DOI: 10.21062/mft.2024.041  http://www.journalmt.com

# Rolling Bearing Fault Diagnosis based on Multi-scale Entropy Feature and Ensemble Learning

Mei Zhang (0000-0003-2013-7613), Zhihui Wang (0009-0006-0681-0137)*, Jie Zhang (0009-0002-3933-5548)
School of Electrical and Information Engineering, Anhui University of Science & Technology, Huainan, 232001, China. E-mail: *1729728067@qq.com

**Aiming at the problem of fault feature extraction and fault identification of rolling bearings, a fault recognition technology based on multi-scale entropy features and ensemble learning is proposed. First, the vibration signal is decomposed by using the variational mode decomposition method, and the cross-correlation coefficient method is used to reconstruct the signal for denoising. Then, to improve the efficiency of feature extraction of rolling bearings, a Refined Composite Multiscale Reverse Permutation Entropy (RCMRPE) based-feature extraction method is proposed. Third, a fault diagnostic model that leverages Stacking-CatBoost ensemble learning to improve the accuracy of rolling bearing defect identification is provided. Finally, relevant experiments are conducted for signal denoising, feature extraction and fault recognition. The proposed fault detection model and the RCMRPE entropy extraction method are compared with common machine learning models and common entropy extraction methods respectively. The experimental results show that the feature extraction error based on RCMRPE is small and can comprehensively reflect the actual faults of bearings; the fault diagnosis model based on Stacking-CatBoost ensemble learning has significantly better diagnostic performance than other models, with accuracy and recall rates above 99%.**

**Keywords:** Rolling bearing, Fault diagnosis, Feature extraction, Variational mode decomposition algorithm, RCMRPE, Stacking- CatBoost model

## 1 Introduction

Many mechanical equipment regards rolling bearings as essential components, and the smooth operation of the equipment depends on the condition of rolling bearings. To maintain the safe operation of mechanical equipment, it is necessary to accurately diagnose the faults of rolling bearings.

The application of vibration signal detection methods in rolling bearing fault diagnosis is currently one of the hotspots in bearing fault diagnosis. How to extract effective features from complex and noisy vibration signals is the key to fault diagnosis. Due to complex working conditions of mechanical equipment, vibration signals are mixed with interference noises, and signal denoising is principal. Common signal denoising methods include Kalman filtering, wavelet decomposition, and variational mode decomposition. Salunkhe et al. [1] used VMD technology to extract rolling bearing features and filter out data noise. Nassef et al. [2] used the VMD method to process the signals and identify the defects from the vibration signals of rolling bearings. Nouioua et al. [3] demonstrated the decomposition and envelope demodulation of VMD, which can efficiently identify a wide range of bearing faults.

After noise reduction, the fault features of rolling bearings are still submerged in complex and redundant information, and it is crucial to extract fault features. Entropy based feature extraction can fully and accurately characterize the faults of rolling bearings. Based on fine composite multi-scale fuzzy entropy, Xu and Tse [4] used support vector machines for fault diagnosis and fault extraction of rolling bearings. Zhang Y. et al. [5] calculated the MPE of the optimal inherent mode function and reconstructed the signals of each mode MPE into feature vectors for fault identification. Zheng et al. [6] used fine generalized composite multivariate multi-scale reverse dispersion entropy to extract the fault characteristics of bearings

Nowadays, deep learning methods are widely used for bearing fault diagnosis. Li et al. [7] suggested using the optimal ensemble deep migration network (OEDTN) to identify the faults of rolling bearings problems from unlabeled data. This method benefits from ensemble learning, domain adaptation, and parameter transfer learning. Zhang et al. [8] presented a group of self-learning convolutional auto-encoders (STL-CAEs). The model can effectively extract valuable faults from less labeled data. Imane et al. [9] used the random forest ensemble learning algorithm to train a model that can classify faults based on selected features. This method can effectively solve the problem of speed variation in bearing fault detection. Li et al. [10] proposed a reinforced ensemble deep migration-based learning network for multi-source domain identification of faults. Diao et al. [11] used the algorithm of LeNet5 combined with Light GBM to diagnose bearing faults, which can achieve high diagnostic accuracy. Xia et al. [12] proposed a multi-level bearing fault detection approach based on ensemble learning and EMD. It combines the Bagging and AdaBoost algorithms to establish a multi-level fault diagnosis framework.

To improve the accuracy of fault diagnosis for rolling bearings in mechanical equipment, this study provides a fault diagnosis method based on bearing vibration signals.VMD is used for signal denoising, and the features are extracted based on Refined Composite Multiscale Reverse Permutation Entropy (RCMRPE).To achieve accurate identification of bearing faults, a Stacking-CatBoost integrated learning based-fault diagnostic model is applied.

## 2 Methods

To improve the accuracy of bearing fault diagnosis, VMD-RCMRPE-Ensemble Learning is used to extract feature information from nonlinear and non-stationary rolling bearing fault signals. Its network structure is shown in Figure 1.

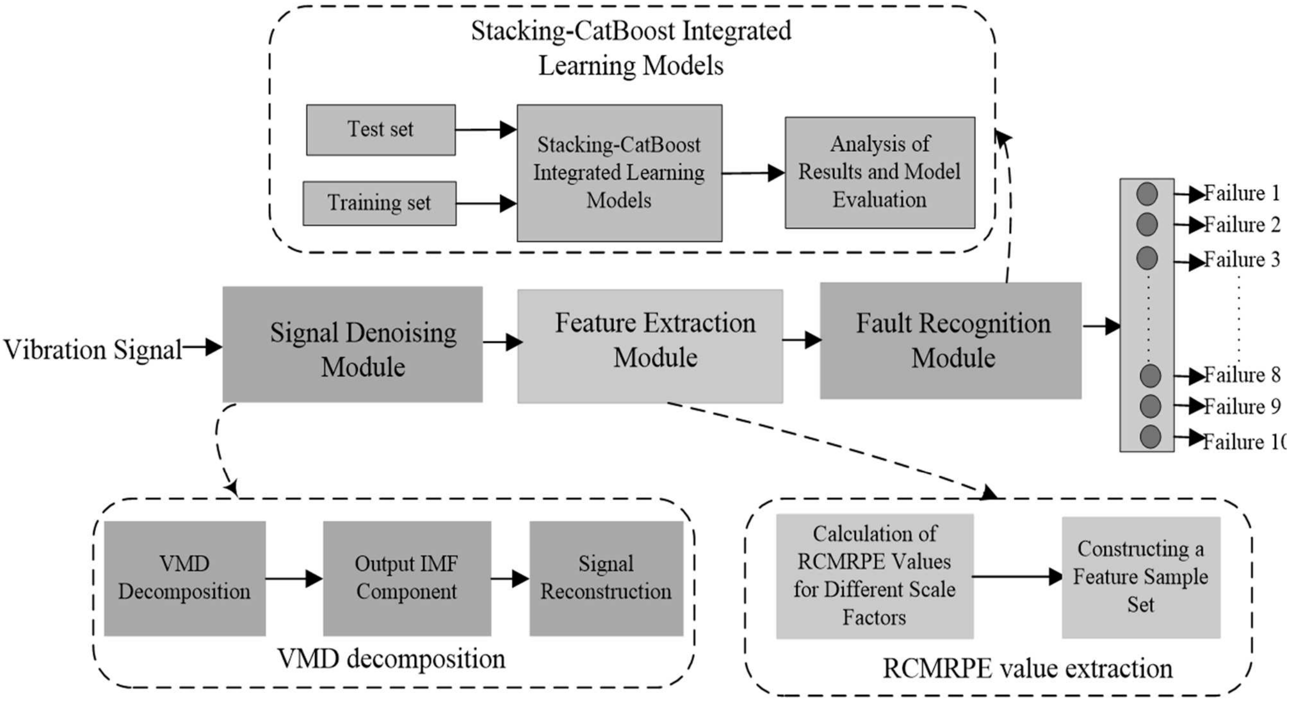


***Fig. 1*** *Network structure for fault diagnosis of rolling bearing*

The fault diagnosis method for rolling bearings based on VMD-RCMRPE-integrated learning first uses the VMD algorithm to decompose the vibration signal and effectively remove interference components. Then the signal is reconstructed using the cross-correlation coefficient method to reconstruct the signals for signal denoising. Based on the proposed RCMRPE feature extraction method, the RCMRPE entropy of different bearing fault state samples is extracted, effectively eliminating redundant information in the samples and constructing a feature sample set. The feature samples are randomly separated into two groups: training set and testing set. The training set is used to train the defect diagnostic model to obtain the trained model through Stacking ensemble learning. The trained fault diagnosis model is then applied to the testing set, and the output of the model is used to establish fault categories and damage levels. Finally, the output results are analyzed and the model is evaluated.

