processes

MDPI

Article

Variable-Speed Bearing Fault Diagnosis Based on BDVMD, FRTSMFrBSIE, and Parameter-Optimized GRU-MHSA

Jie Ma *, Jun Wei, Qiao Li and Lei Xia

Mechanical and Electrical Engineering School, Beijing Information Science and Technology University, Beijing 100192, China; 2022020076@bistu.edu.cn (J.W.); lq15978774548@163.com (Q.L.); 18662700279@163.com (L.X.)
* Correspondence: mjbeijing@163.com

Abstract: To address the challenges of feature extraction and low classification accuracy in fault diagnosis of variable-speed rolling bearings, this paper proposes an intelligent fault diagnosis method based on bandwidth division variational mode decomposition (BDVMD), fractional domain time-shift multiscale fractional Boltzmann-Shannon interaction entropy (FRTSMFrBSIE), and parameter-optimized gated recurrent unit with multi-head self-attention (GRU-MHSA). First, the BDVMD is introduced to decompose and reconstruct signals, obtaining high-quality reconstructed fault signals. Next, the FRTSMFrBSIE is proposed to calculate the entropy of the reconstructed signals and generate a fault feature dataset. Subsequently, the improved dung beetle optimization (IDBO) algorithm is applied to optimize the parameters of the GRU-MHSA model, adaptively determining its optimal configuration. Finally, the fault feature dataset is input into the optimized model for fault classification, achieving a classification accuracy of 98.75%. Experiments conducted on the Ottawa bearing dataset validate the proposed method, and the results demonstrate its effectiveness and superiority in feature extraction and fault classification.

Keywords: variable-speed rolling bearings; fault diagnosis; BDVMD; FRTSMFrBSIE; IDBO-GRU-MHSA



Academic Editor: Jie Zhang

Received: 7 January 2025
Revised: 31 January 2025
Accepted: 9 February 2025
Published: 11 February 2025

Citation: Ma, J.; Wei, J.; Li, Q.; Xia, L. Variable-Speed Bearing Fault Diagnosis Based on BDVMD, FRTSMFrBSIE, and Parameter-Optimized GRU-MHSA. *Processes* **2025**, *13*, 498. https://doi.org/10.3390/pr13020498

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Rolling bearings are critical components of mechanical systems and are widely employed in various industrial equipment and machinery. Their safety and reliability directly influence the operational efficiency and stability of the system [1]. However, due to their unique structural design and harsh operating conditions, bearings frequently endure heavy loads and friction, rendering them susceptible to failures that may result in severe safety incidents and economic losses [2,3]. Traditional research on bearing fault diagnosis has primarily concentrated on constant-speed conditions. However, variable-speed conditions are more prevalent in practical scenarios, presenting significant challenges due to their complexity and uncertainty [4]. Therefore, investigating bearing fault diagnosis under variable-speed conditions is essential to better address real-world needs and holds substantial theoretical and practical value.

Rolling bearing vibration signals are often accompanied by various resonance and noise components, necessitating the effective processing of fault vibration signals. Consequently, signal decomposition methods have emerged as critical tools in vibration signal analysis [5]. Empirical mode decomposition (EMD) has been widely applied in bearing fault diagnosis. However, it suffers from mode mixing and endpoint effects [6]. To address these issues, researchers have proposed advanced algorithms such as ensemble empirical mode

decomposition (EEMD), robust empirical mode decomposition (REMD), and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN). Nonetheless, these methods fail to fundamentally address the mode mixing problem [7–9]. To overcome these limitations, the local mean decomposition (LMD) algorithm was proposed. However, its decomposition accuracy is heavily influenced by the sampling frequency, making it challenging to achieve consistent precision [10,11]. Dragomiretskiy et al. proposed a variational mode decomposition (VMD) algorithm, which can effectively extract harmonic and impulsive components from signals. Despite its advantages, the performance of VMD is significantly influenced by the selection of the number of decomposition modes and penalty factors [12,13]. Nassef et al. [14] proposed the use of sailfish optimization (SFO) to optimize VMD parameters, while Yan et al. [15] employed the cuckoo search algorithm (CSA) to determine relevant parameters. These swarm optimization algorithms demonstrate strong adaptability but are often associated with high computational costs.

After signal preprocessing, the extraction of fault feature information constitutes another critical step in fault diagnosis. Entropy, an effective metric for signal complexity and uncertainty, has been widely employed in feature extraction [16]. Commonly used methods include Shannon entropy (ShEn), approximate entropy (ApEn), and sample entropy (SampEn). However, these entropy measures are unable to capture information across multiple scales of a sequence [17–19]. Costa et al. [20] proposed multiscale entropy (MSE), which evaluates the complexity of signals across different time scales by constructing multiscale sequences. However, MSE is sensitive to the length of the original sequence, exhibiting increased susceptibility to errors under short-sequence conditions [21]. To address this issue, researchers have introduced time-shift operations to refine the construction of multiscale sequences, enhancing adaptability to short sequences while improving robustness and stability [22]. For example, Li et al. [23] proposed modified hierarchical time-shift multiscale amplitude-aware permutation entropy (MHTSMAAPE), which enhances entropy stability while extracting more detailed features. In 2014, Machado [24] leveraged the advantages of fractional-order calculus to propose the new generalized fractional entropy, which enhances signal sensitivity and provides new perspectives for related research. For instance, Li et al. [25] employed fractional fuzzy dispersion entropy (FrFuzzDE) for fault diagnosis of rolling bearings, achieving excellent diagnostic performance under constant-speed conditions. However, its applicability to more complex variable-speed conditions remains unexplored. Further research is required to develop entropy-based methods capable of efficiently extracting fault features and accurately distinguishing fault types in rolling bearing vibration signals under complex variable-speed conditions.

In addition to high-quality fault feature information, an accurate fault classifier is equally indispensable. In recent years, deep learning has achieved remarkable success in the field of fault diagnosis [26]. For instance, Liu H. et al. [27] proposed an autoencoder based on recurrent neural networks (RNNs) for rolling bearing fault diagnosis. Ting Huang et al. [28] introduced a CNN-LSTM approach to address fault diagnosis in chemical processes. These studies demonstrate the widespread application of RNNs in fault diagnosis. RNNs utilize inter-layer connections to retain previous information and use it to influence subsequent outputs, providing strong representational capabilities for complex feature data. However, RNNs are prone to vanishing and exploding gradient problems [29]. The gated recurrent unit (GRU) effectively mitigates these issues by introducing gating mechanisms. Nonetheless, GRU exhibits limitations in capturing long-term dependencies when processing global contextual information [30]. To address this, the multi-head self-attention (MHSA) mechanism is integrated, enhancing the model's capacity to capture global dependency information [31]. Despite these advancements, setting the parameters of the GRU-MHSA model remains challenging. The improved

dung beetle optimization (IDBO) algorithm, a novel swarm intelligence optimization approach with excellent optimization capabilities, adaptively searches for the model's optimal parameters [32,33]. Based on this, the IDBO-GRU-MHSA method is proposed to achieve more precise fault classification.

In summary, this paper presents a variable-speed rolling bearing fault diagnosis method based on bandwidth division variational mode decomposition (BDVMD), fractional domain time-shift multiscale fractional Boltzmann-Shannon interaction entropy (FRTSMFrBSIE) and parameter-optimized GRU-MHSA. The main contributions of this paper are as follows:

1. A parameter-setting approach for VMD using the weighted spectral trend (WST) is proposed, addressing the challenge of parameter selection in VMD. Additionally, the Gini index (GI) is utilized to filter high-quality components and reconstruct the signal.
2. By leveraging the advantages of fractional-order calculus and time-shift multiscale entropy, the FRTSMFrBSIE method is introduced to enhance the capability of Boltzmann-Shannon interaction entropy in representing fault feature information.
3. The IDBO-GRU-MHSA method is proposed to address the limitations of traditional GRU in capturing complex sequential features and the difficulty of parameter setting in GRU-MHSA, thereby improving fault classification performance.
4. An intelligent fault diagnosis method for variable-speed rolling bearings based on BDVMD-FRTSMFrBSIE and parameter-optimized GRU-MHSA is presented, demonstrating excellent fault diagnosis effectiveness.

The remainder of this paper is organized as follows: Section 2 introduces the fundamental theories of BDVMD, FRTSMFrBSIE and IDBO-GRU-MHSA. Section 3 elaborates on the proposed method, including specific steps and a flowchart. Section 4 presents experimental validations of the proposed method, along with comparative analyses for verification. Finally, Section 5 concludes the paper and outlines future research directions.

2. Theoretical Foundations

2.1. Bandwidth Division Variational Mode Decomposition

The WST method is formulated by combining the spectral trend of each frequency band with a weighting factor. The spectral trend effectively partitions the frequency bands of each mode, while the weighting factor accentuates the content of the bands by employing energy density as a parameter to quantify signal strength. This method not only efficiently delineates the frequency bands but also reduces the impact of noise on the band division results [34]. The integration of BDVMD and the WST method effectively addresses the challenges associated with parameter selection in VMD. The spectral trend can be expressed as:

$$s_t(f) = F\left(F^T F + \lambda I\right)^{-1} F^T s, s = [S(f_0), \ldots, S(f_{N-1})]^T \tag{1}$$

where $S(f)$ represents the discrete Fourier spectrum, F denotes the Fourier basis matrix, λ is the regularization coefficient, I is the identity matrix, and N is the number of spectral samples. The weighted factor is obtained by calculating the spectral energy density of each frequency band as follows:

$$W_l = \frac{\sum_{i=1}^{B_{l+1}-B_l} \left[S\left(F_{B_{l+1}}\right)\right]^2}{B_{l+1} - B_l}, l = 1, 2, \ldots, L-1 \tag{2}$$

where B_l denotes the boundary layer. The WST is defined as:

$$WST(f) = W_l \times S_t(f), f_{B_l} \leq f \leq f_{B_{l+1}} \tag{3}$$

Based on the results of the WST, the decomposition level K is determined, and the penalty factors α corresponding to each mode are computed. The expression for calculating the penalty factor is as follows:

$$\alpha_k = \beta\left(\frac{B_w - B_b}{B_w}\right) \tag{4}$$

where B_b represents the bandwidth corresponding to a specific mode, B_w denotes the sum of the bandwidths of all modes, and β is the initial penalty factor. According to reference [35], β is set to 500.

The determined parameters are provided as input to the VMD algorithm. By solving the variational constraint problem, the center frequency of each frequency band is estimated to complete the signal decomposition process. The variational constraint model can be expressed as follows:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial \left[\left(\partial(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad s.t. \sum_k u_k = f \tag{5}$$

where f represents the input signal, $\{u_k\}$ denotes the intrinsic mode functions (IMFs) obtained from the decomposition, and $\{\omega_k\}$ indicates the center frequencies of each component. By introducing a quadratic penalty term and the Lagrange multiplier operator to solve the variational model, it can be expressed as follows:

$$L(\{u_k\},\{\omega_k\},\lambda) = \alpha \sum_k \left\| \partial \left[\sigma(t) + \frac{j}{\pi t} * u_k(t) \right] e^{-jw_k t} \right\|_2^2 + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \tag{6}$$

where α represents the penalty factor, and λ is the Lagrange multiplier operator. The alternating direction method of multipliers is employed for iteration, and the frequency spectrum of each component $\hat{u}_k^{n+1}$ and the center frequency ω_k^{n+1} are respectively expressed as follows:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{7}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_i(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_i(\omega)|^2 d\omega} \tag{8}$$

The iteration of the VMD is halted when the relative error e is less than the convergence tolerance ε, which satisfies Equation (9), thereby completing the decomposition process.

$$e = \frac{\sum_k \left\| \hat{u}_k^{n+1} - \hat{u}_k^n \right\|_2^2}{\left\| \hat{u}_k^n \right\|_2^2} < \varepsilon \tag{9}$$

2.2. Fractional Domain Time-Shift Multiscale Fractional Boltzmann-Shannon Interaction Entropy

2.2.1. Fractional Boltzmann–Shannon Interaction Entropy

Boltzmann–Shannon interaction entropy (BSIE) is a novel entropy measure that combines frequency-based histograms and measure-based geometric partitions, providing a

parameter-unbiased entropy estimation method [36]. BSIE offers a standardized entropy measure that is not influenced by sample size, enabling effective comparisons of entropy across samples of varying sizes, thereby effectively addressing the bias issues present in traditional entropy estimation. Moreover, BSIE demonstrates a robust capability for handling outliers, preventing their interference with entropy estimation, and enhancing the robustness of the entropy measure.

Considering the advantages offered by the BSIE, this study further incorporates the characteristics of fractional calculus to extend the BSIE into fractional orders. The fractional Boltzmann–Shannon interaction entropy (FrBSIE) not only retains the benefits of the BSIE but also integrates fractional orders to enhance flexibility and adaptability, allowing for more effective capture of local features and dynamic changes in the data, thus accommodating diverse data distributions. The FrBSIE is defined as follows:

$$\begin{aligned} FrBSIE &= 1 - JSD^{\alpha}(\mathbf{p} \parallel \mathbf{q}) \\ &= 1 - \frac{D_{KL}^{\alpha}(\mathbf{p}\|\mathbf{m}) + D_{KL}^{\alpha}(\mathbf{q}\|\mathbf{m})}{2} \end{aligned} \tag{10}$$

$$D_{KL}^{\alpha}(\mathbf{p} \parallel \mathbf{m}) = \sum_i \left\{ -\frac{(p/m)_i^{-\alpha}}{\Gamma(\alpha+1)} [\ln(p/m)_i + \psi(1) - \psi(1-\alpha)] \right\} p_i \tag{11}$$

$$D_{KL}^{\alpha}(\mathbf{q} \parallel \mathbf{m}) = \sum_i \left\{ -\frac{(q/m)_i^{-\alpha}}{\Gamma(\alpha+1)} [\ln(q/m)_i + \psi(1) - \psi(1-\alpha)] \right\} p_i \tag{12}$$

where p represents the equiprobable histogram distribution, q denotes the quantized geometric distribution, and m is the average distribution of p and q. The Γ function is the gamma function, and the ψ function is the digamma function. The parameter α represents the fractional order. When α = 0, the FrBSIE degenerates into the traditional BSIE.

