

## Research Article

# A Simultaneous Fault Diagnosis Method Based on Cohesion Evaluation and Improved BP-MLL for Rotating Machinery

Yixuan Zhang,<sup>1</sup> Rui Yang ,<sup>1</sup> Mengjie Huang ,<sup>1</sup> Yu Han,<sup>1</sup> Yiqi Wang,<sup>1</sup> Yun Di,<sup>1</sup> Dongke Su,<sup>1</sup> and Qidong Lu<sup>2</sup>

<sup>1</sup>*Xi'an Jiaotong-Liverpool University, Suzhou, China*

<sup>2</sup>*Weihai Beiyang Electric Group, Weihai, China*

Correspondence should be addressed to Rui Yang; [r.yang@xjtlu.edu.cn](mailto:r.yang@xjtlu.edu.cn)

Received 6 June 2021; Revised 22 August 2021; Accepted 20 October 2021; Published 8 November 2021

Academic Editor: Jun Zhu

Copyright © 2021 Yixuan Zhang 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.

In this paper, an improved simultaneous fault diagnostic algorithm with cohesion-based feature selection and improved backpropagation multilabel learning (BP-MLL) classification is proposed to localize and diagnose different simultaneous faults on gearbox and bearings in rotating machinery. Cohesion evaluation algorithm selects high sensitivity feature parameters from time and frequency domain in high-dimensional vectors to construct low-dimensional feature vectors. The BP-MLL neural network is utilized for fault diagnosis by classifying the feature vectors. An effective global error function is proposed in BP-MLL neural network by modifying distance function to improve both generalization ability and fault diagnostic ability of full-labeled and nonlabeled situations. To demonstrate the effectiveness of the proposed method, simultaneous fault diagnosis experiments are conducted via wind turbine drivetrain diagnostics simulator (WTDDS). The experiment results show that the proposed method has better overall performance compared with conventional BP-MLL algorithm and some other learning algorithms.

## 1. Introduction

Rotating machinery is a power transmission device in various mechanical equipment and also has been an indispensable part in industrial applications. Components in the rotation machinery including rotor, rotating shaft, bearing, and gearbox are all under arduous work and, thus, are subject to performance degradations and mechanical failures [1, 2]. However, any failure of key components in rotating machinery will cause serious accidents with high economic losses [3]. Therefore, the accurate detection of mechanical fault locations and types in rotating machinery is highly needed.

Currently, there are two types of quantitative analysis fault diagnostic methods for rotating machinery: model-based and data-driven diagnosis [4]. Model-based methods implement dynamic process models in the form of mathematical formulas and parameters; however, describing models by mathematical structure can be difficult and inefficient because of the more and more complicated industrial processes [5]. Compared with model-based fault

diagnosis, data-driven methods tend to transfer diagnostic problems to the pattern recognition problems. The data-driven methods are mainly composed of multivariate statistical analysis, such as regression [6] and principal component analysis [7], and machine learning methods, such as support vector machine (SVM) [8], random forest [9], neural networks [10–14], and transfer learning [15, 16].

Neural networks have been commonly used for intelligent fault diagnosis due to the powerful capabilities of pattern classification and function approximation [17]. On the basis of learning strategies, diagnostic algorithms based on neural networks can be classified into supervised and unsupervised learning. Backpropagation neural network is one of the most popular supervised learning strategies; in the 20th century, researches in [18–20] all proved the effectiveness of backpropagation neural networks in fault diagnosis. Additionally, compared with SVM, neural networks have a higher classification accuracy for fault diagnosis of rotor bearing systems [17]. Modifications to conventional backpropagation neural networks are also recommended to

address the problem of fault diagnosis. Meireles et al. pointed out that radial basis function (RBF) networks offer advantages of higher training speed and an easier optimization of performance over conventional neural networks for fault diagnosis [21]. Wu and Chow developed a RBF network-based system to induct machine faults and propose the cell-splitting grid algorithm so that the architecture of RBF network is automatically determined [22]. This proposed system can detect unbalanced electrical and mechanical faults under different working environment.

Unsupervised networks have different architectures such as self-organizing neural networks whose structures are adaptively determined to realize that all nodes in a neighborhood have similar output to an input when stable. The method based on self-organizing maps (SOM) proposed in [23] is not only able to detect the bearing faults, but also locates them and evaluates the failure extent. Jounela et al. developed a process monitoring system based on SOM associated with heuristic rules to detect machine malfunctions [24]. All in all, neural networks are capable of classifying arbitrary regions in space which makes it a good choice for fault diagnosis [25]. The recognition of simultaneous multiple faults is also discussed in [26]; the diagnostic performance by using single-fault recognition techniques may be limited: (1) fault isolation operations can be difficult since noise in the measured signals can obscure a particular fault feature; (2) a large training set is required, which is difficult and time-consuming to collect; (3) the choice of the most suitable classifier is still vague in engineering practice.

Multilabel learning methods are usually adopted during the detection and diagnosis of simultaneous fault in rotating machinery. Three main groups of multilabel learning strategies are data transformation, adaptation, and ensemble of classifiers [26]. The basic idea of data transformation methods is to turn the multilabel problems to other known learning problems according to [27]; one of the representative algorithms is the binary relevance which converts the original multilabel dataset to binary dataset [28]. Adaptation methods improve the conventional classification algorithms and directly employ the adapted algorithms for learning on multilabel data [29]. The kernel-dependent SVM in [8] is utilized to select features and to realize simultaneous fault detection of continuous processes. Zhang et al. extended  $k$  nearest neighbor (KNN) to a multilabel learning approach, named ML-KNN [30]. In detail, maximum a posteriori (MAP) principle is employed to determine the label set after  $k$  nearest neighbors are recorded for each instance in the training set. In terms of an ensemble of classifiers, Zhong et al. advanced a new probabilistic framework that combines multiple classifiers with a new ensemble method to realize simultaneous fault diagnosis with only single-fault data trained [31]. The first multilabel learning algorithm derived from the feed-forward neural network is proposed in [32], named back-propagation multilabel learning (BP-MLL). This neural network is optimized by minimizing the differences between the actual outputs and desired outputs on each training example. One of the most popular error functions is the sum-of-squares error functions, but BP-MLL applies a novel error function that does an exponential operation on the differences

between the outputs of labeled units and unlabeled units to capture characteristics of multilabel learning, i.e., yield output of labeled unit larger than that of unlabeled unit, and then a threshold function is used to determine a label set associated with each instance. BP-MLL neural network has been applied to assist medical syndrome diagnosis [33,34]. Multilabel text categorization systems based on BP-MLL neural networks are developed to classify multilabel documents [32]. Moreover, the prediction model based on BP-MLL in [35] is applied to estimate the types of sustainable flood retention basins.

