



## OPEN ACCESS

## EDITED BY

Yaoyao Wang,  
Nanjing University of Aeronautics and  
Astronautics, China

## REVIEWED BY

Junsheng Cheng,  
Hunan University, China  
Jie LING,  
Nanjing University of Aeronautics and  
Astronautics, China

## \*CORRESPONDENCE

Yang Shi,  
✉ shiyang20250331@163.com

RECEIVED 21 August 2025

REVISED 28 October 2025

ACCEPTED 28 October 2025

PUBLISHED 25 November 2025

## CITATION

Shi Y (2025) Improved multi-scale divergence  
entropy combined with extreme learning  
machine classifier for rotating machinery  
fault recognition.  
*Front. Mech. Eng.* 11:1690084.  
doi: 10.3389/fmech.2025.1690084

## COPYRIGHT

© 2025 Shi. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#). The use,  
distribution or reproduction in other forums is  
permitted, provided the original author(s) and  
the copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic practice.  
No use, distribution or reproduction is  
permitted which does not comply with these  
terms.

# Improved multi-scale divergence entropy combined with extreme learning machine classifier for rotating machinery fault recognition

Yang Shi\*

Mechanical Engineering Department, Shanxi Engineering Vocational College, Taiyuan, China

**Introduction:** As the core equipment in industrial production, rotating machinery bearings play a critical role. However, traditional feature extraction algorithms for vibration signals are susceptible to noise interference and inaccurate in extracting complex features. Meanwhile, traditional fault classification algorithms face challenges such as high dependence on feature quality and insufficient generalization ability.

**Methods:** For vibration signal feature extraction, an improved multi-scale divergence entropy method is proposed. It integrates multi-scale sample entropy and divergence entropy to enhance the discrimination of signal features. For fault classification, a regularized extreme learning machine (ELM) model is developed, where regularization constraints are introduced to avoid pathological matrices.

**Results:** When using the refined composite multi-scale divergence entropy for feature extraction, setting the scale to 20 minimized the entropy value and achieved the highest classification accuracy of 98.79%. For the regularized ELM model, adopting the Softplus function as the activation function and setting the neuron number to 17 led to the lowest loss rate and the highest average classification accuracy of  $93.98\% \pm 0.94\%$ . Additionally, the model exhibited a relatively short running time of only 400 ms.

**Discussion:** The results indicate that the improved multi-scale divergence entropy effectively enhances the robustness and accuracy of feature extraction under noise interference. The regularized ELM model improves both classification accuracy and computational efficiency compared to traditional algorithms. This proposed method not only advances the classification accuracy of rotating machinery faults but also provides new technical support for machine fault prevention work, demonstrating potential for practical industrial applications in fault diagnosis systems.

## KEYWORDS

rotating machinery, feature extraction, fault identification, MDE, RELM

## 1 Introduction

Rotating machinery has become a guarantee for sustainable production in various industries due to its efficient energy conversion and transmission capabilities. However, due to long-term exposure to complex working conditions, it is difficult to avoid faults in various parts of rotating machinery (Liu D. et al., 2023). Among them, bearing failures are more

serious. To ensure the normal operation of the machine, fault detection and prevention are extremely important. Traditional methods include time-domain analysis, frequency-domain analysis, and time-frequency analysis (Tao et al., 2023). Time-domain analysis is simple and intuitive, but it is difficult to reveal complex fault characteristics. Frequency-domain analysis displays the frequency components, but cannot reflect the changes in frequency over time. Although time-frequency analysis combines the advantages of both, it has high computational complexity and is susceptible to noise interference. Traditional fault classification methods, involving artificial neural networks, Support Vector Machines (SVM), etc. rely on manually designed features, have high requirements for feature extraction, and lack generalization ability (Gangsar et al., 2021). In recent years, technologies such as deep learning, machine learning, and the Internet of Things have been applied in fault recognition, achieving high-precision classification through automatic feature extraction (Tama et al., 2023; Purohit and Dave, 2023; Jiao et al., 2022). However, these methods still have issues such as requiring much labeled data and limited performance in small sample scenarios. Therefore, there is an urgent need to develop a more comprehensive fault recognition system to accurately extract and classify fault features.

Currently, most scholars have introduced multi-scale entropy into the field of fault recognition. Multi-scale entropy evaluates the complexity of time series by analyzing their entropy values at different time scales, which is commonly used to analyze vibration signals. Li et al. built a refined composite variable step multi-scale model to extract only single time scale information from nonlinear dynamic indices. Compared with other commonly used entropies, this model could detect series nonlinear dynamic changes with high stability and variability (Li Y. et al., 2024). To solve the fault feature extraction in multi-channel vibration signal synchronization, Wang et al. combined mutation embedding multi-scale diversity entropy feature extraction with random forest classifier. Through simulation and experimental verification, the multi-channel feature extraction capability of this method was the best (Wang et al., 2021). Kafantaris et al. introduced a hierarchical entropy framework to address the information loss caused by multiple entropy quantization algorithms. Three algorithm variants of hierarchical multi-scale dispersed entropy were proposed. The results indicated that variants prioritized channel processing and had low computation time, and could also utilize prior knowledge to extract new information (Kafantaris et al., 2022). To address the multi-scale diversity entropy ignoring high-frequency information, Wang et al. combined hierarchical diversity entropy with a random forest classifier. Compared with various existing entropy methods, this model had the best feature extraction ability and could effectively solve the early fault signal feature extraction (Wang et al., 2022).

Extreme Learning Machine (ELM) has fast learning speed, strong generalization ability, and is easy to implement. Many scholars have applied it to fault classification (Liu X. et al., 2023). In response to the electrical fault problem of external short circuits, Yang R et al. used an ELM-based thermal model to obtain the temperature behavior of lithium-ion batteries under external short-circuit conditions. The model had higher computational efficiency and better fitting (Yang et al., 2021). To solve the fault diagnosis for

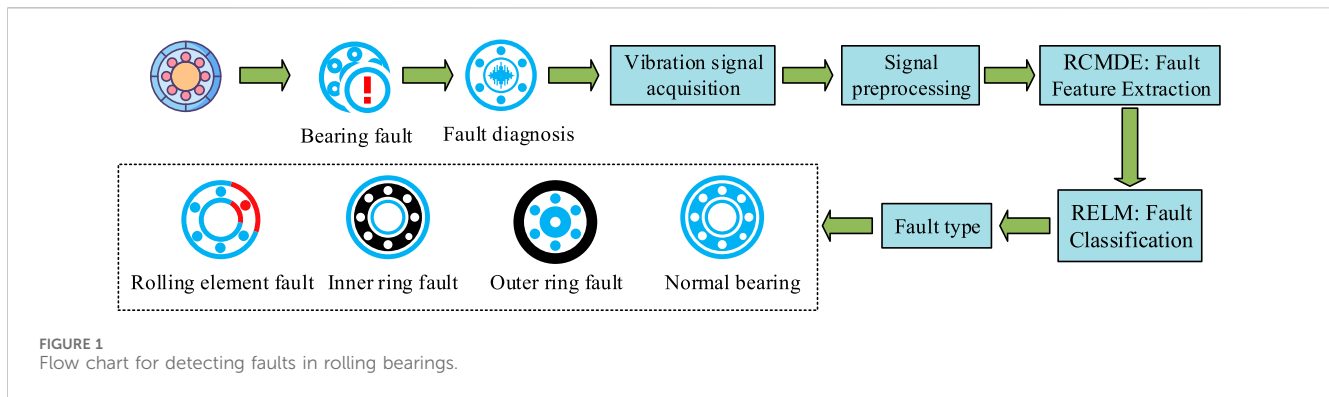
steam turbines, Dhini et al. proposed an ELM-radial basis functional network model. Although the proposed model had lower accuracy, its computational speed far exceeded that of backpropagation neural networks (Dhini et al., 2022). Xie et al. proposed a fusion thermal model that integrated multiple thermal effects and constructed a pseudo distributed model for diagnosing internal short circuits in batteries using thermal behavior. Experiments were conducted on 18,650 lithium-ion batteries. The results showed that this scheme could effectively identify internal short-circuit faults in low-grade batteries, with a state misjudgment rate of only 3.13% (Xie and Yao, 2021). To address the impact of strong noise on acoustic sensor data, Zhou Y et al. built a new method that combined acoustic sensor signal feature parameters with a double-layer angular kernel ELM. This method outperformed other advanced methods based on sound sensor data (Zhou et al., 2022).

