

Research Article

Fault Prediction of Centrifugal Pump Based on Improved KNN

YunFei Chen , Jianping Yuan, Yin Luo , and Wenqi Zhang

Research Center of Fluid Machinery Engineering and Technology, Jiangsu University, Zhenjiang 212000, China

Correspondence should be addressed to Yin Luo; luoyin@ujs.edu.cn

Received 31 July 2021; Revised 30 September 2021; Accepted 4 October 2021; Published 21 October 2021

Academic Editor: Antonio Giuffrida

Copyright © 2021 YunFei Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To effectively predict the faults of centrifugal pumps, the idea of machine learning k-nearest neighbor algorithm (KNN) was introduced into the traditional Mahalanobis distance fault discrimination, and an improved centrifugal pump fault prediction model of KNN based on the Mahalanobis distance is proposed. In this method, the Mahalanobis distance is used to replace the distance function in the conventional KNN algorithm. Grid search and cross-validation are used to determine the optimal K value of the prediction model. A centrifugal pump test rig was established to solve three common faults of centrifugal pumps: cavitation, impeller damage, and machine seal damage, and the method was verified. The results show that this method can effectively distinguish the specific fault types of centrifugal pumps based on vibration signals, and the fault prediction accuracy of the off-balance condition is up to 82%. This study provides a novel idea and method for centrifugal pump fault prediction and diagnosis and avoids the interaction between parameters when monitoring multiple parameters.

1. Introduction

Centrifugal pump is an important fluid conveying equipment; it is widely used in various fields and plays a significant role in the development of the national economy. Therefore, it is particularly necessary to ensure the normal and stable operation of the centrifugal pump.

In the process of centrifugal pump operation, the earliest indication of failure is usually the abnormal vibration signal. Some minor equipment damage or defects will cause abnormal vibration of centrifugal pumps, such as seal damage, cavitation, and impeller damage. Although the centrifugal pump can transmit information to the outside world through acoustic emission [1], temperature, oil pressure, pressure pulsation, and other signals, all parameter monitoring methods have their limitations. While vibration signals can timely and accurately transmit the operating state information of the centrifugal pump and can be monitored in a long-term and stable manner. Therefore, vibration signal monitoring has become the most commonly used method in centrifugal pump fault diagnosis. Approaches to extract features from mechanical systems based on time-domain data were proposed in [2, 3]. The vibration signal of the centrifugal pump can be analyzed and the fault of the

centrifugal pump can be performed. In the centrifugal pump monitoring, in addition to time-domain parameters, there are frequency-domain parameters. Xue et al. [4] proposed a fault prediction system. The vibration signal of a centrifugal pump is analyzed in the amplitude domain and time domain, and the characteristic structure of the signal in the frequency domain is analyzed by the fast Fourier transform (FFT) signal analysis method. The failure of the centrifugal pump was analyzed based on the database of typical vibration failure cases.

With the continuous development of computer algorithms and artificial intelligence, clustering algorithms, neural network, and other technologies have a lot of positive impact on the traditional fault monitoring methods, making fault monitoring more accurate and intelligent [5–10]. Jun Du et al. [11] proposed a clustering diagnosis method based on statistical average relative power difference (ARPD) for the faults of the aircraft hydraulic pump, such as swashplate eccentricity and the increase of the gap between piston and slider. By effectively enhancing the fault characteristics of these two kinds of faults, the ARPD calculated from vibration signals is used to complete the hypothesis testing. To extract the weak signal fault characteristics of aeroengine intermediate shaft bearing effectively, Jing et al. [12]

introduced a tolerance idea into the traditional adaptive genetic algorithm and proposed a variational mode decomposition (VMD) method based on TAGA-VMD. Machine learning and neural network methods have also become a research hotspot recently [13–18]. Tai-Ming Tsai and Wei-Hui Wang [19] addressed dealing with these signals to establish the database of input-output relations by using several neural network models through learning algorithms. Meanwhile, to set up an online diagnosis network, the learning speed, and accuracy of three kinds of networks, the backpropagation (BNP), radial basis function (RBF), and adaptive linear (ADALINE) neural networks have been compared and assessed. For online diagnosis, the BPN method is recommended. Yang et al. [20] proposed a fault diagnosis scheme for rotating machinery using hierarchical symbolic analysis and a convolutional neural network. The method performs superior diagnosis capacity with simple network architecture. Kong and Chen [21] proposed a new combined method based on wavelet transformation, fuzzy logic, and neuronetworks for fault diagnosis of a triplex. The failure characteristics of the fluid and dynamics can be divided by the wavelet transform in different scales at the same time. Therefore, the characteristic variables can be constructed making use of the coefficients of the Edgeworth asymptotic spectrum expansion formula and fuzzified to train the neuronetwork to identify the faults of the fluid and dynamic of the triplex pump in the fuzzy domain.

In conclusion, there are many parameters to monitor the operation of the centrifugal pump, which makes it difficult to have a unified judgment standard for fault diagnosis of the centrifugal pump. This paper proposes a centrifugal pump fault monitoring system based on an improved KNN algorithm [22] based on the Mahalanobis distance. Firstly, the ReliefF algorithm [23] was utilized to carry out weight analysis on the time-domain and frequency-domain features [24] and parameters commonly used in centrifugal pump monitoring. Then, the parameters with greater influence on centrifugal pump fault monitoring were selected. The characteristic space was developed by using the characteristic parameters under different fault conditions and normal conditions. The KNN algorithm was improved and the centrifugal pump working conditions were predicted by using Mahalanobis distance [25, 26] instead of Euclidean distance. Finally, the proper K value is selected by grid search [27–29] to establish the centrifugal pump fault prediction system, so that the operation of the centrifugal pump can be evaluated.

2. Materials and Methods

The improved KNN prediction model based on the Mahalanobis distance is given in Figure 1, which is mainly divided into two parts: feature engineering and prediction model.

3. Feature Engineering Based on ReliefF Algorithm

There are several parameter indexes to evaluate the operation condition of the centrifugal pump; if all parameters

are selected at the same time, it will be difficult to judge the operating condition of the centrifugal pump because of the mutual influence between the parameters, and the operation fault cannot be accurately predicted. The ReliefF algorithm can choose the top four parameters indexes with the highest contribution. The commonly used time-domain and frequency-domain indexes are divided into absolute indexes and relative indexes. Absolute indicators are dimensionless eigenvalues and have units, such as peak value, mean value, and root mean square. Relative indexes are dimensionless eigenvalues without units, such as kurtosis and margin.

