

Research Article

A Framework for Final Drive Simultaneous Failure Diagnosis Based on Fuzzy Entropy and Sparse Bayesian Extreme Learning Machine

Qing Ye,¹ Hao Pan,¹ and Changhua Liu²

¹School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430000, China

²Yangtze University College of Technology and Engineering, Jingzhou 430023, China

Correspondence should be addressed to Qing Ye; yq0712@163.com

Received 6 October 2014; Revised 4 January 2015; Accepted 19 January 2015

Academic Editor: J. Alfredo Hernandez

Copyright © 2015 Qing Ye 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.

This research proposes a novel framework of final drive simultaneous failure diagnosis containing feature extraction, training paired diagnostic models, generating decision threshold, and recognizing simultaneous failure modes. In feature extraction module, adopt wavelet package transform and fuzzy entropy to reduce noise interference and extract representative features of failure mode. Use single failure sample to construct probability classifiers based on paired sparse Bayesian extreme learning machine which is trained only by single failure modes and have high generalization and sparsity of sparse Bayesian learning approach. To generate optimal decision threshold which can convert probability output obtained from classifiers into final simultaneous failure modes, this research proposes using samples containing both single and simultaneous failure modes and Grid search method which is superior to traditional techniques in global optimization. Compared with other frequently used diagnostic approaches based on support vector machine and probability neural networks, experiment results based on F_1 -measure value verify that the diagnostic accuracy and efficiency of the proposed framework which are crucial for simultaneous failure diagnosis are superior to the existing approach.

1. Introduction

With sustained increase of work condition complexity, simultaneous failures occur more frequently in final drive which is the pivotal part of car and seriously affect running status and comfort and safety of car. Final drive is mainly consisting of a pair of gears which are meshing together when car runs. Owing to the complex structure, a certain function disorder in final drive usually stems from more than one single failure at the same time which is called simultaneous failure. Traditional manual technology cannot accomplish simultaneous failure diagnosis (SFD). This paper is focusing on final drive simultaneous failure diagnosis which is essential for auto manufacturer and maintenance industry.

Failure diagnosis by using vibration signal is almost the most frequently used approach because vibration signal is relatively precise and accurate against diagnosis based on sound. It can be divided into three main steps: feature extraction, training diagnostic models, and failure mode

identification. The vibration signal collected from final drive has the characteristics of being nonlinear and nonstationary, and it is enclosed with a lot of uncorrelated and superfluous information. It is impossible to extract valid failure mode information from original vibration signal because of the noise and interference embedded in it. The frequently used preprocessed methods include wavelet analysis [1, 2], wavelet package transform (WPT) [3, 4], and empirical mode decomposition (EMD) [5]. Wavelet package transform is suitable for nonstationary vibration signal by decomposing original signal into several subfrequency bands which contains different failure information and effectively reduces noise interference.

Data contained in preprocessed signal is high-dimensional so that it cannot be directly inputted into diagnostic system. Feature extraction has a deep effect on accuracy and reliability of failure diagnosis. Recently, researchers have introduced entropy into the field of feature extraction including approximate entropy, sample entropy [6], and fuzzy entropy [7]. Compared with approximation entropy

and sample entropy which are based on Heaviside step function which is mutational at the classification boundary, fuzzy entropy eliminates the influence of baseline drift of data and guarantees the entropy to vary smoothly and continuously with similarity tolerance [8] so that it is excellent in measuring complexity and self-similarity of the preprocessed vibration signal and fully reflecting changes of the vibration performance of mechanical equipment [9].

In recent years, many machine learning methods are applied in failure diagnosis including support vector machine (SVM) [10], artificial neural networks (ANN) [11], extreme learning machine (ELM) [12, 13], and kernel extreme learning machine (KELM) [14]. ELM is single-hidden-layer feedforward neural networks and without human intervention in tuning parameters which differs from SVM and ANN and makes it superior in high generalization and less learning time. KELM apply kernel function to ELM to improve generalization and nonlinear approximation ability [15]. However, computational cost and memory cost of KELM are high with regard to large scale problem. Recently, Bayesian methods are employed into ELM to learn the output weights by estimating the probability distribution of output with high generalization. Soria-Olivas and Gómez-Sanchis proposed Bayesian extreme learning machine [16] for linear regression without solving classification problem. Sparse Bayesian extreme learning machine (SBELM) [17] is a novel method for finding the sparse representatives of hidden layer output weights by imposing a hyperparameter on each weight. During learning phase, SBELM tunes some output weights into zero to obtain compact model. In summary, SBELM has the advantages of probability output, high generalization, sparsity, and fast training speed. To solve the problem of simultaneous failure diagnosis, a proper classifier has to offer the probability of all possible failures. In this research, the proposed framework constructs classifiers based on paired SBELM in which each classifier based on SBELM is trained by a pair of single failure samples. The paired SBELM effectively reflect the probability distribution of failure modes. In general, only single failure samples are used for constructing diagnostic models. Since it is impossible to collect all combinations of existing single failure modes for training, the proposed framework can effectively solve the practical bottleneck in simultaneous failure diagnosis. With the purpose of recognizing simultaneous failure modes, use both single and simultaneous failure samples and Gird search method to generate optimal decision threshold which could convert probability result of classifier into final multiple failure modes. Considering that partial matching is valid and instrumental in simultaneous failure diagnosis, this research adopts F_I -measure to evaluate the performance of the proposed framework.

This paper is organized as follows: Section 2 presents the proposed framework. Section 3 presents the experiment setup, data acquisition, and preprocessing. The results of experiment are discussed in Section 4. Finally, a conclusion is given in Section 5.