However, the computation in BP-MLL neural networks is complex; according to [32], the total training cost of BP-MLL is  $O(W \cdot I \cdot n)$ , where  $W$  represents the total number of weights and bias,  $I$  is the number of training instances, and  $n$  is the total number of training epochs. Furthermore, the distance between relevant labels and irrelevant ones in conventional BP-MLL is represented by subtraction, which may be not obvious enough to be observed. Thus, two new distance functions that enhance pairwise labels discrimination to improve BP-MLL algorithm are proposed in [36]. Besides the problems mentioned above, the conventional BP-MLL algorithm is also not applicable for scenarios with full-labeled or nonlabeled situations. Multilabel classification assigns each instance with multiple categories that reflect properties of a data-point such as topics relevant to a document. A text might be about any of politics, education, specialties, or finance at the same time or none of these. Assume that a set of labels is organized and associated with each instance; if an instance is relevant to all the labels, then all labels in the set will be marked, so called full-labeled situation. Similarly, if an instance does not have connection to any labels in the set, it will be considered as nonlabeled situation. Modifications for BP-MLL algorithm made in [37] avoid failures under these two situations by taking differences between the rank values and the thresholds into account; besides, experimental performance of the modified algorithm is better shown on the same dataset as that in [32]. However, none of the existing literature specifically addresses full-labeled or nonlabeled situations, which may cause serious problems in practical application as computational errors may happen during network learning process in current approaches. As a result, the normal working rotating machinery would be misdiagnosed as a faulty one. Therefore, in this paper, we proposed an improved BP-MLL algorithm with a novel global error cost function and regularization term enhancing the generalization ability. Additionally, the cohesion evaluation algorithm based on standard deviation analysis is applied to obtain more comprehensive signal information and improve the adaptive ability of dynamic models.

Based on the above literature review and discussions, the main contributions of this paper are declared as follows: (1) a new global error function is proposed to deal with the problem of full-label and nonlabel learning situations; (2) a fault diagnosis method based on the improved BP-MLL and cohesion evaluation is proposed; (3) the problem of multilabel gearbox and bearing fault diagnosis in rotating machine under different working and environmental conditions is investigated.

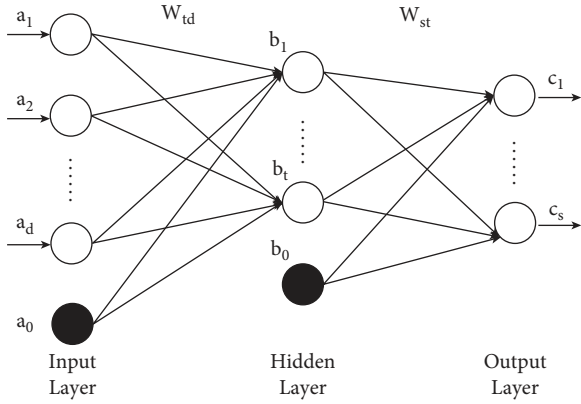


FIGURE 1: The architecture of BP-MLL model.

The structure of this paper is as follows. Section 2 discusses the preliminaries and formulates the problem. In Section 3, the proposed method is introduced. In Section 4, hardware experiments and comparative studies are carried out to verify the effectiveness of the method. Section 5 concludes this paper.

## 2. Preliminaries and Problem Formulation

**2.1. BP-MLL.** Suppose the training set is composed of  $I$  multilabel instances, i.e.,  $\{(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \dots, (X_i, Y_i)\}$ ,  $i = 1, 2, \dots, I$ . Each  $X_i$  is a  $d$ -dimension feature vector and  $Y_i$  is the associated set of labels. The BP-MLL architecture is shown in Figure 1, where  $d$  input neurons correspond to a feature vector,  $s$  output neurons represent  $s$  labels in  $Y_i$ , and the hidden layer has  $t$  hidden units. Each layer is fully connected with the next layer, with the weights  $W = [W_{td}, W_{st}]$  ( $d = 1, 2, \dots, N; t = 1, 2, \dots, M; s = 1, 2, \dots, J$ ). The number of hidden layers may be more than one in different neural network structures.

The error function proposed in [32] is

$$E = \sum_{i=1}^I E_i = \sum_{i=1}^I \frac{1}{|Y_i| \|\bar{Y}_i\|} \sum_{(k,l) \in Y_i \times \bar{Y}_i} e^{-(c_k^i - c_l^i)}. \quad (1)$$

This error cost function reflects the relationship between relevant labels and irrelevant ones by calculating the difference between them:

$$\text{distance} = \text{func}(c_k^i, c_l^i) = c_k^i - c_l^i. \quad (2)$$

Specifically,  $\bar{Y}_i$  is the complementary set of  $Y_i$  and  $|\cdot|$  measures the cardinality of a set.  $c_k^i$  represents the output of the network on one label belonging to the instance ( $k \in Y_i$ ) and  $c_l^i$  represents the one not belonging to it ( $l \in \bar{Y}_i$ ). Apparently, the larger the difference is, the smaller value of the error function of BP-MLL algorithm is, so that labels in  $Y_i$  will get greater neural network outputs than those not in. Therefore, when the training set covers sufficient information to disseminate the learning problem, the trained neural network will eventually distinguish the relevant labels from irrelevant ones.

**2.2. Problem Formulation.** Consider the following uncertain cases in the diagnostic system:

- (1) There are not any labels for one instance, indicating all components in the rotating machinery run perfectly such that  $|Y_i| = 0$ .
- (2) All the components are broken down such that  $|\bar{Y}_i| = 0$ , where  $\bar{Y}_i$  is the complementary set of  $Y_i$ .

When the diagnostic system applies error function Eq. (1), either uncertain case would cause mathematical failures, in the full-labeled case:

$$E = \sum_{i=1}^I E_i = \sum_{i=1}^I \frac{1}{|Y_i| \|\bar{Y}_i\|} \sum_{k \in Y_i} e^{-c_k^i}. \quad (3)$$

Firstly, because of  $|\bar{Y}_i| = 0$ , the denominator  $|Y_i| \|\bar{Y}_i\| = 0$ ; and furthermore, the value of  $c_k^i$  would toward infinity based on the property of exponential function. Similarly, when there do not exist any labels for a specific instance,  $|Y_i| = 0$  leads to an unreasonable denominator and besides,  $c_l^i$  is approaching infinity as well:

$$E = \sum_{i=1}^I E_i = \sum_{i=1}^I \frac{1}{|Y_i| \|\bar{Y}_i\|} \sum_{l \in \bar{Y}_i} e^{c_l^i}. \quad (4)$$

According the chain rule and the gradient descent rule for updating weights, the mathematical formulations are shown as below:

$$\begin{aligned} W_{st} &= W_{st} + \Delta W_{st}, \\ \Delta W_{st} &= -\alpha \frac{\partial E_i}{\partial W_{st}}, \\ &= -\alpha \frac{\partial E_i}{\partial Sc_s} \frac{\partial Sc_s}{\partial W_{st}}, \\ &= -\alpha \frac{\partial E_i}{\partial c_s} \frac{\partial c_s}{\partial Sc_s} \frac{\partial Sc_s}{\partial W_{st}}, \end{aligned} \quad (5)$$

where  $\alpha$  is the learning rate,  $W_{st}$  represents the weights from the hidden layer to the output layer,  $Sc_s$  represents the weighted sum, and  $c_s$  is the actual output of  $s$ -th output unit.

Apparently, there exists nondifferentiability in Eq. (5) since the value of  $c_s$  tends to infinity. Therefore, our approach is to optimize the error function to develop a fault diagnostic system such that it would not be affected by uncertain cases such as nonlabeled and full-labeled situation. In the next section, the improved error function that tolerates uncertain cases with high generalization ability is introduced.

