Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 11, No. 3, 2025

Acoustic Emission Signal Diagnosis of Rolling Bearing Faults Based on CEEMDAN and MESSA-SVM

Zhang Xu *, Yu Yang

Shenyang University of Technology, Shenyang Liaoning, 110870, China
* **Corresponding author**: Zhang Xu (Email: 15524501591@163.com)

Abstract: Acoustic emission signals generated during rolling bearing failures reflect their health conditions, serving as crucial indicators for to address the limitations in feature extraction and classification accuracy of traditional methods in acoustic emission signal to address the limitations in feature extraction and classification accuracy of traditional methods in acoustic emission signal, this paper proposes a fault diagnosis method based on CEEMDAN and MESSA-SVM. Initially, CEEMDAN is employed to perform multi-scale Initially, CEEMDAN is employed to perform multi-scale decomposition of fault acoustic emission signals, obtaining Intrinsic Mode Function (IMF) components, followed by selecting critical IMF components The IMF components are selected through kurtosis-correlation feature screening. Permutation entropy is then calculated to construct feature vectors. To optimize SVM classifier parameters, a Multi-strategy Sparrow Search Algorithm (MESSA) incorporating Euclidean distance population diversity, Gaussian perturbation strategy, and fitness difference guidance is designed. This algorithm demonstrates excellent global optimization capability in standard test functions. This algorithm demonstrates excellent global optimization capability in standard test functions. Experimental results indicate that the MESSA-SVM classifier achieves 95% accuracy. Through comparative analysis with other algorithms, the advantages of this method over alternatives include the advantages of this method over alternatives are validated, demonstrating its effectiveness and reliability in bearing fault acoustic emission signal diagnosis.

Keywords: Bearing Fault Diagnosis; Acoustic Emission; CEEMDAN; Permutation Entropy; Sparrow Search Algorithm; Support Vector Machin.

1. Introduction

As an important rotating component in mechanical equipment, rolling bearings have a direct impact on the performance and service life of the equipment. However, due to the complex long-term working environment, bearings are prone to wear, fatigue, cracks and other faults, which may lead to equipment downtime or even major safety accidents. Therefore, the real-time monitoring and fault diagnosis of rolling bearing operation state has important theoretical significance and engineering application value[1] .

Acoustic Emission (AE), as a highly sensitive non-destructive testing method, can capture high-frequency elastic wave signals at the early stage of microcrack formation or expansion within the material, providing an important means for early detection of bearing failures. Compared with vibration signals, acoustic emission signals are characterized by high frequency range and rapid response, which can reflect the health status of bearings earlier and more accurately. However, due to the complexity of the acoustic emission signal and its susceptibility to environmental noise interference, how to extract effective features and accurately classify fault types is still a difficult and hot spot in current research.

In recent years, bearing fault diagnosis methods combining advanced signal processing techniques and machine learning algorithms have gradually received widespread attention[2] . Techniques such as time-frequency analysis, wavelet transform, and empirical modal decomposition (EMD) are able to extract rich time-frequency features from acoustic emission signals[3] ; at the same time, the introduction of algorithms such as support vector machines (SVMs) and deep learning provides stronger model expression capabilities for classification and identification of fault patterns[4] . However, these methods still face problems such as high computational complexity and insufficient model robustness in practical applications.

Adaptive White Noise Complete Empirical Mode Decomposition (CEEMDAN) is an improved signal processing method proposed by scholars Maria Alejandra Torres et al. based on the traditional Ensemble Empirical Mode Decomposition (EEMD) algorithm[5] . CEEMDAN effectively solves several problems in traditional EEMD, such as mode aliasing, incompleteness of decomposition, and large signal reconstruction error, by adaptively adjusting the noise component and reconstructing the signal during the introduction of white noise.

Rolling bearing fault diagnosis essentially belongs to the category of pattern recognition, in which support vector machine (SVM) is widely used as a classification technique [6]. In order to improve the performance of the classification model, it is usually necessary to tune the parameters with the help of optimization algorithms to obtain the best parameter configuration. The Support Vector Machine (SSA-SVM) classifier optimized with the Improved Sparrow Search Algorithm (ISSA) is able to effectively search for the global optimal solution [7] with high classification accuracy and good recognition ability. The sparrow search algorithm performs global optimal search by simulating the sparrow's foraging behavior and danger perception mechanism. However, the traditional SSA suffers from the problem of easily falling into local optimality.

This paper takes the faulty bearing acoustic emission signal as the research object, and centers on the rolling bearing fault detection and diagnosis problem, and proposes a rolling bearing acoustic emission fault diagnosis method based on the

CEEMDAN decomposition algorithm, feature extraction algorithm, and the improved MESSA-SVM classifier model. First, the collected acoustic emission signals are decomposed using the CEEMDAN algorithm to obtain a series of intrinsic modal functions (IMFs); subsequently, the key IMF components are screened out according to the magnitude of the crag. Then, the time-domain alignment entropy features[7] are extracted from these IMF components to construct the feature vector set. Finally, the feature vector collections under different fault states are inputted into the improved MESSA-SVM model for training and classification to realize intelligent fault diagnosis of rolling bearing fault acoustic emission signals[8] .

Through experimental verification, the method proposed in this paper can significantly improve the accuracy and efficiency of fault detection, providing theoretical support and practical basis for the popularization and application of acoustic emission technology in actual engineering.

2. Acoustic Emission Signal Processing

2.1. CEEMDAN Principles

The CEEMDAN algorithm introduces an adaptive white noise-assisted decomposition process, and this strategy can significantly reduce the modal aliasing in EMD processing. Meanwhile, CEEMDAN also has smaller reconstruction error and higher decomposition efficiency[10] . Its specific algorithm steps are as follows:

Define the cavitation noise time series containing white noise as$x(t)$,$v_i(t)$ is the white noise sequence,γ_0 is the standard deviation of the white noise sequence, and$i = 1,2,\ldots N$ is the number of trials. The signal after adding white noise is:

$$X_i(t) = x(t) + \gamma_0 v_i(t) \tag{1}$$

The signal$X_i(t)$ after the addition of white noise is decomposed with EMD to obtain a set of intrinsic modal functions (IMFs). These modal components are averaged to obtain the first IMF of the CEEMDAN decomposition:

$$IMF_1(t) = \frac{1}{N}\sum_N^{i=1} I\,MF_t^1(t) \tag{2}$$

Calculate the first residual component:

$$r_1(t) = x(t) - IMF_1(t) \tag{3}$$

White noise$v_i(t)$ is added to the first residual$r_1(t)$ fori trials. On each trial, decompose the signal$r_1(t) + \gamma_1 v_i(t)$ until it is decomposed into first order components. The second modal componentIMF_2 is obtained:

$$IMF_2(t) = \frac{1}{N}\sum_N^{i=1} E_1\{r_1(t) + \gamma_1 E_1[v_i(t)]\} \tag{4}$$

where$E_j[\cdot]$ denotes thej th order component generated by the EMD method.

