A Rolling Bearing Fault Diagnosis Technique Based On Time-Shifted Multi-Scale Attention Entropy and SSA-KELM

Bing Wang, Yu Wang and Xiong Hu

Mechanical Engineering, College of Logistics Engineering, Shanghai Maritime University, Shanghai, China.
E-mail: wangb@shmtu.edu.cn

(Received 27 July 2024; accepted 15 October 2024)

To improve the fault diagnosis efficiency of rotating mechanisms such as rolling bearings, a rolling bearing fault diagnosis technique based on time-shifted multi-scale attention entropy and sparrow search algorithm optimized kernel-extreme learning machine (abbreviated as TSMATE-SSA-KELM) is proposed. Firstly, in the feature extraction part, to solve the issue of insufficient coarse-graining of multi-scale attention entropy (abbreviated as MATE), time-shifted multi-scale attention entropy (abbreviated as TSMATE) is introduced to construct multi-dimensional fault feature vectors. Secondly, the sparrow search algorithm (abbreviated as SSA), which has a fine optimization ability and fast convergence speed, is introduced to optimize the regularization and kernel function parameters of kernel-extreme learning machine (abbreviated as KELM), and the optimal SSA-KELM model is established. Finally, an instance analysis on the Jiangnan University bearing dataset and the bearing dataset of Case Western Reserve University (CWRU) is carried out, and the results show that the proposed technique can achieve 99.85% and 100% accuracy compared with different features and models (MATE, SVM, ELM, etc.). It has potential engineering applications with fast computation speed and high diagnostic efficiency.

1. INTRODUCTION

Rolling bearings are the critical rotating support component in mechanical equipment. They are prone to failure under complex and harsh working conditions and have become a weak link to the healthy operation of mechanical equipment. Both online monitoring and fault diagnosis are very important for enhancing whole machine operation reliability and avoiding significant economic property losses. With the continuous advancement of deep learning in bearing signal fault diagnosis, an increasing number of algorithms have been applied to mechanical fault diagnosis. Zhao et al.[1] proposes a generalized graph convolutional auto-encoder algorithm, which can not only extract sufficient generalized high-dimensional features but also calculate classification loss. Zhu et al.[2] presented a multi-scale convolutional neural network (MSCNNN), which resolves the smearing issue of classification features under different working conditions and improves the performance of diagnostic models. Li et al.[3] proposed a novel matrix-form classifier named LSISMM to obtain more comprehensive fault information. Zhao et al.[4] proposed a Model-Assisted Multi-source Fusion Hypergraph Convolutional Neural Network (MAMF-HGCN) to address the few-shot intelligent fault diagnosis of EHA. Zhao[5] proposed the adaptive activation function is added into the 1-D convolutional neural network (1dCNN) to enlarge the heterogeneous distance and narrow the homogeneous distance of samples. Meimiao et al.[6] propose to decompose the non-stationary signals of rolling bearings with CEEMD method, perform T-distributed Stochastic Neighbor Embedding (t-SNE) clustering visualization analysis, and finally apply kernel extreme learning machine for diagnosis. These studies demonstrate the immense potential and innovative directions of intelligent algorithms in the field of mechanical fault diagnosis. Although the existing methods that combine signal processing with intelligent algorithms have improved diagnostic accuracy, there is still room for improvement in terms of the automation level of feature extraction, the generalization ability of the models, and other aspects.

Feature extraction is the primary step for fault diagnosis, and the key lies in extracting parameters that can characterize different failure modes.[7] Due to the complex operating environment, the vibration signal of mechanical equipment often presents a non-linear and non-stationary character. Thus, traditional character parameters based on time-domain and frequency-domain analysis may not be able to characterize failure modes sensitively. With the deepening of the study of feature extraction, parameters based on information entropy have been proposed and gradually applied to fault diagnosis and prediction. For example, sample entropy, approximate entropy, fuzzy entropy, permutation entropy and so on.[8–10] The calculation of the above information entropy is based on phase space reconstruction with spatial distance calculation and statistics, and with the increase in signal length, the calculation efficiency will decrease dramatically.

To overcome the shortcomings of the above techniques, Yang[11,12] proposed a new tool for measuring signal complexity — attention entropy. Unlike traditional entropy parameters, which focus on the frequency distribution of the whole data in a time series, the attention entropy only focuses on the frequency

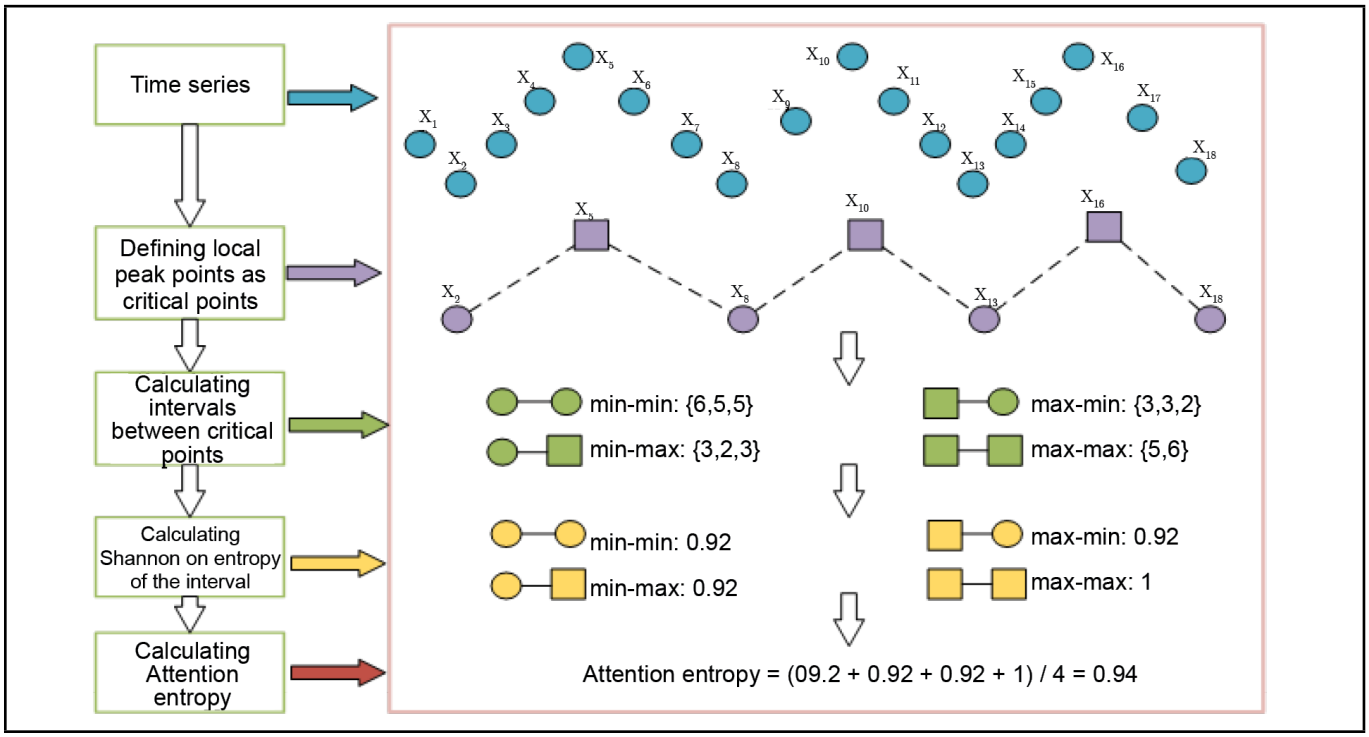


Figure 1. Flowchart of attention entropy.

distribution of the interval between the peak points of the series. Therefore, the attention entropy shows some advantages including fewer hyper-parameters, shorter running time, and strong robustness to the time series length. Considering that attention entropy cannot measure signal complexity at multiple scales, attention entropy and multiscale entropy are combined, and multiscale attention entropy (MATE) is constructed. Meanwhile, to solve the problem of signal information loss for coarse-graining in multiscale entropy, the time-shift technique[13] is introduced, and time-shift multiscale attention entropy is proposed(abbreviated as TSMATE) to extract bearing fault features effectively. By introducing the concept of time shift, TSMATE avoids the coarse-graining process, enabling a more flexible preservation of information within the original time series. By performing translation operations on the time series at different scales, TSMATE can more comprehensively reveal the dynamic characteristics of the time series, thereby enhancing the richness of fault information and making fault features more pronounced. Additionally, TSMATE exhibits good noise resistance and robustness.