2.2.2. FRTSMFrBSIE

Multiscale fractional Boltzmann-Shannon interaction entropy (MFrBSIE), the multiscale version of FrBSIE, incorporates a coarse-graining process to quantify irregularity by averaging the original time series at non-overlapping time points, thereby estimating entropy across multiple scales. However, as the scaling factor increases, coarse-graining may shorten the effective length of the time series, potentially omitting important information within the sequence. To address this limitation, this study introduces a time-shifting technique to enhance the traditional MFrBSIE, resulting in the proposed time-shift multiscale fractional Boltzmann-Shannon interaction entropy (TSMFrBSIE) method.

Given the complexity of variable-speed rolling bearing signals and to extract signal features more comprehensively and effectively, this study further incorporates the theory of fractional-order Fourier transforms to extend the TSMFrBSIE into different fractional domains, resulting in the FRTSMFrBSIE. This method aims to comprehensively extract fault features of variable-speed rolling bearings from multiple domains and scales, thereby improving the accuracy and reliability of fault diagnosis. The computational process is as follows:

1. Set the fractional order range to [0, 1] with a step size of 0.2 to perform the fractional Fourier transform (FRFT). The signal is transformed into the p order domain, and the signal y(t) after undergoing the FRFT transformation is expressed as follows:

$$y_{\beta}(u) = F^{p}[y(t)] = \int_{-\infty}^{\infty} y(t) K_{\beta}(t,u) dt \tag{13}$$

where p represents the order of the FRFT,

$$K_\beta(t,u) = \begin{cases} \sqrt{\frac{1-j\cot\beta}{2\pi}} \exp\left\{\left[j\frac{(t^2+u^2)}{2}\right]\cot\beta - tu\csc\beta\right\}, \beta \neq n\pi \\ \delta(t-n), \beta = 2n\pi \\ \delta(t+n), \beta = (2n\pm 1)\pi \end{cases} \tag{14}$$

The range of β is from 0 to $\pi/2$, which corresponds to the order p ranging from 0 to 1. As p is varied from 0 to 1, the signal transitions progressively from the time domain to the frequency domain. Consequently, the FRFT effectively captures the time-frequency characteristics of the signal at different orders.

2. For the sequence $X = \{x_i, i = 1, 2, \ldots, N\}$ obtained from the p order domain transformation, it is divided into k subsequences. A new time-shifted subsequence $Y_{\alpha,k}$, is constructed as follows:

$$Y_{\alpha,k} = \left\{x_\alpha, x_{\alpha+k}, x_{\alpha+2k}, \ldots, x_{\alpha+k[(N-\alpha)/k]}\right\} \tag{15}$$

where k represents the time interval, and α denotes the initial time.

3. For the scale τ, the FrBSIE for each time-shifted sequence $Y_{\alpha,k}$ is calculated according to Equation (10), resulting in the FrBSIE values for τ coarse-grained sequences. By averaging the τ time-shifted fractional Boltzmann-Shannon interaction entropy (TSFrBSIE) values, obtaining the TSMFrBSIE, expressed as follows:

$$TSMFrBSIE(X,\tau,m,\lambda) = \frac{1}{\tau}\sum_{k=1}^{\tau} TSFrBSIE\left(Y_{\alpha,k}^{(\tau)}, m, \lambda\right) \tag{16}$$

where m represents the embedding dimension, while τ indicates the delay time.

4. By combining the TSMFrBSIE values from different fractional domains, the FRTSMFrBSIE is obtained, expressed as follows:

$$FRTSMFrBSIE = \left[TSMFrBSIE_{p_0}, TSMFrBSIE_{p_{0.2}}, \ldots, TSMFrBSIE_{p_1}\right] \tag{17}$$

From the outlined steps, it is evident that the coarse-graining process of TSMFrBSIE is more comprehensive, rendering it less sensitive to data length and, consequently, superior to MFrBSIE. Furthermore, the incorporation of fractional orders provides greater adaptability for different signals. Moreover, processing signals across various fractional orders improves the comprehensiveness of the extracted feature entropy values.

2.3. Improved Dung Beetle Optimization

Dung beetle optimization (DBO) parameters lack randomness, which makes it prone to getting trapped in local optima when faced with complex multi-peak problems. This limitation affects the speed and accuracy of the algorithm. By introducing randomness through Levy flight, the algorithm enhances its global search capability and mitigates the risk of premature convergence. The updated mathematical expression for the improved beetle-rolling ball algorithm is given as follows:

$$x_i(t+1) = x_i(t) + C_1 l_v b_1 \alpha k x_i(t-1) + b\Delta x \tag{18}$$

where x_i(t) represents the position information of the i-th beetle at the t-th iteration, C_1 is a random vector following a normal distribution, l_v is the Levy flight parameter, α is a

constant, k represents the deviation coefficient, and Δx denotes the light intensity. Female beetles choose the spawning area, and the position of the beetle is defined as:

$$B_i(t+1) = X^* + C_2 l_v b_1 [B_i(t) - Lb^*] + b_2 [B_i(t) - Ub^*] \tag{19}$$

where X^* represents the local optimal position of the population in this iteration, C_2 denotes a random vector with a value in the range (0, 1), where Lb^* and Ub^* are lower and upper bounds of the safety region, respectively. b_1 and b_2 represent random vectors. Some small dung beetles that have grown up emerge from underground to search for food, and the updated position is:

$$x_i(t+1) = x_i(t) + C_1 \left[x_i(t) - Lb^b \right] + C_2 \left[x_i(t) - Ub^b \right] \tag{20}$$

where Lb^b and Ub^b represent the lower and upper bounds of the optimal foraging area, respectively. During the foraging stage of the DBO algorithm, there is a lack of global search capability and diversity. To address this, t-distribution is used to influence the foraging behavior of dung beetles. The t-distribution includes a degree of freedom parameter whose probability density function is defined as:

$$p_t(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \times \Gamma\left(\frac{n}{2}\right)} \times \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \tag{21}$$

For each dung beetle, a random number is generated between 0 and 1. For dung beetles whose random number is less than the mutation probability, the t-distribution mutation is performed. The position update formula is:

$$x_i(t+1) = x_i(t) + x_i(t) \cdot trnd(n) \tag{22}$$

where $x_i\ (t+1)$ represents the new position of the i-th dung beetle after t-th iterations, $x_i(t)$ represents the optimal value currently searched, $trnd(n)$ represents the t-distribution with the current number of iterations being degrees of freedom. Dung beetles steal dung balls from others, and the thief's location update can be described as:

$$x_i(t+1) = X^* + S \cdot g \left[|x_i(t) - X^*| + \left| x_i(t) - X^b \right| \right] \tag{23}$$

where S represents a constant value, g represents a random vector that obeys normal distribution, and X^b denotes the optimal global position.

2.4. Gated Recurrent Unit with Multi-Head Self-Attention

The GRU has demonstrated superior performance in processing time-series data; however, it still has certain limitations, particularly in handling long-term dependencies. To address these shortcomings, this study incorporates the MHSA mechanism based on the GRU, aiming to capture key features of the data more accurately and filter important information. This approach enhances the model's representation capability and generalization performance.

The structure of the GRU-MHSA model is shown in Figure 1. It consists of a Sequence Input layer (SequenceInput), a GRU layer, a ReLU activation layer, a Flatten layer, a MHSA layer, a Fully Connected layer (Fc), a Softmax layer, and a Classification Output layer (Classoutput). The SequenceInput layer receives the sequential data for the neural network, with a sequence length of 60, and performs data normalization during input processing. The GRU layer captures dependencies within the sequential data and effectively manages

the flow of information through its gating mechanism. The ReLU activation function applies a thresholding operation to each input element, enhancing the model's non-linear characteristics. The Flatten layer maps high-dimensional features into a one-dimensional vector to facilitate subsequent processing. The MHSA assigns different weights to various parts of the input features. Computing an attention weight matrix effectively enhances the model's ability to capture global dependencies. The Fc layer maps input features to the output class dimensions. The Softmax layer calculates the probability for each class and outputs the classification results. Finally, the Classoutput layer compares predicted values with the target values to compute the loss.

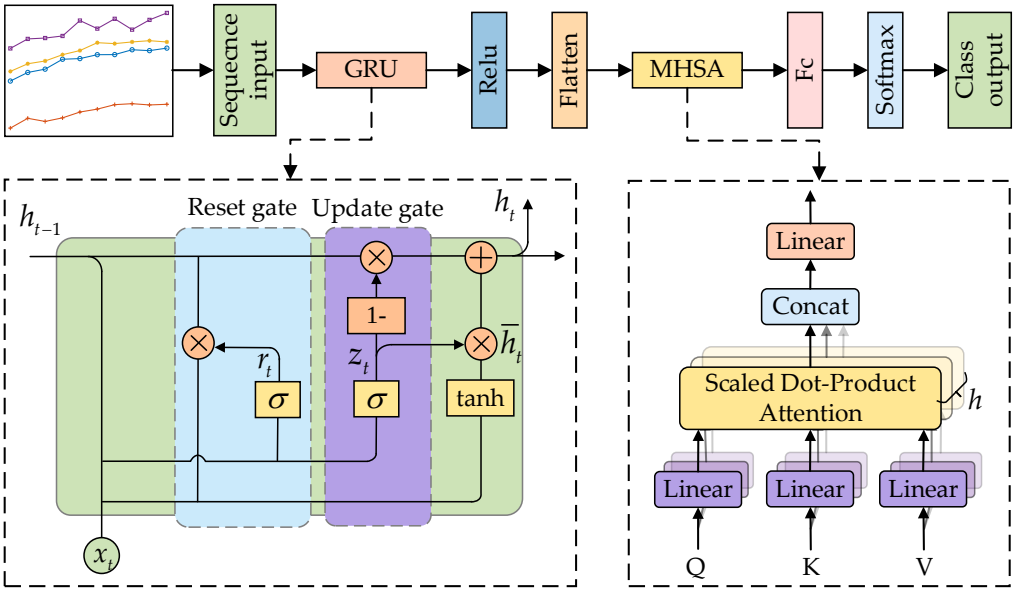


Figure 1. Structure of the GRU-MHSA model.

The GRU addresses the limitations of RNN in handling long-term dependencies and the vanishing gradient problem. Through its gating mechanism, which consists of a reset gate and an update gate, the GRU effectively captures dependencies within sequential data while simplifying the architecture compared to long short-term memory (LSTM) networks, thus avoiding the complexity associated with state transfer. This mechanism not only achieves performance comparable to that of LSTM but also reduces computational resource requirements.

Let the current input information be x_t. The reset gate r_t and update gate z_t determine the degree to which the previous information is ignored and retained, respectively. W_r, W_z, W_h, and U_r, U_z, U_h represent the weight matrices corresponding to x_t and h_{t-1}, the hidden state from the previous time step. h_{t-1}, h_t, $\overline{h_t}$ denote the hidden state from the previous time step, the current time step hidden state, and the candidate hidden state, respectively. The sigmoid function is denoted as σ. The forward propagation of the GRU is computed as follows:

$$\begin{cases} r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \\ z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \\ \overline{h}_t = \tanh[W_h x_t + U_h(r_t \odot h_{t-1}) + b_h] \\ h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \overline{h}_t \end{cases} \tag{24}$$

where b_r, b_z, and b_h are the bias terms.

The MHSA structure significantly enhances the overall performance of the model by processing different parts of the input sequence in parallel, making it more effective at capturing global dependencies within time-series data. The core components of the MHSA model include h parallel attention heads. Each attention "head" performs distinct linear mappings on the GRU output vectors to generate a set of query matrices Q, key matrices K, and value matrices V. Each head processes the input through linear transformations independently. The results are then concatenated and further processed through another linear layer to produce the final output.

$$\begin{cases} Attention(Q,K,V) = soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V \\ head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \\ MultHead(Q,K,V) = Concat(head_1, head_2, \ldots, head_h)W^o \end{cases} \tag{25}$$

where d_k represents the dimension of the matrices, and W_i^Q, W_i^K, and W_i^V are the weight matrices for Q, K, and V, respectively, while W^o is the weight matrix for the linear transformation. MHSA effectively attends to multiple positions in the input sequence simultaneously to capture various relationships, significantly enhancing the model's representation capability and generalization performance.

The preceding system analysis highlights the GRU-MHSA model, which integrates the multi-head self-attention mechanism to enhance the model's feature extraction and representation capabilities, providing robust support for classification tasks.

3. The Proposed Method

The process of the proposed method is illustrated in Figure 2. The specific process is as follows:

Step 1: Data Acquisition. Vibration signals of variable-speed rolling bearing faults from the University of Ottawa are utilized. The experimental verification involves signals under the conditions of acceleration followed by deceleration, which include normal state, inner ring fault, outer ring fault, and ball fault.

Step 2: Data Processing. Due to the high sampling frequency and excessive data points, the signals are downsampled for processing. Subsequently, the processed signals undergo BDVMD decomposition to obtain multiple IMF components. High-quality components are selected based on the GI criterion, and the signals are reconstructed accordingly. A sample length of 2048 points is used for each data sample, and 100 groups of data are randomly extracted from the reconstructed signals of each category, resulting in a total of 400 sample datasets.