## 3. Proposed Fault Diagnosis Method

Figure 2 illustrates the steps of fault diagnosis in rotating machinery. Firstly, the signals from the time-frequency domain are collected from multiple channels under different working conditions. Secondly, to fully grasp the characteristics of the signal and enhance the recognition ability of the fault diagnosis system, the cohesion evaluation algorithm is employed to pick out feature parameters with high

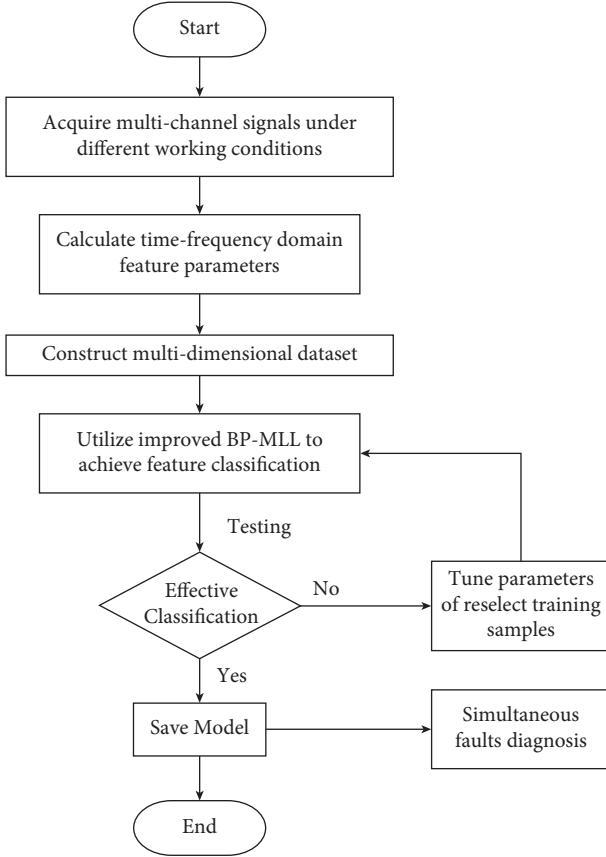


FIGURE 2: Training flow chart of the proposed method.

sensitivity to form the sensitive feature vector. Finally, an improved BP-MLL neural network is trained and utilized to classify the constructed feature vectors to make the system have dynamic model adaptability.

**3.1. Feature Selection.** Compared with conventional algorithms, the cohesion evaluation algorithm based on standard deviation analysis can combine multiple signals to obtain more comprehensive signal information and achieve the

purpose of improving the accuracy of fault diagnosis [10]. The distance assessment technique is described as follows:

Assume a set of  $d$ -dimensional feature vector has  $J$  different classes and the index of samples for each category is  $i$ :

$$\{q_{s,m,i}, \quad s = 1, 2, \dots, J; m = 1, 2, \dots, M; i = 1, 2, \dots, I\}, \quad (6)$$

where  $s$ ,  $m$ , and  $i$  are positive integers, and  $q_{s,m,i}$  reflects the  $m$ -th feature parameter of the  $i$ -th instance of the  $s$ -th category.

Table 1 is the specific operational steps of the cohesion evaluation algorithm where steps 1–3 reflect the intracategory standard deviation computation.  $c$  is a characteristic parameter, in which different characteristic parameters represent different practical meanings and  $u$  reflects the weight of the corresponding position neuron. The classification can be improved by reducing average intracategory standard deviation  $clt_j^{\text{inner}}$  and the intracategory standard deviation  $f_j^{\text{inner}}$ . Steps 4–8 represent the standard deviation computation of the feature distance, where the larger standard deviation of the feature distance  $stc_{s,j}$  and the smaller impurity measure of intercategory cohesion difference  $f_j^{\text{outer}}$  are more favorable for classification. Steps 9–10 determine the sensitivity of each feature parameter.

Figure 3 represents the cohesion evaluation process using the parameters in Table 1 as the horizontal and vertical coordinates, where  $clt^{\text{inner}}$  and  $stc$ , respectively, represent the size of circle radius and the position of circle center. In Figure 3, the intracategory standard deviation in class 3 and class 4 are easily overlapped and the distances between the points in each class are similar, resulting in a small average intercategory cohesion difference  $clt_j^{\text{outer}}$ , which is not conducive for distinguishing; in contrast, classes 1 and 2 belong to the easy classification feature parameter class. Overall, the cohesion evaluation algorithm can reflect the internal dispersion of the data and compare the detail of data differences. According to the steps of the cohesion evaluation algorithms in Table 1, the sensitivity weighting factor can be calculated as

$$\beta_m = \frac{1}{\left[ f_m^{\text{inner}} / \max(f_g^{\text{inner}}) + f_m^{\text{outer}} / \max(f_g^{\text{outer}}) + e_m^{\text{inner}} / \max(e_g^{\text{inner}}) + e_m^{\text{outer}} / \max(e_g^{\text{outer}}) \right]}. \quad (7)$$

The sensitivity factor is

$$\eta_m = \beta_m \frac{clt_m^{\text{outer}} + \text{pro} \cdot d_m^{\text{outer}}}{clt_m^{\text{inner}} + \text{pro} \cdot d_m^{\text{inner}}}, \quad (8)$$

where  $\text{pro}$  is the proportional adjustment coefficient.

The input feature vector of classification neural network is constructed by selecting parameters with large sensitivity factor according to equation (7):

$$\begin{aligned} \nu &= \text{sort}(\eta), \\ X_i &= [\nu_1, \nu_2, \dots, \nu_d], \end{aligned} \quad (9)$$

where  $\text{sort}(\cdot)$  sorts the  $m$  feature parameters in descending order, and first  $d$  ( $d < m$ ) high sensitive parameters construct a  $d$ -dimension input feature vector.

**3.2. Feature Classification.** The conventional BP-MLL algorithm captures correlation between relevant labels and irrelevant ones by using distance function which calculates the difference between them. The error function accumulates the differences in each instance and then normalize the summation by the total number of pairwise labels, i.e.,  $|Y_i| \| \bar{Y}_i |$ . As a result, with the increase in distances, the value

TABLE 1: Cohesion evaluation.

Step	Process parameter	Expression
1	Intracategory standard deviation	$\sigma_{m,i} = \sqrt{\sum_{s=1}^J (q_{s,m,i} - u_{m,i})^2 / J - 1}$
2	Average intracategory standard deviation	$clt_i^{inner} = 1/d \sum_{m=1}^d \sigma_{m,i}$
3	Difference of intracategory standard deviation	$f_i^{inner} = \max(clt_{m,i}) / \min(clt_{c,i})$
4	Distance of each feature	$cd_{s,r,m,i} =  q_{s,m,i} - q_{r,m,i} $
5	Quadratic sum of feature distance	$qs = \sum_{s,r=1}^d (cd_{s,r,m,i} - d_{m,i})^2$
6	Standard deviation of feature distance (intracategory cohesion)	$stc_{m,i} = \sqrt{qs / J(J-1) - 1}$
7	Average intercategory cohesion difference	$clt_i^{outer} = \sum_{m,c=1}^d  stc_{m,i} - stc_{c,i}  / d(d-1)$
8	Imparity measure of intercategory cohesion difference	$f_i^{outer} = \max( stc_{m,i} - stc_{c,i} ) / \min( stc_{k,i} - stc_{z,i} )$
9	Cohesion weighting factor	$c\omega_l = 1 / f_i^{inner} / \max(f_g^{inner}) + f_j^{outer} / \max(f_g^{outer})$
10	Cohesion factor	$\gamma_i = c\omega_l clt_i^{outer} / clt_i^{inner}$

