

Research Article

Fault Diagnosis Method of Check Valve Based on Multikernel Cost-Sensitive Extreme Learning Machine

Jun Ma,¹ Jiande Wu,^{2,3} and Xiaodong Wang^{2,3}

¹Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming 650500, China

²Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China

³Engineering Research Center for Mineral Pipeline Transportation YN, Kunming 650500, China

Correspondence should be addressed to Jiande Wu; wjiande@kmust.edu.cn

Received 7 July 2017; Accepted 8 November 2017; Published 28 December 2017

Academic Editor: Yanan Li

Copyright © 2017 Jun Ma 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.

Check valve is one of the most important components and most easily damaged parts in high pressure diaphragm pump, which is a typical representative of reciprocating machinery. In order to ensure the normal operation of the pump, it is necessary to monitor its running state and diagnose fault. However, in the fault diagnosis of check valve, the classification models with single kernel function can not fully interpret the classification decision function, and meanwhile unreasonable assumption of diagnostic cost equalization has a significant impact on classification results. Therefore, the multikernel function and cost-sensitive mechanism are introduced to construct the fault diagnosis model of check valve based on the multikernel cost-sensitive extreme learning machine (MKL-CS-ELM) in this paper. The comparative test results of check valve for high pressure diaphragm pump show that MKL-CS-ELM can obtain fairly or slightly better performance than ELM, CS-ELM, MKL-ELM, and multikernel cost-sensitive support vector learning machine (MKL-CS-SVM). At the same time, the presented method can obtain very high accuracy under imbalance datasets condition and effectively overcome the weakness of diagnostic cost equalization and improve the interpretability and reliability of the decision function of classification model. It, therefore, is more suitable for the practical application.

1. Introduction

High pressure diaphragm pump is the most important equipment for high concentration slurry pipeline transportation. Its working condition is directly related to whether the pump can be restarted after stopping and whether it will produce accelerated flow in batch transportation. Check valve is the core and the easiest damaged component of the high pressure diaphragm pump. In order to ensure the normal operation of the pump, it is necessary to monitor its running state and diagnose fault [1]. So, the research of condition monitoring and fault diagnosis of the high pressure diaphragm pump has important practical significance in promoting development of slurry pipeline transportation field.

However, the fault characteristics of reciprocating machinery are difficult to extract because of its complex structure, multiple excitation sources, unstable operation, and so on [2]. In order to complete the condition monitoring

and fault diagnosis of reciprocating machineries effectively, both domestic and foreign scholars have introduced the fault diagnosis methods of rotating machinery into the fault diagnosis of reciprocating machinery and made many valuable research results [3–5]. Ogle and Morrison [6] analyzed the failure accident of diaphragm pump and found that the environmental stress cracking of diaphragm is one of the main reasons for the diaphragm pump failure. The research results have provided effective theoretical support for accident prevention and pipeline maintenance and greatly reduced maintenance costs. In recent years, wavelet transform and Fourier transform, information entropy, neural network, bispectrum analysis, feature fusion, evidence theory, chaos theory, fractal theory, decision tree, and SVM have been widely applied to the fault diagnosis of reciprocating machinery, and many significant research achievements have been obtained [7–16]. Yet compared with the fault diagnosis of rotating machinery, there are still many research

contents to be improved: (1) the data sample size of reciprocating machinery is huge and a great deal of multisource heterogeneous information is held within them due to the influence of complex structure, multiple excitation source, multiple wearing parts, coupling of the signal, and strong nonlinearity of reciprocating machinery. It is not reasonable to use a single kernel function (such as radial basis function kernel, polynomial kernel function) for processing all the samples, and it is unable to explain the signal completely. Consequently, it is an inevitable choice to combine multiple kernel functions to achieve better processing results [17–19]. (2) It is impossible for fault diagnosis models to get ideal classification results when datasets of fault diagnosis are not balanced (the fault samples are far less than the normal samples) and the diagnostic cost is unequal (e.g., the diagnostic cost between “the normal state which is identified as a fault state” and “the fault state which is identified as a normal state” is quite different; the former will only result in an “invalid” examining and repair for operator, but the latter will result in major safety incidents), so the hypothesis deficiency of minimum classification error and diagnostic cost equalization in the existing classification model need to be overcome [20]. (3) At present, BP neural network and SVM are relatively mature classification learning methods and play important role in the fault diagnosis of reciprocating machinery. But, BP neural network has the problems of easily falling into local minimum, being not convergence, and so on. Meanwhile, the optimization calculation load of SVM increases with the optimization parameters and data sample size. And many parameters will be optimized to get the optimization SVM classification model. It, therefore, is one of the hot topics to explore new classification method which has the advantages of fast training speed and fewer optimization parameters to obtain global optimal solution [21].

In recent years, ELM is widely used because of its effectiveness, high speed, being easy for implementation, and multiclassification in the related fields of machine learning [22–24]. Moreover, the modified ELM models can validly solve the problems of imbalance sample and obtain better performance [25–28]. Therefore, the modified ELM methods have become the main research direction. For one thing, the transfer function of the original hidden layer based on random feature mapping will be substituted for the more efficient transfer functions. Then, the sigmoid function and radial basis function (RBF) [29–32] which are widely used in neural networks have been introduced into ELM and obtained better experiments results. For another thing, how to improve classification performance of ELM under multisource heterogeneous data and information fusion is also one of the latest research trends of the modified ELM classification models. Liu et al. [33] proposed the multikernel ELM (MKL-ELM) combined with the multikernel learning with ℓ_p constraints. Compared with traditional ELM, the MKL-ELM can solve these issues, including the selection and optimization of multikernel function, the application of multisource heterogeneous data processing method, and information fusion method in the classification. But, in [33], the researcher does not consider the impact of classification cost on the classification model. So, the cost-sensitive mechanism

was introduced into the conventional ELM [34], and a new classification model based on cost-sensitive is proposed to conquer the drawback of diagnostic cost equalization. But, it is not very effective in dealing with the multisource heterogeneous data and information fusion because of the restriction of single and permanent kernel during the subsequent processing.

With the intensive study of ELM theory and application, the MKL-ELM and CS-ELM have greatly promoted the development of ELM. But there is still plenty of room for improvement and extension. This is typically shown in two aspects: (1) how to select the most appropriate cost-sensitive method; (2) how to construct more general multikernel function which can be widely used in fault diagnosis field. Based on the points discussed above, the multikernel function and cost-sensitive mechanism are introduced into ELM to construct the fault diagnosis model based on MKL-CS-ELM for check valve of high pressure diaphragm pump in this paper.

This paper has the following main contributions. First, the advantages, shortcomings, and the application ranges of oversampling, undersampling, and threshold adjusting are analyzed to provide theoretical support for the choice of cost-sensitive methods. Second, a new fault diagnosis method based on MKL-CS-ELM is proposed to diagnose the check valve faults of high pressure diaphragm pump. Third, the comparison experiments of ELM, CS-ELM, MKL-ELM, MKL-CS-SVM, and MKL-CS-ELM are carried out, and the effectiveness of the proposed MKL-CS-ELM method is verified.

The remainder of this paper is organized as follows. Section 2 describes the fundamental theory of ELM, MKL-ELM, cost-sensitive learning, and evaluation index of classification model. Section 3 presents the implementation process of the proposed method in detail. Section 4 elaborates experimental process. Section 5 shows the experimental results analysis. Section 6 offers the discussion and conclusion.

2. Related Work

2.1. Extreme Learning Machine (ELM). From the classification optimization point of view, the principle of ELM is similar to SVM and LSSVM, whose goal is to obtain the minimum training error and maximum classification margin or generalization ability. So, on the basis of SVM principle analysis, the optimized mathematical model of ELM is described as follows [35]:

$$\begin{aligned} \min_{\beta, \xi} \quad & \frac{1}{2} \|\beta\|_F^2 + \frac{C}{2} \|\xi\| \\ \text{s.t.} \quad & \beta' \phi(x_i) = y_i - \xi_i, \quad \forall i. \end{aligned} \quad (1)$$

In (1), $\beta \in \mathfrak{R}^{|\phi(\cdot) \times T|}$ stands for the connecting weighting coefficients between hidden layer and output layer, $\|\cdot\|_F$ is Frobenius norm, C represents regularization parameter or penalty factor which achieves the balance between the minimum training error and maximum classification margin, $\xi_i = [\xi_{1i}, \xi_{2i}, \dots, \xi_{Ti}]'$, ($1 \leq i \leq n$) is the i th column of the error matrix $\xi \in \mathfrak{R}^{T \times n}$, β' represents the transpose of

matrix β (similarly hereinafter), $\phi(x_i)$ is the output function of hidden layer for the input neuron x_i , $\{(x_i, y_i)\}_{i=1}^n$ represents the given training set, and $y_i = [0, \dots, 0, 1, 0, \dots, 0]^t \in \{0, 1\}^l$ represents the case that sample x_i belongs to the classification label t . n and T are the number of training samples and categories, respectively. According to KKT (Karush Kuhn Tucker) theory, the analysis solution of (1) is calculated and the detailed solution process can be read in [36]. The solution of the output weight β^* is solved using the Moore-Penrose $\widehat{\Phi}^+$:

$$\begin{aligned} \beta^* &= \widehat{\Phi}^+ Y = \Phi^T \left(\frac{I}{C} + \Phi \Phi^T \right)^{-1} Y' \quad n < l, \\ \beta^* &= \widehat{\Phi}^+ Y = \left(\frac{I}{C} + \Phi \Phi^T \right)^{-1} \Phi^T Y' \quad l < n. \end{aligned} \quad (2)$$

In (2), the output matrix of the hidden layer is $\Phi = [\phi(X_1), \dots, \phi(X_n)]^t \in \mathfrak{R}^{n \times |\phi(\cdot)|}$, the output result of ELM classification model is $Y = [y_1, \dots, y_n] \in \mathfrak{R}^{T \times n}$, and I is identity matrix.

For a given new sample x , the output decision function $f(x)$ of the ELM is shown as follows:

$$f(x) = \beta^{*t} \phi(x). \quad (3)$$

2.2. Multikernel Extreme Learning Machine (MKL-ELM). The common definition of multikernel function is the linear combination of basic kernel function. So, the combination coefficient of optimal kernel function and the maximum margin of ELM are the key and core of the MKL-ELM [33]. A typical form of multikernel function is shown in

$$\kappa(\cdot, \cdot; \gamma) = \sum_{p=1}^m \gamma_p \kappa_p(\cdot, \cdot). \quad (4)$$

In (4), $\{\kappa_p(\cdot, \cdot)_{p=1}^m\}$ represents m basic kernel functions.

For the convenience of processing and computing, the combination coefficients γ_p of basic kernel function satisfy restricting condition $\sum_{p=1}^m \gamma_p = 1$. The feature mapping of (4) is shown in

$$\phi(\cdot; \gamma) = [\sqrt{\gamma_1} \phi_1(\cdot), \sqrt{\gamma_2} \phi_2(\cdot), \dots, \sqrt{\gamma_m} \phi_m(\cdot)]. \quad (5)$$

In (5), $\phi(\cdot, \gamma)$ and $\{\phi(\cdot)_{p=1}^m\}$ are the high dimensional feature mapping of $\kappa(\cdot, \cdot; \gamma)$ and $\{\kappa_p(\cdot, \cdot)_{p=1}^m\}$, respectively.

