24}$$

$$a(t) = o(t) \odot tanh(c(t)) \tag{25}$$

Structurally, compared with the one-way LSTM network, the Bi-LSTM network adopts a two-way loop structure of forward and backward propagation. Viewing the problem from the flow of time, Bi-LSTM adds the data flow from the future into the past based on the one-way flow of LSTM data from the past into the future, and the hidden layer for the past and the hidden layers for the future are independent of each other, so LSTM can better explore the time series feature of data.

The structure of the Bi-LSTM network is shown in Figure 4.

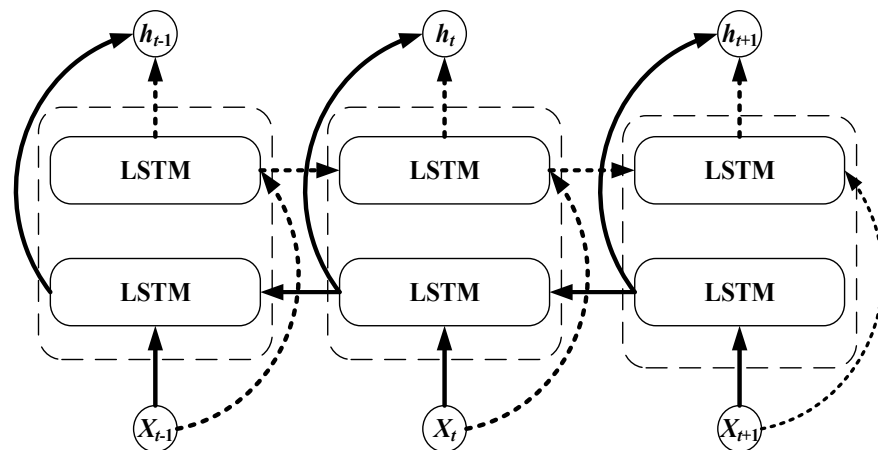



Figure 4. The structure of the Bi-LSTM network.

Forward propagation has two steps: (1) move from left to right and calculate from the original time step to the ultimate time step; move from right to left and calculate from the last time step to the ultimate time step. At time step *t*, the forward layer and the backward layer in the Bi-LSTM use the same input X_t, and they jointly generate the output and state. However, the two LSTM network layers do not share the bias vector parameters and weight matrix. The formula can be expressed as:

$$h_t^1 = f\left(W^1 \cdot X_t + U^1 \cdot h_{t-1}\right) \tag{26}$$

$$h_t^2 = f\left(W^2 \cdot X_t + U^2 \cdot h_{t+1}\right) \tag{27}$$

$$Y_t = soft\,max\left(V^1 h_t^1 + V^2 h_t^2\right) \tag{28}$$

where f is the hidden-layer activation function; W^1 and W^2 are weight matrix parameters from the intromission layer to the hidden layer; U^1 and U^2 are weight parameters from the hidden layer to the hidden layer, V^1 and V^2 are weight parameters from the hidden layer to the outgoing layer. The weight matrix and bias vector parameters of Bi-LSTM are adjusted by backpropagation according to the error function output by Y_t theory.

3. The Proposed Method
3.1. Model Optimization of Variational Mode Decomposition

The decomposition order of VMD is determined by the *K* value. A too-large *K* value will cause signal decomposition fault, while a too-small *K* value will lead to incomplete signal decomposition or signal frequency aliasing. The punishment factor *α* is a key parameter for converting the confined variational problem into an unconfined problem, and it mainly affects the bandwidth of the IMF component obtained by VMD decomposition and the convergence rate of the algorithm. The punishment factor is inversely proportional to the bandwidth of the IMF component. This paper uses the GWO algorithm to optimize *K* and *α*.

The GWO algorithm is advanced by observing the grey wolf population's preying strategy in nature. In the grey wolf group, the grey wolf is classified into four classes from high to low according to the fitness value, which are *α* wolf, *β* wolf, *δ* wolf, and *ω* wolf. Among them, the *α* wolf is the optimum result, the *β* wolf and *δ* wolf are the second and third optimum results, and the *ω* wolf is the candidate result. The use of the grey wolf pack predation mechanism to solve the target solution space involves three steps: tracking, encircling, and attacking prey, and they are mapped to the solution to the mathematical model problem, i.e., the fitness from good to bad corresponds to the *α* wolf, *β* wolf, and

δ wolf in the wolf population, and the other fitness individuals correspond to the ω wolf. The specific algorithm applied to the mathematical model can be represented as:

$$D = \left| C * X_p - X_{(t)} \right| \tag{29}$$

$$X_{(t+1)} = X_p(t) - A * D \tag{30}$$

where D indicates the relative location between the grey wolf and the target; t is the current number of iterations; C and A are coefficients, X_p is the location of the target, and $X_{(t)}$ is the location of the current gray wolf. The values of coefficients C and A are determined by:

$$C = 2r_1 \tag{31}$$

$$A = 2ar_2 - a \tag{32}$$

$$a = 2 - 2 * \frac{t}{T_{max}} \tag{33}$$

where r_1 and r_2 are random numbers, the value of a linearly reduces from 2 to 0 with the number of iterations t, and T_{max} is the maximum number of iterations.

After the gray wolf population completes the positioning and tracking of the prey location, it needs to complete the enclosure of the prey and map it to the mathematical model. The location of ω can be updated by calculating the locations of α, β, and δ. The update formula is represented as:

$$\begin{cases} D_\alpha = \left| C_1 \cdot X_\alpha - X \right| \\ D_\beta = \left| C_2 \cdot X_\beta - X \right| \\ D_\delta = \left| C_3 \cdot X_\delta - X \right| \end{cases} \tag{34}$$

where D_α, D_β, and D_δ denote the distance between α, β, and δ and the current iterated grey wolf individual; X_α, X_β, and X_δ indicate the current location of α, β, and δ, respectively; C_1, C_2, and C_3 are the corresponding coefficients; X denotes the location of the gray wolf individual in the current iteration.