The essence of fault diagnosis is a multi-classification problem for fault patterns. Huang[14] proposed the Kernel Extreme Learning Machine (KELM) in 2014; the single hidden layer feed-forward neural network has some advantages, such as a simple structure, fast operation speed, strong generalization ability, etc. Liu[15] demonstrated that when the sample size is large, the accuracy of KELM and SVM models do not differ much; however, the calculation complexity of the KELM model is much smaller. Research shows that the accuracy of the KELM model depends on the selection of kernel parameters and regularization coefficients. Some optimization algorithms are introduced for the selection of model parameters, including particle swarm optimization,[16,18] firefly algorithm[17] and so on. In the study of the swarm intelligence optimization algorithm, the sparrow search algorithm (SSA) was proposed by Xue Jiankai[19] in 2020. Compared with the above methods, the SSA method has the advantages of high search accuracy, fast convergence speed, good stability, and robustness compared with traditional algorithms.[20] Based on this consideration, SSA is introduced in the optimization of KELM model parameters and constructs the SSA-KELM fault diagnosis model. The integration of SSA-KELM, where SSA (Sparrow Search Algorithm) is utilized to optimize the regularization and kernel function parameters of KELM (Kernel Extreme Learning Machine), significantly enhances the performance of KELM. This optimization enables KELM to better adapt to diverse datasets and practical application scenarios. By combining the global search capability of SSA with the kernel method advantages of KELM, SSA-KELM maintains a rapid training speed while also bolstering the model's generalization ability. As a result, SSA-KELM becomes more robust and reliable when dealing with complex and dynamic datasets, offering a compelling solution for handling intricate machine learning challenges.

In summary, leveraging the advantages of feature extraction through TSMATE and the optimized model of SSA-KELM, we propose a bearing fault diagnosis technique based on TSMATE-SSA-KELM. Above all, a fault diagnosis method based on TSMATE and SSA-KELM (abbreviated as TSMATE-SSA-KELM) is proposed. Datasets from Jiangnan University and Case Western Reserve University (CWRU) are used to analyze the influence of parameters and verify the model's effectiveness. The paper is organized as follows: Section II proposed the feature extraction method based on TSMATE. In Section III, SSA is introduced into the parameter optimization of KELM. The whole fault diagnosis procedure

is discussed in Section IV and the technique is verified with a bearing dataset in Section V. Finally, the conclusion of this paper is given in Section VI.

2. FEATURE EXTRACTION BASED ON TIME-SHIFTED MULTI-SCALE ATTENTION ENTROPY

2.1. Attention Entropy (ATE)

Attention entropy has the advantages of robustness to time series length and no hyperparameters, etc. The flowchart of the attention entropy is shown in Fig. 1, and the main steps can be summarized as follows:

1. If each point in the time series is considered a system, its state change can be regarded as an adjustment to the environment. The peak points can effectively characterize the changes in the upper and lower bounds of the local state, so the local peak points are defined as key points.

2. Set as key points according to four different strategies, {min-min}, {min-max}, {max-min}, and {max-max}, and calculate the number of interval points between neighboring key points.

3. Calculating Shannon entropy of neighbouring key point intervals, the specific formula is as follows.

$$H(x) = -\sum_{x=1}^{k} p(x)\log_2 p(x); \tag{1}$$

where $p(x)$ is the probability of x occurring and b is the number of interval point species.

4. The mean value of Shannon's entropy was defined by the four different strategies, which are called attention entropy.

2.2. Time-shifted Multiscale Attention to Antropy (TSMATE)

Multiscale entropy was proposed by Costa et al.[21] to measure signal complexity at multiscale metrics. In this paper, multiscale entropy and attention entropy are combined, and a time-shifted rule is introduced to propose the technique of TSMATE, which overcomes the problems of signal information loss for coarse-graining in multiscale entropy. The specific calculation process of TSMATE is as follows:

1. Splitting time-series signal with length N into k subsequences as follows:

$$Y_k^{\beta} = \left\{x_{\beta}, x_{\beta+k}, x_{\beta+2k}, \ldots, x_{\beta+k[(N-\beta)/k]}\right\}; \tag{2}$$

where $\boldsymbol{x}$ is the sample point of the original signal; $\boldsymbol{k}$ is the number of segmentation subsequences; Y_k^{β} is the first β subsequence.

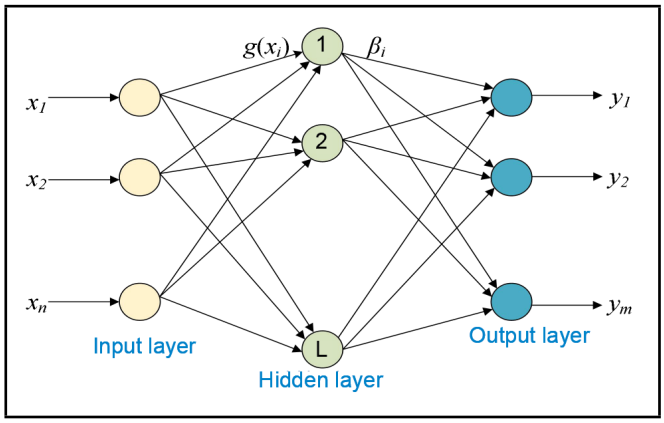


Figure 2. ELM network topology

2. Calculating the attention entropies of the time-series signals at all scales and define the mean of these attention entropies at that scale K as the:

$$TSMA(k) = \frac{1}{k}\sum_{\beta=1}^{k} ATE(Y_k^{\beta}); \tag{3}$$

where $ATE(Y_k^{\beta})$ is the value of the entropy of attention of Y_k^{β}.

3. Calculating TSMA(K) for all K using Eq. (3) and use the set of these values as TSMATE.

3. KELM MODEL OPTIMIZATION WITH SPARROW SEARCH ALGORITHM

3.1. Theory of Sparrow Search Algorithm

Sparrow Search Algorithm (Sparrow Search Algorithm) obtains the optimal solution by simulating the foraging and anti-predation behaviors of sparrows, which has the advantages of fast convergence speed and high stability.[22] Let there are $\boldsymbol{N}$ sparrows in the D-dimensional space, and the position of the $\boldsymbol{i}$th sparrow in the D-dimensional search space is $X_{id} = [X_{i1}, \ldots, X_{id}, \ldots, X_{iD}]$, where $i = 1, 2, ..., N$. At each iteration, the position distributions of the discoverer, the follower, and the detector are updated according to Eqs. (4), (5), and (6):

$$X_{ij}^{t+1} = \begin{cases} X_{ij}^{t}\exp\left(\frac{-i}{\alpha\cdot iter_{\max}}\right), & R_2 < ST \\ X_{ij}^{t} + QL, & R_2 \geq ST \end{cases}; \tag{4}$$

where: $\boldsymbol{t}$ is the number of iterations; $iter_{\max}$ is the maximum number of iterations; $\boldsymbol{\alpha}$ is a uniform random number between (0,1]; $\boldsymbol{Q}$ is a random number obeying a normal distribution; $\boldsymbol{L}$ is a 1*dimensional matrix whose elements are all $\mathbf{1}$; $R_2 \in [0, 1]$ is the warning value, and $R_2 \in [0.5, 1]$ denotes the safety value. When $R_2 < ST$, no predator is detected, the detector searches widely and guides the population to obtain higher fitness; when $R_2 \geq ST$, the detector finds a predator and releases a signal, the population immediately engages in anti-predator behavior and moves closer to the safe zone.