3.1. Relief Algorithm. The Relief algorithm was first proposed by Kira and was initially restricted to the classification of two types of data. Later, with the progressing of the Relief algorithm, it became a feature weight algorithm. To address its limitations, Kononeill expanded it in 1994 and obtained the ReliefF algorithm, which can give different weights to features according to the correlation of each feature and category. ReliefF algorithm randomly selects a sample R from the training set D and then looks for k -nearest neighbor samples from the samples comparable to R , which is called Near Hits. Looking for the nearest neighbor samples that are different from R is called Near Misses and then update the weight of each feature according to the following rules, as shown in

$$W(A) = W(A) - \sum_{j=1}^k \frac{\text{diff}(A, R, M_j)}{(mk)} + \sum_{c \neq \text{class}} \frac{[P(C)/1 - P(\text{class}(R)) \sum_{j=1}^k \text{diff}(A, R, P_j(C))]}{(mk)}. \quad (1)$$

In the above formula, M_j is the J_{th} sample of the same kind, P_j is the J_{th} adjacent sample in the different class C , m is the sampling number, $\text{diff}(A, R, R_2)$ represents the difference between samples R_1 and R_2 in the feature space A , and its calculation formula are shown as follows:

$$\text{diff}(A, R, R_2) = \begin{cases} \frac{|R_1[A] - R_2[A]|}{\max(A) - \min(A)} & \text{if } A \text{ is continuous,} \\ 0 & \text{if } A \text{ is discrete and } R_1[A] = R_2[A], \\ 1 & \text{if } A \text{ is discrete and } R_1[A] \neq R_2[A]. \end{cases} \quad (2)$$

ReliefF function is called repeatedly for n times. Finally, a better weight of each feature index can be obtained by averaging the results of n pairs of calculations.

3.2. Centrifugal Pump Fault Working Characteristic Index. The flow rate of the test centrifugal pump under standard working conditions is $Q_d = 50\text{m}^3/\text{h}$. Axial vibration signals

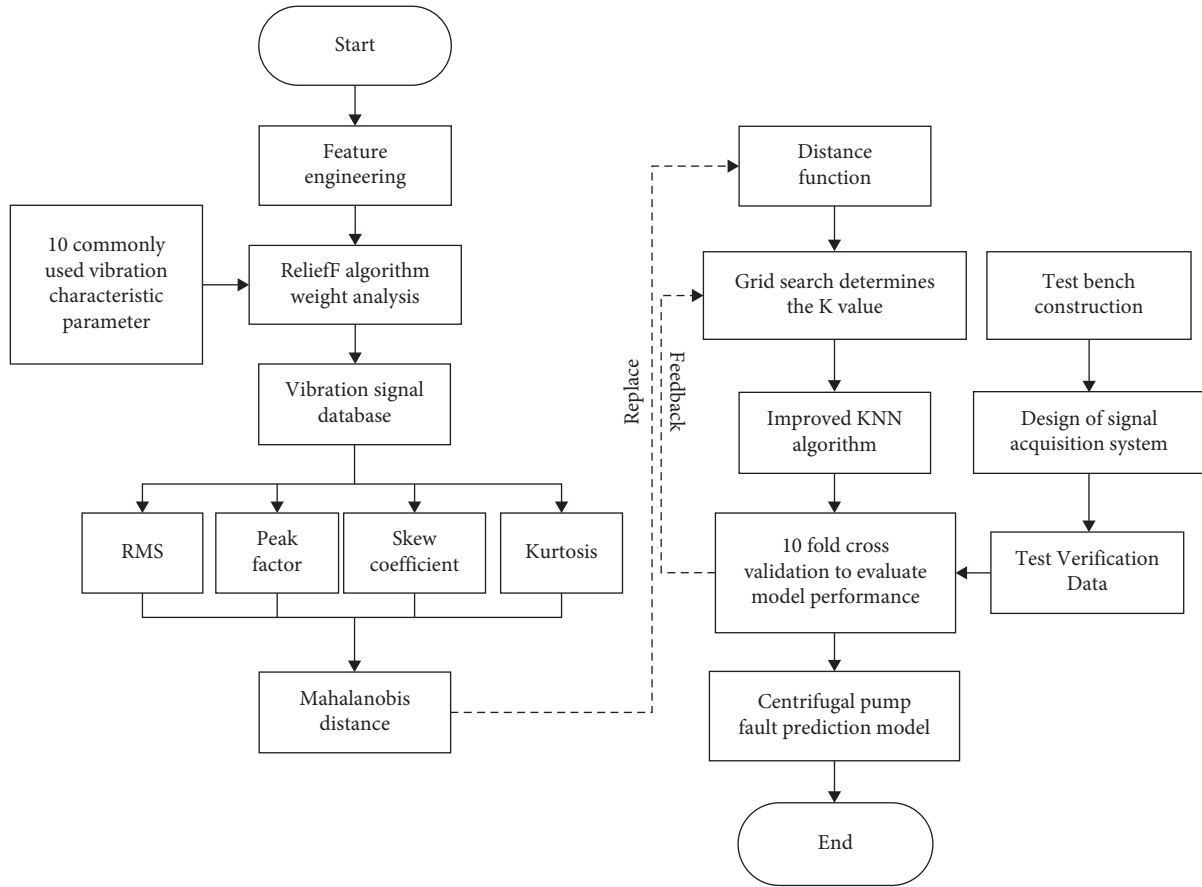


FIGURE 1: Flowchart of the prediction model.

under standard working conditions are collected. Weight analysis of the following 10 characteristic parameters [9, 10] in the time domain and frequency domain is carried out according to the Relief algorithm, as shown in Table 1.

After the operation, the Relief algorithm was used to filter 10 characteristic parameters. The weight of each feature parameter is calculated by *Python*. The larger the weight is, the greater the contribution of the feature parameter to the correct classification of the data is. The characteristic parameters are more reliable for centrifugal pump monitoring. Weights of the 10 characteristic indicators are obtained as shown in Table 2, and the histogram of descending order is drawn as shown in Figure 2.

According to the analysis results of the Relief algorithm, the four parameters that contribute the most to the centrifugal pump fault diagnosis classification are root mean square, peak factor, skewness coefficient, and kurtosis. To reduce the dimension and data redundancy of features and facilitate the training of KNN, in the next step, this paper takes the four feature indexes with the largest weight as the eigenvalues of the dataset. The weight of the other six parameters is very small; not dealing with them will not affect the result of fault prediction.

4. Improved Centrifugal Pump Fault Prediction Model of KNN Based on the Mahalanobis Distance

Due to its simplicity and effectiveness, the KNN algorithm is usually the first choice to solve any classification problem. In essence, fault monitoring can be regarded as classifying the collected operation status signals. However, two factors can degrade KNN performance. The key of the KNN algorithm is the selection of K values and the calculation of the distance between samples. First, KNN determines the similarity between two samples using a distance function. Second, the accuracy is sensitive to the neighborhood size K value.

4.1. Mahalanobis Distance. In general, the KNN algorithm uses the Euclidean distance to calculate the sample distance, but the Euclidean distance has an obvious disadvantage, which equates to the differences between different attributes of the sample. Because of the correlation between discrete monitoring parameters of centrifugal pumps, Euclidean distance cannot meet the practical requirements of centrifugal pump fault prediction.

TABLE 1: Time-domain feature.

