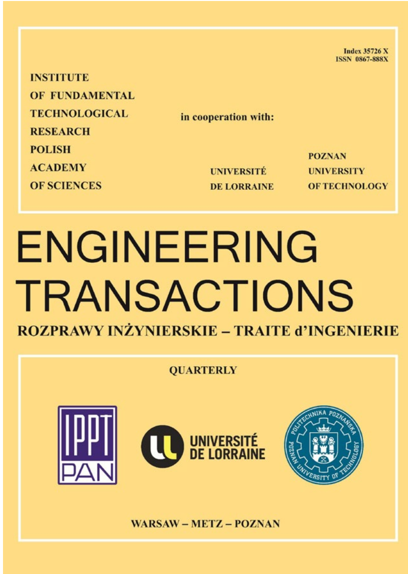
JOURNAL PRE-PROOF



Title: Research on an improved VMD-RCMDE multi-feature fusion diagnosis method for external pump faults

Author(s): Liang Ge, Nvqing Zhou, Nengji Jiang, Lie Li, Xiaojun Wang, Yi Teng

DOI: https://doi.org/10.24423/engtrans.2026.3655

Journal: *Engineering Transactions*

ISSN: 0867-888X, e-ISSN: 2450-8071

Publication status: In press

Received: 2025-09-08

Revised: 2026-07-16

Accepted: 2026-07-16

Published pre-proof: 2026-07-27

Please cite this article as:

Ge L., Zhou N., Jiang N., Li L., Wang X., Teng Y., Research on an improved VMD-RCMDE multi-feature fusion diagnosis method for external pump faults, *Engineering Transactions*, 2026, https://doi.org/10.24423/engtrans.2026.3655

Research on an improved VMD-RCMDE multi-feature fusion diagnosis method for external pump faults

Liang Ge[a,b], Nvqing Zhou[a], Nengji Jiang[c], Lie Li[c], Xiaojun Wang[c], Yi Teng[a]

[a] *College of Electromechanical Engineering, Southwest Petroleum University*

[b] *National Key Laboratory of Reservoir Geology and Development Engineering-Southwest Petroleum University*

[c] *Baikouquan Oil Production Plant, Xinjiang Oilfield Branch, PetroChina*

Email address of corresponding author: cgroad@swpu.edu.cn

ABSTRACT: This paper proposes a multi-feature fusion fault diagnosis method based on the global search whale optimization algorithm(GSWOA) to address the issues of single feature extraction, low detection efficiency, and insufficient accuracy in the actual operation of external pumps in oil and gas gathering and transmission. This method optimizes the variational mode decomposition(VMD) and the refined composite multi-scale dispersion entropy(RCMDE). First, GSWOA optimizes the VMD parameters to decompose the pressure signal and extract its intrinsic mode function(IMF) components during external pump operation. Second, to improve the accuracy and stability of feature extraction, RCMDE values are calculated for the pressure and temperature signals. Then, the extracted features are fused with the vibration signals to construct a multi-source feature fusion model. Finally, a classifier for fault identification and diagnosis of the constructed feature model is used. This classifier is a least squares support vector machine(LS-SVM) optimized by the snow geese algorithm(SGA). Experimental results demonstrate that the improved VMD-RCMDE multi-feature fusion diagnostic method achieves a 91% fault identification accuracy, which is at least 3.5% higher than other methods and significantly enhances the accuracy and robustness of external pump fault diagnosis. These results demonstrate the feasibility of efficiently and reliably diagnosing external pump faults and have significant theoretical and engineering application value.

KEYWORDS: External pump; Fault diagnosis; Variational mode decomposition (VMD); refined composite multi-scale dispersion entropy(RCMDE); Least squares support vector machine(LS-SVM) ; Snow geese algorithm(SGA)

1. Introduction

With the rapid development of the oil and gas industry, external pumps play a central role in crude oil gathering and transportation, directly impacting pipeline pressure stability and the safe operation of the entire system. Operating continuously under complex conditions, external pumps are prone to fluctuations in parameters such as pressure, temperature, and vibration. These variations can trigger issues such as abnormal liquid discharge and supply, or pressure stabilization failure, severely compromising production efficiency and equipment reliability. Therefore, real-time monitoring of external pump operational status and fault information, coupled with timely maintenance interventions, is crucial for ensuring efficient operation and creating favorable conditions for crude oil transportation. This approach also effectively enhances the operational efficiency of oil transfer stations[1].

External pumps exhibit diverse failure patterns, making it challenging to accurately diagnose specific faults through single-parameter detection alone. Comprehensive assessment requires simultaneous integration of multi-dimensional parameter characteristics. Multi-source data fusion technology offers an effective solution for this challenge, garnering widespread attention and application. Scholars such as Yu et al.[2]established the core framework of this technology by analyzing information hierarchy and functional classification dimensions, laying a solid theoretical foundation for subsequent research. Dong et al. [3] conducted a review from the perspective of fusion algorithms, advancing the technology's international dissemination. The core application of multi-feature fusion technology in external pump fault diagnosis lies in achieving effective decomposition of raw signals and precise extraction of key features.

Signal decomposition techniques have experienced iterative evolution from experience-driven approaches to model optimization. While early Empirical Mode Decomposition (EMD) could adaptively process non-stationary signals, it suffered from inherent limitations such as mode aliasing and endpoint effects[4]. The VMD algorithm proposed by Dragomiretskiy et al. effectively overcomes EMD's limitations by constructing a variational model to adaptively separate modal components, emerging as the current mainstream signal decomposition method[5]. However, VMD's decomposition effectiveness relies on the preset number of modes K, making it prone to over- or under-decomposition phenomena[6][7]. To address this, Ge et al. [8]combined the Differential Evolution (DE) algorithm with the Grey Wolf Optimization (GWO) algorithm, introducing a novel fitness function that significantly enhances VMD's fault diagnosis capabilities. Rostaghi and Azami proposed the Dispersion Entropy (DE) algorithm, which provides significantly lower computational complexity than Sample Entropy while preserving amplitude information that is ignored by Permutation Entropy[9]. However, Conventional dispersion entropy is limited to single-scale analysis and cannot effectively capture complexity variations across multiple temporal scales[10]. Based on this limitation, Azami et al. [11] proposed multi-scale dispersion entropy (MDE), which expands the analysis scale through coarse-graining. Yet, its use of equidistant segmentation for averaging tends to lose statistical information and is susceptible to computational bias due to the influence of initial point positions. To overcome these limitations, Lv et al. [12] employed refined composite multiscale dispersion entropy (RCMDE). By utilizing multi-initial-point coarsening and probabilistic mean calculation, the RCMDE effectively reduces information loss and computational bias, demonstrating superior accuracy and stability in nonlinear signal feature extraction.

After processing fault signals, accurately determining whether a fault has occurred and its evolution state is the core objective for achieving equipment condition monitoring and proactive early warning. Therefore, the fault diagnosis phase constitutes the critical closed-loop of the entire monitoring system. The emergence of fault diagnosis technology carries both historical context and significant practical implications. The rise of intelligent technologies has propelled its development into a new phase, with extensive research conducted by scholars worldwide on fault diagnosis for large rotating machinery[13]. Among various intelligent diagnostic methods, the LSSVM has gained widespread application and optimization in rotating machinery fault diagnosis due to its excellent adaptability to small samples and nonlinear data[14]. Compared to traditional Support Vector Machines (SVM), LSSVM converts inequality constraints into equality constraints, significantly reducing computational complexity and accelerating convergence speed[15]. Addressing the sensitivity issues of LSSVM's core parameters (regularization parameter C and kernel parameter g), researchers have proposed multiple optimization strategies. Liang et al. [16] investigated different cavitation margins and applied a genetic algorithm-optimized LSSVM to cavitation margin detection, achieving precise diagnosis of cavitation faults.