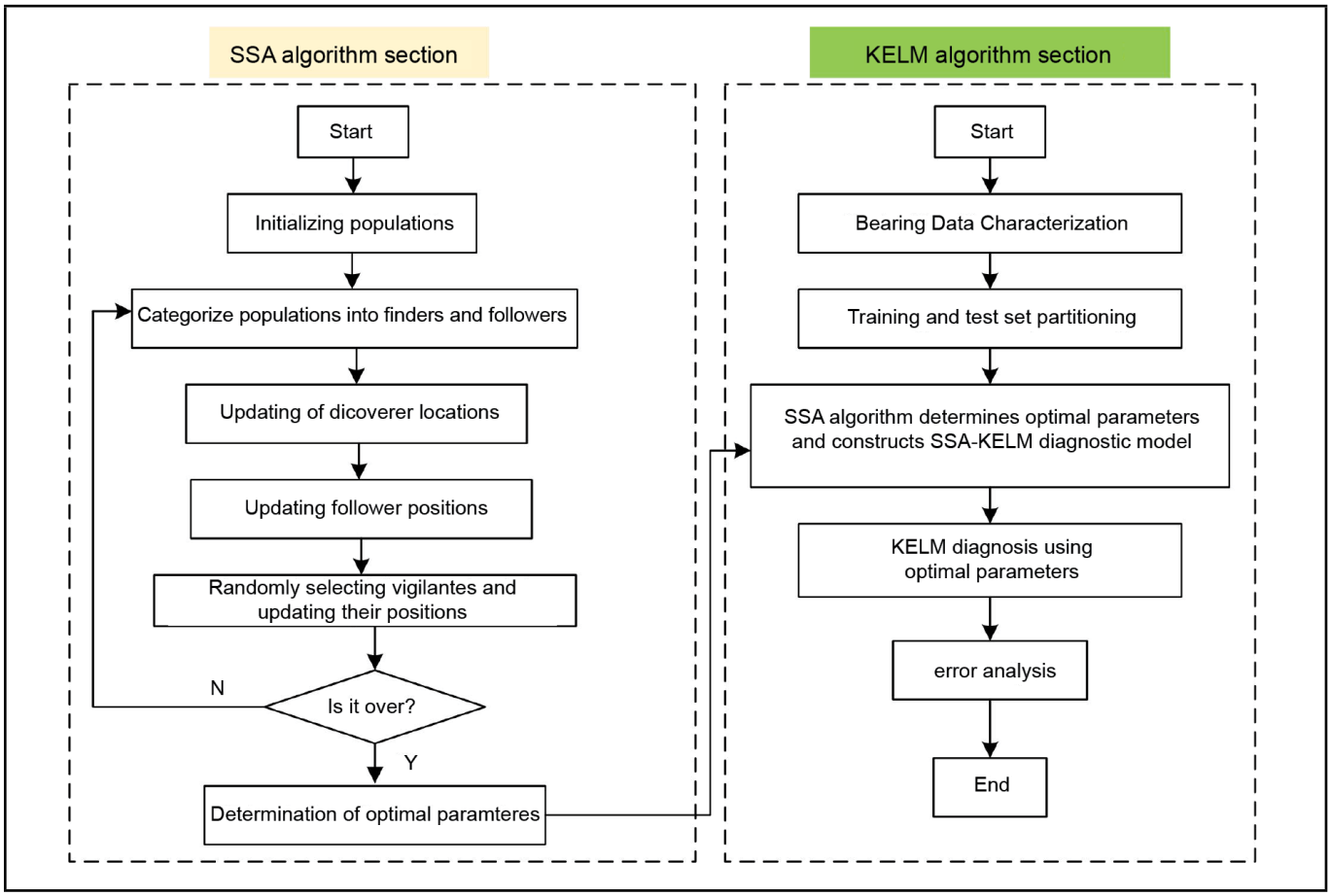


Figure 3. The flowchart of SSA-KELM.

$$X_{id}^{t+1} = \begin{cases} Qexp\left(\frac{xw_d^t - X_{id}^t}{i^2}\right), \ i > \frac{n}{2} \\ xb_d^{t+1} + \frac{1}{D}\sum_{j=1}^{D}\left(rand\{-1,1\} \cdot \left|X_{id}^t - xb_d^{t+1}\right|\right), \ i \le \frac{n}{2} \end{cases}; \tag{5}$$

where xw_d^t is the worst position of the sparrow in d-dimension for the $\boldsymbol{t}$th iteration; xb_d^{t+1} is the optimal position of the sparrow in d-dimension for the $(\boldsymbol{t+1})$th iteration. When $i > \frac{n}{2}$, the $\boldsymbol{i}$th follower searches near the current optimal position $\boldsymbol{xb}$.

$$x_{id}^{t+1} = \begin{cases} xx_d^t + \beta(x_{id}^t - xb_d^t), \ f_i \neq f_g \\ x_{id}^t + K\frac{x_{id}^t - xw_d^t}{|f_i - f_w| + \varepsilon}, \ f_i = f_g \end{cases}; \tag{6}$$

where: $\boldsymbol{\beta}$ and K are step control parameters: $\boldsymbol{\beta}$ is a random number with mean 0 variance; $K \in [-1,1]$ is the moving direction of the sparrow; $\boldsymbol{\varepsilon}$ is a very small number to avoid the denominator being 0; f_i, f_g, w are the current individual sparrow, the optimal and the worst fitness values, respectively. When $f_i \neq {}_g$, the sparrow is at the edge of the population and is highly vulnerable to predator attacks; when $f_i = f_g$, the sparrow realizes the danger and approaches other sparrows to adjust its search strategy.

3.2. Theory of KELM

3.2.1. Extreme Learning Machine (ELM)

The ELM algorithm is derived from a single hidden layer feedforward neural network, the network topology is shown in Fig. 2, with x_i and y_i denoting the inputs and outputs of the structural network. The number of neurons in the input, hidden, and output layers are n, l, and m, respectively. b_i is the threshold value of the neurons in the hidden layer, and the weights of the connections between the input and hidden layers correspond to ω_i and β_i. The input and output layers are represented by x_i and y_i.[23]

Suppose that for a training set (x_i, t_i), where $x_i \in R^n$, $t_i \in R^m$, and $\boldsymbol{T}$ are the desired outputs with sample size $\boldsymbol{Q}$, and denoting the hidden layer neuron activation function by $\boldsymbol{g(x)}$, the output expression:

$$\sum_{i=1}^{Q}\beta_i g_i\left(\omega_i x_j + b_i\right) = y_i, \ j = 1, 2\ldots, Q. \tag{7}$$

When the feedforward neural network approximates Q samples with zero error, $\sum_{j=1}^{Q}\|t_i - y_i\| = 0$, simplification yields: $H\beta = T$; where: $\boldsymbol{H}$ is denoted as the hidden layer output matrix and the detailed expression is:

$$H\left(\omega_1, \omega_2, \cdots \omega_l, b_1, b_2, \cdots, b_l, x_1, x_2, \cdots, x_l\right) = \begin{bmatrix} g(\omega_1 x_1 + b_1) & \cdots & g(\omega_l x_1 + b_l) \\ \vdots & \ddots & \vdots \\ g(\omega_1 x_Q + b_1) & \cdots & g(\omega_l x_Q + b_l) \end{bmatrix}. \tag{8}$$

Calculate the least squares solution of $\min_\beta \|H\beta - T\|$ to obtain the optimal link weights β:

$$\beta = \mathbf{H}^{+}\mathbf{T}; \tag{9}$$

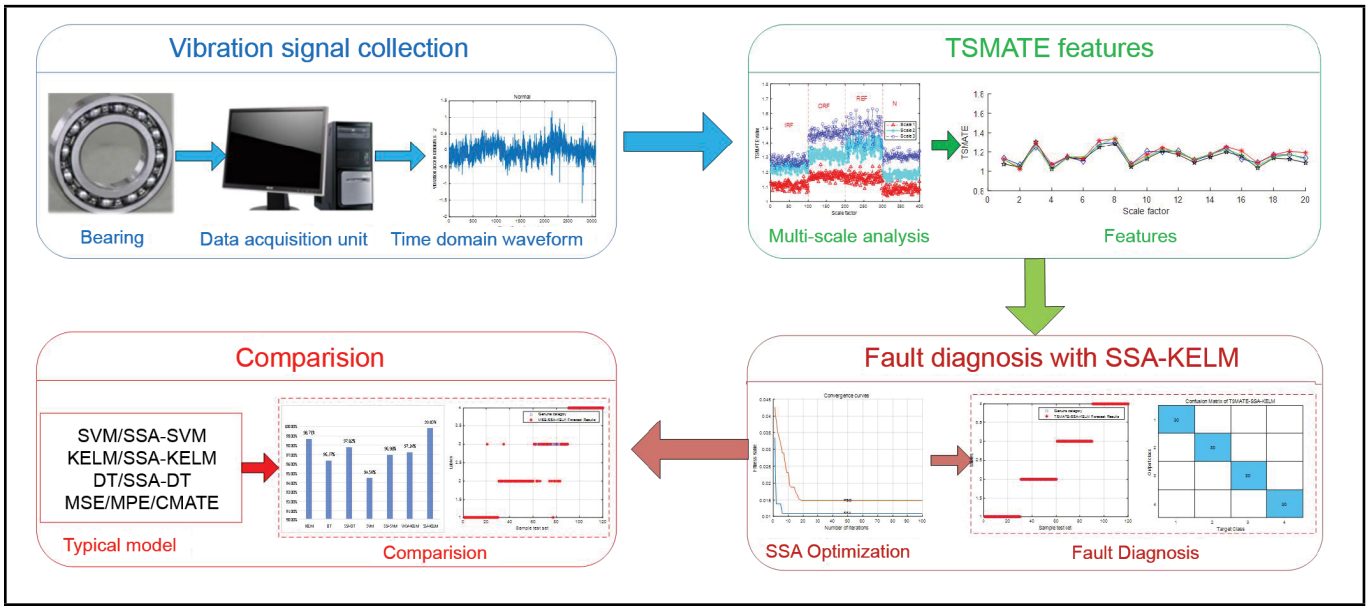


Figure 4. The flowchart of fault diagnosis based on TSMATE-SSA-KELM.

HH^T often exhibits non-singularity when solving the pseudo-inverse matrix $H^+ = H^T(HH^T)^{-1}$. A penalty coefficient C is introduced to deviate the eigenroots of HH^T from the zero value, where **1** is the diagonal matrix, and the final weights of the solution are found:

$$\beta = H^T\left(\frac{1}{C} + HH^T\right)^{-1} T. \tag{10}$$

The output model of ELM is obtained:

$$f(x_i) = h(x_i)\,\beta = h(x_i)H^T\left(\frac{1}{C} + HH^T\right)^{-1} T; \tag{11}$$

where $h(x_i)$ is the implicit layer output and x_i is the $\boldsymbol{i}^{th}$ sample of $\boldsymbol{Q}$.