When the prey stops moving, the wolf completes the hunt by attacking. The location of a grey wolf individual can be determined by:

$$\begin{cases} X_1 = X_\alpha - A_1 \cdot D_\alpha \\ X_2 = X_\beta - A_2 \cdot D_\beta \\ X_3 = X_\delta - A_3 \cdot D_\delta \end{cases} \tag{35}$$

$$X_{(t+1)} = \frac{X_1 + X_2 + X_3}{3} \tag{36}$$

where X_1, X_2, and X_3 indicate the step length and direction of ω wolves forward to α, β, and δ, respectively; $X_{(t+1)}$ indicates the final location of ω wolves after the iteration.

The maximum value of the product of Pearson correlation p and kurtosis value K_v is taken as the optimized objective function value. The specific formula is presented below:

$$K_v = \frac{\sum_{i=1}^{N} \left(|X_i| - \bar{X} \right)^4}{\partial^4} \tag{37}$$

$$f_{max} = P \cdot K_v \tag{38}$$

3.2. Multi-Scale Feature Optimization Based on Improved Coarse-Graining

To improve the capability of traditional multi-scale arrangement features, this paper starts with the coarse-grained step and establishes an improved multi-scale feature. Taking scale 3 as an example (as shown in Figure 5), different from the original algorithm, under the same time scale τ, our method can obtain τ sets of time series after coarse-grained

processing to solve the mutation problem at the 'breakpoint', while there is only one set of time series in the classical algorithm. The mathematical process can be expressed as:

1. The given signal is processed by the improved coarse-grained process to obtain τ sets of time series:

$$\begin{cases} Z_i^{(\tau)} = \left\{ y_{i,1}^{(\tau)}, y_{i,2}^{(\tau)}, \cdots \right\} \\ y_{i,j}^{(\tau)} = \frac{\sum_{f=0}^{\tau-1} x_{f+i+\tau(j-1)}}{\tau} \end{cases}, i = 1, 2, \cdots, \tau \qquad (39)$$

2. For each set of new coarse-grained time series $Z_i^{(\tau)} | (i = 1, 2, \cdots, \tau)$, its characteristic value in time and frequency domains are obtained, and then the average value of τ time series eigenvalues is calculated to obtain the eigenvalues under the time scale τ.

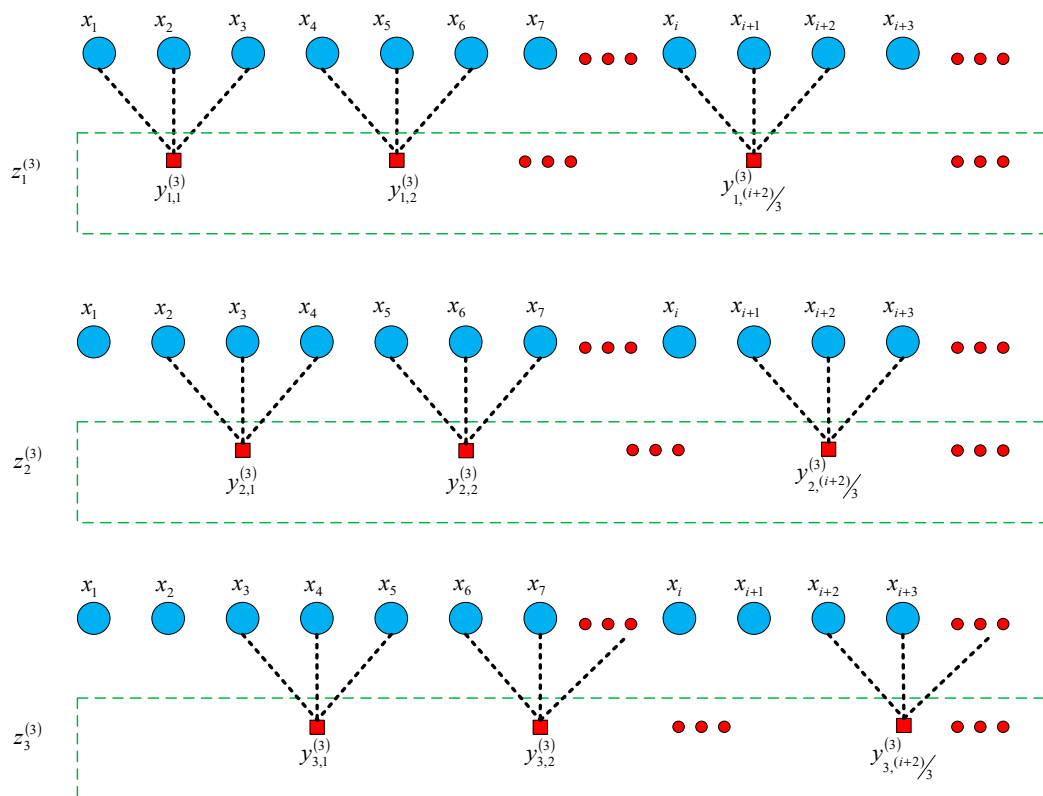



Figure 5. The time-frequency characteristic method at a scale of 3.

3.3. Parameter Optimization of Long Short-Term Memory Neural Network

In the process of model training, the optimal set of hyperparameters can be selected to improve the model diagnosis accuracy and performance. The hyperparameters include the number of hidden layers, the learning rate, and the number of hidden layer neurons in the Bi-LSTM prediction model. The hyper-parameter selection of the Bi-LSTM prediction model depends on manual training and is adjusted multiple times according to experience. However, the training of the Bi-LSTM prediction model requires a certain amount of time, which will inevitably decline the work efficiency. Moreover, the hyper-parameter combination of the prediction model selected by this method is not optimal, which will affect the accuracy and effectiveness of the prediction model. To address this issue, the whale algorithm is employed to optimize the hidden layer neural network and learning rate of the Bi-LSTM fault model to enhance the generalization capability and estimation accuracy of the model.

The whale algorithm is a mathematical model developed by based imitating the behaviors of humpback whales to surround, dabble, and search for prey. It has the advantages

of less parameter adjustment, simple operation, and a strong ability to jump out of local optima.