In summary, research scholars have studied the diagnosis based on various intelligent algorithms, such as the genetic algorithm, in fault diagnosis, respectively, and have theoretically analyzed and researched various analytical methods in signal decomposition and feature extraction, which provide a solid academic foundation for the fault diagnosis of external pumps. However, the application of these fault diagnosis methods on external pumps is very limited, and the accuracy and effectiveness of the existing techniques are insufficient. Therefore, this paper proposes a multi-feature fusion fault diagnosis method for external pumps based on improved VMD-RCMDE. Through the fusion of multiple algorithms and features, this study can analyze multiple fault factors and determine multiple fault types at the same time, obtaining more accurate calculation processes and diagnosis results, this study is based on a large number of fault data obtained from the actual testing, which is more convincing and effective, and it has an important application value in the production and operation of the external pumps of the oil transfer station.

2. Principles and Methods

2.1 GSWOA-VMD signal decomposition methods

The VMD is a typical time-frequency analysis method capable of decomposing a non-stationary time series into a number of eigenmode functions with different frequency characteristics. In this study, the VMD is used to decompose the critical pressure signals of the external pump of an oil transfer station under abnormal operating conditions in the frequency domain to explore its potential fault characteristics. However, the VMD algorithm needs to preset two core parameters, the decomposition modal number K and the penalty factor α , whose values directly affect the decomposition accuracy and modal distribution effect. Since the traditional VMD lacks an adaptive parameter adjustment mechanism, it is difficult to realize optimal decomposition under actual complex working conditions, thus limiting its diagnostic accuracy and robustness in engineering applications.

To overcome the above limitations, this study optimizes the VMD algorithm with a more efficient GSWOA based on the global search strategy, which obtains the optimal combination of the optimized decomposition scale K and the penalty factor α during the whale position updating[17][18], and the optimal combination of the decomposition scale K and the penalty factor α is obtained, which can be used for the diagnosis of the whale position update, thus solving the adaptive problem in the VMD decomposition process.WOA is a swarm intelligence optimization algorithm that mimics the hunting process of whales in nature. It can be divided into three phases: encircling the prey phase, bubble-net attacking phase, and random prey hunting phase. In each phase, the position of the whale is updated. The process of problem-solving using the whale optimization algorithm involves representing the position of each whale as a feasible solution and obtaining the optimal solution by continuously updating the position of the whales. In contrast, GSWOA optimizes WOA by adopting an adaptive weighting strategy, a variable spiral position updating strategy, and an optimal domain perturbation strategy to improve the convergence speed and global search capability of the traditional whale optimization algorithm[19].

The GSWOA-VMD optimization algorithm combines the advantages of the whale algorithm based on the global search strategy and the variational modal decomposition, which can adaptively adjust the search strategy and improve the algorithm's optimization ability. Its algorithm flowchart is shown in **Figure 1**.

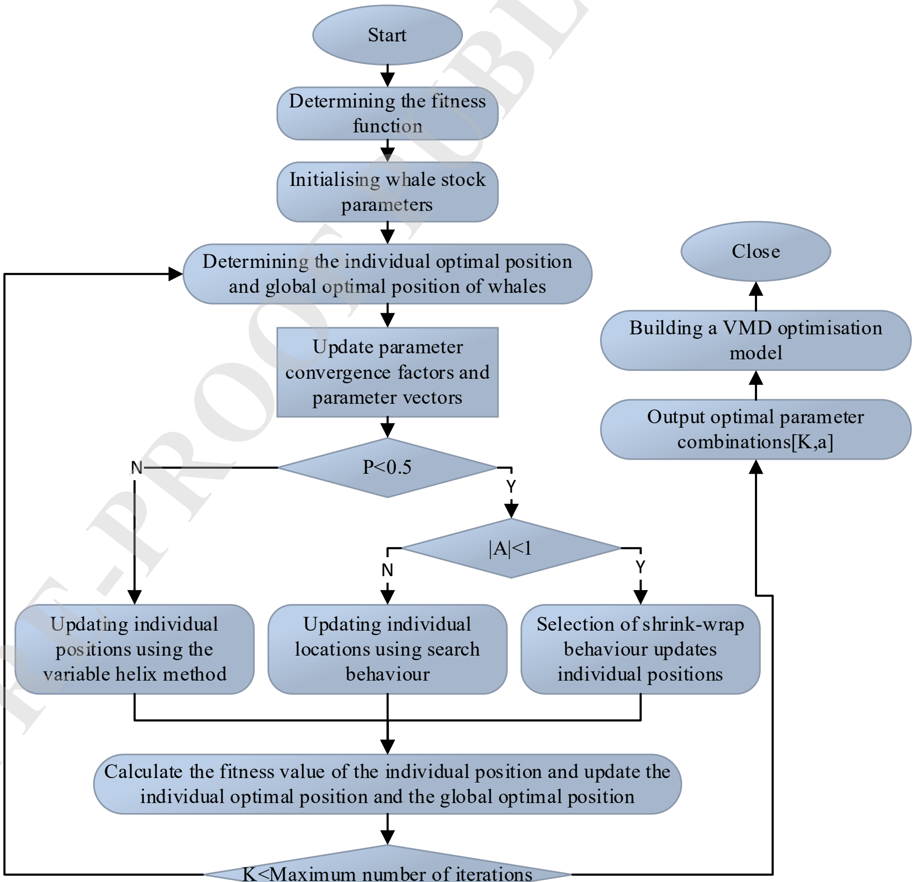


Figure 1. Flow chart of GSWOA optimization of VMD parameters

2.2 RCMDE signal feature extraction methods

The RCMDE is a refinement of the MDE, which is more rapidly computed and is more suitable for analyzing long time series such as fault signals[20][21]. When the RCMDE method is employed to compute the RCMDE with a scale factor τ, the original time series is segmented into non-overlapping intervals of length τ, starting from the initial interval [1, τ]. The mean value of each segment is computed, and these means are subsequently sorted to generate τ coarse-grained time series. In the computational procedure, the probability distribution of the dispersion patterns π across all sequences is first determined. The average probability for each pattern is then calculated, followed by the computation of RCMDE using Equation (2.1). The solution steps are as follows:

The time series u is coarsely granulated, and the hth coarse-grained approximation signal $\alpha_{h,\gamma}^{\tau}=\left\{\alpha_{h,1}^{\tau},\ \ \alpha_{h,2}^{\tau},\cdots\right\}$ is obtained as Eq. (2.1):

$$\alpha_{h,\gamma}^{\tau}=\frac{1}{\tau}\sum_{k=h+\tau(\gamma-1)}^{h+\gamma\tau-1}u_k \tag{2.1}$$

In equation (2.1), γ is an integer from 1 to N, N is the ratio of length to scale factor, k represents an integer from 1 to τ .

For different scale factors τ , the RCMDE is defined in equation (2.2):

$$\begin{gathered}RCMDE(X,m,c,d,\tau)=-\sum_{\pi=1}^{c^m}\overline{P}\left(\pi_{v_0v_1\cdots v_{m-1}}\right)\log_2\left(\overline{P}\left(\pi_{v_0v_1\cdots v_{m-1}}\right)\right)\\ \overline{P}\left(\pi_{v_0v_1\cdots v_{m-1}}\right)=\frac{1}{\tau}\sum_{h=1}^{\tau}P_h^{\tau}\end{gathered} \tag{2.2}$$

In equation (2.2), $\overline{P}\left(\pi_{v_0v_1\cdots v_{m-1}}\right)$ represents the mean probability of the dispersion pattern π for the coarse-grained approximation signal $\alpha_{h,\gamma}^{\tau}$; m represents the embedding dimension; c represents the number of classes; d represents the time delay, typically set to 1.

A set of simulated signals is used to verify the advantages of the RCMDE in feature extraction. Thirty sets of random noise and Gaussian white noise with a data length of 3000 are compared, and the RCMDE means of the two types of noise over 15 scale factors are computed separately and compared with the MDE, the mean-square error (MSE), and the multiscale permutation entropy(MPE). In the RCMDE vs. the MDE, the embedding dimension is set to be 3, the category to be 6, and the time delay to be 1; MSE is set to have an embedding dimension of 3, and a similarity tolerance of 0.15; and MPE sets the embedding dimension to 3 and scale factor 15, and the results are shown in **Figure 2**.