Step 3: Feature Extraction. The reconstructed signals undergo FRFT with a range set to [0, 1] and a step size of 0.2, resulting in six signals across different fractional domains. The TSMFrBSIE values of the signals in each fractional domain are calculated to construct the FRTSMFrBSIE feature dataset. This forms a feature vector set with dimensions of 1×60. For each fault type, the feature dataset has dimensions of 100×60.

Step 4: Parameter Initialization. The parameters of the IDBO algorithm and the relevant parameters of the GRU-MHSA model are determined and initialized.

Step 5: The parameters of the GRU-MHSA model are optimized using IDBO. The optimal parameter settings for minimizing the fitness function of the GRU-MHSA model are searched within the defined hyperparameter range. The optimization process checks whether the maximum number of iterations has been reached; if so, the parameters and corresponding fitness values of the optimal solution are output; otherwise, optimization continues.

Step 6: Fault Classification and Recognition. The optimization results are employed to set the parameters of the GRU-MHSA model. The training and testing datasets are input into the model, yielding the final fault diagnosis classification results.

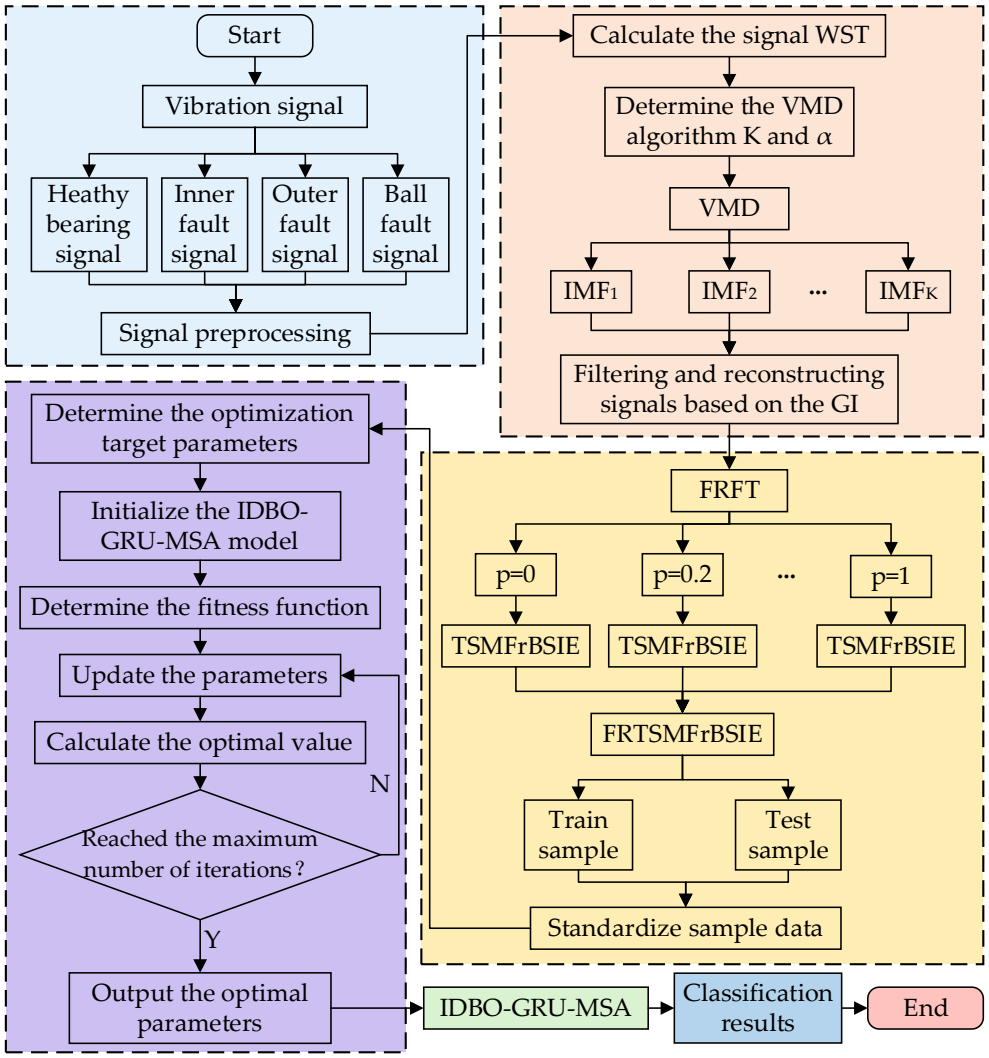


Figure 2. Flowchart for fault diagnosis of variable-speed rolling bearings.

4. Experimental Validation

4.1. Data Acquisition

This section employs the variable-speed bearing vibration signals from the University of Ottawa for experimental validation. The dataset is sourced from the mechanical fault simulator (MFS-PK5MT), and the main structure of the test rig is illustrated in Figure 3.

The test rig is driven by a three-phase motor through a coupling, with the rotational speed adjusted by a variable frequency AC drive. The experimental bearings are subjected to four conditions: healthy, inner ring fault, outer ring fault, and rolling element fault, with specific data presented in Table 1. A vibration acceleration sensor is employed to collect the bearing vibration signals, while an incremental encoder is utilized to measure the rotational speed of the shaft. The experimental sampling frequency is set at 200 kHz, with a sampling duration of 10 s. Four operating conditions are designed for the experiment, including acceleration, deceleration, a sequence of increase followed by a decrease in speed, and a sequence of decrease followed by an increase in speed.

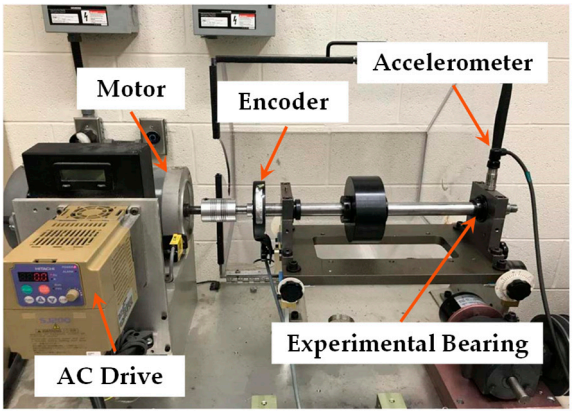


Figure 3. Structure of the experimental platform.

Table 1. Bearing parameters.

Bearing Type	Pitch Diameter	Ball Diameter	Number of Balls	Contact Angle
ER16K	38.52 mm	7.94 mm	9	0°

This study selects four types of signals under the condition of acceleration followed by deceleration for analysis and validation. This operational condition, characterized by a curved frequency variation, is more effective in demonstrating the generalization capability of the proposed method. Due to the large volume of data and high sampling rate, directly analyzing the raw signals may lead to memory overload and reduced computational efficiency. Therefore, the signals are subjected to downsampling, yielding a processed sampling frequency of 10 kHz, with a total of 100,000 sample points. The time-domain waveforms for each health condition are presented in Figure 4.

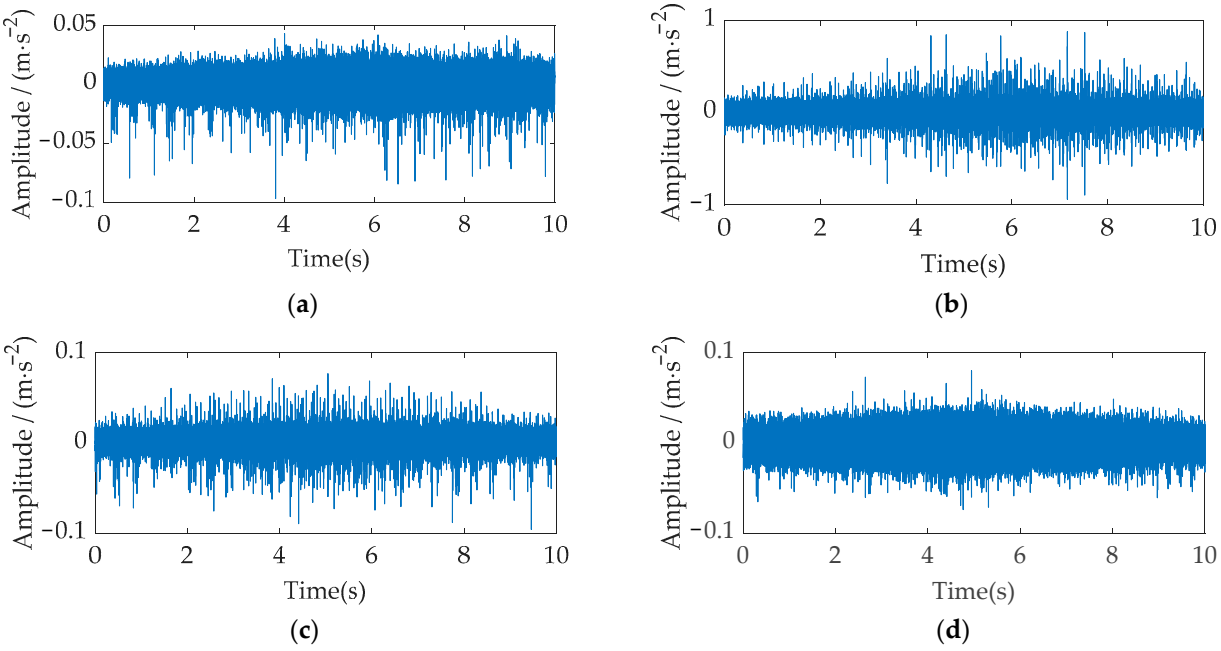


Figure 4. The time-domain waveform of bearing vibration signals during acceleration and deceleration: (**a**) Healthy condition; (**b**) Inner race fault; (**c**) Outer race fault; (**d**) Rolling element fault.

4.2. Signal Decomposition

As shown in Figure 4, there is significant noise interference in the experimental signals. To improve the accuracy and reliability of fault diagnosis, it is necessary to process the collected experimental signals. Therefore, in this section, the proposed BDVMD method is applied to decompose the signals from various health conditions. The GI is then used to select high-quality components, followed by signal reconstruction.

Taking the inner race fault signal as an example, the WST of the experimental signal is first calculated. Based on the WST results shown in Figure 5, the decomposition level K is determined to be 4. The corresponding penalty factors for each component are calculated as 390.19, 434.54, 418.42, and 256.84, respectively. Subsequently, the determined decomposition level and penalty factors are input into the VMD algorithm. The algorithm processing time is 22.41 s.

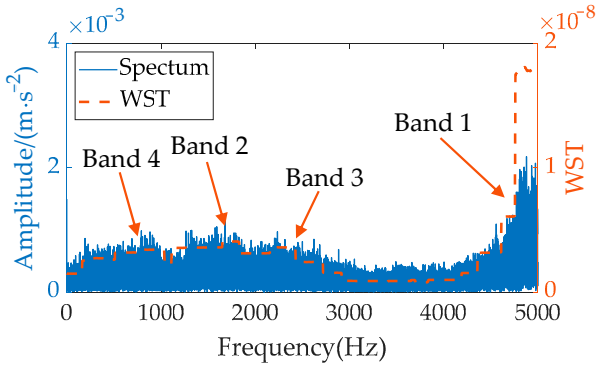
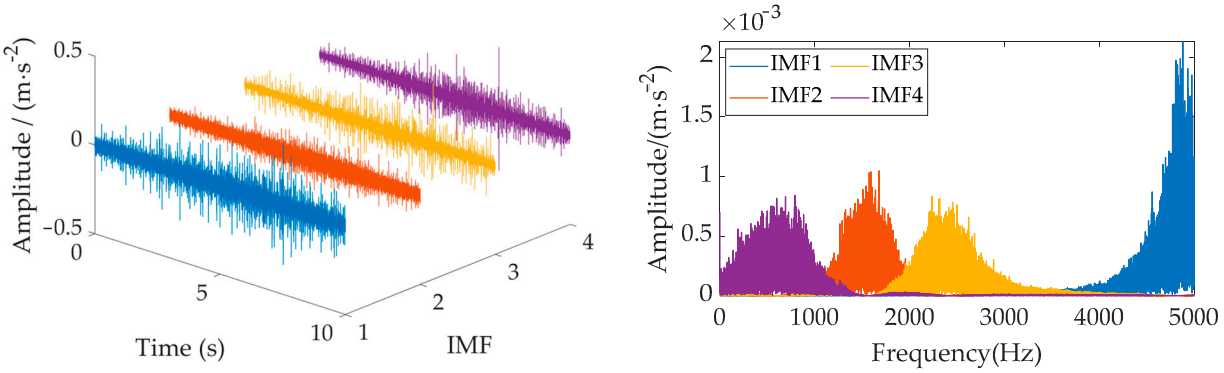


Figure 5. Frequency spectrum of WST results.

The time-domain waveforms and frequency spectra of each IMF component are shown in Figure 6. As observed in Figure 6a, BDVMD decomposes the signal into four IMF components, which corresponds to the number of primary frequency bands identified by the WST. Among these, the IMF3 and IMF4 components exhibit prominent impact features with minimal noise interference. From Figure 6b, it can be observed that the WST successfully identifies the effective frequency bands of the dominant components, thereby validating the effectiveness of this method.



(**a**) (**b**)

Figure 6. BDVMD decomposition results: (**a**) Time-domain waveform of decomposed IMFs; (**b**) Frequency spectrum of decomposed IMFs.

Finally, the GI values of each IMF are calculated, resulting in values of 0.73, 0.69, 0.81, and 0.79, respectively. The IMFs with GI values greater than the average, IMF3 and IMF4, are selected for signal reconstruction. The reconstructed signal is shown in Figure 7. From the figure, it can be observed that the amount of noise contained in the signal has been reduced.