The first stage is:

$$D = |C \cdot X^*(t) - X(t)| \tag{40}$$

$$X(t+1) = X^*(t) - A \cdot D \tag{41}$$

where D denotes the distance vector between the current optimal solution and the search body; t denotes the current iteration times; X^* denotes the location vector of the current optimal solution; X denotes the location vector of the search volume.

The second stage is:

$$X(t+1) = D' \cdot e^{bl} \cdot cos(2\Pi l) + X^*(t) \tag{42}$$

where $D' = |X^*(t) - X(t)|$ denotes the distance vector between the searching individual and the target prey; b denotes the constant coefficient of logarithmic spiral shape; l denotes a random number in the interval $[-1, 1]$.

The third stage is:

$$D = |C \cdot X_{rand} - X(t)| \tag{43}$$

$$X(t+1) = X_{rand} - A \cdot D \tag{44}$$

where X_{rand} denotes the location of a random search body in the current population.

The flowchart of using the whale algorithm to optimize the hyperparameters of the Bi-LSTM diagnostic model is shown in Figure 6.

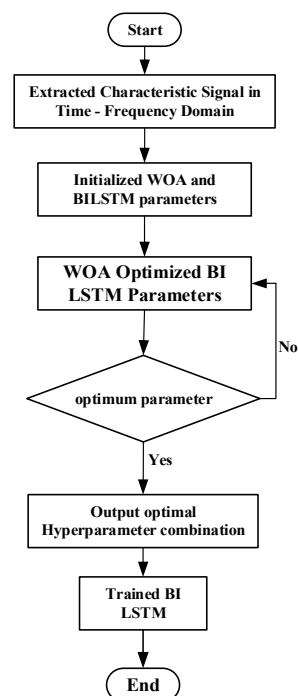



Figure 6. The flowchart of using the whale algorithm to optimize the hyperparameters of the Bi-LSTM diagnostic model.

4. Experiments and Results

Firstly, the experimental data are obtained by the Simulink simulation program in MATLAB, and the collected data are sorted out. Secondly, the effectiveness of the improved method is verified by ablation experiments and comparative analysis.

4.1. Data Acquisition

Through the Simulink in MATLAB, it is simulated that the inter-turn short-circuit fault occurs in the a-phase stator winding after the PMSM runs normally for 0.5 s. The simulation model is shown in Figure 7. The stator current waveform is shown in Figure 8. Figure 8a presents the stator current simulation diagram of the permanent magnet synchronous motor from start-up to normal operation to the occurrence of inter-turn short circuit fault. Figure 8b presents the local amplification diagram of 0.4 s before the period in Figure 8a. Figure 8c is the local amplification diagram of 1 s after the period in Figure 8a. It can be seen from Figure 8b that under normal circumstances, the three-phase current is a periodic waveform. In this paper, phase a stator current is selected as the fault feature to detect the occurrence of a fault.

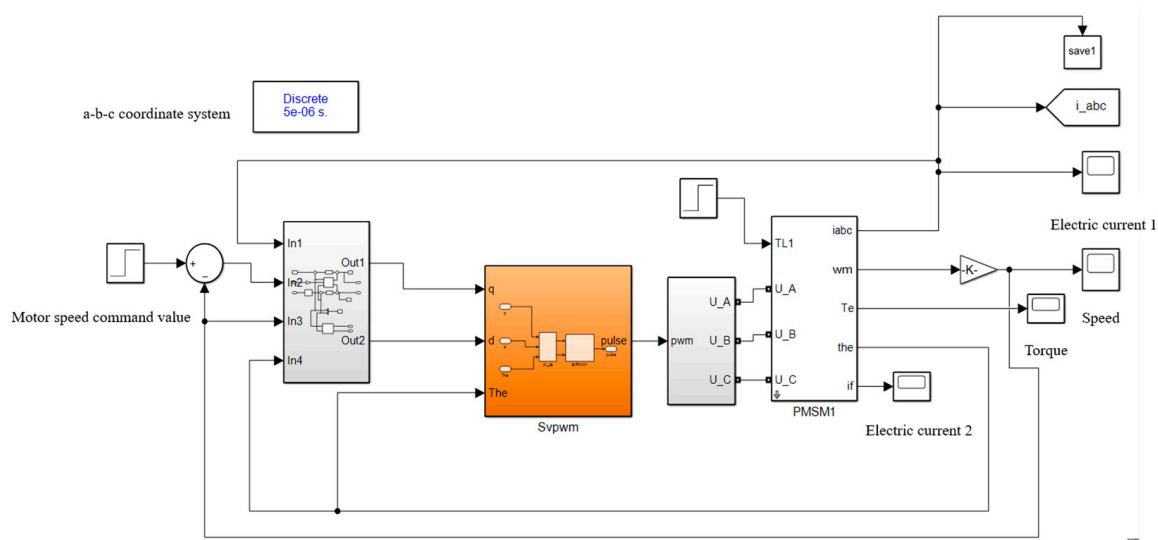



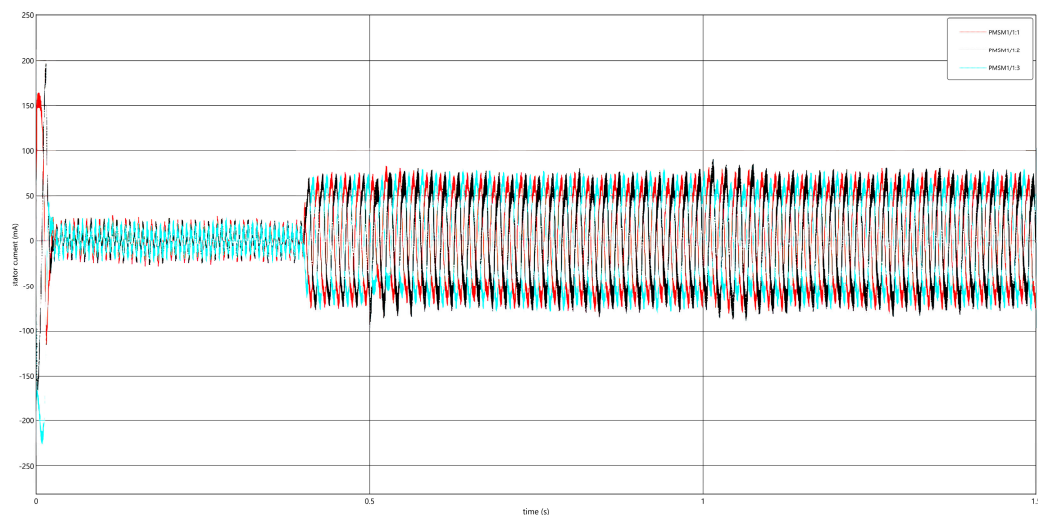
Figure 7. The simulation diagram of the permanent magnet synchronous motor a phase stator winding inter-turn short-circuit.