In the construction process of multikernel function, RBF kernel function, Laplace kernel function, and inverse-distance kernel function are selected as basic kernel functions.

RBF kernel function: $k(x_i, x_j)$

$$= \exp\left(\frac{-\|x_i - x_j\|^2}{\sigma}\right)$$

Laplace kernel function: $k(x_i, x_j)$

$$= \exp\left(\frac{-\|x_i - x_j\|}{\sqrt{\sigma}}\right)$$

Inverse-distance kernel function: $k(x_i, x_j)$

$$= \frac{1}{(\|x_i - x_j\| / \sqrt{\sigma} + 1)}.$$

(6)

In (6), σ is the parameter of kernel function. In this paper, the value of σ is calculated by $\sigma = 2^t \sigma_0$ ($t \in \{-2, -1, 0, 1, 2\}$). At the same time, σ_0 represents the mean Euclidean distance between the samples.

In order to insure that the final solution and combination kernel function of multikernel optimal problem are subject to the boundedness and symmetric positive semidefinite, respectively, the ℓ_q norm is used as the constraint condition of the combination coefficient γ of the multikernel function. The different value of q in the ℓ_q norm represents different constraint norm. According to the theoretical basis of multikernel SVM [37, 38] and (5), the theoretical expression of conventional MKL-ELM is described as follows:

$$\begin{aligned} \min_{\gamma} \min_{\beta, \xi} \quad & \frac{1}{2} \|\beta\|_F^2 + \frac{C}{2} \sum_{i=1}^n \|\xi_i\|^2 \\ \text{s.t.} \quad & \beta^t \phi(x_i; \gamma) = y_i - \xi_i, \quad \forall i = 1, \dots, n, \\ & \sum_{p=1}^m \gamma_p^q = 1, \quad \gamma_p \geq 0, \quad \forall p = 1, \dots, m. \end{aligned} \quad (7)$$

In (7), the connecting weighting coefficient β is $\beta = [\beta_1, \dots, \beta_m] \in \mathfrak{R}^{(|\phi_1(\cdot)| + \dots + |\phi_m(\cdot)|) \times T}$ and $\beta_p \in \mathfrak{R}^{(|\phi_p(\cdot)|) \times T}$ ($p = 1, \dots, m$) is the connecting weighting of the p th basic kernel function.

Substituting (5) into (7), the expression of MKL-ELM is obtained and shown in

$$\begin{aligned} \min_{\gamma} \min_{\beta, \xi} \quad & \frac{1}{2} \|\beta\|_F^2 + \frac{C}{2} \sum_{i=1}^n \|\xi_i\|^2 \\ \text{s.t.} \quad & \sum_{p=1}^m \sqrt{\gamma_p} \beta_p^t \phi_p(x_i) = y_i - \xi_i, \quad \forall i = 1, \dots, n, \\ & \sum_{p=1}^m \gamma_p^q = 1, \quad \gamma_p \geq 0, \quad \forall p = 1, \dots, m. \end{aligned} \quad (8)$$

If $\tilde{\beta} = [\tilde{\beta}_1, \tilde{\beta}_2, \dots, \tilde{\beta}_m]$ is equal to $[\sqrt{\gamma_1} \beta_1, \sqrt{\gamma_2} \beta_2, \dots, \sqrt{\gamma_m} \beta_m]$, then (8) can be simplified to

$$\begin{aligned} \min_{\gamma} \min_{\beta, \xi} \quad & \frac{1}{2} \sum_{p=1}^m \frac{\|\tilde{\beta}_p\|_F^2}{\gamma_p} + \frac{C}{2} \sum_{i=1}^n \|\xi_i\|^2 \\ \text{s.t.} \quad & \sum_{p=1}^m \tilde{\beta}_p^t \phi_p(x_i) = y_i - \xi_i, \quad \forall i = 1, \dots, N, \\ & \sum_{p=1}^m \gamma_p^q = 1, \quad \gamma_p \geq 0, \quad \forall p = 1, \dots, m. \end{aligned} \quad (9)$$

Equation (9) is similar to the expression of ELM. So, the Lagrangian function of MKL-ELM can be calculated:

$$L(\tilde{\beta}, \xi, \gamma) = \frac{1}{2} \sum_{p=1}^m \frac{\|\tilde{\beta}_p\|_F^2}{\gamma_p^q} + \frac{C}{2} \sum_{i=1}^n \|\xi_i\|^2 - \sum_{t=1}^T \sum_{i=1}^n \alpha_{it} \left(\sum_{p=1}^m \tilde{\beta}_p' \phi_p(x_i) - y_{ti} - \xi_{ti} \right) + \tau \left(\sum_{p=1}^m \gamma_p^q - 1 \right). \quad (10)$$

In (10), $\alpha \in \mathfrak{R}^{n \times T}$ and τ are the Lagrangian multiplier. Then, KKT optimization condition is calculated and shown in

$$\tilde{\beta}_p = \gamma_p^q \sum_{t=1}^T \sum_{i=1}^n \alpha_{it} \phi_p(x_i), \quad \forall p, \quad \xi_{ti} = \frac{\alpha_{ti}}{C}, \quad \forall t, \forall i, \quad (11)$$

$$\sum_{p=1}^m \tilde{\beta}_p' \phi_p(x_i) = y_i - \xi_i.$$

The matrix form of (11) is expressed in

$$\left(\kappa(\cdot, \cdot; \gamma) + \frac{I}{C} \right) \alpha = Y'. \quad (12)$$

In (12), the compound kernel function $\kappa(\cdot, \cdot; \gamma)$ represents $\kappa(x_i, x_j; \gamma) = \phi(x_i; \gamma)' \phi(x_j; \gamma) = \sum_{p=1}^m \gamma_p K(x_i, x_j)$. Then, the solution of α is shown as follows:

$$\alpha = \left(\kappa(\cdot, \cdot; \gamma) + \frac{I}{C} \right)^{-1} Y'. \quad (13)$$

At the same time, the combination coefficient γ of multi-kernel function can be calculated by the derivative of

$$\gamma_p^{\text{new}} = \frac{\|\tilde{\beta}_p\|_F^{2/(1+q)}}{\left(\sum_{p=1}^m \|\tilde{\beta}_p\|_F^{2q/(1+q)} \right)^{1/q}}, \quad \forall p. \quad (14)$$

The sparse MKL-ELM constrained ℓ_1 norm is given by $q = 1$ in (14). The optimal parameter of α^* and γ^* is calculated by iterative optimization methods. Now, for a given sample x , the output decision function $f(x)$ of MKL-ELM is expressed as follows:

$$f(x) = [f_1(x), \dots, f_T(x)]. \quad (15)$$

In (15), the component of $f(x)$ represents $f_c(x) = \sum_{i=1}^n \alpha_{ic}^* \sum_{p=1}^m K_p(x_i, x)$, $1 \leq c \leq T$.

2.3. Cost-Sensitive Methods. The cost-sensitive methods largely fall into three groups [39]: constructing the cost-sensitive classification model directly, establishing the cost-sensitive classification model using the Bayesian risk theory, and building the cost-sensitive classification model by

TABLE 1: The cost matrix of binary classification.

Class attributes	True positive	True negative
Predict positive	$C(1, 1)$ or T_p	$C(1, 0)$ or F_p
Predict negative	$C(0, 1)$ or F_N	$C(0, 0)$ or T_N

changing the samples distribution. The latter two methods are emphatically introduced [40].

Assuming that the number of given the class labels of training set is C and the number of training samples in each category is N_i , the classification cost is defined as follows.

(1) $\text{Cost}[i, c]$ ($i, c \in \{1, \dots, C\}$) is described as the misclassification cost where the category I is misclassified as category c . Then, $\text{Cost}[i, i] = 0$ is obvious.

(2) $\text{Cost}[i]$ ($i \in \{1, \dots, C\}$) represents the total cost function of category i , namely, $\text{Cost}[i] = \sum_{c=1}^C \text{Cost}[i, c]$.

The cost expressions of oversampling, undersampling, and threshold adjusting by definition are discussed as follows.

In oversampling and undersampling, the cost expression of N_k^* is defined by

$$N_k^* = \left[\frac{\text{Cost}[k]}{\text{Cost}[\lambda]} N_\lambda \right], \quad (16)$$

where N_k^* is the number of categories k . λ represents the resample category of oversampling and undersampling, which is calculated by (17) and (18), respectively:

$$\lambda = \arg \min_j \frac{(\text{Cost}[j] / \min_c \text{Cost}[c]) N_{\arg \min_c \text{Cost}[c]}}{N_j}, \quad (17)$$

$$\lambda = \arg \max_j \frac{(\text{Cost}[j] / \max_c \text{Cost}[c]) N_{\arg \max_c \text{Cost}[c]}}{N_j}. \quad (18)$$

However, the realization principle of threshold adjusting can be interpreted as follows:

$$O_i^* = \eta \sum_{c=1}^C O_i \text{Cost}[i, c]. \quad (19)$$

In (19), O_i ($i \in \{1, \dots, C\}$) is the actual output of different output nodes of ELM, and it satisfies constraint condition $\sum_{i=1}^C O_i = 1$, $0 < O_i < 1$. η is the normalization coefficient of $\sum_{i=1}^C O_i^* = 1$. At the same time, the output of threshold adjusting also satisfies constraint condition $\sum_{i=1}^C O_i^* = 1$, $0 < O_i^* < 1$.

2.4. The Evaluation Indicators of Classification Model. The evaluation indicators of binary classification and multiclassification are introduced to validate the effectiveness of the proposed method in the section.

2.4.1. The Cost-Sensitive Evaluation Indicators of Binary Classification. In binary imbalanced learning, the cost matrix is shown in Table 1. It is generally recognized that the cost of correct classification is defined as $T_p = T_N = 0$.

Based on Table 1, the cost-sensitive evaluation indicators of binary classification are defined as follows.

```

(1) Input:  $x, y, \{K_p\}_{p=1}^m, q, C$ 
(2) Output:  $f_c(x), \alpha, \gamma$ 
(3) Initialization:  $\gamma = \gamma^0, t = 0, S^* = x, S_k (\forall x_k \in S_k \subset x)$ 
% The resample of the cost-sensitive dataset
(4) Calculate  $N_k^*$  and  $\lambda$  based on Eq. (16) and (17)
(5) Judgement
(6) If  $N_k^* > N_k$ ,
(7) The number of resample in the  $S_k$  is  $N_k^* - N_k$ 
(8) Updating  $S^*$ 
(9) Updating the sample  $x \leftarrow S^*$ 
(10) End
% Construct the multi-kernel function
(11) Repeat
(12) Calculate  $K(\cdot, \cdot; \lambda) = \sum_{p=1}^m \gamma_p^t K_p$ .
(13) Updating  $\alpha^t$  based on Eq. (13).
(14) Updating  $\gamma^{t+1}$  based on Eq. (14).
(15) Iteration  $t = t + 1$ ;
(16) Until  $\max\{|\gamma^{t+1} - \gamma^t|\} \leq 1e - 4$ 
(17) Calculate the optimal output function  $f_c(x)$  of the MKL-CS-ELM based on Eq. (15).