3.2.2. Kernel Extreme Learning Machine (KELM)

Because ELM randomly sets the hidden layer threshold and weights easily lead to unstable prediction results,[24] to improve the generalization ability of the ELM model, the RBF kernel function is introduced on its basis to map the input samples to the high-dimensional feature space, and at the same time to solve the problem of the low-dimensional linear indivisibility. The kernel matrix is used to replace the random matrix of ELM:

$$HH^T = \Omega_{ELM} = \begin{bmatrix} K(x_1,x_1) & \cdots & K(x_1,x_Q) \\ \vdots & \ddots & \vdots \\ K(x_Q,x_1) & \cdots & K(x_Q,x_Q) \end{bmatrix}. \tag{12}$$

The losses and weights of KELM are calculated from Eqs. (10), (11) and (12) as follows:

$$f(x_i) = \begin{bmatrix} K(x,x_1) \\ \cdot \\ \cdot \\ \cdot \\ K(x,x_Q) \end{bmatrix}^T \left(\frac{1}{C} + \Omega_{ELM}\right)^{-1}; \tag{13}$$

Table 1. Description of experimental JN dataset.

Fault type	Sampling length	Numbers of Sampling data	Label Type
Inner ring failure(IRF)	3072	100	1
Outer ring failure(ORF)	3072	100	2
Rolling element failure(REF)	3072	100	3
Normal(N)	3072	100	4

$$\Gamma = \left(\frac{1}{C} + \Omega_{ELM}\right)^{-1} y. \tag{14}$$

The expression for the Gaussian kernel function chosen in this paper is as follows:

$$K(x,x_i) = e^{-\frac{\|x-x_i\|^2}{2\sigma^2}}. \tag{15}$$

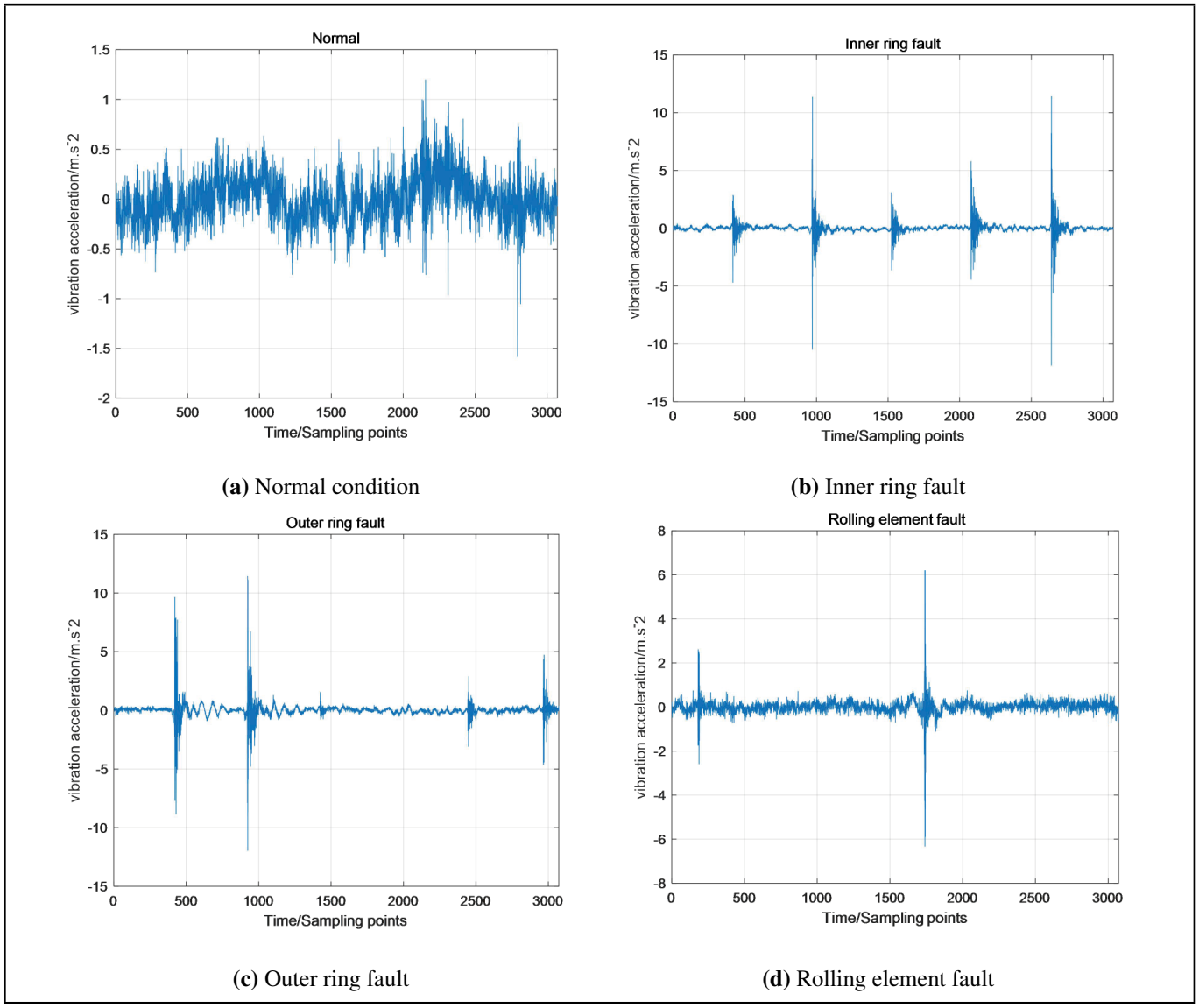
3.3. Flow of SSA Optimized KELM

The KELM neural network training process is based on a certain degree of randomness, which leads to the need for more hidden layer neurons in the computation, which in turn will increase the amount of computation and make the kernel limit learning machine consume more computing time in the testing process. In this paper, the regularization coefficients and kernel parameters of the kernel limit learning machine are optimized using the sparrow search algorithm, and the diagnostic model for bearing fault types is constructed after the optimal parameters are finally determined. The flowchart of SSA-KELM is shown in Fig. 3.

4. FAULT DIAGNOSIS MODEL BASED ON TSMATE-SSA-KELM

To improve the accuracy of rolling bearing fault diagnosis, a fault diagnosis technique based on TSMATE-SSA-KELM was proposed and the basic flowchart is shown in Fig. 4 below.

The technique mainly included the following steps:



(a) Normal condition

(b) Inner ring fault

(c) Outer ring fault

(d) Rolling element fault

Figure 5. Time domain waveforms for different conditions of rolling bearings.

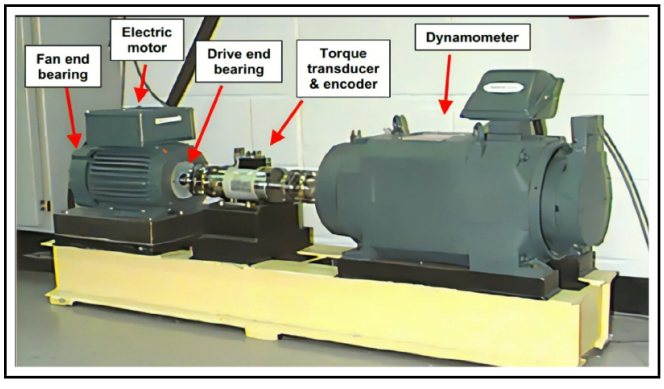


Figure 6. CWRU bearing data test bench.

1. Vibration signal acquisition. Monitoring and collecting vibration signals of rolling bearings under different conditions as a fault diagnosis dataset.

2. Fault feature extracting based on TSMATE. In turn, D_i is subjected to multi-scale analysis for Each group of samples. The attention entropy at different scales is calculated to extract multi-dimension TSMATE feature vectors, which are used to represent different conditions of rolling bearings. In this paper, the maximum analytical scale is set as $scale = 20$ to get abundant information of signal.