Evaluation index	Formula	Significance of evaluation
RMS	$X_{rms} = \sqrt{1/N \sum_{i=1}^N x_i^2}$	Represents the average amount of signal amplitude based on a unit of time
Kurtosis coefficient	$C_K = 1/N \sum_{i=1}^N (x_i - \bar{x})^4 / \sigma^4 - 3$	It can express the difference between the signal and its average value at this time
Peak factor	$C = X_{peak} / X_{rms}$	Dynamic expression of the damage to the surface of the machine seal
Pulse factor	$I = X_{peak} / \bar{x}$	The ratio of the peak value to the mean value of the signal, the criterion for judging the presence of an impact in the signal
Waveform factor	$S = X_{rms} / \bar{x}$	Based on the upper root mean square value and the mean value ratio of the time-domain
Kurtosis	$K_v = (1/N) \sum_{i=1}^N (x_i - \bar{x})^4 / ((1/N) \sum_{i=1}^N x_i^2)^2$	The amplitude of the seal acoustic emission signal is described. If the kurtosis is too large, it means that there is a dry friction
Margin coefficient	$df = x_{max} / ((1/N) \sum_{i=1}^N \sqrt{ x_i })^2$	The ratio of the signal root means a square value to its mean value, used to evaluate the degree of component damage
Skew coefficient	$C_s = 1/N \sum_{i=1}^N (x_i - \bar{x})^3 / \sigma^3$	The asymmetry degree of acoustic emission signal is described
Center frequency	$FC = \sum_{i=1}^N f_i p_i / \sum_{i=1}^N p_i$	Denotes the change in the position of the power spectrum barycenter and describes the change of the energy ratio of each frequency component
MSF	$MSF = \sum_{i=1}^N f_i^2 p_i / \sum_{i=1}^N p_i$	Represents the discrete state of power spectrum energy distribution and describes the composition of signal frequency components

TABLE 2: Table of characteristic parameter weights.

Feature	Weight	Feature	Weight
RMS	0.3274	MSF	0.0990
Peak factor	0.2869	Kurtosis coefficient	0.0264
Skew coefficient	0.2331	Center frequency	0.0202
Kurtosis	0.1235	Waveform factor	0.0043
Margin coefficient	0.1008	Pulse factor	0.0021

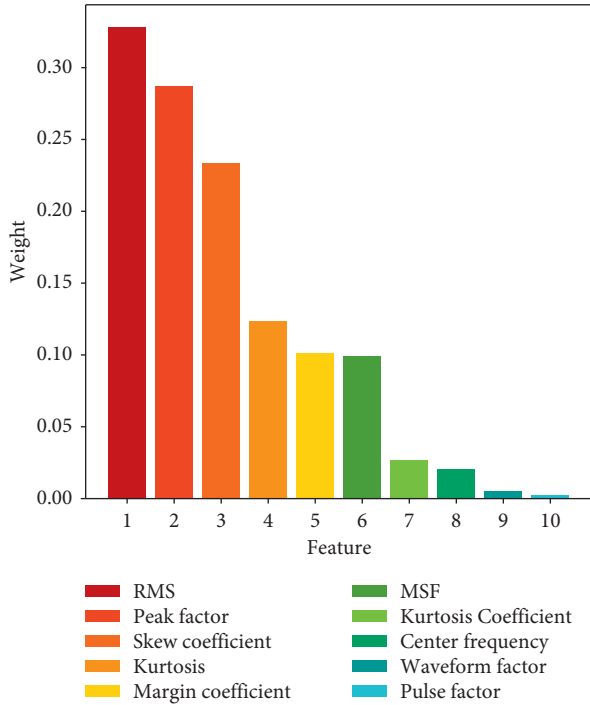


FIGURE 2: Diagram of characteristic parameter weight.

The Mahalanobis distance was suggested by P.C. Mahalanobis, an Indian statistician, which represents the distance between a point and a distribution. It is an effective method to calculate the similarity between two unidentified sample sets. Unlike Euclidean distance, it takes into consideration the relationship between various properties and is dimensionless. For an average of $\mu = (\mu_1, \mu_2, \mu_3, \dots, \mu_p)^T$, the covariance matrix S for several variables $x = (x_1, x_2, x_3, \dots, x_p)^T$, the Mahalanobis distance formula is shown in

$$D_M(x) = \sqrt{(x - \mu)^T S^{-1} (x - \mu)}. \quad (3)$$

To sum up, the Mahalanobis distance can easily measure the distance between the observed samples and the known sample set, so it is very suitable for the field of fault diagnosis. The improved KNN algorithm can be obtained by replacing the original Euclidean distance function of the KNN algorithm with the Mahalanobis distance.

4.2. Data Analysis. According to the weight analysis results of the Relief algorithm, four parameters, that is, root mean square, peak factor, skewness coefficient, and kurtosis, were statistically analyzed. We collect the vibration signal data of centrifugal pump impeller damage, centrifugal pump cavitation, centrifugal pump seal damage, and centrifugal pump under normal operation. Then, we calculate the root mean square, peak factor, skewness coefficient, and kurtosis under each operating condition. Due to the dimensionality and value range of these four characteristic parameters, the consequences are often not good if the algorithm is directly applied to these data, so the data should be scaled in the same proportion to make it fall within a specific interval. Standard deviation standardization, otherwise known as

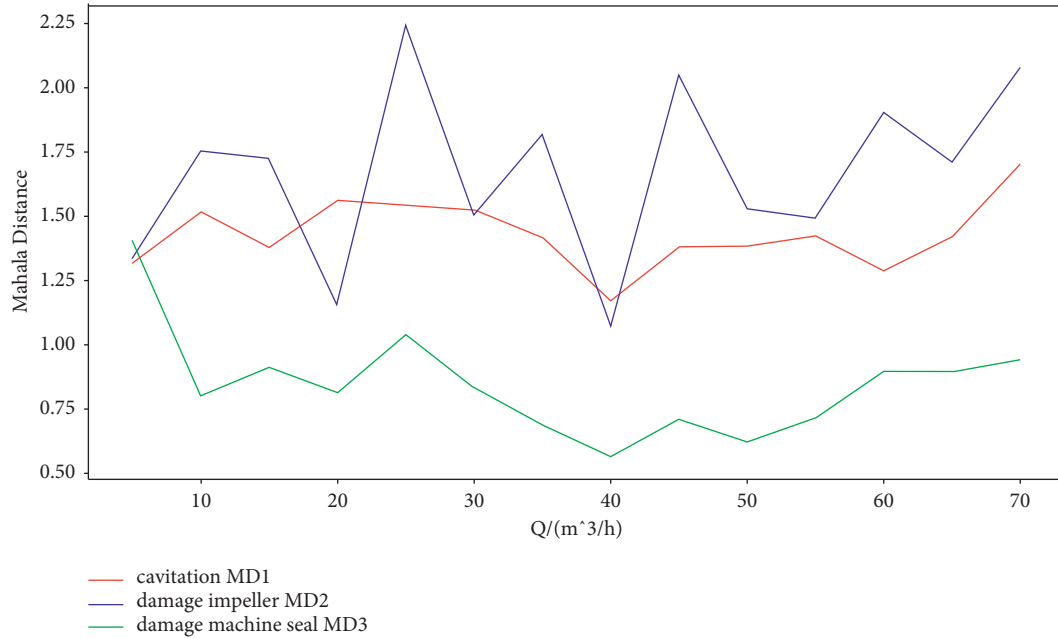


FIGURE 3: Mahalanobis distance diagram under different working conditions.

zero-mean standardization or score standardization, is the most extensive data standardization method at present. After standard deviation normalization, the standard deviation of the data is 1, the mean value is 0, and the tangent is under the normal distribution. The conversion formula is presented in

$$X^* = \frac{X - \bar{X}}{\sigma}, \quad (4)$$

where X is the mean of the original data and σ is the standard deviation of the original data. Standard deviation normalization is used for root mean square, peak factor, skewness coefficient, and kurtosis before calculating the Mahalanobis distance.