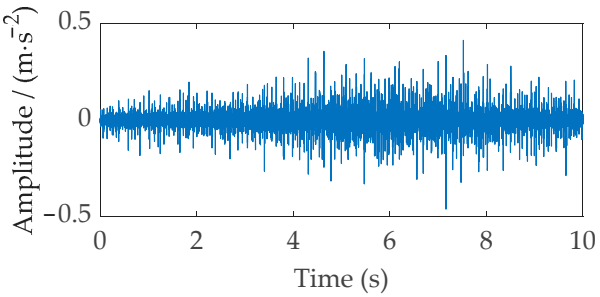


Figure 7. The time-domain waveform of the reconstructed signal.

To validate the effectiveness of the BDVMD method, this section introduces a comparative analysis with VMD, REMD, and LMD. To quantify the denoising effect, noise levels of −3 dB, 0 dB, and 3 dB were added to the signal, and the signal-to-noise ratio (SNR) changes of the reconstructed signals after processing by each algorithm were calculated. The results are shown in Table 2. Under conditions of higher noise interference, the DBVMD algorithm significantly improves the signal's SNR. As the noise decreases, DBVMD still maintains an advantage. Although the gap between DBVMD and other methods is narrowed, it remains the best choice. This indicates that the DBVMD method is effective in reducing noise in the signal while highlighting fault information, thereby demonstrating the validity of the method.

Table 2. SNR of signals processed by different methods.

Method	NULL	BDVMD	VMD	REMD	LMD
SNR (dB)	−3.0231	0.6283	0.4679	−1.1966	−1.2035
SNR (dB)	−0.0323	1.7276	1.6447	0.3793	0.4146
SNR (dB)	2.9832	3.2838	3.2394	2.9919	3.0433

The signals of the four health states are processed using the BDVMD method, resulting in reconstructed vibration signals for each fault. A batch processing approach is employed to randomly select a starting point from the vibration signals of each health state, extracting 2048 points as one data sample. Each fault type consists of 100 samples, totaling 400 samples across the four fault types. One set of data samples is selected for visualization, and its time-domain waveform is illustrated in Figure 8.

4.3. Feature Extraction

The parameters for the FRTSMFrBSIE algorithm are determined. The embedding dimension m is a critical parameter involved in the reconstruction of the phase space. A small value of m may lead to loss of information during the reconstruction process, while an excessively large m significantly increases the computational burden and necessitates a longer sample length. Considering these factors and in conjunction with existing literature, the embedding dimension is set to $m = 3$. The delay time τ has a minimal impact on the entropy value; thus, τ of 1 is typically chosen. To adequately extract the entropy values of the vibration signals across different scale factors, the scale factor is set to 10.

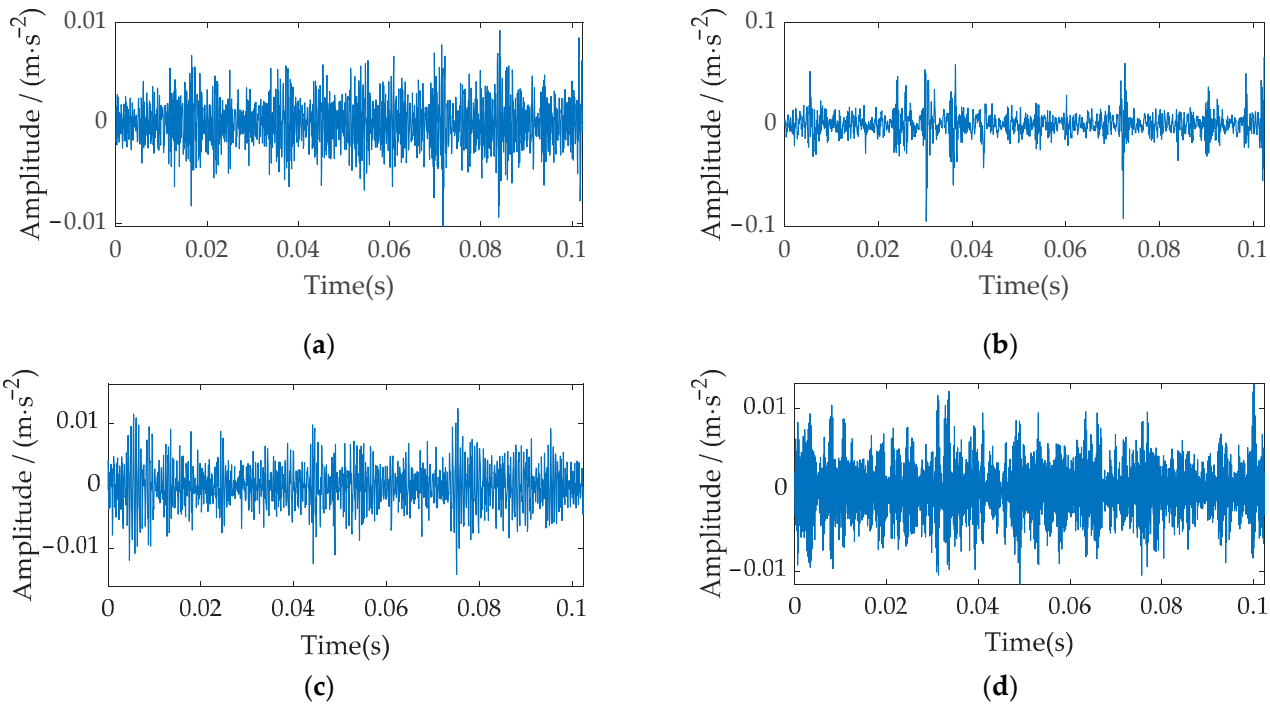


Figure 8. Time-domain waveforms of sample signals for different health conditions of rolling bearings: (**a**) Healthy condition; (**b**) Inner race fault; (**c**) Outer race fault; (**d**) Rolling element fault.

The fractional-order factor *fo* is a critical parameter that significantly influences the distinctiveness of features. Identifying the fractional-order factor with the most prominent discriminative ability is, therefore, essential. A set of vibration signals was randomly selected, and their FrBSIE values were calculated, as shown in Figure 9. In this figure, FrBSIE1 to FrBSIE4 correspond to the FrBSIE values of healthy, inner race fault, outer race fault, and ball fault conditions, respectively. When the fractional-order factor *fo* falls within the ranges of −0.3 to 0.4 or 0.7 to 0.9, the FrBSIE values exhibit notable discriminative ability among different fault signals.

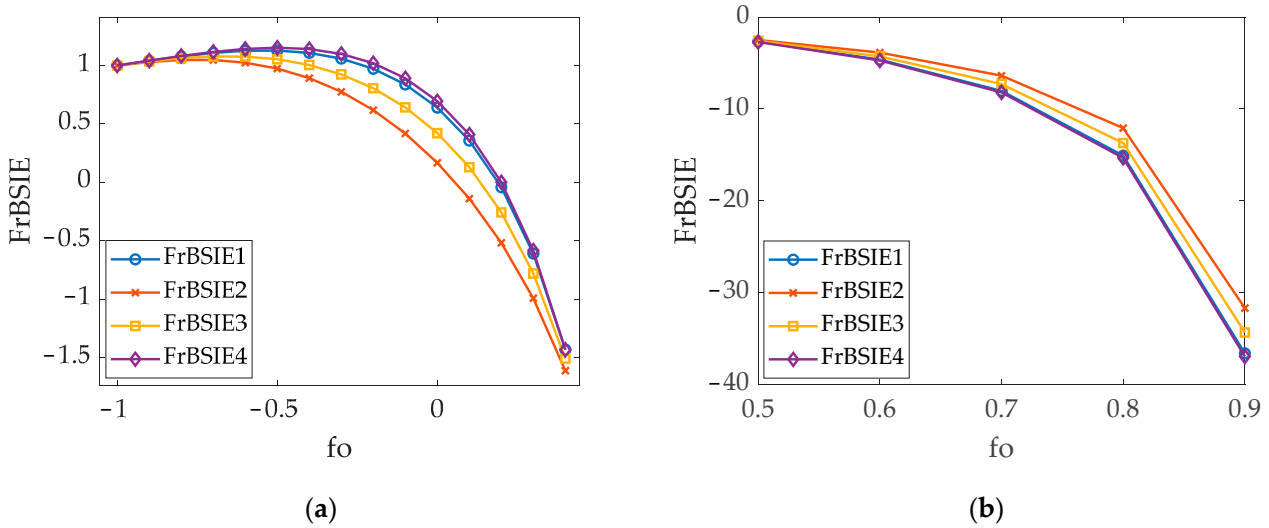


Figure 9. FrBSIE values of sample signals: (**a**) −1 to 0.4; (**b**) 0.5 to 0.9.

The FrBSIE values exhibit significant fluctuations under different fractional factors. To quantify the impact of fractional factors on the FrBSIE values of various fault vibration signals, this study introduces the coefficient of variation (CV) as a quantitative metric. Since some entropy values are negative, CV is redefined as:

$$CV = \frac{\sigma}{|\mu|} \tag{26}$$

where σ represents the standard deviation of the data, indicating the degree of dispersion, and $|\mu|$ represents the absolute mean, reflecting the central tendency of the data. Data samples from the initial, middle, and final segments of the vibration signals were randomly selected, and the CV values under different fractional-order factors *fo* were calculated. The results are shown in Figure 10. A higher CV indicates greater variability relative to the mean, signifying stronger discriminative potential. As can be seen from the figure, the CV values are relatively high within the fractional-order range of -0.1 to 0.3. Considering this observation in conjunction with the results in Figure 8, the fractional-order factor *fo* is determined to be 0.1.

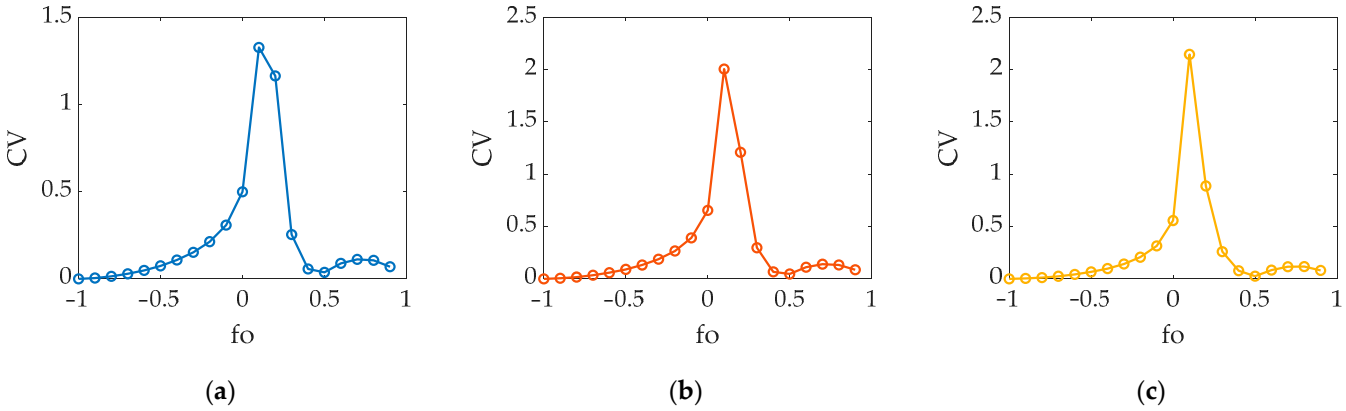


Figure 10. CV values under different fractional-order factors: (**a**) The initial segment of the signal; (**b**) The middle segment of the signal; (**c**) The final segment of the signal.

The FRTSMFrBSIE values for various fault sample signals were calculated by setting the fractional-order factor FRFT to vary from 0 to 1 in increments of 0.2. This yielded the FRTSMFrBSIE values for the sample data across different fractional domains. Each data sample has a dimension of 1×60, and each health condition forms a fault-characteristic dataset with dimensions of 100×60, resulting in an overall fault-characteristic dataset with dimensions of 400×60. The time required to calculate the entropy values of the four types of fault sample signals is as follows: Healthy state: 6.78 s; Outer race fault: 6.77 s; Inner race fault: 6.52 s; Ball fault: 6.81 s.

A random selection of sample data was chosen for visualization, with the results presented in Figure 11. In this figure, Fault1 represents the healthy state, Fault2 indicates the inner race fault, Fault3 signifies the outer race fault, and Fault4 corresponds to the rolling element fault. The entropy values of the fault signals demonstrate a consistent trend across the six fractional domains. There is a significant discriminative ability among the entropy values of different fault types within various fractional domains, with only a few entropy values exhibiting overlap. This indicates that FRTSMFrBSIE possesses strong extraction capabilities and discrimination ability when confronted with signals from different fault types.