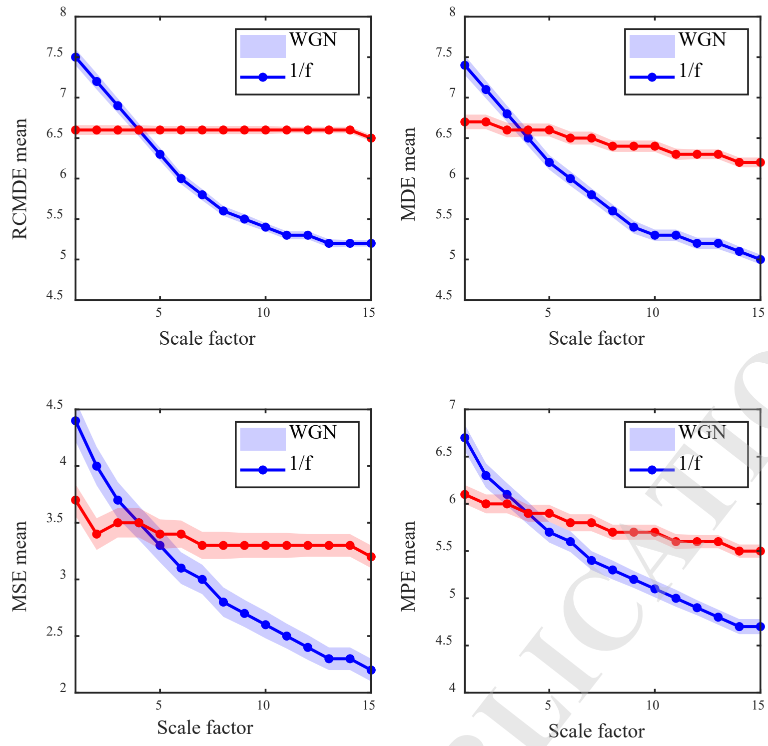


Figure 2. Results of the four approaches

As shown in **Figure 2**, at small scales, the entropy values of white Gaussian noise (WGN) calculated using all four entropy methods are higher than those of 1/f noise, indicating that WGN exhibits greater irregularity and complexity at small scales. As the scale factor increases, the RCMDE value of WGN gradually decreases from approximately 7.5 to around 5.2, while the RCMDE value of 1/f noise changes more gradually, remaining largely between 6.6 and 6.5. Therefore, the two types of noise remain distinctly distinguishable across all scales. The figure further shows the $\pm 1\sigma$ bands corresponding to the mean values of each method. It can be seen that RCMDE has the narrowest standard deviation band across the entire scale range, indicating that it exhibits smaller result fluctuations and better stability at different scales; at the same time, a large difference in entropy values is consistently maintained between the two types of noise, suggesting that RCMDE can more effectively distinguish between different types of noise. In contrast, the standard deviation bands for MDE, MPE, and MSE are relatively wide, and the curves exhibit significant fluctuations across certain scales, indicating that their stability is relatively weak and their ability to distinguish between different noise types is inferior to that of RCMDE.

2.3 Multi-feature fusion methods

In conclusion, this paper proposes a multi-feature fusion fault diagnosis method based on GSWOA-VMD and SVM. The method utilizes the pressure signal as the primary feature to extract fault feature vectors, serving as the core diagnostic basis. Meanwhile, temperature and vibration signals are incorporated as auxiliary features. To eliminate discrepancies in magnitude and dimensionality, feature parameters of diverse types and dimensions are uniformly preprocessed using Z-score standardization. Subsequently, the standardized auxiliary features are fused with the pressure-derived feature vector, enabling the integration of multi-dimensional fault information and thereby providing a more comprehensive foundation for accurate fault diagnosis. The methods and steps of parameter fusion are as follows:

(1) Scale decomposition of the pressure feature parameters using the improved VMD algorithm of GSWOA to obtain m IMF components.

(2) The RCMDE values of the m pressure IMF components are extracted and form the dispersion entropy feature vector P.

$$P = [p_1 p_2 \dots p_m] \tag{2.3}$$

In equation (2.3), $p_1 p_2 \dots p_m$ represents the corresponding RCMDE value for each IMF component.

(3) In addition to the pressure feature parameter, temperature is also a major parameter reflecting the fault. Calculate the RCMDE values of the temperature parameters and form the temperature feature matrix T , $T = [t_1 t_2 \dots t_n]$, n represents the number of temperature parameters. In some cases, it is necessary to fuse the temperature with other parameters to form a comprehensive criterion. The temperature dispersion entropy eigenvector T is fused with the pressure eigenvector P to form the characteristic parameter matrix PT.

$$PT = [p_1 p_2 \cdots p_m t_1 t_2 \cdots t_n] \tag{2.4}$$

In equation (2.4), $t_1 t_2 \cdots t_n$ represents the RCMDE value corresponding to the temperature characteristic parameter.

(4) The vibration amplitude serves as a critical parameter for assessing the operational status of the external pump. Owing to its unique data characteristics, it is designated as $V = [v_1 v_2 \cdots v_k]$, k represents the number of vibration parameters. By integrating the vibration amplitude with the temperature dispersion entropy feature vector T, a combined TV feature parameter matrix is established:

$$TV = [t_1 t_2 \cdots t_n v_1 v_2 \cdots v_k] \tag{2.5}$$

In equation (2.5), $v_1 v_2 \cdots v_k$ represents the vibration amplitude parameter.

The flow of the fusion of the fault characteristic parameters PTV of the external pump is shown in **Figure 3**.

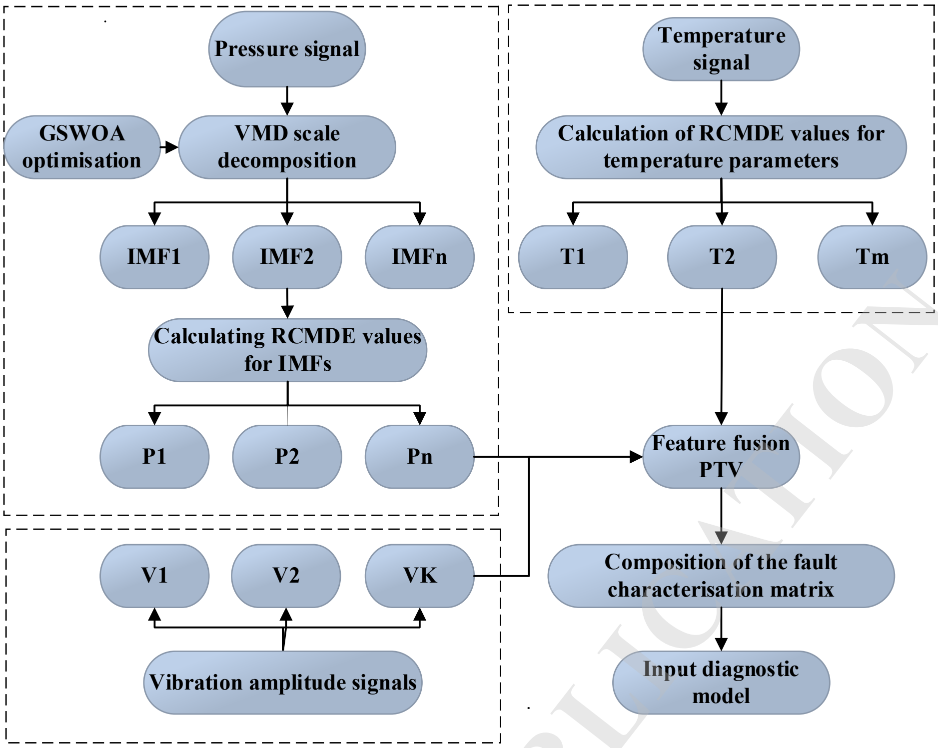


Figure 3. Fusion Model of Fault Characteristics of External Pumps

2.4 SGA-LSSVM Fault Diagnosis Methodology

After completing the feature vector extraction of the external pump signals, the pattern recognition of faults becomes the last step in the intelligent diagnosis of external pumps. Research has shown that support vector machine has a great advantage in fault diagnosis based on the signals of machine and pump equipment, and have an important position in the field of pattern recognition[22].

Therefore, this chapter proposes the SGA optimized LSSVM fault diagnosis model by combining the characteristics of the external pump signals, and the optimal set of feature vectors obtained is input into the diagnostic model for the pattern recognition of the faults. The SGA algorithm is inspired by the migratory behaviors of the snow geese and imitates the unique "herringbone" and "straight line" shaped flight patterns observed during their migration[23]. In this section, the Snow Geese algorithm is used to optimize the parameters of the LS-SVM model for pattern recognition of the fault characteristics of the external pump.