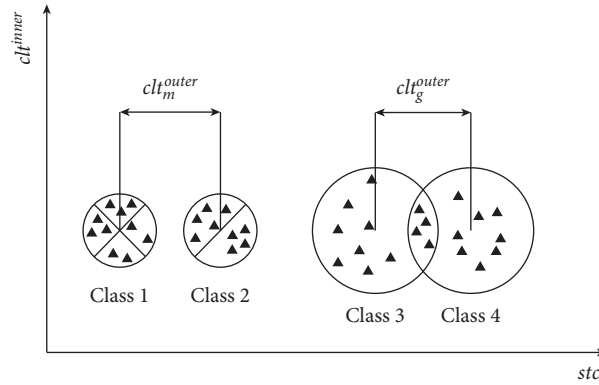


FIGURE 3: Cohesion evaluation.

of error function equation. (1) in BP-MLL algorithm will be smaller and smaller, which helps to rank labels belonging to an instance higher than those not belonging to.

$$\text{distance} = \text{func}(c_k^i, c_l^i) = c_k^i - c_l^i, \quad (10)$$

where  $c_k^i$  represents the output of relevant labels of  $i$ -th instance and  $c_l^i$  is the output of those irrelevant ones.

Nevertheless, as mentioned in Section 2, the conventional algorithm does not take full-labeled or nonlabeled situations into account. Full-label in an instance is shown in Figure 4 and the nonlabeled situation implies an instance does not have any marked labels. Mathematical failures such as an unreasonable denominator would occur during training or the trained neural network cannot attain an acceptable classification result on unseen cases if not considering those two situations. The most direct way is to modify the distance function so that the algorithm can let labels be as close to targets as possible while considering the characteristics of multilabel learning. Therefore, in this paper, modifications are made on the distance function to handle this problem. If an instance is not marked by any labels, that is,  $|Y_i| = 0$ , then there only exist irrelevant labels so that the distance function would be modified as

$$\text{distance} = \text{func}(-1, c_l^i) = -1 - c_l^i, \quad (11)$$

$$\text{if } |Y_i| = 0.$$

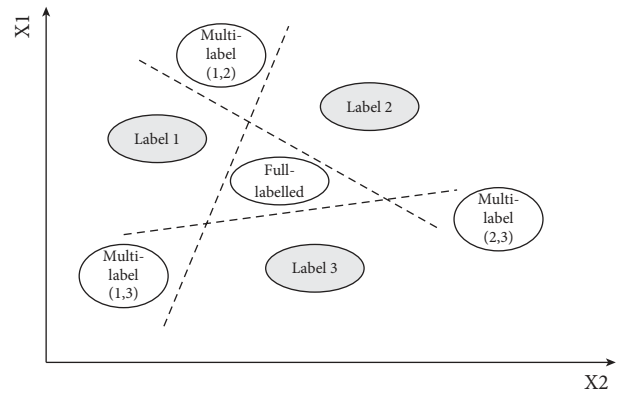


FIGURE 4: Labels for different instances.

Similarly, if all labels are marked in an instance, then these labels are all relevant so that the distance between 1 and these labels should be as small as possible:

$$\text{distance} = \text{func}(c_k^i, 1) = c_k^i - 1, \quad (12)$$

$$\text{if } |\pi| = 0.$$

The error function in BP-MLL algorithm is visualized in Figure 5, along with the modifications marked as red solid line and blue dotted line. From the perspective of image, two new distances functions based on full-labeled and

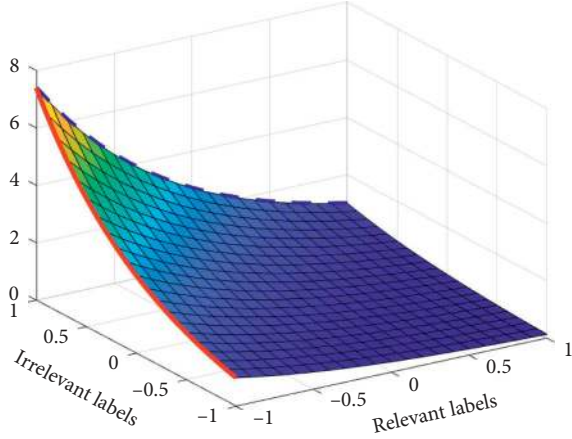


FIGURE 5: The error function of BP-MLL algorithm.

nonlabeled situations are derived from original error function, which indicates these would not generate conflicts in different types of training samples. Therefore, the improved error function in BP-MLL algorithm preserves the ability of discriminating relevant labels and irrelevant ones; meanwhile, it has capability of dealing with full-labeled or nonlabeled situations. Apart from the modifications on distance function, a regularization term is added to enhance the generalization ability. The main contribution in this paper is the improvement of BP-MLL algorithm to let it concentrate on both the correlations between different labels and the occurrence of empty sets by the following global error function:

$$E = \sum_{i=1}^I \frac{\sum_{(k,l) \in Y_i \times \bar{Y}_i} e^{-(\bar{c}_k - \bar{c}_l)}}{\max(1, |Y_i|) \max(1, |\bar{Y}_i|)} + \frac{\beta}{2} \sum_{d=0}^N \sum_{t=0}^M \sum_{s=1}^J (W_{td}^2 + W_{st}^2), \quad (13)$$

where  $\beta$  is the regularization coefficients and

$$\bar{c}_k^i = \begin{cases} c_k^i, & |Y_i| \neq 0, \\ -1, & |Y_i| = 0, \end{cases} \quad (14)$$

$$\bar{c}_l^i = \begin{cases} c_l^i, & |\bar{Y}_i| \neq 0, \\ 1, & |\bar{Y}_i| = 0 \end{cases}$$

Remark  $c_k$  represents the actual output of  $k$ -th output unit and this label belongs to this instance ( $k \in Y_i$ ), while  $c_l$  represents the output of one label that does not belong to this instance ( $l \notin Y_i$ ). Due to the property of the exponential term, the bigger the difference between  $c_k$  and  $c_l$  is, the smaller the global error is.

In conventional BP-MLL algorithm, saturation may occur due to the choice of activation function: sigmoid or hyperbolic tangent (tanh). Moreover, due to the exponential computation in the error function, using these two functions can also lead to high time complexity. Therefore, to avoid vanishing gradient problem and to reduce the difficulty of calculation, the improved method proposed in this paper adopts leaky rectified linear unit (Leaky ReLU) as the activation function:

$$h(x) = \begin{cases} x, & \text{if } x > 0, \\ \alpha_0 x, & \text{if } x \leq 0, \end{cases} \quad (15)$$

where  $\alpha_0 \in (0, 1)$ .