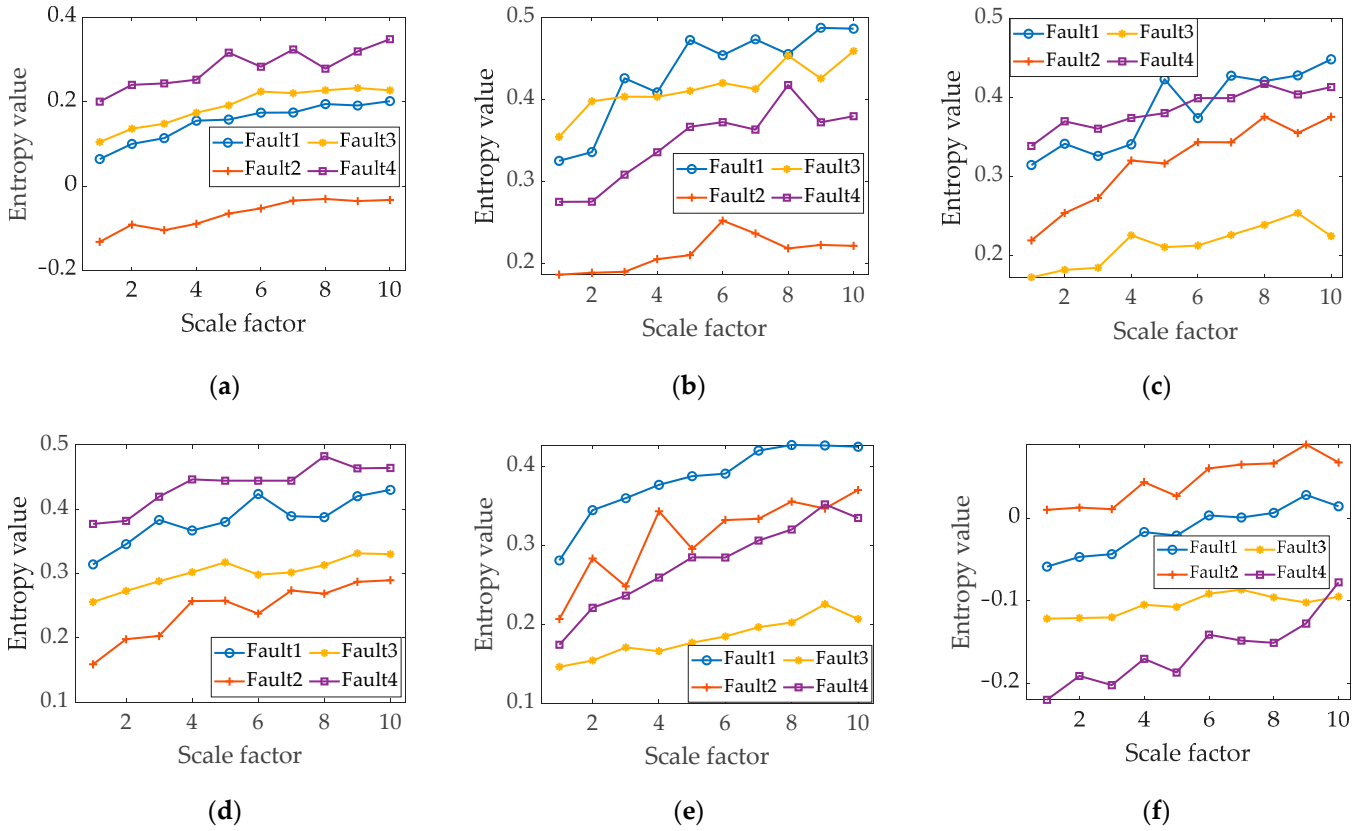


Figure 11. FRTSMFrBSIE values: (**a**) p = 0; (**b**) p = 0.2; (**c**) p = 0.4; (**d**) p = 0.6; (**e**) p = 0.8; (**f**) p = 1.

To validate the effectiveness and superiority of FRTSMFrBSIE, this section introduces fractional domain multiscale fractional Boltzmann-Shannon interaction entropy (FRMFrBSIE) and fractional domain time-shift multiscale fractional sample entropy (FRTSMFrSE) for comparative verification. The two entropies exhibit the most significant discrimination when *fo* is also set to 0.1, while the remaining parameters are consistent with previous settings. The FRMFrBSIE and FRTSMFrSE for this group of sample data were calculated for comparative purposes, with the results presented in Figures 12 and 13, respectively.

As can be seen from the figures, both FRMFrBSIE and FRTSMFrSE exhibit overlapping interference among the entropy values of different fault types across different fractional domains, showing no significant discrimination. This lack of differentiation is detrimental to the classification and judgment of fault types.

4.4. Fault Classification

The parameter settings and optimization process are as follows. First, the feature dataset is divided into training and testing sets in a ratio of 6:4. Second, the IDBO is configured with a maximum iteration count of 10 and a population size of 30. IDBO is employed to optimize the hidden layer nodes, learning rate, and regularization coefficient of the GRU-MHSA model, with their respective optimization ranges set to [10, 80], [0.0001, 0.1], and [0.0001, 0.1]. The maximum iteration count for this optimization process is set to 500, with the number of MHSA heads set to 8 and the output dimension for each attention headset to 64. In this paper, the fitness function value, denoted as *fitness*, is defined as:

$$fitness = \left[1 - \frac{\sum_{i=1}^{N} I(T_{sim}(i) = T_{train}(i))}{N}\right] \times 100 \quad (27)$$

where T_{sim} represents the predicted label vector of the model, T_{train} represents the true label vector, and N denotes the total number of samples. I is the indicator function, where $I = 1$ if $T_{sim}(i) = T_{train}(i)$, and $I = 0$ otherwise. The term $\sum_{i=1}^{N} I(T_{\text{sim}}(i) = T_{\text{train}}(i))$ represents the number of correctly classified samples.

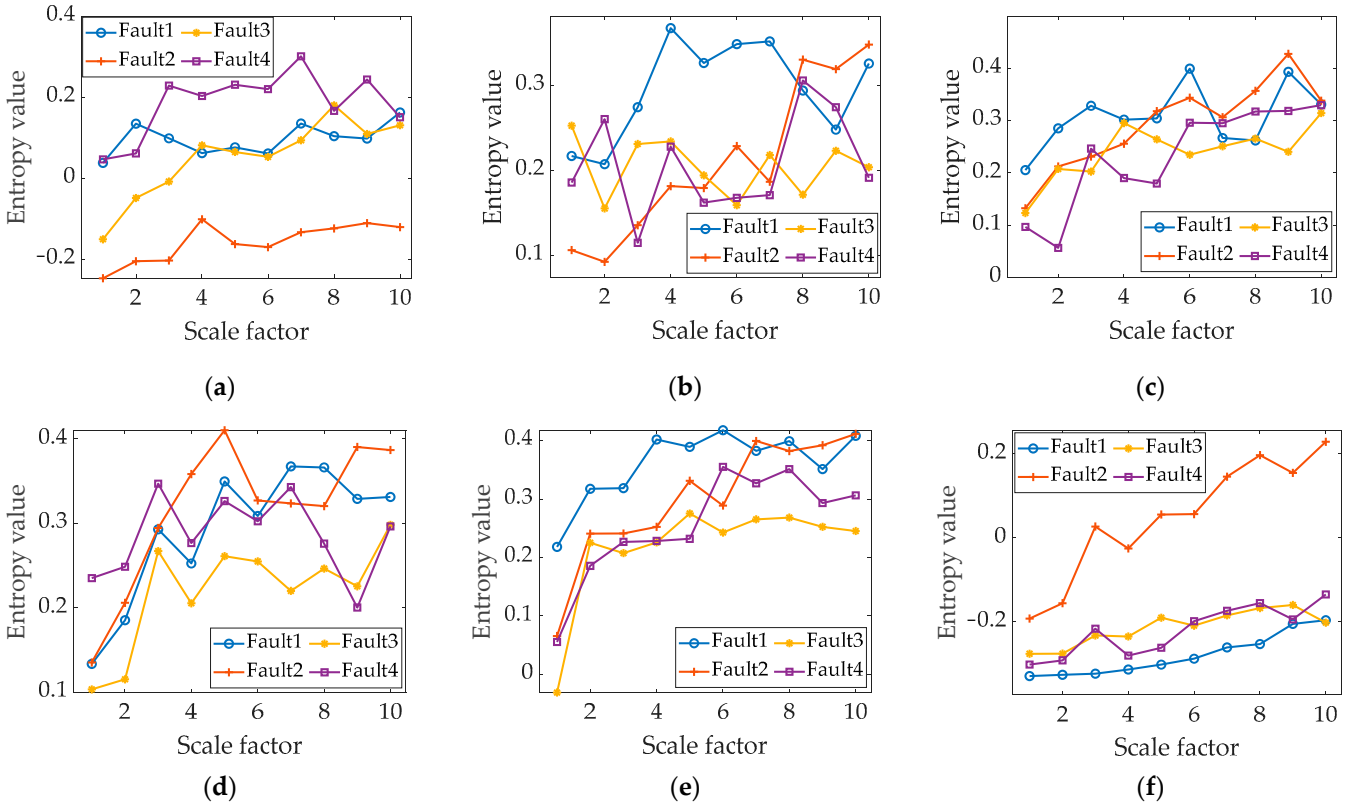


Figure 12. FRMFrBSIE values: (**a**) $p = 0$; (**b**) $p = 0.2$; (**c**) $p = 0.4$; (**d**) $p = 0.6$; (**e**) $p = 0.8$; (**f**) $p = 1$.

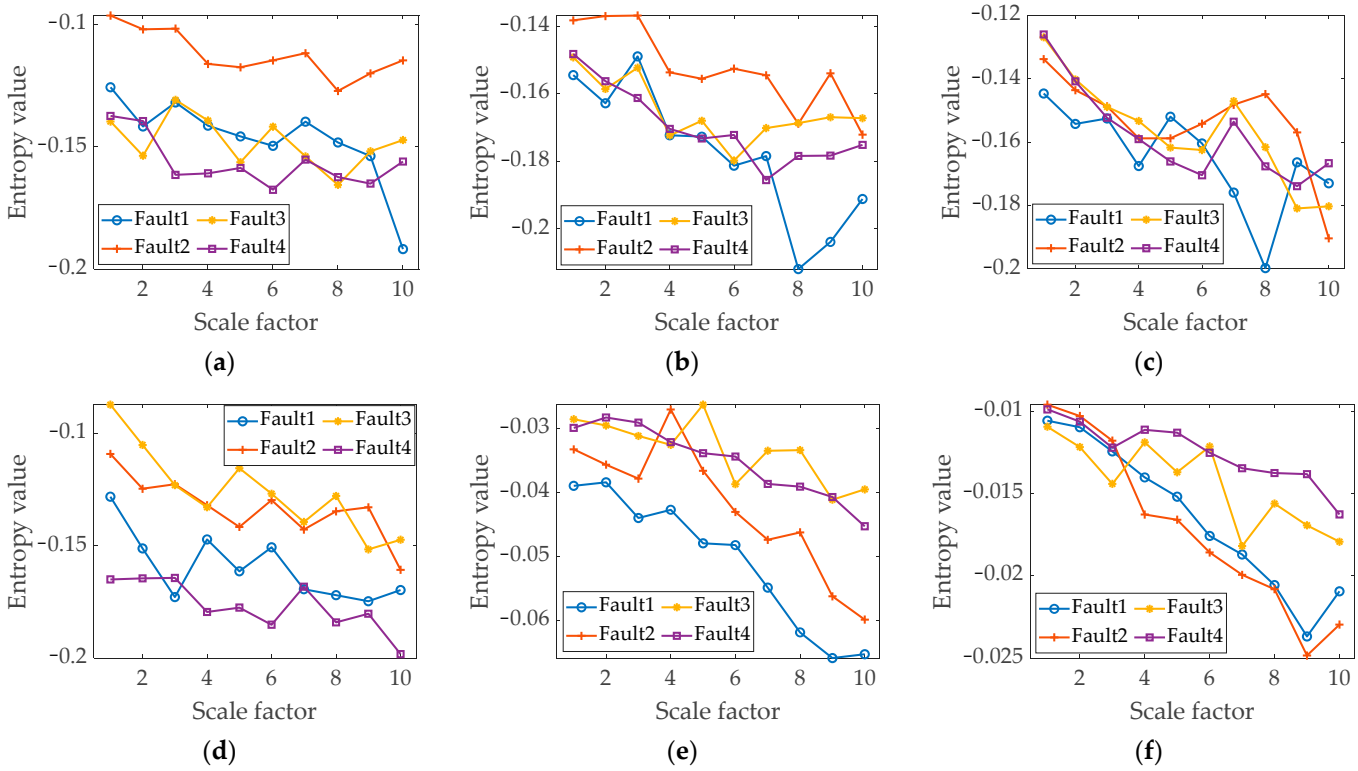


Figure 13. FRTSMFrSE values: (**a**) $p = 0$; (**b**) $p = 0.2$; (**c**) $p = 0.4$; (**d**) $p = 0.6$; (**e**) $p = 0.8$; (**f**) $p = 1$.

In this section, the improved raccoon optimization algorithm (ICOA) [37] and hunting prey optimization (HPO) [38] are introduced for comparative analysis with the IDBO algorithm. The optimization results are shown in Figure 14. Due to the significant distinguishability of the fault feature dataset, the model can efficiently fit the fault feature data of various types. All three algorithms are able to find the optimal fitness function value, which is 0, at which point the classification accuracy of the model's training set reaches 100%. The model accurately fits and classifies the four types of fault training sets, achieving optimal diagnosis. The IDBO algorithm demonstrates excellent optimization speed and accuracy, rapidly finding the optimal values for the GRU-MHSA model parameters. It reaches the optimal fitness function value in the second iteration, showing a significant advantage over the other two algorithms.

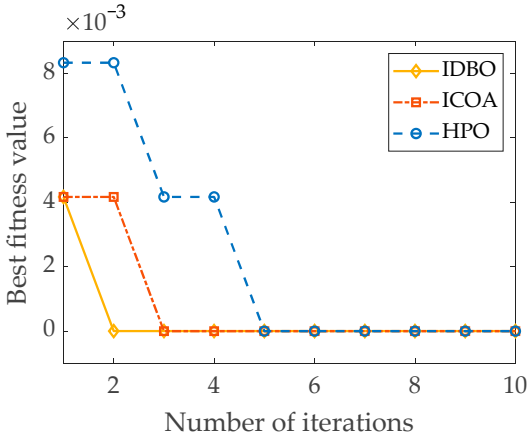


Figure 14. Convergence curve.