To address the problem of unbalanced rolling bearing data and difficulty in fault diagnosis under variable speed conditions, Li S et al. used interference attribute projection to process segmented variable speed data; and proposed an adaptive clustering weighted oversampling method to process unbalanced data. The effectiveness of this method was verified through comparative experiments on two data sets (Li et al., 2024a). To address the problem of unbalanced and uneven distribution of wind turbine blade icing data, Li S et al. proposed a center jump hoist method. This method combines improved clustering-based oversampling and lightweight gradient hoisting to predict blade icing. The results show that the accuracy and running time are significantly improved on two SCADA data sets (Li et al., 2023). To address the problem of abnormal deviation and abnormal accuracy drop in helicopter transmission system vibration data, Fan C et al. proposed a variable scale multilayer perceptron. The results show that this method is superior to the Transformer and MLP models on bearing and gear data sets (Fan et al., 2024). To address the problem of few label samples, data imbalance, complex transfer learning training, and weak interpretation in bearing fault diagnosis, Li S et al. proposed a center jump hoist method for semi-supervised intelligent fault identification of unbalanced data bearings. The results show that the superiority of the proposed method is verified on three bearing data sets with different balancing rates (Li et al., 2024b).

In summary, many scholars have proposed many solutions in feature extraction and fault classification, with certain achievements and progress. However, these methods still have problems such as insufficient model generalization ability and high requirements for features. In response to the above issues, a new classification model is proposed, which is based on the features obtained from the Refined Composite Multi-scale Divergence Entropy (RCMDE) and uses a Regular Extreme Learning Machine (RELM) to classify faults to address the recognition task of rotating machinery faults. The innovation of the research lies in combining multi-scale and Divergence Entropy (DE) to discriminate fault features, and introducing regular constraints into ELM to solve the pathological matrix, which is conducive to improving the recognition accuracy of rotating machinery faults.

## 2 Methods and materials

The fault signals of rolling bearings are generally vibration signals. Relying on the vibration signals, RCMDE is taken to



extract multi-scale features and output the features in vector form. Then, the feature vectors are input into RELM. Based on its fast learning algorithm and good generalization ability, the input fault feature vectors are trained to establish a mapping relationship between fault features and fault types, further accurately classify and diagnose the types of rolling bearing faults.

## 2.1 Feature extraction of rotating machinery faults based on MDE

The healthy operation of rolling bearings is related to the stability and reliability of mechanical equipment. In actual complex working conditions, rolling bearings are prone to various faults. It is extremely important to detect and repair their faults timely. The detection process for rolling bearing faults is presented in Figure 1.

In Figure 1, if a rolling bearing malfunctions, it needs to be detected. The fault type is classified relying on the fault signal, so that targeted repairs can be made. Common rolling bearing failures include outer ring failure, inner ring failure, and rolling element failure. The feature extraction commonly uses traditional entropy methods like Shannon entropy, Sample Entropy (SE), Permutation Entropy (PE), etc. However, these methods have limitations in single scale analysis and low feature discrimination (Wang and Sun, 2021). Therefore, the study proposes the Multi-scale Divergence Entropy (MDE) to extract features of signals from rotating mechanical bearings. This method has the ability to fuse multi-scale features and quantify differences through multi-scale distributions to more comprehensively characterize the hierarchical structure of fault signals. MDE is based on Multi-scale Sample Entropy (MSE) and DE. The calculation of DE is as follows. For any given time series  $Z = \{z_1, z_2, \dots, z_M\}$ , assuming that the embedding dimension is  $n$ , the reconstructed phase space  $W(n)$  based on the phase space embedding theory is shown in Equation 1.

$$W(n) = \{w_1(n), w_2(n), \dots, w_{M-n+1}(n)\}$$

$$= \begin{bmatrix} z_1 & z_2 & \dots & z_n \\ z_2 & z_3 & \dots & z_{n+1} \\ \vdots & \vdots & & \vdots \\ z_{M-n+1} & z_{M-n+2} & \dots & z_M \end{bmatrix} \quad (1)$$

In Equation 1,  $M$  is the length of the data.  $w_i(n) = \{z_i, z_{i+1}, \dots, z_{i+n}\}$  is a series of spatial trajectories composed of

time series, and  $i \in (1, M - n + 1)$ . The Cosine Similarity (CS) set  $E(n)\{e_1, \dots, e_{M-n}\}$  can be obtained by calculating the CS between adjacent phase trajectories. CS is a physical quantity that measures the similarity between two trajectories, with a value range of  $[-1, 1]$ . The CS  $e_{ij}$  is shown in Equation 2.

$$e(w_i(n), w_j(n)) = \frac{\sum_{k=1}^n w_i(k) \times w_j(k)}{\sqrt{\sum_{k=1}^n w_i(k)^2} \times \sqrt{\sum_{k=1}^n w_j(k)^2}} \quad (2)$$

In Equation 2,  $w_i(n)$  and  $w_j(n)$  are two trajectories in phase space. The CS set  $E(n)$  is presented in Equation 3.

$$E(n) = \{e_1, \dots, e_{M-n}\} = \{e(w_1(n), w_2(n)), e(w_2(n), w_3(n)), \dots, e(w_{M-n}(n), w_{M-n+1}(n))\} \quad (3)$$

In Equation 3,  $E(n)$  can be used to describe the overall similarity relationship between adjacent trajectories in phase space. The distance calculation between spatial trajectories is shown in Figure 2.

In Figure 2a, if  $M - n + 1$  trajectories are written in matrix form, SE and Fuzzy Entropy (FE) calculate the distance between trajectories in the upper triangular matrix region without diagonals, which means that the process traverses the distance between multiple pairs of trajectories except for self-matching. From Figure 2b, when calculating the distance between trajectories, DE only calculates the elements in the upper row of the diagonal, indicating that it only calculates the distance between adjacent trajectories, rather than traversing the distance between all trajectories. Subsequently,  $J(J_1, J_2, \dots, J_s)$  is used to represent the range  $[-1, 1]$  of  $s$  small cells, and the state probability  $Q(q_1, q_2, \dots, q_s)$  of each small cell is calculated. Based on  $Q$ , the DE can be obtained, as shown in Equation 4.

$$DE(n, s) = -\frac{1}{\ln s} \sum_{k=1}^s q_k \ln q_k \quad (4)$$

In Equation 4,  $q_k$  is the probability of the CS falling within the  $k$ -th interval. MSE divides the time series into several multi-scale sequences through a coarse-grained process. For time series  $Z = \{z_1, z_2, \dots, z_M\}$ , the multi-scale time series is shown in Equation 5.

$$c_j^\alpha = \frac{1}{\alpha} \sum_{i=(j-1)\alpha+1}^{j\alpha} z_i, \quad 1 \leq j \leq M/\alpha \quad (5)$$

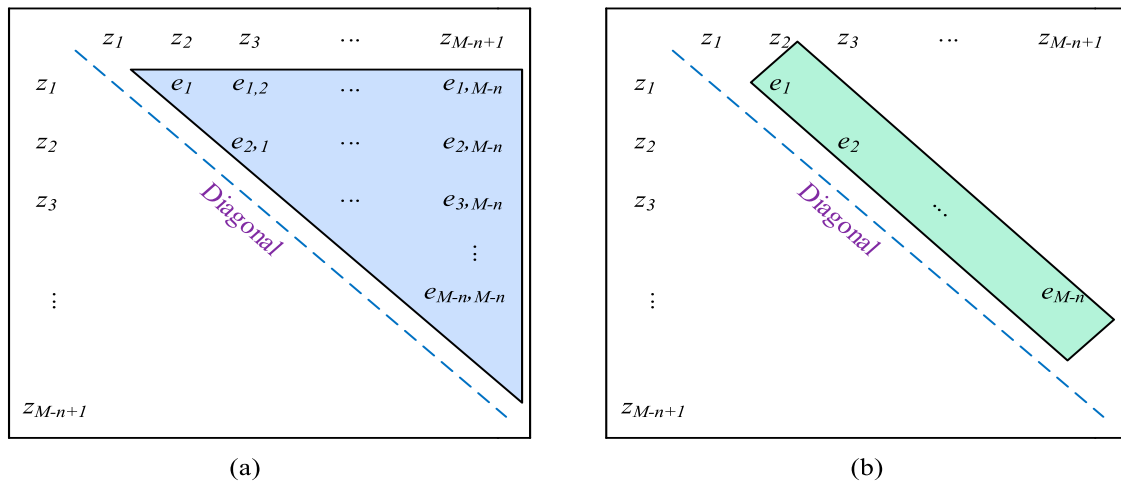


FIGURE 2  
Calculation method for track distance. (a) Sample entropy and fuzzy entropy; (b) Divergence entropy.