Let  $Sc_s$  represent the weighted sum; then the actual output of  $s$ -th output unit is

$$c_s = h(Sc_s). \quad (16)$$

In this paper, the gradient descent rule is adopted to adjust weights and bias until the global error converges to an acceptable value so that the updated  $W_{st}$  is

$$W_{st} = W_{st} + \Delta W_{st},$$

$$\Delta W_{st} = -\alpha \left( \frac{\partial E_i}{\partial W_{st}} + \beta W_{st} \right), \quad (17)$$

$$= -\alpha \frac{\partial E_i}{\partial Sc_s} \frac{\partial Sc_s}{\partial W_{st}} - \alpha \beta W_{st},$$

where  $\alpha$  is the learning rate, and define  $\text{temp}_s$  as

$$\text{temp}_s = -\frac{\partial E_i}{\partial Sc_s} = -\frac{\partial E_i}{\partial c_s} \frac{\partial c_s}{\partial Sc_s}. \quad (18)$$

Let  $h' = \partial c_s / \partial Sc_s$ ; then

$$h' = \frac{\partial c_s}{\partial Sc_s} = \begin{cases} 1, & \text{if } Sc_s > 0, \\ \alpha_0, & \text{if } Sc_s \leq 0. \end{cases} \quad (19)$$

Substituting equations (9) and (14) into (13):

$$\text{temp}_s = -\frac{\partial \left[ \sum_{(k,l) \in Y_i \times \bar{Y}_i} \exp(-(\bar{c}_k - \bar{c}_l)) / \max(1, |Y_i|) \max(1, |\bar{Y}_i|) \right]}{\partial c_s} h'.$$

$$\text{temp}_s = \begin{cases} \frac{1}{|Y_i|} e^{-(c_s-1)} h', & \text{if } |\bar{Y}_i| = 0, \\ -\frac{1}{|\bar{Y}_i|} e^{-(-1-c_s)} h', & \text{if } |Y_i| = 0, \\ \left[ \frac{1}{|Y_i| |\bar{Y}_i|} \sum_{l \in \bar{Y}_i} e^{-(c_s - c_l)} \right] h', & \text{if } s \in Y_i, \\ \left[ -\frac{1}{|Y_i| |\bar{Y}_i|} \sum_{k \in Y_i} e^{-(c_k - c_s)} \right] h', & \text{if } s \in \bar{Y}_i. \end{cases} \quad (20)$$

So equation (12) can be rewritten as

$$\Delta W_{st} = \alpha \text{temp}_s \left[ \frac{\partial \sum_{t=1}^M b_t \cdot W_{st}}{\partial W_{st}} \right] - \alpha \beta W_{st}, \quad (21)$$

$$= \alpha \text{temp}_s b_t - \alpha \beta W_{st}.$$

A preset threshold  $\varepsilon$  is used for classification of each instance and fault detection, represented by res:

$$\text{res} = \begin{cases} 1, & \text{if } c_s > \varepsilon, \\ -1, & \text{if } c_s \leq \varepsilon. \end{cases} \quad (22)$$

The proposed algorithm for the multifault diagnosis of rotating machinery is summarized as Table 2 and the main contributions of the proposed algorithm are as follows: (1) optimize the BP-MLL algorithm by improved error cost function and regularization term; (2) propose a fault diagnostic algorithm based on the improved BP-MLL and cohesion evaluation; (3) perform an experimental study on the multilabel gearbox and bearing fault diagnosis in rotating machine under various working and environmental conditions.

## 4. Experimental Results

**4.1. Experimental Platform.** In this paper, wind turbine drivetrain diagnostics simulator (WTDDS) produced by SpectraQuest, USA, is used as the experimental platform. Figure 6 illustrates its operation diagram, in which label **a** represents the torque sensor set on the shaft, **b** and **d** are the vibration sensors, fixed above and to the left of the parallel shaft gearbox, respectively, and **c** is the pressure sensor. These sensors are connected to a multichannel signal acquisition device which can aggregate signals to a computer and convert them into voltage signals.

In Figure 7, the reference number 4 is a single-phase motor to power the entire system. The reference number 3 is a parallel shaft gearbox which can transmit the kinetic energy of the motor to the planetary gearbox by coupling with the motor. In the planetary gearbox (referred to by 2), four planetary gears are rotated under the traction of the driving wheel that can transmit kinetic energy to the load brake referred to by 1. The windmill (referred to by 6) then can be driven by the next stage of rotating shaft after the braking action of the load brake.

**4.2. Experimental Setup.** In reality, the motor frequency is affected by the mechanical structure and wind speed, and the load is related to the generator structure and voltage. To collect accurate and effective data, the combination of different motor frequency and load voltage is used to simulate different working conditions. Through the software Lab-View, the input voltage of the load controller and the speed controller is adjusted to realize manual control of shaft load and motor speed. Table 3 summarizes the six operating conditions considered and set in this experiment:

The proposed method is applied to classify the five types of faults in gears, which are ball bearing, outer bearing, inner bearing, chipped tooth, and missing tooth, respectively, as shown in Figure 8. A label of five-digit binary code represents the expected output value. In the experiment, the length of the feature vector for the signal segment is 2048, and the sampling time is 6.4s, and the sampling frequency is 5120 Hz. The categories of simultaneous faults and the samples assigned to training and testing procedures are shown in Table 4.

## 4.3. Experimental Results

**4.3.1. Classification of Different Simultaneous Faults.** The multilabel algorithms (improved BP-MLL, BP-MLL, and ML-KNN) and the conventional classification technique (BP neural network) were performed on the same data sets. In this paper, all experiments have been performed on a computer with 8G RAM and Intel® Core™ i5-7200U CPU @ 2.70 GHz. The improved BP-MLL neural network has two hidden layers, the numbers of neurons in input, first hidden, second hidden, and output layers are 32, 72, 12, and 5, respectively. To approach the optimal solution and make the algorithm converge, the learning rate in improved BP-MLL neural network is set as  $0.95^{\text{iter}} * \alpha_0$ .

Table 5 summarizes the classification accuracy of different types of simultaneous faults and the total training time in terms of various algorithms. The ML-KNN algorithm required the least time to train and obtained a relatively high classification accuracy, but the classification accuracy of certain cases cannot be guaranteed, such as 00000 and 00101. Although the basic BP technique trained the neural network fast, it is limited to nonlabeled situation discriminations and fails to classify the second faulty types. As mentioned in Section 2, conventional BP-MLL algorithm is unable to deal with nonlabeled situations. Additionally, conventional BP-MLL spent more time than improved BP-MLL to train the same network. This is because the proposed method applied Leaky ReLU as the activation function that helps to reduce time consumption for calculation during the gradient descent process by judging the weighted sum in error function equation (9) firstly. The faults on outer bearing with chipped tooth confuses BP, BP-MLL, and ML-KNN. However, all algorithms for comparison can detect the faults on outer bearing with missing tooth. In general, the proposed method can achieve the accuracy of 100% with less training time than conventional BP-MLL method. Although the training time of the proposed method is longer than ML-KNN method, this is acceptable as the training process is conducted offline and the classification accuracy of the proposed method is consistently higher than ML-KNN method.

As a nonlinear dimensionality reduction algorithm, T-distribution stochastic neighbor embedding (t-SNE) technique proposed in [38] uses conditional probability to express the similarity of distance between data points, which is very applicable for dimensionality reduction. To visualize classification results directly, the three-dimensional mapping results are shown in Figure 9. Although data points in conventional BP-MLL have clear borders, the third type of simultaneous faults cannot be detected correctly due to the missing classification of one specific fault. Additionally, there are only a few points in ML-KNN; this is because the value of each output unit is the probability of each label. The predicted data instances after training in improved BP-MLL algorithm are basically distributed around their centers.

TABLE 2: Proposed fault diagnosis algorithm.