### 2.1 Signal Denoising Module

The VMD algorithm is an emerging time-frequency signal analysis and estimation algorithm. Compared with the EMD algorithm, it has a more solid foundation of mathematical theory, can maximize the suppression of multi-component aliasing and decompose signals more effectively.

The working principle of variational mode decomposition can be understood as the process of establishing a mathematical model for variational constraint problems and continuously iterating to find the optimal solution. The constructed variational problem is to find the most suitable solution in the modal component functions of the original signal $f(t)$ k u ($k$), so that the sum of the estimated bandwidth of each function is minimized, and the sum of all functions is equal to the original signal [13-14]. Before solving the multiple modal functions, the variational modal problem is constructed. The specific process is as follows:

(1) Using the Hilbert transform to demodulate each modal function $u_k(t)$, the unilateral spectrum is as follows:

$$J(t)=\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t) \quad (1)$$

Where:

$J(t)$…The unilateral spectrum of each function,

$*$…The convolution operation.

(2) Mixing each function spectrum with the estimated center frequency, multiplying the unilateral spectrum with the exponential term, the fundamental frequency band corresponding to each function spectrum is solved as follows:

$$q(t)=\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-j\omega_k t} \quad (2)$$

Where:

q(t)…The fundamental frequency band of the final modal function.

(3) Calculate the L2norm of the gradient and estimate the bandwidth of the signal:

$$p(t)=\left\|\partial(t)\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-j\omega_k t}\right\|_2^2 \quad (3)$$

(4) The mathematical model of the constrained variational problem is obtained as follows:

$$\begin{cases}\min\left\{\sum_k\left\|\partial(t)\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-j\omega_k t}\right\|_2^2\right\} \\ s.t.\sum_k u_k=f(t)\end{cases} \quad (4)$$

Where:

$f(t)$…The original input signal,

$u_k(t)$…The modal component function.

The above four computational technologies can be used to create mathematical models for variational problems. To determine the optimal signal decomposition result, the variational problem is solved in the next step.

The following mathematical model is obtained by transforming Eq.(4) into an unconstrained problem:

$$\begin{aligned}L\left(\{u_k\},\{\omega_k\},\lambda\right)=\alpha\sum_k\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-j\omega_k t}\right\|_2^2 \\ +\left\|f(t)-\sum_k u_k(t)\right\|_2^2+\left\langle\lambda(t),f(t)-\sum_k u_k(t)\right\rangle\end{aligned} \quad (5)$$

The multiplicative operator alternating direction method is used to solve the quadratic optimization problem:

$$\hat{u}_k^{n+1}(\omega)=\frac{\hat{f}(\omega)-\sum_{i\neq k}\hat{u}_i(\omega)+\frac{\hat{\lambda}(\omega)}{2}}{1+2\alpha\left(\omega-\omega_k\right)^2} \quad (6)$$

Where:

$a$…The penalty factor.

At the same time, the center frequency of the solution is also transferred to the frequency domain to obtain:

$$\omega_k^{n+1}=\frac{\int_0^\infty \omega|\hat{u}(\omega)|^2\,d\omega}{\int_0^\infty |\hat{u}(\omega)|^2\,d\omega} \quad (7)$$

Where:

$\omega_k^{n+1}$…The spectral center of each function.

According to Eq. (6) and Eq. (7), λ is calculated as follows:

$$\hat{\lambda}^{n+1}(\omega)=\hat{\lambda}^n(\omega)+\tau\left(\hat{f}(\omega)-\sum\hat{u}_k^{n+1}(\omega)\right) \quad (8)$$

Where:

$\tau$…The tolerance of noise.

When Eq. (8) satisfies the following conditions, the iteration is stopped:

$$\sum_k \frac{\left\| \hat{u}_k^{n+1} - \hat{u}_k^n \right\|_2^2}{\left\| \hat{u}_k^n \right\|_2^2} < \varepsilon \qquad (9)$$

Where:
$\varepsilon$…The iterative convergence error.

VMD transforms the decomposition process into a non-recursive mathematical model for solving variational problems. This signal decomposition method has a strong theoretical foundation. VMD uses multiple Wiener filter banks to analyze signals and has stronger robustness than EMD. At the same time, the convergence error ε is used to control the solution and the central frequency of each modal component, so that the sampling effect is much smaller than the empirical mode decomposition method, thereby reducing the impact of the sampling frequency on the algorithm. To some extent, VMD can prevent decomposition errors, allowing each modal component to fully highlight significant features in the signal, which helps to diagnose fault models in the future.

Each modal component is generated after VMD decomposition, and then the signal is reconstructed using the correlation coefficient method. In the field of fault diagnostics, correlation coefficients are often used as a measure of signal tightness evaluation [15]. It can measure the closeness of one component containing relevant information to another component, and quantitatively express the degree of interdependence between two signals in numerical form.

To further eliminate redundant information and improve the sensitivity of feature parameters to bearing fault conditions, effective components are selected based on the tightness between the components and the original signal. The correlation coefficient can be calculated as follows:

$$\begin{cases} \lambda_i = \dfrac{Cov(u,u_i)}{\sqrt{v(u\ )}\sqrt{v(u_i)}} \\ \lambda_m = \dfrac{1}{k}\sum\limits_{i=1}^{k} \lambda_i \end{cases} \qquad (10)$$

Where:
$u$ and $u_i$…The sample signals,
$v(u)$…The variance of $u$,
$v\ (u_i)$…The variance of $u_i$,
$\lambda_i$…The degree of association among the original signal and the signal component,
$\lambda_m$…The threshold for selecting effective components.

The degree of similarity between any two components can be precisely determined by their correlation coefficients with the original signal. Selecting highly correlated components for signal reconstruction can effectively eliminate irrelevant noise interference components, which is beneficial to subsequent feature extraction and fault diagnosis.

## 2.2 Feature Extraction Module

According to the above noise reduction techniques, the noise signal in the original signal is basically suppressed. However, vibration signals are often mixed with other complex redundant information. It is necessary to extract the features of bearing vibration signals, eliminate redundant information, and obtain state feature quantities that can reflect corresponding faults. When diagnosing rolling bearing faults, the entropy algorithm-based feature extraction method is better in extracting the fault features of rolling bearings [16].

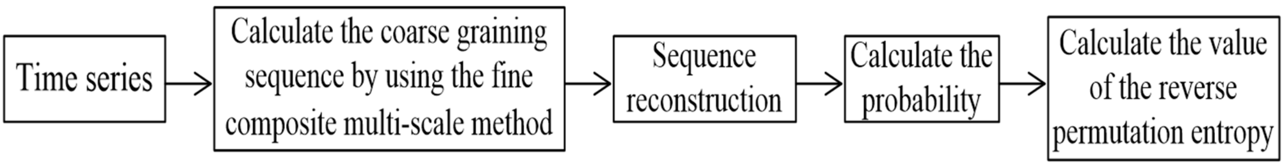


***Fig. 2*** *The principle of the RCMRPE algorithm*

To enrich the extraction of failure state features from vibration signals, an RCMRPE feature extraction method is proposed to introduce the fine composite multi-scale method and the reverse dispersion entropy (RDE) method into the entropy calculation. The Refined Composite Multiscale Reverse Permutation Entropy (RCMRPE) is obtained by improving the Multiscale Permutation Entropy (MPE).

The RCMRPE algorithm is a further refinement of MPE, and its principle is shown in Figure 2.

Under the condition of a scaling factor $s$, $s$ refined composite sequences are constructed sequentially from different initial starting points of the original time series $u$. That is, multiple refined sequences with different starting points and lengths are obtained continuously in a loop to ensure that the required results. The obtained RCMRPE entropy value can fully reflect the correlation between adjacent elements in the initial time series. When calculating the multi-scale permutation entropy of each refined sequence, the probability pattern of each subsequence and the

Euclidian distance of the embedded dimension factorial are used instead of the traditional entropy calculation method. The average RPE entropy of all refined sequences under *s*-scale is calculated to obtain the RCMRPE entropy at different scales. The following are the steps involved in the calculation:

(1) Instead of the traditional coarse-graining method, a fine composite multiscale method is used to calculate the k-th coarse-grained sequence $x_k^{(s)}=\{x_{k,1}^{(s)}, x_{k,2}^{(s)},\ldots\}$, based on the original signal *u*:

$$x_{k,j}^{(s)} = \frac{1}{s}\sum_{b=k+s(j-1)}^{k+js-1} u_b \tag{11}$$

Where:
$k$=1,2,…$s$,
$j$=1,2,…$L/s$,
$s$…The selected scale factor,
$L$…The data length.