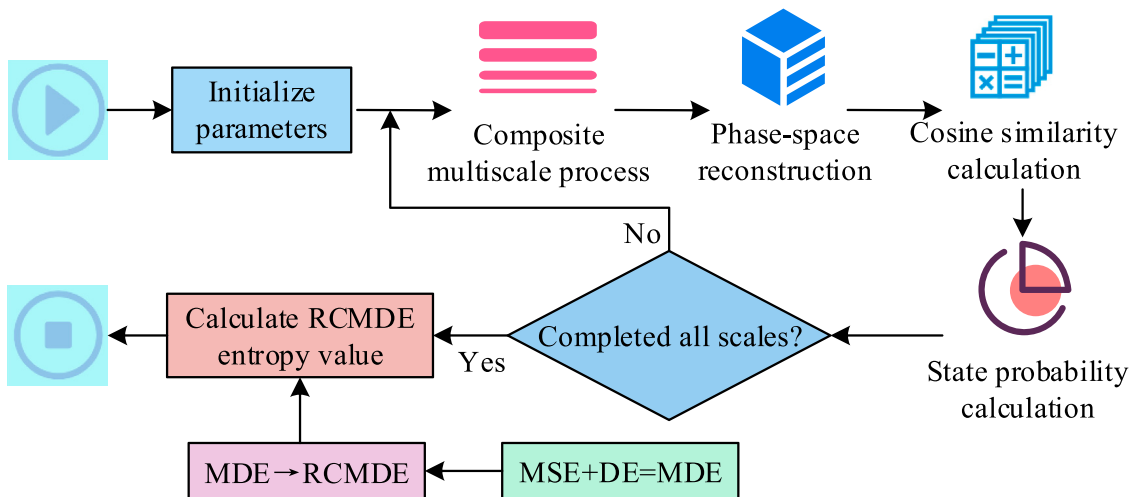


FIGURE 3  
Algorithm flowchart of RCMDE.

In Equation 5,  $\alpha$  is the scale factor, with a positive integer. The SE value corresponding to each multi-scale time series is shown in Equation 6.

$$MSE(Z, \alpha, n, p) = SE(c_j^\alpha, n, p) \quad (6)$$

In Equation 6,  $p$  represents the tolerance parameter. The study combines MSE with DE to propose MDE, which divides time series into multi-scale time series through coarsening. Multi-scale time series use windows of various lengths to perform sliding average processing on the original time series. The multi-scale time series of MDE for time series  $Z = \{z_1, z_2, \dots, z_M\}$  is shown in Equation 7.

$$b_j^\alpha = \frac{1}{\alpha} \sum_{i=j}^{j\alpha} z_i, \quad 1 < j < M - \alpha + 1 \quad (7)$$

The calculation for MDE obtained from  $b_j^\alpha$  is shown in Equation 8.

$$MDE(Z, \alpha, n, s) = DE(b_j^\alpha, n, s) \quad (8)$$

In response to the entropy instability of MDE at high scales, further improvements are made to MDE. The RCMDE is proposed. RCMDE accurately estimates the state probability distribution in phase space, thereby improving the stability and reliability of MDE in large-scale analysis. Equation 9 displays the probability of the CS falling between each small cell  $J$ .

$$q_k^\alpha = \frac{1}{\alpha} \sum_{j=1}^{\alpha} q_{k,j}^\alpha \quad (9)$$

In Equation 9,  $q_{k,j}^\alpha$  represents the probability that the CS falls in the  $k$ -th small cell in the  $j$ -th multi-scale time series. The entropy value of RCMDE can be obtained from  $q_k^\alpha$ , as shown in Equation 10.

$$RCMDE(Z, n, s) = -\frac{1}{\ln s} \sum_{k=1}^s q_k^\alpha \ln q_k^\alpha \quad (10)$$

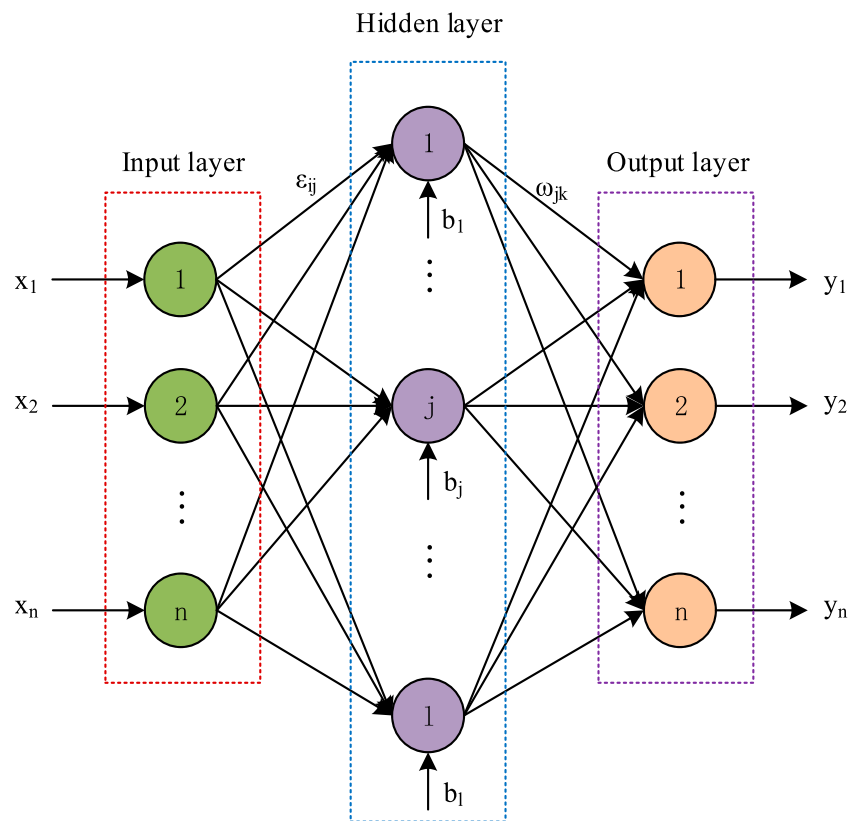


FIGURE 4  
Schematic of ELM.

Figure 3 displays the RCMDE.

Figure 3 shows the detailed calculation process of RCMDE. Firstly, the original time series is divided and averaged through a composite multi-scale process to obtain a multi-scale time series. Multi-scale time series are reconstructed in phase space. The 1D time series is transformed into phase space to analyze its dynamic characteristics. Next, the CS between adjacent trajectories in phase space is calculated to measure the similarity between trajectories. The state probability is calculated based on the CS calculation results. Finally, the entropy value of RCMDE is calculated through a fine composite multi-scale process to accurately analyze the multi-scale characteristics in rotating machinery.

## 2.2 Fault recognition of rotating machinery based on ELM

After using the RCMDE algorithm for fault signal feature extraction, the type of fault is identified. ELM is selected to complete the fault identification task. ELM is based on feedforward neural networks. It is extensively applied in regression, clustering, classification, feature learning, etc (Yu et al., 2023). Figure 4 presents the ELM (Mohan and Senthilkumar, 2022).