Training stage	
1	Obtain original sampling signals from multichannel sensors as training data.
2	Compute feature parameters for all the channels to construct a high-dimensional feature vector.
3	Compute sensitivity weighting factor $\beta$ and sensitivity factor $\eta$ using equations (6) and (7) and Table 1. Select parameters of high sensitivity factor to construct the sensitive feature vector.
4	Use the feature vector as input vector and modify weights and bias by using equations (10), (13), and (14) until the trained model can meet the test requirement of high accuracy or the maximum number of training epochs has been reached
Diagnosis Stage	
1	Construct the sensitive feature vector of new data.
2	Compute the outputs for each testing instance by epoch as shown in equations (11) and (10).
3	Use a preset threshold to classify each instance for fault diagnosis using equation (17)

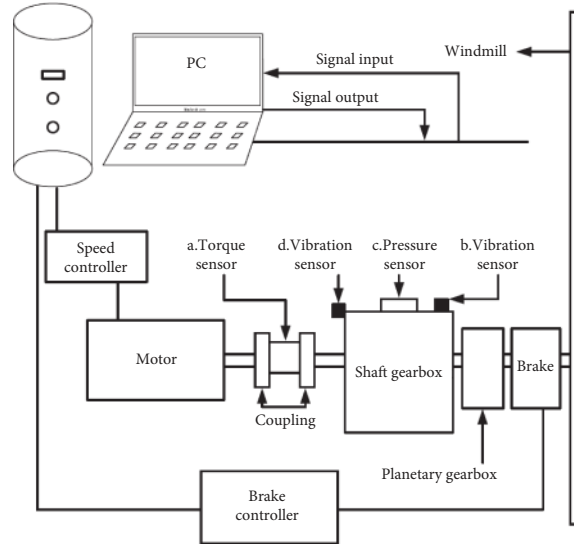


FIGURE 6: The operation diagram of WTDDS.

4.3.2. *Comparison on Different Algorithms.* To compare the performance of the proposed method with other conventional ones, the following six evaluation metrics are used to measure the classification results [39].

F1-score is also known as balanced F score, which is defined as the harmonic mean of precision and recall:

$$F1 - \text{score} = \frac{1}{I} \sum_{i=1}^I \frac{2|Z_i \cap Y_i|}{|Z_i| + |Y_i|}, \quad (23)$$

where  $Z_i$  denotes the predicted labels and  $Y_i$  is the desired value of  $s$ -th label in  $i$ -th instance. Recall is the fraction of the correct labels expected from the actual labels, while precision is the fraction of labels correctly classified from the expected positive labels, averaged on all instances:

$$\begin{aligned} \text{Recall} &= \frac{1}{I} \sum_{i=1}^I \frac{|Z_i \cap Y_i|}{|Y_i|}, \\ \text{Precision} &= \frac{1}{I} \sum_{i=1}^I \frac{|Z_i \cap Y_i|}{|Z_i|}. \end{aligned} \quad (24)$$

Hamming loss is used to investigate the misclassification of an instance on a single label; i.e., the correlation label does not appear in the predicted label set or the irrelevant label appears in the predicted label set. The smaller the value of Hamming loss is, the better the system performance is:

$$\text{Hamming loss} = \frac{1}{I} \sum_{i=1}^I \text{XOR}(Y_i, Z_i), \quad (25)$$

where XOR represents exclusive or.

Ranking loss evaluates the fraction of pairs of labels that are misclassified for the instance. The lower the values of this metric are, the better the performance is:

$$\text{Ranking loss} = \frac{1}{I} \sum_{i=1}^I \frac{1}{|Y_i| \|\bar{Y}_i\|} |E|, \quad (26)$$

where  $|E|$  is the size of error-set and  $E = \{(\lambda, \lambda') | \tau_i(\lambda) \leq \tau_i(\lambda'), (\lambda, \lambda') \in Y_i \times \bar{Y}_i\}$ , and  $\tau_i(\cdot)$  is the true value of output before being labeled.



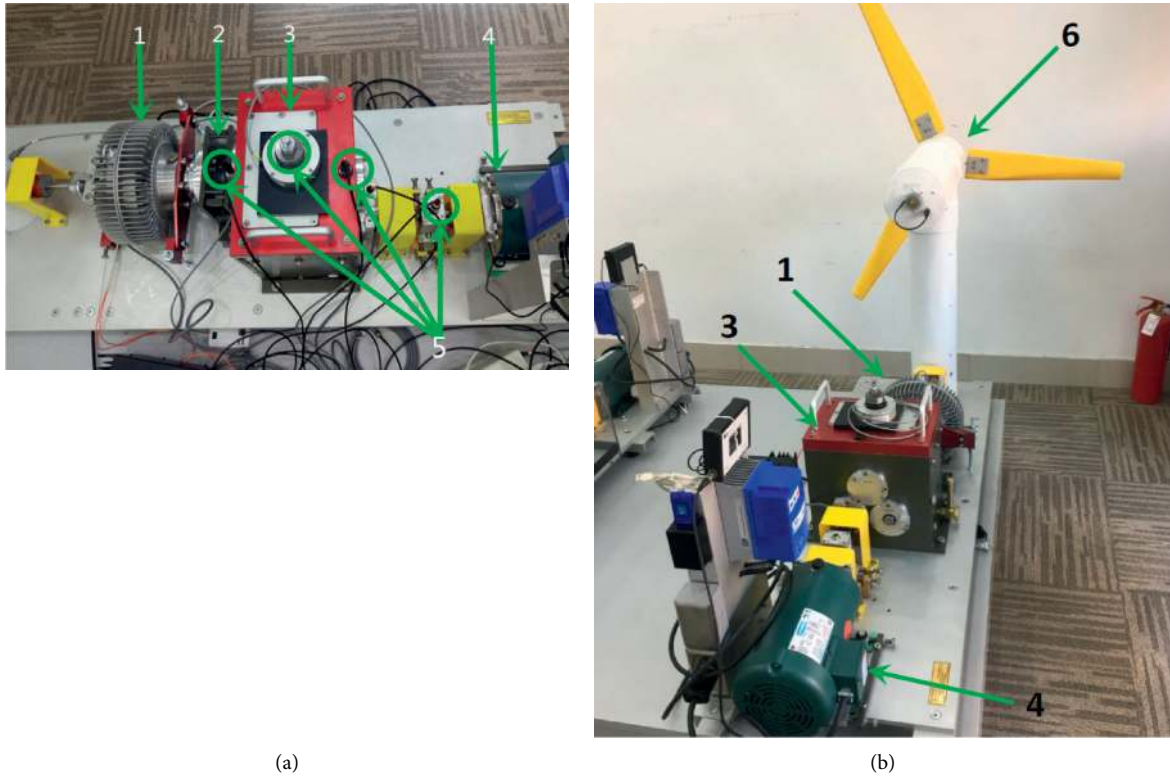


FIGURE 7: The structure of WTDDS (1, brake; 2, planetary gearbox; 3, shaft gearbox; 4, motor; 5, four sensors (from left to right: vibration, force, vibration, and torque); 6, windmill).

TABLE 3: Working conditions settings.

Motor frequency (Hz)	Load voltage (V)	Condition category
6	8	A
6	5	B
10	8	C
10	5	D
14	8	E
14	5	F



FIGURE 8: Five types of gears.

TABLE 4: Samples and labels.

Fault category	Number of training samples	Number of testing samples	Category code
Normal	48	10	00000
Ball bearing + missing tooth	48	10	10010
Outer bearing + missing tooth	48	10	00110
Outer bearing + chipped tooth	48	10	00101
Inner bearing + chipped tooth	48	10	01001

TABLE 5: Accuracy for different simultaneous faults.