(**a**)

Figure 8. *Cont.*

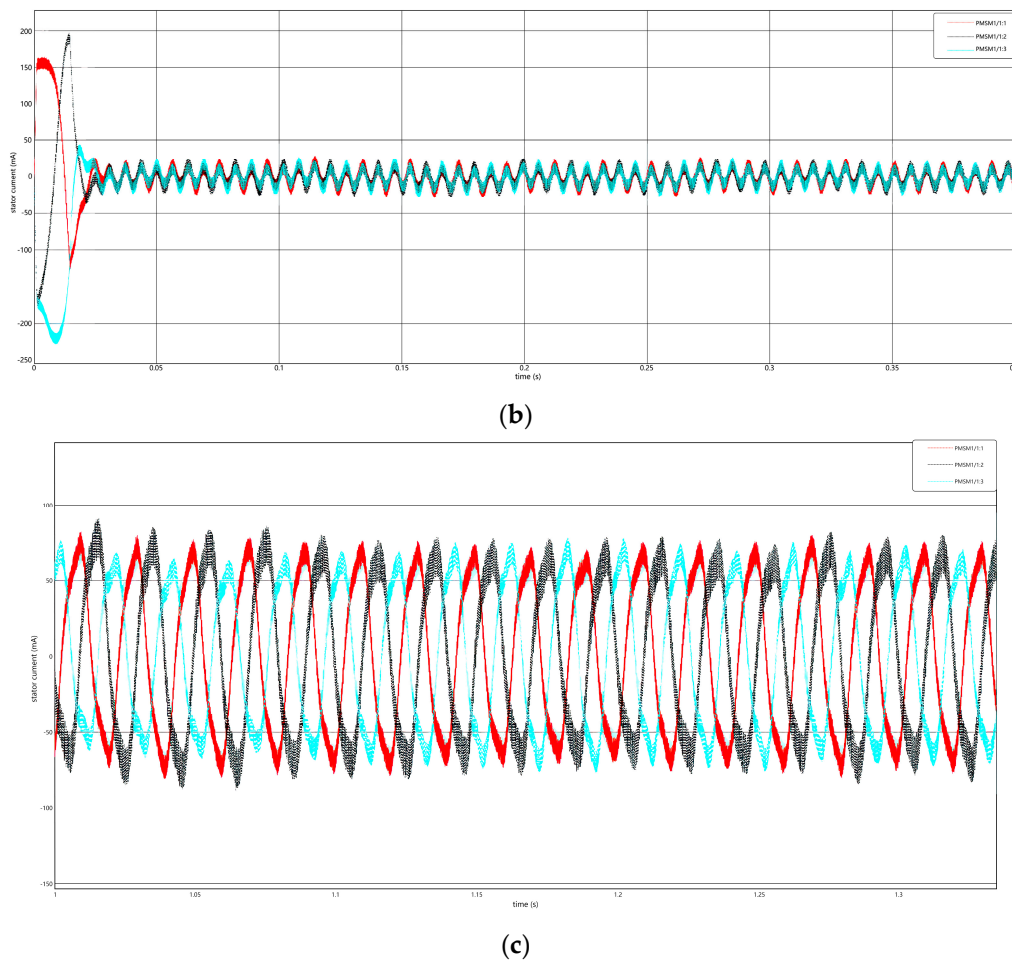



Figure 8. (**a**) The waveform of the stator current signal. (**b**) The stator current. (**c**) The simulation of the stator current with a short-circuit turn ratio of 0.2.

Under normal circumstances, there is no short-circuit current in the PMSM, so its value is zero; after an inter-turn short-circuit fault occurs, a short-circuit loop is generated, thereby generating a short-circuited current. The above simulation results show that the stator current of the PMSM changes significantly after the inter-turn short-circuit fault occurs.

4.2. Ablation Experiments

4.2.1. The Effectiveness of the VMD Algorithm Based on GWO Optimization

This paper proposes a VMD decomposition method based on GWO optimization to remove noise in the signal. By comparing the results of inter-turn short circuit fault diagnosis using the improved VMD decomposition method with those using the traditional VMD decomposition method, the effectiveness of the improved algorithm is verified. It should be noted that the feature extraction part adopts the unimproved multi-scale feature extraction method, and the fault diagnosis part adopts the unimproved Bi-LSTM neural network model.

Since no failure occurs in the first 0.5 s of the dataset, it was deleted. Then, 300 sets of data are collected for each of the five faults, with 2000 data in each group. The sample data of the training dataset and the test dataset are obtained by dividing each fault sample by a ratio of 5:1. For the VMD decomposition method based on GWO optimization, the initial population of the GWO algorithm is set to 20, and the maximum number of iterations is 10. The decomposition number and punishment factors are selected in the ranges of [2, 12] and [1000, 3000]. Finally, the optimized IMF decomposition number $K = 6$, and the punishment factor $\alpha = 1564$.