After data preparation, the Mahalanobis distances between the samples under normal operation of the centrifugal pump and the samples in the case of cavitation, the sample with damaged impeller, and the sample with damaged seal were calculated, which were demonized as MD_1 , MD_2 , and MD_3 , respectively. Figure 3 shows the Mahalanobis distances under different centrifugal pump fault conditions after the addition of the four characteristic parameters according to the ReliefF weights and Table 3 shows the Mahalanobis distances under different centrifugal pump fault conditions after the addition of the four characteristic parameters according to ReliefF weights.

By Mahalanobis distance diagnosis, it was found that, in the whole test, the maximum value of MD_2 was found, and MD_2 , MD_1 , and MD_3 were decreased in turn. Centrifugal pumps operate at a low flow rate or operate over the design point; with the increase of flow, the Mahalanobis distance between the centrifugal pump failure data and normal data at runtime is growing. The Mahalanobis distance reached 2.244143, indicating that the impeller damage is very serious;

TABLE 3: Partial Mahalanobis distance data.

$Q(m^3/h)$	$MD1$	$MD2$	$MD3$
5	1.315011	1.333311	1.402389
10	1.514888	1.753904	0.798257
15	1.377642	1.724976	0.909592
20	1.560997	1.150543	0.812094
25	1.544954	2.244143	1.038043
30	1.522004	1.501566	0.830493
35	1.415861	1.818289	0.685511
40	1.170261	1.068531	0.559391
45	1.379778	2.051609	0.706578
50	1.382423	1.526468	0.618459
55	1.422706	1.489968	0.714366
60	1.285552	1.903851	0.895176
65	1.418641	1.708746	0.892477
70	1.702013	2.077338	0.939009

in this condition, the centrifugal pump operating condition is very poor. According to the analysis of the Mahalanobis distance, it can be seen that using the Mahalanobis distance can effectively judge the operation fault of the centrifugal pump, and it is appropriate and effective in improving the KNN algorithm.

4.3. Improved KNN Algorithm. KNN (K-nearest neighbor) method, originally proposed by Cover and Hart in 1968, is a machine learning algorithm. The idea of this method is very intuitive: if most of the K-nearest likeness (that is, the nearest neighbor of the feature space) samples of a sample in the feature space belong to a certain category, then the sample also belongs to this category. In the classification decision, the method only determines the category of the samples to be distributed according to the category of the nearest one or several samples.

The selection of distance function and the determination of K value are the main factors that affect the performance and result of the KNN algorithm. The Mahalanobis distance has proved through data analysis that it can effectively replace Euclidean distance, and the effect is good. For KNN algorithm parameter K , if you choose smaller values of K , it is equivalent to use smaller training instances in the field of the forecast, reducing the approximation learning error. Only training instances that are close to the input instances are useful for the predicted results, but the downside is the fact that the study of estimation error will increase; the result is sensitive to the neighbor instance of the point. If the adjacent instance points happen to be noisy, the prediction will be wrong. This means that as K values decrease, the model becomes more complex and easier to overfit. If a larger value of K is selected, it will be tantamount to using training examples in a larger field to make predictions. The advantage is the case that the estimation error of learning can be reduced, but the approximate error will increase. In other words, with the increase of K value, the prediction accuracy of input instances will be reduced, and the overall model will become simple. The approximate error can be understood as the training error of the existing training set. A small approximate error can make a better prediction of the existing training set, but its performance on the unknown test sample set is poor. The estimation error can be seen as the test error of the test set. A small estimation error indicates a satisfactory ability to predict the unknown data, and the model at this time is closer to the optimal model.

As the observation data, based on the data under normal working conditions, with marker distance as the radius, cavitation, and impeller damaged, we draw machine seal damage case sample points. As there are too many sample points, only 100 of them are drawn, as shown in Figure 4. Because the sample points are too concentrated, the normal working condition points can be seen when the central area is enlarged, as shown in Figure 5. Taking the yellow asterisk points in Figure 4 as an example, the schematic diagram of $K=1$, $K=10$, and $K=100$ is drawn. It can be seen that if the K value is too small, the sample points near the prediction point are too small, and it is easy to misjudge the running status of the centrifugal pump. When the K value is too large, there are too many sample points in the range, which will also increase the difficulty of centrifugal pump operating condition prediction.

To determine the appropriate K value, the automatic parameter tuning method is approved in this paper. Grid Research is the most widely used automated parameter optimization method at present. A grid search evaluates all possible combinations of parameter values to calculate the best combination, using cross-validation to evaluate the merits of K values. Starting with a small value of K , increasing the value of K , and then calculating the variance of the result, we can finally determine a suitable value of K . In this paper, K values were selected from 400, 800, 1600, 1800, 2000, 2200, 3000, and 5000. To make a better evaluation of the model performance, we measure the accuracy of the evaluation and the standard deviation from the actual results at the same time and avoid the occurrence of overfitting; the K -fold cross-validation method can be used. In this paper, 10-fold cross-validation was used. As shown in

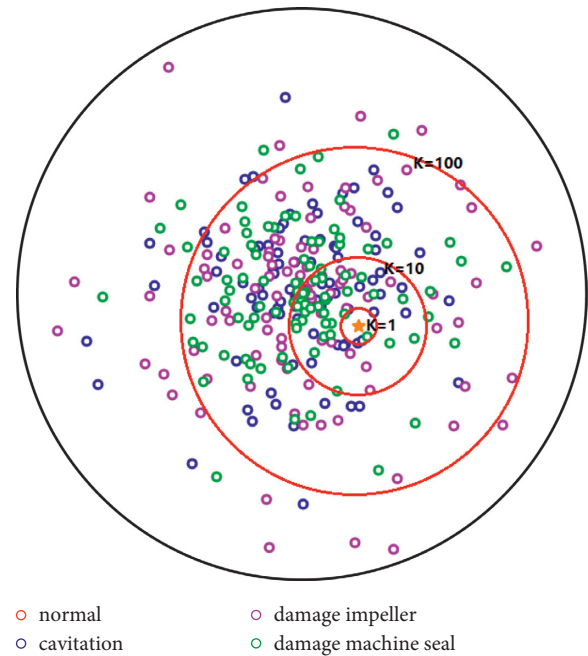


FIGURE 4: Schematic diagram of sample point distribution and K values.