(2) For each scaling factor *s*, according to the principle of reverse permutation entropy, the entropy of each fine composite multi-scale reverse permutation is defined as follows:

$$H_{RCMRPE}(m) = \Delta^2 = \sum_{i=1}^{q}\left(P_i - \frac{1}{m!}\right)^2 \tag{12}$$

Where:
$P_i$…The probability of subsequence occurrence,
$m$…The embedding dimension of the reconstruction matrix.

The RCMRPE entropy has a minimum value of zero. A high entropy number indicates a high degree of correlation between the components of time series.

This statistical method of fine composite multi-scale reverse entropy takes into account the complex change of time series at multiple scales, overcomes the shortcomings of traditional multi-scale coarse-graining, and can dig deep into the relationship between adjacent elements in the time series, so as to reduce entropy calculation deviation for better feature extraction.

## 2.3 Fault Identification Module

To accurately determine the faults of rolling bearings, a Stacking-CatBoost model is proposed, which adopts the Stacking ensemble learning framework structure and combines the multi-model primary classifier with the secondary classifier CatBoost for fault diagnosis.

The Stacking ensemble learning model is a high-level heterogeneous ensemble learning framework that integrates multiple optimal learning models [17]. Its heterogeneous model combination method provides a more optimal solution to actual complex classification problems [18-19]. Figure 3 shows the Stacking ensemble learning model structure.

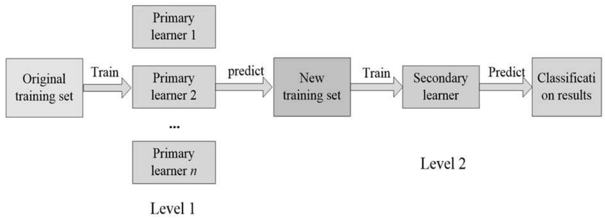


***Fig. 3*** *Model structure of Stacking ensemble learning*

The integrated learning consists of two layers. At the beginning, the original training set is used to train many primary learning models, which forms the first layer of Stacking. The second layer model is further trained on this basis to obtain the final classification result.

The Stacking-CatBoost model is constructed by combining the CatBoost model with the Stacking ensemble learning, and its structural block is shown in Figure 4.

As shown in Figure 4, the Stacking-CatBoost ensemble learning consists of two layers: the first layer classification model includes five classification models: Extra Trees, Random Forest, Ada Boost, GBDT and SVM, and the second layer classification model include CatBoost. The following are the precise work steps:

- Step 1: Use the original training set to train several primary learning models;
- Step 2: Combine the prediction results of the first-layer primary learning model to form a new training set;
- Step 3: Train and test the second-layer model using the newly created data set created by the first classifier.

To minimize the risk of model overfitting, a new training set is generated using the K-fold cross-validation method. Taking 5-fold cross-validation as an example, the specific principle is as follows: First, the training samples are decomposed into 5 small samples for 5-fold cross-validation training of models. Then for each primary learning model, each model is trained5 times in sequence, retaining one-fifth of the samples as the verification set for each training to validate the trained model, and outputting the prediction results of the validation set to form a new training set. After each training, the testing set is predicted, and the prediction results of the 5 testing sets are averaged to form a new testing set. Finally, the new sample sets constructed by each model are composed together, and the new sample set is the input sample of the secondary classification model.

Unlike traditional single classification models, the Stacking ensemble learning method integrates multiple powerful and different learning models into a secondary classifier, and uses the classification prediction results of these models as input. This improves the overall classification performance of the model and provides a new idea for fault diagnosis of complex rolling bearings.

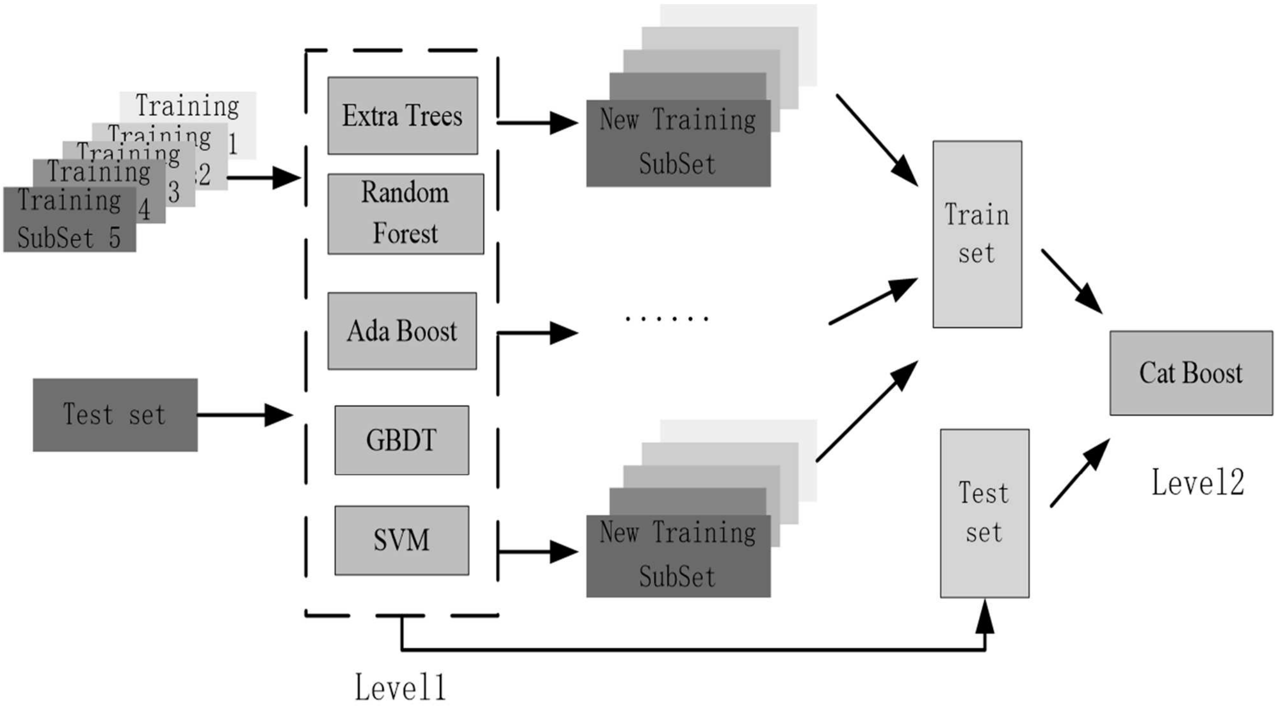


***Fig. 4*** *Structure block diagram of the Stacking- CatBoost model*

## 3 Results and analysis

To confirm the feasibility of the previously proposed strategy, experiments on bearing defect detection methods using the bearing dataset provided by Case Western Reserve University in the United States are conducted.

### 3.1 Dataset

The deep groove ball rolling bearing 6205-2RS JEM SKF at the driving end and the deep groove ball rolling bearing 6203-2RS JEM SKF at the fan end are the two types of bearings used in this data set. The rolling bearing data at the driving end serves as the basis for research and debate.

The types of faults are classified based on their location and degree of damage. The dataset includes the normal state and 9 fault states of rolling bearings. The specific types are shown in Table 1.

All experimental data are processed in the Matlab environment. As an example of a serious failure of a motor bearing at the driving end, Fig. 5 shows the time-domain waveforms of the normal and faulty states of the bearing.

From Figure 5, the time-domain data waveforms of the rolling element fault state and the normal state of the bearing can be obtained. Both are chaotic multi-component modulated pulse signals, and the characteristics of the vibration mechanism cannot be directly derived from the time-domain data. Therefore, effective denoising and feature extraction methods must be used to filter out useful information for subsequent bearing fault diagnosis.

***Tab. 1*** *Fault types and labels of rolling bearings*

| Labels | Fault types | Fault size (mm) |
|---|---|---|
| 1 | normal | 0 |
| 2 | Minor failure of inner ring | 0.1778 |
| 3 | Minor failure of rolling elements | 0.1778 |
| 4 | Minor failure of outer ring | 0.1778 |
| 5 | Moderate failure of the inner ring | 0.3556 |
| 6 | Moderate failure of rolling element | 0.3556 |
| 7 | Moderate failure of the outer ring | 0.3556 |
| 8 | Severe failure of inner ring | 0.5334 |
| 9 | Severe failure of rolling element | 0.5334 |
| 10 | Severe failure of outer ring | 0.5334 |

ISSN 1213–2489
e-ISSN 2787–9402

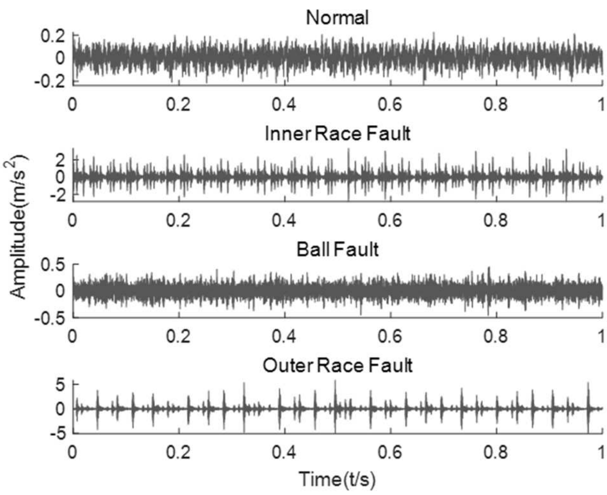


*Fig. 5* *Time domain data of different fault samples*

## 3.2 Signal Denoising Based on VMD

Here, the fault data of rolling elements is used as an example for VMD decomposition. Figure 6 shows the decomposed time-domain diagram.