The training set is tested by the trained model, and then the results are compared with the original results. Finally, the ratio of the predicted correct value to the total number is calculated to obtain the correct rate of the training set, as shown in Figure 9. Figure 9a presents the result of using the unimproved VMD decomposition method, and Figure 9b presents the result of using the improved VMD decomposition method. It can be seen from the figure, the accuracy of the model before VMD improvement becomes stable after 1200 iterations, and the accuracy of the model after VMD improvement becomes stable after 1000 iterations.

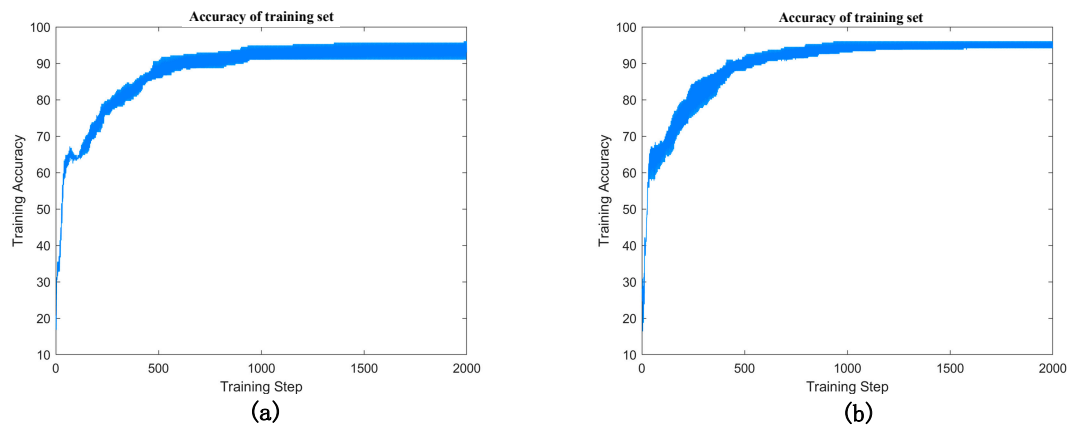



Figure 9. The accuracy in the training process. (**a**) The accuracy in the training process by unimproved VMD. (**b**) The accuracy in the training process by improved VMD.

Figure 10 shows the confusion matrix of the fault classification model on the test set in the experiment. Figure 10a presents the confusion matrix of using the unimproved VMD decomposition method, and Figure 10b presents the confusion matrix of using the improved VMD decomposition method. It can be seen from Figure 10 that for fault category 2, both the unimproved and improved models can make a correct diagnosis. For fault category 1, both the unimproved and the improved VMD make three diagnostic errors. For fault category 3, both the unimproved and improved VMD make one diagnostic error. For fault category 4, the unimproved VMD makes four diagnostic errors, while the improved VMD makes three diagnostic errors. For fault category 5, the unimproved VMD makes nine diagnostic errors, while when the improved VMD makes six diagnostic errors. The final fault identification accuracy of the two methods is 93.2% and 94.8% respectively.

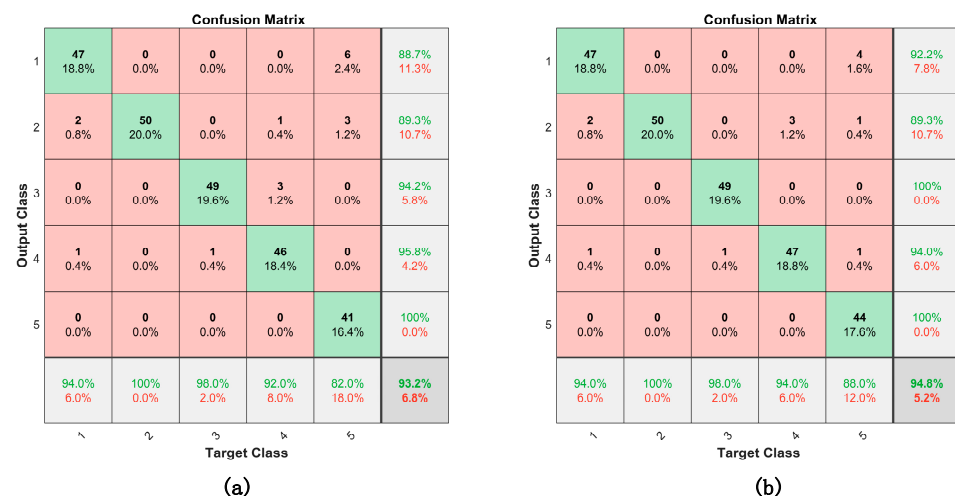



Figure 10. The confusion matrix. (**a**) The confusion matrix in the training process by the unimproved VMD. (**b**) The confusion matrix in the training process by the improved VMD.

The t-SNE method is used to obtain the 2D scatter plot, which can demonstrate the fault classification effect more intuitively, as shown in Figure 11. Figure 11a presents the 2D scatter plot using the unimproved VMD decomposition method, and Figure 11b presents the 2D scatter plot using the improved VMD decomposition method. The results in Figure 11 indicate that the distance between and within the VMD classes before and after the improvement is not very different, and the effect is not good.

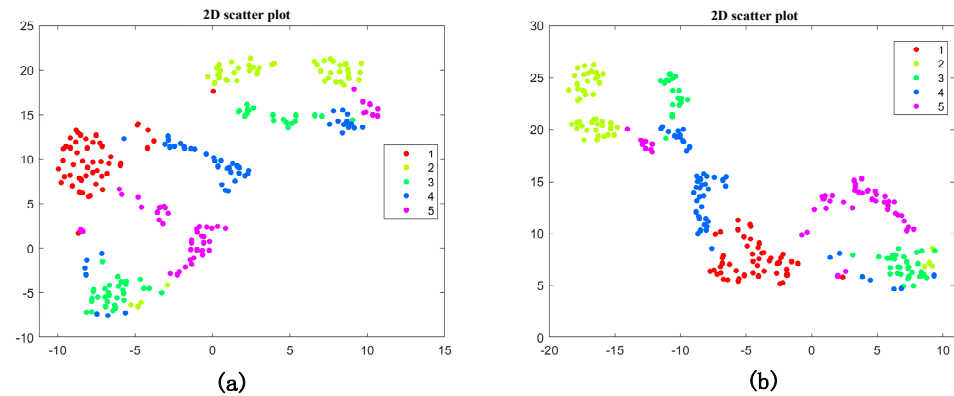



Figure 11. The 2D scatter plot. (**a**) The 2D scatter plot in the training process by the unimproved VMD. (**b**) The 2D scatter plot in the training process by the improved VMD.