Fault category	Algorithms			
	Improved BP-MLL	BP	BP-MLL	ML-KNN
00000	100.00%	66.67%	—	89.58%
10010	100.00%	0.00%	100%	100.00%
00110	100.00%	100.00%	100%	100.00%
00101	100.00%	83.33%	18.75%	83.33%
01001	100.00%	95.83%	100%	100.00%
Grand total	100.00%	69.17%	79.69%	94.58%
Training time (s)	105.1355	25.4463	787.2271	3.3972

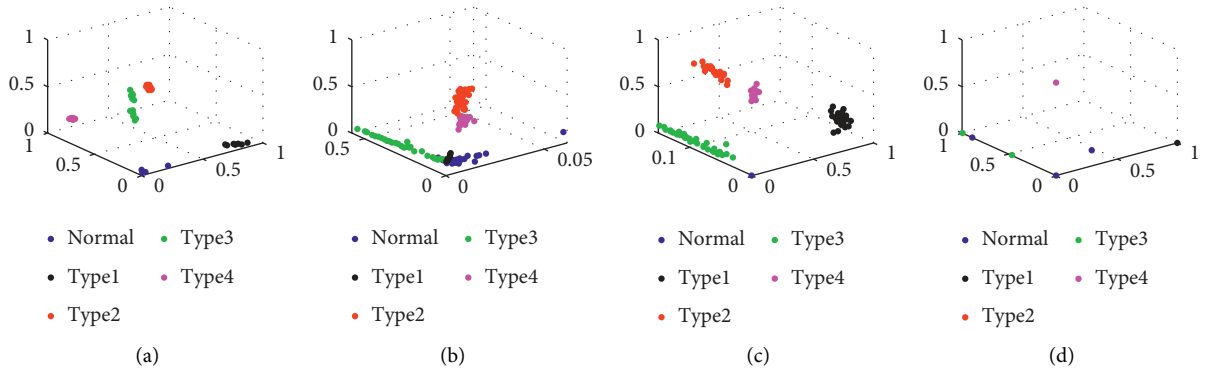


FIGURE 9: The classification results of different algorithms.

TABLE 6: Performance comparison of different algorithms.

Evaluation metrics	Algorithms			
	Improved BP-MLL	BP	BP-MLL	ML-KNN
Average precision	1	0.9155	0.9905	1
Hamming loss	0	0.1033	0.0406	0.0217
One error	0	0	0	0
F1-score	1	0.8967	0.9594	0.9783
Coverage	0.6	1.2292	1.0625	0.6
Ranking loss	0	0.1311	0.0104	0

Average Precision, which is also called classification accuracy or exact match ratio, computes the percentage of instances whose predicted labels are exactly the same as the actual corresponding set of labels:

$$AP = \frac{1}{I} \sum_{i=1}^I \frac{1}{|Y_i|} \sum_{\lambda' \in Y_i} \frac{|\tau_i(\lambda') \leq \tau_i(\lambda)|}{\tau_i(\lambda)}. \quad (27)$$

One error describes the possibility that top-ranked labels in one instance are not the actual labels in the proper set. The smaller the value of one error is, the better the system performance is

$$\text{One - error} = \frac{1}{I} \sum_{i=1}^I \left[ \arg \max_{\lambda \in y} \tau_i(\lambda) \notin y_i \right]. \quad (28)$$

Coverage measures how far the traversal all the labels is in the ranking averagely associated with an instance. The smaller the value is, the better the performance is

$$\text{Coverage} = \frac{1}{I} \sum_{i=1}^I \max_{\lambda \in Y_i} \tau_i(\lambda) - 1. \quad (29)$$

Table 6 summarizes the performance of each algorithm based on the evaluation metrics. All four algorithms get high f1-score: 1, 0.8967, 0.9594, and 0.9783, separately; particularly, improved BP-MLL obtained the highest one. The values of average precision are all above 0.9. Additionally, all these classifiers can recognize the relevant labels in each instance which leads to one error at 0. Improved BP-MLL and ML-KNN shared the same figures in coverage and ranking loss. Nevertheless, values of coverage for basic BP and conventional BP-MLL algorithms are twice as high as those for both improved BP-MLL and ML-KNN. In comparison, the Hamming loss of improved BP-MLL is slightly lower than ML-KNN, which is because ML-KNN shows misclassifications on single labels in some instances. Although values of different metrics vary, multilabel learning

algorithm can predict the simultaneous faults more effectively, among which improved BP-MLL outperforms the other compared two methods.

## 5. Conclusion

In this paper, an improved fault diagnostic method based on cohesion evaluation and improved BP-MLL classification is proposed. Compared with the conventional single-fault diagnosis, the problem of simultaneous faults occurring on gearbox and bearings of rotating machinery under different environmental conditions is investigated. On the basis of BP-MLL, this paper proposes a new global error function to deal with full-label and nonlabel learning situations through modifying its distance function and enhancing the generalization ability. Experiments conducted on WTDDS show that the proposed method is superior to conventional methods under six performance evaluation metrics. Although this paper has achieved good experimental results, there are still limitations such as the current algorithm being supervised learning only. Therefore, further studies can be focused on improvements such as semisupervised learning based on partial labels and transfer learning based on different working conditions.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Disclosure

An earlier version of this manuscript has been presented in 2020 IEEE 29th International Symposium on Industrial Electronics as An Improved Simultaneous Fault Diagnosis Method Based on Cohesion Evaluation and BP-MLL for Rotating Machinery.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was partially supported by National Natural Science Foundation of China (61603223), Jiangsu Provincial Qinglan Project, Research Development Fund of XJTU (RDF-18-02-30, RDF-20-01-18), Key Program Special Fund in XJTU (KSF-E-34), and the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (20KJB520034).