Calculate the second residual component:

$$r_2(t) = r_1(t) - IMF_2(t) \tag{5}$$

The above steps are repeated until the residual signal obtained is a monotonic function and the decomposition cannot be continued. At this point,k modal components are obtained and the final residual component is$R(t)$. The signal$X(t)$ can be expressed as:

$$x(t) = \sum_k^{k=1} I\,MF_k + R(t) \tag{6}$$

2.2. Cliff Criterion

Cliff is a statistical index used to describe the steepness of the probability distribution. For bearing acoustic emission signal, when its normal operation signal probability distribution is approximately normal distribution, the periodic impact is not significant; and when the bearing failure, due to the rolling body and the fault point of the periodic collision effect, the signal probability distribution tends to be steeper, the crag value increases significantly. The mathematical expression for this statistic of crag is:

$$K = \frac{E(x-u)^4}{\sigma^4} \tag{7}$$

Where$x$ is the acoustic emission signal,u is the mean value of the signal, andσ is the standard deviation of the signal.

However, the bearing acoustic emission signals contain fault information in addition to the superimposed background noise generated by the operation of the mechanical system and vibration interference from other components. In this study, a normalization method is used to process the IMF component clarity values to attenuate the influence effect of the original signal clarity:

$$ku = \frac{K(IMF_i)}{K(x)} \tag{8}$$

2.3. Pearson's Coefficient

The Pearson correlation coefficient is a statistic that measures the degree of linear correlation between two variables[11] . For bearings containing localized faults, their acoustic emission signals are mainly composed of fault shocks, friction vibrations and background noise components. The stronger the correlation between the IMF component obtained by CEEMDAN decomposition and the original signal, the more likely the component carries valid fault information. The mathematical definition of Pearson's correlation coefficient is:

$$r = \frac{Cov(X,Y)}{\sigma(X)\cdot\sigma(Y)} \tag{9}$$

wherer is the Pearson correlation coefficient,$Cov(X,Y)$ is the covariance of the random variables$X$ and$Y$,$\sigma(X)$ is the standard deviation of the random variable$X$, and$\sigma(Y)$ is the standard deviation of the random variable$Y$.

The Pearson correlation coefficient is used to measure the degree of correlation between the IMF component and the original signal, and is calculated as:

$$cc = r(IMFi, x) \tag{10}$$

Where, IMFi is the firsti modal component and$x$ is the original signal.

When a localized fault occurs in a bearing, the resulting shock signal manifests itself in multiple IMF components. These IMF components containing fault information have a strong correlation with the original signal because they retain the main features of the original signal. On the contrary, those IMF components that are mainly composed of noise have a weak correlation with the original signal due to their high randomness. Based on this feature, by calculating the correlation coefficient between the IMF components and the original signal, the components containing fault information can be effectively identified and filtered out.

2.4. Cliff-Correlation Feature-Based Weight Component Screening Approach

Rolling bearing acoustic emission signal has non-stationarity and non-linear characteristics, and its fault information is mainly manifested as transient impact component. Due to the influence of complex working conditions, diverse propagation paths and other factors, it is difficult for a single evaluation index to fully reflect the fault characteristic information contained in the IMF component. Although the crag of transient characteristics can capture the short-term shock characteristics, it is easily affected by noise and may amplify irrelevant signals; while the correlation coefficient of overall characteristics, although it can reflect the global trend of the signals, may be difficult to respond sensitively under weak fault shocks. To overcome the limitations of a single criterion, this paper combines the cliff value and the correlation coefficient to construct a combined evaluation index-multidimensional fault characterization factor P. The multidimensional fault characterization factor P (hereafter referred to as the fault factor) is defined as:

$$P_i = ku_i + cc_i \tag{11}$$

whereku is the normalized cliff value andcc is the Pearson correlation coefficient.

By this method, the component that best reflects the signal fault characteristics can be found among multiple modal components, avoiding the blindness of a single threshold or a fixed criterion and ensuring the dynamic adaptability of the analysis process.

2.5. Permutation Entropy

Permutation Entropy (PE) is a measure of time series complexity proposed by Porto et al. in 2000. Permutation entropy evaluates the complexity and disorder of information in time series data by analyzing the ordering of neighboring points[12] . Different from the traditional entropy theory, the alignment entropy does not depend on the specific value of the time series, but is based on the sorting pattern of the data, so it has strong robustness and can effectively deal with the interference of noise. Based on this, the alignment entropy can effectively measure the complexity and regularity of the acoustic emission fault signal sequence. The value of the alignment entropy is calculated by the following formula:

$$PE(m) = -\sum_{i=1}^{k} p_i \log p_i \tag{12}$$

wherep_i is the probability of occurrence of the first$i$ sorting pattern and$k$ is the total number of sorting patterns.

3. Parameter Optimization based on Improved Sparrow Search Algorithm

3.1. Principle of Sparrow Algorithm

Sparrow Search Algorithm (SSA) is a novel meta-heuristic optimization algorithm proposed in 2020, which is inspired by the foraging behavior and anti-predator strategy of sparrows. The algorithm has attracted attention for its uniqueness, powerful optimization capability, and fast convergence speed, and its basic principle is as follows:

Sparrow populations are mainly composed of three types of individuals: discoverers, followers and scouts. Discoverers are responsible for leading the population to find suitable foraging areas and directions; followers obtain food based on the information provided by the discoverers; in order to maximize the foraging efficiency, the roles of discoverers and followers are dynamically changed in the algorithm[13] .

3.2. MESSA (Multi-Strategy Enhanced Sparrow Search Algorithm)

Sparrow search algorithms perform global optimality search by simulating the foraging behavior and danger perception mechanism of sparrows. However, the traditional SSA suffers from the problem of easy to fall into local optimization, especially in the case of insufficient population diversity, the global search ability of the algorithm may be significantly reduced[14] . In order to improve this situation, this paper utilizes various enhancement strategies to improve the sparrow algorithm.

3.2.1. Population Diversity Optimization Initialization Based on Euclidean Distance

The optimization performance of the Sparrow Search Algorithm (SSA) relies on the diversity of the populations, especially in the distribution stage of the initial populations. For this reason, this study proposes a population diversity retention strategy based on Euclidean distance for SSA to enhance its global search capability and avoid falling into premature convergence.

