

## Research Article

# Fault Diagnosis for Hydraulic Servo System Using Compressed Random Subspace Based ReliefF

Yu Ding <sup>1,2</sup>, Fei Wang <sup>1,2</sup>, Zhen-ya Wang <sup>3</sup>, and Wen-jin Zhang <sup>1,2</sup>

<sup>1</sup>School of Reliability and Systems Engineering, Beihang University, Beijing, China

<sup>2</sup>Science & Technology on Reliability and Environmental Engineering Laboratory, Beijing, China

<sup>3</sup>Research and Development Center, China Academy of Launch Vehicle Technology, Beijing, China

Correspondence should be addressed to Wen-jin Zhang; [buaazwjok@yeah.net](mailto:buaazwjok@yeah.net)

Received 28 October 2017; Revised 5 January 2018; Accepted 14 January 2018; Published 18 February 2018

Academic Editor: Gangbing Song

Copyright © 2018 Yu Ding 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.

Playing an important role in electromechanical systems, hydraulic servo system is crucial to mechanical systems like engineering machinery, metallurgical machinery, ships, and other equipment. Fault diagnosis based on monitoring and sensory signals plays an important role in avoiding catastrophic accidents and enormous economic losses. This study presents a fault diagnosis scheme for hydraulic servo system using compressed random subspace based ReliefF (CRSR) method. From the point of view of feature selection, the scheme utilizes CRSR method to determine the most stable feature combination that contains the most adequate information simultaneously. Based on the feature selection structure of ReliefF, CRSR employs feature integration rules in the compressed domain. Meanwhile, CRSR substitutes information entropy and fuzzy membership for traditional distance measurement index. The proposed CRSR method is able to enhance the robustness of the feature information against interference while selecting the feature combination with balanced information expressing ability. To demonstrate the effectiveness of the proposed CRSR method, a hydraulic servo system joint simulation model is constructed by HyPneu and Simulink, and three fault modes are injected to generate the validation data.

## 1. Introduction

Hydraulic servo system plays a crucial role in electromechanical systems, like engineering machinery, metallurgical machinery, ships, and other equipment. Failures of hydraulic servo system caused by severe and complex conditions may lead to catastrophic accidents and enormous economic losses. Fault diagnosis based on monitoring and sensory signals is able to classify the current state of complex systems, which plays a key role in performance evaluation [1]. Feature set extracted from signals is an important index to reflect the fault mechanism and performance evolution laws. The quality of feature set plays a key role in improving the generalization ability of fault identification [2]. The common feature extraction methods are time-frequency index extraction, wavelet analysis, Hilbert transform, Duffing oscillator, and so on. Despite their respective applicable conditions and limitations, those methods are able to mine the health characteristics of the system from multiaspect [3, 4]. Same as

machine learning, features extracted from images, speeches, and other signals often have certain correlations and hidden mutual influences. Information expressed by a single feature is usually inadequate, which can be greatly improved when the single feature is aggregated with others [5]. Similarly, due to the nonlinearity, instability, and nonconformity of complex electromechanical systems, the expression of the information on individual feature is often one-sided. Thus, a new challenge is how to utilize those features more effectively and efficiently, in other words, how to obtain the feature set that expresses the information sufficiently by eliminating the redundant and negatively correlated features [6–9].

To tackle the challenge mentioned above, on the premise of existing feature extraction techniques, feature processing techniques including feature selection and dimension reduction have gradually become an important research focus. Both feature selection and dimension reduction can reduce the scale of feature set by obtaining a set of principal variables. Such techniques often use a variety of feature

extraction methods to integrate the features into a comprehensive representation of the signal [10]. For the purpose of enhancing the expressing ability of core information on multiclass feature sets, spatial transformation or importance measurement methods are used [11]. Such methods are able to reduce redundancy existing in features and improve learning efficiency while retaining the performance advantages. The data transformation may be linear, such as principal component analysis (PCA). But many nonlinear dimension reduction techniques also exist. The common feature processing techniques include linear dimensionality reduction methods (FDA, LPP, etc.), kernel function based dimension reduction methods (KPCA, KFDA, etc.), manifold based dimension reduction methods (Isomap, LLE, MDS, etc.), filter method based feature selection methods (Relief, Focus, information gain, CBFS, etc.), wrapper method based feature selection methods (genetic algorithm, distribution estimation, differential algorithm, etc.), and embedded method based feature selection methods (SVN-RFE, RF, etc.) [12–14].

For the nonlinear signals of complex electromechanical systems, although the dimension reduction methods can reduce the scale of input features for fault diagnosis, they change the basic attributes of the feature set. Such situation makes it difficult to give a clear understanding of the obtained feature subset. Meanwhile, existing feature selection methods sort the importance degree of the features according to the independent feature evaluation result. They ignore the interaction among features, which would lead to information loss in processing the data of electromechanical systems [15]. Aiming at the shortcomings of feature selection and dimension reduction techniques, such as low expandability, unclear evaluation indexes, and strong tendency, this study proposes compressed random space based ReliefF (CRSR) method. Based on ReliefF method, CRSR introduces ensemble strategy on feature level based on compressed random space. Furthermore, CRSR optimizes the objective function using information entropy and fuzzy theory. The main contributions are as follows:

- (i) This study analyzes the feasibility of the ReliefF based feature selection architecture for the fault diagnosis of complex electromechanical system. Meanwhile, the basic mechanism of measuring the contribution of the features based on ReliefF is also demonstrated.
- (ii) By converting the assessment process of ReliefF, which takes the entire feature set as object, into the construction and ensemble process of feature subspace, CRSR can improve the global optimization ability of ReliefF.
- (iii) Considering ReliefF based feature selection as a problem of maximizing the distance, CRSR replaces the traditional spatial distance with fuzzy membership degree, which is able to obtain a robust and steady objective function.

This paper is structured as follows: In Section 2, ReliefF method based feature selection structure is introduced. Then feature integration method based on compressed random subspace is described. Objective function optimization

method based on information entropy and fuzzy theory is also introduced. Section 3 presents the overall diagnosis procedure for hydraulic servo systems, the details of construction and fault injection of the hydraulic servo system, and analysis and comparison on feature selection results using the proposed CRSR method which are discussed numerically. Section 4 concludes the paper.

## 2. Related Theories

*2.1. ReliefF Method Based Feature Selection Structure.* ReliefF is the extension of Relief method by estimating probabilities more reliably, which is able to handle incomplete and multiclass data sets while the complexity remains the same [16–19]. By calculating the distances between the sample distributions, ReliefF can obtain the correlation weight coefficients of the features which is similar to Relief.

For a specific feature from the feature set, if its difference in same class is much smaller than that in different class, it is considered that this feature contributes to class discrimination [20]. Given a sample set  $S = \{s_1, \dots, s_i, \dots, s_n\}$  ( $i \in [1, n]$ ) with  $n$  instances, each sample,  $s_i = \{s_{i1}, \dots, s_{it}, \dots, s_{im}\}$  ( $t \in [1, m]$ ), has  $m$ -dimensional features. Meanwhile, the samples in  $S$  only belong to two classes which are tagged as  $c_i \in (C_1, C_2)$ . The difference between each two samples ( $s_i$  and  $s_j$ ) on feature  $t$  is defined as

$$\text{diff}(t, s_i, s_j) = \begin{cases} 1, & s_{it} \neq s_{jt} \\ 0, & s_{it} = s_{jt} \end{cases} \quad (1)$$

where the attribute of feature  $t$  is discrete value. If the attribute of feature  $t$  is continuous value,

$$\text{diff}(t, s_i, s_j) = \left| \frac{s_{it} - s_{jt}}{\max_t - \min_t} \right|. \quad (2)$$

Features extracted from condition monitoring data of electromechanical system mostly are continuous data. Meanwhile  $\max_t$  and  $\min_t$  represent the maximum value and minimum value of the entire sample, respectively. The closest same-class instance of sample  $s_i$  is called “near-hit (NH),” and the closest different-class instance of sample  $s_i$  is called “near-miss (NM).” Meanwhile, the weight of feature  $t$  is denoted as  $\omega_t$ , and  $\omega_t$  is updated by

$$\omega_t = \omega_{t\text{-old}} - \text{diff}(t, s_i, \text{NH}) + \text{diff}(t, s_i, \text{NM}), \quad (3)$$

where the initial value of  $\omega_t$  is 0, and  $\omega_{t\text{-old}}$  is the last value of  $\omega_t$ .

For reducing the randomness in feature evaluation, the whole process needs to repeat  $r$  times to obtain the average value being the final weight. Although Relief method is very efficient in estimating the quality of attributes, it cannot deal with incomplete data and is limited to two-class problem. Thus ReliefF method is utilized in the paper to deal with the multiclass classification and regression problems for continuous data.