4.2.2. The Effectiveness of the Improved Coarse-Grained Multi-Scale Feature Extraction Method

In this paper, an improved coarse-grained multi-scale feature extraction method is proposed. The reconstructed signal is processed by improved coarse-grained processing at different scales. The effectiveness of the improved algorithm is verified by comparing the results of inter-turn short circuit fault diagnosis using the improved multi-scale feature extraction method with those of using the traditional multi-scale feature extraction method. It should be noted that the data preprocessing part adopts the improved VMD decomposition method, and the fault diagnosis part adopts the unimproved Bi-LSTM neural network model.

After the decomposed signal is obtained by the improved VMD decomposition method, the decomposed signal is reconstructed. The decomposed signal is reconstructed by kurtosis value, and the decomposed components with more impact information can be screened out. Table 1 shows the kurtosis values of six signals decomposed by VMD from three samples for each of the five faults. In reference [18], the decomposition signal with a kurtosis value greater than 3 is superimposed and reconstructed. The reconstructed signal is subjected to different scales of improved coarse-grained processing to obtain its distribution at different scales of 1–10. For each new set of coarse-grained time series, its time-frequency domain eigenvalues are obtained, and then the average value of the eigenvalues of τ time series is calculated. The 19 time-frequency domain eigenvalues under the time scale τ can be obtained. There are 14 time domains, which are the maximum value, minimum value, average value, peak-peak value, average value of absolute value, variance, standard deviation, kurtosis, skewness, root mean square, waveform factor, peak factor, pulse factor, and margin factor. There are five frequency domains, which are the center of gravity frequency, mean square frequency, root mean square frequency, frequency variance, and frequency standard deviation. Each scale has 19 features, and there are 190 features in 10 scales. The PCA was used to extract the main features and normalize them by the Map minmax function.

Figures 12–14 show the training set accuracy, confusion matrix, and 2D scatter plot obtained by using the improved multi-scale feature extraction method for inter-turn short circuit fault diagnosis.

Table 1. The kurtosis values of five faults after VMD.

Failure Mode	Sample Label	Kurtosis Value after VMD					
	S1	6.24	1.52	25.71	1.53	23.65	8.96
1	S2	3.18	3.38	1.52	12.87	1.51	23.94
	S3	3.42	3.08	1.52	12.38	1.52	16.32
	S4	3.05	3.40	1.52	12.13	1.51	29.16
2	S5	2.89	1.52	22.57	1.54	24.08	13.42
	S6	2.98	1.53	20.10	1.54	20.91	16.10
	S7	3.03	3.39	1.53	11.95	1.52	31.91
3	S8	3.44	2.98	1.53	11.97	1.52	38.54
	S9	3.46	2.81	1.53	11.44	1.52	30.73
	S10	3.57	1.53	19.02	1.55	15.55	21.01
4	S11	3.18	1.53	17.95	1.55	16.87	19.15
	S12	3.02	3.36	1.53	12.25	1.52	21.24
	S13	5.20	1.52	52.68	9.03	1.54	21.00
5	S14	6.42	1.53	48.69	9.23	1.54	20.98
	S15	2.38	1.53	54.71	9.05	1.54	21.98

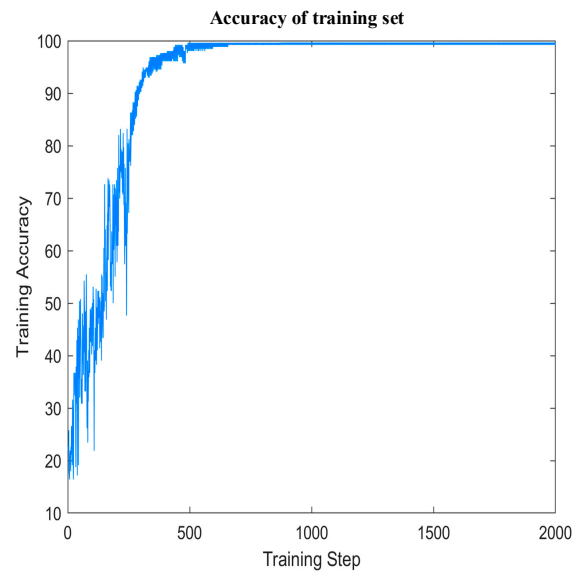



Figure 12. The accuracy of the training set using the improved multiscale feature extraction.

Figure 13. The confusion matrix of using the improved multiscale feature extraction.

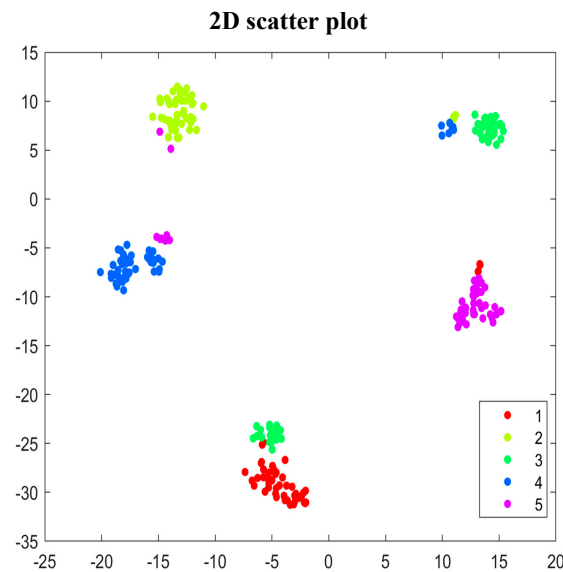



Figure 14. The 2D scatter plot using the improved multiscale feature extraction.