```

ALGORITHM 1: Oversampling example.

The classification accuracy of positive samples (AP):

$$AP = \frac{|T_P|}{|T_P| + |F_N|}. \quad (20)$$

The classification accuracy of negative samples (AN):

$$AN = \frac{|T_N|}{|T_N| + |F_P|}. \quad (21)$$

Global classification accuracy (Accuracy):

$$Accuracy = \frac{|T_P| + |T_N|}{|T_P| + |F_N| + |T_N| + |F_P|}. \quad (22)$$

In (20)~(22), $|\cdot|$ represents module operation.

2.4.2. The Cost-Sensitive Evaluation Indicator of Multiclassification. The cost-sensitive evaluation indicator of multiclassification is more complicated than binary classification. The indicator of robustness r_α referred to in [41] is introduced to describe the classification performance in multiclassification. The robustness indicator r_α is calculated by

$$r_\alpha = \frac{AV_{Cost_\alpha}}{\max_i AV_{Cost_i}}. \quad (23)$$

In (23), AV_{Cost_α} is average cost of method α . $\max_i AV_{Cost_i}$ represents the maximum average cost of the designed method. The indicator of robustness r_α is lower, and the robust performance of the method is better.

3. Classification Method of Imbalance Sample Distribution Based on MKL-CS-ELM

The main procedure of proposed MKL-CS-ELM method involves data preprocessing (data normalization and feature extraction), construction of multikernel function, and

cost-sensitive learning. The brief process of MKL-CS-ELM is shown in Figure 1. The detailed process of the proposed method is described in Algorithms 1 and 2. The oversampling process refers to Algorithm 1 and the principle of undersampling is similar to oversampling. Algorithm 2 is the implementation process of threshold adjusting method.

4. Experimental Description

4.1. The Principle of Check Valve and Experiment Platform

4.1.1. The Principle of Check Valve. The check valve completed a process of feeding and discharging in every stroke of the diaphragm pump. Assume that the stroke coefficient of diaphragm pump is 50 r/min and the reciprocating action of inlet and outlet check valve will be 72000 times when it is in the normal operation for one day. Therefore, the check valve is core component of frequent motion in diaphragm pump, and it also turns into one of the most important reasons for the check valve failure. The high pressure diaphragm pump and the failure check valve for mineral slurry pipe transportation with solid-liquid two-phase flow are shown in Figure 2.

In Figure 2, the check valve of the high pressure diaphragm pump is a cone-valve and its simple structure is shown in Figure 3. And “spool-spring” forms a weakly damped oscillation system. There are two reasons for the vibration of the system: one is external factor (resonance); the other is caused by its own characteristics. When the frequency of the external excitation source is an integral multiple of the natural frequency of the valve system, the resonance of the whole system will occur during work. So, the different running states of the check valve can be effectively judged by analyzing the vibration signal of the check valve.

4.1.2. Vibration Data Acquisition Experiment Platform. Figure 4 is the experiment platform of check valve. The

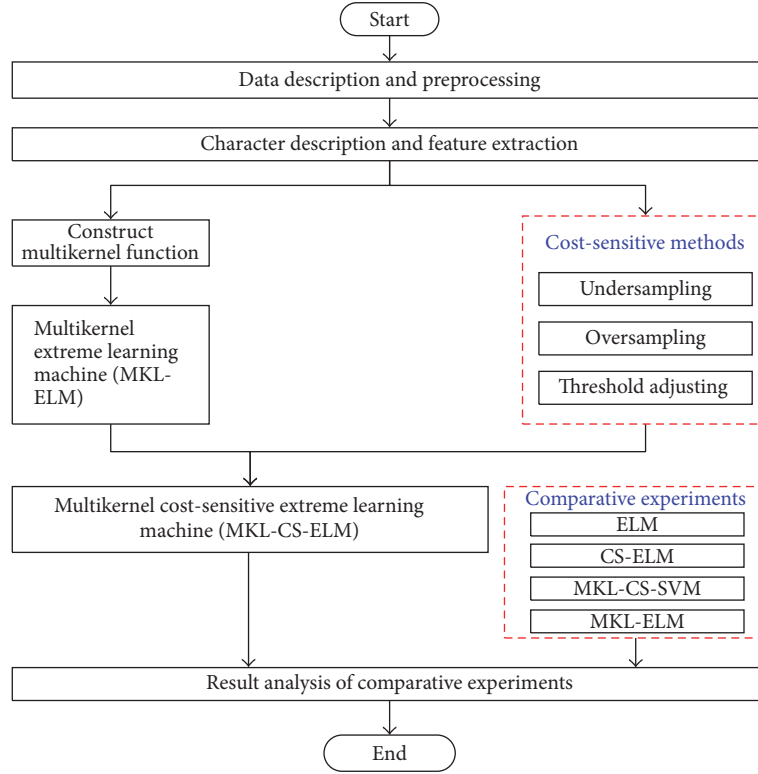


FIGURE 1: The realization process of MKL-CS-ELM.

- (1) Input: $x, \gamma, \{K_p\}_{p=1}^m, q, C$
- (2) Output: $f_c(x), \alpha, \gamma$
- (3) Initialization: $\gamma = \gamma^0, t = 0, S^* = x, S_k (\forall x_k \in S_k \subset x)$
% Construct the multi-kernel function
- (4) Repeat
- (5) Calculate $K(\cdot, \cdot; \lambda) = \sum_{p=1}^m \gamma_p^t K_p$.
- (6) Updating α^t based on Eq. (13).
- (7) Updating γ^{t+1} based on Eq. (14).
- (8) Iteration $t = t + 1$;
- (9) Until $\max\{\gamma^{t+1} - \gamma^t\} \leq 1e - 4$
- (10) Calculate the optimal output function $f_c(x)$ of the MKL-CS-ELM based on Eq. (15)
% Cost-sensitive Learning by Threshold Adjusting
- (11) Calculate the output O_i^* of MKL-CS-ELM based on the Eq. (19).
- (12) Calculate the optimal output $f_c(x) = \arg \max_i o_i^*$ of MKL-CS-ELM.

ALGORITHM 2: Threshold adjusting example.

three-cylinder diaphragm pump includes 3 pairs of check valves, which means that it includes 3 inlet check valves and 3 outlet check valves. So, in the process of data acquisition, the six PCB 352C33 accelerometers are installed on the check valve housing to collect vibration data by a PXI-3342. The data sampling frequency f_s is 2560 Hz and the data point N is 20480.

4.2. Experimental Setting. The data attributes of check valve and classification information are defined as in Table 2.

Based on the data characteristics in Table 2, the three kinds of cost matrixes are introduced and defined as follows [42].

(1)

$$\begin{aligned}
 & 1 < \text{Cost}[i, j] \leq 10 \quad j = c \cap j \neq i, \\
 & \text{Cost}[i, j \neq c] = 1 \quad j \neq i, \\
 & \text{Cost}[i] = \text{Cost}[i, c] \quad j \neq c, \\
 & \text{Cost}[c] = 1.
 \end{aligned} \tag{24}$$

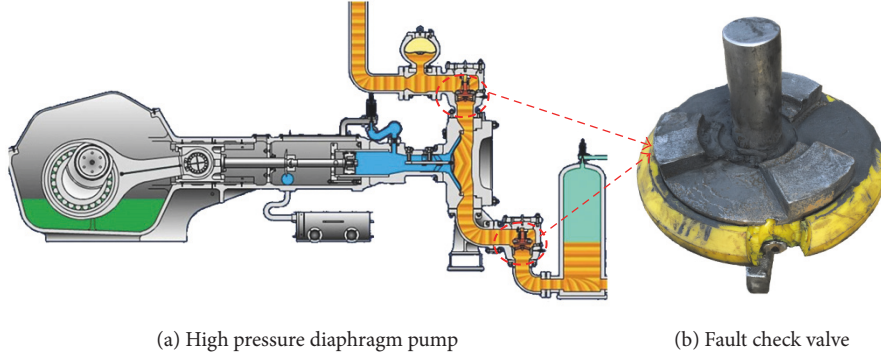


FIGURE 2: The high pressure diaphragm pump and fault check valve.

TABLE 2: The description of the experimental datasets.

Datasets	Data length	Attribute	Classes	Classification distribution
Check valve	20480	9	3	Normal condition (NC)/stuck valve fault (NK)/abrasion fault (NM)

(2)

$$\begin{aligned}
 1 \leq \text{Cost}[i, j] = H_i \leq 10 \quad j \neq i, \\
 \text{Cost}[i] = H_i, \\
 \exists H_i = 1.
 \end{aligned} \tag{25}$$

(3)

$$\begin{aligned}
 1 \leq \text{Cost}[i, j] \leq 10 \quad j \neq i, \\
 \text{Cost}[i] = \sum_{c=1}^C \text{Cost}[i, c], \\
 \exists \text{Cost}[i, j] = 1.
 \end{aligned} \tag{26}$$

4.3. The Feature Extraction of Wavelet Packet Energy Entropy.

Figure 5 shows the time and frequency waveform of the vibration signal for the check valve under 3 different operating conditions, including normal condition (NC), stuck valve fault (NK), and abrasion fault (NM). From point of the time domain and frequency domain waveform, it can be seen that the abnormal check valve has occurred, but further reasons or categories can not be obtained. In order to realize the automatic identification of the different running states of the check valve, it is necessary to extract the effective characteristics of the running state and then construct the state identification model.

The feature extraction makes full use of the advantage of wavelet packet and entropy in this paper. The third-layer wavelet packet energy distribution coefficient and energy entropy are extracted as characteristic parameters of the following classification model [42]. The selection of feature extraction method is based on the following points to consider.

(1) It is by using wavelet packet technique that the vibration signal of check valve can be mapped to wavelet-basis functions without information loss and has the superior ability in localization analysis of nonstationary signal.

(2) Entropy is introduced into depicting the operation state characteristics for check valve. This is mainly because the more disordered the system is, the greater the entropy becomes. And then, we can extract sensitive and transient features to describe the operation state of check valve.

The steps of feature extraction are listed below.

(1) Signal decomposition and reconstruction: the vibration signal of check valve is analyzed by three layers' wavelet packet transform to get the wavelet coefficients of the third-layer decomposition. In this paper, "db10 wavelet" is chosen as basic wavelet-basis function, which is mainly because "db10 wavelet" can well reflect the sensitive and transient features of vibration signal of check valve.

(2) Extraction feature vector: the wavelet packet energy distribution coefficient P_{3i} of reconstructed signals of the third-layer wavelet packets coefficients and energy entropy H compose the feature vector $T = [P_{30}, P_{31}, \dots, P_{37}, H]$. P_{3i} and H can be calculated as follows:

$$P_{3i} = \frac{\sqrt{E(a_{3i})}}{\sum_{i=1}^L \sqrt{E(a_{3i})}}, \tag{27}$$

$$H = - \sum_{i=1}^L p_{3i}^2 \log p_{3i}^2, \tag{28}$$

where L denotes the number of component signals ($L = 8$) and $E(a_{3i})$ represents the energy of the reconstruction signal of third-layer wavelet coefficients.