For a multiclass classification problem, assuming that the samples in  $S$  belong to multiple classes and the tags for  $S$  are

$c_i \in (C_1, \dots, C_l)$ , ReliefF updates  $\omega_t$  on sample  $s_i$  by taking  $k$  near hits (NHS) and  $k$  near misses (NMS) into consideration,

$$\omega_t = \omega_{t\text{-old}} - \frac{\sum_{j=1}^k \text{diff}(t, s_i, \text{NHS}_j)}{(r * k)} + \frac{\sum_{c \neq \text{class}(s_i)} [(p(c_i) / (1 - p(\text{class}(s_i)))) \sum_{j=1}^k \text{diff}(t, s_i, \text{NMS}_j)]}{(r * k)}, \quad (4)$$

where  $p(c_i)/(1 - p(\text{class}(s_i)))$  represents the ratio of the entire samples in class  $c_i$  to all the heterogeneous samples in  $s_i$ . Furthermore, ReliefF method equalizes the differentiation of NHs and calculates the average differences between  $s_i$  and other classes on feature  $t$  to evaluate the classification ability of the samples nearby.

**2.2. Feature Integration Method Based on Compressed Random Subspace.** The purpose of CRSR method is finding the balance between the differences and the correlations of features. Specifically, in the premise of fully mining the correlations of features using ReliefF method, CRSR method is applied to make each feature subset keep a certain degree of difference. Based on random subspace and feature sorting strategy, the feature sets with higher contribution can be obtained in various feature combinations [21, 22].  $RS_1$  and  $RS_2$  are denoted as two random subspaces, so the difference, denoted as  $D\_RS(RS_1, RS_2)$ , can be calculated as

$$D\_RS(RS_1, RS_2) = |RS_1 \cup RS_2| - |RS_1 \cap RS_2|, \quad (5)$$

where symbol  $|\cdot|$  denotes the dimension of random subspaces.

The right side of (5) obtains the noncoincident features of  $RS_1$  and  $RS_2$ .  $D\_RS(RS_1, RS_2)$  is plus one when there is an unrepeat feature  $f_D$ . The average difference between two random subspaces on all the features is defined as

$$E(D\_RS) = \sum_{q=1}^Q [P(f_D \in RS_1) P(f \notin RS_2) + P(f_D \notin RS_1) P(f \in RS_2)], \quad (6)$$

where  $Q$  is the dimension after feature ordering compression. It can be simplified as  $Q = m - m_1$ . Concretely, according to sorting strategy,  $RS_1$ , the probability of the first  $m_0$  strongly related features being selected is  $2m_0/Q$ , and the relatively poorly related  $m_0$  features being selected are  $m_1/(Q - m_0)$ . The average difference of feature evaluation based on compressed random subspace method, which is  $E(D\_RS)$ , can be denoted as follows:

$$\begin{aligned} E(D\_RS) &= 2 \left( \left( \frac{1}{2} Q \right) \frac{m_0}{(1/2)Q} \left( 1 - \frac{m_0}{(1/2)Q} \right) \right. \\ &\quad \left. + (Q - m_0) \frac{m_1}{Q - m_0} \left( 1 - \frac{m_1}{Q - m_0} \right) \right) \\ &= 2 \left( \frac{m_0(Q - 2m_0)}{Q} + \frac{m_1(Q - m_1 - m_0)}{Q - m_0} \right). \end{aligned} \quad (7)$$

Equation (7) shows that the ranking result can balance the difference of ReliefF to determine the dominant features that

which is different from Relief. Similarly, the weight of feature  $t$ , which is  $\omega_t$ , can be updated through

are crucial for classification, which would improve the feature selection efficiency.

Based on compressed random subspace method, this study proposes redundancy analysis method from statistics to reduce the redundancy of feature. The features are checked in pairs using redundancy analysis method [23, 24]. Firstly, two sets of feature vectors,  $F_1$  and  $F_2$ , are selected randomly from the feature set obtained by ReliefF. Then, the selected feature vectors are regarded as independent variable and dependent variable separately. The covariance matrixes, denoted as  $U$  and  $V$ , can be calculated, respectively, as

$$\begin{aligned} U &= E \{ (F_1 - E[F_1]) (F_1 - E[F_1])^T \}, \\ V &= E \{ (F_2 - E[F_2]) (F_2 - E[F_2])^T \}, \end{aligned} \quad (8)$$

where  $E[X]$  represents mathematical expectation of vector  $X$ . Then, the correlation coefficient of  $F_1$  and  $F_2$ , denoted as  $\rho(U, V)$ , can be formulated as

$$\rho(U, V) = \frac{\text{Cov}(U, V)}{\sqrt{\text{Var}(U)} \sqrt{\text{Var}(V)}}, \quad (9)$$

where  $\text{Cov}(U, V)$  denotes the covariance of  $U$  and  $V$ . If the correlation coefficient is greater than a presetting threshold, only the one with larger weight from  $F_1$  and  $F_2$  will be added to the final selected feature set. It is noticed that redundancy analysis based on matrix transformation focuses on the correlation between features instead of the similarity of data. Thus, CRSR can reduce the influence of numerical confusion existing in feature selection. Furthermore, compared with traditional methods based on data similarity, CRSR is able to obtain higher confidence level.

**2.3. Objective Function Optimization Method Based on Information Entropy and Fuzzy Theory.** From the aspect of maximizing the distance, Relief method can be seen as a distance optimization algorithm using feature weighting method [25]. Under this condition, the optimization objective function, denoted as  $\max_{\omega} J$ , can be described as [26]

$$\max_{\omega} J = \sum_{i=1}^n \rho_i(\omega), \quad (10)$$

where  $\rho_i(\omega)$  denotes the distance of the  $i$ th sample. Based on (4),  $\rho_i(\omega)$  can be converted as

$$\rho_i(\omega) = \sum_{t=1}^m \omega_t \cdot \left( \sum_{c \notin \text{class}(s_i)} \frac{p(c_i)}{1 - p(\text{class}(s_i))} \sum_{j=1}^k \text{diff}(t, s_i, \text{NMS}_j) - \sum_{j=1}^k \text{diff}(t, s_i, \text{NHS}_j) \right). \quad (11)$$

For complex electromechanical system signals, two problems occur when using (10) as the optimization objective function of CRSR. One problem is that the objective function concentrates the weight on one or some of the features, which leads to the result that assessment value of the remaining features tends to 0. Meanwhile, (10) regards the samples with stochastic volatility and noise similar to the normal sample. Another problem is the lack of consideration on the influence from the quality of samples on the feature selection process.

Aiming at the first problem, the information entropy theory is proposed based on compressed random subspace method, which combines the maximization of entropy together with maximization of distance to reduce the over tendency problem of the existing ReliefF method. After adding a sample estimation factor  $\kappa_i$ , the optimization objective function is denoted as

$$\max_{\omega, \kappa} J = \sum_{i=1}^n \kappa_i \rho_i(\omega). \quad (12)$$

Supposing that  $\kappa_i$  and  $\omega_i$  follow probability distribution, Shannon entropy is used to adjust the sample distribution, as shown below:

$$J_e(\kappa_i) = - \sum_{i=1}^n \kappa_i \ln \kappa_i, \quad (13)$$

$$J_e(\omega_i) = - \sum_{i=1}^m \omega_i \ln \omega_i.$$

Aiming at the second problem, the fuzzy membership degree is chosen to replace the traditional nearest neighbor distance. The fuzzy membership degree has the advantage of being insensitive to sample fluctuations and noise and the ability of updating adaptively while changing the feature weights [27]. In a sample space of same class, the fuzzy membership degree of the feature  $s_{it}$ , denoted as  $\nu_{ij}$ , can be calculated as [28, 29]

$$\nu_{ij} = \frac{(Q_{\nu,ij})^{1/(1-m)}}{\sum_{i=1}^{|\text{NHS}|} \left( (Q_{\nu,ij})^{1/(1-m)} \right)}, \quad (14)$$

where  $|\text{NHS}|$  is the sample set of same class, and  $Q_{\nu,ij}$  is the fuzzy difference between feature  $s_{it}$  and feature  $s_{jt}$ , as shown below:

$$Q_{\nu i} = \sum_{j=1}^M \omega_j \text{diff}(t, s_i, s_j), \quad s_j \in \text{NHS}. \quad (15)$$

Then the same fuzzy distance, denoted as  $m\_diff(t, s_i, \text{NHS})$ , is calculated as

$$m\_diff(t, s_i, \text{NHS}) = \sum_{s_j \in \text{NHS}} \nu_{ij} \text{diff}(t, s_i, s_j). \quad (16)$$