In the original SSA, the population individuals are randomly initialized, which may lead to too much concentration of individuals in the solution space and reduce the exploration efficiency of the population. To solve this problem, this study introduces a population screening mechanism based on Euclidean distance, which calculates the two-by-two distances of individuals in the population and excludes the individuals whose distances are smaller than a set thresholdδ , so as to construct a uniformly distributed population in the solution space.

The Euclidean distance is calculated as:

$$d(X_i, X_j) = \sqrt{\sum_{n}^{k=1} (X_i^k - X_j^k)^2} \tag{13}$$

WhereX_i andX_j are two individuals in the sparrow population and$n$ is the problem dimension. When the distance between two individuals$d(X_i, X_j) < \delta$, one of them is removed to ensure the population diversity. Through this screening method, the solution space coverage ability of SSA in the initialization stage can be significantly enhanced, laying a good foundation for the subsequent optimization process.

3.2.2. Gaussian Perturbation Strategy (Improvement of SSA Discoverer Stage Update Mechanism)

In Sparrow Search Algorithm (SSA), the individual discoverer is mainly responsible for global exploration[15] . In order to improve the global search capability and avoid the algorithm from falling into local optimum, this study proposes a Gaussian perturbation strategy. This strategy enhances the diversity of search behaviors when updating discoverer individuals by introducing perturbations based on Gaussian distributions, and optimizes the balance between exploration and exploitation at different stages by adaptively adjusting the magnitude of perturbations.

The probability density function of a Gaussian distribution (also called normal distribution) is.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{14}$$

whereμ is the mean andσ is the standard deviation. This distribution is widely used in nature.

While the traditional updating approach in the discoverer phase relies only on the local information of the current solution, the Gaussian perturbation strategy breaks the existing local optimum by introducing random noise to simulate more diverse search behaviors and enhance the global exploration capability. The Gaussian perturbation amplitude is dynamically adjusted with iterations, using a larger perturbation amplitude at the beginning of the search for extensive search, and a smaller perturbation amplitude at the later stages of the algorithm to refine the search and enhance the local search capability.

In the improved Gaussian perturbation strategy, a Gaussian distributed perturbation term is introduced to replace the individual discoverer's position update method:

$$X_i^{t+1} = X_i^t \cdot e^{-\frac{t}{T_{\max}}} + \alpha \cdot N(0,1) + \beta_0 \left(1 - \frac{t}{T_{\max}}\right) \cdot G(0,1) \tag{15}$$

$N(0,\sigma)$denotes a standard Gaussian distribution (mean 0, standard deviationσ of the perturbation.

X_1 is the current global optimal solution andX_{end} is the current worst solution.

In the early stages of the algorithm (when the number of iterations is low), the perturbation magnitude is large, allowing the individual discoverer to explore the solution space extensively.

As the number of iterations increases (i.e., the search enters a later stage), the magnitude of the perturbation decreases, allowing the search to converge to a locally optimal solution and enhancing the fine development capability.

Specifically, the perturbation magnitudeσ is adjusted with the number of iterations as follows:

$$\sigma_{\text{iter}} = \sigma_{\max} \cdot \left(1 - \frac{\text{iter}}{\text{Max iter}}\right) \tag{16}$$

where$\sigma_{\max}$ is the initial perturbation amplitude,Max_iter is the maximum number of iterations, and σ_{iter} is the perturbation amplitude of the current iteration.

3.2.3. Adaptation Difference Bootstrap Strategy Based on Global Optimum

The main goal of the joiner phase is to exploit potential search regions. However, the updating strategy of traditional sparrow algorithm joiner relies more on local information (e.g., the location of the optimal solution), and the joiner updating strategy does not make enough use of the global optimal solution, which leads to the limitation of its exploitation ability. In this study, we propose an adaptation difference guidance strategy based on the global optimum to enhance the convergence speed and exploitation capability of the algorithm by strengthening the influence of the global optimal solution.

When the difference between an individual's fitness and the global optimal solution is large, the joiner will move more substantially in the direction of the global optimal solution, thus accelerating the search process. In the early stages of the algorithm (when the number of iterations is small), the larger difference in fitness guides the individuals to explore the solution space extensively, which enhances the global search capability.

And as the number of iterations increases (i.e., entering the later stages of search), the fitness difference gradually decreases and the step size shrinks accordingly, which makes the individual search gradually converge to the global optimal solution and improves the fineness of local exploitation. This strategy effectively balances the ability of global exploration and local exploitation, which helps the algorithm accelerate convergence near the global optimal solution and avoid falling into the local optimal solution.

The formula for the difference in fitness is as follows:

$$\Delta f_j = | f_j - f_{\text{GBest}} | \tag{17}$$

where the fitness of the individual j isf_j and the fitness of the global optimal solution isf_{GBest} .

Combine the fitness difference with the global optimal solution to guide the search step size of an individual. Assume that the step size ΔX_j is determined by the positional difference between the fitness difference and the global optimal solution with the following formula:

$$\Delta x_j = \Delta f_j \cdot \gamma \cdot \varepsilon \cdot \left(1 - \frac{t}{T_{\max}}\right) \tag{18}$$

whereγ is the step adjustment factor and$\varepsilon \sim U(0,1)$ is a uniformly distributed random number.

Based on the step size guided by the fitness difference, the update formula for the individualj is.

$$X_{\text{new}}(j) = X(j) + \Delta x_j \cdot \left(X_{\text{GBest}} - X(j)\right) \tag{19}$$

Where: $X(j)$ is the current solution location of the individual j ,X_{GBest} is the location of the global optimal solution, andΔX_j is the bootstrap step size.

3.3. Algorithm Testing and Analysis

In order to test the improved effectiveness of MESSA (Enhanced Multi-Strategy Sparrow Search Algorithm), which incorporates population diversity optimization initialization, Gaussian perturbation strategy, and updating strategy guided by fitness difference, this paper compares it with the Sparrow Search Algorithm (SSA) and several other search algorithms, namely: the Sparrow Algorithm SSA; the Gray Wolf Optimization Algorithm GWO; and the Particle Swarm Algorithm PSO.

For a comprehensive performance evaluation, 14 benchmark test functions covering different features were selected for the study. These sets of evaluation functions contain single-peak, multi-peak, and dimension-specific multimodal functions. Among them, the multi-peak function can effectively verify the global search capability of the algorithm and its performance in jumping out of the local optimum due to the inclusion of a number of local optimal solutions, while the single-peak function is mainly used to evaluate the convergence efficiency and local optimality search performance of the algorithm. The detailed function parameters and characteristics are shown in Table 1.