Finally, the optimized parameters obtained are fed into the model for training and testing. The training process took 9.213 s, the prediction time for the training set was 0.173 s, and the prediction time for the testing set was 0.021 s. The inference time is reasonable and meets the computational efficiency requirements for real-time industrial applications. Figure 15a presents the loss curve, which shows a stable decrease in loss value as the number of training epochs increases. This indicates that the model fits the training data effectively and can learn the features from the training samples while reducing prediction error. Figure 15b displays the accuracy curve, where the model achieved a classification accuracy of 100% on the training samples, consistent with the results of the optimization iteration curve. This phenomenon confirms that the IDBO-GRU-MHSA model has successfully fitted the training data.

The classification results of the proposed IDBO-GRU-MHSA model are shown in Figure 16, with the corresponding visualization results presented in Figure 17. As illustrated, the IDBO-GRU-MHSA model achieved a classification accuracy of 100% on the training samples and a classification accuracy of 98.125% on the testing samples, with only 3 healthy instances being misclassified as rolling element faults. The model demonstrates excellent fitting capabilities on the training samples while maintaining a high level of accuracy on the testing samples, showing the model's reliability and effectiveness.

To validate the effectiveness of the IDBO-GRU-MHSA model, this section conducts three comparative experiments: IDBO-GRU, GRU-MHSA, and GRU. The parameter settings for IDBO remain consistent with those previously outlined, while the GRU parameters are configured as follows: hidden layer nodes are set to 40, the learning rate is 0.01, and the regularization coefficient is 0.08. The results of these comparative experiments are illustrated in Figures 18–20.

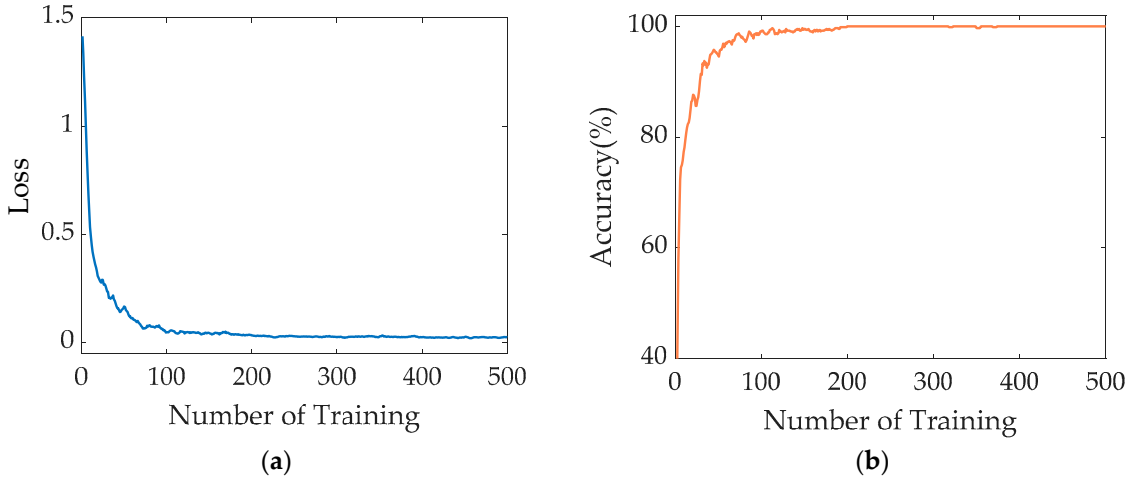


Figure 15. Training curves of the IDBO-GRU-MHSA model: (**a**) Loss curve; (**b**) Accuracy curve.

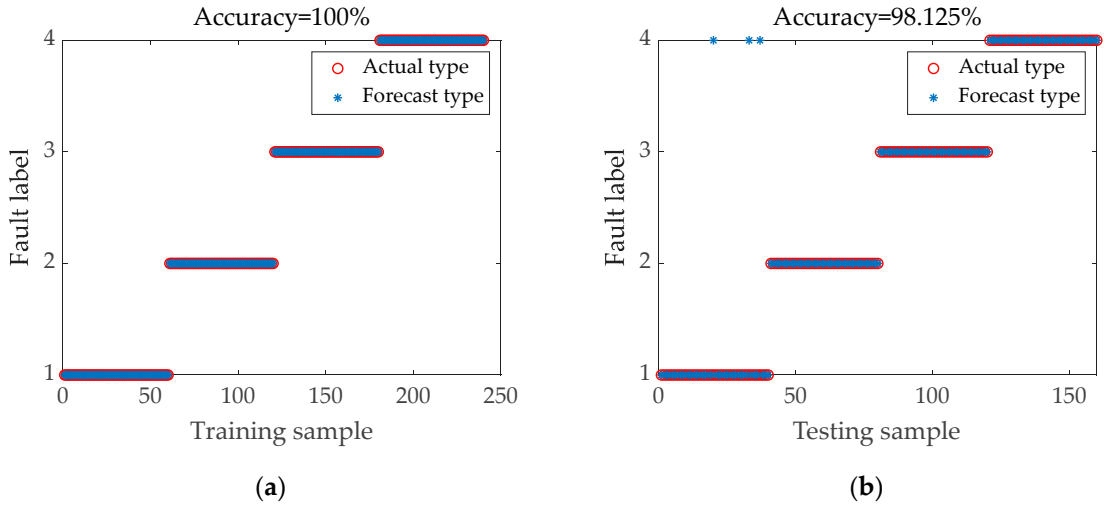


Figure 16. Fault diagnosis results of the proposed model: (**a**) Diagnosis results on the training set; (**b**) Diagnosis results on the test set.

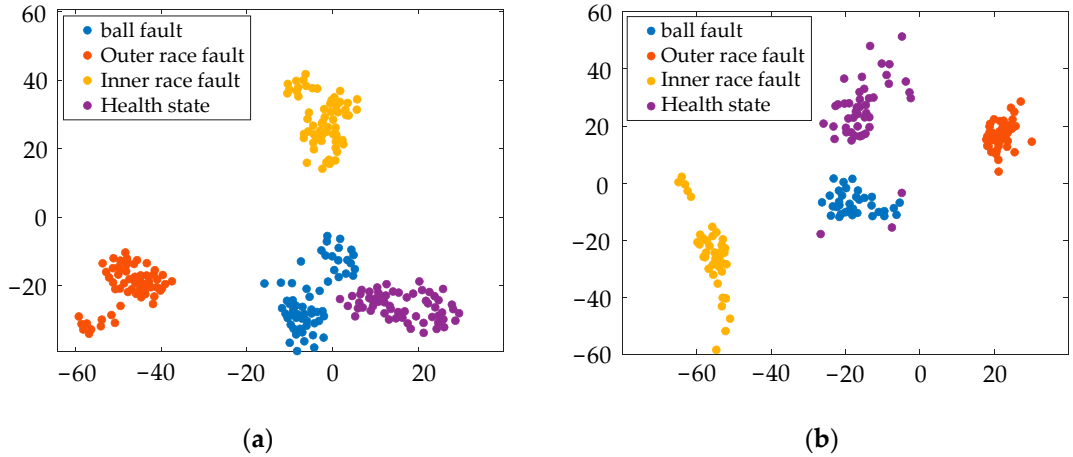


Figure 17. Visualization of fault diagnosis results of the proposed model: (**a**) Visualization of diagnosis results on the training set; (**b**) Visualization of diagnosis results on the test set.

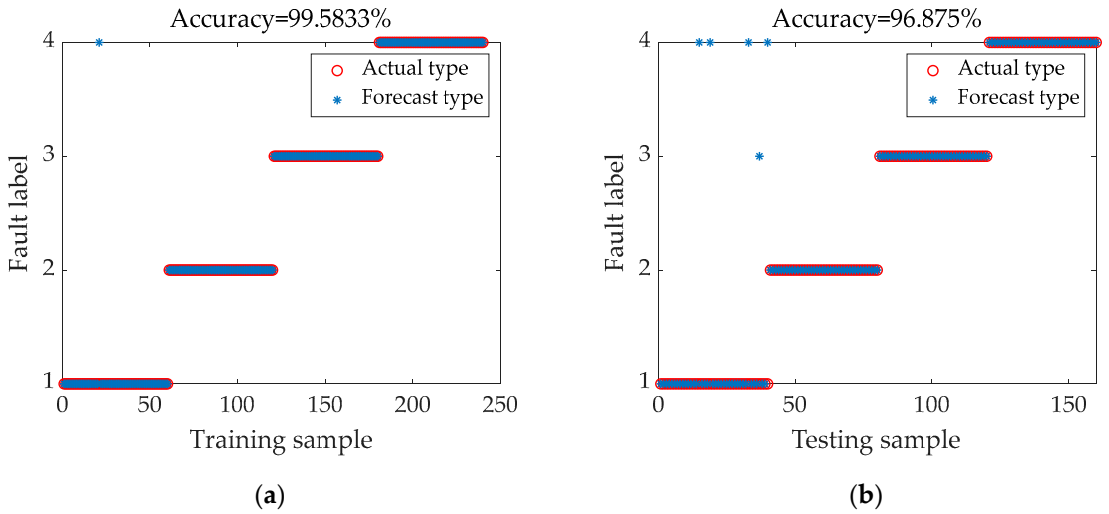


Figure 18. Fault diagnosis results of the IDBO-GRU model: (**a**) Diagnosis results on the training set; (**b**) Diagnosis results on the test set.

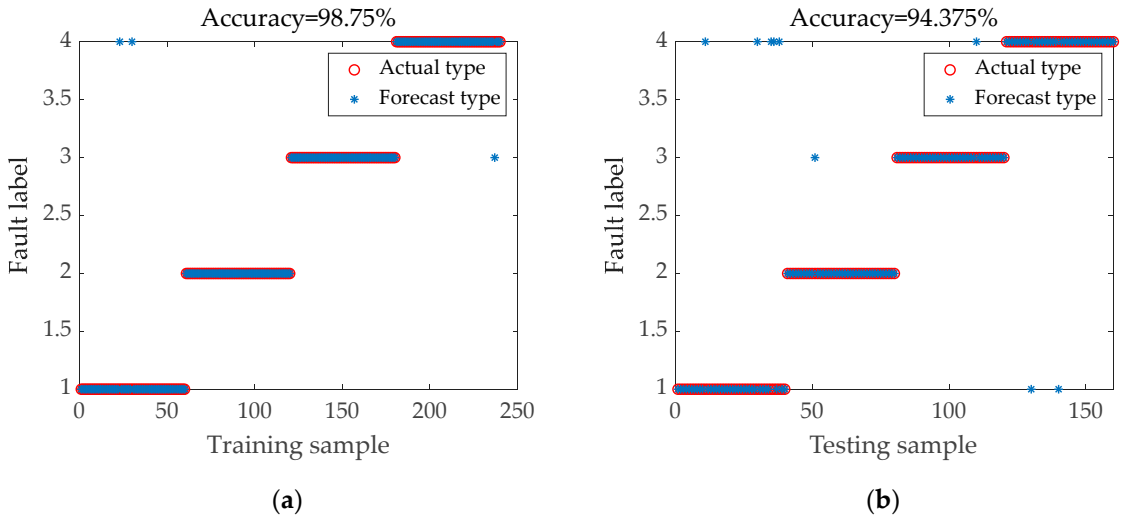


Figure 19. Fault diagnosis results of the GRU-MHSA model: (**a**) Diagnosis results on the training set; (**b**) Diagnosis results on the test set.

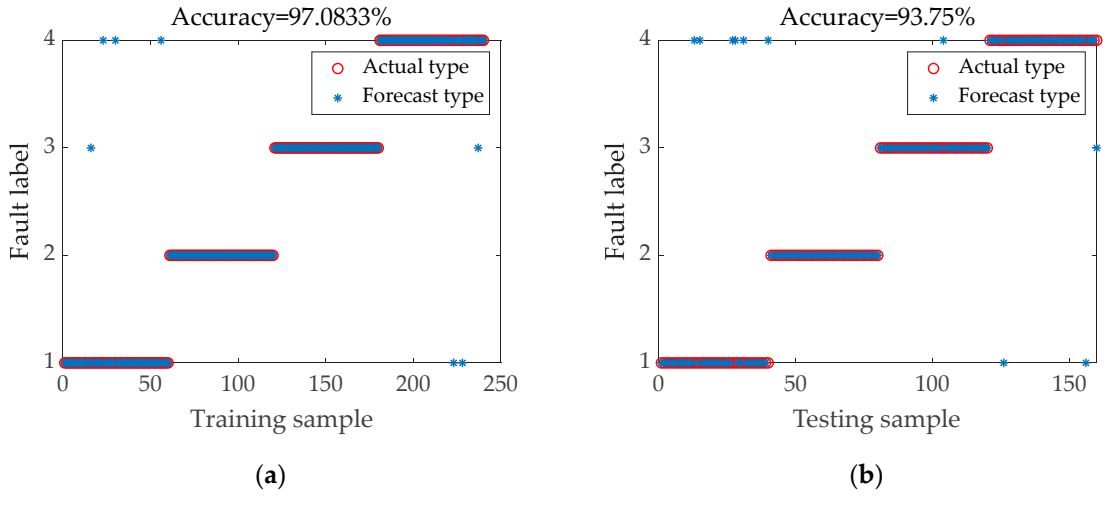


Figure 20. Fault diagnosis results of the GRU model: (**a**) Diagnosis results on the training set; (**b**) Diagnosis results on the test set.