According to the definition of feature extraction in (27) and (28), the feature vectors of check valve can be calculated. Because of the limited space, partial features (not all features) are shown in Table 3. Compared to the normal check valve with the fault check valve, the operating conditions will be easily distinguished based on the wavelet packet energy distribution coefficient P and energy entropy H . It shows that the feature extraction method based on wavelet packet energy entropy is effective and reliable.

TABLE 3: The wavelet packet energy entropy features.

Fault class	Feature parameters								
	P_0	P_1	P_2	P_3	P_4	P_5	P_6	P_7	H
NC	0.4521	0.2296	0.0693	0.0663	0.0365	0.0304	0.0779	0.0379	0.5848
NM	0.8257	0.1458	0.0066	0.0115	0.0001	0.0001	0.0067	0.0035	0.3453
NK	0.9263	0.0464	0.0132	0.0046	0.0001	0.0002	0.0060	0.0032	0.1469

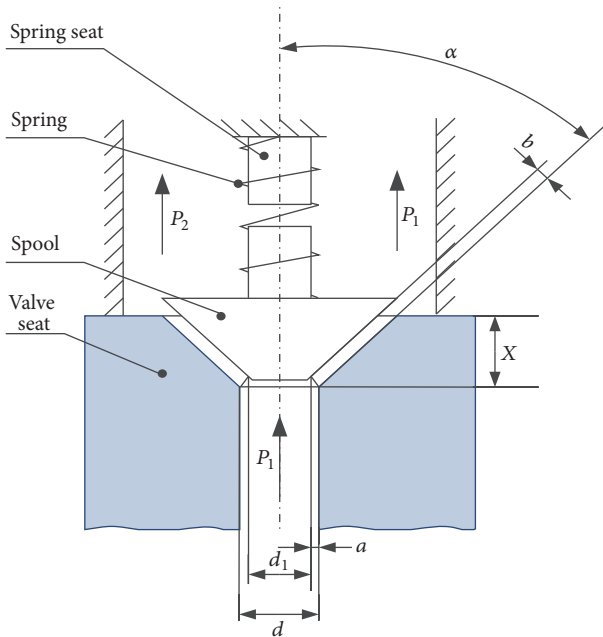


FIGURE 3: The structure diagram of cone-shaped check valve.

TABLE 4: The cost matrix of check valve.

Fault classes	NC	NK	NM
NC	0	1	1
NK	R	0	1
NM	R	1	0

5. Discussion of Experimental Results

Based on the definition of cost functions, the diagnosis cost matrix is constructed and shown in Table 4. The value of diagnostic cost R is from 1 to 5 ($R \in [1, 5]$) and increases by certain step length (usually 0.5) in the experiments.

In the experimental processing, the 110 data samples are collected, including 70 NC data samples, 20 NK data samples, and 20 NM data samples. The data samples of the check valve will be processed by combining the cost matrix shown in Table 4 with theoretical illustration of oversampling, undersampling, and threshold adjusting in Section 2.3. Then, the fault diagnosis classification models of ELM, CS-ELM, MKL-ELM, MKL-CS-ELM, and MKL-CS-SVM are constructed. The experimental results of binary classification and multiclassification for check valve are elaborated as follows in detail.

5.1. The Experimental Results Analysis of Binary Classification for Check Valve. In the binary classification experiments, the datasets of NC and NK are selected as the test data. The cost matrix is consistent with Table 4. The experimental results are described as follows.

5.1.1. The Experimental Results of Oversampling. The data sample distribution of oversampling is calculated and shown in Table 5 according to cost matrix in Table 4, (16), and (17). The 90 data samples are collected, 54 samples are selected as training samples, and the remaining 36 samples as test samples. Then the recognition results of classification models are presented in Figure 6.

As seen in Figure 6, some conclusions can be observed, including the following: (1) In the cost-sensitive processing of oversampling, the AP of CS-ELM, MKL-CS-SVM, and MKL-CS-SVM increases at first then decreases with increasing cost R , the AN increases at first then reaches steady state with increasing cost R , and the global classification accuracy (Accuracy) increases at first and then decreases with increasing cost R . (2) The recognition results of ELM and MKL-ELM method do not change with increasing cost R , which is mainly because the data distribution of the mentioned ELM and MKL-ELM does not also change. Therefore, it is only for the comparison of experimental results and independent of the diagnostic cost R . (3) In CS-ELM, MKL-CS-SVM, and MKL-CS-ELM method, the optimal recognition effect is obtained when the diagnostic cost is $R = 2.5$. (4) Compared with the ELM and MKL-ELM methods without diagnostic cost, the diagnostic cost can improve the accuracy and reliability of classification models in CS-ELM, MKL-CS-SVM, and MKL-CS-SVM method. (5) From the experimental results, we can also see that the multikernel learning mechanism is also helpful to further improve the diagnostic performance of the classification models. At the same time, Figure 6 also shows CS-ELM and MKL-CS-ELM are more sensitive to the cost R than the MKL-CS-SVM.

5.1.2. The Experimental Results of Undersampling. The data sample distribution of undersampling is calculated based on the Table 4, (16), and (18). Then, the recognition results of the above-mentioned classification models are displayed in Figure 7.

As shown in Figure 7, some conclusions can be obtained, which are similar to the results of oversampling methods. Moreover, the classification performance of mentioned classification models for check valve is slightly poor in undersampling. The experimental results found that the major



FIGURE 4: The experiment platform of check valve.

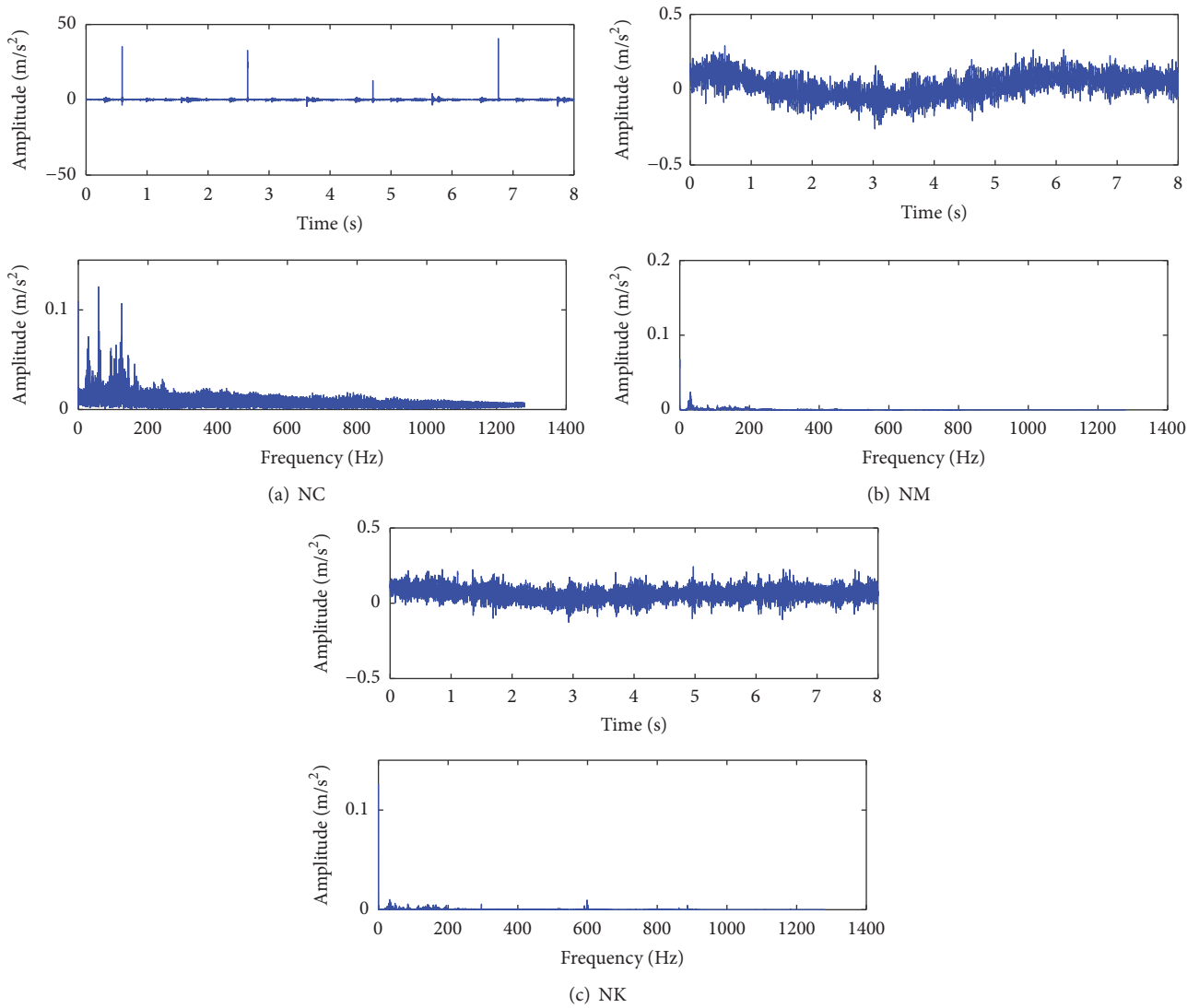


FIGURE 5: The analysis results of time domain and frequency domain for different work conditions.