The experimental platform is configured as follows: CPU model: HP HP EliteBook 840 G9, Intel Core i7-1255U main frequency 2.7GHz System memory: 16GB System platform: Windows 11 64-bit version Development tools: Matlab

R2018(a)

In terms of parameter configuration, MESSA adopts the same basic parameter settings as SSA, and all the algorithms involved in the comparison are tested under uniform experimental conditions. In order to ensure the statistical reliability of the experimental results and reduce the impact of errors caused by random factors, each algorithm was run independently for 30 times. The experimental results are summarized in Table 2, where the optimal results are marked in bold font

Table 1. Test Functions

serial number	function (math.)	Single/Multiple Peak	dimension (math.)	domain (math.)	theoretical optimum
F1	Sphere	Single peak	30	[-100, 100]	0
F2	Schwefel'problem2.22	Single peak	30	[-10, 10]	0
F3	Schwefel'problem 1.2	Single peak	30	[-100, 100]	0
F4	Schwefel' problem2.21	Single peak	30	[-100, 100]	0
F5	Generalized Roscnbrocks' Funtion	Single Peak	30	[-30, 30]	0
F6	Step Function	Single Peak	30	[-100, 100]	0
F7	Quartic Function	Single peak	30	[-1.28, 1.28]	0
F8	Generalized Schwefel' problem 2.2	Multiple peaks	30	[-500, 500]	-12569.5
F9	Generalized Rastrigin's Function	Multi-peak	30	[-5.12, 5.12]	0
F10	Ackley's Function	Multi-peak	30	[-32, 32]	0
F11	Ceneralized Criewank Function	Multi-peak	30	[-600, 600]	0
F12	Ceneralized Penalized Function1	Multi-peak	30	[-50, 50]	0
F13	Ceneralized Penalized Function2	Multi-peak	30	[-50, 50]	0
F14	Shekel 's Foxholes Function	Fixed-dimensional multimodal	30	[-65.536, 65.536]	1

Table 2. Functional simulation test results

function (math.)	arithmetic	optimum value	average value	(statistics) standard deviation	function (math.)	arithmetic	optimum value	average value	(statistics) standard deviation
F1	SSA	3.67×10^{-10}	2.03×10^{-7}	3.28×10^{-7}	**F8**	SSA	-8717.03	-7704.774	7.62×10^{2}
	MESSA	$\mathbf{0.00 \times 10^{0}}$	$\mathbf{0.00 \times 10^{0}}$	$\mathbf{0.00 \times 10^{0}}$		MESSA	**-12569.49**	**-12569.48**	$\mathbf{5.46 \times 10^{-3}}$
	GWO	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}		GWO	-418.982	-418.983	1.94×10^{-9}
	PSO	3.70×10^{-18}	3.03×10^{-17}	3.61×10^{-17}		PSO	-10238.66	-9355.68	4.34×10^{2}
F2	SSA	5.27×10^{-5}	6.22×10^{-4}	7.91×10^{-4}	**F9**	SSA	1.22×10^{-9}	5.37×10^{-7}	7.51×10^{-7}
	MESSA	$\mathbf{0.00 \times 10^{0}}$	$\mathbf{0.00 \times 10^{0}}$	$\mathbf{0.00 \times 10^{0}}$		MESSA	$\mathbf{0.00 \times 10^{0}}$	$\mathbf{0.00 \times 10^{0}}$	$\mathbf{0.00 \times 10^{0}}$
	GWO	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}		GWO	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}
	PSO	4.51×10^{-11}	1.84×10^{-10}	1.20×10^{-10}		PSO	1.59×10^{1}	2.11×10^{1}	4.02×10^{0}
F3	SSA	3.77×10^{-9}	1.44×10^{-6}	2.81×10^{-6}	**F10**	SSA	8.37×10^{-6}	1.15×10^{-4}	1.18×10^{-4}
	MESSA	$\mathbf{0.00 \times 10^{0}}$	$\mathbf{0.00 \times 10^{0}}$	$\mathbf{0.00 \times 10^{0}}$		MESSA	$\mathbf{8.88 \times 10^{-16}}$	$\mathbf{8.88 \times 10^{-16}}$	$\mathbf{0.00 \times 10^{0}}$
	GWO	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}		GWO	9.01×10^{-16}	9.01×10^{-16}	0.00×10^{0}
	PSO	8.76×10^{-2}	1.27×10^{0}	1.91×10^{0}		PSO	2.28×10^{-10}	9.41×10^{-10}	5.95×10^{-10}
F4	SSA	5.54×10^{-6}	5.81×10^{-5}	5.20×10^{-5}	**F11**	SSA	8.79×10^{-12}	8.37×10^{-9}	9.87×10^{-9}
	MESSA	$\mathbf{0.00 \times 10^{0}}$	$\mathbf{0.00 \times 10^{0}}$	$\mathbf{0.00 \times 10^{0}}$		MESSA	$\mathbf{0.00 \times 10^{0}}$	$\mathbf{0.00 \times 10^{0}}$	$\mathbf{0.00 \times 10^{0}}$
	GWO	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}		GWO	0.00×10^{0}	0.00×10^{0}	0.00×10^{0}
	PSO	8.57×10^{-4}	3.11×10^{-3}	1.67×10^{-3}		PSO	0.00×10^{0}	1.43×10^{-2}	1.01×10^{-2}
F5	SSA	2.86×10^{-1}	2.88×10^{-1}	5.71×10^{-2}	**F12**	SSA	6.52×10^{-2}	1.27×10^{-1}	3.82×10^{-2}
	MESSA	3.09×10^{-12}	1.27×10^{-8}	2.77×10^{-8}		MESSA	$\mathbf{1.45 \times 10^{-15}}$	3.94×10^{-12}	4.34×10^{-12}
	GWO	$\mathbf{0.00 \times 10^{0}}$	$\mathbf{0.00 \times 10^{0}}$	$\mathbf{0.00 \times 10^{0}}$		GWO	1.85×10^{-15}	6.28×10^{-13}	9.08×10^{-13}
	PSO	1.69×10^{1}	2.99×10^{1}	2.01×10^{1}		PSO	4.22×10^{-20}	1.32×10^{-18}	1.89×10^{-18}
F6	SSA	3.39×10^{0}	3.72×10^{0}	4.72×10^{-1}	**F13**	SSA	7.93×10^{-1}	1.03×10^{0}	1.46×10^{-1}
	MESSA	$\mathbf{3.74 \times 10^{-14}}$	1.46×10^{-11}	2.06×10^{-11}		MESSA	$\mathbf{7.60 \times 10^{-13}}$	$\mathbf{1.07 \times 10^{-11}}$	$\mathbf{1.23 \times 10^{-11}}$
	GWO	1.31×10^{-13}	5.73×10^{-13}	6.25×10^{-13}		GWO	3.77×10^{-12}	4.05×10^{-11}	3.24×10^{-11}
	PSO	9.52×10^{-6}	1.04×10^{-5}	7.29×10^{-7}		PSO	4.35×10^{-7}	9.21×10^{-6}	1.35×10^{-5}
F7	SSA	1.63×10^{-5}	1.83×10^{-4}	1.72×10^{-4}	**F14**	SSA	9.98×10^{-1}	9.98×10^{-1}	7.21×10^{-11}
	MESSA	$\mathbf{8.50 \times 10^{-7}}$	1.19×10^{-5}	1.16×10^{-5}		MESSA	$\mathbf{9.98 \times 10^{-1}}$	$\mathbf{9.98 \times 10^{-1}}$	$\mathbf{3.20 \times 10^{-11}}$
	GWO	1.31×10^{-6}	2.33×10^{-6}	2.05×10^{-6}		GWO	9.98×10^{-1}	9.98×10^{-1}	2.35×10^{-10}
	PSO	1.42×10^{-3}	2.58×10^{-3}	7.67×10^{-4}		PSO	9.98×10^{-1}	9.98×10^{-1}	5.21×10^{-11}