Comparing Figure 12 with Figure 11b, it can be seen that the accuracy of the unimproved ordinary time-frequency domain feature extraction model tends to be stable after 1000 iterations, and the accuracy of the improved multi-scale time-frequency domain feature extraction model tends to be stable after 800 iterations. Comparing Figure 13 with Figure 12b, it can be seen that for fault category 2, the model can make a correct diagnosis. For fault category 1, the model based on ordinary time-frequency domain feature extraction makes three diagnostic errors, and the model based on improved multi-scale feature extraction makes two diagnostic errors. For fault category 3, the model based on ordinary time-frequency domain feature extraction makes one diagnostic error, and there is no error after using the improved multi-scale feature extraction. For fault category 4, both methods make three diagnostic errors. For fault category 5, the model based on ordinary time-frequency domain feature extraction makes six diagnostic errors, and the model based on multi-scale feature extraction makes four diagnostic errors. The final fault identification accuracy of the two methods is 94.8% and 96.4%, respectively. By comparing Figure 14 with Figure 13b, it can be seen that the improved multi-scale time-frequency domain feature extraction model has a larger inter-class distance and smaller intra-class distance, which does not achieve the desired effect.

4.2.3. The Effectiveness of the Fault Diagnosis Model Based on the Improved Bi-LSTM Neural Network

In this paper, a Bi-LSTM neural network model based on WOA optimization is proposed to realize the fault diagnosis of inter-turn short circuits. By comparing the fault diagnosis results of the improved Bi-LSTM neural network with those of the unimproved Bi-LSTM neural network, the effectiveness of the improved algorithm is verified. It should be noted that the data preprocessing part adopts the improved VMD decomposition method, and the feature extraction part adopts the improved multi-scale feature extraction method.

The number of hidden-layer nodes and learning rate in the Bi-LSTM prediction model are hyperparameters. During model training, the diagnosis accuracy and performance of the model can be improved by selecting the best set of hyperparameters through hyperparameter optimization. The WOA optimization algorithm is used to optimize the number of hidden layer nodes and the learning rate of the Bi-LSTM network to reduce the diagnostic error. The LSTM network has one hidden layer and one fully connected layer. The selection ranges of the number of hidden-layer nodes and the learning rate parameter are [100, 300] and [0.001, 1], respectively. The population and iteration times of the whale algorithm are 10 and 50, respectively.

Figures 15–17 show the accuracy of the training set, the confusion matrix, and the 2D scatter plot obtained by using the improved Bi-LSTM neural network for inter-turn short circuit fault diagnosis.

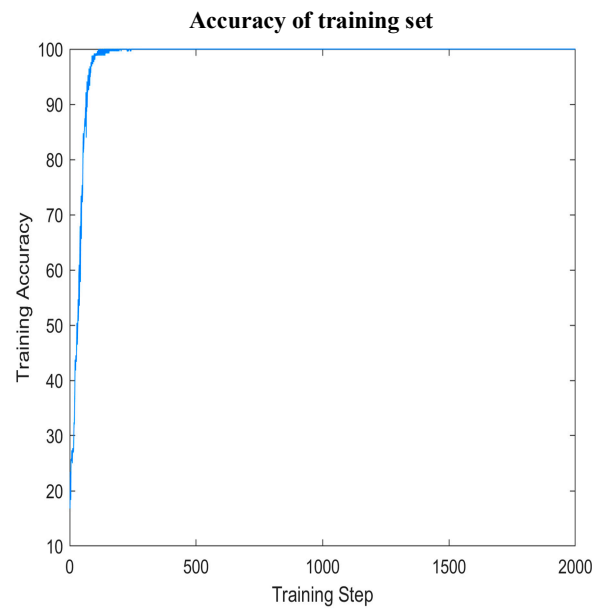



Figure 15. The accuracy of the training set using the improved Bi-LSTM network.

Figure 16. The confusion matrix using the improved Bi-LSTM network.

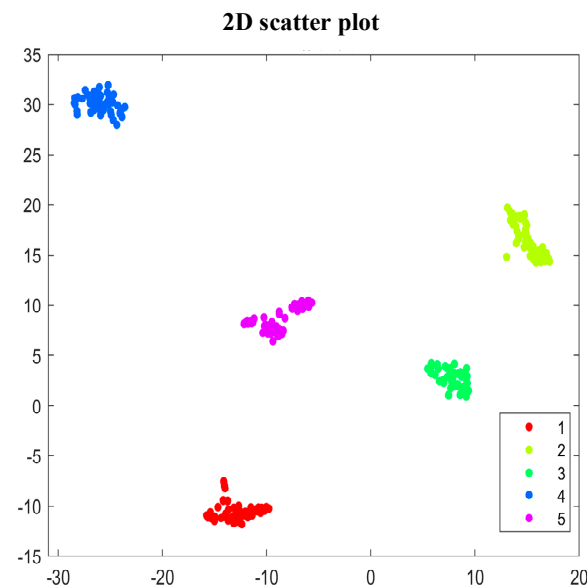



Figure 17. The 2D scatter plot using the improved Bi-LSTM network.

Comparing Figures 12–14, it can be seen that the improved VMD decomposition method, the improved multi-scale feature extraction method, and the improved Bi-LSTM fault diagnosis model have been greatly improved in all aspects. It can be seen from the whole training iteration process that the accuracy of the model tends to be stable after 300 iterations. The accuracy of fault identification is as high as 100%, the distance between classes becomes larger, the distance within classes becomes smaller, and it has high accuracy and robustness. The proposed fault diagnosis model based on multi-scale space can effectively classify PMSM inter-turn short circuit faults of different degrees.

5. Conclusions and Future Works

In this paper, the inter-turn short circuit fault simulation program of PMSM is designed, and the stator current data is collected by the Simulink simulation program. Then, the inter-turn short circuit fault diagnosis model of the permanent magnet synchronous motor based on multi-scale space is proposed. The experimental results indicate that the proposed method achieves a good diagnosis effect for different degrees of inter-turn short circuit fault of the permanent magnet synchronous motor, and the fault diagnosis accuracy is as high as 100%. The conclusions of this paper are as follows.