The LS-SVM is an improved algorithm based on SVM, which simplifies the inequality constraints in the SVM method into equality constraints, decomposes the multi-output problem into m binary classification problems, reduces the complexity of the solution, ensures that the final solution can be obtained at a faster speed[24], and has a stronger generalization ability. However, the settings of the regularization parameter γ and the squared bandwidth δ_2 of the RBF kernel function in the model play a decisive role in the performance of the classifier, and no better standard has been produced on the selection of this parameter, so it is necessary to further optimize the parameters by using the SGA in order to improve the classification effect of the LS-SVM model.

The parameters of the LS-SVM model are optimized using the Snow Geese Algorithm in the following steps:

(1) Initialize the individual location parameters: set the particle population size, the solution space boundary (u_b, l_b) and dimension, and create a two-dimensional matrix $X = [\gamma, \delta_2]$ to represent the location of the individual.

(2) Establish the LS-SVM model based on the Gaussian RBF kernel by adjusting the two parameters of γ and δ_2 to train the external pump fault feature samples, see equation (2.6):

$$Model = trainlssvm\left(\left\{x_{train}, y_{train}, f', \gamma, \delta_2, RBF - \text{kernel}'\right\}\right) \tag{2.6}$$

(3)In each iteration, the LS-SVM model is evaluated using five-fold cross-validation conducted exclusively on the training dataset. In each fold, four subsets are used for model training and the remaining subset is used for validation. The mean validation error is used as the fitness value f_i according to Equation (2.7). The independent test set is not involved in the parameter optimization process.

$$f_i = \frac{1}{m_{val}} \sum_{j=1}^{m_{val}} \left(y_{val,j} - \tilde{y}_{i,j}\right)^2 \tag{2.7}$$

In Equation (2.7), m_{val} denotes the number of validation samples in each cross-validation fold; $y_{val,j}$ and $\tilde{y}_{i,j}$ represent the actual and predicted labels of the j-th validation sample, respectively; and f_i represents the fitness value of the i-th individual.

(4) By comparison with the values of the fitness function, the velocity V and position X of the individual are updated.

(5) Determine whether the preset number of iterations or the difference between neighboring generations satisfies the termination condition of the minimum error limit. Return to step (2) if it is not satisfied, and output the optimal parameter values γ and δ_2 if they are satisfied.

(6) Bring the γ and δ_2 obtained from the optimization search into the LS-SVM model for training again, then the optimized model can be obtained.

The γ and δ_2 parameters of LS-SVM are globally optimized by SGA parameters, which reduces the randomness of the parameters. When the data feature set is larger, using SGA to globally optimize the LS-SVM parameters for model construction and training, the final model recognition effect will have higher accuracy and faster convergence speed.

3. Results and Discussion

3.1 Improved VMD performance analysis

To demonstrate the advantages of the improved VMD in decomposing external pump signals, we selected the signal from an external pump at an oil transfer station in a certain oilfield in Xinjiang for analysis. Taking an instance of abnormal fluctuations in the inlet pipeline pressure of the external pump over a specific period as an example, we compared the decomposition performance of the VMD optimized using GSWOA with that of the VMD optimized using a genetic algorithm (GA), and conducted a quantitative analysis using three evaluation metrics: the correlation coefficient, energy distribution, and relative RMS deviation.

(1) correlation coefficient

The correlation coefficients is an important evaluation index to screen whether the decomposition of IMFs is effective or not. Let the original signal be $x(t)$ and the decomposed signal be $c_i(t)$, and calculate the correlation coefficient ρ_i between $c_i(t)$ and $x(t)$ as in equation (3.1):

$$\rho_i\left[c_i(t),x(t)\right]=\frac{\operatorname{cov}\left[c_i(t),x(t)\right]}{\sigma\left(c_i(t)\right)\times\sigma\left(x(t)\right)} \tag{3.1}$$

In equation (3.1), cov is the covariance of the original signal with respect to each component, σ represents the standard deviation of the original signal and each component.

(2) Energy distribution

The energy distribution is the degree of energy preservation, calculate the energy preservation degree of the VMD decomposition signal can reflect the good or bad decomposition effect. A good decomposition effect indicates that more energy of the original signal is preserved in the decomposed IMF, and the energy preservation degree is closer to 1; on the contrary, it indicates that more energy is leaked after the signal is decomposed.

$$IE=\frac{\sum_{t=0}^{T}\sum_{i=1}^{m}\left|c_i(t)\right|^2}{\sum_{t=0}^{T}\left|x(t)-r_n(t)\right|^2} \tag{3.2}$$

In equation (3.2), $r_n(t)$ represents the residual term.

(3) Relative RMS deviation

The relative root mean square (RMS) deviation is used to evaluate the consistency between the overall RMS energy of the decomposed IMF components and that of the original signal. It is defined as the relative difference between the combined RMS of all IMF components and the RMS of the original signal, as shown in Equation (3.3). A smaller value indicates that the decomposed components preserve the overall RMS energy characteristics of the original signal more effectively.

$$\theta = \frac{\left| \sqrt{\sum_{i=1}^{n} RMS_i^2} - RMS_{original} \right|}{RMS_{original}} \tag{3.3}$$

In equation (3.3), $RMS = \sqrt{\sum_{j=1}^{m} s^2(j)/m}$ represents the root mean square value of the IMF, $RMS_{original}$ represents the root mean square value of the original signal.

Table 1. Quantitative comparison results of improved VMD algorithms

Comparison term	GSWOA-VMD	GA-VMD
Decomposition layer	9	10
Maximum correlation coefficients	0.9013	0.8127
Energy distribution	0.9552	0.6685
Relative RMS deviation	0.0226	0.0824

The quantitative comparison results between the GSWOA-improved VMD algorithm and the GA-improved VMD algorithm are shown in Table 1. As shown in the table, GSWOA-VMD achieves a maximum correlation coefficient of 0.9013, which is higher than the value of 0.8127 obtained by GA-VMD. This indicates that the selected IMF component obtained by GSWOA-VMD has a stronger correlation with the original inlet pipeline pressure signal and better preserves the characteristics of the original signal. In terms of energy distribution, the value obtained by GSWOA-VMD is 0.9552, with a deviation of only 0.0448 from 1, whereas the value obtained by GA-VMD is 0.6685, with a deviation of 0.3315 from 1. Therefore, GSWOA-VMD exhibits better energy-preservation performance. In addition, the relative RMS deviation of GSWOA-VMD is 0.0226, which is lower than the value of 0.0824 obtained by GA-VMD. This result indicates that the IMF components generated by GSWOA-VMD maintain better consistency with the overall RMS energy of the original signal. Therefore, the GSWOA-improved VMD method demonstrates better overall decomposition performance for the external pump pressure signal.