As indicated in the figures, the model optimized by IDBO exhibits superior classification accuracy compared to the GRU model, which is attributed to its more reasonable parameter settings. Furthermore, after incorporating the MHSA, the GRU-MHSA model enhances its ability to capture global information, further improving classification accuracy. These comparative experiments reaffirm the effectiveness of the IDBO-GRU-MHSA model in classification tasks.

4.5. Comparative Experiments

Ten groups of comparative experiments were conducted to validate the effectiveness of the proposed method. The number of hidden nodes, learning rate, and regularization coefficient for the LSTM model are set to 40, 0.01, and 0.05, respectively, while the BP model has 40 hidden nodes and a learning rate of 0.01. Each model was trained 10 times, and the average, best, and standard deviation values of the test set results, as well as the training and testing times for each model, were recorded. The specific methods and experimental results for each model are shown in Table 3.

Table 3. The specific methods and experimental results for each model.

Method	Signal Decomposition	Feature Value	Classification Model	Average (%)	Best (%)	Std	Time (s)
Proposed Method	DBVMD	FRTSMFrBSIE	IDBO-GRU-MHSA	98.063	98.750	0.438	9.239
Model 1	DBVMD	FRTSMFrBSIE	IDBO-LSTM-MHSA	97.125	98.125	0.848	13.904
Model 2	DBVMD	FRTSMFrBSIE	LSTM	93.875	96.625	1.335	8.799
Model 3	DBVMD	FRTSMFrBSIE	BP	92.625	94.375	1.474	7.106
Model 4	DBVMD	TSMFrBSIE	IDBO-GRU-MHSA	93.563	97.500	2.725	8.010
Model 5	DBVMD	FRMFrBSIE	IDBO-GRU-MHSA	92.063	95.625	1.855	9.612
Model 6	DBVMD	FRTSMFrSE	IDBO-GRU-MHSA	94.063	97.500	1.795	10.149
Model 7	REMD	FRTSMFrBSIE	IDBO-GRU-MHSA	87.875	91.875	1.899	9.655
Model 8	LMD	FRTSMFrBSIE	IDBO-GRU-MHSA	85.250	90.625	2.439	9.861
Model 9	REMD	FRMFrBSIE	LSTM	79.938	83.750	2.155	8.581
Model 10	LMD	FRTSMFrSE	BP	82.275	86.875	2.568	7.404

The proposed model achieved the highest average accuracy and best accuracy among all models. Additionally, it exhibited the smallest standard deviation, indicating more stable prediction results while maintaining reasonable computation times. When compared with Models 1, 2 and 3, the IDBO-GRU-MHSA model demonstrated higher classification accuracy and superior stability. Although Model 4 achieves relatively good optimal diagnostic accuracy, its standard deviation is large, and the average accuracy is unsatisfactory. The features have not undergone FRFT processing, and the extracted feature entropy values are limited to time-domain signals. Compared to the proposed method, the fault feature information contained in the feature dataset is less comprehensive. This limitation may contribute to the larger standard deviation observed in its diagnostic accuracy. Compared with Models 5 and 6, the results confirmed that calculating the FRTSMFrBSIE values to construct the feature dataset enabled better differentiation of various fault types, leading to improved diagnostic performance. When compared with Models 7 and 8, the results validated that BDVMD effectively reduces noise interference while retaining relatively complete fault vibration signal information. The experimental results of Model 9 and Model 10 demonstrate that in the fault diagnosis of rolling bearings under variable speed conditions, each stage, including signal denoising, feature extraction, and classification, significantly impacts the final diagnostic accuracy. Compared to the proposed method, the diagnostic ac-

curacy of these two models is relatively poor. Overall, the proposed model exhibits superior performance across all aspects and demonstrates the best comprehensive capabilities.

The fractional-order factor *fo* is a key parameter that influences feature distinguishability. To validate the effectiveness of determining *fo* in FRTSMFrBSIE based on CV, different values of *fo* were set in the range of −0.9 to 0.9, with a step size of 0.1. Multiple fault feature datasets were constructed accordingly. These datasets were input into the IDBO-GRU-MHSA model for training and testing, with each model being run 10 times to calculate the average accuracy. The overall diagnostic accuracy results on both the training and testing sets are shown in Figure 21.

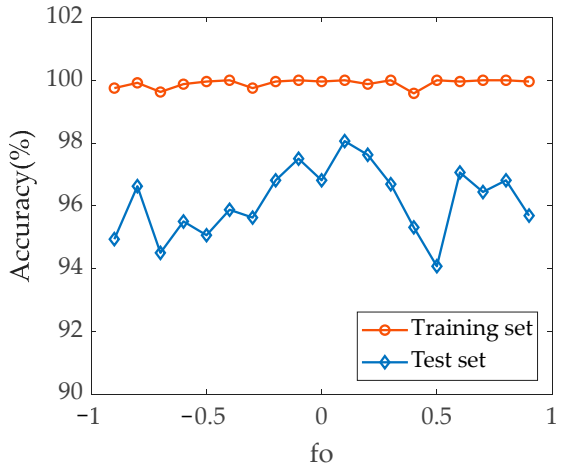


Figure 21. The average classification accuracy of the training and testing sets under different fractional order factors.

As can be seen from the figure, the training set accuracy remains above 98%, indicating that the classification model is able to effectively fit the datasets under different fractional-order factors. When *fo* is set to 0.1, the average fault diagnosis accuracy reaches 98.063%. However, when *fo* is 0.5, the fault diagnosis accuracy drops to 94.075%. When *fo* is set to −0.8 and 0.6, the fault diagnosis accuracy of the testing set increases. These results align with the overall trends in the FrBSIE and CV values shown in Figures 9 and 10, thereby validating the feasibility of determining *fo* in FRTSMFrBSIE based on CV.

In practical operating conditions, the bearing environment is complex, and the vibration signals collected by sensors often contain noise interference. To evaluate the noise resistance and robustness of the proposed fault diagnosis method, Gaussian white noise of varying intensities was added to the fault vibration signals for validation. In the test, the SNR values ranged from −10 dB to 10 dB, with a step size of 2 dB. Each model was run 10 times, and the average accuracy was calculated to assess the classification accuracy of the training and testing sets under different noise intensities. The results are shown in Figure 22.

Under different SNR values, the classification accuracy of the training set consistently remained above 95%, indicating that the model has good fitting ability under noise interference. However, when the SNR value was between −10 dB and −6 dB, representing a high-noise environment, the diagnostic accuracy of the testing set fell below 85%, indicating that high-intensity noise interferes with the model's performance. As the SNR value increased, the diagnostic accuracy of the testing set improved substantially, reaching 88.938% when the SNR was −4 dB. As the noise gradually weakened, the model's diagnostic accuracy progressively increased, and when the SNR value exceeded −2 dB, the diagnostic accuracy surpassed 90%. When the SNR value was greater than 4 dB, the model's diagnostic accuracy remained at a high level and was almost unaffected by noise, staying above

95%. These results demonstrate that the model exhibits strong noise resistance performance in noisy environments.

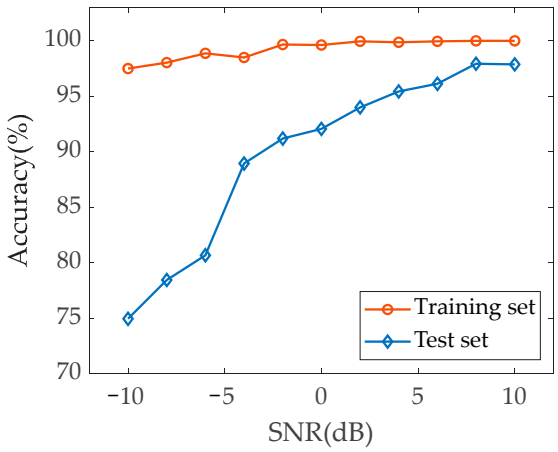


Figure 22. The average classification accuracy of the training and testing sets under different levels of noise interference.

To investigate the variation in diagnostic accuracy of the proposed method under different fault data volumes, this study conducts an analysis of data scarcity [39]. The number of fault samples per class is set within the range of 20 to 150, increasing in increments of 10 to construct multiple fault feature datasets. The datasets are then divided into training and test sets in a 6:4 ratio, and each dataset configuration is independently run 10 times to compute the average diagnostic accuracy. The diagnostic accuracy results for the training and test sets are presented in Figure 23.

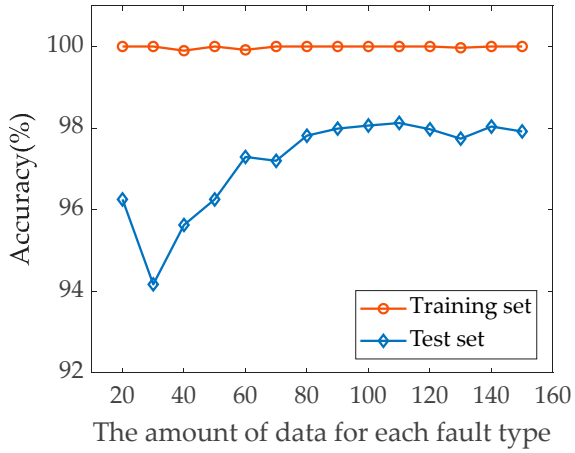


Figure 23. The average classification accuracy of the training and testing sets under data volumes.

As shown in the figure, the training accuracy of the model remains relatively stable and consistently high. When the number of data samples is below 60, the average diagnostic accuracy on the test set falls below 97%, indicating that the model exhibits limited generalization ability under data scarcity. As the sample size increases, the diagnostic accuracy of the test set shows an overall upward trend. When the sample size reaches 80 or higher, the test set accuracy stabilizes at approximately 98%, suggesting that the model achieves satisfactory diagnostic accuracy and generalization capability at this stage.

To evaluate the model's adaptability, this study introduces bearing fault vibration signals under different operating conditions for testing. Specifically, the following operating condition signals were selected: the bearing healthy state vibration signal with acceleration

followed by deceleration, the inner race fault vibration signal during acceleration, the outer race fault vibration signal during deceleration, and the ball fault vibration signal with deceleration followed by acceleration. The corresponding time-domain waveforms are shown in Figure 24.

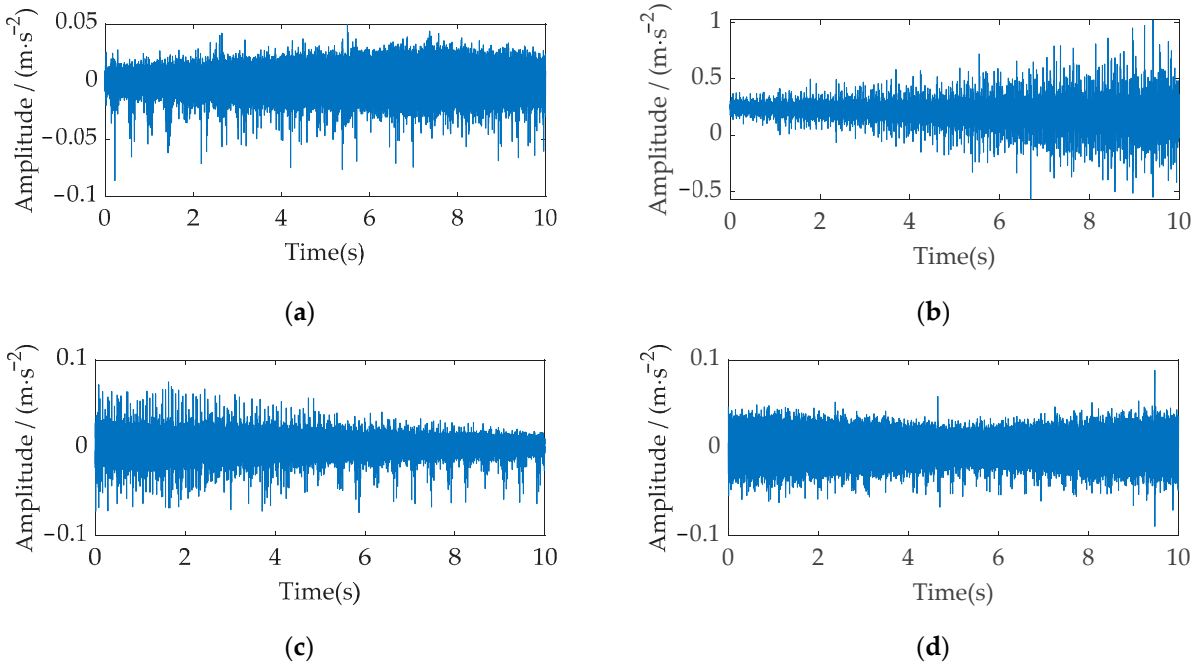


Figure 24. Time-domain waveforms of bearing vibration signals under different operating conditions: (**a**) Accelerating and then decelerating bearing healthy state; (**b**) Accelerating inner race fault; (**c**) Decelerating outer race fault; (**d**) Decelerating and then accelerating ball fault.

The signals were denoised, and a feature entropy dataset was constructed according to the signal processing methods and fault feature dataset construction methods outlined in Sections 4.2 and 4.3. The processed data were then input into the IDBO-GRU-MHSA model for fault diagnosis, and the classification results are shown in Figure 25.