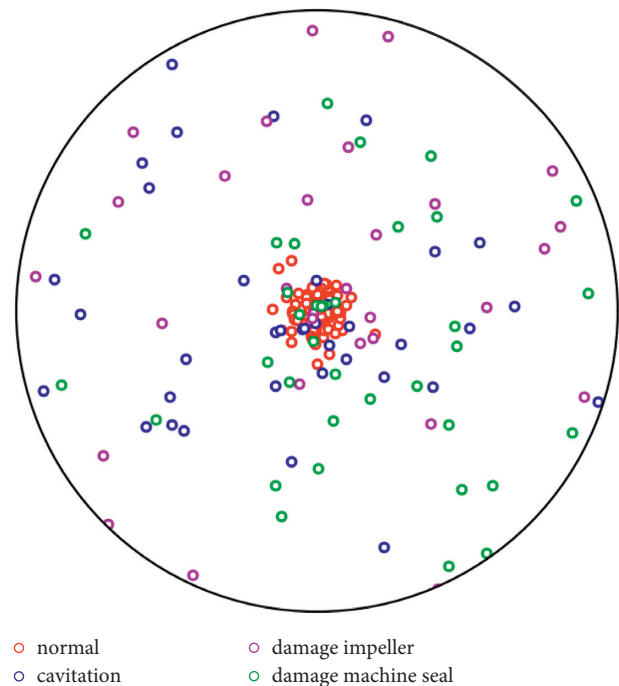


FIGURE 5: Schematic diagram of sample point distribution.

Table 4, the dataset is divided randomly into 10 different subsets, each subset is called a fold, and then, the model is trained and evaluated 10 times, onefold is selected for evaluation each time, and the other nine folds are used for training. The output result is an array containing 10 evaluation scores, and the average value of this array is the evaluation of the model performance. Meanwhile, the robustness of the model can be verified by observing the evaluation score of each training.

TABLE 4: Grid search cross-verifies the prediction accuracy in each case.

K value	400	800	1600	1800	2000	2200	3000	5000
Set 1	0.6183	0.6185	0.8183	0.8190	0.9187	0.8182	0.7186	0.6164
Set 2	0.7054	0.7049	0.9073	0.9072	0.9072	0.9071	0.8064	0.7058
Set 3	0.7289	0.7288	0.9289	0.9291	0.9283	0.8290	0.8287	0.7294
Set 4	0.7071	0.7097	0.9086	0.9088	0.9094	0.9095	0.8089	0.7091
Set 5	0.5534	0.5529	0.7540	0.7535	0.8537	0.7531	0.6536	0.5516
Set 6	0.6399	0.6385	0.8397	0.8398	0.8398	0.8401	0.7403	0.6399
Set 7	0.7436	0.7437	0.9442	0.9439	0.9440	0.9440	0.8439	0.7438
Set 8	0.5745	0.5755	0.7748	0.7759	0.8759	0.8754	0.6753	0.5749
Set 9	0.4991	0.5009	0.7017	0.7027	0.8025	0.7020	0.6026	0.5033
Set 10	0.5013	0.5025	0.7041	0.7038	0.9044	0.7040	0.6042	0.5066
Mean	0.6271	0.6276	0.8282	0.8283	0.8884	0.8283	0.7282	0.6278
Instability	0.0878	0.0875	0.0873	0.0872	0.0871	0.0874	0.0871	0.0868

It can be observed that when the value of K is increased, the accuracy of fault prediction will increase first, because there are more samples around for reference, and the prediction effect will become better. With the increase of K value, the accuracy will have a maximum value of 0.8884. At this time, the K value of 2000 is the optimal K value, because when there is a further increase of K value, the error rate of prediction will gradually rise, and it will be meaningless to continue to increase the K value.

5. Experimental Verification

In a check to see the validity of the prediction model in fault prediction, a centrifugal pump testbed was built. Conditions of $0.8 Q_d$, $1.0 Q_d$, and $1.2 Q_d$ were selected as the test conditions, which included the centrifugal pump external characteristic test under normal working conditions, cavitation test, machine seal failure test, and impeller damage test. The vibration signals of the centrifugal pump under various fault conditions were collected and processed by the signal acquisition system, and the weight analysis results of the ReliefF algorithm were combined with four characteristic parameters, namely, root mean square, peak factor, skewness coefficient, and kurtosis. The fault prediction model of the KNN centrifugal pump improved with the Mahalanobis distance was used to predict the centrifugal pump working conditions, and the results were compared with the test results.

5.1. Construction and Design of Centrifugal Pump Testbed.

The centrifugal pump test rig built is shown in the schematic diagram and the physical drawing as shown in Figure 6. IS-65-50-160 centrifugal pump, export solenoid valve, driving motor, import manual ball valve, import and export stainless steel water pipe, bellows, storage tank, and so on together constitute the water circulation system. The collective design parameters of the IS-65-50-160 centrifugal pump are given in Table 5.

5.2. Design of Signal Acquisition System. The signal acquisition system is composed of the equipment responsible for collecting the various parameters of the centrifugal pump. The signal acquisition system is mainly composed of the inlet

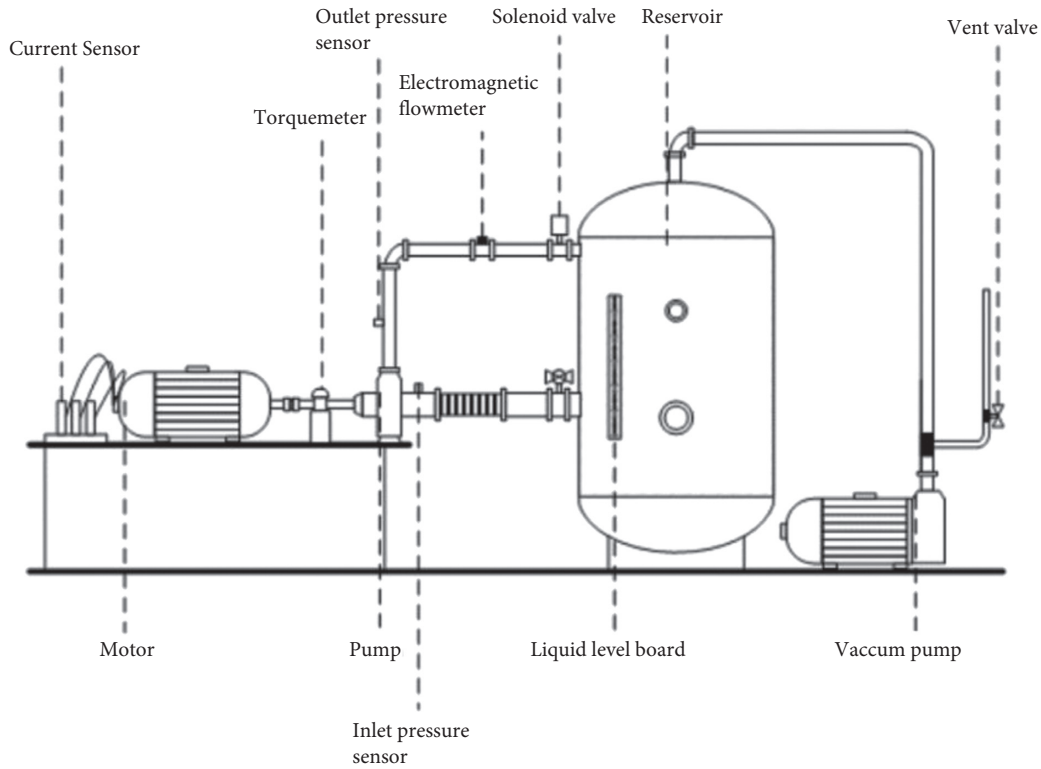
and outlet pressure sensor, an electronic flow meter, NI signal acquisition card, transient speed, torque tester, and resistance. The precise parameters of the instrument and sensor are shown in Table 6.