3.2 Analysis of the Effectiveness of the RCMDE Feature Extraction Method

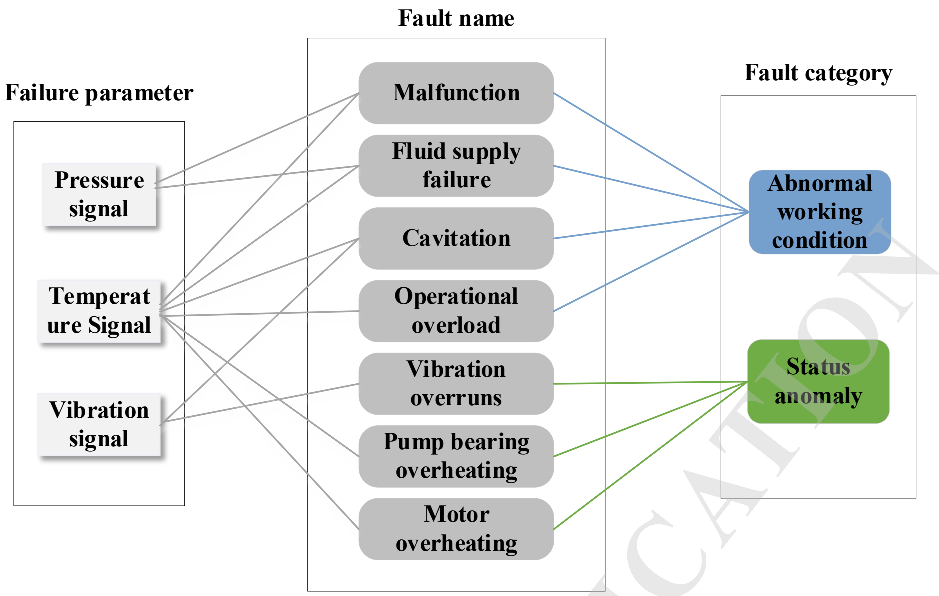


Figure 4. Common Fault Categories and Fault Signal Parameter Relationship Chart

In order to verify the validity of the method in this paper, an oil extraction plant in Xinjiang, for example, an oil transfer station of an external pump, the 11 sensors installed in the external pump body and motor, the distribution of measurement points as shown in **Figure 5**. from the actual investigation, it can be seen, caused by the external pump of the common failure signals are mainly temperature, pressure, and vibration signals, the common types of faults and fault signal parameter relationships as shown in **Figure 4**. Therefore, the required three kinds of fault signals are measured, and these fault signal parameters are used as data for subsequent signal feature extraction and fault diagnosis.

Figure 5. CSDK90-50*10*220kW external pump measurement point layout location

In order to verify the effectiveness of the RCMDE feature extraction, the failure of the liquid supply under abnormal working conditions is taken as an example for processing and analysis. A total of 1024 sets of historical data of inlet pipeline pressure, differential pressure on both sides of the filter, and pump temperature are intercepted, including all the data of the external pump from the normal state to the abnormal liquid supply and then back to normal.

The trend of the inlet pipeline pressure, differential pressure on both sides of the filter, and pump temperature is shown in **Figure 6**.

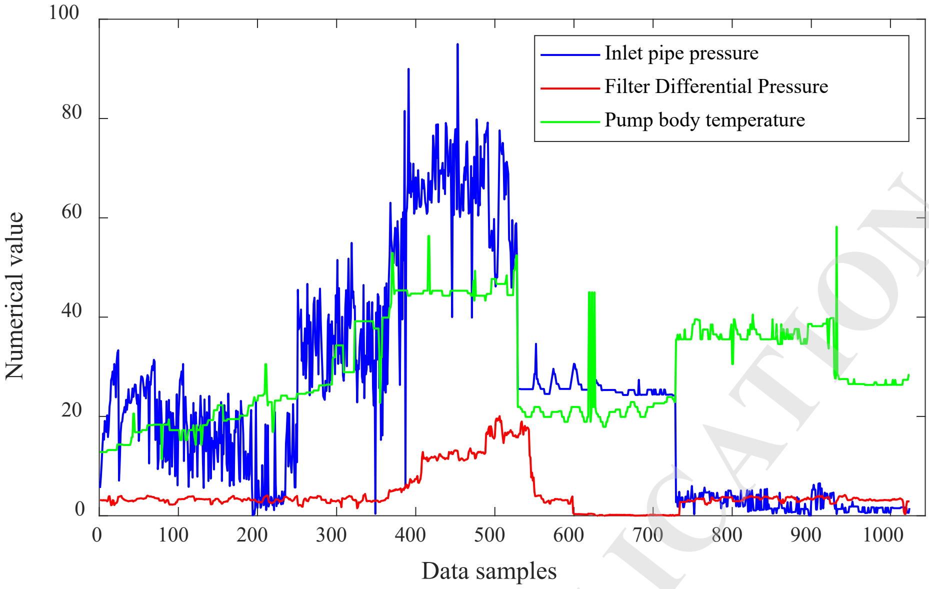


Figure 6. Trend graph of characteristic parameters of fluid supply anomalies

The inlet pipeline pressure and differential pressure data on both sides of the filter were subjected to GSWOA-VMD signal decomposition to obtain the IMFs component, as shown in **Figures 7 and, Figure 8**.

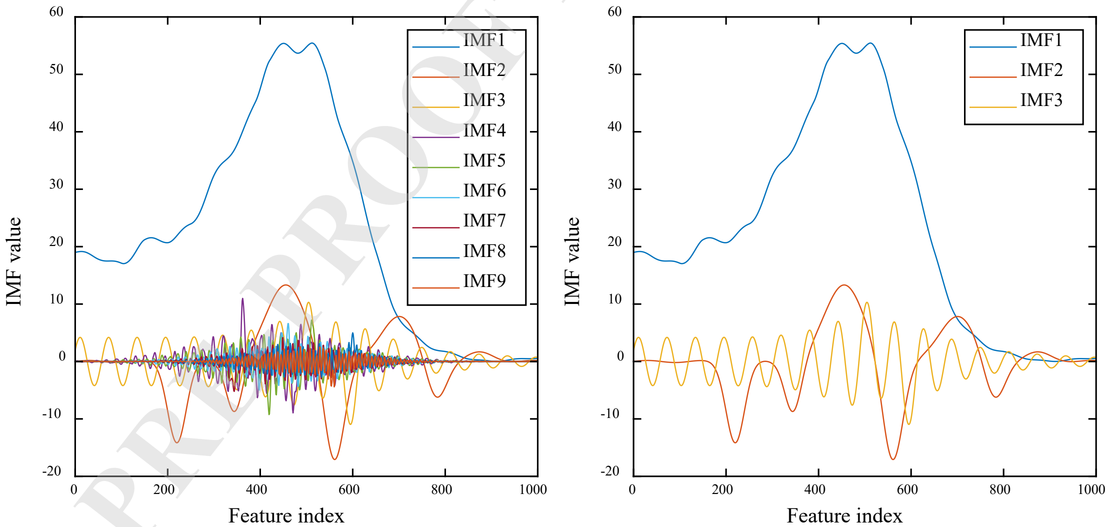


Figure 7. IMFs for inlet pipeline pressure

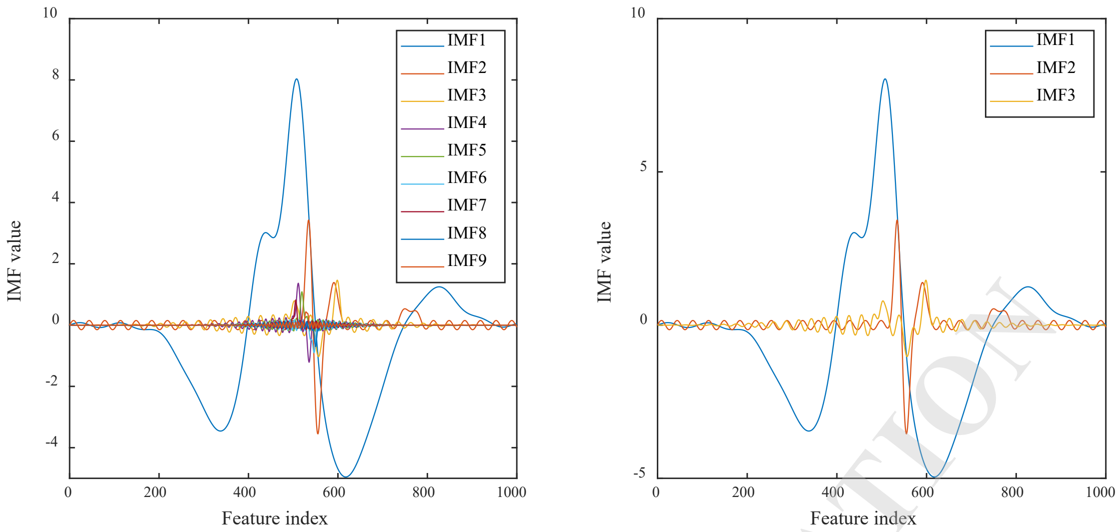


Figure 8. IMFs for differential pressure on both sides of the filter

The first three IMF components that have the strongest correlation with the original signal among the nine layers of IMF components obtained from decomposition are selected, followed by multiscale feature extraction of inlet pipeline pressure, IMFs of differential pressure on both sides of the filter, and pump body temperature using RCMDE, and the results are compared with the extraction results of the MPE. The results are shown in **Figure 9, Figure 10, and Figure 11**.