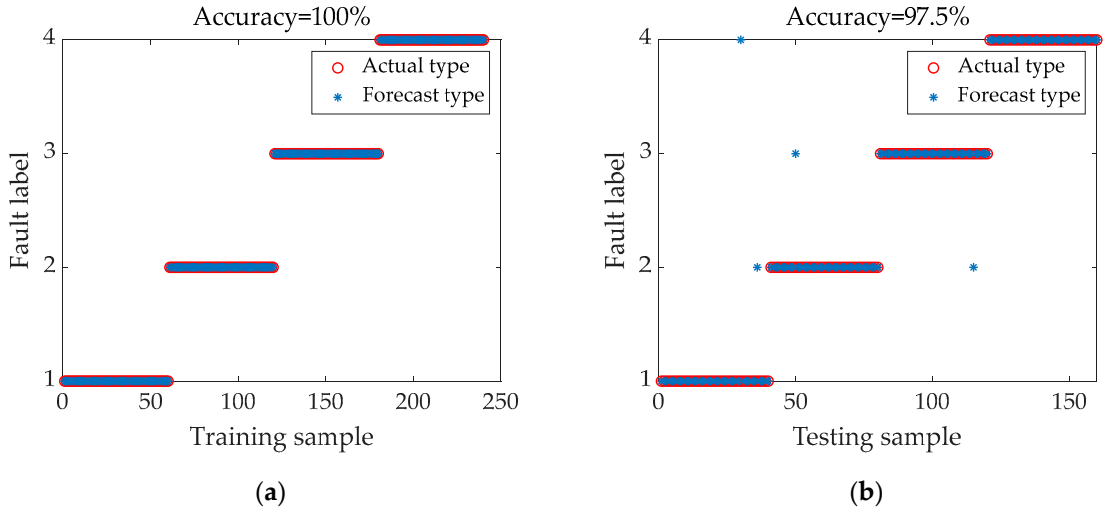


Figure 25. Fault diagnosis results of the proposed model under different operating conditions: (**a**) Diagnosis results on the training set; (**b**) Diagnosis results on the test set.

As can be seen from the figure, the model achieves 100% classification accuracy on the training set, indicating that it can effectively fit bearing fault data under various

operating conditions and accurately predict fault types. In the diagnostic results for the testing set, the model achieved a classification accuracy of 97.5% on unseen data. This indicates that although the model's performance slightly decreased on the testing set, it still maintains strong predictive capability, demonstrating excellent generalization performance and confirming the model's robust classification ability in fault diagnosis.

5. Conclusions

This paper proposes a novel fault diagnosis method for variable-speed rolling bearings based on BDVMD, FRTSMFrBSIE and IDBO-GRU-MHSA. The effectiveness of this approach is validated using the bearing dataset from the University of Ottawa. Experimental results demonstrate that the proposed method achieves a fault diagnosis accuracy of 98.75%. Comparative experiments further confirm the superiority and reliability of the proposed approach. The main contributions and advantages of this study are as follows:

1. The BDVMD algorithm addresses the challenge of parameter selection in the VMD algorithm, effectively denoising the fault vibration signals of variable-speed rolling bearings. Combined with the GI, it enables the accurate selection of high-quality components and signal reconstruction, thereby significantly improving the SNR.
2. The proposed FRTSMFrBSIE method integrates the advantages of fractional calculus and time-shift multiscale entropy, enhancing the ability of Shannon interaction entropy to represent fault feature information. Experimental results indicate that the entropy values of various fault features exhibit excellent distinguishability.
3. The IDBO-GRU-MHSA resolves the limitations of traditional GRU in capturing complex sequence features and the challenges of parameter settings in GRU-MHSA. It demonstrates high accuracy and reliability in fault feature classification.
4. This study proposes an intelligent fault diagnosis method for variable-speed bearings based on BDVMD, FRTSMFrBSIE and IDBO-GRU-MHSA. Through experimental analysis and comparative experiments, the method is shown to exhibit outstanding performance in terms of fault diagnosis accuracy and stability.

The proposed method enables the decomposition of fault signals, feature extraction, and intelligent classification diagnosis for variable-speed rolling bearings, providing a novel approach to bearing fault diagnosis. Moreover, this method exhibits good scalability, enabling the processing of fault vibration signals from components under various operating conditions, extracting feature entropy to construct feature datasets, and performing fault diagnosis classification. However, due to the black-box nature of neural network models, their interpretability is limited. Future research will focus on enhancing the model's interpretability. Additionally, future work could explore the integration of structures such as CNNs to develop multi-channel neural fusion network models. This approach aims to extract and integrate fault feature data under more complex and diverse operating conditions, thereby improving the model's feature extraction capability and, ultimately, increasing both diagnostic accuracy and stability.

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

Funding: This research was funded by the National Natural Science Foundation of China under Grant No. 61973041.

Data Availability Statement: The data are publicly available.

Acknowledgments: We would like to thank the University of Ottawa for providing the bearing dataset under time-varying conditions.

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

References

1. Wei, M.; Jiang, L.; Zhang, D.; Wang, B.; Tu, F.; Jiang, P. Bearing Fault Feature Extraction Method Based on Variational Mode Decomposition of Fractional Fourier Transform. *Russ. J. Nondestruct. Test.* **2022**, *58*, 221–235. [CrossRef]
2. Feng, Z.; Chen, X.; Wang, T. Time-varying demodulation analysis for rolling bearing fault diagnosis under variable speed conditions. *J. Sound. Vib.* **2017**, *400*, 71–85. [CrossRef]
3. Wang, Z.; Yang, J.; Guo, Y. Unknown fault feature extraction of rolling bearings under variable speed conditions based on statistical complexity measures. *Mech. Syst. Signal Process.* **2022**, *172*, 108964. [CrossRef]
4. Ji, M.; Peng, G.; He, J.; Liu, S.; Chen, Z.; Li, S. A Two-Stage, Intelligent Bearing-Fault-Diagnosis Method Using Order-Tracking and a One-Dimensional Convolutional Neural Network with Variable Speeds. *Sensors* **2021**, *21*, 675. [CrossRef]
5. Chen, B.; Zhang, W.; Gu, J.X.; Song, D.; Cheng, Y.; Zhou, Z.; Gu, F.; Ball, A.D. Product envelope spectrum optimization-gram: An enhanced envelope analysis for rolling bearing fault diagnosis. *Mech. Syst. Signal Process.* **2023**, *193*, 110270. [CrossRef]
6. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. Ser. A* **1998**, *454*, 903–995. [CrossRef]
7. Yin, C.; Wang, Y.; Ma, G.; Wang, Y.; Sun, Y.; He, Y. Weak fault feature extraction of rolling bearings based on improved ensemble noise-reconstructed EMD and adaptive threshold denoising. *Mech. Syst. Signal Process.* **2022**, *171*, 108834. [CrossRef]
8. Liu, Z.; Peng, D.; Zuo, M.J.; Xia, J.; Qin, Y. Improved Hilbert-Huang transform with soft sifting stopping criterion and its application to fault diagnosis of wheelset bearings. *ISA Trans.* **2022**, *125*, 426–444. [CrossRef]
9. Liang, Y.; Lin, Y.; Lu, Q. Forecasting gold price using a novel hybrid model with ICEEMDAN and LSTM-CNN-CBAM. *Expert Syst. Appl.* **2022**, *206*, 117847. [CrossRef]
10. Li, Y.; Liang, X.; Yang, Y.; Xu, M.; Huang, W. Early fault diagnosis of rotating machinery by combining differential rational spline-based LMD and K-L divergence. *IEEE Trans. Instrum. Meas.* **2017**, *99*, 3077–3090. [CrossRef]
11. Liu, H.; Fang, H.; Yu, X.; Xia, Y. Adaptive Signal Extraction Based on RLMD-SVD for Water Pipeline Leakage Localization. *IEEE Sens. J.* **2024**, *24*, 26522–26533. [CrossRef]
12. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2013**, *62*, 531–544. [CrossRef]
13. Li, H.; Liu, T.; Wu, X.; Chen, Q. An optimized VMD method and its applications in bearing fault diagnosis. *Measurement* **2020**, *166*, 108185. [CrossRef]
14. Nassef, M.G.A.; Hussein, T.M.; Mokhiamar, O. An adaptive variational mode decomposition based on sailfish optimization algorithm and Gini index for fault identification in rolling bearings. *Measurement* **2021**, *173*, 108514. [CrossRef]
15. Yan, X.; Jia, M. Application of CSA-VMD and optimal scale morphological slice bispectrum in enhancing outer race fault detection of rolling element bearings. *Mech. Syst. Signal Process.* **2019**, *122*, 56–86. [CrossRef]
16. Wang, S.; Li, Y.; Noman, K.; Wang, D.; Feng, K.; Liu, Z.; Deng, Z. Cumulative spectrum distribution entropy for rotating machinery fault diagnosis. *Mech. Syst. Signal Process.* **2024**, *206*, 110905. [CrossRef]
17. Aguayo-Tapia, S.; Avalos-Almazan, G.; Rangel-Magdaleno, J.d.J. Entropy-Based Methods for Motor Fault Detection: A Review. *Entropy* **2024**, *26*, 299. [CrossRef]
18. Pincus, S.M. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 2297–2301. [CrossRef] [PubMed]
19. 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]
20. Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of biological signals. *Phys. Rev. E* **2005**, *71*, 021906. [CrossRef] [PubMed]
21. Wang, Z.; Li, G.; Yao, L.; Cai, Y.; Lin, T.; Zhang, J.; Dong, H. Intelligent fault detection scheme for constant-speed wind turbines based on improved multiscale fuzzy entropy and adaptive chaotic Aquila optimization-based support vector machine. *ISA Trans.* **2023**, *138*, 582–602. [CrossRef]
22. Pham, T.D. Time-shift multiscale entropy analysis of physiological signals. *Entropy* **2017**, *19*, 257. [CrossRef]
23. Li, Z.; Li, L.; Chen, R.; Zhang, Y.; Cui, Y.; Wu, N. A novel scheme based on modified hierarchical time-shift multi-scale amplitude-aware permutation entropy for rolling bearing condition assessment and fault recognition. *Measurement* **2024**, *224*, 113907. [CrossRef]
24. Machado, J.T. Fractional order generalized information. *Entropy* **2014**, *16*, 2350–2361. [CrossRef]

25. Li, Y.; Tang, B.; Geng, B.; Jiao, S. Fractional Order Fuzzy Dispersion Entropy and Its Application in Bearing Fault Diagnosis. *Fractal Fract.* **2022**, *6*, 544. [CrossRef]
26. Raouf, I.; Kumar, P.; Kim, H.S. Deep learning-based fault diagnosis of servo motor bearing using the attention-guided feature aggregation network. *Expert Syst. Appl.* **2024**, *258*, 125137. [CrossRef]
27. Liu, H.; Zhou, J.; Zheng, Y.; Jiang, W.; Zhang, Y. Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA Trans.* **2018**, *77*, 167–178. [CrossRef]
28. Huang, T.; Zhang, Q.; Tang, X.; Zhao, S.; Lu, X. A novel fault diagnosis method based on CNN and LSTM and its application in fault diagnosis for complex systems. *Artif. Intell. Rev.* **2022**, *55*, 1289–1315. [CrossRef]
29. Zhang, S.; Zhang, S.; Wang, B.; Habetler, T.G. Deep learning algorithms for bearing fault diagnostics—A comprehensive review. *IEEE Access* **2020**, *8*, 29857–29881. [CrossRef]
30. Mao, X.; Zhang, F.; Wang, G.; Chu, Y.; Yuan, K. Semi-random subspace with Bi-GRU: Fusing statistical and deep representation features for bearing fault diagnosis. *Measurement* **2021**, *173*, 108603. [CrossRef]
31. Wan, L.; Huang, L.; Ning, J.; Li, C.; Li, K. A novel meta-transfer learning approach via convolutional multi-head self-attention network for few-shot fault diagnosis. *Knowl.-Based Syst.* **2024**, *299*, 112113. [CrossRef]
32. Zhang, H.; Yang, S.; Xu, D. An improved optimization dung beetle algorithm. In Proceedings of the 2023 4th International Symposium on Computer Engineering and Intelligent Communications (ISCEIC), Nanjing, China, 18–20 August 2023; pp. 17–20. [CrossRef]
33. Zhu, F.; Li, G.; Tang, H.; Li, Y.; Lv, X.; Wang, X. Dung beetle optimization algorithm based on quantum computing and multi-strategy fusion for solving engineering problems. *Expert Syst. Appl.* **2024**, *236*, 121219. [CrossRef]
34. Chen, S.; Guo, L.; Fan, J.; Yi, C.; Wang, K.; Zhai, W. Bandwidth-aware adaptive chirp mode decomposition for railway bearing fault diagnosis. *Struct. Health Monit.* **2024**, *23*, 876–902. [CrossRef]
35. Huang, Y.; Lin, J.; Liu, Z.; Wu, W. A modified scale-space guiding variational mode decomposition for high-speed railway bearing fault diagnosis. *J. Sound Vibr.* **2019**, *444*, 216–234. [CrossRef]
36. Diggans, C.T.; AlMomani, A.A.R. Boltzmann–Shannon interaction entropy: A normalized measure for continuous variables with an application as a subsample quality metric. *Chaos* **2023**, *33*, 123131. [CrossRef] [PubMed]
37. Hashim, F.A.; Houssein, E.H.; Mostafa, R.R.; Hussien, A.G.; Helmy, F. An efficient adaptive-mutated coati optimization algorithm for feature selection and global optimization. *Alex. Eng. J.* **2023**, *85*, 29–48. [CrossRef]
38. Naruei, I.; Keynia, F.; Molahosseini, A.S. Hunter–prey optimization: Algorithm and applications. *Soft Comput.* **2022**, *26*, 1279–1314. [CrossRef]
39. Rezazadeh, N.; Perfetto, D.; de Oliveira, M.; De Luca, A.; Lamanna, G. A Fine-Tuning Deep Learning Framework to Palliate Data Distribution Shift Effects in Rotary Machine Fault Detection. *Struct. Health Monit.* **2024**, 14759217241295951. [CrossRef]

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