Figure 6 demonstrates that the components decomposed by the VMD method are close to the original signal, and can accurately depict the fault features of signals. This indicates that the VMD method is superior in reducing signal noise and retaining details.

Figure 7 shows the spectrum of each component after signal decomposition.

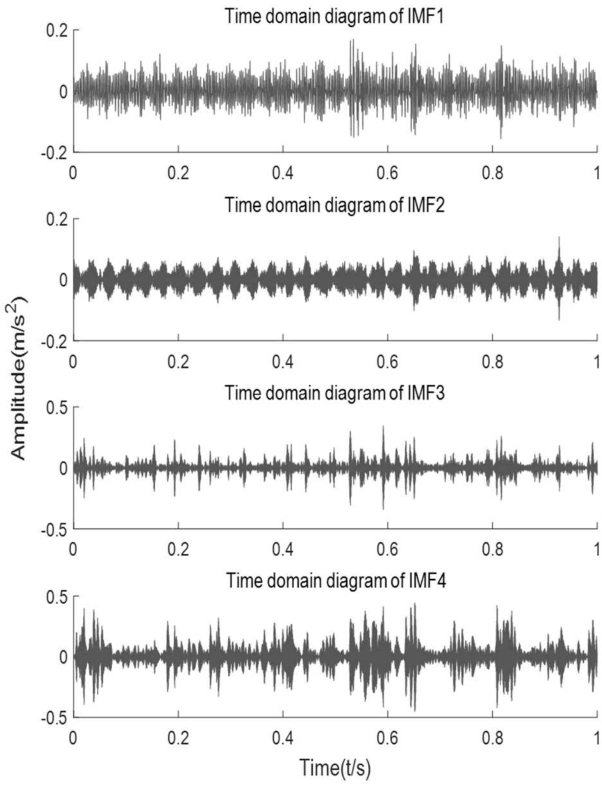


*Fig. 6* *VMD decomposition time-domain diagram of rolling body fault data*

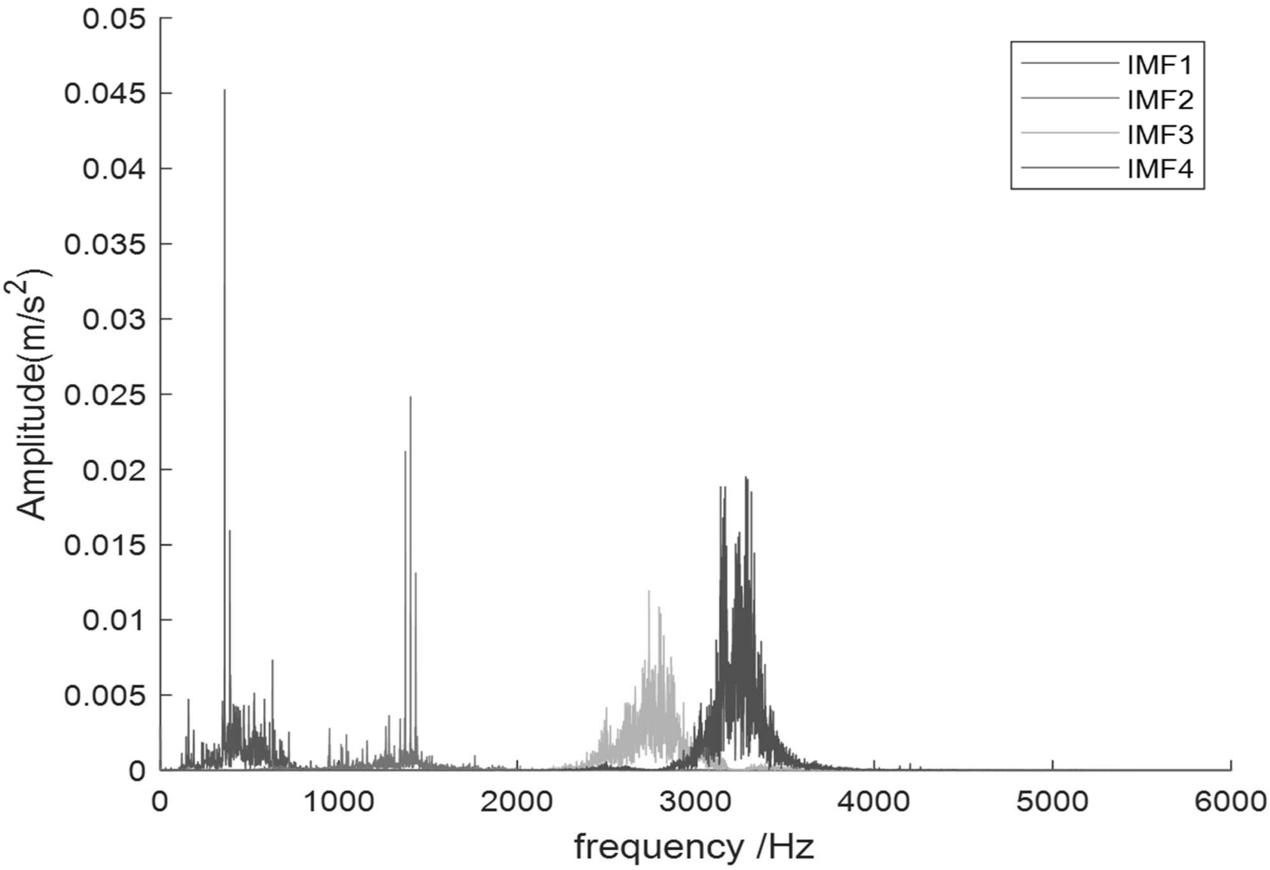


*Fig. 7* *VMD component spectrum of rolling element fault data*

As shown in Figure 7, this method can greatly improve the modal mixing. The signal components obtained from VMD decomposition are smoothly in the spectrum space, and the frequencies of different signal components are different. There is no aliasing in the main frequency between the components. The results show that this method can effectively separate different frequency components from the original signal, obtain the characteristic frequency signals that can represent the fault impact component from the original signal, and eliminate the invalid interference components to achieve signal denoising [20].

The tightness between signal components and the original signal is determined by using the correlation coefficient method for each component after VMD decomposition. The effective signal components are then selected for signal reconstruction to remove noise and interference from the original signal.

After calculation, the correlation coefficients of IMF1~IMF4 are 0.375, 0.265, 0.665, and 0.670 respectively, and the threshold is determined by taking the average of 0.494.The correlation coefficients of Component 3 and Component 4 are greater than the threshold, indicating that Component 3 and Component 4 are closely related to the original signal and contain rich fault impact characteristic signals. Therefore, Component 3 and Component 4 are selected for signal reconstruction, and the remaining components are invalid signals and eliminated.

To compare the reconstruction effect, the time-domain waveforms before and after reconstruction are shown in Figure 8.

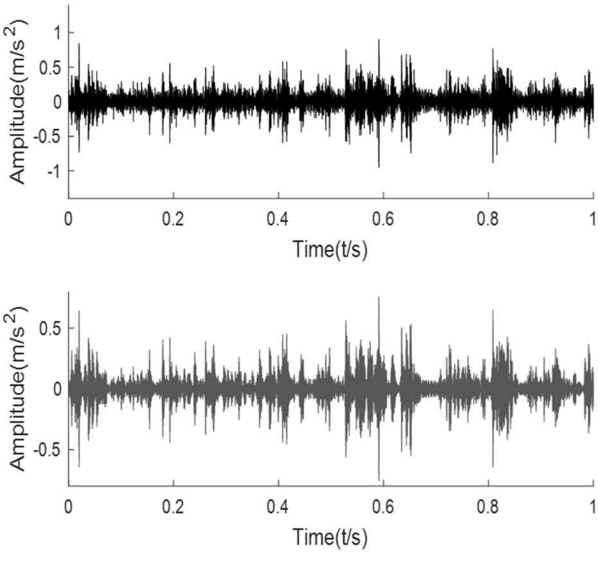


***Fig. 8*** *Time-domain comparison of rolling element fault signals before and after reconstruction*

It can be seen from Figure 8 that the noise interference component in the reconstructed signal is weakened. For example, between 0.8s and 1s, the magnitude of the burr interference information in the reconstructed signal significantly decreases, while the non-stationary fault influence characteristics increase. The results show that after VMD signal decomposition and selecting effective components for signal reconstruction, it has a good signal denoising effect in fault signal processing of rolling bearings.