In order to visually assess the optimization performance of the MESSA algorithm, the convergence trajectories of four typical test functions are selected in Figure 2 for comparative analysis. By processing the logarithmic transformation of the function fitness values, the convergence behavior and dynamic characteristics of each algorithm in the search process are reflected more clearly.

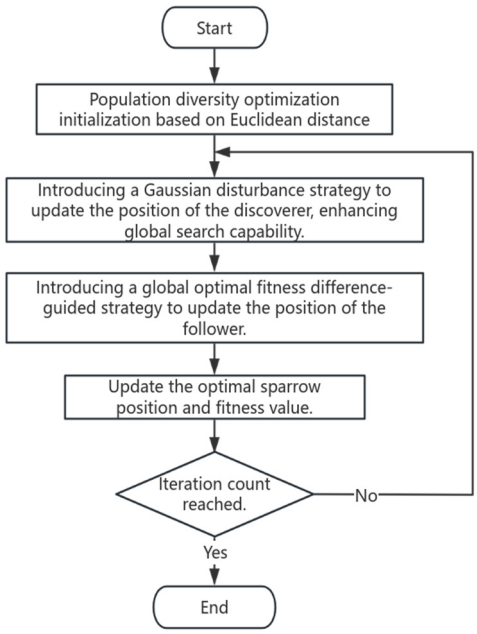
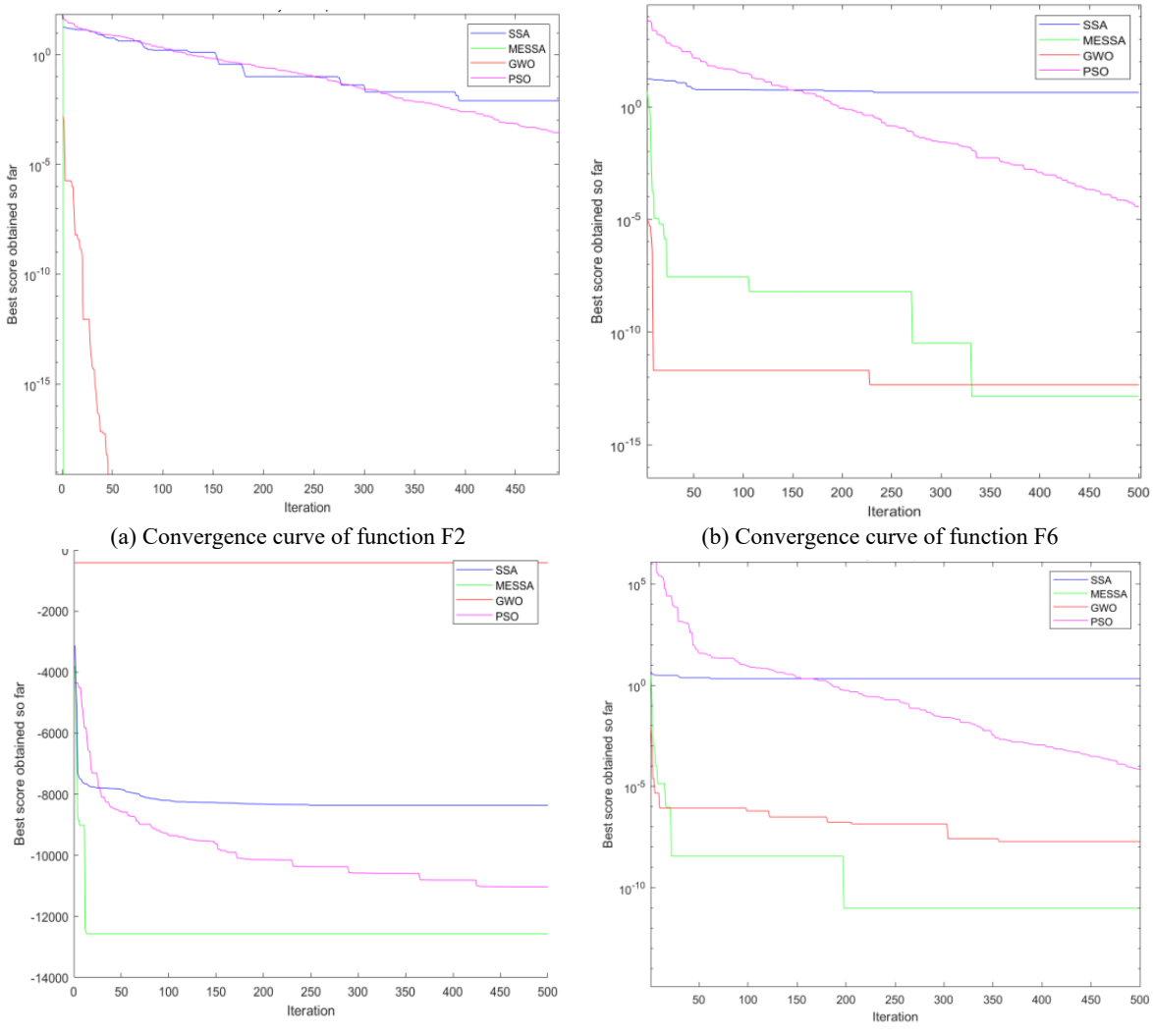


Figure 1. MESSA algorithm flow



(a) Convergence curve of function F2

(b) Convergence curve of function F6

(c) Convergence curve of function F8

(d) Convergence curve of function F13

Figure 2. Convergence plot of test function

By comparing the experimental results, it can be found that the MESSA algorithm achieves exact convergence in all the tests of the single-peak functions F1-F4, which is mainly attributed to the Gaussian perturbation mechanism introduced in the discoverer updating strategy, the Gaussian perturbation provides random fluctuations conforming to the normal distribution for the individual searches, which is more conducive to the careful localized exploration.