In Figure 4, the ELM contains Input Layer (IL), Hidden Layer (HL), and Output Layer (OL). The IL nodes are connected to the HL nodes through weights  $\varepsilon_{ij}$ , which determine the influence of the

input data on the HL nodes. The HL nodes perform nonlinear transformations on input data, which is a key part of implementing complex mappings in networks. Each node has a bias  $b_j$ . The HL nodes are linked to the OL nodes through weights  $\omega_{jk}$ . The OL calculates the final output based on the output of the HL and these weights. Firstly, an ELM network is set with a HL neuron count of  $l$  and an activation function of  $f(x)$ . A training set  $D = \{(u_i, v_i)\}_{i=1}^N$  containing  $N$  different samples is input into the network to train. At this point, the ELM is presented in Equation 11.

$$A\omega = V \quad (11)$$

In Equation 11,  $A$  signifies the HL output matrix.  $\omega$  signifies the weight matrix between the HL and the OL.  $V$  signifies the output matrix. The definitions of  $A$ ,  $\omega$ , and  $V$  are displayed in Equation 12.

$$A = [a_1^T a_2^T \cdots a_N^T]^T = \begin{bmatrix} f(\omega_1, b_1, u_1) & \cdots & f(\omega_n, b_n, u_1) \\ \vdots & \ddots & \vdots \\ f(\omega_1, b_1, u_N) & \cdots & f(\omega_n, b_n, u_N) \end{bmatrix}_{N \times l} \quad (12)$$

$$\omega = [\omega_1^T \omega_2^T \cdots \omega_n^T]^T$$

$$V = [v_1^T v_2^T \cdots v_N^T]^T$$

In Equation 12,  $a_i$  represents the output vector of HL corresponding to the input vector  $u_i$ .  $f(\omega_j, b_j, u_i)$  represents the output value obtained by the activation function  $f(x)$  after the  $i$ -th input vector  $u_i$  passes through the  $j$ -th HL neuron.  $\omega_j$  signifies the corresponding weight.  $b_j$  signifies the corresponding bias. Usually, the HL output matrix  $A$  and the

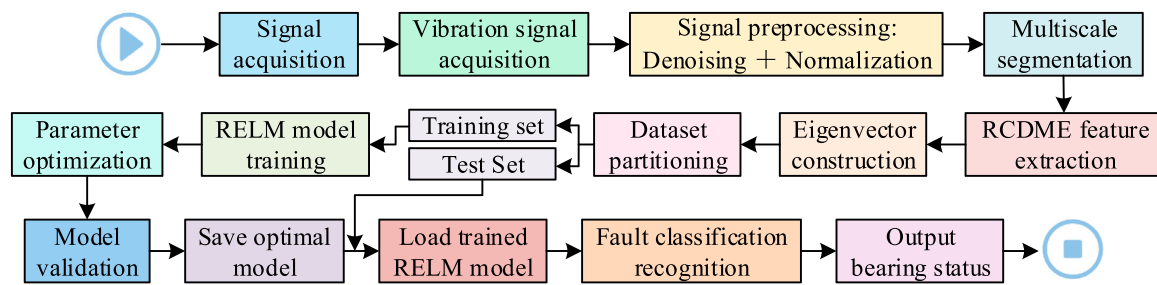


FIGURE 5  
Fault identification flowchart of RCMDE-RELM.

target output matrix  $V$  in Equation 13 are known and fixed. It is only necessary to use generalized inverse theory to solve the minimum norm least squares solution of the weight matrix  $\omega$ . However, the ELM algorithm needs to solve its generalized inverse when  $A^T A$  or  $A A^T$  is non-singular. Otherwise, pathological matrix problems may occur. The study employs RELM to address this issue. The definition of the objective function with regularization constraints is shown in Equation 13.

$$\min_{\omega} \left\{ L_{reg} = \frac{\lambda}{2} \|\varphi_i\|^2 + \frac{1}{2} \|\omega\|^2 \right\} \quad (13)$$

$$a(u_i)^T \omega = v_i^T - \varphi_i^T, i = 1, \dots, N$$

In Equation 13,  $\lambda$  is the regularization parameter.  $\varphi_i$  represents the training error of the  $i$ -th sample. According to KKT theory, if the regularization constrained optimization problem is changed into a dual optimization problem, the objective function is shown in Equation 14.

$$\min_{\omega, \alpha, \varphi} \left\{ L_{dual} = \frac{1}{2} \|\omega\|^2 + \frac{\lambda}{2} \|\varphi_i\|^2 - \sum_{i=1}^N \sum_{j=1}^N \alpha_{ij} (a(u_i)^T \omega_j - v_{ij} + \varphi_{ij}) \right\} \quad (14)$$

In Equation 14,  $\alpha$  is the Lagrange multiplier. Then, based on the optimization conditions of KKT theory, Equation 15 is obtained.

$$\begin{cases} \frac{\partial L_{dual}}{\partial \omega_j} = 0 \rightarrow \omega_j = \sum_{i=1}^N \alpha_i a(u_i)^T \rightarrow \omega = A^T \alpha \\ \frac{\partial L_{dual}}{\partial \varphi_i} = 0 \rightarrow \alpha_i = \lambda \varphi_i \\ \frac{\partial L_{dual}}{\partial \alpha_i} = 0 \rightarrow a(u_i)^T \omega_j - v_i + \varphi_i = 0 \end{cases} \quad (15)$$

The expressions for  $V$  and  $\omega$  can be obtained by solving the above equations, as shown in Equation 16.

$$\begin{cases} V = \left( \frac{I}{\lambda} + A^T A \right) \alpha \\ \omega = A^T \left( \frac{I}{\lambda} + A^T A \right)^{-1} V \end{cases} \quad (16)$$

In Equation 16,  $I$  represents the identity matrix. The output of RELM is displayed in Equation 17.

$$g(u) = (a(u)^T \omega)^T = \left( a(u)^T A^T \left( \frac{I}{\lambda} + A^T A \right)^{-1} V \right)^T \quad (17)$$

RELM inherits the advantages of ELM, such as high computational efficiency and fast training speed, and avoids the pathological matrix and over-fitting by introducing regularization constraints. RCMDE extracts features and classifies the types of faults using the RELM. The fault identification flowchart based on RCMDE-RELM is shown in Figure 5.

In Figure 5, the vibration signals in different states are collected. RCMDE extracts features from the signals. Next, 80% of the extracted features is to train, 20% to test, and the training set will be trained using RELM. Finally, the test set is input into the trained RELM to achieve fault classification and recognition.

## 3 Results

The study verifies the influence of different feature extraction algorithms and different fault recognition models on the rolling bearing fault classification through comparative experiments. By determining appropriate scale factors and activation functions, the accuracy of fault identification accuracy can be improved, thereby better identifying faults in rotating machinery.

### 3.1 Analysis of fault feature extraction results based on RCMDE

In the research on fault recognition in rotating machinery, to fully verify its performance, other advanced methods are compared and validated. The study selected the Paderborn University Bearing Dataset (PU Dataset) to evaluate the classification performance of different feature extraction models (Lessmeier et al., 2016). Four distinct bearing conditions were chosen from the dataset for validation: normal condition, outer race fault, inner race fault, and rolling element fault. These faults were artificially induced through electro-erosion methods such as drilling and pitting. The data specifically originates from bearings of model 6,203, collected under operating conditions of 17,000 rpm and a sampling frequency of 64 kHz. For each condition, 60 samples were randomly selected, with each sample consisting of 2,048 data points. The total sample pool (4 classes  $\times$  60 samples = 240 samples) was randomly divided into training and testing sets in an 8:2 ratio using stratified sampling



TABLE 1 Detailed parameters for testing bearings.

Bearing model	6,203
Rotational speed	17000r/min
Sampling frequency	64k Hz
Inner diameter/outer diameter	17 mm/40 mm
Rolling element diameter	6.75 mm
Number of rolling elements	8
Weight	0.067 kg

principles, ensuring the proportion of each class in both sets matched the overall distribution. Table 1 presents the detailed data.

The study first verifies the optimal scale factor value through experiments, and then compares the RCMDE algorithm with other feature extraction algorithms. The scale factor is an important parameter of RCMDE, and different scale factors have multiple impacts on RCMDE. To verify the effect of scale factor on the classification accuracy, the study sets the range of scale factor values from 1 to 20 while keeping the embedding dimension and symbol number unchanged. The RELM is combined for experimental verification on PU Dataset. Figure 6 displays the RCMDE value and classification accuracy.

In Figure 6a, when the bearing was in a normal state, the value of the scale factor had little effect on the entropy value of the final RCMDE. When the bearing was in a fault state, its entropy value decreased with the increase of the scale factor value, and ultimately decreased to below 0.5. As the value increased, the entropy value of the bearing signal at the rolling element fault decreased the fastest, and decreased to 0.21 when the scale factor value was 20. In Figure 6b, when the scale factor was set to 1, the classification accuracy for bearing faults was only 71.35%. When the scale factor was set to 5, the accuracy was improved to 85.68%. When the scale factor was set to 10 and 15, the accuracy was 90.41% and 94.33%,

respectively. When the scale factor was 20, the model had the highest classification accuracy for fault types, reaching 98.79%. The study considered both entropy and classification accuracy, and ultimately set the scale to 20. Next, the RCMDE algorithm is compared with three commonly used feature extraction algorithms, namely, Daubechies wavelet (DB wavelet), DMeyer wavelet (DM wavelet), and Variational Mode Decomposition (VMD-7), and validated experimentally on the PU dataset using RELM. The average classification accuracy is presented in Figure 7.