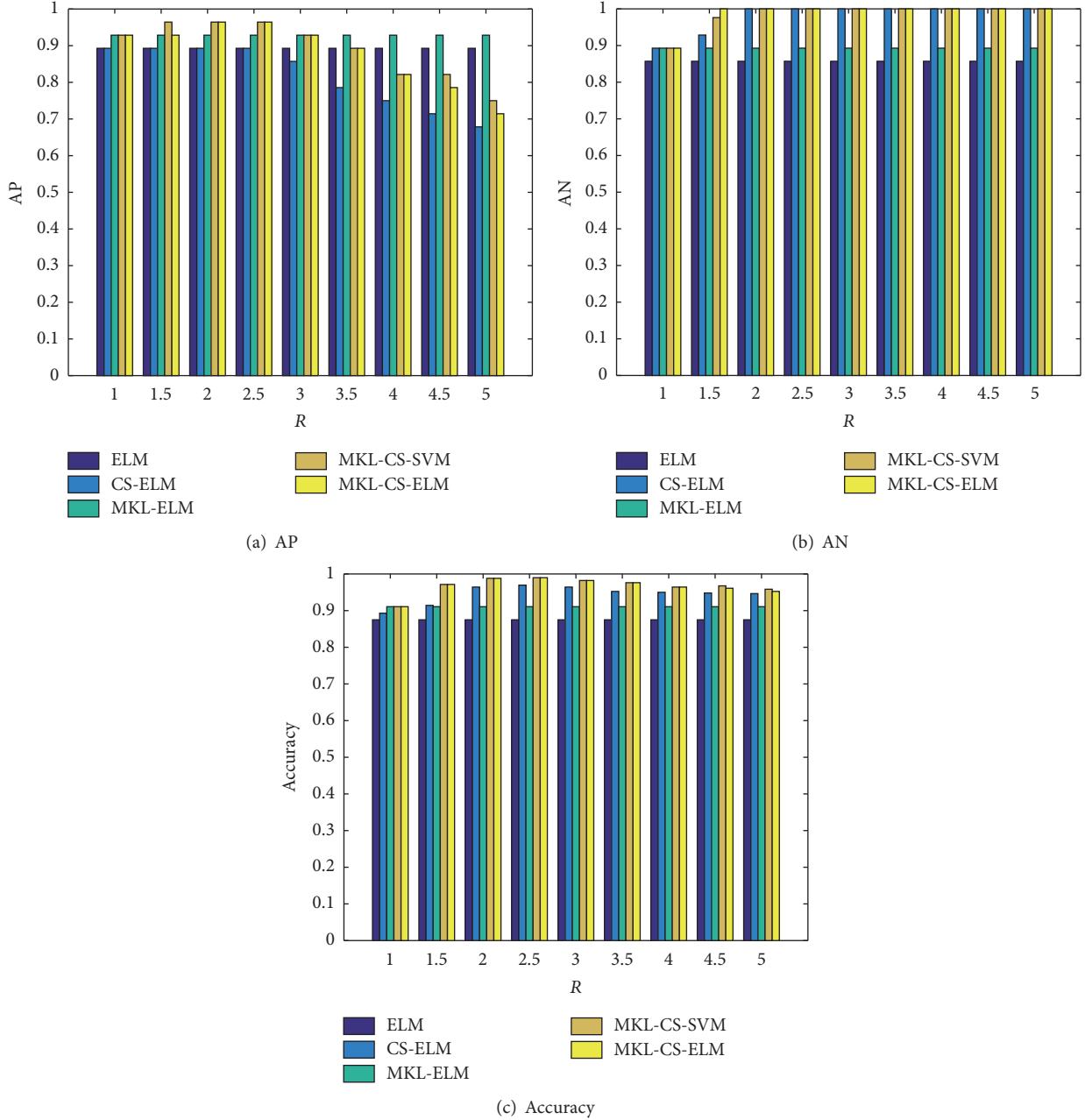


FIGURE 6: The recognition results of five classification models in oversampling.

problems are mostly owing to lack of the enough samples of check valve and the extreme imbalance of sample distribution is caused in the undersampling processing. At the same time, we can also observe an interesting phenomenon that when the sample is very small, the classification results of MKL-CS-ELM are slightly worse than the other classification models. This is probably an indirect argument that the training process of MKL-CS-ELM also needs the sufficient samples and the essence of MKL-CS-ELM is the single-hidden layer feedforward neural network. At the same time, the presented results also indirectly demonstrate the superiority of SVM in classification with smaller samples.

5.1.3. The Experimental Results of Threshold Adjusting. Based on the Table 4 and (19), the recognition results of five classification models in threshold adjusting are presented in Figure 8.

In Figure 8, the classification models of CS-ELM, MKL-CS-SVM, and MKL-CS-ELM can obtain good effects due to the introduction of cost-sensitive learning mechanism. The AN, AP, and Accuracy of aforementioned classification models are significantly improved with the increasing cost R . At the same time, the misclassification and missed diagnosis samples are sharply reduced with the increasing cost R . Compared with the performance of oversampling

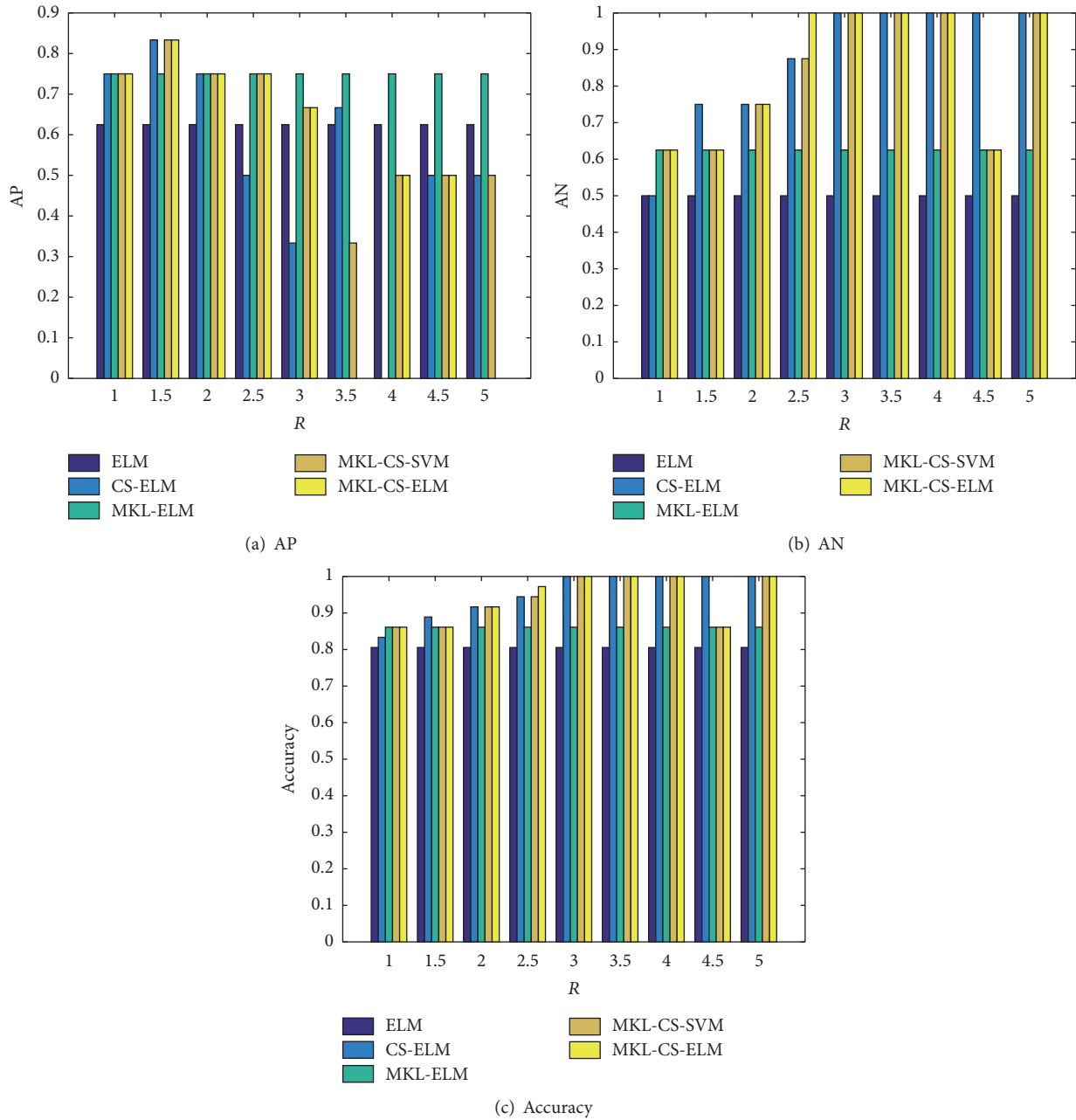


FIGURE 7: The recognition results of five classification models in undersampling.

and undersampling, the experimental results show that the threshold adjusting algorithm can also achieve satisfactory results. Therefore, the cost-sensitive method of threshold adjusting is also one of the effective choices for imbalance and inequality diagnosis cost in binary classification problems.

5.2. The Experimental Results Analysis of Multiclassification for Check Valve. In order to test validity and generalization ability of MKL-CS-ELM, the aforementioned three cost-sensitive methods are applied to identify multioperation states of check valve. Then the effectiveness of the proposed method is verified by multiclassification tests.

5.2.1. The Experimental Results of Oversampling. In the multiclassification experimental processing, the 110 data samples are collected, 66 samples are selected as training samples, and the remaining 44 samples are as test samples. The data sample distribution of oversampling is calculated based on (16) and (17). And the recognition results of classification models are presented in Figure 9.

As seen in Figure 9, the classification accuracy of CS-ELM, MKL-CS-SVM, and MKL-CS-ELM increases with the increasing cost R . On the contrary, the misclassification samples sharply reduce with the increasing cost R . The three classification models of CS-ELM, MKL-CS-SVM, and

TABLE 5: The data sample distribution by oversampling for check valve.

Cost R	Fault classes	Original samples	Processed samples
1	NC	70	70
	NK	20	70
1.5	NC	70	70
	NK	20	105
2	NC	70	70
	NK	20	140
2.5	NC	70	70
	NK	20	175
3	NC	70	70
	NK	20	210
3.5	NC	70	70
	NK	20	245
4	NC	70	70
	NK	20	280
4.5	NC	70	70
	NK	20	315
5	NC	70	70
	NK	20	350
--	--	--	--
--	--	--	--

MKL-CS-ELM can gain the optimal classification performance when the cost R is equal to 2.5 in oversampling processing. Meanwhile, compared with the experimental results illustrated in Figures 9(a), 9(b), and 9(c), some conclusions are summarized as follows: (1) The CS-ELM and MKL-CS-ELM are more sensitive to the cost R than the MKL-CS-SVM. (2) The classification performance of MKL-CS-ELM is slightly better than other above-mentioned classification models. (3) The change regularity of classification accuracy, misclassification, and missed diagnosis samples with the cost R is obtained and shown as follows: (1) the diagnosis cost $R = 2.5$ can be regarded as a demarcation line and inflection point of classification accuracy. (2) The misclassification and missed diagnosis samples drastically reduce when the cost R is less than 2.5. And the misclassification samples are reduced to 0 and reached balanced state when the cost R is greater than 2.5. But the missed diagnosis samples are sharply increased and the classification accuracy is also gradually decreasing. (3) The experimental results show that the above-mentioned cost-sensitive methods are feasible in check valve fault diagnosis of industrial field.

5.2.2. The Experimental Results of Undersampling. Similar to the previous oversampling approach, the data sample distribution of undersampling is calculated. Then, the experimental results of mentioned-above classification models are presented in Figure 10.

As shown in Figure 10, the classification accuracy of multikernel cost-sensitive diagnosis models is obviously decreased due to sharply reducing of data samples in undersampling. But, the misclassification samples can be also effectively restrained (even reduced to 0) by undersampling

when the cost R is equal to 2.5. However, Figure 10 also shows that the undersampling method should not be used in the conditions of the insufficient samples and high-accuracy requirements.

5.2.3. The Experimental Results of Threshold Adjusting. In the same way, the multiclassification recognition results of five mentioned classification models by the threshold adjusting are presented in Figure 11.

As depicted in Figure 11, in threshold adjusting processing, the misclassification samples are reduced to 0 when the cost R is increased to 2.5. The cost-sensitive classification models reach balanced state when the cost R is increased to 2.5, but the missed diagnosis samples and accuracy of CS-ELM, MKL-CS-SVM, and MKL-CS-ELM have no obvious change with the continuous increasing cost R .

5.3. Robust Performance Evaluation of Three Cost-Sensitive Methods for Check Valve. In order to assess the effectiveness of three cost-sensitive classification methods and choose the proper evaluation index for fault diagnosis of check valve, the robust performance evaluation r_α according to the description in Section 2.4.2 is calculated; the change regularity of robust performance index r_α varying with cost R is obtained and shown in Figure 12.