Similarly, the fuzzy difference and the fuzzy membership degree of the heterogeneous sample sets of  $s_j$ , which are denoted as  $Q_{\mu i}$  and  $\mu_{ij}$  separately, can be calculated as follows:

$$Q_{\mu i} = \sum_{j=1}^M \omega_j \text{diff}(t, s_i, s_j), \quad s_j \in \text{NMS}, \quad (17)$$

$$\mu_{ij} = \frac{(Q_{\mu,ij})^{1/(1-m)}}{\sum_{i=1}^{|\text{NMS}|} \left( (Q_{\mu,ij})^{1/(1-m)} \right)}.$$

Therefore, the heterogeneous fuzzy distance, denoted as  $m\_diff(t, s_i, \text{NMS})$ , is calculated as

$$m\_diff(t, s_i, \text{NMS}) = \sum_{s_j \in \text{NMS}} \mu_{ij} \text{diff}(t, s_i, s_j). \quad (18)$$

Based on the fuzzy distances obtained by (16) and (18), updated (11) can be formulated as follows:

$$\rho_{f,i}(\omega) = \sum_{t=1}^m \omega_t \cdot \left( \sum_{c \notin \text{class}(s_i)} \frac{p(c_i)}{1 - p(\text{class}(s_i))} \sum_{j=1}^k m\_diff(t, s_i, \text{NMS}_j) - \sum_{j=1}^k m\_diff(t, s_i, \text{NHS}_j) \right). \quad (19)$$

Based on the formulas above, the objective function, which is  $\max_{\omega, \kappa} J$ , can be denoted as

$$\max_{\omega, \kappa} J = \sum_{i=1}^n \kappa_i \rho_{f,i}(\omega) + \lambda_{\kappa} J_e(\kappa_i) + \lambda_{\omega} J_e(\omega_i), \quad (20)$$

where the first item on the right side is the maximum distance of ReliefF, which is meant to determine the feature set that contributes most to classification. Meanwhile, from the aspect of entropy maximization, the second and the third items denote the sample evaluation operator and Shannon entropy of feature weight, respectively, which are used to avoid the over tendency problem of objective function.  $\lambda_{\kappa}$  and  $\lambda_{\omega}$  are the balance coefficients for adjusting the differences between features. When the maximum point of objective function is

achieved, the constraint condition of the sample evaluation operator is defined below [30]:

$$\kappa_i = \frac{\exp\left(\sum_{j=1}^m \omega_j d_{i,j}/\lambda_\kappa\right)}{\sum_{i=1}^n \exp\left(\sum_{j=1}^m \omega_j d_{i,j}/\lambda_\kappa\right)}, \quad (21)$$

where  $d_{i,j}$  denotes the partial derivative of  $\rho_{f,i}(\omega)$ .

In the process of feature selection for mechanical system, the information entropy and fuzzy theory based optimization objective function ensures that the evaluation process of each subspace using ReliefF is adaptive and robust. Such advantage provides a new thought for feature processing of complex monitoring signal, in other words, under the premise of maintaining the high calculating efficiency of ReliefF method, reducing the bias and redundancy caused by methodological defects and external disturbances [31].

### 3. Method for Hydraulic Servo System Fault Diagnosis Based on CRSR

The CRSR based fault diagnosis method for hydraulic servo system consists of following successive steps: first, the average, standard deviation, skewness, and wavelet singular entropy features are extracted to form a feature matrix as the input of ReliefF model. Second, the initial contribution of the features sampled randomly is measured by calculating the inner-class distance and between-class distance, and the features of high contribution and features of low contribution are determined for fault classification. The  $i$ th iteration operation is based on the result of the previous iteration, and the iteration stops as long as  $i$  reaches the preset threshold. Third, based on the sorting result of the second step, the compressed evaluation of the features is realized using supervised sampling method in current iteration. Finally, keeping the iteration running until certain terms is satisfied to acquire the difference value as the criterion of feature selection. The detailed process of feature selection method using CRSR for hydraulic servo pump is shown in Figure 1.

Compared with the traditional feature processing methods [32–35], the CRSR method is designed to meet the robustness and accuracy requirements of fault diagnosis with smaller feature set and lower resource consumption. This study demonstrates the advantages of the proposed method by extracting a variety of categories of features from the simulation data of hydraulic servo system and optimizes the feature set extracted by CRSR method to verify the feasibility in feature selection technique.

**3.1. Description of the Simulation Environment of Hydraulic Servo System.** The data used in this case is generated in HyPneu and Simulink joint simulation environment. The joint simulation process of hydraulic servo system can be divided into mechanical part and control part considering the characteristics of the hydraulic components [36–38]. The mechanical hydraulic physical part, including hydraulic pumps, servo valves, and actuators, is modeled using component library in HyPneu. The failure is realized by adding the modules of fault injection for the relevant components. The dynamic control

part, including feedback sensors and electronic amplifiers, is established using the relevant model in Simulink. Meanwhile, the control of the hydraulic servo system and the simulation of fault injection can be realized by transmitting the signal data through relevant interface files. Thus, the model architectures of hydraulic servo system constructed using HyPneu and Simulink are shown in Figures 2 and 3, respectively.

As shown in Figure 3, the Simulink model includes input signals, comparison elements, control elements, amplifying elements, and HyPneu module. The gain parameter of the electronic amplifying element is 80, and the control parameter of PID is set as proportional parameter of 1500, integral parameter of 0, and differential parameter of 5. To simulate the actual operating environment of hydraulic system, random noises in the range of  $-0.01$  to  $0.01$  are added on the output of the HyPneu module.

The process of joint simulation based on HyPneu and Simulink is as follows. First, the actuator receives the feedback signal converted by servo valve and right after that drives the load to do reciprocated motion. Then, the real-time displacement information of the load is collected by the sensor and transferred to the control circuit. Finally, the input signal and the amplified displacement signal are compared and inputted into the servo valve to build the negative feedback control logic of closed loop. The joint simulation model has a good ability to match the hydraulic system and to realize the fault injection of the hydraulic system effectively.

**3.2. Fault Injection for Hydraulic Servo System and Multi-dimensional Feature Extraction.** This study simulates four kinds of state of the hydraulic system including normal state, electronic amplifier fault, sensor constant deviation fault, and hydraulic pump wear fault. The detailed fault injection scheme is shown in Table 1.

The input of all fault modes is sinusoidal signal with amplitude being 1 and frequency being  $4\pi$  rad/s. As the expression of the input signal is  $y = \sin(4\pi t)$ , the sampling frequency is 100 Hz and the sampling time is 70 s. The input-output relationships of the hydraulic servo system under circumstances of normal state and three fault states are collected, and every signal of the corresponding state contains 7000 points. The original output signals are shown in Figure 4.

As mentioned above, the input of the proposed CRSR method is a feature matrix, and each column of the matrix represents a feature vector extracted from the original signal. As the prerequisite of feature selection, multidimensional features should be extracted for all of the signals. It should be noted that the cycle of hydraulic servo system signal collected for the feature selection should be an integer multiple of the input waveform cycle. The cycle of the input signal is 0.5 seconds and one input-output cycle contains 50 points, which means the object points of the feature extraction should be a geometric multiple of 50. In addition, in consideration of the fact that pump wear fault is a gradual degradation process and the degree of fault in the early phase may be weak, this study takes the last 4000 points of the original signal to extract the feature matrix. Moreover, a sampling window of 500-point length is used (with 5-point interval) to obtain 700 feature vectors.