In Figure 7a, the DB-RELM model had low accuracy in classifying various types of faults, only about 68%. The average classification accuracy of DM-RELM for outer ring faults was 73.25%, inner ring faults was 74.47%, and rolling element faults was 75.19%. The VMD-7-RELM for three types of faults exceeded than that of the DM-RELM. Compared with the VMD-7-RELM model, the average classification accuracy of RCMDE-RELM for these three types of faults was 4.57%, 7.37%, and 8.88% higher, respectively, at 85.43%, 87.58%, and 90.07%. In Figure 7b, the average classification accuracy of the RCMDE-RELM added with the increase of training samples. After reaching 40 samples, its classification accuracy stabilized at around 88%. The VMD-7-RELM model had low classification accuracy when the sample size was small, with an average classification accuracy of 77.51% when the sample size was 50, significantly lower than that of the RCMDE-RELM. The designed method has high classification accuracy for faults and low requirements for sample size. The study also validates feature extraction methods on the Ionosphere sample set, Cmc sample set, and Wine sample set. The classification accuracy and feature selection results of the RELM model based on four different feature extraction algorithms on each sample set are compared, as shown in Figure 8.

In Figure 8a, on the Ionosphere sample set, the DB-RELM model used 12 features and had a low classification accuracy of 81.84%. The DM-RELM model used 19 features and had a classification accuracy of 86.79%. The VMD-7-RELM model used 15 features and had a classification accuracy of 89.06%. The RCMDE-RELM model took

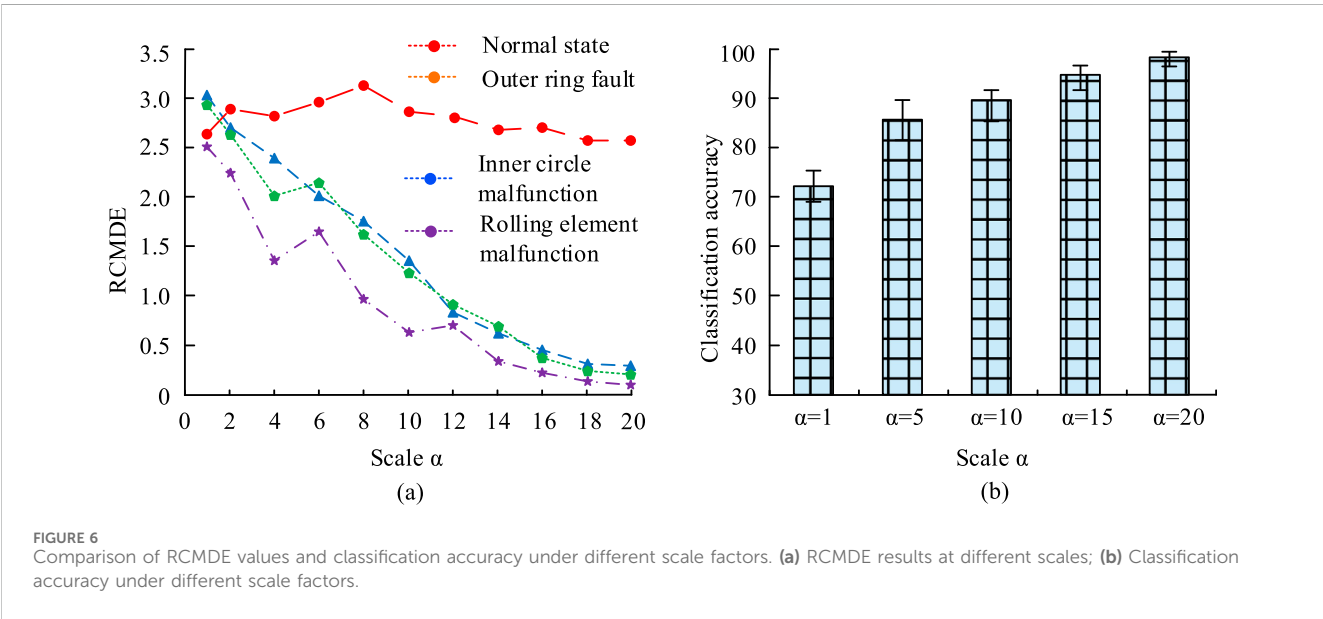


FIGURE 6 Comparison of RCMDE values and classification accuracy under different scale factors. (a) RCMDE results at different scales; (b) Classification accuracy under different scale factors.

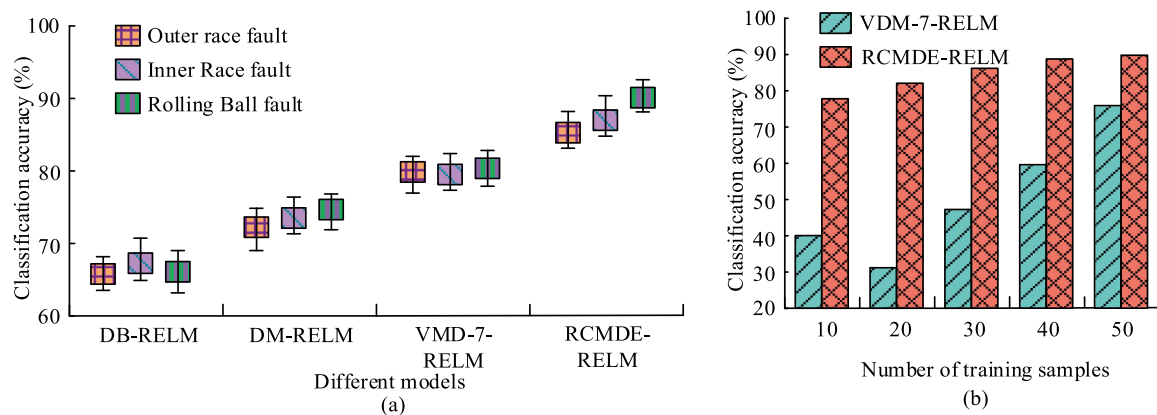


FIGURE 7 The fault classification accuracy of models with different feature extraction algorithms. (a) Average classification accuracy under different models; (b) Comparison of RCMDE and VMD-7 under Different Training Sample Sizes.