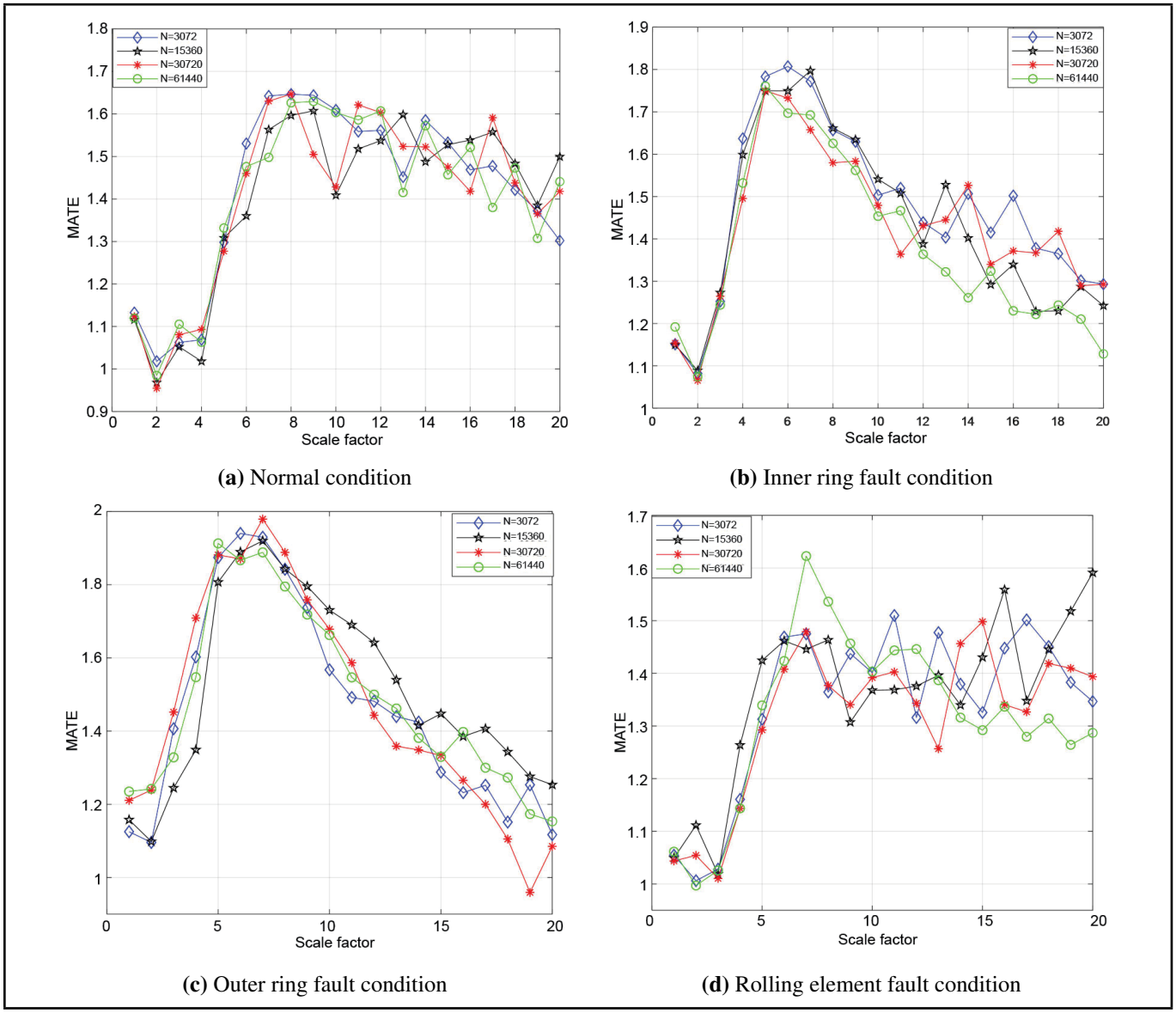
3. SSA is introduced to optimize parameters C and γ of KELM and establish an SSA-KELM fault diagnosis model. Samples are divided into training and testing sets.

4. Sample testing and comparative analysis. Diagnosis on the testing set. The model's accuracy is tested, and different features and models are imported for comparison to verify the method's superiority.

5. INSTANCE ANALYSIS

5.1. Introduction of Experimental Data

The experimental data was obtained from Jiangnan University, which was abbreviated as the JN dataset.[25] The data sampling frequency was 50 kHz, and the rotational speed was 600 r/min. The dataset was divided by 3072 sampling points as



(a) Normal condition

(b) Inner ring fault condition

(c) Outer ring fault condition

(d) Rolling element fault condition

Figure 7. Distribution of MATE features under different scale factor and signal length.

Table 2. Description of experimental CWRU dataset.

Fault type	Sampling length	Numbers of Sampling Data	Label Type	Data
Inner ring failure (IRF)	1024	100	1	105.mat
Outer ring failure (ORF)	1024	100	2	130.mat
Rolling element failure (REF)	1024	100	3	118.mat
Normal(N)	1024	100	4	97.mat

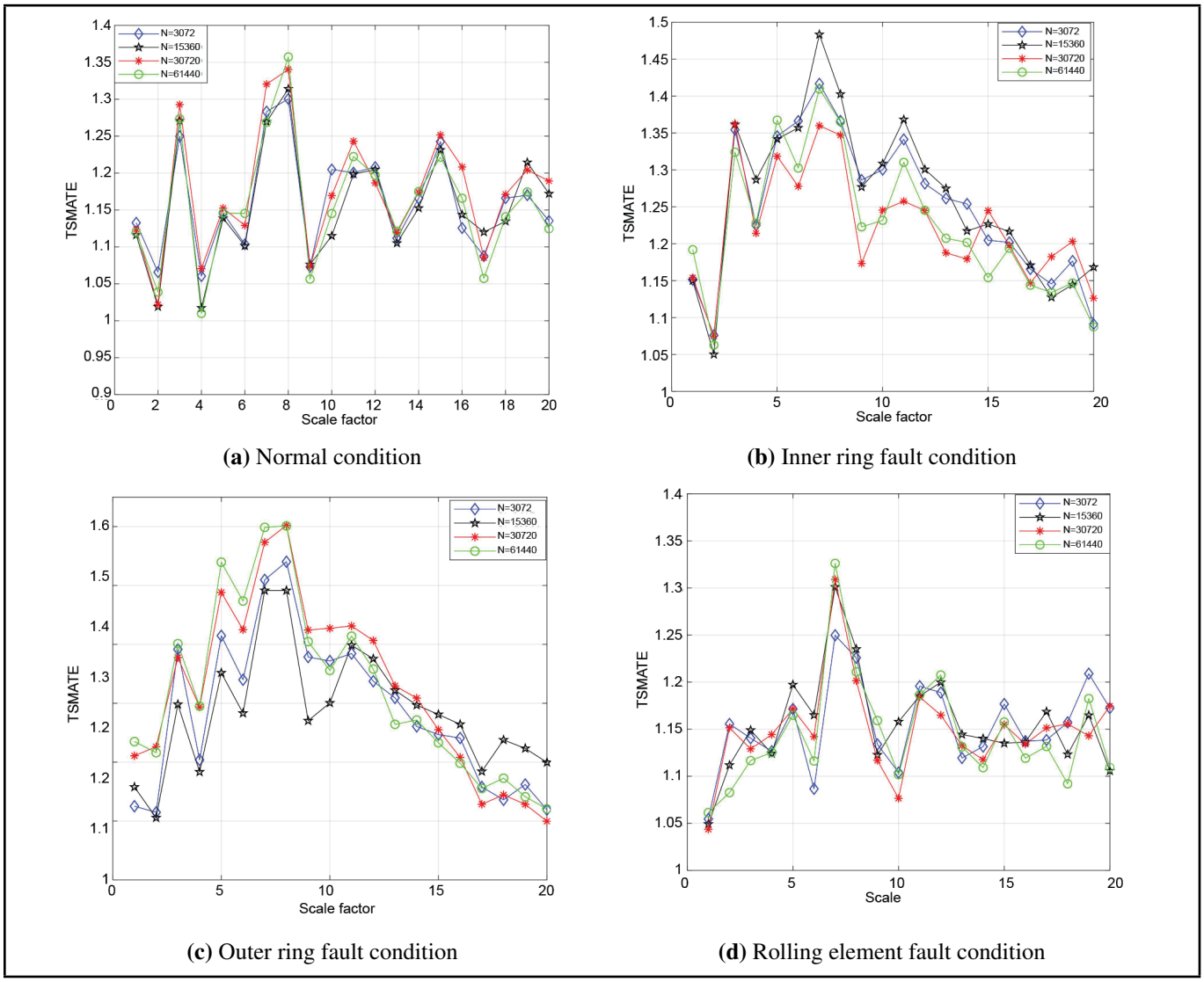
the window width, and four groups of samples with different fault conditions were obtained, including normal (100 groups), inner ring faults (100 groups), outer ring faults (100 groups), and rolling element faults (100 groups). The time-domain waveforms of the four conditions are shown in Fig. 5. The signals show different time-domain amplitude, impact frequency, and distribution characters.

More experimental data was obtained from Case Western Reserve University, which is abbreviated as the CWRU dataset.[25] The drive end SKF6205-2RS type deep groove ball bearing was selected as the research object. When the motor speed was 1797 r/min, the sampling frequency of the vibration signal was 12 KHz. Faults include inner ring, outer ring, and rolling body of the bearing respectively. As shown in Fig. 6, it is the test bench of this data.

5.2. Feature Extraction Based TSMATE

Extracting features for each group of samples based on TSMATE technique and fault feature vectors with the structure of 400×20 is extracted.

To analyze the superiority for time-shifted rule for TSMATE technique under different signal length and analytical scale, the MATE technique was introduced for comparison with different signal length and analytical scale, the fault features under the four conditions are shown in Fig. 7 and Fig. 8. It is clear that the distribution of MATE features is relatively fluctuating com-



(a) Normal condition

(b) Inner ring fault condition

(c) Outer ring fault condition

(d) Rolling element fault condition

Figure 8. Distribution of TSMATE features under different scale factor and signal length.

pared with TSMATE features. Taking the curve with length of $N = 3072$ as an example, the peak-to-peak value which reflects the fluctuation magnitude of MATE curve on the four conditions reaches 0.62587, 0.72526, 0.84429, 0.50396 respectively, while value of TSMATE curve is 0.23989, 0.34159, 0.42558, 0.18831, indicating the TSMATE features has good robustness to scale factors and signal length.