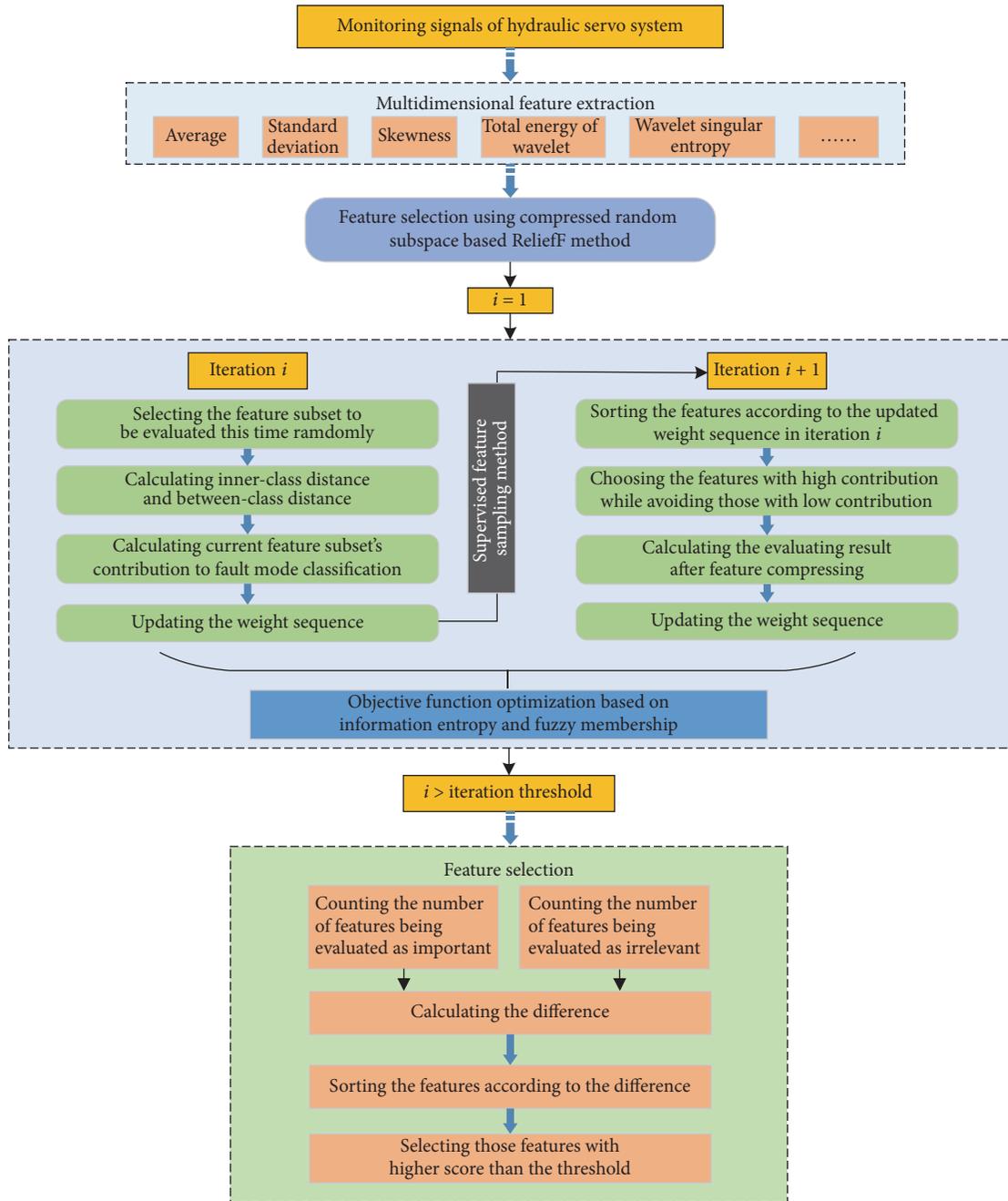


FIGURE 1: Process of CRSR based feature selection method for hydraulic servo pump.

TABLE 1: Fault injection scheme.

Fault mode	Fault phenomenon	Fault injection mode
Normal state	-	-
Electronic amplifier fault	Magnification notably increases	Multiplying the gain by 0.25
Sensor fault	Constant gain	Setting the feedback of the sensor to 0.5
Hydraulic wear fault	Output flow decreases	Setting flow-pressure coefficient of the variable orifice valve to 0.03

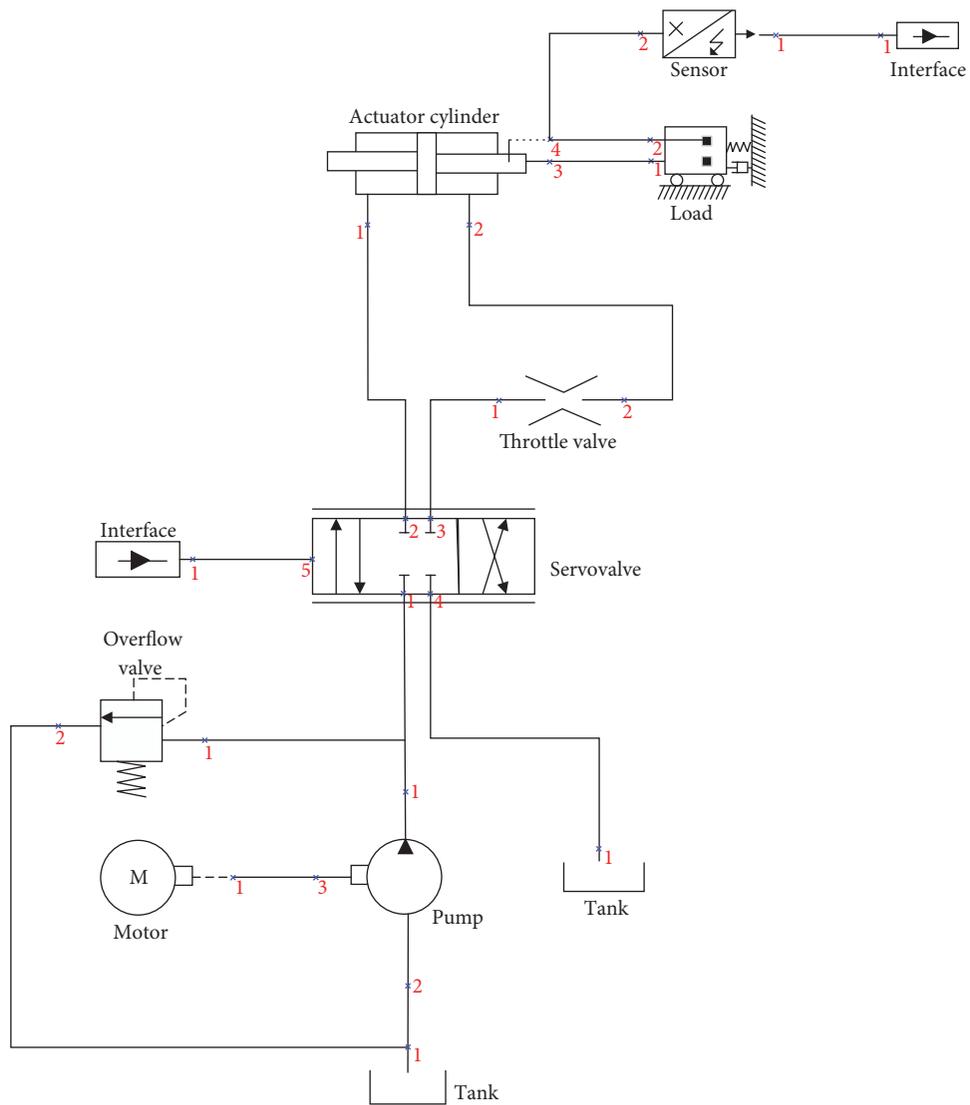


FIGURE 2: The failure injection model of mechanical hydraulic part using HyPneu.

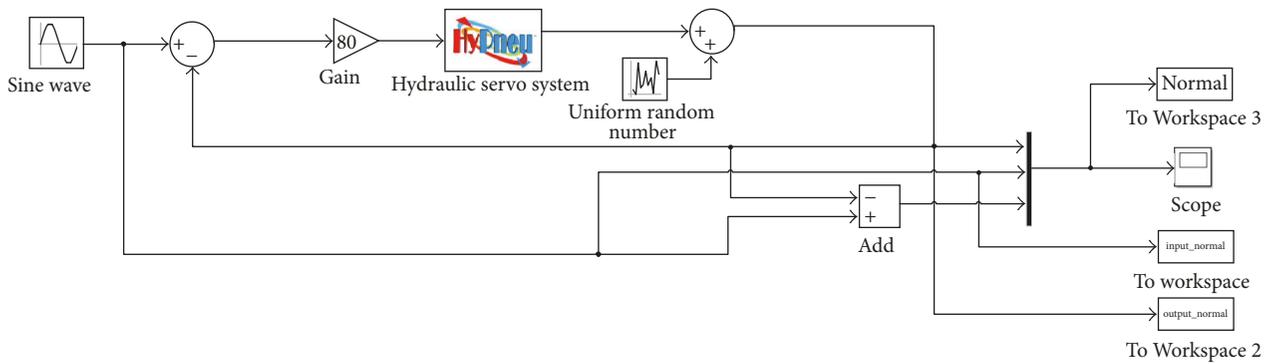


FIGURE 3: The failure injection model of dynamic control part using Simulink.