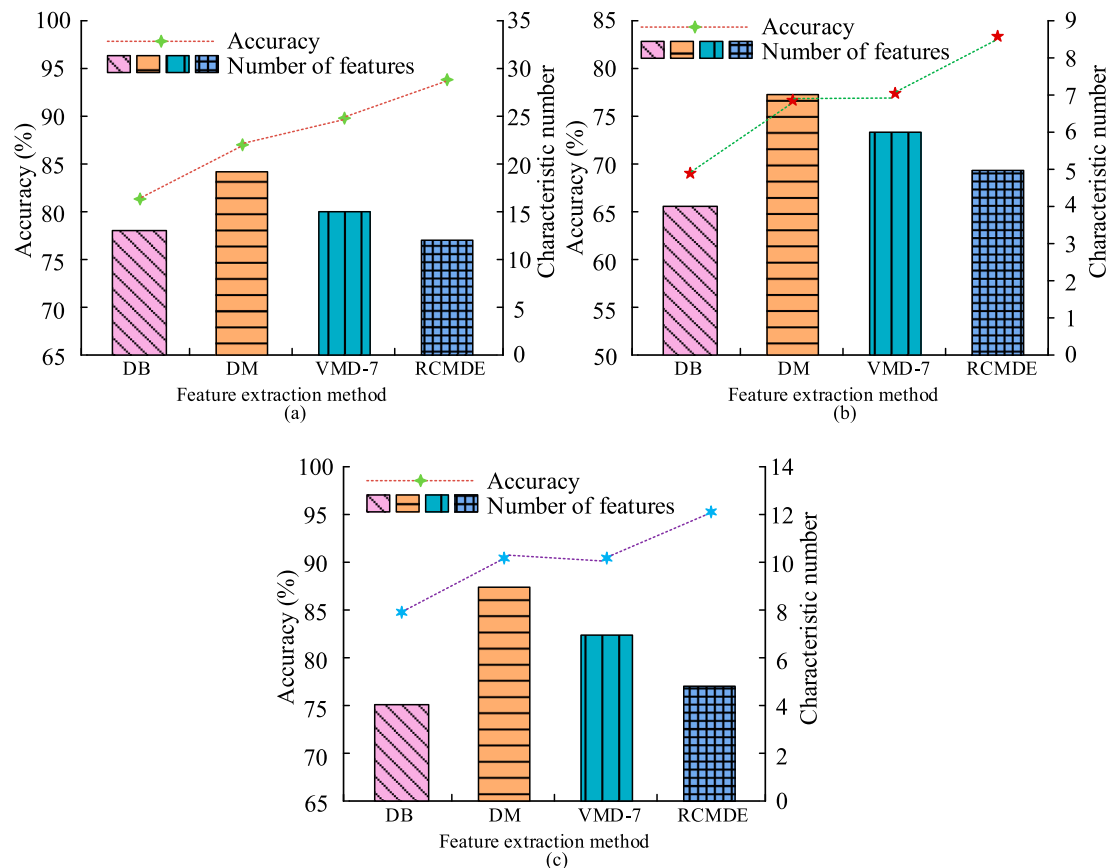


FIGURE 8 Comparison of experimental results of models based on different feature extraction algorithms on various sample sets. (a) Comparison results of Ionosphere sample set experiments; (b) Comparison results of Cmc sample set experiments; (c) Comparison results of Wine sample set experiments.

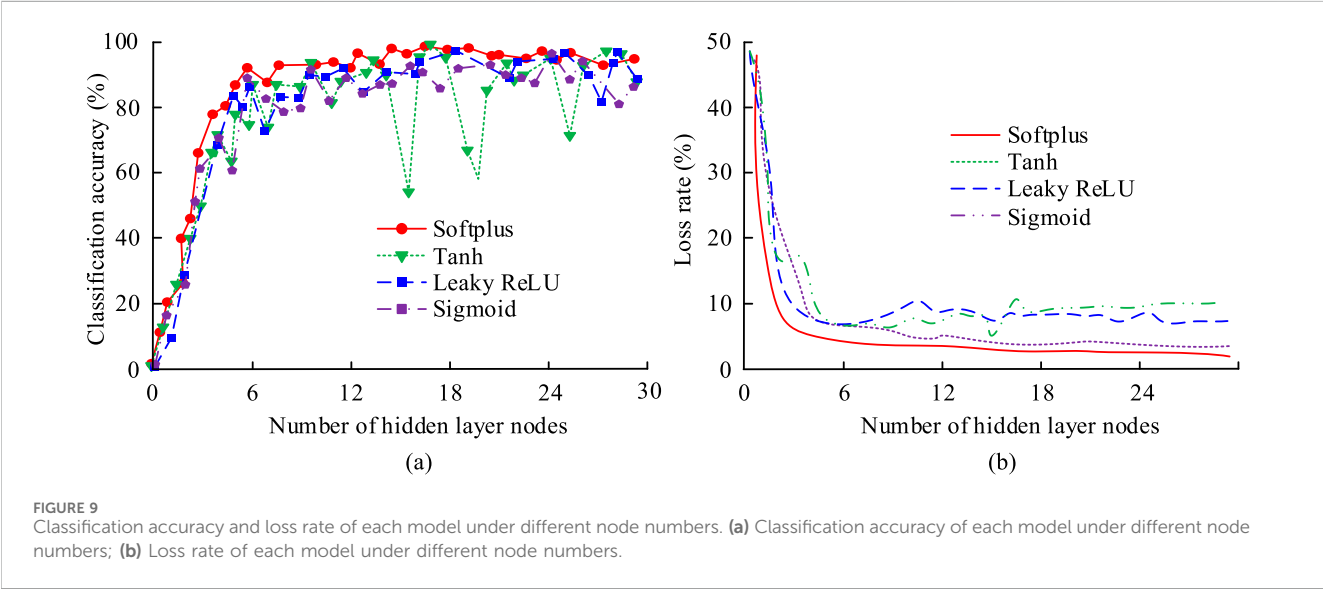
fewer features and had the highest classification accuracy of 93.42%. In Figure 8b, on the Cmc sample set, the DB-RELM model selected 4 features and had a classification accuracy of 68.75%. The DM-RELM model and VMD-7-RELM model selected 7 and 6 features respectively, with classification accuracy of 76.99% and 77.13%,

respectively. The RCMDE-RELM selected 5 features and had the highest classification accuracy of 83.83%. In Figure 8c, on the Wine sample set, the classification accuracy was relatively high. The DB-RELM used 4 features and had a classification accuracy of 84.58%. The DM-RELM and VMD-7-RELM used 9 and 7 features



TABLE 2 Detailed classification accuracy of different entropy-based methods on PU dataset.

Method/Bearing state	Normal (%)	Outer race fault (%)	Inner race fault (%)	Rolling element fault (%)	Overall accuracy (%)
RCMSE-RELM	94.50 ± 0.45	80.15 ± 1.85	82.37 ± 1.65	84.26 ± 1.55	85.32 ± 1.50
RCMPE-RELM	95.80 ± 0.40	81.90 ± 1.70	84.55 ± 1.50	86.18 ± 1.40	87.11 ± 1.25
MDE-RELM	96.20 ± 0.35	83.95 ± 1.40	86.41 ± 1.25	88.95 ± 1.10	88.88 ± 1.05
RCMDE-RELM	98.25 ± 0.25	85.43 ± 1.20	87.58 ± 1.05	90.07 ± 0.95	90.33 ± 0.85



respectively, with classification accuracy of 91.17% and 90.02%, respectively. The RCMDE-RELM used 5 features and achieved a classification accuracy of 95.79%. The proposed method requires fewer features for fault identification and has higher accuracy in classifying various faults of rolling bearings.

To more fairly evaluate the performance advantages of RCMDE in the entropy feature family, the study further compared it with several advanced entropy based feature extraction methods, including fine composite multi-scale sample entropy (RCMSE), fine composite multi-scale permutation entropy (RCMPE), and basic multi-scale divergence entropy (MDE). The parameter settings for all entropy methods remain consistent: embedding dimension  $m = 2$ , tolerance  $r = 0.15 \times$  standard deviation, and scale factor  $\tau = 20$ . The comparison of the results of 10 experiments conducted on the PU dataset using RELM combined with various methods is shown in Table 2.

According to Table 2, the RCMDE-RELM model proposed in the study has better classification accuracy than the comparative methods in all four bearing states, and achieved the highest overall classification accuracy ( $90.33\% \pm 0.85\%$ ). Moreover, while achieving optimal accuracy, the model also has the smallest standard deviation among all methods, indicating that RCMDE-RELM has high stability. Specifically, compared to the basic MDE, RCMDE has improved accuracy in various fault types with smaller fluctuations, verifying the improvement of feature stability by the fine composite process. Meanwhile, compared with RCMSE and RCMPE, the superior performance of RCMDE also reflects the advantage of

divergence entropy in quantifying fault characteristics. This result fully demonstrates that RCMDE can extract more discriminative and robust features from vibration signals.

### 3.2 Analysis of fault identification results based on RCMDE-RELM

To verify the effectiveness of the RCMDE-RELM for fault identification, the activation function of the RELM is first determined through experiments. The study selects four different activation functions and conducts experimental verification on PU Dataset in conjunction with RCMDE. The classification accuracy and loss rate results of the RCMDE-RELM model based on four activation functions under different node numbers are shown in Figure 9.

In Figure 9a, in multiple experiments, as the node increased, the classification accuracy also added. The classification accuracy of models based on Sigmoid and Leaky ReLU was similar, with 82.51% and 83.48%, respectively. The Tanh-based model had high classification accuracy but significant fluctuations, indicating low model stability. The classification accuracy of the Softplus was 93.53%, with relatively smooth fluctuations. In Figure 9b, the Tanh-based model only converged when the number of nodes increased to 18, with a loss rate of 10.17%. The models based on Leaky ReLU and Sigmoid gradually stabilized after the number of nodes increased to 12, reaching 8.36% and 5.11%, respectively. The model based on Softplus converged to 3.47% after 6 nodes.