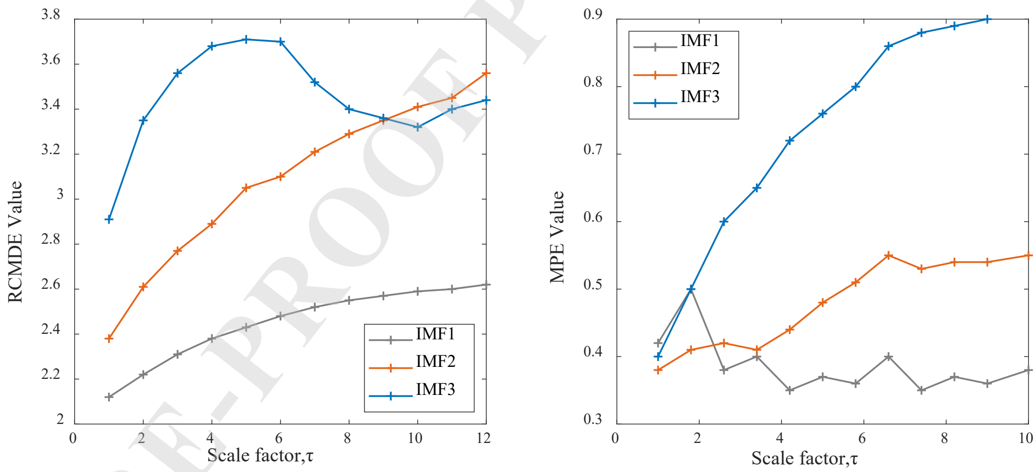


Figure 9. Entropy characterization of IMFs for inlet pipeline pressure

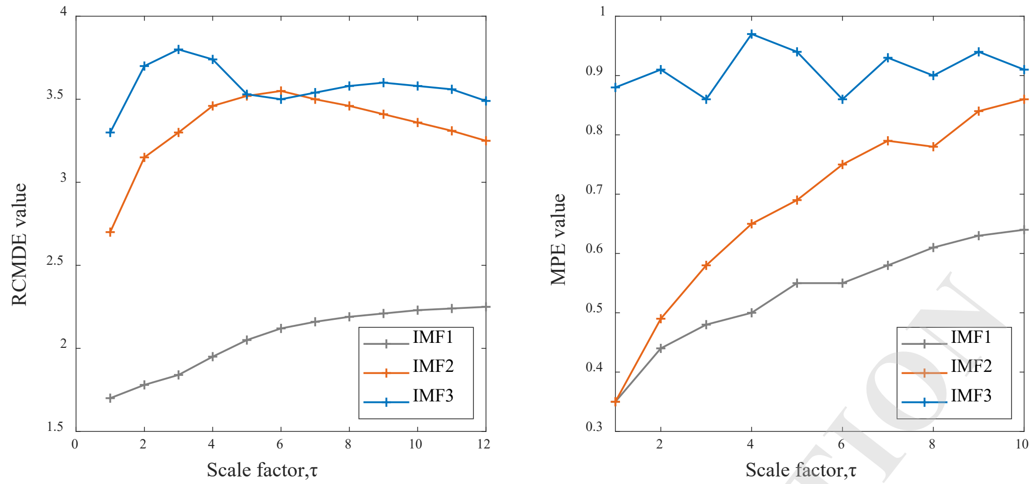


Figure 10. Entropy characterization of IMFs for differential pressure on both sides of the filter

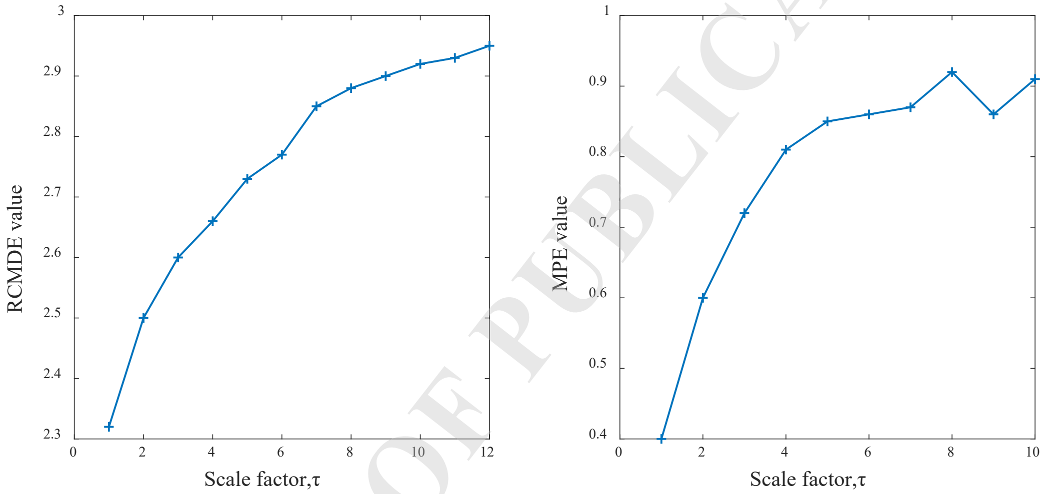


Figure 11. Entropy characterization of pump body temperature

As shown in Figures 9–11, RCMDE generally exhibits smoother variations across scale factors and clearer separation among the IMF components of the two pressure signals than MPE. For the pump-body temperature signal, the RCMDE values also show a more continuous multiscale trend, whereas the MPE curve presents more pronounced local fluctuations at several scale factors. These results indicate that RCMDE provides relatively stable and discriminative multiscale features, thereby providing a more reliable feature representation for subsequent fault classification.

3.3 Analysis of the methodological validity of this study

For the four cases of abnormal working conditions, the characteristic parameters of abnormal liquid discharge, abnormal liquid supply, Cavitation, and operation overload are selected for testing. The characteristic parameters include inlet pipeline pressure, pump body temperature, outlet pipeline pressure, differential pressure on both sides of the filter, non-drive end bearing temperature, drive end bearing temperature, motor bearing extension end temperature, motor bearing fan end temperature, non-drive end bearing vibration, drive end bearing vibration, and

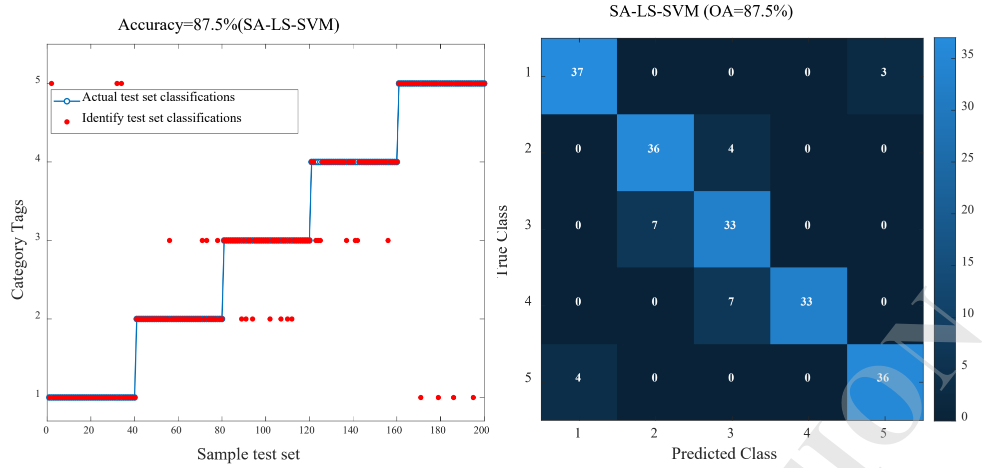
motor stator temperature, i.e., the model input is 11-dimensional parameters. The entropy values of 12 intrinsic mode functions (IMFs) derived from the decomposition of three pressure characteristic parameters, along with the entropy values of five temperature characteristic parameters, were selected. These were subsequently combined with two vibration parameters, resulting in a fused dataset with a total dimensionality of 19.
A total of 1000 data samples were selected, comprising 200 sets for each of the normal condition and four abnormal conditions. Each sample set includes 19 characteristic parameters. To ensure objective model evaluation, the dataset was stratified by operational condition categories. Within each category, 160 samples were allocated to the training set and the remaining 40 to the test set. The resulting training feature matrix has dimensions of 800×19, denoted as C^(800×19), while the test feature matrix is 200×19, denoted as C^(200×19). The five operational states are labeled and encoded numerically from 1 to 5.