In the complex multi-peak function F5-F7 tests, the advantages of the improved algorithm are more obvious. Taking F6 as an example, MESSA obtains the optimal solution of 3.74×10^{-14} . This significant performance improvement mainly stems from the innovative design of the joiner updating strategy: a dynamic step-size adjustment mechanism based on the fitness difference, which enables the individual to adaptively adjust the searching step size according to the gap with the global optimal solution, and it performs well in a complex function space like F6.

Especially on the high-dimensional function F8, MESSA reaches the global optimum value of -12569.49 and the standard deviation is only 5.46 × 10^-3, this result is significantly better than the other compared algorithms. Meanwhile the lower standard deviation proves that the algorithm has strong stability, thanks to the diversity preserving initialization method based on Euclidean distance which effectively maintains the dispersion of the population in the high-dimensional space and avoids premature convergence caused by premature aggregation of the population.

Meanwhile, the advantage of MESSA is more obvious in functions with complex terrain features such as F10-F13. Especially in the F12 function, the optimal value obtained by MESSA is improved by 12 orders of magnitude compared to the original SSA algorithm. This significant improvement can be attributed to the optimization of the joiner updating strategy in the improved algorithm: by combining the global optimal solution information with the fitness difference, the convergence speed is guaranteed while maintaining a strong exploitation capability.

However, it is worth noting that MESSA does not perform as well as GWO on F5, mainly due to the fact that MESSA's updating strategy relies heavily on the difference from the current optimal solution to guide the search. This mechanism works well on most functions as it maintains good convergence. However, on multi-peak complex functions such as F5, over-reliance on the current optimal solution rather limits the algorithm's ability to cross local peaks. GWO guides the population search through α, β, δ three leaders jointly[16] , and this multi-leader mechanism has a natural advantage when dealing with highly nonlinear multi-peak functions such as F5. To summarize, MESSA has high stability and significant performance in searching optimization, which indicates that MESSA has strong competitiveness compared with the comparison algorithms, but there are still some shortcomings.

4. CEEMDAN and MESSA-SVM based Fault Diagnostic Models

The fault diagnosis system in this study incorporates the CEEMDAN method for effective decomposition of faulty bearing acoustic emission signals, while utilizing arrangement entropy to construct a feature file. Finally, the support vector machine is optimized by enhanced multi-strategy sparrow search algorithm to construct an efficient classification model for fault diagnosis [17]. The specific steps of this fault diagnosis model are as follows.

Step 1: Acoustic Emission Experiment (AEE) to collect the input fault acoustic emission signal data to ensure that the signal is acquired under appropriate sampling conditions.

Step 2: The raw acoustic emission signals are decomposed using the CEEMDAN method to obtain the IMF components for different fault types.

Step 3: Calculate the importance of each component according to the weights of the craggy-correlation features, select the top 5 components with P-value, calculate the entropy eigenvalue of their alignment, and construct the feature file.

Step 4: Divide the feature vector dataset obtained in Step 3 into a training set and a test set, and use the training set in conjunction with the MESSA algorithm for parameter optimization and training of the SVM.

Step 5: The optimized MESSA-SVM model is tested for fault classification using the partitioned test set and the final fault classification results are output.

5. Experimental Validation

5.1. Data Selection

The physical diagram of the bearing acoustic emission detection experimental platform of this study is shown in Figure 3. The experiment uses NU205EM cylindrical roller bearings as the test object, configured with a two-channel acoustic emission signal acquisition program, the sensor arrangement location was the upper surface of the test bearing housing and the neighboring shaft system platform area. The data acquisition equipment is PCI-2 acoustic emission system (made in USA). The experimental design included three bearing conditions: cage damage, outer ring defects, and inner ring failures, and the tests were conducted at three rotational speeds: 600, 900, and 1200 r-min-¹. A total of three classes of labeled data were constructed for the experiments, each class containing 100 sets of samples, and the training and test sets were divided according to the ratio of 8:2, i.e., 80 sets of each class were selected for training, and the remaining 20 sets were used for test validation.

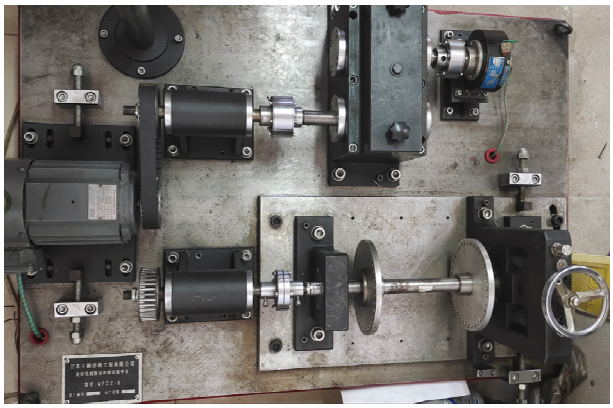


Figure 3. Acoustic emission monitoring test setup

5.2. Feature Extraction

First, taking the bearing acoustic emission signal 600r inner ring fault as an example, the acoustic emission data are decomposed by CEEMDAN to obtain multiple IMF

components. The first 5 components with the largest P-value are calculated and selected by using the weighted component screening method of cliff-correlation feature, as shown in Table 3. The top 5 IMF components are selected to contain the vast majority of the effective information of the original signal, but also effectively avoid the problems of time-consuming and long redundancy in diagnosis caused by too much component data[17] .

Table 3. IMF Component P-values

IMF component	IMF1	IMF2	IMF3	IMF4
P-value	8.81	5.17	0.33	0.17
IMF component	IMF5	IMF6	IMF7	IMF8
P-value	-0.60	0.13	-1.52	-1.51

The arrangement entropy is solved for the first five IMF components selected by using the P value, and then normalized to construct the feature vectors to obtain the feature vector set of100×5. The IMF1 arrangement entropy of 8 sets of data for each of the 3 faulty operating states of rolling bearings is illustrated as an example, see Fig. 4.