5.3. Centrifugal Pump Characteristic Test Procedure.

Offset conditions and cavitation are the most common unstable conditions in the operation of the pump. Tests were carried out to diagnose cavitation and deviation conditions. The experimental steps of centrifugal pump external characteristics are as follows:

- (1) Open the inlet pipeline, ball valve to maximum; open the storage tank vent valve; let the internal pressure of the storage tank be close to the atmosphere.
- (2) Open the centrifugal pump unit and check the water circulation system to avoid the leakage of the system affecting the subsequent test; check the stability of the data acquisition card, and check each sensor one by one.
- (3) The flow of the centrifugal pump under standard working conditions was set at Q_d , and the actual pump flow was regarded as an independent variable. The abscissa of external characteristics was divided into 14 working conditions, and the working conditions from 0 to $1.3 Q_d$ were measured, respectively.
- (4) Set the specific parameters of the centrifugal pump, by controlling the solenoid valve to adjust the specified flow rate of the centrifugal pump, so that the centrifugal pump runs under the specified flow parameters. When the centrifugal pump runs in this working condition and is stable, there is synchronous collection and recording of several parameters such as flow rate, input current, inlet and outlet pressure, speed, and torque of the centrifugal pump.
- (5) Repeat the test until 14 working conditions are recorded and the test is completed. Then, calculate the efficiency, head, and shaft power under the corresponding working conditions, and finally, draw the external characteristic curve.

The implementation steps of the cavitation characteristic test of a centrifugal pump are as follows:



(a)



(b)

FIGURE 6: (a) Schematic diagram. (b) Physical drawing of the testbed.

TABLE 5: Design parameters of centrifugal pump for test.

Parameter	Unit	Value
Impeller inlet diameter	D_1 (mm)	74
Impeller outlet diameter	D_2 (mm)	174
Blade outlet width	b_2 (mm)	12
Number of blades	Z	6
Rated flow	Q_d (m^3/h)	50
Rated head	H_d (m)	34
Rated speed	n (r/min)	2980
Rated power	H (%)	72.8
Specific speed	n_s	89.51

- (1) Open the valve connecting the water storage tank with the atmosphere, so that the internal pressure of the water storage tank is on the brink of the atmospheric pressure, and then close the water storage tank vent valve.
- (2) Check the running state of the water circulation system, sensor equipment, and data acquisition card, start the centrifugal pump unit, and simulate the test steps of full flow bias condition for operation.
- (3) Adjust the analog voltage of the data acquisition card so that the centrifugal pump can run at the specified flow rate.
- (4) Open the vacuum pump to reduce the inlet pressure.
- (5) When the head drops by 0.05 m, close the vacuum pump. When the measured data tend to be stable, turn on the vacuum pump and wait for the next round of data collection, waiting for a 10% drop in head to complete the collection.
- (6) Adjust the flow rate so that the centrifugal pump works in each flow rate, respectively. Repeat the above steps to obtain the cavitation characteristic data of the centrifugal pump in each flow rate.

TABLE 6: Specific parameters of each instrument and sensor.

Instrument name	Parameter	Value (unit)
NI USB-signal data acquisition card	Analog input	4-way, (kS/s), 16-bit resolution
	Analog output	4-way, (kS/s), 16-bit resolution
	Voltage range	-10(V)~+10 (V)
	Digital I/O interface	4-way
LDW-SIN-DW-Z electromagnetic fluidizer	MWP	1.6 (MPa)
	Supply voltage	220 (V)
	Measuring flow range	0~120 (m ³ /h)
	Error	0.16 (%)
SGDN-dynamic torque measuring device	Output current signal	4-20 (mA)
	Output frequency	5~15 (kHz)
	Error	0.3%
WIKAS-transducer	Measuring flow range	0-120 (N*m)
	Measuring range	Inlet 0~1.6 (bar abs pressure)/outlet 1~4 (bar)
	Current output	4-20 (mA)
	Mains input	10~30 (V)
	Error	0.2%

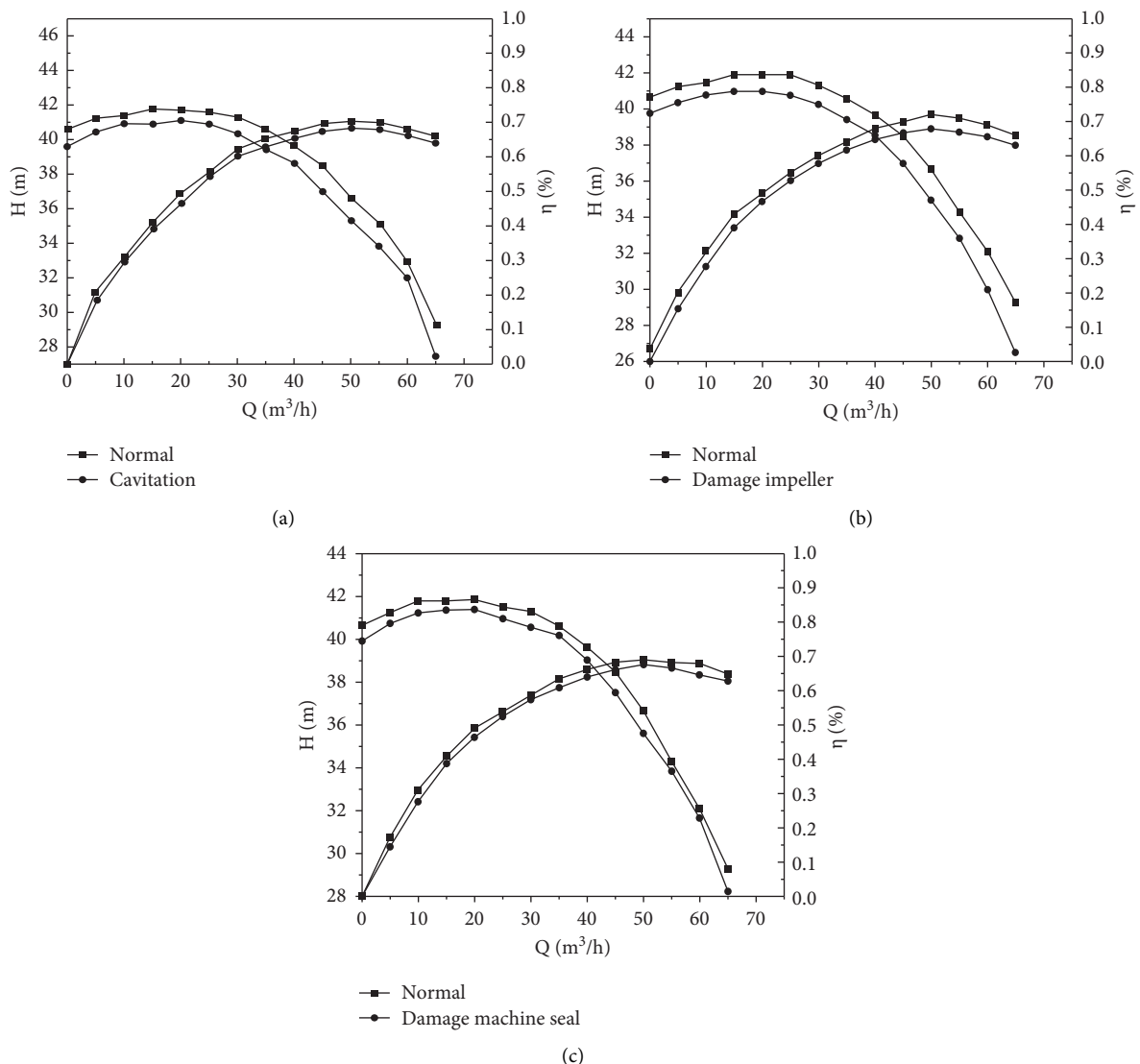


FIGURE 7: (a) External characteristics of cavitation and normal working conditions. (b) External characteristics of damage impeller and normal working conditions. (c) External characteristics of damage machine seal and normal working conditions.