Keeping the multi-scale analysis part consistent, different fault feature extraction techniques are introduced for comparison including multi-scale permutation entropy (MPE), multi-scale sample entropy (MSE), and multi-scale attention entropy (MATE). The features on any three scales including scale 1, scale 7 and scale 11 are selected for contrast in Fig. 9. TSMATE features have the lightest aliasing phenomena, which indicates that the feature vectors can distinguish different failure conditions.

In addition, Table 3 shows the running time of MPE, MSE, MATE, and TSMATE to compare the operation efficiency of the proposed method. TSMATE and MATE have a greater computing speed than the other two techniques. The introduction of the time-shifting rule takes more time for the TSMATE technique than for MATE. Above all, the proposed TSMATE technique is efficient and stable.

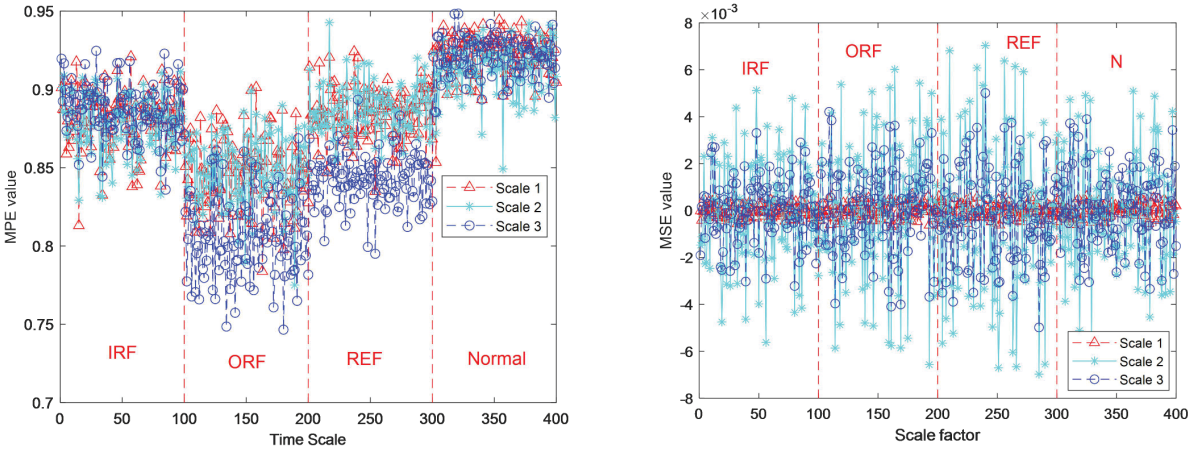
Table 3. Compaison with the running time for different feature extraction tecniques.

Feature extraction techniques	Data length	Runing time/s
MPE	400*3072	36.73
MSE	400*3072	64.10
MATE	400*3072	12.42
TSMATE	400*3072	**36.81**

5.3. Diagnosis Based on SSA-KLEM

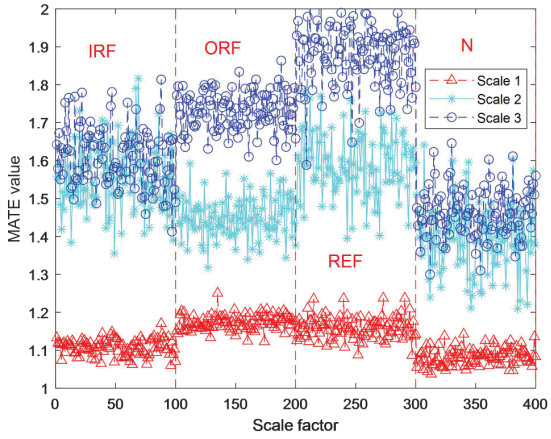
Dividing the obtained features matrix into training and testing samples according to the ratio of 7:3. SSA was introduced to optimize the regularization coefficient C and kernel function parameter γ for KELM model. The parameters of the SSA algorithm are set as follows: the number of sparrow populations was set as 20, the maximum iterations number was set as 100, the range of C was set as $[0.1, 50]$ and γ is set as $[0.1, 10]$. The optimized SSA-KELM model is obtained while the parameters are optimized as $C = 49.52, \gamma = 1.87$.

Fig. 10 shows the confusion matrix of the testing samples before and after optimization. The proposed TSMATE-SSA-

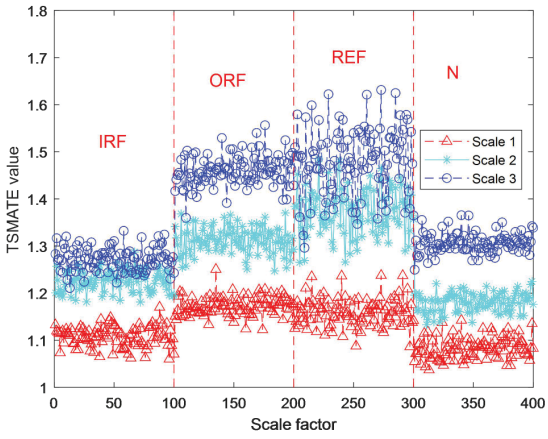


(a) Multi-scale arrangement entropy

(b) Multi-scale sample entropy



(c) Multi-scale attention to entropy



(d) Time-shifted multiscale attention to entropy

Figure 9. Distribution of different multi-scale entropy measures.

KELM reached an accuracy of 100% from 98.33% on account of the introduction of the SSA algorithm.

To compare the effectiveness of different optimization algorithms on KELM, this paper optimizes KELM separately using Whale Optimization Algorithm (WOA), Particle Swarm Optimization (PSO), and Firefly Algorithm (FA), and obtains their convergence curves. As shown in Fig. 11, among the four different algorithms, SSA exhibits the fastest convergence capability and can reach the minimum value quickly, indicating that SSA optimization has the best convergence effect.

To avoid the influence on sample division and the randomness of diagnosis results, 10-CV was introduced for sample division, and the diagnosis process is repeated 100 times. The accuracy of the two models is shown in Fig. 12, with the mean values of the two datasets reaching 99.85% and 100% respectively. Obviously, the proposed model has lower error and higher accuracy.

5.4. Comparative Analysis

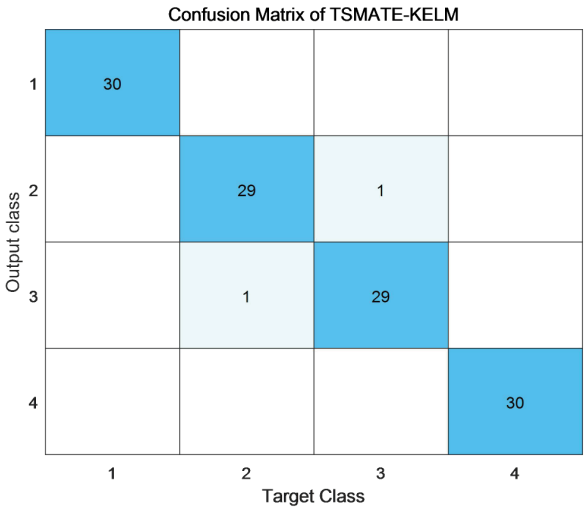
To verify the effectiveness and superiority of the technique proposed in the paper, comparative analyses are carried out in terms of feature selection and diagnosis model respectively.

5.4.1. Comparation of feature selection

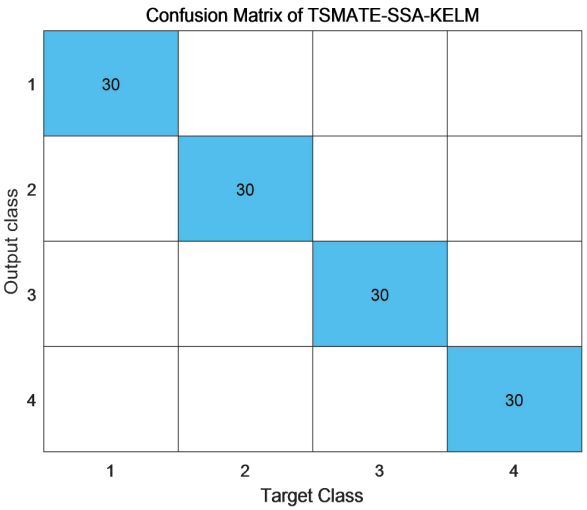
Keeping the SSA-KELM model consistent, different fault features were extracted and fed into the model for training and testing including MPE, MSE, MATE, (CMATE, Composite multiscale attention entropy), (RCMATE, Fine composite multiscale attention entropy). The diagnosis results are shown in Fig. 13. It is apparent that the TSMATE-SSA-KELM model has the best effect, and the accuracy reaches 100%, mainly because the method is independent of the signal length and has good robustness.