During the optimization of γ and δ^2, five-fold cross-validation was performed within the 800-sample training set. After the optimal parameters were determined, the final LS-SVM model was retrained using all 800 training samples and evaluated on the independent 200-sample test set.

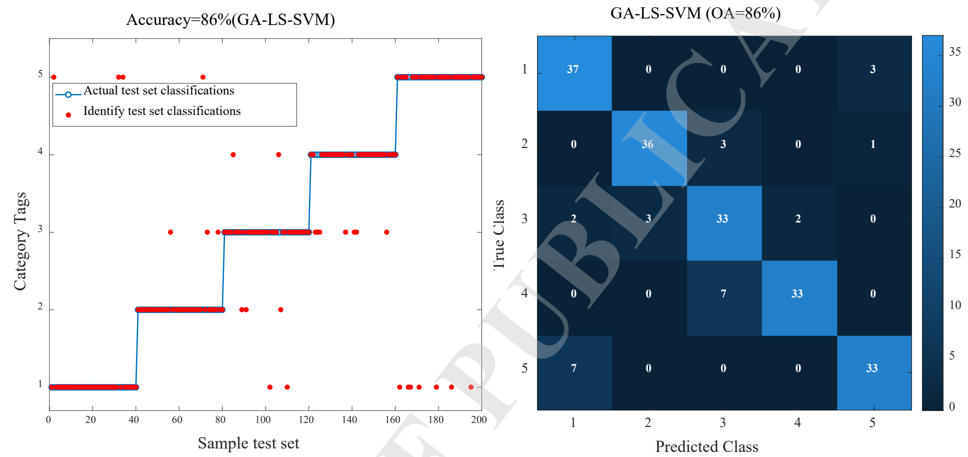
Table 2. Model input feature information

Type of working condition	Sample size	Eigenvector (math.)	Status tag
Normal state	1-200	19	1
Abnormal liquid discharge	201-400	19	2
Abnormal liquid supply	401-600	19	3
Cavitation	601-800	19	4
Operating overload	801-1000	19	5

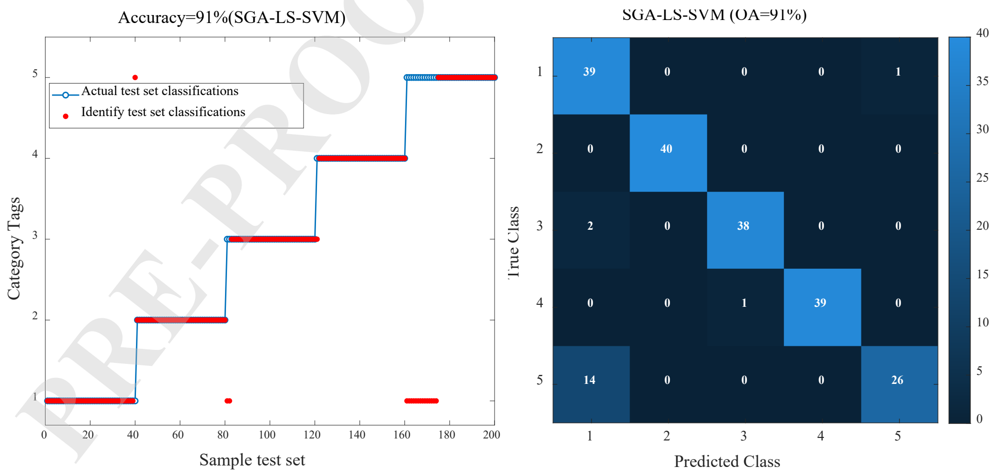
The pressure signals are first modally decomposed using the improved VMD algorithm of GSWOA, and then the entropy features of the pressure IMFs and temperature signals are extracted using the RCMDE and fused with the vibration parameters to form the feature set for fault diagnosis input to the model for pattern recognition. The model input feature signals are shown in **Table 2**.



(a)



(b)



(c)

Figure12. Pattern recognition results and confusion matrices of LS-SVM optimized by SA(a), GA(b), and SGA(c)

The LS-SVM models optimized by SGA, GA, and simulated annealing (SA) are used to perform pattern recognition on the feature set. The recognition results and corresponding confusion matrices are shown in **Figure 12**, and the diagnostic result accuracy of each optimization algorithm is shown in **Table 3**. The recognition accuracy for overloading is as low as 65% because the characteristic parameters for overloading encompass both the motor and the pump body. The wide variety of parameter types, combined with mutual interference from mechanical structures, results in highly complex variations in the raw signals, thereby reducing the model's ability to accurately identify this operating condition.

Table 3. SGA, GA, and SA optimized the LS-SVM model recognition accuracy statistics

Fault diagnosis model	Test sample	Accuracy(%)	Overall Accuracy(%)
SGA-LSSVM	Normal state	97.5	91
	Abnormal liquid discharge	100	
	Abnormal liquid supply	95	
	Cavitation	97.5	
	Operating overload	65	
GA-LSSVM	Normal state	92.5	86
	Abnormal liquid discharge	90	
	Abnormal liquid supply	82.5	
	Cavitation	82.5	
	Operating overload	82.5	
SA-LSSVM	Normal state	92.5	87.5
	Abnormal liquid discharge	90	
	Abnormal liquid supply	82.5	
	Cavitation	82.5	
	Operating overload	90	

The optimization results and search times of the three optimization methods for the regularization parameter γ of LS-SVM and the squared bandwidth δ_2 in the RBF kernel function are shown in **Table 4**.

Table 4. Three algorithms to optimize LS-SVM results

Optimization algorithm	Regularization parameter γ	Square bandwidth δ_2	Optimization time(s)	Accuracy（%）
SGA	21.604	2.475	181	91
GA	18.955	12.378	127	86
SA	47.753	1.571	196	87.5

From **Table 4**, it can be concluded that the optimization time of the SGA optimization algorithm is 181s, which is relatively moderate compared to the other two optimization algorithms, and it will not be too fast to cause a decrease in accuracy or too slow to reduce the operational efficiency. From the recognition results of the three algorithms in Fig. 12, it can be seen that the SGA optimization algorithm has the least part of deviation between the actual test set and the recognition test set, and from Tables 3 and 4, it can be seen that the accuracy rate of SGA is 91%, which is improved by 5% and 3.5% compared with GA and SA respectively. It can be seen that the recognition ability of the LS-SVM classifier optimized by SGA is better, and it is the most effective for the recognition of the results after signal decomposition and feature extraction of the external pump in this paper, with the highest accuracy of fault diagnosis.

4. Conclusions

This paper aims to address common failure problems that occur during the operation of external pumps in oil transfer stations. To this end, it investigates a signal decomposition method based on GSWOA-optimized VMD and RCMDE feature extraction. It also constructs an information dataset that integrates multiple features and carries out fault identification experiments and analyses based on an LS-SVM fault diagnosis model that is optimized by SGA. The research conclusions are as follows:

(1) This paper introduces the GSWOA to jointly optimize the modal number K and the penalty factor α in VMD decomposition. The optimized parameters are then applied to VMD decomposition to obtain more representative IMF components. This solves the problems of strong parameter dependence and lack of adaptivity in the traditional VMD method. Comparing the decomposition effects of GSWOA-VMD and GA-VMD verifies the improved method's advantages and accuracy in signal decomposition.

(2) In terms of feature extraction, the RCMDE method is used to extract entropy features from the IMF component of the pressure signal. At the same time, the RCMDE features of the temperature and vibration amplitude signals are combined to construct a fault feature dataset that integrates information from multiple sources. This provides reliable data support for subsequent fault mode identification. RCMDE features were further compared with traditional MPE extraction results. The results showed that RCMDE is more suitable than MPE for extracting external pump fault features in terms of expressing multi-scale information and classifying differences.