1. The GWO algorithm is used to adaptively select the k value and α value in VMD decomposition, and the maximum value of the product of Pearson correlation p and kurtosis value K_v is taken as the optimized objective function value, to realize the extraction of weak signals in inter-turn short circuit faults.
2. The improved coarse-grained multi-scale feature extraction improves the performance of traditional multi-scale arrangement features.
3. The WOA is adopted to optimize the number of hidden-layer nodes and the learning rate of the Bi-LSTM network, which improves the diagnostic accuracy and solves the problem of the hyper-parameter configuration of the deep learning model.
4. In practical engineering applications, a new fault diagnosis method is proposed to make fault diagnosis more intelligent.

This paper only studies the stator winding inter-turn short circuit fault of PMSM. In the future, we will conduct more extensive experimental research on other fault types of PMSM, expand the data set, and apply this fault diagnosis method to actual industrial production.

Author Contributions: Conceptualization, F.M.; methodology, F.M., L.Q. and S.Y.; software, F.M., L.Q. and S.Y.; validation, Y.C., H.X. and S.L.; data curation, L.Q. and S.Y.; writing—original draft preparation, F.M.; writing—review and editing, F.M., L.Q. and S.Y.; visualization, F.M.; project administration, L.Q. and S.Y.; funding acquisition, L.Q. and S.Y. All authors have read and agreed to the published version of the manuscript.

Funding: General Projects of National Natural Science Foundation of China (Project number: 51875270), Industry—University Research Project of Jiangsu (Project number: No. BY2020031), Graduate In-novation Program (Project number: SJCX22_1887).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is unavailable due to privacy or ethical restrictions.

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

References

1. Hu, Y. Model Predictive Torque Control Strategy for Marine Permanent Magnet Synchronous Propulsion Motor. Master's Thesis, Wuhan University of Technology, Wuhan, China, 2019. (In Chinese)
2. Yu, C.; Qi, L.; Sun, J.; Jiang, C.; Su, J.; Shu, W. Fault diagnosis technology for ship electrical power system. *Energies* **2022**, *15*, 1287. [CrossRef]
3. Zhang, H.; Liu, Y.; Zhao, X.; Hu, Q.; Liu, D. Fuzzy wavelet network intelligent predictive controller for vacuum injection molding. *Comput. Integr. Manuf. Syst.* **2010**, *16*, 2647–2652. (In Chinese)
4. Zhang, L.; Yang, M. Fault diagnosis of permanent magnet synchronous motor based on mixup-LSTM. *Electr. Switch.* **2022**, *60*, 58–62. (In Chinese)
5. Xue, S.; He, Q.; Pan, J.; Huang, X. Research on dynamic eccentricity fault diagnosis method of permanent magnet synchronous motor based on GA-SVM. *Zuhe Jichuang Yu Zidonghua Jiagong Jishu* **2022**, 99–103. (In Chinese)
6. Zhao, S.; Song, Q.; Zhang, Y.; Zhang, W. Mechanical fault detection of permanent magnet synchronous motor based on improved DFA and LDA. *J. Beijing Inst. Technol.* **2023**, *43*, 61–69. (In Chinese)
7. Huang, W.; Du, J.; Hua, W.; Fan, Q. An open-circuit fault diagnosis method for PMSM drives using symmetrical and DC components. *Chin. J. Electr. Eng.* **2021**, *7*, 124–135. [CrossRef]
8. Tang, S.; Chen, X.; Zheng, S. Fault diagnosis method of motor bearing based on attention and multi-scale convolution neural network. *Electr. Technol.* **2020**, *21*, 32–38. (In Chinese)
9. Li, F.; Honglin, L.; Shuiqing, X. Research on open circuit fault diagnosis of PMSM Inverter with SDAE-FFNN network. *Chongqing Ligongxue Xuebao* **2013**, 1–9. (In Chinese)
10. Chen, Z.; Liang, K.; Peng, T.; Wang, Y. Multi-condition PMSM fault diagnosis based on convolutional neural network phase tracker. *Symmetry* **2022**, *14*, 295. [CrossRef]
11. Zhang, Y.; Lei, Y. Data anomaly detection of bridge structures using convolutional neural network based on structural vibration signals. *Symmetry* **2021**, *13*, 1186. [CrossRef]
12. Lee, H.; Jeong, H.; Kim, S.W. Detection of interturn short-circuit fault and demagnetization fault in IPMSM by 1-D convolutional neural network. In Proceedings of the 2019 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), Macao, China, 1–4 December 2019; pp. 1–5.
13. Man, Y.; Yang, L.; Yan, L.; Cui, G. Detection method of incipient inter-turn short circuit fault of PMSM based on VMD. *Electr. Mechines Control Appl.* **2022**, *49*, 66–74. (In Chinese)
14. Wang, Y. Fault Diagnosis of Interturn Short Circuit in Permanent Magnet Synchronous Motor Based on Stacked Sparse Autoencoder. Master's Thesis, Jiangsu University of Science and Technology, Zhenjiang, China, 2021. (In Chinese)
15. Gu, J.; Jiang, T.; Zhu, H. Multi-objective discrete grey wolf optimization algorithm for job shop energy saving scheduling. *Comput. Integr. Manuf. Syst.* **2021**, *27*, 2295–2306. (In Chinese)
16. Xu, F.; Fang, Y.; Zhang, R. Cluster fault diagnosis of PCA-GG rolling bearing based on EEMD fuzzy entropy. *Comput. Integr. Manuf. Syst.* **2016**, *22*, 2631–2642. (In Chinese)
17. Liu, J.; Ma, Y.; Li, Y. Improved whale algorithm for solving engineering design optimization problems. *Comput. Integr. Manuf. Syst.* **2021**, *27*, 1884–1897. (In Chinese)
18. Wang, J.; Li, F. Review of signal processing methods in fault diagnosis for machinery. *Noise Vib. Control.* **2013**, *33*, 128–132. (In Chinese)