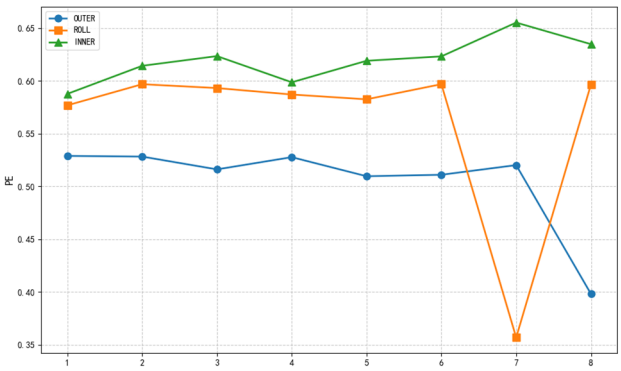


Figure 4. Entropy value of IMF1 arrangement in different states

5.3. SVM Recognition of Bearing Faults

During training, the outer ring bearing fault features are labeled as 1, rolling element fault features are labeled as 2, and inner ring fault features are labeled as 3. In order to highlight the advantages of the craggy-correlation feature-based weight component screening method proposed in this paper, three groups of comparative experiments are designed in this study, which are identified using ordinary SVM. The first group uses EEMD decomposition and directly selects the alignment entropy of the first 5 IMF components as features; the second group applies crag-correlation feature screening after EEMD decomposition

Selection method, by calculating the crag and the Pearson correlation coefficient of each IMF with the original signal, the five components with the largest P-value are screened out, and the arrangement entropy of these five sensitive feature components is calculated as the features. The third group applies the crag-correlation feature screening method after CEEMDAN decomposition to screen the arrangement entropy of the sensitive feature components as features. The features of the above three experiments are input into the SVM classifier, and the experimental results are shown in Table 4.

The experimental results show that the screening method based on crag-correlation features can effectively improve the accuracy of feature selection, especially in the case of higher signal complexity, and has better diagnostic performance.

Table 4. SVM classification accuracy for different feature extraction methods

Feature Extraction Methods	Outer ring failure	Rolling body failure	Inner ring failure	accuracy
EEMD	11	11	19	68.3%
EEMD vs. P-value	14	10	18	70.0%
CEEMDAN with P-value	20	17	15	86.7%

5.4. Comparison of the Effect of SVM with Different Classifiers

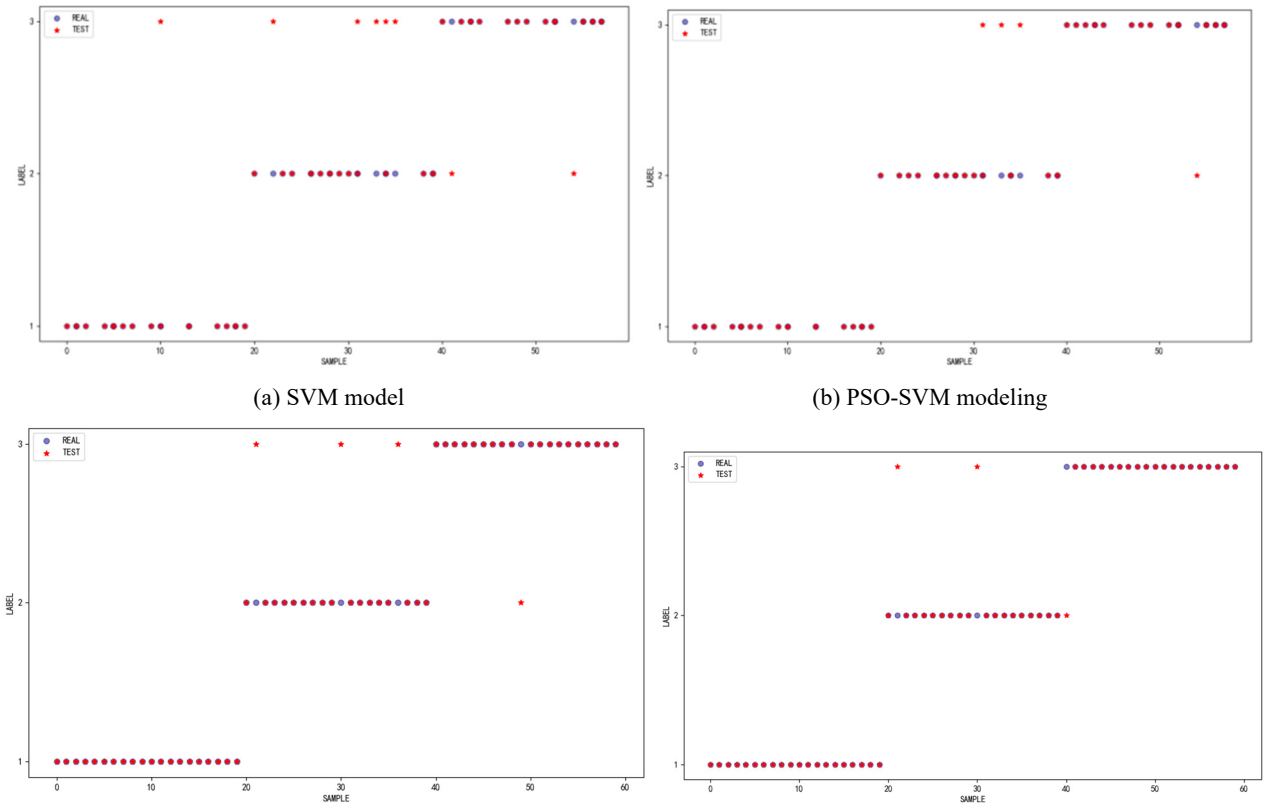
In order to further demonstrate the advantages of SVM classifier under the optimization of enhanced multi-strategy sparrow search algorithm over other excellent algorithms, the feature vectors are inputted into the following classification models for comparison: SVM, PSO-SVM, SSA-SVM, and MESSA-SVM, respectively. The number of populations of each classification model is uniformly set to 30, and the number of iterations is 50 times. The recognition result graph is shown in Fig. 5, SVM model has 8 groups of samples failed to recognize correctly, PSO-SVM and SSA-SVM models have 4 groups of samples recognized incorrectly respectively, which is significantly improved compared with traditional SVM. And the improved MESSA-SVM model has only 3 groups not recognized correctly among 60 groups of test objects, and the overall recognition accuracy reaches 95%, which has a significant advantage over other comparison algorithms. A comparison of the experimental results is shown in Table 3, which indicates that the MESSA-SVM model shows outstanding performance advantages in the bearing fault diagnosis task.