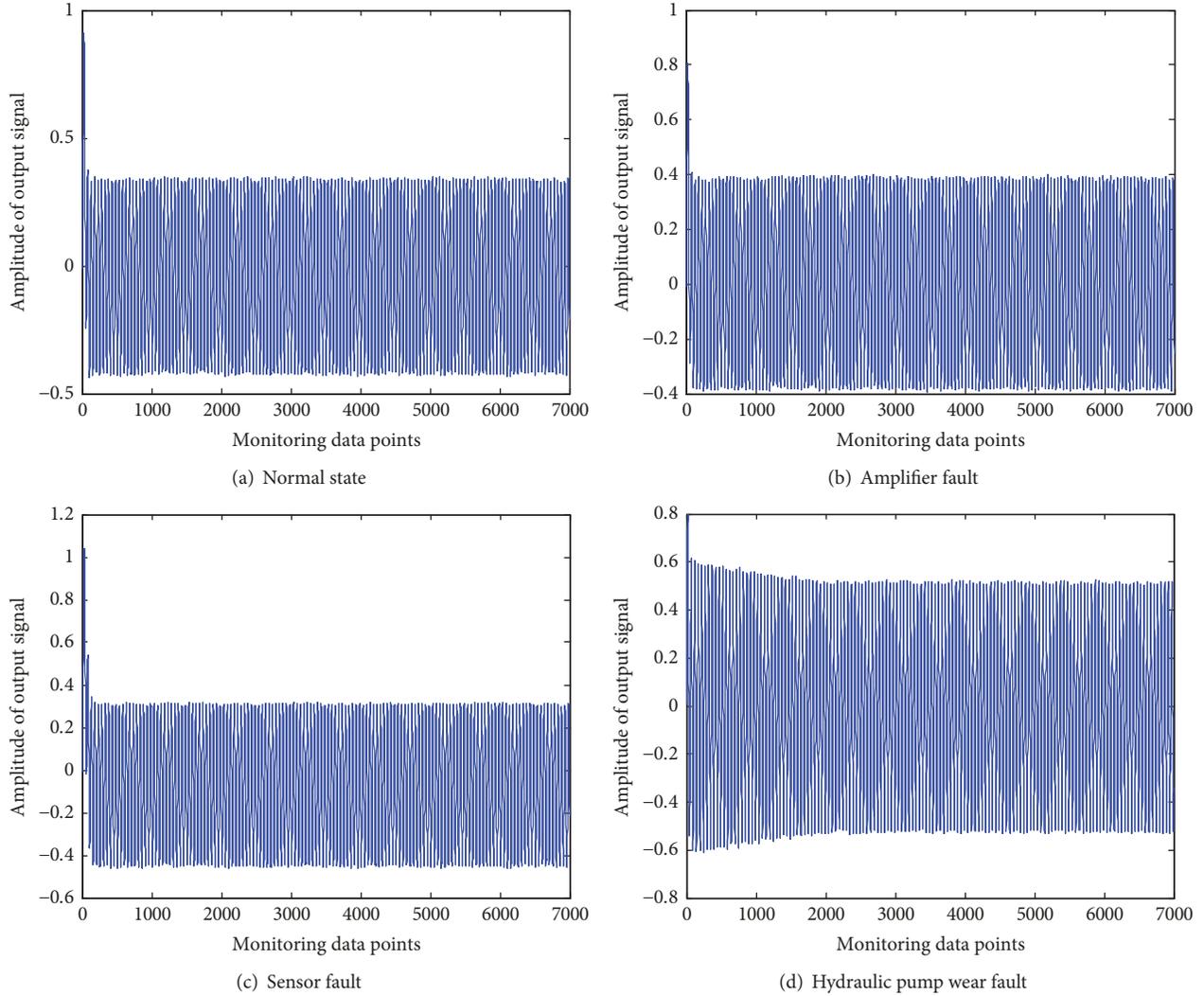


FIGURE 4: Original output signals of the hydraulic servo system.

Furthermore, this study adds 15 dB Gaussian white noise to all the original signals to validate the performance of CRSR method. The details of the extracted features are shown in Table 2.

As can be seen from Table 2, the quantitative range of different fault features fluctuates much, which could lead to the bias problem in the feature selection process of ReliefF. Therefore, this study normalizes the features to the range of zero to one. The visualization of the feature vectors of different states is shown in Figure 5.

As is shown in Figure 5, data confusion exists in the features of different fault modes, which makes it difficult to select satisfied feature subsets through artificial observation. Thus, CRSR method is designed to select the feature combination with higher contribution to the fault diagnosis adaptively.

**3.3. Analysis and Comparison on Feature Selection Result.** The input of CRSR feature selection model is a  $9 \times 700$  matrix acquired by regularizing the 9 sets of the obtained feature vectors. One of the advantages of CRSR is the ability

to calculate the distances of either same class or different classes adaptively and iteratively. Based on the ReliefF feature selection method, the corresponding distribution vectors of feature weight can be obtained.

During the iteration process of ReliefF, feature selection constraint is established using compressed stochastic subspace method, which optimizes the performance and efficiency of CRSR. The parameters are as follows: number of iterations: 30; feature base of the subspace: 6. And the threshold condition of feature selection is that the difference between a feature being selected as “one with high contribution” and “one with low contribution” is greater than 10. With  $F_1-F_9$  representing the feature parameters in Figure 5, the statistical result of the features being selected as “one with high contribution” and “one with low contribution” is shown in Table 3.

It can be seen from Table 3 that the feature combination selected by CRSR method contains average ( $F_1$ ), standard deviation ( $F_2$ ), mean square root ( $F_3$ ), crest factor ( $F_5$ ), and the maximum amplitude of FFT ( $F_6$ ). Although wavelet

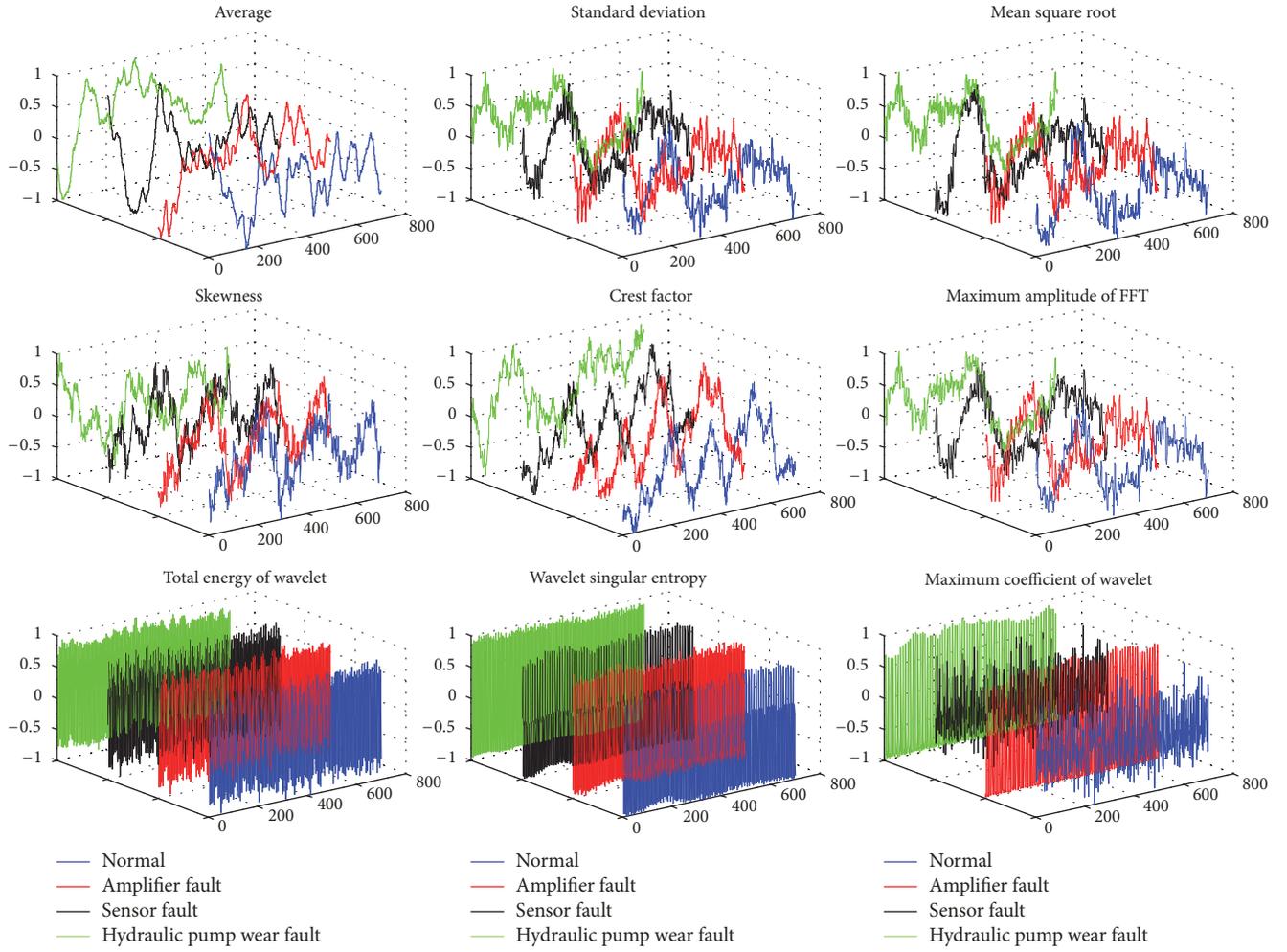


FIGURE 5: Feature vectors of the hydraulic servo system in different states.