Figure 12 shows the comparative tests of robust performance evaluation in three cost-sensitive methods. The robust performance index r_α of the undersampling is biggest. That is to say, when the sample distribution is very imbalanced, it is not suitable to adopt the cost-sensitive method of undersampling. Moreover, in CS-ELM, MKL-CS-SVM, and MKL-CS-ELM method, the robust performance index r_α in

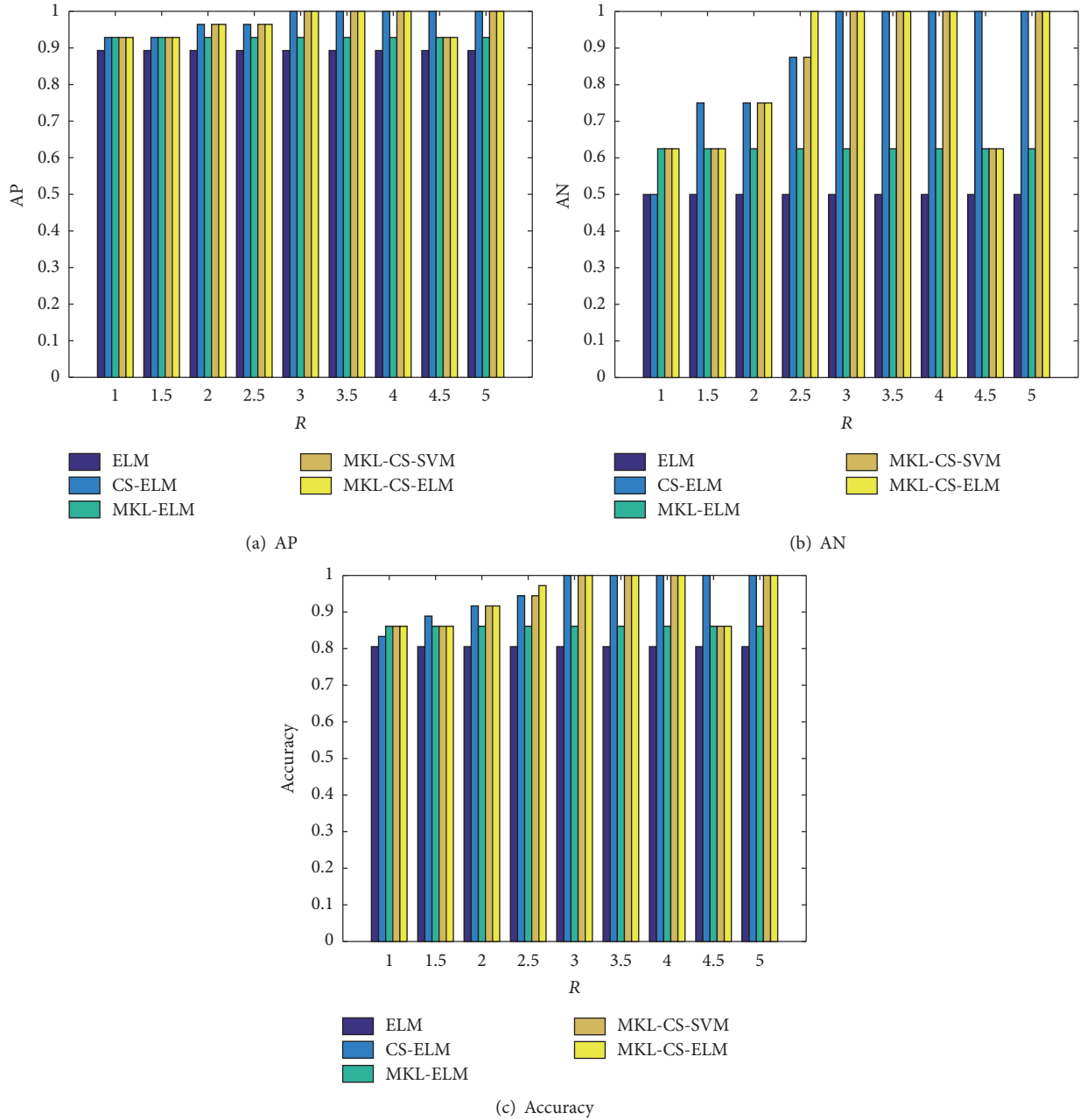


FIGURE 8: The recognition results of five classification models in threshold adjusting.

oversampling decreases at first then increases with increasing cost R and the robust performance index r_α in threshold adjusting decreases at first and then reaches steady state with increasing cost R . At the same time, Figure 12 also shows that the robust performance index r_α of oversampling is smaller than the threshold adjusting when the diagnosis cost R is less than 2.5, and then the change trend is reversed when the diagnostic cost R is greater than 2.5. Therefore, the oversampling and threshold adjusting are more appropriate cost-sensitive methods in multioperation states recognition of check valve.

6. Discussion and Conclusion

6.1. Discussion. High pressure diaphragm pump is often used as the core power equipment in slurry pipeline transportation, and its operating conditions are extremely complex. Therefore, it is critical to improve state recognition accuracy for ensuring operation safety and stability. However, the check valve is the core component of the high pressure diaphragm pump, and it is one of the most easily damaged and frequently replaced parts. Meanwhile, in the developed data acquisition system of check valve, the vibration data with

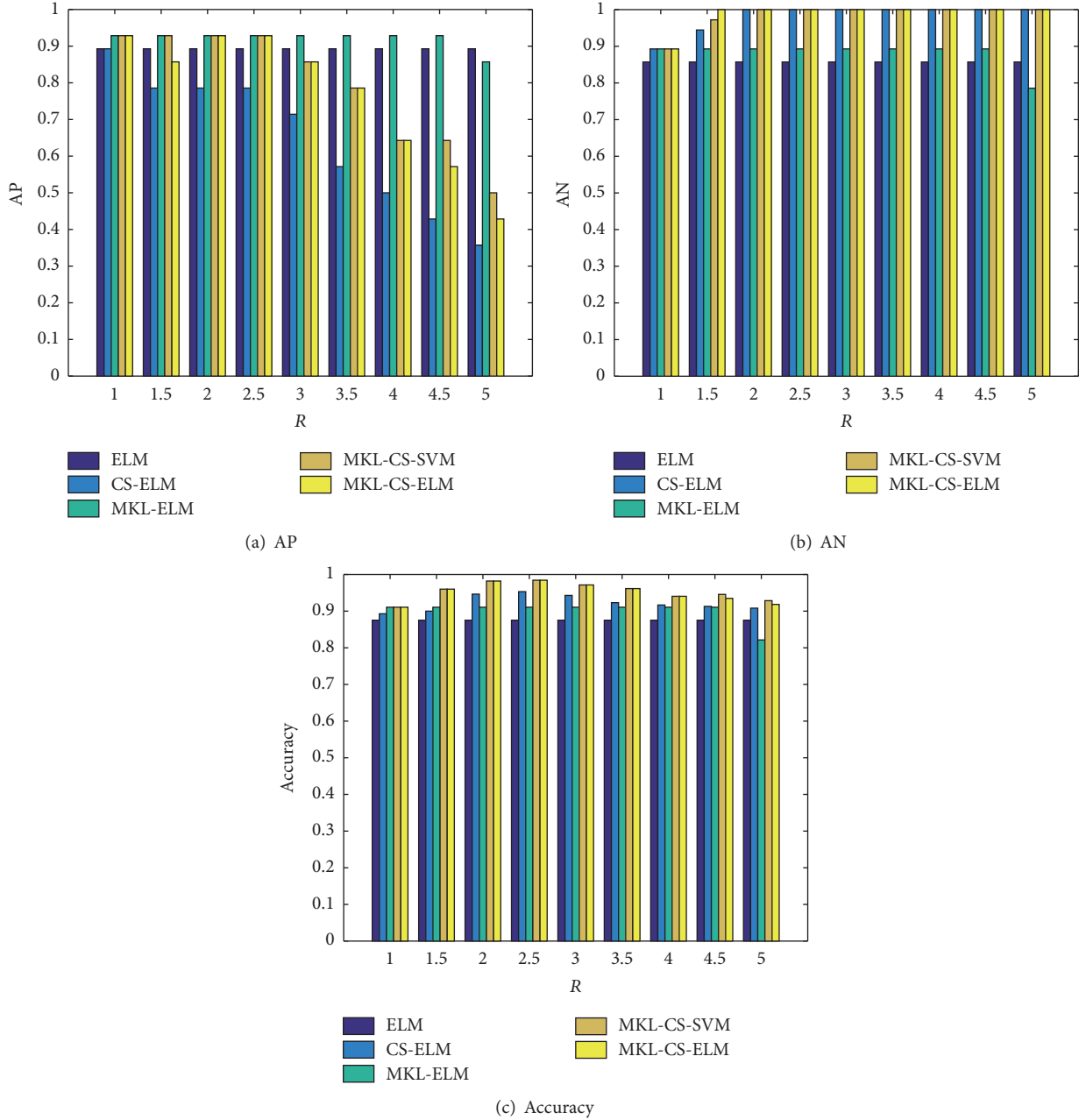


FIGURE 9: The recognition results of five classification models in oversampling.

normal operation has been collected in most of the time; on the contrary, the vibration data of fault time and fault state accounted for less. Therefore, it is of great significance to identify the operation state of the check valve effectively under the condition of complex operation and information asymmetry. Inspired by multikernel learning and cost-sensitive analysis, a fast diagnosis method of check valve based on MKL-CS-ELM is proposed. The presented MKL-CS-ELM method can complete the rapid positioning and analysis of the check valve fault and provide theoretical support for the adjustment and optimization in operation conditions of check valve during the follow-up operation.

(1) The multikernel learning mechanism is introduced to realize the multikernel projection of nonlinear and nonstationary data, which can overcome the limitation of incomplete information characterized with the single kernel function effectively and improve the ability to represent signals. Three kinds of common kernel function are used to construct multikernel classification model during the experiment. The introduction of multikernel learning can improve the recognition accuracy of classification model effectively through the analysis of MKL-ELM and ELM. In this case, what kind of kernel function and how many kernel functions are selected still lack normative choice mechanism. Therefore,