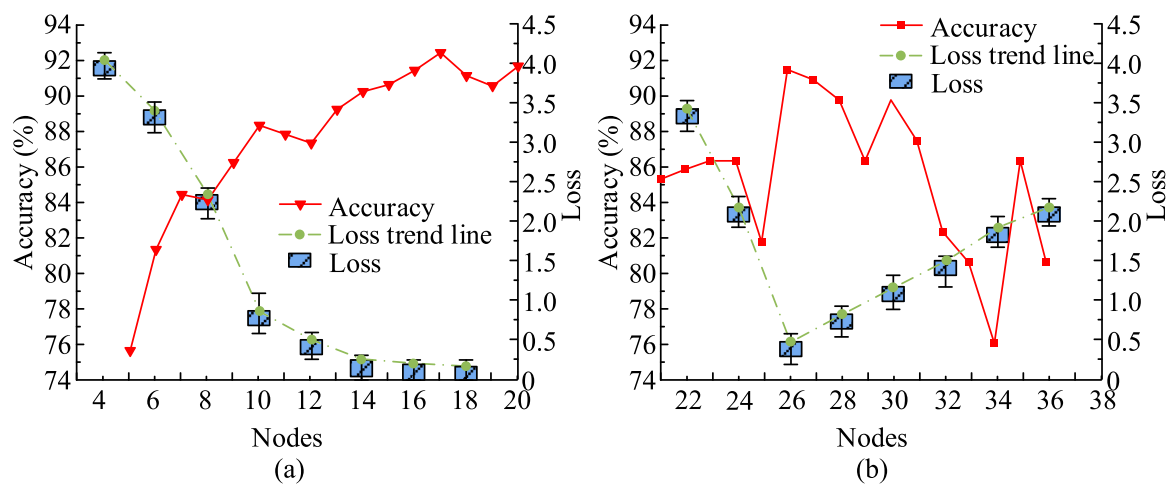


FIGURE 10  
Average classification accuracy and loss under different numbers of neurons. (a) Low node count; (b) High node count.

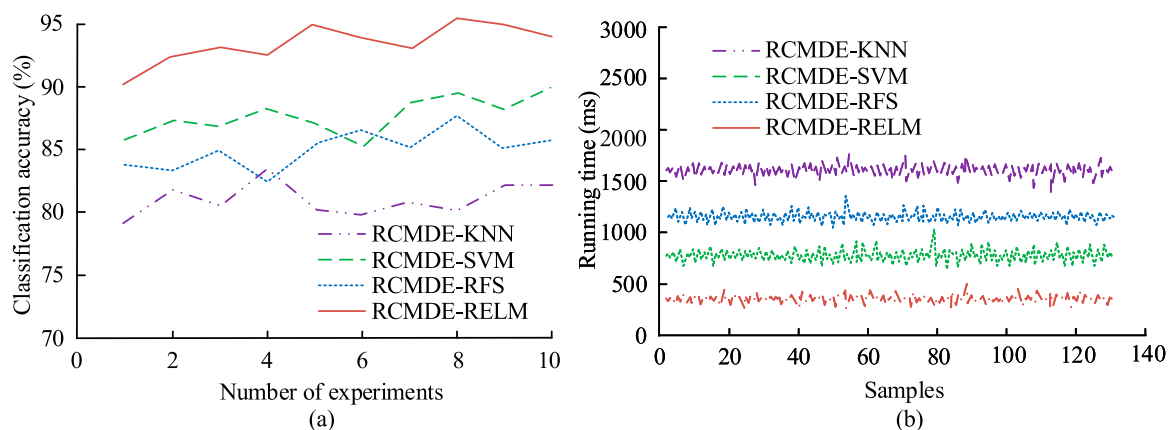


FIGURE 11  
Classification accuracy and runtime of each model. (a) Classification accuracy of each model; (b) The runtime of each model.

Therefore, the study selects the Softplus function as the activation function for RELM. After determining the activation function, the quantity of neurons is determined using the five-fold cross test method. The average classification accuracy and loss rate of 10 experiments are taken as evaluation results, as presented in Figure 10.

From Figure 10a, the average classification accuracy increased with the node when the neuron was 0–20, and reached the highest classification accuracy of  $92.31\% \pm 1.21\%$  when the node was 17. The average loss rate gradually decreased with the increase of the node, and dropped to  $74.25\% \pm 1.43\%$  when the number of nodes was 18. In Figure 10b, the average classification accuracy fluctuated between 20 and 36 nodes, reaching a maximum of  $91.89\% \pm 1.18\%$  at 26 nodes. The average loss rate showed a downward trend in the range of 20–26 and an upward trend in the range of 26–36. Therefore, the study sets the number of neurons to 17. Finally, the study compares RELM with several other commonly used methods for fault classification and conducts experimental verification on the PU Dataset using RCMDE. The selected comparison methods are

K-Nearest Neighbors (KNN), SVM, and Random Forest System (RFS), with 10 experiments conducted for each method. Figure 11 presents the classification accuracy and running time.

In Figure 11a, the average classification accuracy of the RCMDE-KNN was relatively low, only  $81.07\% \pm 2.03\%$ . The RCMDE-SVM was  $84.76\% \pm 1.78\%$ . The RCMDE-RFS slightly exceeded that of the other two categories, at  $87.34\% \pm 1.49\%$ . The RCMDE-RELM had the highest average classification accuracy, reaching  $93.98\% \pm 0.94\%$ . According to Figure 11a, the running time of RCMDE-KNN, RCMDE-SVM, and RCMDE-RFS was 1600 m, 1200 m, and 750 m, respectively. The running time of the RCMDE-RELM was significantly lower than the other three types of models, only 400 m. The designed method has better model performance and higher accuracy in fault classification of rolling bearings. To verify the generalization ability of the proposed method, further experiments were conducted on the bearing dataset of Case Western Reserve University (CWRU) (Smith and Randall, 2015). This dataset contains multiple types of faults (inner ring, outer ring, rolling

TABLE 3 Detailed classification performance of each model on the CWRU bearing dataset.

Method/Bearing state	Normal (%)	Outer race fault (%)	Inner race fault (%)	Rolling element fault (%)	Overall accuracy (%)	Runtime (ms)
RCMDE-KNN	85.2 ± 1.5	76.3 ± 2.5	75.1 ± 2.8	77.1 ± 2.3	78.45 ± 2.15	520
RCMDE-SVM	88.5 ± 1.3	80.4 ± 2.2	79.2 ± 2.4	82.2 ± 1.9	82.60 ± 1.90	650
RCMDE-RFS	90.3 ± 1.1	83.5 ± 1.8	82.4 ± 2.0	85.3 ± 1.6	85.72 ± 1.55	850
RCMDE-RELM	95.0 ± 0.8	91.2 ± 1.3	90.8 ± 1.5	90.0 ± 1.4	91.50 ± 1.10	420

element faults) and different load conditions (0HP, 1HP, 2HP, 3HP). Similarly, RCMDE is used for feature extraction, RELM is used for classification, and training testing is split into 80%–20%. The experimental results are shown in Table 3.

According to Table 3, the proposed RCMDE-RELM model not only achieved the highest overall classification accuracy (91.50% ± 1.10%) on the CWRU dataset, but also maintained a stable recognition accuracy of over 90% for all single fault categories and normal states, significantly better than other comparative models. At the same time, the standard deviation of the RCMDE-RELM model is minimized, which proves that the method has excellent stability. In terms of operational efficiency, the RCMDE-RELM model only takes 420 m, which is much lower than other models and achieves the best balance between accuracy and speed.

## 4 Conclusion and discussion

Aiming at the difficulty in extracting fault features and low recognition accuracy of rolling bearings, the RCMDE and RELM were proposed. In the feature extraction experiment, the RCMDE scale factor was first determined. As the scale factor increased, the entropy value decreased. When the scale was 20, the entropy value dropped below 0.5, and the classification accuracy reached 98.79%. Therefore, the scale was set to 20. Compared with other algorithms, DB-RELM and DM-RELM performed poorly. RCMDE-RELM had an average classification accuracy of 4.57%, 7.37%, and 8.88% higher than VMD-7-RELM for various types of faults, and could still maintain high accuracy even with a small sample size. This result was similar to the conclusion obtained by Wang S et al., which may be due to the relatively simple calculation process of RCMDE, without requiring prior knowledge (Wang et al., 2023). Meanwhile, RCMDE had higher sensitivity to different types of fault characteristics and better noise resistance. When verifying the effectiveness of the RELM model, Softplus was determined as the activation function with 17 neurons. The model had a classification accuracy of 92.31% ± 1.21% and the lowest loss rate. On the PU Dataset, compared with other classification models, RCMDE-KNN, RCMDE-SVM, and RCMDE-RFS have lower average classification accuracy and longer running time. The average classification accuracy of RCMDE-RELM is 93.98% ± 0.94%, running for only 400 m. On the CWRU dataset, this method also achieved a high accuracy of 91.50% ± 1.10%, demonstrating its good generalization ability, which is similar to the conclusion obtained by Liu (2023). The reason for obtaining this result may be that RELM has better processing ability for high-dimensional data, faster training speed, and can flexibly adjust parameters to adapt to different tasks.