TABLE 2: Extracted features of the output signals of hydraulic servo system.

Feature	Feature magnitude			
	Normal	Amplifier fault	Sensor fault	Hydraulic pump wear fault
Average	-0.0446~-0.0411	0.0004~0.0036	-0.0734~-0.0698	-0.0100~-0.0063
Standard deviation	0.2715~0.2728	0.2745~0.2756	0.2715~0.2727	0.3068~0.3080
Mean square root	0.2746~0.2760	0.2742~0.2754	0.2806~0.2819	0.3067~0.3079
Skewness	0.0080~0.0148	0.0028~0.0099	0.0082~0.0149	0.0085~0.0174
Crest factor	1.4603~1.4608	1.4620~1.4685	1.4599~1.4675	1.8089~1.8189
Maximum amplitude of FFT	95.8092~96.288	96.8611~97.2778	95.8319~96.2598	107.5320~107.94
Total energy of wavelet	99.4925~99.746	99.4695~99.7350	99.5243~99.7664	98.5909~99.0743
Wavelet singular entropy	0.7682~0.8193	0.7701~0.8113	0.7857~0.8401	0.8579~0.9178
Maximum coefficient of wavelet	0.0012~0.0035	0.0005~0.0669	0.0004~0.0034	0.0005~0.0543

TABLE 3: Statistical results based on compressed stochastic subspace.

Feature	$F_1$	$F_2$	$F_3$	$F_4$	$F_5$	$F_6$	$F_7$	$F_8$	$F_9$
High contribution	28	26	22	0	12	26	3	14	1
Low contribution	0	0	0	17	1	0	2	6	21
Difference	28	26	22	-17	11	26	1	8	-20

TABLE 4: Times of feature selections based on compressed random subspace.

Dependence analysis	$F_1$	$F_2$	$F_3$	$F_5$	$F_6$
$F_1$	1	0.42	0.69	0.17	0.16
$F_2$	-	1	0.56	0.09	0.44
$F_3$	-	-	1	0.31	0.61
$F_5$	-	-	-	1	0.82
$F_6$	-	-	-	-	1

TABLE 5: Fault diagnosis results with different combinations of features.

Feature combination	Fault diagnosis accuracy of the 10-fold cross validation method (%)									
$F_1 + F_2 + F_3 + F_6$	94.62	95.17	94.29	94.40	93.81	94.11	93.54	95.12	95.01	93.98
$F_1 + F_2 + F_3 + F_5 + F_6$	94.48	93.61	93.01	93.77	94.15	95.25	93.99	94.58	94.62	93.71
$F_1 + F_2 + F_3 + F_5 + F_6 + F_7 + F_8$	92.04	93.62	93.11	93.25	91.37	94.54	93.27	92.32	93.29	93.41
$F_1 + F_3 + F_4 + F_5$	93.02	93.57	92.15	92.89	92.46	94.19	93.32	93.55	92.53	92.77
$KPCA_1 \sim KPCA_4$	93.08	93.78	93.07	93.57	92.17	93.04	92.80	92.42	93.48	93.36
$LLE_1 \sim LLE_4$	94.10	93.59	94.51	94.22	93.98	94.10	93.99	93.44	93.98	94.78
$F_1 \sim F_9$	92.67	91.80	93.54	92.47	92.11	93.64	92.97	93.62	94.01	93.00

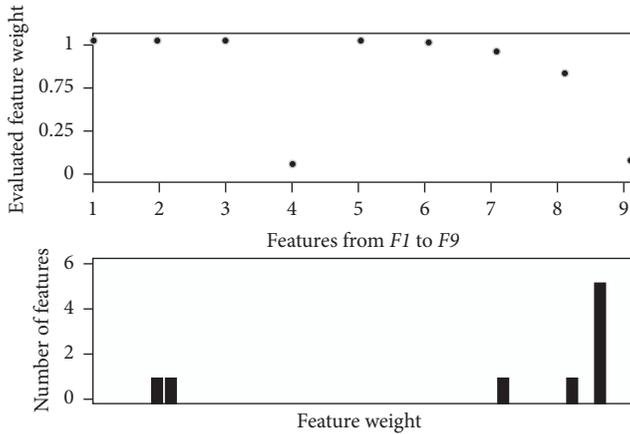


FIGURE 6: Statistical results based on compressed stochastic subspace.

singular entropy ( $F_8$ ) was selected as “one with high contribution” for 14 times during the iteration, it was selected as “one with low contribution” for 6 times. Such result reveals that wavelet singular entropy is able to contribute to classification sometimes, but the ensemble performance is not stable. Even in some feature subspace matrixes, it exerts a negative impact on the classification. The final statistical result of feature weights assessed by CRSR is shown in Figure 6.

As is shown in Figure 6, if the weight basis for ReliefF is set to 0.75, the feature combination selected by CRSR method will append features  $F_7$  (total energy of wavelet) and  $F_8$  (wavelet singular entropy). However, according to the statistical results from the CRSR model, the contribution of these two features in the integration process is unstable. In particular, as for the total energy of wavelet, the reason why its average weight coefficient is high is that the weight coefficient reaches 0.92 and 0.88 in the 8th and 22nd iterations, respectively, while in the remaining subspaces

its contributions are lower than the average slightly. Such situation reveals that the total energy of wavelet could not meet the requirement of stability in feature selection. The result shows that the CRSR method is more reasonable than the traditional ReliefF, which reflects the advantages of ensemble learning in generalization.

Based on the feature selection result using CRSR, redundancy analysis for the feature sequence is carried out in this study. The threshold of correlation coefficient is set to 0.8, and the correlation matrix of feature is shown in Table 4.

It can be seen from Table 4 that the correlation coefficient of peak factor ( $F_5$ ) and the maximum amplitude of FFT transform ( $F_6$ ) are greater than the preset threshold, which indicates that redundancy exists in the fault diagnosis information provided by them. Therefore, only the maximum amplitude of FFT transform ( $F_6$ ) is retained. In summary, the features selected by CRSR are average ( $F_1$ ), standard deviation ( $F_2$ ), mean square root ( $F_3$ ), and the maximum amplitude of FFT ( $F_6$ ).

The purpose of introducing CRSR method is improving the performance of fault diagnosis. In other words, for a classifier, the diagnostic performance using selected feature set should not be less than that using the original feature set. The comparative models used in this study contain the classical ReliefF algorithm, the Mean Impact Value (MIV) algorithm, the Locally Linear Embedding (LLE) algorithm, and Kernel Principal Component Analysis (KPCA). In this study, the feature sets selected by different feature selection methods are used as inputs for the classifier based on Radial Basis Function (RBF) neural network. The ratio of training samples to test samples is 50%. The fault diagnosis accuracy of the hydraulic servo system is obtained by using the 10-fold cross validation method as shown in Table 5.

In Table 5, the first row ( $A_1$ ) represents the final feature set selected by CRSR; the second row ( $A_2$ ) represents the feature set selected by CRSR before the redundancy analysis; the third row ( $A_3$ ) represents the feature set selected by

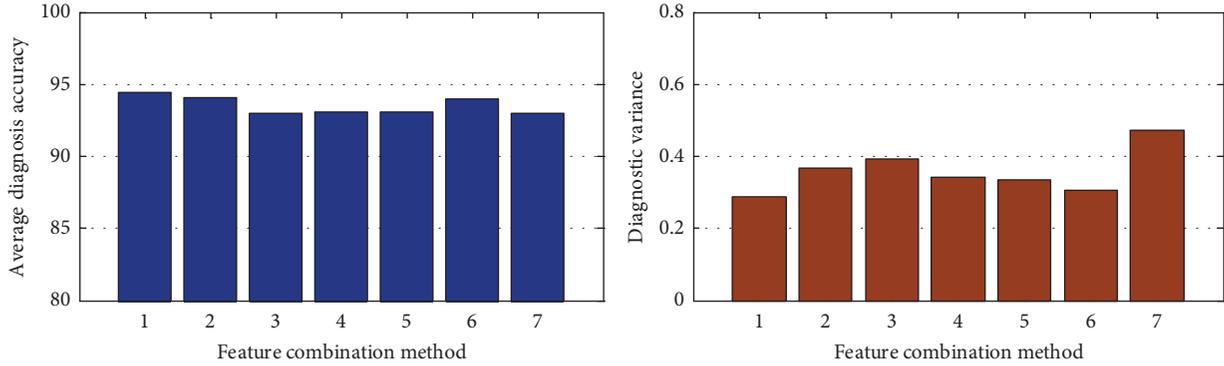


FIGURE 7: Fault diagnosis of hydraulic system under different combinations of features.