(3) The SGA is used to optimize the regularization parameter γ and the bandwidth parameter δ_2 of the radial basis kernel function for the LS-SVM model, and to construct a high-precision fault identification classifier. Test samples are fed into the trained model for fault identification. The results show that the SGA-optimized LS-SVM model is significantly more accurate than the GA- and SA-optimized models, reaching 91% accuracy compared to 86% and 87.5%, respectively. Furthermore, the comparison of fault identification accuracy among the three optimization algorithms demonstrates the superiority of the proposed SGA in optimizing the LS-SVM parameters, thereby improving the fault diagnosis performance of the developed method.

5. Acknowledgments

This study was partially supported by the Nanchong-Southwest Petroleum University Municipal and University Science and Technology Strategic Cooperation Project (No. 23XNSYSX0022/0026) and the Sichuan Provincial Science and Technology Program (No. 2023ZHCG0020) and Sichuan Provincial International, Hong Kong, Macao, and Taiwan Science and Technology Innovation Cooperation Project (No. 2025YFHZ0218).

References

[1] Ding H., Lei W., Li H T., Research on maintenance and countermeasures of the external pump in Oil transmission Station [in Chinese], *Modern Business Trade Industry*, (18):194, 2014, https://ss.zhizhen.com/detail_38502727e7500f269ce4af58a839b663171d30a9415e44bf1921b0a3ea255101fc1cf1fbb4666ae6a39c6becd35a17d8189f6edd49fcbbbc38762bd7a7e16d1eab13b643b53bc101a3d8df49be4dcda5.

[2] Yu W X., Yong S W., Guo G R., Comments on Multisensor Information Fusion [in Chinese], *Journal of National University of Defense Technology*, (3): 1-11,1994, https://ss.zhizhen.com/detail_38502727e7500f26a9dc1098c5271258ec6cca749bfe38291921b0a3ea255101fc1cf1fbb4666ae6e80c46f2cb6a01e35b47735b0fe92e355758cafb8235d8f19f9ab2474a8142af83a8a5002f7cc9d0.

[3] Dong J., Zhuang D F., Huang Y H., Fu J Y., Advances in Multi-Sensor Data Fusion: Algorithms and Applications, *Sensors (Basel, Switzerland)*, **9**(10): 7771-7784, 2009, doi: 10.3390/s91007771.

[4] Wang M Z., Xu W X., Chen S H., Feng J., On the end effect and end condition methods of empirical mode decomposition with cubic spline, *Communications in Nonlinear Science and Numerical Simulation*, **153**: 109532, 2026, doi: 10.1016/j.cnsns.2025.109532.

[5] Dragomiretskiy K., Zosso D., Variational mode decomposition(Article). *IEEE Transactions on Signal Processing*, **62**(3): 531-544, 2014, doi: 10.1109/TSP.2013.2288675.

[6]Li H P., Zhou F X., Xu B., Yan B K., Zhou F Q., Application of Auto-Regulative Sparse Variational Mode D ecomposition in Mechanical Fault Diagnosis, *Electronics*, **12**(14): 3081, 2023, doi: 10.3390/electronics1214 3081.

[7]LAKIKZA A., CHEGHIB H., KAHOUL N., Optimized variational mode decomposition for improved beari ng fault diagnosis and performance evaluation, *Archive of Mechanical Engineering*, **71**(4): 467-495, 2024, d oi: 10.24425/ame.2024.152615.

[8]Ge L., Fan W., Xiao X T., Gan F J., Lai X., Deng H X., Huang Q., Rolling bearing fault diagnosis method ba sed on improved variational mode decomposition and information entropy, *Engineering Transactions*, **70**(1): 23-51, 2022, doi: 10.24423/EngTrans.1390.20220207.

[9]MOSTAFA R., HAMED A., Dispersion Entropy: A Measure for Time-Series Analysis, *IEEE Signal Processing Letters*, **23**(5): 610-614, 2016, doi: 10.1109/LSP.2016.2542881.

[10]Li Y X., Wu J X., Zhang S., Tang B Z., Lou Y L., Variable-Step Multiscale Fuzzy Dispersion Entropy: A Novel Metric for Signal Analysis, *Entropy*, **25**(7): 997, 2023, doi: 10.3390/e25070997.

[11]HAMED A., MOSTAFA R., DANIEL A., JAVIER E., Refined composite multiscale dispersion entropy and its application to biomedical signals, *IEEE transactions on bio-medical engineering*, **64**(12): 2872-2879, 2017, doi: 10.1109/TBME.2017.2679136.

[12]Lv J., Sun W L., Wang H W., Zhang F., Coordinated approach fusing RCMDE and sparrow search algorithm-based SVM for fault diagnosis of rolling bearings, *Sensors (Basel, Switzerland)*, **21**(16): 5297, 2021, doi: 10.3390/s21165297.

[13]Su N Q., Xiong J B., Zhang Q H., Huang C L., Research methods of the rotating machinery fault diagnosis [in Chinese], *Machine Tool & Hydraulics*, **46**(7):133-139, 2018, https://ss.zhizhen.com/detail_38502727e7500f26653bf4977b61f2a6a0ea1c615f8cb2cb1921b0a3ea255101fc1cf1fbb4666ae6e68055c50ae2c7314e1390bd32f089133359d92948e2f39d091fdadd9491d269db8cf161021442d1.

[14]Xia X., Wang X L., Fault diagnosis of planetary gearbox based on hierarchical refined composite multiscale fuzzy entropy and optimized LSSVM, *Entropy (Basel, Switzerland)*, **27**(5): 512, 2025, doi: 10.3390/e27050512.

[15]Wang H M., Chan L., Wavelet packet transform-assisted least squares support vector machine for gear wear degree diagnosis, *Mathematical Problems in Engineering*, **2021**(39): 1-9, 2021, doi: 10.1155/2021/9889933.

[16]Liang C., Zhou Y L., Yang N., Cavitation fault diagnosis of centrifugal pump based on wavelet multi-scale chaotic characteristic parameters, *Journal of Vibration and Shock*, **40**(21):106-112,134, 2021, doi: 10.13465/j.cnki.jvs.2021.21.015.

[17] Wu Z Z., Song F., Whale optimization algorithm based on improved spiral update position model, *Systems Engineering-Theory & Practice*, **39**:2928-2944, 2019, doi: 10.12011/1000-6788-2018-2156-17.

[18] Zheng Y., Hu J Z., Jia M P., Xu F Y., Tong Q J., A method for rolling bearing fault feature extraction based on parametric optimization VMD, *Journal of Vibration and Shock*, **39**(21):195-202, 2019, doi: 10.13465/j.cnki.jvs.2020.21.026.

[19] Hu Q., Zhou H T., Wang C C., Zhou C X., Shen J P., He P., Time-frequency fusion features-based GSWOA-KELM model for gear fault diagnosis, *Lubricants*, **12**(1): 10, 2024, doi: 10.3390/lubricants12010010

[20] Sun J H., Analysis and diagnosis of rail traffic motor bearing based on adaptive multi-scale dispersion entropy, *Machine Design & Research*, **36**(6), 2020, doi:10.13952/j.cnki.jofmdr.2020.0243.

[21] Li C Z., Zhen J D., Pan H Y., Liu Q Y., Fault diagnosis method of rolling bearings based on refined compositeMultiscale dispersion entropy and support vector machine. *China Mechanical Engineering*, **30**(14): 1713-1719, 2019, doi: 10.3969/j.issn.1004-132X.2019.014.012.

[22] Yin Z., Hou J., Recent advances on SVM-based fault diagnosis and process monitoring in complicated indu strial processes, *Neurocomputing*, 174:643-650, 2016,doi: 10.1016/j.neucom.2015.09.081.

[23] Tian A Q., Liu F F., Lv H X., Snow Geese Algorithm: A novel migration-inspired meta-heuristic algorithm for constrained engineering optimization problems, *APPLIED MATHEMATICAL MODELLING*, **126**: 327-347, 2024, doi: 10.1016/j.apm.2023.10.045.

[24] Zhang W., Zhang Z Z., Yao Q., Zhang X., Liu D., Liu Z S., Meng H M., Qin W M., Fault diagnosis of wind turbine generator system based on PMI-LSSVM, *Journal of Physics: Conference Series*, 2095(1): 012009, 2021, doi: 10.1088/1742-6596/2095/1/012009.