## References

- [1] I. El-Thalji and E. Jantunen, "A summary of fault modelling and predictive health monitoring of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 60-61, pp. 252–272.
- [2] Y. Lei, J. Lin, M. J. Zuo, and Z. He, "Condition monitoring and fault diagnosis of planetary gearboxes: a review," *Measurement*, vol. 48, pp. 292–305, 2014.
- [3] D. Cabrera, A. Guamán, S. Zhang et al., "Bayesian approach and time series dimensionality reduction to LSTM-based model-building for fault diagnosis of a reciprocating compressor," *Neurocomputing*, vol. 380, pp. 51–66, 2020.
- [4] J.-H. Zhong, J. Zhang, J. Liang, and H. Wang, "Multi-fault rapid diagnosis for wind turbine gearbox using sparse bayesian extreme learning machine," *IEEE Access*, vol. 7, pp. 773–781, 2019.
- [5] R. Maamouri, M. Trabelsi, M. Boussak, and F. M'Sahli, "Mixed model-based and signal-based approach for open-switches fault diagnostic in sensorless speed vector controlled induction motor drive using sliding mode observer," *IET Power Electronics*, vol. 12, no. 5, pp. 1149–1159, 2019.
- [6] D. Lieftucht, M. Völker, C. Sonntag, U. Kruger, G. W. Irwin, and S. Engell, "Improved fault diagnosis in multivariate systems using regression-based reconstruction," *Control Engineering Practice*, vol. 17, no. 4, pp. 478–493, 2009.
- [7] M. Pirra, E. Gandino, A. Torri, L. Garibaldi, and J. M. Machorro-López, "PCA algorithm for detection, localisation and evolution of damages in gearbox bearings," *Journal of Physics: Conference Series*, vol. 305, Article ID 012019, 2011.
- [8] M. Onel, C. A. Kieslich, Y. A. Guzman, and E. N. Pistikopoulos, "Simultaneous fault detection and identification in continuous processes via nonlinear support vector machine based feature selection," in *Computer Aided Chemical Engineering*, M. R. Eden, M. G. Ierapetritou, and G. P. Towler, Eds., vol. 44pp. 2077–2082, 2018.
- [9] Z. Wang, M. Zhong, R. Yang, and Y. Liu, "An improved random forest algorithm of fault diagnosis for rotating machinery," in *Proceedings of the 2019 CAA Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS)*, pp. 12–17, IEEE, Xiamen, China, July 2019.
- [10] Q. Lu, R. Yang, M. Zhong, and Y. Wang, "An improved fault diagnosis method of rotating machinery using sensitive features and RLS-BP neural network," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 4, pp. 1585–1593, 2019.
- [11] Y. Xue, D. Dou, and J. Yang, "Multi-fault diagnosis of rotating machinery based on deep convolution neural network and support vector machine," *Measurement*, vol. 156, Article ID 107571, 2020.
- [12] R. Yang, P. V. Er, Z. Wang, and K. K. Tan, "An RBF neural network approach towards precision motion system with selective sensor fusion," *Neurocomputing*, vol. 199, pp. 31–39, 2016.
- [13] R. Yang, M. Huang, Q. Lu, and M. Zhong, "Rotating machinery fault diagnosis using long-short-term memory recurrent neural network," *IFAC-PapersOnLine*, vol. 51, no. 24, pp. 228–232, 2018.
- [14] R. Yang, K. K. Tan, A. Tay et al., "An RBF neural network approach to geometric error compensation with displacement measurements only," *Neural Computing & Applications*, vol. 28, no. 6, pp. 1235–1248, 2017.
- [15] Z. Wan, R. Yang, and M. Huang, "Deep transfer learning-based fault diagnosis for gearbox under complex working conditions," *Shock and Vibration*, vol. 2020, Article ID 8884179, 13 pages, 2020.
- [16] Z. Wan, R. Yang, M. Huang, N. Zeng, and X. Liu, "A review on transfer learning in EEG signal analysis," *Neurocomputing*, vol. 421, pp. 1–14, 2021.

- [17] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: a review," *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, 2018.
- [18] L. Ungar, B. Powell, and S. Kamens, "Adaptive networks for fault diagnosis and process control," *Computers & Chemical Engineering*, vol. 14, no. 4, pp. 561–572, 1990.
- [19] M. A. Kramer and J. A. Leonard, "Diagnosis using back-propagation neural networks-analysis and criticism," *Computers & Chemical Engineering*, vol. 14, no. 12, pp. 1323–1338, 1990.
- [20] G. M. Knapp and H.-P. Wang, "Machine fault classification: a neural network approach," *International Journal of Production Research*, vol. 30, no. 4, pp. 811–823, 1992.
- [21] M. R. G. Meireles, P. E. M. Almeida, and M. G. Simoes, "A comprehensive review for industrial applicability of artificial neural networks," *IEEE Transactions on Industrial Electronics*, vol. 50, no. 3, pp. 585–601, 2003.
- [22] S. Wu and T. W. S. Chow, "Induction machine fault detection using som-based RBF neural networks," *IEEE Transactions on Industrial Electronics*, vol. 51, no. 1, pp. 183–194, 2004.
- [23] S. Haroun, A. Nait Seghir, and S. Touati, "Feature selection for enhancement of bearing fault detection and diagnosis based on self-organizing map," in *Recent Advances in Electrical Engineering and Control Applications*, pp. 233–246, Springer International Publishing, New York, NY, USA, 2017.
- [24] S.-L. Jämsä-Jounela, M. Vermasvuori, P. Endén, and S. Haavisto, "A process monitoring system based on the Kohonen self-organizing maps automation in Mining, Mineral and Metal Processing," *Control Engineering Practice*, vol. 11, no. 1, pp. 83–92, 2003.
- [25] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin, "A review of process fault detection and diagnosis," *Computers & Chemical Engineering*, vol. 27, no. 3, pp. 327–346, 2003.
- [26] A. Dineva, A. Mosavi, M. Gyimesi, I. Vajda, N. Nabipour, and T. Rabczuk, "Fault diagnosis of rotating electrical machines using multi-label classification," *Applied Sciences*, vol. 9, no. 23, p. 5086, 2019.
- [27] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 8, pp. 1819–1837, 2014.
- [28] M. R. Boutell, J. Luo, X. Shen, and C. M. Brown, "Learning multi-label scene classification," *Pattern Recognition*, vol. 37, no. 9, pp. 1757–1771, 2004.
- [29] A. Elisseeff and J. Weston, "A kernel method for multi-labelled classification," in *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, pp. 681–687, Granada, Spain, December 2002.
- [30] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: a lazy learning approach to multi-label learning," *Pattern Recognition*, vol. 40, no. 7, pp. 2038–2048, 2007.
- [31] J.-H. Zhong, P. K. Wong, and Z.-X. Yang, "Fault diagnosis of rotating machinery based on multiple probabilistic classifiers," *Mechanical Systems and Signal Processing*, vol. 108, pp. 99–114, 2018.
- [32] M. Zhi-Hua Zhou and Z. Zhou, "Multilabel neural networks with applications to functional genomics and text categorization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 10, pp. 1338–1351, 2006.
- [33] K. Zhang, H. Ma, Y. Zhao, H. Zan, and L. Zhuang, "The comparative experimental study of multilabel classification for diagnosis assistant based on Chinese obstetric emrs," *Journal of Healthcare Engineering*, vol. 2018, Article ID 7273451, 9 pages, 2018.
- [34] G.-P. Liu, G.-Z. Li, Y.-L. Wang, and Y.-Q. Wang, "Modelling of inquiry diagnosis for coronary heart disease in traditional Chinese medicine by using multi-label learning," *BMC Complementary and Alternative Medicine*, vol. 10, no. 1, p. 37, 2010.
- [35] Q. Yang, J. Shao, M. Scholz, C. Boehm, and C. Plant, "Multi-label classification models for sustainable flood retention basins," *Environmental Modelling & Software*, vol. 32, pp. 27–36, 2012.
- [36] W. Long, K. Zhang, H. Ma, D. Yue, and L. Zhuang, "Neural network multi-label learning based on enhancing pairwise labels discrimination for obstetric auxiliary diagnosis," in *Proceedings of the 2018 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, pp. 297–2977, Zhengzhou, China, October 2018.
- [37] R. Grodzicki, J. Mańdziuk, and L. Wang, "Improved multi-label classification with neural networks," *Parallel Problem Solving from Nature - PPSN X*, vol. 5199, no. 9, pp. 409–416, 2008.
- [38] K. Bunte, S. Haase, M. Biehl, and T. Villmann, "Stochastic neighbor embedding (SNE) for dimension reduction and visualization using arbitrary divergences," *Neurocomputing*, vol. 90, pp. 23–45, 2012.
- [39] E. Gibaja and S. Ventura, "A tutorial on multi-label learning," *ACM Computing Surveys*, vol. 47, p. 04, 2015.