5.4.2. Comparison of diagnostic models

Taking TSMATE as the fault features, typical supervised learning methods including DT (Decision Tree), KELM (Kernel-Extreme Learning Machine) and SVM (Support Vector Machine) were introduced for model comparison. Critical parameters of the three models are also optimized with SSA algorithm. The default values of parameters were set first; for the SVM model, the kernel function parameter is set as 50, and the penalty factor was set as 0.5. The key parameter was set as $num\_split = 10$ for the DT model, the kernel function parameter was set as 2 and the penalty factor was set as 4 for the KELM model. The average value of accuracy for one hundred times is shown in Fig. 14. The three supervised



(a) TSMATE-KELM model



(b) TSMATE-SSA-KELM model

Figure 10. Confusion matrix of different models.

learning methods all have good classification accuracy, and the accuracy is improved after optimization with SSA. Comparatively, the KELM model has a superior performance with an optimized accuracy of 99.85%. According to Table 4, it can be found that compared with other models, TSMATE-SSA-KELM has the smallest average error rate and the average F1 value is the closest to 1. However, in terms of consumption time, TSMATE-SSA-KELM is slightly longer than other models, but overall, the efficiency of this model is still very high.

It has been proved that the advantages of the proposed TSMATE-SSA-KELM method lie in its high accuracy of fault diagnosis for mechanical bearings, small error, relatively stable diagnosis results, short time consumption and high efficiency.

6. CONCLUSION

A rolling bearing fault diagnosis technique based on TSMATE-SSA-KELM is proposed. The validity and superiority are verified with the Jiangnan University dataset. The experimental results show that the proposed technique has potential engineering applications with fast computation speed and

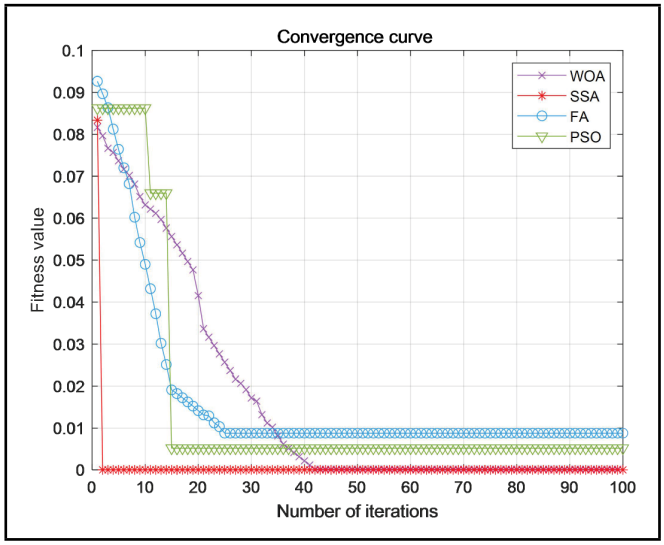


Figure 11. Convergence curves of different optimization algorithms.

Table 4. Comparison of errors and efficiency among various models.

Models	Average Error /%	Average F1	Time /s
KELM	1.29	0.979	48.01
DT	3.63	0.953	42.94
SSA-DT	2.18	0.980	60.85
SVM	5.5	0.932	53.72
SSA-SVM	3.02	0.961	76.88
WOA-KELM	2.76	0.974	78.61
PSO-KELM	1.31	0.974	62.17
SSA-KELM	**0.15**	**0.991**	**64.03**

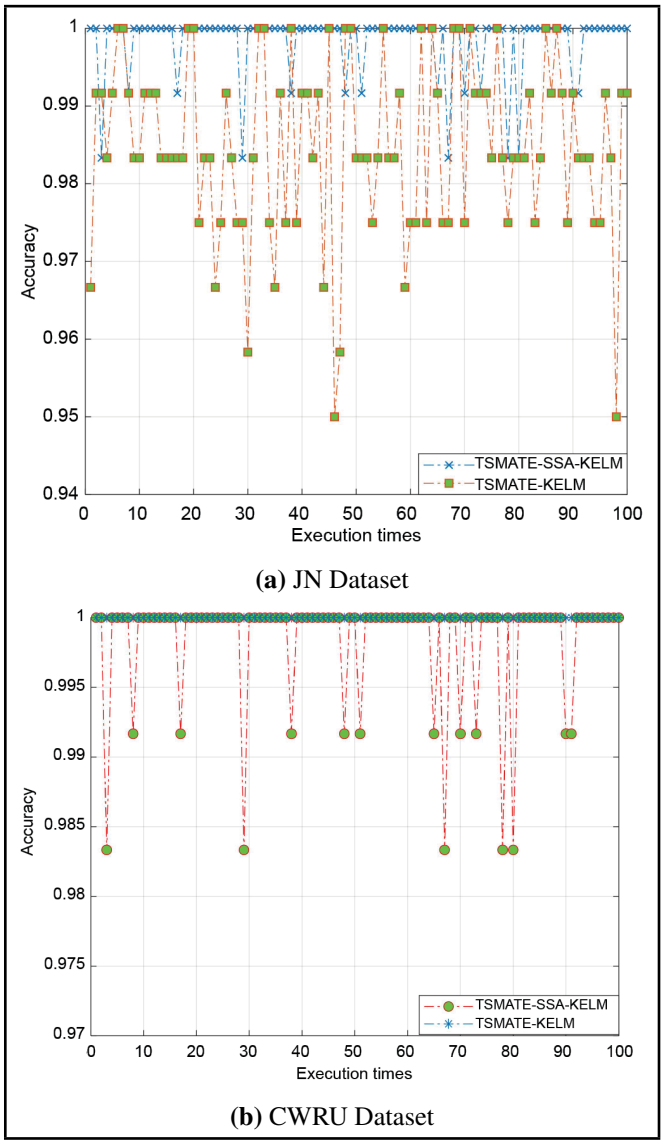
high diagnostic efficiency. The following three conclusions are obtained.

1. Aiming at the information loss issue of the MATE technique, the time-shifting rule is introduced in the TSMATE method. After comparison, TSMATE is basically unaffected by the signal length and has good robustness.

2. Comparing the distribution graph with different multiscale entropy, the proposed TSMATE technique is efficient and stable.

3. The optimization of KELM using the SSA algorithm can avoid the blindness of parameter selection, thus improving fault diagnosis accuracy. The KELM model performs better under minor sample conditions than similar supervised learning methods. The proposed method is higher than the conventional multiscale entropy.

4. The proposal of the TSMATE-SSA-KELM algorithm brings a new combination of algorithms to the field of machine learning. The effectiveness of the method in the field of fault diagnosis has been proven through JN dataset and CWRU dataset. It demonstrates the feasibility and effectiveness of combining different optimization algorithms with machine learning models to enhance performance. This combination not only enriches the ma-



(a) JN Dataset

(b) CWRU Dataset

Figure 12. The accuracy of diagnosis results obtained by 100 times.

chine learning algorithm library but also provides inspiration for future algorithmic innovations. However, the training time of the model in this paper is relatively long, so the next step of the research is to shorten the training time of the model while ensuring the accuracy and robustness.

FUNDINGS

This research received specific grant from National High Technology Research Development Plan (2013AA041106), National Natural Science Foundation (62073213), China Postdoctoral Science Foundation(2014M561458), Shanghai Natural Science Foundation of China (23ZR1426700), Shanghai Engineering Technology Research Center Construction projects(20DZ2253300).

DECLARATION OF CONFLICT

The authors declare no conflict of interest in preparing this article.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the Jiangnan University at http://www.52phm.cn/datasets/bear/Bearing-data-set-of-Jiangnan-University.html.

REFERENCES

[1] Zhao, X., Hu, Y., Liu, J., et al. A novel intelligent multicross domain fault diagnosis of servo motor-bearing system based on Domain Generalized Graph Convolution Autoencoder. *Structural Health Monitoring*, 2024. https://doi.org/10.1177/14759217241262722

[2] Zhu, X., Zhao, X., Yao, J., Deng, W., Shao, H. and Liu, Z. Adaptive Multiscale Convolution Manifold Embedding Networks for Intelligent Fault Diagnosis of Servo Motor-Cylindrical Rolling Bearing Under Variable Working Conditions, in *IEEE/ASME Transactions on Mechatronics*, **29**, 2230–2240, (2024). https://doi.org/10.1109/TMECH.2023.3314215

[3] Li, X., Shao, H., Lu, S., Xiang, J., and Cai, B. Highly Efficient Fault Diagnosis of Rotating Machinery Under Time-Varying Speeds Using LSISMM and Small Infrared Thermal Images, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, **52**, 7328–7340, (2022). https://doi.org/10.1109/TSMC.2022.3151185

[4] Zhao, X., Zhu, X., Liu, J., Hu, Y., Gao, T., Zhao, L., Yao, J., and Liu, Z. Model-Assisted Multi-source Fusion Hypergraph Convolutional Neural Networks for intelligent few-shot fault diagnosis to Electro-Hydrostatic Actuator, *Information Fusion*, **104**, 102186, 1566–2535, (2024). https://doi.org/10.1016/j.inffus.2023.102186