### 3.3 Feature Extraction Based on RCMRPE

The RCMRPE feature extraction technique is used to extract the vibration signal after denoising. To verify the effectiveness and feasibility of the feature extraction method based on RCMRPE, experiments using different entropy algorithms are used here to compare with the proposed algorithm. The commonly used entropy algorithms include Multiscale Approximate Entropy (MAE), Multiscale Fuzzy Entropy (MFE), Multiscale Permutation Entropy (MPE), and more. For comparison, two experiments are conducted: one is the feature extraction of bearing signals with the same fault size and different fault types under different entropy algorithms; the other is the feature extraction of bearings with the same fault types and different sizes of faults under different entropy algorithms.

#### 3.3.1 Comparative Experiments on Feature Extraction under Different Fault Types

For bearing signals with the same fault size and different fault types, four entropy algorithms, MAE, MFE, MPF, and RCMRPE are used for feature extraction. Here, a fault signal with a fault size of 0.1778mm is taken as an example. There are four fault types: normal, inner-ring fault, rolling element fault, and outer-ring fault. The scale factor ranges from 0 to 15, and four entropy algorithms are used to for experiments. The entropy standard deviation results of these algorithms are shown in Figure 9.

According to Figure 9, the entropy calculation errors of the four entropy calculation methods are all less than 0.04. The entropy standard deviation of MAE fluctuates largely, followed by MFE and MPE. The fluctuation amplitude of the entropy standard deviation of RCMRPE is the smallest, with the most stable entropy. At low scales, the fluctuation range of the RCMRPE entropy errors of the four bearing states is small [21]. Inner-ring faults, rolling element faults, and outer-ring faults have steady entropy errors as the scale factor increases.

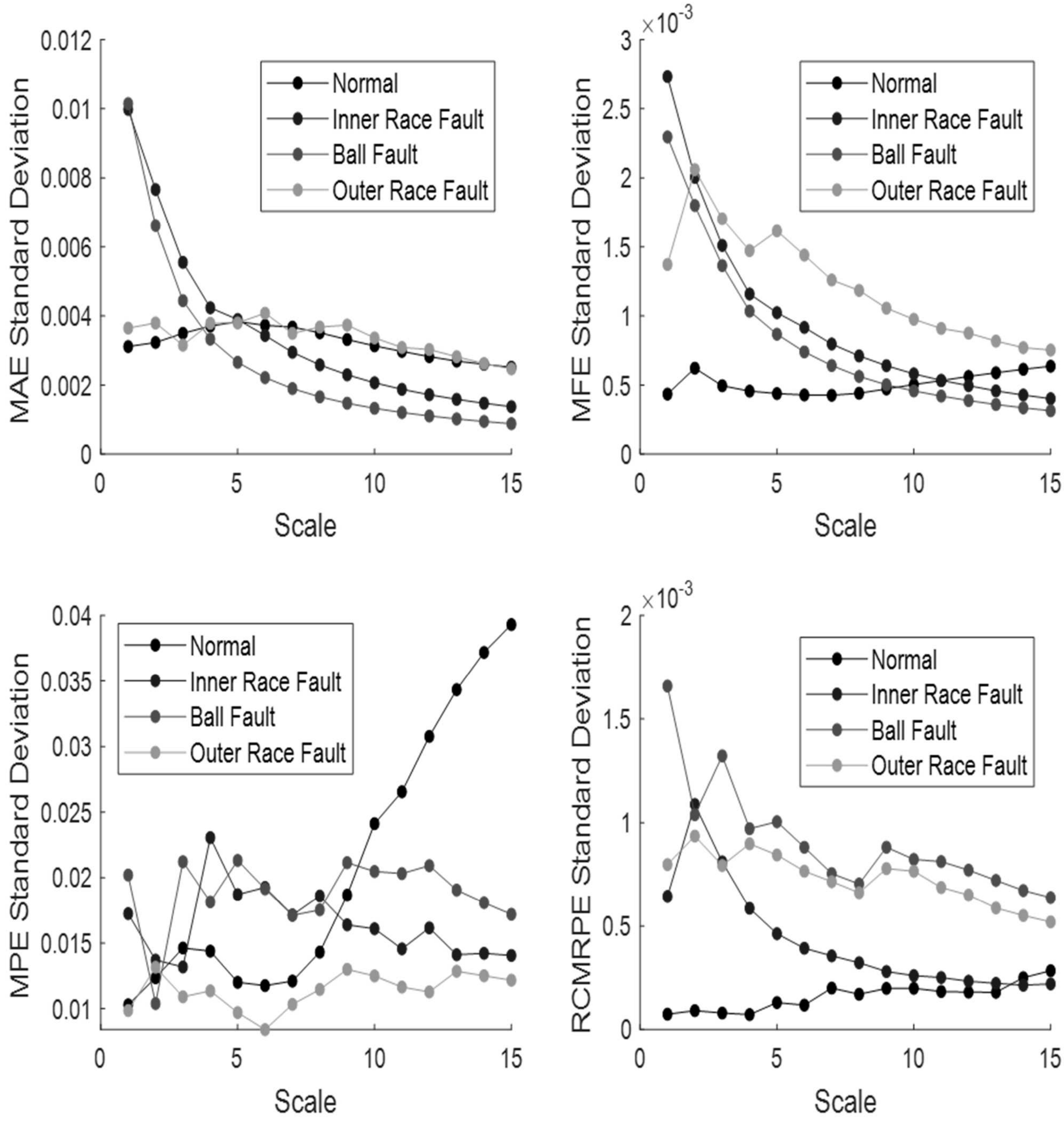


***Fig. 9*** *Entropy error results of different entropy algorithms under different fault types*

### 3.3.2 Comparative Experiments on Feature Extraction under Different Sizes

For bearing signals with the same fault types and different fault sizes, four entropy algorithms, MAE, MFE, MPF, and RCMRPE, are used for feature extraction. Taking the inner-ring fault as an example, the fault sizes are 0mm, 0.1778mm, 0.3556mm, and 0.5334mm respectively. The entropy standard deviation curve of each entropy value algorithm is shown in Figure 10.

It can be observed from Figure 10 that under different fault sizes, the errors of MAE, MFE, MPE and RCMRPE are all lower than 0.015, but that of MAE and MPE changes significantly. Especially when the fault size is 0.1778mm, the entropy error of MAE declines linearly with unstable variation. In the entropy error distribution curve of RCMRPE, the error value distribution of the four-fault size states is relatively stable, and the fluctuation amplitude of the error in each state is relatively small. This indicates that the RCMRPE entropy value calculation method has small errors in the feature extraction process and can fully reflect the actual fault information of motor bearings.