Source	SS	df	MS	Chi-sq	Prob > Chi-sq
Groups	14948.2	6	2491.36	36.1	2.63857e - 06
Error	13624.9	63	216.27		
Total	28573	69			

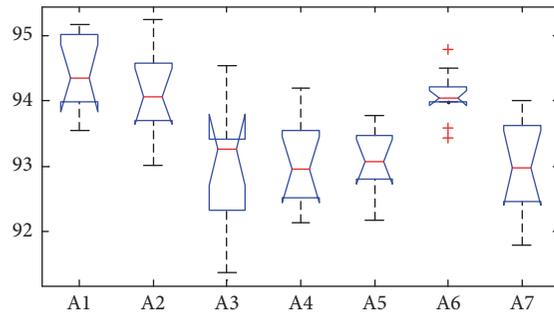


FIGURE 8: Nonparametric Kruskal-Wallis significance test results.

the classical ReliefF method; the fourth line ( $A_4$ ) represents the feature set selected by MIV; the fifth row ( $A_5$ ) and the sixth row ( $A_6$ ) are the high-dimensional feature sequences obtained by dimension reduction using KPCA and LLE, respectively, where the dimension reduction target is set to be same as the feature numbers determined by CRSR. The last row ( $A_7$ ) is the collection of original features. Taking the feature combinations mentioned above as input, the average and variance of fault diagnosis accuracy of hydraulic servo system calculated are shown in Figure 7, respectively.

According to the RBF network fault diagnostic results from Table 5 and Figure 7, the following can be obtained.

(1) Compared with using the original feature set directly, the feature selection methods are able to improve the precision of fault diagnosis based on RBF classifier, as the averages of the 10-fold cross validation from  $A_1$  to  $A_7$  are 94.405%, 94.217%, 93.022%, 93.045%, 93.077%, 94.069%, and 92.983%, respectively. The result indicates that the feature selection process plays a positive role in the classification task using high-dimensional data.

(2) Compared with the original ReliefF method, the introduction of compressed random subspace method eliminates those features with unstable contribution in the process of subspace integration and improves the performance and efficiency of feature selection. Eventually, higher fault diagnosis accuracy can be achieved with fewer input features.

(3) The variances calculated for  $A_1$  (with redundancy analysis) and  $A_2$  (without redundancy analysis) are 0.29 and 0.37, respectively. It can be seen that although the average diagnostic accuracy of them is close, the redundancy analysis

can optimize the information repetition among the features and reduce the computational resource consumption.

(4) Compared with the MIV-based feature selection method, CRSR has a great advantage in the diagnosis performance, which indicates that it has higher confidence in feature selection. Compared with KPCA and LLE, although the latter two (especially the LLE reduction algorithm) achieve high diagnostic accuracy as well, the variance of the 10-fold cross validation results for KPCA and LLE is higher than that of CRSR (0.29 and 0.31, resp.), which indicates that a certain degree of volatility exists in the results obtained by dimension reduction algorithms. In addition, due to the lack of clear interpretability and the existence of ambiguity in optimal target parameter setting, it is difficult to determine the best reduction method among the dimension reduction algorithms. Thus adaptive distance metrics based CRSR method has better applicability.

To further illustrate the rationality of CRSR method, the diagnostic results were tested by rank-based nonparametric Kruskal-Wallis test, which is based on fault diagnosis accuracy as a basic indicator in Table 5. The Kruskal-Wallis test was used to determine whether the different features sets have significant influence on the diagnostic performance. The significance threshold is set to 0.05, and the test results returned by Kruskal-Wallis function are shown in Figure 8. The  $p$  value returned by Kruskal-Wallis function is 0.00002 which is far less than 0.05. It indicates that the different combinations of features make significant impact on the fault diagnosis performance for hydraulic servo system.

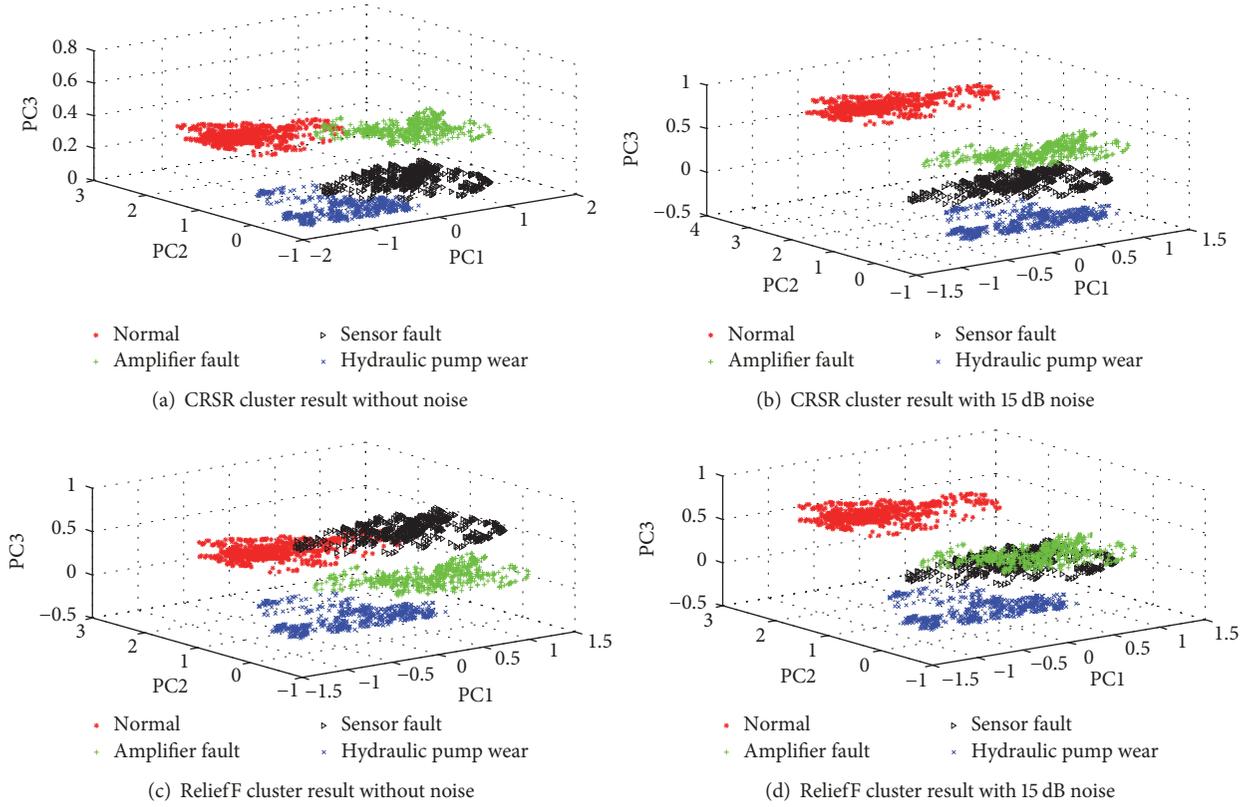


FIGURE 9: Fault feature vector clustering results under different noise conditions.

Moreover, another improvement of CRSR is that the target optimization function is constructed based on the information entropy and the fuzzy theory. Compared with the distance measurement method used in the classical ReliefF model, the robustness of the feature selection process is improved in a complex environment. To illustrate the improvement of CRSR mentioned above, this study analyzes the original output signal and the output signal with 15 dB noise. Meanwhile, PCA is used to project two sets of the selected feature sequences into three-dimensional space. Furthermore, the  $K$ -means method is used to cluster the feature sets of different fault modes. The results are shown in Figure 9.

It can be seen from Figure 9 that the feature qualities selected by CRSR and ReliefF are higher before the noise is added, and the feature sets corresponding to each failure mode can be clearly distinguished by  $K$ -means method. After adding 15 dB of noise, due to the influence of external disturbances, the data points of the amplifier gain fault and the sensor constant deviation fault are obviously confused in the clustering result of ReliefF, which indicates that the feature differentiation declines. However, in the result of CRSR method, spatial distribution boundaries between the four fault modes are still identifiable, which proves that the proposed objective function optimization method has good robustness against noise. Thus, CRSR method is a promising technique for feature selection and subsequent fault diagnosis of hydraulic servo systems.

#### 4. Conclusion

This study presents a fault diagnosis scheme for hydraulic servo system using compressed random subspace based ReliefF (CRSR) method. Based on the feature selection structure of ReliefF, the proposed CRSR method employs feature integration rules in the compressed domain and substitutes information entropy and fuzzy membership for traditional distance measure index. The advantage of the proposed method lies in the ability of determining the feature set with the better generalization performance and the less resource consumption. As a data-driven method, CRSR could be practical and flexible in engineering.