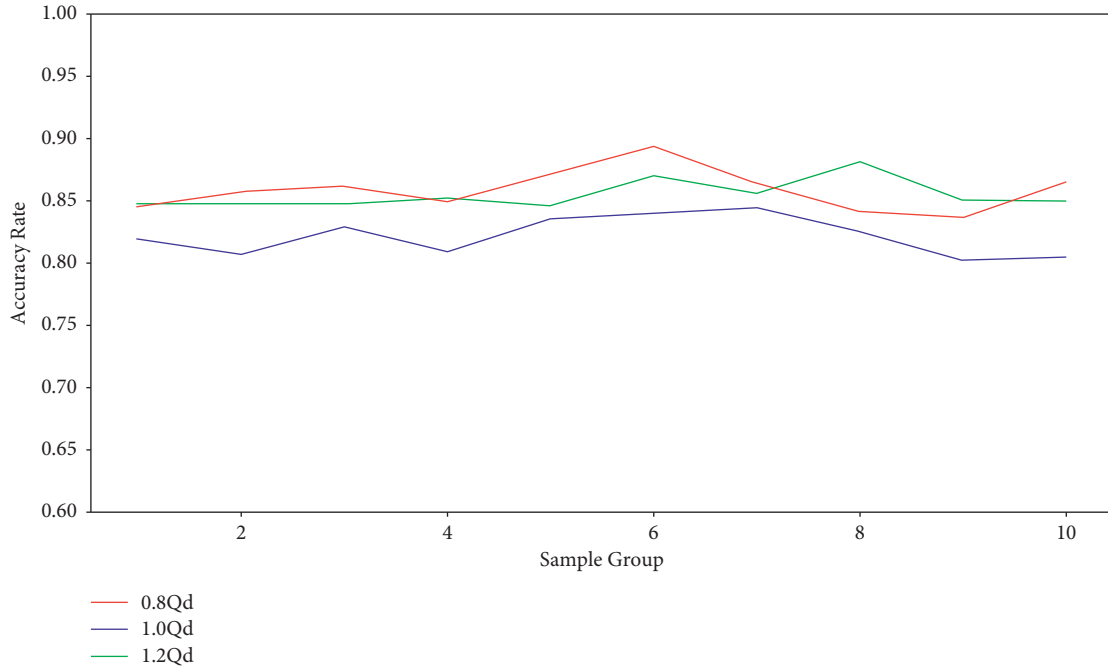


FIGURE 8: Model accuracy of group verification under various working conditions.

Use normal impeller, damaged impeller, and damaged machine seal to carry out the above tests, respectively. According to the experimental results, the external characteristic curves of the centrifugal pump under cavitation and normal conditions are drawn, as shown in Figure 7.

In Figure 7(a), by comparing the external characteristics of the pump under the two conditions, it is found that the cavitation phenomenon of the centrifugal pump is serious, and the damage of the inlet blade will produce a large number of cavities, and the flow disorder will occur when the fluid flows through, resulting in the decrease of head and efficiency. In Figure 7(b), it can be found that the head and efficiency of the centrifugal pump have significantly decreased after the impeller is damaged, which is caused by the unstable flow in the pump. In Figure 7(c), machine seal damage leads to serious system leakage, which will also make the flow in the pump more unstable. Centrifugal pump head and efficiency decline.

5.4. Predictive Model Validation Steps

Step 1. The data of the above tests were collated to compute the root mean square, peak factor, skewness coefficient, and kurtosis required for the prediction.

Step 2. Carry out data cleaning and standard deviation standardization for the collated data, and combine the four characteristic parameters according to the weight analysis results of the ReliefF algorithm. Then, input them into the fault prediction model for faulty judgment.

Step 3. Compare the output of the prediction model with the actual test centrifugal pump state. Under the flow of $0.8 Q_d$, $1.0 Q_d$, and $1.2 Q_d$, 100,000 data were collected, respectively,

and divided into 10 groups for fault prediction accuracy as shown in Figure 8.

The test results demonstrate that the improved KNN centrifugal pump fault prediction model can realize the monitoring of four kinds of centrifugal pump operation faults. The prediction model has superior prediction speed and accuracy for unknown samples. Under the condition of $0.8 Q_d$, the average prediction accuracy is 0.859, and the highest prediction accuracy is 0.871. Under the condition of $1.0 Q_d$, the average prediction accuracy is 0.821, and the highest prediction accuracy is 0.844. Under the condition of $1.2 Q_d$, the average prediction accuracy is 0.855, and the highest prediction accuracy is 0.881. The results show that the centrifugal pump runs stably under the design condition. The vibration amplitude is smaller than that of off-working conditions, and the prediction accuracy is relatively low. However, the vibration of the centrifugal pump is exacerbated under the off-working condition, and the prediction accuracy of the prediction model is higher. The fault diagnosis method has good accuracy and practicability because the water pump mostly operates in off-working conditions in actual use. Some verification results are presented in Table 7.

6. Results and Discussion

- (1) Using the ReliefF weight analysis algorithm not only can be the intuitive observation of vibration monitoring parameters of centrifugal pump failure prediction weight proportion but also can remove the redundant sieve parameters. At the same time, there are too many parameters that can be solved and influence each other because the failure prediction model is too complex and leads to the fact that the failure prediction results are not accurate.

TABLE 7: Verification results.

Flow Q	Set1	Set2	Set3	Set4	Set5	Set6	Set7	Set8	Set9	Set10
0.8 Q_d	0.845	0.857	0.862	0.849	0.871	0.893	0.864	0.841	0.837	0.865
1.0 Q_d	0.819	0.807	0.828	0.809	0.835	0.840	0.844	0.825	0.802	0.804
1.2 Q_d	0.847	0.848	0.847	0.852	0.846	0.870	0.856	0.881	0.850	0.849

- (2) The distance function in the KNN algorithm is replaced by the Mahalanobis distance, which avoids the disadvantage that the Euclidean distance cannot describe the correlation between variables and eliminates the dimensionless influence among various vibration parameters. At the same time, it is proved that the Mahalanobis distance is an excellent performance and application in fault diagnosis.
- (3) The improved KNN centrifugal pump fault prediction model can effectively identify the centrifugal pump fault, comply with the requirements of centrifugal pump fault diagnosis, and predict the centrifugal pump fault situation in time and quickly. It also avoids the difficulty of centrifugal pump fault analysis caused by the simultaneous monitoring of multiple parameters, so that the centrifugal pump fault monitoring has a unified judgment method.
- (4) In the practical application, the sample size and sample types used by the training model can be increased according to the needs of users, to further enhance the accuracy of model prediction and the types of predicted faults.