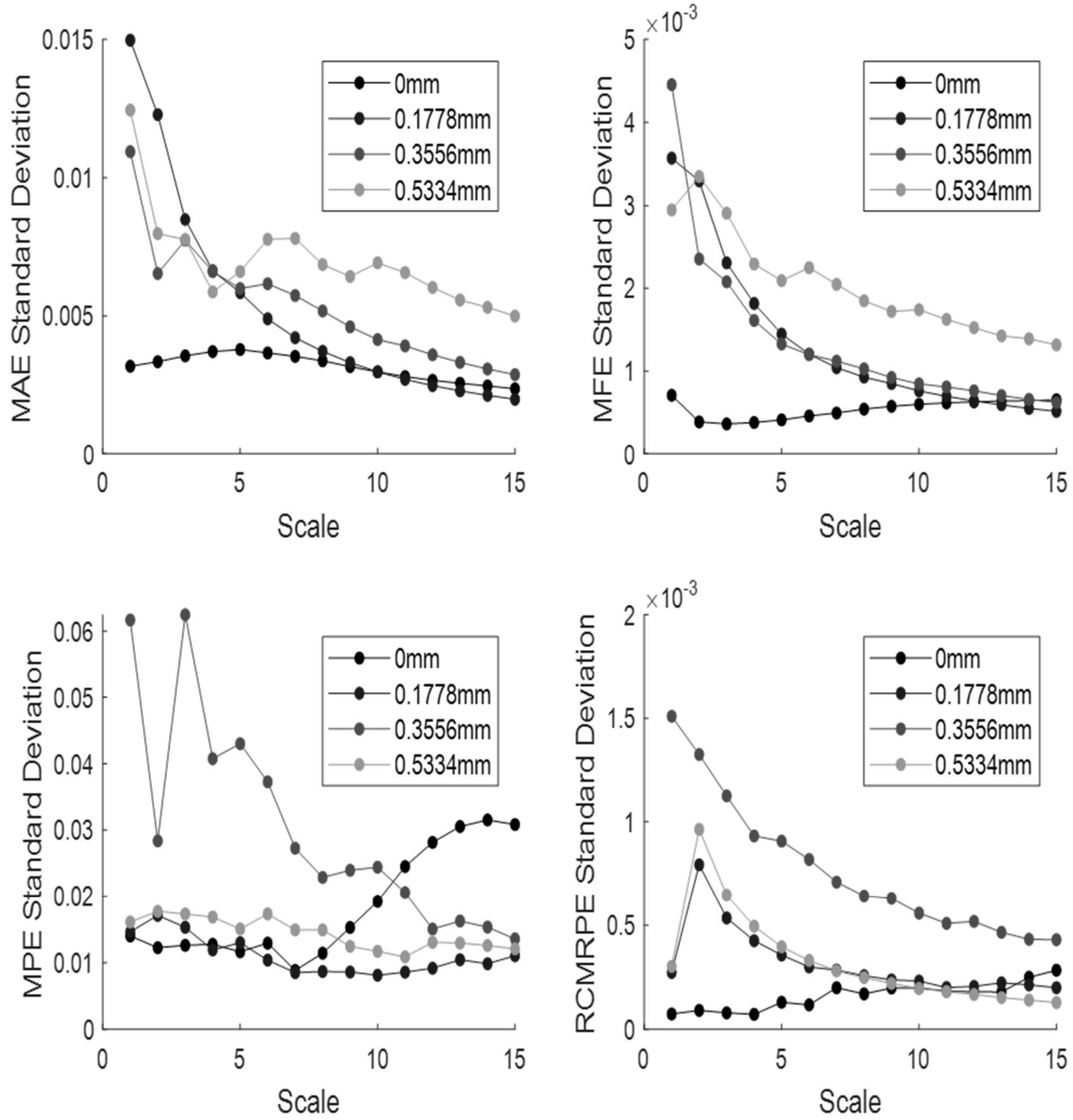


***Fig. 10*** *The entropy error results of different entropy algorithms under different fault sizes*

### 3.4 Fault Identification based on Stacking-CatBoost Ensemble Learning

To investigate the performance of the Stacking-CatBoost ensemble learning model, the single models such as Extra Trees (ET), Random Forest (RF), Ada Boost, Gradient Boosting Decision Tree (GBDT), and Support Vector Machine (SVM) are selected for comparison with the Stacking-CatBoost model [22-26].

After feature extraction, the feature set is separated into a training set and a testing set with a ratio 7:3 for model training and testing. Two experiments are conducted: one is to identify faults on signals of the same fault size under the same working conditions, and the other is to identify faults on signals of the same fault size under different working conditions.

#### 3.4.1 Model-Evaluation Index

This paper mostly employed *Accuracy, Precision, Recall, F1_Score, Kappa coefficient, and Jaccard coefficient* as model evaluation indexes, and calculated as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (13)$$

Where:
$TN$…The accurate forecast of a negative sample,

*FP*…The incorrect prediction of a positive sample,

*FN*…Wrong prediction of a negative sample,

*TP*…The positive sample with the correct prediction.

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

$$F1\_Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{16}$$

$$\begin{cases} Kappa = \dfrac{p_0 - p_e}{1 - p_e} \\ p_e = \dfrac{\sum_{i=1}^{c} a_i \times b_i}{n \times n} \end{cases} \tag{17}$$

Where:

*c*…The total number of categories,

*n*…The total number of samples,

$a_i$…The true number of samples in the *i* category,

$b_i$…The predicted number of samples in the *I* category,

$P_0$…The overall classification accuracy.

The *Jaccard coefficient* measures the similarities and differences between samples by predicting the distance between the label and real label of the samples. The larger the *Jaccard coefficient*, the higher the similarity between the two samples.

### 3.4.2 Fault Diagnosis Experiment under the Same Working Conditions

To further study the superiority of the bearing fault diagnosis method based on the Stacking-CatBoost model, the models, Extra Trees, Random Forest, Ada Boost, GBDT, SVM, and Stacking-Cat Boost are used to classify the feature sample set.

Under the same working conditions and different faults, the fault identification experiments based on different models are carried out. Here, the working condition is selected with a load of 2 HP, and 300 samples are randomly selected for fault diagnosis experiments. The diagnosis results are shown in Figure 11.

It can be seen from Figure 11 that the fault diagnosis effect of Stacking-Cat Boost model is the best. There are 9 kinds of faults identified by the model completely and only a few parts of 1 kind of faults (class 5) are incorrectly identified. The error rate of ET model is relatively high, and the 5th, 6th, 9th and 10th faults all have diagnostic errors. The error rate of RF model mainly focuses on the 2th, 5th, 7th, and 10th faults. For the Ada Boost model, there are errors in the identification of class 2, 5, 6 faults, while class 10 errors are not correctly identified. The fault diagnosis error rates of GBDT and SVM models mainly are mainly concentrated in the 5th and 9th, 10th faults.

***Tab. 2*** *Performance indicators of various fault diagnosis models under the same working condition*

| Model | Accuracy | Accuracy | Recall | F1 value | Kappa coefficient | Jaccard coefficient |
|---|---|---|---|---|---|---|
| ET | 0.9533 | 0.9531 | 0.9553 | 0.9516 | 0.9479 | 0.9108 |
| RF | 0.9667 | 0.969 | 0.9659 | 0.9667 | 0.9628 | 0.9355 |
| Adaboost | 0.8800 | 0.8379 | 0.8814 | 0.8499 | 0.8666 | 0.7857 |
| GBDT | 0.9567 | 0.9618 | 0.9584 | 0.9552 | 0.9517 | 0.9169 |
| SVM | 0.94 | 0.9426 | 0.9606 | 0.9414 | 0.9332 | 0.8868 |
| Stacking - CatBoost | 0.9900 | 0.9900 | 0.9900 | 0.9895 | 0.9888 | 0.9802 |

It can be seen from Table 2 that under different fault types, the evaluation indicators of the Stacking-CatBoost model are all greater than 98%. The accuracy, precision and recall rates are almost 99%, and F1 and Kappa coefficient are also greater than98.5%. The fault identification results are significantly higher than the single models such as Extra Trees, Random Forest, and AdaBoost. It proves the feasibility and effectiveness of defect diagnosis technology based on the Stacking-CatBoost model. Its diagnosis results are more accurate, with better generalization ability and stronger classification performance.

The performance indicators of each model are shown in Table 2.