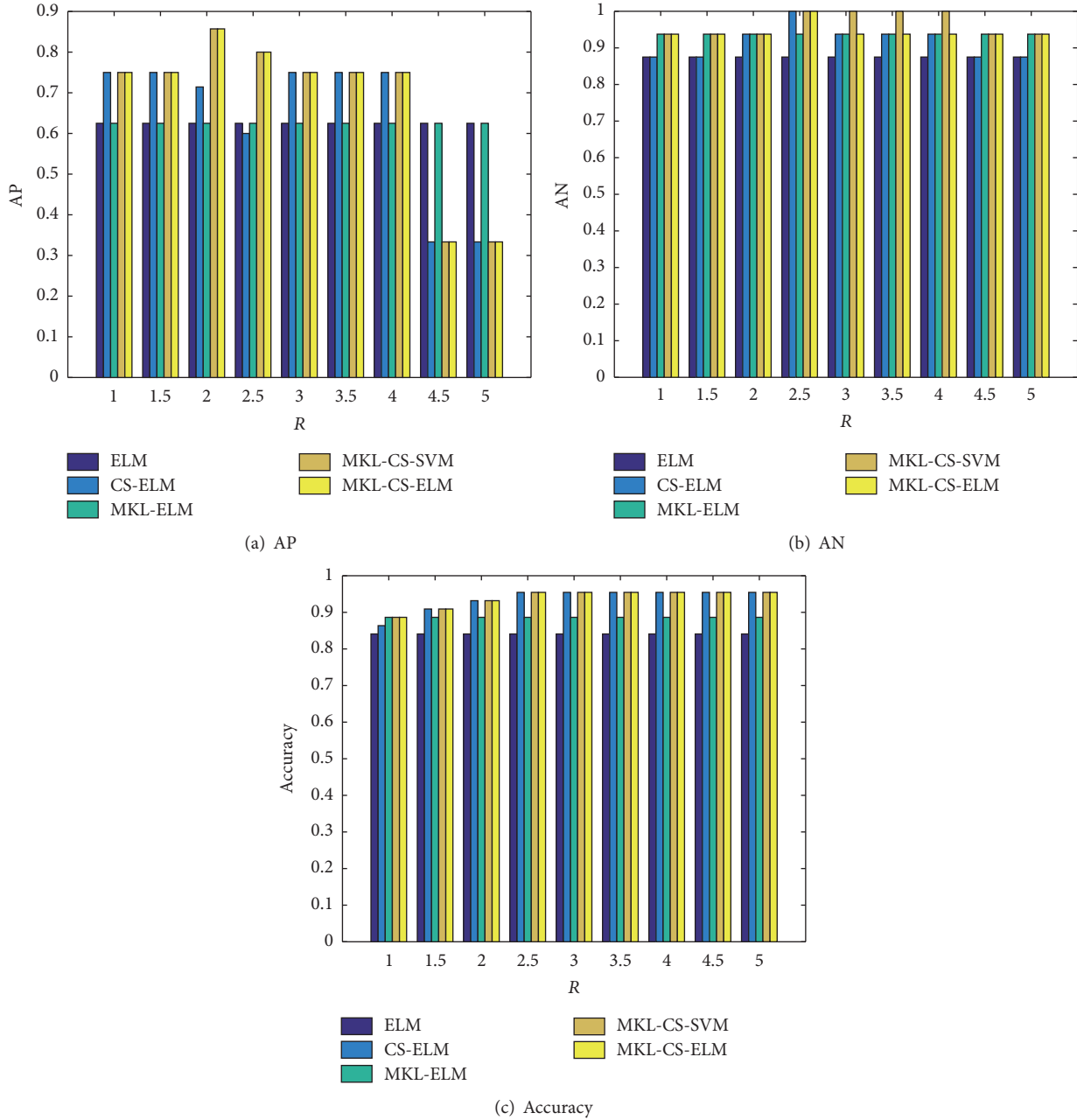


FIGURE 10: The recognition results of five classification models in undersampling.

we need to combine the signal characteristics and previous empirical rules about the selection to the kernel function so as to complete the selection of the effective kernel function and construct the multikernel function.

(2) In order to overcome the deficiency of assuming that the classification cost is equal through the classification model and improve the actual adaptability of the model, the paper makes the choice of the common cost-sensitive processing methods to construct CS-ELM model. The effectiveness of the introduction to cost-sensitive mechanism has been demonstrated through the binary classification and multiclassification recognition results; the experimental

results when using three kinds of cost-sensitive methods have also been compared with each other in different situations to provide theoretical support and guidance for the selection of cost-sensitive method. However, the cost of diagnosis needs to be moderate through the experimental comparison; otherwise it will reduce the overall recognition accuracy of the classification model.

6.2. Conclusion. The fault diagnosis model of MKL-CS-ELM based on the multikernel learning and cost-sensitive learning is constructed, and the datasets of check valve are used to verify the effectiveness of the proposed method. By

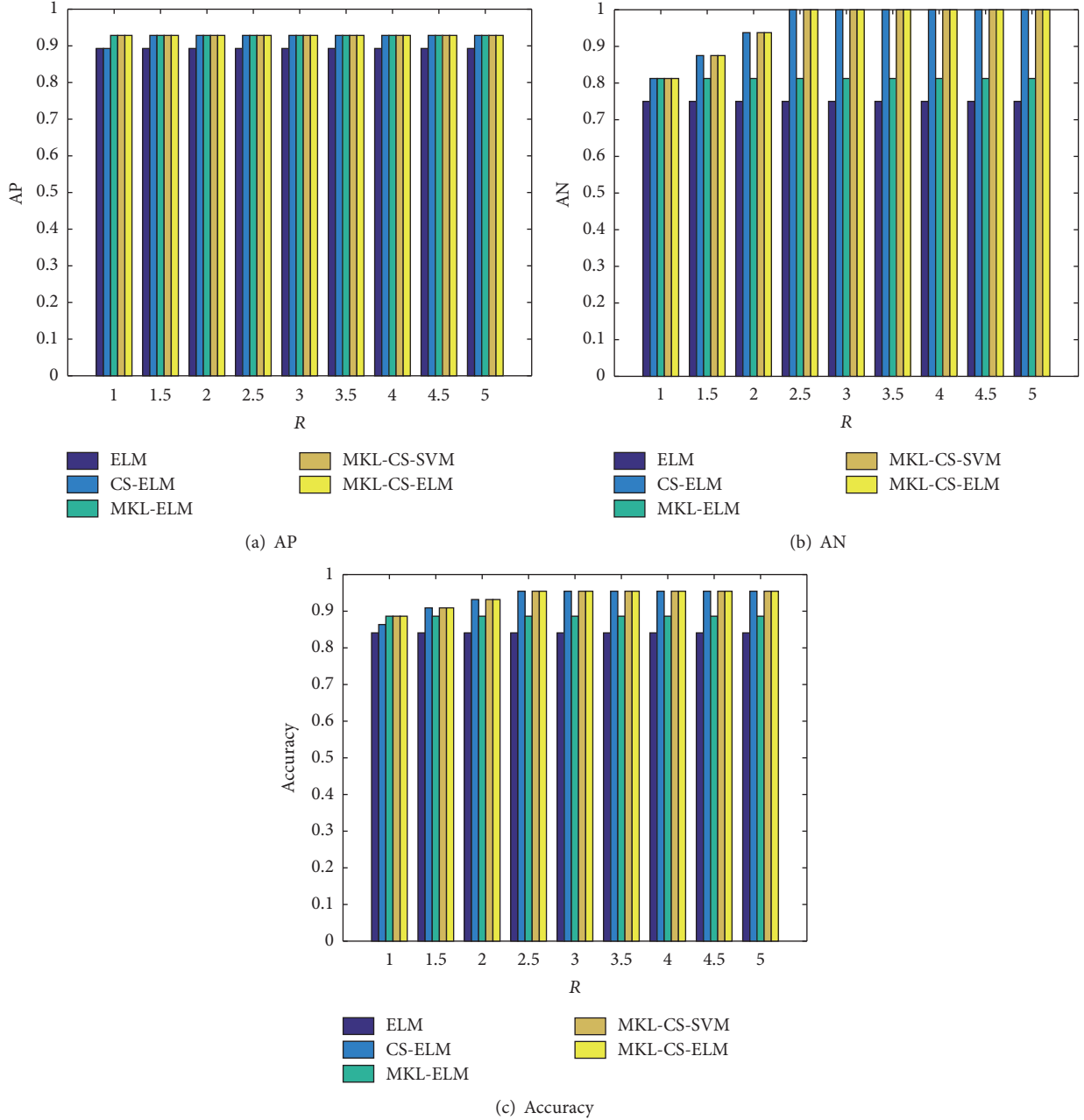


FIGURE 11: The recognition results of five classification models in threshold adjusting.

comparative tests, some conclusions can be summarized as follows.

(1) The MKL-CS-ELM can gain fair or better performance than the other classification models, including ELM, CS-ELM, MKL-ELM, and MKL-CS-SVM.

(2) The comparative analysis of robust performance evaluation r_α demonstrates that the oversampling and threshold adjusting cost-sensitive method are more appropriate choice in multiclassification application of check valve.

(3) The study of three cost-sensitive methods shows that, by selecting the appropriate cost R , the constructed

classification model can reduce the misclassification rate, achieve the balance between misclassification rates, miss diagnosis rate, and accuracy, and also improve the overall reliability of the classification model.

(4) The overall experimental results of the check valve show that the theory of multikernel learning and cost-sensitive learning can effectively overcome the disadvantage of the sample distribution imbalance and diagnostic cost equalization supposed in the conventional classification model and improve the accuracy and reliability of classification models.

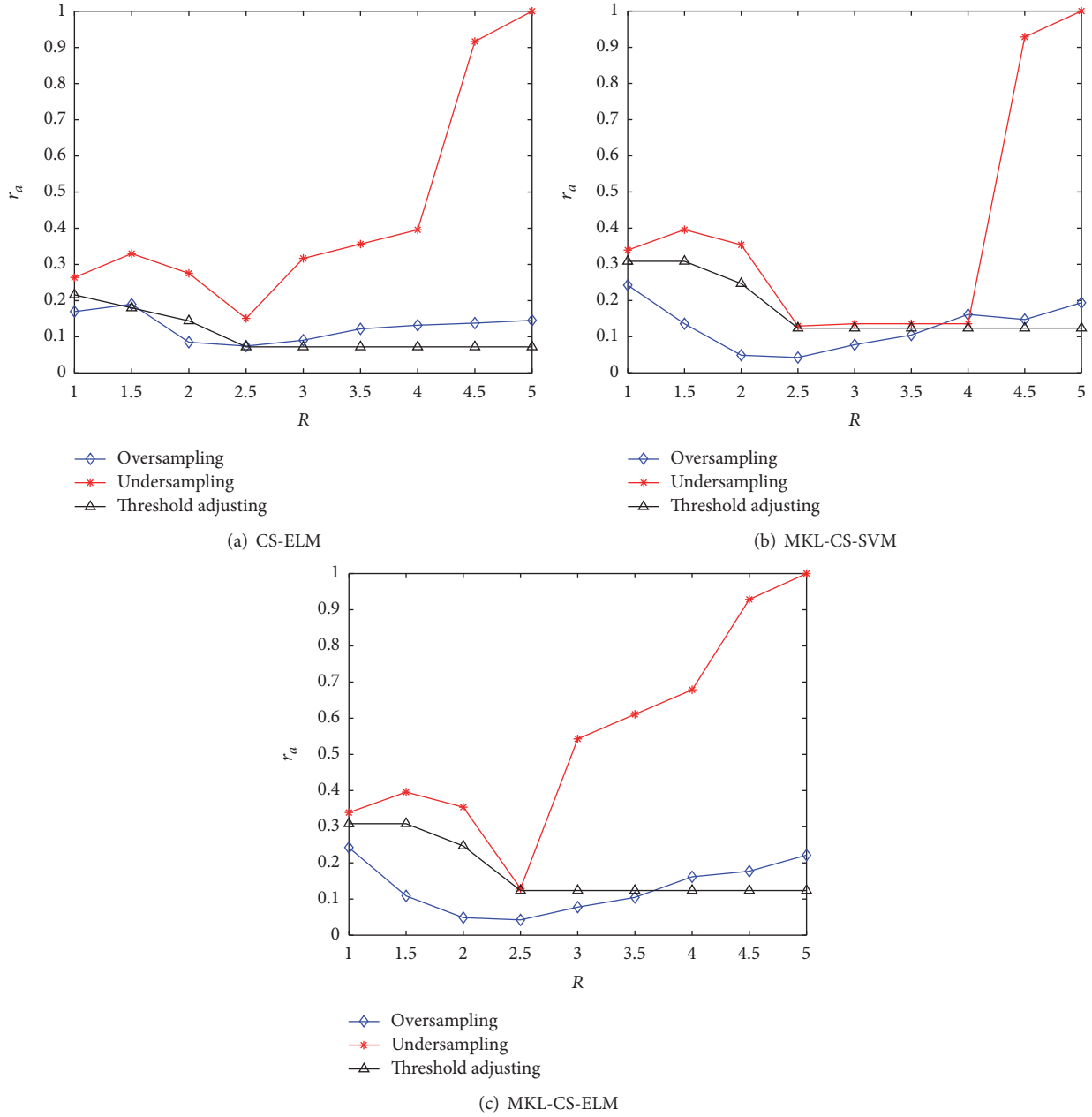


FIGURE 12: The performance comparisons of robust performance evaluation for different cost-sensitive methods.