To demonstrate the effectiveness of the proposed CRSR method, validation data of three fault modes is generated through a hydraulic servo system joint simulation model. Comparing with existing feature reduction and feature selection methods, the result indicates that the feature selection process plays a positive role in the classification task using high-dimensional data, and CRSR based fault diagnosis method has higher average accuracy and smaller variance. Meanwhile, the compressed random subspace method can eliminate those features with unstable contribution in the process of subspace integration and improve the performance and efficiency of feature selection. Besides, due to the robustness and stability of the information entropy and fuzzy theory based objective function optimization, the result shows that CRSR method is more suitable for fault diagnosis problem under noisy conditions.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research was supported by the National Natural Science Foundation of China [Grant nos. 51605014, 51105019, and 51575021], the Technology Foundation Program of National Defense [Grant no. Z132013B002], and the Fundamental Research Funds for the Central Universities [Grant no. YWF-16-BJ-J-18].

## References

- [1] C. Lu, H. Yuan, L. Tao, and H. Liu, "Performance assessment of hydraulic servo system based on bi-step neural network and autoregressive model," *Journal of Vibroengineering*, vol. 15, no. 3, pp. 1546–1559, 2013.
- [2] X. He, "Fault diagnosis approach of hydraulic system using FARX model," in *Proceedings of the 2011 International Conference on Advanced in Control Engineering and Information Science, CEIS 2011*, pp. 949–953, China, August 2011.
- [3] H. Liu, X. Wang, and C. Lu, "Rolling bearing fault diagnosis under variable conditions using Hilbert-Huang transform and singular value decomposition," *Mathematical Problems in Engineering*, vol. 2014, Article ID 765621, 10 pages, 2014.
- [4] Z. Wang, C. Lu, Z. Wang, and J. Ma, "Health assessment of rotary machinery based on integrated feature selection and Gaussian mixed model," *Journal of Vibroengineering*, vol. 16, no. 4, pp. 1753–1762, 2014.
- [5] S.-Q. Ren, Y.-W. Fu, X. Li, and Z.-W. Zhuang, "Feature selection based on classes margin," *Journal of Software*, vol. 19, no. 4, pp. 842–850, 2008.
- [6] M. Van and H.-J. Kang, "Two-stage feature selection for bearing fault diagnosis based on dual-tree complex wavelet transform and empirical mode decomposition," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 230, no. 2, pp. 291–302, 2016.
- [7] T. W. Rauber, F. de Assis Boldt, and F. M. Varejão, "Heterogeneous feature models and feature selection applied to bearing fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 1, pp. 637–646, 2015.
- [8] M. Cerrada, R. V. Sánchez, D. Cabrera, G. Zurita, and C. Li, "Multi-stage feature selection by using genetic algorithms for fault diagnosis in gearboxes based on vibration signal," *Sensors*, vol. 15, no. 9, pp. 23903–23926, 2015.
- [9] K. H. Hui, C. S. Ooi, M. H. Lim, M. S. Leong, and S. M. Al-Obaidi, "An improved wrapper-based feature selection method for machinery fault diagnosis," *PloS one*, vol. 12, no. 12, 2017.
- [10] J. Lin and L. Qu, "Feature extraction based on morlet wavelet and its application for mechanical fault diagnosis," *Journal of Sound and Vibration*, vol. 234, no. 1, pp. 135–148, 2000.
- [11] I. K. Fodor, A survey of dimension reduction techniques. *Neoplasia* 7.5 (2002): 475-485.
- [12] H. Peng, X. Chen, and W. Xu, "Application of PCA feature extraction and SVM multi-classification on sensor fault diagnosis," *Journal of Data Acquisition and Processing*, vol. 25, no. 1, pp. 111–116, 2010.
- [13] Z.-B. Zhu and Z.-H. Song, "A novel fault diagnosis system using pattern classification on kernel FDA subspace," *Expert Systems with Applications*, vol. 38, no. 6, pp. 6895–6905, 2011.
- [14] D. E. Goldberg, *Genetic Algorithm in Search, Optimization, and Machine Learning*, vol. xiii.7, 1989.
- [15] I. Iguyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
- [16] N. Spolaôr, E. A. Cherman, M. C. Monard, and H. D. Lee, "ReliefF for multi-label feature selection," in *Proceedings of the 2nd Brazilian Conference on Intelligent Systems, BRACIS 2013*, pp. 6–11, Brazil, October 2013.
- [17] Y. Huang, P. J. McCullagh, and N. D. Black, "An optimization of ReliefF for classification in large datasets," *Data & Knowledge Engineering*, vol. 68, no. 11, pp. 1348–1356, 2009.
- [18] I. Kononenko, M. Robnik, and S. U. Pompe, "ReliefF for estimation and discretization of attributes in classification, regression, and ILP problems," *Artificial Intelligence: Methodology, Systems, Applications*, 1996.
- [19] L.-X. Zhang, J.-X. Wang, Y.-N. Zhao, and Z.-H. Yang, "A novel hybrid feature selection algorithm: using ReliefF estimation for GA-Wrapper search," in *Proceedings of the International Conference on Machine Learning and Cybernetics*, vol. 1, pp. 380–384, IEEE, Xi'an, China, November 2003.
- [20] F. Yang, W. Cheng, R. Dou, and N. Zhou, "An improved feature selection approach based on ReliefF and Mutual Information," in *Proceedings of the 2011 International Conference on Information Science and Technology, ICIST 2011*, pp. 246–250, China, March 2011.
- [21] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, 1998.
- [22] T. K. Ho, "Nearest neighbors in random subspaces," in *Advances in pattern recognition (Sydney, 1998)*, vol. 1451 of *Lecture Notes in Comput. Sci.*, pp. 640–648, Springer, Berlin, 1998.
- [23] R. Duangsoithong and T. Windeatt, "Relevance and Redundancy Analysis for Ensemble Classifiers," in *Machine Learning and Data Mining in Pattern Recognition*, vol. 5632 of *Lecture Notes in Computer Science*, pp. 206–220, Springer, Berlin, Germany, 2009.
- [24] I. Kononenko, "Analysis and extension of RELIEF," in *Proceedings of the The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, 1994.
- [25] Y. Sun, "Iterative RELIEF for feature weighting: algorithms, theories, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1035–1051, 2007.
- [26] N. Karayiannis, "MECA: maximum entropy clustering algorithm," in *Proceedings of the 1994 IEEE 3rd International Fuzzy Systems Conference*, pp. 630–635, Orlando, FL, USA.
- [27] D.-M. Zhao, J.-H. Wang, W. U. Jing, and J.-F. Ma, "Using fuzzy logic and entropy theory to risk assessment of the information security," in *Proceedings of the International Conference on Machine Learning and Cybernetics (ICMLC '05)*, vol. 4, pp. 2448–2453, IEEE, Guangzhou, China, August 2005.
- [28] R.-P. Li and M. Mukaidono, "Maximum-entropy approach to fuzzy clustering," in *Proceedings of the IEEE International Conference on Fuzzy Systems, 1995. International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and the Second International Fuzzy Engineering Symposium IEEE Xplore*, vol. 4, pp. 2227–2232, March 1995.

- [29] W. Xizhao and S. An, "Research on learning weights of fuzzy production rules based on maximum fuzzy entropy," *Journal of Computer Research Development*, vol. 43, no. 4, pp. 673–678, 2006.
- [30] P. Langley, "Selection of Relevant Features in Machine Learning," in *Proceedings of the Aaai Fall Symposium on Relevance*, 144, 140 pages, 1994.
- [31] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [32] K. Kira and L. A. Rendell, "Feature selection problem: traditional methods and a new algorithm," in *Proceedings of the Proceedings Tenth National Conference on Artificial Intelligence - AAAI-92*, pp. 129–134, July 1992.
- [33] H. Liu, J. Li, and L. Wong, "A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns," *Genome Informatics*, vol. 13, pp. 51–60, 2002.
- [34] T. Li, C. Zhang, and M. Ogihara, "A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression," *Bioinformatics*, vol. 20, no. 15, pp. 2429–2437, 2004.
- [35] M. H. Nguyen and F. De la Torre, "Optimal feature selection for support vector machines," *Pattern Recognition*, vol. 43, no. 3, pp. 584–591, 2010.
- [36] L. Jiang and J. J. Zhang, "Dynamic Characteristics Simulation of Hydraulic Position Servo-system Based on AMESim," *Mechanical Engineering Automation*, 2007.
- [37] L. Wan X, B. Ding H, and Z. Zhou, "Study on dynamical simulation of hydraulic servo system based on AMESim and simulink," *Coal Mine Machinery*, 2007.
- [38] J. Cheng, W. Liu, and Z. Zhang, "Modeling and simulation for the electro-hydraulic servo system based on Simulink," in *Proceedings of the 2011 International Conference on Consumer Electronics, Communications and Networks, CECNet 2011*, pp. 466–469, China, April 2011.