The analysis in Table 5 shows that the classification accuracy of ordinary SVM model is 86.7%, which is the worst classification effect, SSA - SVM model and PSO - SVM have improved the accuracy compared with SVM classifier, and the classification accuracy is 93.3%, and the classification accuracy of MESSA - SVM classifier is improved by 1.7% based on SSA - SVM, and the classification accuracy is 95% at the highest. In summary, the MESSA - SVM classifier shows more superior performance in bearing fault diagnosis.

6. Conclusion

In order to accurately identify rolling bearing faults based on acoustic emission signals, this paper proposes a weight component screening method based on craggy-correlation features and a feature extraction process of CEEMDAN arrangement entropy, as well as a rolling bearing acoustic emission fault diagnosis method of MESSA-SVM. Through theoretical analysis and experimental verification, the following main conclusions are drawn:

(1) A weight component screening method based on the crag-correlation feature is proposed to construct a multidimensional fault characterization factor P by calculating the crag value and Pearson's correlation coefficient of IMF components, and experiments have proved that the key components containing fault information are effectively identified and screened out, which improves the fault identification accuracy.



(a) SVM model

(b) PSO-SVM modeling

(c) SSA-SVM modeling

(d) MESSA-SVM modeling

Figure 5. Diagram of fault recognition results

Table 5. Comparison of results

classification model	Number of test samples	Accuracy/ (%)
SVM	60	86.7
PSO-SVM	60	93.3
SSA-SVM	60	93.3
MESSA-SVM	60	95

(2) Enhanced multi-strategy sparrow search algorithm (MESSA) is designed, which effectively improves the global search ability and convergence performance of the algorithm by introducing population diversity optimization initialization based on Euclidean distance, Gaussian perturbation strategy, and update strategy guided by fitness difference. Algorithm comparison experiments on 14 standard test functions show that MESSA has significant advantages.

(3) The alignment entropy feature vectors are input into four classifiers, SVM, PSO-SVM, SSA-SVM and MESSA-SVM, respectively, for fault recognition. The results show that the recognition accuracy of the MESSA-SVM classifier reaches 95%, which is 8.3 percentage points higher than that of the ordinary SVM and 1.7 percentage points higher than that of the PSO-SVM and SSA-SVM, which verifies the superiority of the MESSA algorithm in bearing fault diagnosis.

References

[1] YU Yang, HE Ming, LIU Bo, et al. Recognition of rolling bearing fault acoustic emission signal based on GA-IDBN[J]. Nondestructive Testing, 2020, 42(01):31-36.

[2] HUANG Hongchen, HAN Zhennan, ZHANG Qianqian, et al. Rolling bearing fault identification based on Laplace feature mapping[J]. Vibration and Shock 2015, 34(5):128-134+144.

[3] XIE X P, CHENG S J, LI X Y. Vibration diagnosis and optimization of industrial robot based on TPA and EMD methods [J]. Computer Modeling in Engineering & Sciences, 2023, 135(3)

[4] VAN H B, HE D, QU Y. On the use of spectral averaging of acoustic emission signals for bearing fault diagnostics[J]. 2014, 136 (6): 1 - 13.

[5] M. E. Torres, M. A. Colominas, et al. "A complete ensemble empirical mode decomposition with adaptive noise," 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 2011, pp. 4144-4147.

[6] Xiang Mingsheng, Feng Kun, Jia Shaohui, et al. Multi-fault diagnosis method of centrifugal pump based on EMD-AE preferred features under strong noise conditions[J]. Vibration and Shock, 2024, 43(23):66-74.

[7] WEI Wei, WANG Zhihai, LIU Xiaoqin, et al. Acoustic emission diagnosis of rolling bearing faults based on MDFF and ISSA[J]. Vibration and Shock, 2023, 42(07):65-76.

[8] HE Cheng, WU Tao, GU Runwei, et al. Rolling bearing fault diagnosis based on composite multiscale permutation entropy and reverse cognitive fruit fly optimization algorithm - Extreme learning machine[J]. Measurement, 2021, 173: 108636.

[9] YU Yang, LI Eddie, YANG Ping, et al. Feature extraction of rolling bearing fault acoustic emission signal with improved wavelet threshold function and ACEWT method[J]. Vibration and Shock, 2023, 42(17): 194-202.

[10] Lv Fengxia, Miao Yi, Bie Fengfeng, et al. Application of ICEEMDAN and GS-SVM algorithms in rolling bearing acoustic emission fault diagnosis[J]. Noise and Vibration Control, 2022, 42(06):92-97+129.

[11] Dang J., Luo Y., Tian Lulin, et al. Optimization-based VMD fusion of information entropy and FA-PNN for wind turbine

gearbox fault diagnosis[J]. Journal of Solar Energy, 2021, 42(01):198-204.

[12] DING Jiaxin, WANG Zhenya, YAO Lizang, et al. Generalized composite multiscale weighted arrangement entropy and parameter optimized support vector machine for rolling bearing fault diagnosis[J]. China Mechanical Engineering, 2021, 32(02):147-155.

[13] Wang Y, Zheng HZ, Huang X, et al. Sparrow search optimization algorithm based on multi-stage scheduling framework[J]. Journal of Electronics, 2024, 52(09):3086-3096.

[14] YANG X, ZHANG T, LI YM, et al. ISSA optimized SVM for rolling bearing fault diagnosis of electric motors[J]. Electronic Measurement Technology, 2023, 46(15):186-192.

[15] LI Xinyan, JIN Wuyin. Research on rolling bearing fault diagnosis based on optimized support vector machine with improved sparrow algorithm[J]. Vibration and Shock, 2023, 42(06): 106-114.

[16] HU Xuan, LI Chun, YE Kehua, et al. Application of improved gray wolf algorithm optimized support vector machine in wind turbine gearbox fault diagnosis[J]. Mechanical Strength, 2021, 43(06):1289-1296.

[17] YUAN Xianfeng, YAN Zichen, ZHOU Fengyu, et al. Rolling bearing fault diagnosis by SSAE and IGWO-SVM[J]. Vibration. Test and Diagnosis, 2020, 40(02):405-413+424.

[18] Li R, Fan YG. Fault diagnosis of high-pressure diaphragm pump check valve based on CEEMDAN multi-scale arrangement entropy and SO-RELM[J]. Vibration and Shock, 2023, 42(05):127-135.