Abbreviations

ELM: Extreme learning machine
 MKL-ELM: Multikernel ELM
 MKL-CS-ELM: Multikernel cost-sensitive ELM
 RBF: Radial basis function
 KKT: Karush Kuhn Tucker
 NK: Stuck valve fault
 C: Regularization parameter
 AP: The classification accuracy of positive samples
 $\{\kappa_p(\cdot, \cdot)_{p=1}^m\}$: The number of basic kernel functions is m

$\kappa(\cdot, \cdot; \gamma)$: The typical form of multikernel function
 $\{\phi(\cdot)\}_{p=1}^m$: The high dimensional feature mapping of $\{\kappa_p(\cdot, \cdot)_{p=1}^m\}$
 SVM: Support vector machine
 CS-ELM: Cost-sensitive ELM
 MKL-CS-SVM: Multikernel cost-sensitive SVM
 LSSVM: Least squares SVM
 NC: Normal condition
 NM: Abrasion fault
 Accuracy: Global classification accuracy
 AN: The classification accuracy of negative samples

- γ_p : The combination coefficients of basic kernel functions
- $\phi(\cdot, \gamma)$: The high dimensional feature mapping of $\kappa(\cdot, \cdot; \gamma)$.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (51765022, 61663017, and 51169007) and Science & Research Program of Yunnan Province (2015ZC005).

References

- [1] E. Rogatsky, M. Gallitto, and D. T. Stein, "A novel performance test for outlet check valve function in HPLC pumps," *Life Science Instruments*, vol. 28, no. 6, pp. 30-31, 2010.
- [2] B. Yang and J. Zhang, "The automatic diagnosis technology of the one-way valve malfunction of the reciprocating piston diaphragm pump," *Technology Information*, vol. 7, pp. 116-117, 2012.
- [3] B. Liu and S.-F. Ling, "On the selection of informative wavelets for machinery diagnosis," *Mechanical Systems and Signal Processing*, vol. 13, no. 1, pp. 145-162, 1999.
- [4] W. Li, F. Gu, A. D. Ball, A. Y. T. Leung, and C. E. Phipps, "A study of the noise from diesel engines using the independent component analysis," *Mechanical Systems and Signal Processing*, vol. 15, no. 6, pp. 1165-1184, 2001.
- [5] N. Lawrence and H. Y. P. Kortekaas, "DECSIM—a PC-based diesel engine cycle and cooling system simulation program," *Mathematical and Computer Modelling*, vol. 33, no. 6-7, pp. 565-575, 2001.
- [6] R. A. Ogle and D. T. Morrison, "Investigation of an acid spill caused by the failure of an air-operated diaphragm pump," *Process Safety Progress*, vol. 20, no. 1, pp. 41-49, 2001.
- [7] W. Liu and T. Chen, "Research on condition monitoring and trend prediction of reciprocating pump based on grey-neural network," *Journal of Safety Science & Technology*, vol. 3, no. 1, pp. 79-84, 2013.
- [8] C. Wang, M. Gan, and C. Zhu, "Non-negative EMD manifold for feature extraction in machinery fault diagnosis," *Measurement*, vol. 70, pp. 188-202, 2015.
- [9] S. Lu, J. Wang, and Y. Xue, "Study on multi-fractal fault diagnosis based on EMD fusion in hydraulic engineering," *Applied Thermal Engineering*, vol. 103, pp. 798-806, 2016.
- [10] S. Guo, Y. C. Xu, X. S. Li, R. Tao, K. Li, and M. Gou, "Research on roller bearing with fault diagnosis method based on EMD and BP neural network," *Advanced Materials Research*, vol. 1014, no. 1014, pp. 501-504, 2014.
- [11] H. Li, Y. Hu, F. Li, and G. Meng, "Succinct and fast empirical mode decomposition," *Mechanical Systems and Signal Processing*, vol. 85, pp. 879-895, 2017.
- [12] S.-W. Fei, "Fault diagnosis of bearing based on wavelet packet transform-phase space reconstruction-singular value decomposition and SVM classifier," *Arabian Journal for Science and Engineering*, vol. 42, no. 5, pp. 1967-1975, 2017.
- [13] S. Guoqi, S. McLaughlin, X. Yongcheng, and P. White, "Theoretical and experimental analysis of bispectrum of vibration signals for fault diagnosis of gears," *Mechanical Systems and Signal Processing*, vol. 43, no. 1-2, pp. 76-89, 2014.
- [14] Z. Zheng, W. Jiang, Z. Wang, Y. Zhu, and K. Yang, "Gear fault diagnosis method based on local mean decomposition and generalized morphological fractal dimensions," *Mechanism and Machine Theory*, vol. 91, article no. 2479, pp. 151-167, 2015.
- [15] Z. Meng and L.-L. Li, "Rolling bearing fault diagnosis based on local characteristic-scale decomposition and morphological fractal dimension," *Acta Metrologica Sinica*, vol. 37, no. 3, pp. 284-288, 2016.
- [16] J. K. Lee, T. Y. Kim, H. S. Kim, J.-B. Chai, and J. W. Lee, "Estimation of probability density functions of damage parameter for valve leakage detection in reciprocating pump used in nuclear power plants," *Nuclear Engineering and Technology*, vol. 48, no. 5, pp. 1280-1290, 2016.
- [17] D. P. Lewis, T. Jebara, and W. S. Noble, "Nonstationary kernel combination," in *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*, pp. 553-560, ACM, Pittsburgh, Pa, USA, June 2006.
- [18] S. O. Cheng, A. J. Smola, and R. C. Williamson, "Learning the kernel with hyperkernels," *Journal of Machine Learning Research (JMLR)*, vol. 6, no. 1, pp. 1043-1071, 2005.
- [19] J. He, S.-F. Chang, and L. Xie, "Fast kernel learning for spatial pyramid matching," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pp. 1-7, June 2008.
- [20] M.-Z. Tang, C.-H. Yang, W.-H. Gui, and Y.-F. Xie, "Cost-sensitive probabilistic neural network with its application in fault diagnosis," *Control and Decision*, vol. 25, no. 7, pp. 1074-1078, 2010.
- [21] Y. Wang, F. Cao, and Y. Yuan, "A study on effectiveness of extreme learning machine," *Neurocomputing*, vol. 74, no. 16, pp. 2483-2490, 2011.
- [22] C. Yang, K. Huang, H. Cheng, and Y. Li, "Haptic identification by ELM-controlled uncertain manipulator," *IEEE Transactions on Systems Man Cybernetics Systems*, vol. 47, no. 8, pp. 2398-2409, 2017.
- [23] R. Zhang, Y. Lan, G.-B. Huang, and Z.-B. Xu, "Universal approximation of extreme learning machine with adaptive growth of hidden nodes," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 2, pp. 365-371, 2012.
- [24] A. Lendasse, Q. He, Y. Miche, and G.-B. Huang, "Advances in extreme learning machines (ELM2012)," *Neurocomputing*, vol. 128, no. 6, pp. 1-3, 2014.
- [25] Q. Yu, Y. Miche, E. Eirola, M. van Heeswijk, E. Séverin, and A. Lendasse, "Regularized extreme learning machine for regression with missing data," *Neurocomputing*, vol. 102, no. 2, pp. 45-51, 2013.
- [26] T. Wang, D. Zhao, and Y. Feng, "Two-stage multiple kernel learning with multiclass kernel polarization," *Knowledge-Based Systems*, vol. 48, no. 2, pp. 10-16, 2013.
- [27] Z. Liang, S. Xia, Y. Zhou, and L. Zhang, "Training L_p norm multiple kernel learning in the primal," *Neural Networks*, vol. 46, no. 5, pp. 172-182, 2013.
- [28] X. Chao and Y. Peng, "A Cost-sensitive Multi-criteria quadratic programming model," *Procedia Computer Science*, vol. 55, pp. 1302-1307, 2015.
- [29] C. Yang, X. Wang, Z. Li, Y. Li, and C. Su, "Teleoperation control based on combination of wave variable and neural networks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 8, pp. 2125-2136, 2017.

- [30] C. Yang, J. Luo, Y. Pan, Z. Liu, and C. Su, "Personalized variable gain control with tremor attenuation for robot teleoperation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. pp, no. 99, pp. 1–12, 2017.
- [31] C. Yang, X. Wang, L. Cheng, and H. Ma, "Neural-learning-based telerobot control with guaranteed performance," *IEEE Transactions on Cybernetics*, vol. 47, no. 10, pp. 3148–3159, 2017.
- [32] C. Yang, Y. Jiang, Z. Li, W. He, and C.-Y. Su, "Neural control of bimanual robots with guaranteed global stability and motion precision," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1162–1171, 2017.
- [33] X. Liu, L. Wang, G. B. Huang, J. Zhang, and J. Yin, "Multiple kernel extreme learning machine," *Neurocomputing*, vol. 149, part A, pp. 253–264, 2015.
- [34] E. Zheng, C. Zhang, X. Liu, H. Lu, and J. Sun, "Cost-sensitive extreme learning machine," in *International Conference on Advanced Data Mining and Applications*, pp. 478–488, Springer, Berlin, Germany, 2013.
- [35] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 2, pp. 513–529, 2012.
- [36] G. Huang, G.-B. Huang, S. Song, and K. You, "Trends in extreme learning machines: a review," *Neural Networks*, vol. 61, pp. 32–48, 2015.
- [37] Z. Xu, R. Jin, H. Yang, and M. R. Lyu, "Simple and efficient multiple kernel learning by group Lasso," in *International Conference on Machine Learning*, pp. 1175–1182, 2010.
- [38] H. Yang, Z. Xu, J. Ye, I. King, and M. R. Lyu, "Efficient sparse generalized multiple kernel learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 22, no. 3, pp. 433–446, 2011.
- [39] W. Wu and J. Hu, "Fault diagnosis based on cost-sensitive transduction inference," *Chinese Journal of Scientific Instrument*, vol. 31, no. 5, pp. 1023–1028, 2010.
- [40] Z. H. Zhou and X. Y. Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 1, pp. 63–77, 2006.
- [41] K. M. Ting, "An instance-weighting method to induce cost-sensitive trees," *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, no. 3, pp. 659–665, 2002.
- [42] J. Wang, Q. Liu, and D. Qinhu, "Nonlinear membership function established by single-shot clustering method," *Journal of Zhengzhou University (Engineering Science)*, vol. 33, no. 2, pp. 28–30, 2012.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