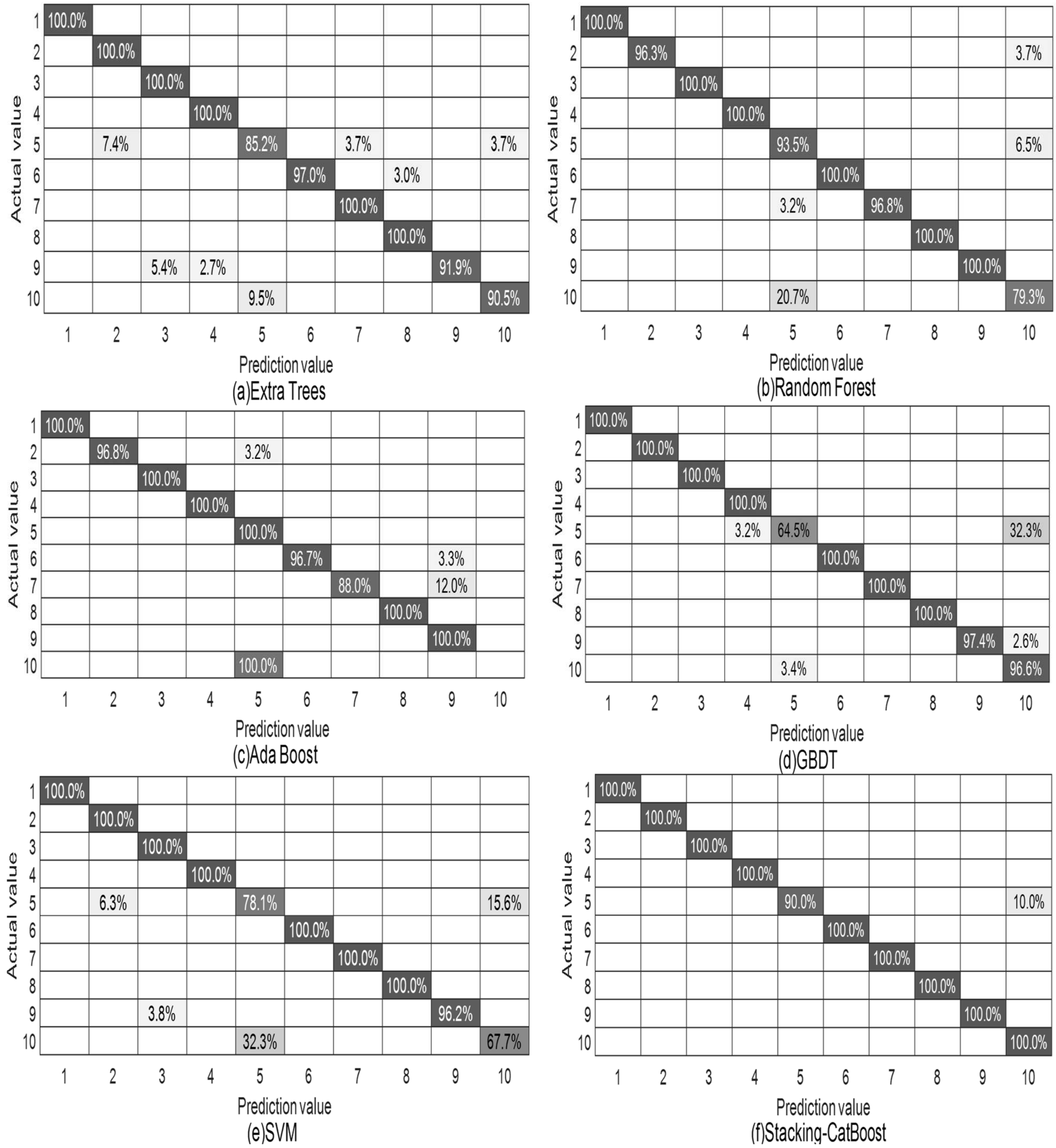


***Fig. 11*** *Confusion Matrix for Fault Diagnosis of different models under the Same Operating Condition*

### 3.4.3 Fault diagnosis experiments under different working conditions

The data from four distinct working conditions is used to further confirm the effectiveness of the model under various working conditions and the same faults. Taking the defect size of 0.1778mm as an example, Table 3 shows its four possible operating states.

***Tab. 3*** *Four different working conditions of the same fault size*

| Fault size/mm | Motor load /HP | Motor speed /(r/min) |
|---|---|---|
| 0.1778 | 0 | 1797 |
| 0.1778 | 1 | 1772 |
| 0.1778 | 2 | 1750 |
| 0.1778 | 3 | 1730 |

The faults of different models are identified under the same working conditions and different fault sizes. 1,000 samples are randomly selected for testing, and the diagnostic results are shown in Figure 12.

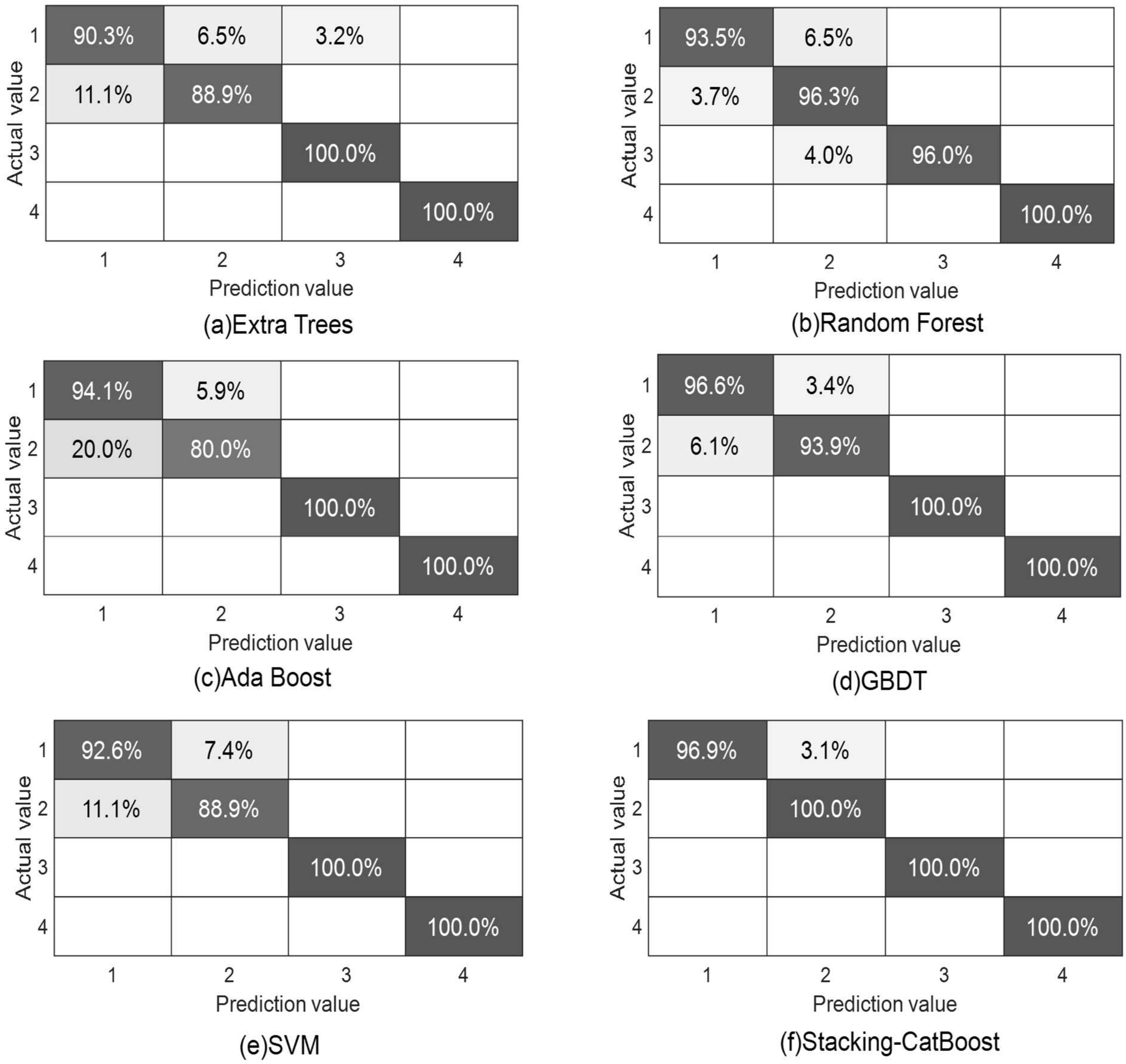


***Fig. 12** Confusion Matrix for Fault Diagnosis of different models under Different Working Conditions*

According to Figure 12, under different working conditions and the same fault size, the fault diagnosis performance of the Stacking-Cat Boost model is the best. The accuracy of the Stacking-CatBoost model remains greater than 99%, which also verifies the effectiveness of the Stacking-Cat Boost ensemble learning method.

The performance indicators of each model are shown in Table 4.

***Tab. 4** Performance indicators of various fault diagnosis models under different working conditions*

| Models | Accuracy | Accuracy | Recall | F1 | Kappa coefficients | Jaccard coefficients |
|---|---|---|---|---|---|---|
| ET | 0.9500 | 0.9470 | 0.9480 | 0.9473 | 0.9328 | 0.9048 |
| RF | 0.9667 | 0.9658 | 0.9646 | 0.9647 | 0.9552 | 0.9355 |
| Adaboost | 0.9333 | 0.9413 | 0.9353 | 0.9365 | 0.9104 | 0.8750 |
| GBDT | 0.9750 | 0.9755 | 0.9762 | 0.9757 | 0.9666 | 0.9512 |
| SVM | 0.9583 | 0.9540 | 0.9537 | 0.9537 | 0.9434 | 0.9200 |
| Stacking - CatBoost | 0.9917 | 0.9922 | 0.9922 | 0.9921 | 0.9889 | 0.9835 |

According to Table 4, under different working conditions and the same fault size, compared with other single learning models, the Stacking- CatBoost model has the most ideal fault diagnosis effect, with the evaluation indicators such as accuracy and precision are all greater than 0.99. It indicates the ability of the Stacking-CatBoost model to determine the faults of rolling bearings in various operational scenarios.