[5] Zhao, X. et al. Intelligent Fault Diagnosis of Gearbox Under Variable Working Conditions With Adaptive Intraclass and Interclass Convolutional Neural Network, in *IEEE Transactions on Neural Networks and Learning Systems*, **34**, 6339–6353, (2023). https://doi.org/10.1109/TNNLS.2021.3135877

[6] Mao, M., Xiao, w., Chen, X., Wang, J., and Wang, L. Fault diagnosis method of rolling bearing based on CEEMD energy entropy and extreme learning machine, *Journal of Ordnance Equipment Engineering*, **44**, 279–285, (2023). https://doi.org/10.1016/j.isatra.2023.05.014

[7] Ruan, W., Ma, Z., and Li, Y. Rolling bearing fault diagnosis method based on variational modal layout gradient and material group optimization support machine, *Journal of Jinan University (Natural Science Edition)*, **32**(04), 291–296, (2018). https://doi.org/10.13349/j.cnki.jdxbn

[8] Zhao, Y., Cu, L., Li, K. et al. Fault diagnosis of rolling bearings based on improved MPE and KELM, *Noise and Vibration Control*,**42**(01), 125–131, (2022). https://doi.org/10.1155/2021/9933137

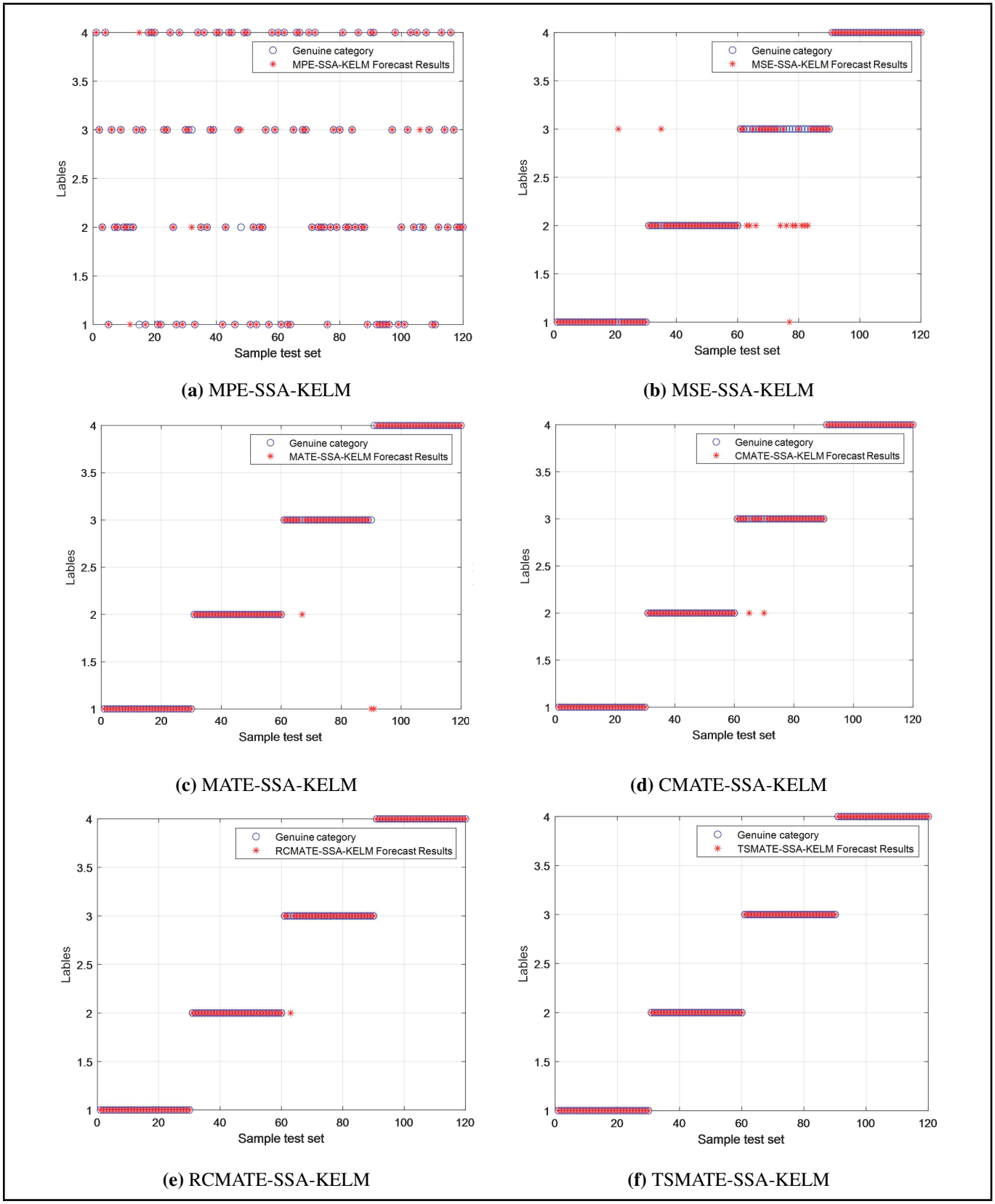


Figure 13. Comparison of diagnosis results for different models

[9] Li, X. and Jin, W. Research on rolling bearing fault diagnosis based on optimized support vector machine with improved sparrow algorithm, *Vibration and Shock*, **42**(06), 106–114, (2023). https://doi.org/10.13465/j.cnki.jvs

[10] Xu, F., Tse, P. W. T, Fang, Y., and Liang, J. A fault diagnosis method combined with compound multiscale permutation entropy and particle swarm optimization—support vector machine for roller bearings diagnosis, *Proceedings of the Institution of Mechanical Engineers*, Part J: Journal of Engineering Tribology, **233**(4), (2019). https://doi.org/10.1177/1350650118788929

[11] Bandt, C., and Pompe, B. Permutation entropy:

Figure 14. Fault classification accuracy for different models.

A natural complexity measure for time series, *Physical Review Letters*, **17**, 21–24, (2022). https://doi.org/10.1103/PhysRevLett.88.174102

[12] Yang, J. W., Choudhary, G. I., Rahardja, S., et al. Classification of interbeat interval time-series using Attention Entropy, *IEEE Transactions on Affective Computing*, (2020), https://doi.org/10.1109/TAFFC.2020.3031004

[13] Pham, T. D. Time-Shift multi-scale entropy analysis of physiological signals, *Entropy*, **19**(257), (2017). https://doi.org/Time-Shift multi-scale entropy analysis of physiological signals

[14] Huang, G. An Insight into Extreme Learning Machines: Random Neurons, Random Features and Kernels, *Cognitive computation*, **6**(3), (2014). https://doi.org/10.1007/s12559-014-9255-2

[15] Liu, X., Gao, C., and Li, P. A comparative analysis of support vector machines and extreme learning machines, *Neural Networks*, **33**, (2012). https://doi.org/10.1016/j.neunet.2012.04.002

[16] Tu, Y., Wang, J., and Zhu, Q. Application of PSO-KELM model to predict hydrological time series, *China Rural Water Conservancy and Hydropower*, **07**, 21–24, (2018).

[17] Dun, P., Liu, C., and Wang, F. Rolling bearing fault diagnosis based on sparse autoencoder and FA-KELM, *Noise and Vibration Control*, **38**, 678–682, (2018).

[18] Yang, X., Guan, W., Liu, Y. et al. Wind power interval prediction method based on particle swarm optimization with kernel limit learning machine model, *Chinese Journal of Electrical Engineering*, **35**, 146–153, (2015). https://doi.org/10.13334/j.0258-8013.pcsee.2015.S.020

[19] Xue, J. and Shen, B. A novel swarm intelligence optimization approach: sparrow search algorithm, *Systems Science & Control Engineering*, **8**, (2020). https://doi.org/10.1080/21642583.2019.1708830

[20] Huang, C., Sun, Z., Zhang, B. et al. Research on neural network recommendation algorithm based on optimization of sparrow algorithm, textitIntelligent Computer and Application, **13**, 95–102, (2023).

[21] Costa, M., Goldberger, A. L., and Peng, C. K . Multiscale entropy analysis of complex physiologic time series, *Physical Review Letters*, **89**, 068102, (2002). https://doi.org/10.1103/PhysRevLett.89.068102

[22] Hu, Q., He, Z., Zhang, Z. et al. Intelligent diagnosis of early faults based on boosted wavelet packet transform and integrated support vector machine, *Journal of Mechanical Engineering*, **08**, 16–22, (2006).

[23] Sisi, Y. Study on fault diagnosis method of Power Transformer based on hybrid feature selection and PSO-KELM, *Hunan University*, (2020). https://doi.org/10.27135/d.cnki.ghudu

[24] Wang, H., Wang, Y., and Ji, Z. Short-term wind power prediction based on SAIGM-KELM, *Power System Protection and Control*, **48**, 78–87, (2020). https://doi.org/10.3390/su141710779

[25] Li, K., Pin, G. X., Wang, H. et al. Sequential fuzzy diagnosis method for motor roller bearing in variable operating conditions based on vibration analysis, *Sensors*, **13**, 8013–8041, (2013). https://doi.org/10.3390/s130608013