In summary, the RCMDE-RELM can effectively achieve the fault recognition and classification of rotating machinery bearings, providing new theoretical and technical support for fault prevention of rotating machinery in the industrial field. However, the study only take a dataset for experimental verification, and the fault signal data in real environments is extremely complex. Therefore, in the future, fault signals can be collected from field scenarios to apply the method from theory to practice.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

YS: Conceptualization, Data curation, Investigation, Methodology, Writing – original draft, Writing – review and editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The research is supported by 2024 Shanxi Province Higher Education Science and Technology Innovation Project, Research on Intelligent Fault Diagnosis Method for Rudder System Based on Machine Learning (2024L535).

## Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Dhini, A., Surjandari, I., Kusumoputro, B., and Kusiak, A. (2022). Extreme learning machine–radial basis function (ELM-RBF) networks for diagnosing faults in a steam turbine. *J. Industrial Prod. Eng.* 39 (7), 572–580. doi:10.1080/21681015.2021.1887948
- Fan, C., Peng, Y., Shen, Y., Guo, Y., Zhao, S., Zhou, J., et al. (2024). Variable scale multilayer perceptron for helicopter transmission system vibration data abnormality beyond efficient recovery. *Eng. Appl. Artif. Intell.* 133, 108184. doi:10.1016/j.engappai.2024.108184
- Gangsar, P., Pandey, R. K., and Chouksey, M. (2021). Unbalance detection in rotating machinery based on support vector machine using time and frequency domain vibration features. *Noise and Vib. Worldw.* 52 (4-5), 75–85. doi:10.1177/0957456521999836
- Jiao, J., Li, H., Zhang, T., and Lin, J. (2022). Source-free adaptation diagnosis for rotating machinery. *IEEE Trans. Industrial Inf.* 19 (9), 9586–9595. doi:10.1109/tii.2022.3231414
- Kafantaris, E., Lo, T. Y. M., and Escudero, J. (2022). Stratified multivariate multiscale dispersion entropy for physiological signal analysis. *IEEE Trans. Biomed. Eng.* 70 (3), 1024–1035. doi:10.1109/tbme.2022.3207582
- Lessmeier, C., Kimotho, J. K., Zimmer, D., and Sestro, W. (2016). Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: a benchmark data set for data-driven classification PHM society European conference.
- Li, S., Peng, Y., and Bin, G. (2023). Prediction of wind turbine blades icing based on CJBm with imbalanced data. *IEEE Sensors J.* 23 (17), 19726–19736. doi:10.1109/jsen.2023.3296086
- Li, Y., Jiao, S., Deng, S., Geng, B., and Li, Y. (2024). Refined composite variable-step multiscale multimapping dispersion entropy: a nonlinear dynamical index. *Nonlinear Dyn.* 112 (3), 2119–2137. doi:10.1007/s11071-023-09145-8
- Li, S., Peng, Y., Shen, Y., Zhao, S., Shao, H., Bin, G., et al. (2024a). Rolling bearing fault diagnosis under data imbalance and variable speed based on adaptive clustering weighted oversampling. *Reliab. Eng. and Syst. Saf.* 244, 109938. doi:10.1016/j.res.2024.109938
- Li, S., Peng, Y., Bin, G., Shen, Y., Guo, Y., Li, B., et al. (2024b). Research on bearing fault diagnosis method based on cjbm with semi-supervised and imbalanced data. *Nonlinear Dyn.* 112 (22), 19759–19781. doi:10.1007/s11071-024-10073-4
- Liu, G. (2023). The application of fault diagnosis techniques and monitoring methods in building electrical systems–based on ELM algorithm. *J. Meas. Eng.* 11 (4), 388–404. doi:10.21595/jme.2023.23357
- Liu, D., Cui, L., and Wang, H. (2023). Rotating machinery fault diagnosis under time-varying speeds: a review. *IEEE Sensors J.* 23 (24), 29969–29990. doi:10.1109/jsen.2023.3326112
- Liu, X., Zhang, Z., Meng, F., and Zhang, Y. (2023). Fault diagnosis of wind turbine bearings based on CNN and SSA–ELM. *J. Vib. Eng. and Technol.* 11 (8), 3929–3945. doi:10.1007/s42417-022-00793-5
- Mohan, V., and Senthilkumar, S. (2022). IoT based fault identification in solar photovoltaic systems using an extreme learning machine technique. *J. Intelligent and Fuzzy Syst.* 43 (3), 3087–3100. doi:10.3233/jifs-220012
- Purohit, J., and Dave, R. (2023). Leveraging deep learning techniques to obtain efficacious segmentation results. *Archives Adv. Eng. Sci.* 1 (1), 11–26. doi:10.47852/bonviewaas32021220
- Smith, W. A., and Randall, R. B. (2015). Rolling element bearing diagnostics using the case Western reserve university data: a benchmark study. *Mech. Syst. signal Process.* 64, 100–131. doi:10.1016/j.ymssp.2015.04.021
- Tama, B. A., Vania, M., Lee, S., and Lim, S. (2023). Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals. *Artif. Intell. Rev.* 56 (5), 4667–4709. doi:10.1007/s10462-022-10293-3
- Tao, H., Qiu, J., Chen, Y., Stojanovic, V., and Cheng, L. (2023). Unsupervised cross-domain rolling bearing fault diagnosis based on time-frequency information fusion. *J. Frankl. Inst.* 360 (2), 1454–1477. doi:10.1016/j.jfranklin.2022.11.004
- Wang, Z., and Sun, Y. J. (2021). Research on rolling bearing fault feature extraction based on entropy feature. *Ann. Math. Phys.* 4 (1), 066–073. doi:10.17352/amp.000025
- Wang, X., Si, S., and Li, Y. (2021). Variational embedding multiscale diversity entropy for fault diagnosis of large-scale machinery. *IEEE Trans. Industrial Electron.* 69 (3), 3109–3119. doi:10.1109/tie.2021.3063979
- Wang, X., Si, S., and Li, Y. (2022). Hierarchical diversity entropy for the early fault diagnosis of rolling bearing. *Nonlinear Dyn.* 108 (2), 1447–1462. doi:10.1007/s11071-021-06728-1
- Wang, S., Li, Y., Si, S., and Noman, K. (2023). Enhanced hierarchical symbolic sample entropy: efficient tool for fault diagnosis of rotating machinery. *Struct. Health Monit.* 22 (3), 1927–1940. doi:10.1177/14759217221116417
- Xie, J., and Yao, T. (2021). Quantified assessment of internal short-circuit state for 18 650 batteries using an extreme learning machine-based pseudo-distributed model. *IEEE Trans. Transp. Electrification* 7 (3), 1303–1313. doi:10.1109/tte.2021.3052579
- Yang, R., Xiong, R., Shen, W., and Lin, X. (2021). Extreme learning machine-based thermal model for lithium-ion batteries of electric vehicles under external short circuit. *Engineering* 7 (3), 395–405. doi:10.1016/j.eng.2020.08.015
- Yu, S., Tan, W., Zhang, C., Tang, C., Cai, L., and Hu, D. (2023). RETRACTED: power transformers fault diagnosis based on a meta-learning approach to kernel extreme learning machine with opposition-based learning sparrow search algorithm. *J. Intelligent and Fuzzy Syst.* 44 (1), 455–466. doi:10.3233/jifs-211862
- Zhou, Y., Sun, B., Sun, W., and Lei, Z. (2022). Tool wear condition monitoring based on a two-layer angle kernel extreme learning machine using sound sensor for milling process. *J. Intelligent Manuf.* 33 (1), 247–258. doi:10.1007/s10845-020-01663-1