## 4 Conclusions

To ensure safe operation of rolling bearings in mechanical equipment, a bearing failure detection method based on multi-scale entropy feature and ensemble learning is proposed. The conclusions are drawn as follows:

- (1) The VMD method is used to decompose the fault signal of a rolling bearing. According to the correlation coefficient method, the effective component is screened to reconstruct the signal. The invalid component is eliminated for signal denoising, which lays a foundation for the subsequent feature extraction.
- (2) The fault features of bearings are extracted using the RCMRPE entropy method, and compared to other multi-scale entropy algorithms such as MAE, MFE, and MPE. The experimental results show that RCMRPE entropy has the smallest standard deviation and the most stable fluctuation amplitude, which can fully reflect the actual faults of motor bearings.
- (3) The Stacking-CatBoost model is used for bearing fault diagnosis, and its accuracy and recall rate are both greater than 99%, which is significantly better than other single models. To this end, the model has higher accuracy and more reliable and effective performance.

## Acknowledgement

***This work was supported by National Natural Science Foundation of China under Grant 52374154.***

## References

[1] SALUNKHE, V. G., DESAVALE, R. G., KHOT, S. M., & YELVE, N. P., (2023). A Novel Incipient Fault Detection Technique for Roller Bearing Using Deep Independent Component Analysis and Variational Modal Decomposition. *Journal of Tribology*, 145(7): 074301. https://doi.org/10.1115/1.4056899

[2] NASSEF, M. G. A., HUSSEIN, T. M., & MOKHIAMAR, O., (2021). An adaptive variational mode decomposition based on sailfish optimization algorithm and Gini index for fault identification in rolling bearings. *Measurement*, 173: 108514. https://doi.org/10.1016/j.measurement.2020.108514

[3] NOUIOUA, I., YOUNES, R., MRABTI, A., MEDDOUR, I., & ALIA, S., (2023). Self-Organizing Maps and VMD for Accurate Diagnosis of Bearing Defects. *Journal of Vibration Engineering & Technologies*. https://doi.org/10.1007/s42417-023-01195-x

[4] XU, F., &TSE, P. W., (2019). A method combining refined composite multiscale fuzzy entropy with PSO-SVM for roller bearing fault diagnosis. *Journal of Central South University*, 26(9): 2404-2417. https://doi.org/10.1007/s11771-019-4183-7

[5] ZHANG Y., LI Y., (2023). Rolling bearing fault diagnosis based on multi-scale permutation entropy and IWOA-SVM. *Electronic Measurement Technology*, 46(19): 29-34. https://doi.org/10.19651/j.cnki.emt.2212286

[6] ZHENG, J., CHEN, Y., TONG, J., & PAN, H., (2023). RGCMMRDE and Its Applications in Rolling Bearing Fault Diagnosis. *China Mechanical Engineering*, 34(11): 1315-1325. https://doi.org/10.3969/j.issn.1004-132X.2023.11.007

[7] LI, X., JIANG, H., WANG, R., & NIU, M., (2021). Rolling bearing fault diagnosis using optimal ensemble deep transfer network. *Knowledge-Based Systems*, 213: 106695. https://doi.org/10.1016/j.knosys.2020.106695

[8] ZHANG, Y., WANG, J., ZHANG, F., LV, S., ZHANG, L., JIANG, M., & SUI, Q., (2022). Intelligent fault diagnosis of rolling bearing using the ensemble self-taught learning convolutional auto-encoders. *IET Science, Measurement & Technology*, 16(2): 130-147. https://doi.org/10.1049/smt2.12092

[9] IMANE, M., RAHMOUNE, C., ZAIR, M., & BENAZZOUZ, D., (2023). Bearing fault detection under time-varying speed based on empirical wavelet transform, cultural clan-based optimization algorithm, and random forest classifier. *Journal of Vibration and Control*, 29(1-2): 286-297. https://doi.org/10.1177/10775463211047034

[10] LI, X., JIANG, H., XIE, M., WANG, T., WANG, R., & WU, Z., (2022). A reinforcement ensemble deep transfer learning network for rolling bearing fault diagnosis with Multi-source domains. *Advanced Engineering Informatics*, 51: 101463-101480. https://doi.org/10.1016/j.aei.2021.101480

[11] DIAO, N.K., MA, H.X., LIU, F., (2022). Bearing Fault Diagnosis Method Based on Improved LeNet5 and LightGBM. *Foreign Electronic Measurement Technology*, 41(1): 140-145. https://doi.org/10.19652/j.cnki.femt.2103308

[12] XIA, T., ZHUO, P., XIAO, L., DU, S., WANG, D., & XI, L., (2021). Multi-stage Fault Diagnosis Framework for Rolling Bearing Based on OHF Elman AdaBoost-Bagging Algorithm. *Neurocomputing*, 433: 237-251. https://doi.org/10.1016/j.neucom.2020.10.003

[13] LI, C., LIU, Y., LIAO, Y., & LIU, W., (2022). Adaptive single-mode variational mode decomposition and its applications in wheelset bearing fault diagnosis. *Measurement Science and Technology*, 33(12): 125008. https://doi.org/10.1088/1361-6501/ac856c

[14] JIANG, X., WANG, J., SHEN, C., SHI, J., HUANG, W., ZHU, Z., & WANG, Q., (2021). An adaptive and efficient variational mode decomposition and its application for bearing fault diagnosis. *Structural Health Monitoring*, 20(5): 2708-2725. https://doi.org/10.1177/1475921720970856

[15] STANCEKOVA, D, RUDAWSKA, A, MRÁZIK, J, TURIAN, F., (2020). Comparision of the bearing rings deformation after heat treatment. *Manufacturing Technology*, 20(5):677-683. Doi: 10.21062/mft.2020.096

[16] LANDAUSKAS, M., CAO, M., & RAGULSKIS, M., (2020). Permutation entropy-based 2D feature extraction for bearing fault diagnosis. *Nonlinear Dynamics*, 102: 1717-1731. https://doi.org/10.1007/s11071-020-06014-6

[17] KHAN, P. W., BYUN, Y. C., & JEONG, O. R., (2023). A stacking ensemble classifier-based machine learning model for classifying pollution sources on photovoltaic panels. *Scientific Reports*, 13(1): 10256. https://doi.org/10.1038/s41598-023-35476-y

[18] MELLIT, A., ZAYANE, C., BOUBAKER, S., & KAMEL, S., (2023). A Sustainable Fault Diagnosis Approach for Photovoltaic Systems Based on Stacking-Based Ensemble Learning Methods. *Mathematics*, 11(4): 936. https://doi.org/10.3390/math11040936

[19] AL-AZZWI, Z. H. N., & NAZAROV, A. N., (2023). Brain Tumor Classification based on Improved Stacked Ensemble Deep Learning Methods. *Asian Pacific journal of cancer prevention: APJCP*, 24(6): 2141-2148. https://doi.org/10.31557/APJCP.2023.24.6.2141

[20] ŤAVODOVÁ, M, VARGOVÁ, M, STANČEKOVÁ, D, HAJDÚCH, A, MRÁZIK, J., (2022). Evaluation of the Influence of Process Parameters on the Mechanical Properties of Castings during High Pressure Die Casting. *Manufacturing Technology*, 22(6):764-770. Doi: 10.21062/mft.2022.079

[21] CHANG, Z, HU, L., (2021). Damage assessment of the rolling bearing based on the rigid-flexible coupling multi-body vibration model. *Manufacturing Technology*, 1(3): 340-348. Doi: 10.21062/mft.2021.048

[22] IMANE, M., RAHMOUNE, C., & BENAZZOUZ, D., (2023). Rolling bearing fault feature selection based on standard deviation and random forest classifier using vibration signals. *Advances in Mechanical Engineering*, 15(4): 1-18. https://doi.org/10.1177/16878132231168503

[23] OMOREGBEE, H. O., & HEYNS, P. S., (2019). Fault Classification of Low-Speed Bearings Based on Support Vector Machine for Regression and Genetic Algorithms Using Acoustic Emission. *Journal of Vibration Engineering & Technologies*, 7(7): 455-464. https://doi.org/10.1007/s42417-019-00143-y

[24] ADLER, A. I., & PAINSKY, A., (2022). Feature Importance in Gradient Boosting Trees with Cross-Validation Feature Selection. *Entropy*, 24(5): 687. https://doi.org/10.3390/e24050687

[25] IQBAL, M., & MADAN, A. K., (2024). Bearing Fault Diagnosis in CNC Machine Using Hybrid Signal Decomposition and Gentle AdaBoost Learning. *Journal of Vibration Engineering & Technologies*, 2(12): 1621–1634. https://doi.org/10.1007/s42417-023-00930-8

[26] MASTELINI, S. M., NAKANO, F. K., VENS, C., & DE LEON FERREIRA, A. C. P., (2023). Online Extra Trees Regressor. *IEEE Transactions on Neural Networks and Learning System*, 34(10): 6755-6767. https://doi.org/10.1109/TNNLS.2022.3212859