Data Availability

There are no publicly archived datasets.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] L. Gao, L. Dong, J. Gao, S. Wang, and W. Liu, "Acoustic emission-based small leak detection of propulsion system pipeline of sounding rocket," *Shock and Vibration*, vol. 2020, Article ID 8875939, 9 pages, 2020.
- [2] B. Peeters and G. D. Roeck, "Reference-based stochastic subspace identification for output-only modal analysis," *Mechanical Systems and Signal Processing*, vol. 13, no. 6, pp. 855–878, 1999.
- [3] M. I. Basseville, A. Benveniste, M. Goursat, L. Hermans, L. Mevel, and H. V. D. Auweraer, "Output-only subspace-based structural identification: from theory to industrial testing Practice1," *Journal of Dynamic Systems, Measurement, and Control*, vol. 123, no. 4, pp. 668–676, 2001.
- [4] Z. Xue, X. Cao, and T. wang, "Vibration test and analysis of centrifugal pump," *Journal of Drainage and Irrigation Machinery Engineering*, vol. 36, pp. 466–471, 2018.
- [5] H. Sun, Q. Si, N. Chen, and S. Yuan, "HHT-based feature extraction of pump operation instability under cavitation conditions through motor current signal analysis," *Mechanical Systems and Signal Processing*, vol. 139, Article ID 106613, 2020.
- [6] J. Yu, "Machinery fault diagnosis using joint global and local/nonlocal discriminant analysis with selective ensemble learning," *Journal of Sound and Vibration*, vol. 27, pp. 2301–2311, 2009.
- [7] J. P. Peck and J. Burrows, "On-line condition monitoring of rotating equipment using neural networks," *Energy Processing Canada*, vol. 86, pp. 0319–5759, 1994.
- [8] Z. Yuan, T. Zhou, J. Liu, C. Zhang, and Y. Liu, "Fault diagnosis approach for rotating machinery based on feature importance ranking and selection," *Shock and Vibration*, vol. 2021, Article ID 8899188, 17 pages, 2021.
- [9] M. Battarra and E. Mucchi, "Incipient cavitation detection in external gear pumps by means of vibro-acoustic measurements," *Measurement*, vol. 129, pp. 51–61, 2018.
- [10] S. Gao, Z. Hu, X. Luo, and H. Wang, "Research on fault diagnosis technology of centrifugal pump blade crack based on PCA and GMM," *Measurement*, vol. 173, Article ID 108558, 2021.
- [11] J. Du, S. Wang, and H. Zhang, "Layered clustering multi-fault diagnosis for hydraulic piston pump," *Mechanical Systems and Signal Processing*, vol. 36, no. 2, pp. 487–504, 2013.
- [12] J. Tian, S. G. Wang, J. Zhou, T. A. Yan, Y. W. Zhang, and C. W. Fei, "Fault diagnosis of intershaft bearing using variational mode decomposition with TAGA optimization," *Shock and Vibration*, vol. 2021, Article ID 8828317, 11 pages, 2021.
- [13] S. F. Yuan and F. L. Chu, "Support vector machines-based fault diagnosis for turbo-pump rotor," *Mechanical Systems and Signal Processing*, vol. 20, no. 4, pp. 939–952, 2006.
- [14] J. Wang and H. Hu, "Vibration-based fault diagnosis of pump using fuzzy technique," *Measurement*, vol. 39, no. 2, pp. 176–185, 2006.
- [15] M. Yang, Y. Jiang, J. Huang, X. Zhang, and L. Han, "Application of fault diagnosis of seawater hydraulic pump based on transfer learning," *Shock and Vibration*, vol. 2020, Article ID 9630986, 8 pages, 2020.
- [16] L. Song, P. Chen, and H. Wang, "Automatic decision method of optimum symptom parameters and frequency bands for intelligent machinery diagnosis: application to condition diagnosis of centrifugal pump system," *Advances in Mechanical Engineering*, vol. 2014, Article ID 603408, 13 pages, 2014.
- [17] R. Tiwari and D. J. Bordoloi, "Aakash Dewangan, "Blockage and cavitation detection in centrifugal pumps from dynamic pressure signal using deep learning algorithm", " *Measurement*, vol. 173, Article ID 108676, 2021.
- [18] S. Zhao, K. Chang, E. Wang, B. Li, K. Wang, and Q. Wu, "Fault diagnosis of oil pumping machine retarder based on sound texture-vibration entropy characteristics and gray wolf optimization-support vector machine," *Shock and Vibration*, vol. 2020, Article ID 2709384, 13 pages, 2020.
- [19] T.-M. Tsai and W.-H. Wang, "Diagnosis of mechanical pumping system using neural networks and system parameters analysis," *Journal of Mechanical Science and Technology*, vol. 23, no. 1, pp. 124–135, 2009.
- [20] Y. Yang, H. Zheng, Y. Li, M. Xu, and Y. Chen, "A fault diagnosis scheme for rotating machinery using hierarchical

- symbolic analysis and convolutional neural network,” *ISA Transactions*, vol. 91, pp. 235–252, 2019.
- [21] F. Kong and R. Chen, “A combined method for triplex pump fault diagnosis based on wavelet transform, fuzzy logic and neuro-networks,” *Mechanical Systems and Signal Processing*, vol. 18, no. 1, pp. 161–168, 2004.
- [22] S. Uddin, R. Islam, S. A. Khan et al., “Distance and density similarity based enhanced k -NN classifier for improving fault diagnosis performance of bearing,” *Shock and Vibration*, vol. 2016, Article ID 3843192, 11 pages, 2016.
- [23] W. Jiang, Y. Wang, and Z. Wang, “Fault feature dimension reduction method combined ReliefF algorithm with correlation calculation and its application,” *Chinese Hydraulics & Pneumatics*, vol. 12, pp. 18–24, 2015.
- [24] U. E. Akpudo and H. Jang-Wook, “A multi-domain diagnostics approach for solenoid pumps based on discriminative features,” *IEEE Access*, vol. 8, Article ID 175020, 2020.
- [25] S. Sun, “Segmentation-based adaptive feature extraction combined with Mahalanobis distance classification criterion for heart sound diagnostic system,” *IEEE Sensors Journal*, vol. 21, no. 9, Article ID 11009, 2021.
- [26] H. Sarmadi, A. Entezami, B. S. Razavi, and K. Yuen, “Ensemble learning-based structural health monitoring by Mahalanobis distance metrics,” *Structural Control and Health Monitoring*, vol. 28, 2021.
- [27] T. T. Wong and P. Y. Yeh, “Reliable accuracy estimates from K -fold cross validation,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 8, pp. 1586–1594, 2020.
- [28] H. Zhao, H. Liu, Y. Jin, X. Dang, and W. Deng, “Feature extraction for data-driven remaining useful life prediction of rolling bearings,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021.
- [29] W. Deng, S. Shang, X. Cai, H. Zhao, Y. Song, and J. Xu, “An improved differential evolution algorithm and its application in optimization problem,” *Soft Computing*, vol. 25, no. 7, pp. 5277–5298, 2021.