2. The Proposed Framework for Final Drive Simultaneous Failure Diagnosis

2.1. Feature Extraction Based on Wavelet Package Transform and Fuzzy Entropy

2.1.1. Wavelet Package Transform. In failure diagnosis, one of the key points is extraction of features from original vibration signal which is nonstationary for mechanical equipment. Wavelet package transform (WPT) is an extended form of wavelet transform to analyse nonstationary and non-linear signal and to supply better partition of frequency band because the same frequency bandwidths can provide good resolution regardless of high and low frequencies [18]. As a multiresolution analysis method, WPT can effectively pre-process nonstationary vibration signal in both time domain and frequency domain. Two-scale equation of WPT is shown below, in which h_{0k} and h_{1k} represent the filter coefficients:

$$\begin{aligned} x_{2n}(t) &= \sqrt{2} \sum_{k \in \mathbb{Z}} h_{0k} x_n(2t - k), \\ x_{2n+1}(t) &= \sqrt{2} \sum_{k \in \mathbb{Z}} h_{1k} x_n(2t - k). \end{aligned} \quad (1)$$

The recursion formula of wavelet package coefficient is

$$d_k^{j+1,2n} = \sum_l h_{0(2l-k)} d_l^{j,n}, \quad d_k^{j+1,2n+1} = \sum_l h_{1(2l-k)} d_l^{j,n}. \quad (2)$$

2.1.2. Fuzzy Entropy. When failure occurred in final drive, the complexity of oscillation feature will change; hence we should extract representative features containing in the signal. Fuzzy entropy is an extension of Shannon entropy and fuzzy sets [19]. The procedure of fuzzy entropy is described as follows.

(1) Consider a time series with the length of N $\{x(i) : 1 \leq i \leq N\}$. For given m , n , and r , construct a vector set in the form of $\{X_i^m, i = 1, 2, \dots, N - m + 1\}$ in which each vector contains m sequential elements starting from $x(i)$ shown as

$$X_i^m = \{x(i), x(i+1), \dots, x(i+m-1)\} - x_0(i), \quad (3)$$

$$i = 1, 2, \dots, N - m + 1,$$

where $x_0(i)$ is the average of vector X_i^m .

(2) Define the distance d_{ij}^m between X_i^m and X_j^m where $i, j = 1, 2, \dots, N - m, i \neq j$ as follows:

$$d_{ij}^m = \max_{k \in (0, m-1)} \{ | [x(i+k) - x_0(i)] - [x(j+k) - x_0(j)] | \}. \quad (4)$$

(3) Calculate similarity between X_i^m and X_j^m using fuzzy function:

$$D_{ij}^m = \mu(d_{ij}^m, n, r) = e^{-\ln 2(d_{ij}^m/r)^n}. \quad (5)$$

(4) Define function ϕ^m as follows:

$$\phi^m(n, r) = \frac{1}{N - m} \sum_{i=1}^{N-m} \left(\frac{1}{N - m + 1} \sum_{j=1, j \neq i}^{N-m} D_{ij}^m \right). \quad (6)$$

(5) Change m to $m + 1$ and repeat step (1) to (4):

$$\phi^{m+1}(n, r) = \frac{1}{N-m} \sum_{i=1}^{N-m} \left(\frac{1}{N-m+1} \sum_{j=1, j \neq i}^{N-m} D_{ij}^{m+1} \right). \quad (7)$$

(6) Fuzzy entropy of sequence $\{x(i) : 1 \leq i \leq N\}$ is defined as follows:

$$\text{FuzzyEn}(m, n, r) = \lim_{N \rightarrow \infty} [\ln \phi^m(n, r) - \ln \phi^{m+1}(n, r)]. \quad (8)$$

(7) If the length N is finite, $\text{FuzzyEn}(m, n, r)$ can be changed as follows:

$$\text{FuzzyEn}(m, n, r) = \ln \phi^m(n, r) - \ln \phi^{m+1}(n, r). \quad (9)$$

2.2. Sparse Bayesian Extreme Learning Machine (SBELM). Given a preprocessed data set $D = (x_i, t_i)$, $i = 1, \dots, N$, $x_i \in R^n$, $t_i \in R^m$. The output function of ELM with L hidden nodes is shown as follows:

$$f(x) = \sum_{i=1}^L \beta_i h_i(x) = \beta \cdot H(x), \quad (10)$$

where $\beta = [\beta_1, \dots, \beta_L]^T$ is output weight connecting hidden nodes and output nodes; $H(x) = [h_1(x), \dots, h_L(x)]$ is the hidden layer output matrix for input x in which $h_i(x)$ is the hidden output of the i th hidden node. Equation (10) can be written as follows:

$$H\beta = T, \quad (11)$$

where T is the training data target matrix. SBELM learns output weight by using Bayesian method instead of by calculating Moore-Penrose generalized inverse of H [17]. The hidden layer output H becomes the input of SBELM. Treat each training sample as an independent Bernoulli event so that probability $p(t | x)$ satisfies Bernoulli distribution. Apply sigmoid function to convert the predicted output $\mathcal{Y}(h; \beta)$ as follows:

$$\sigma(\mathcal{Y}(h; \beta)) = \frac{1}{(1 + \exp(-\mathcal{Y}(h; \beta)))}. \quad (12)$$

The likelihood function of sample set is expressed as follows:

$$p(t | H, \beta) = \prod_{j=1}^N \sigma(\mathcal{Y}(h; \beta))^{t_j} [1 - \sigma(\mathcal{Y}(h; \beta))]^{1-t_j}, \quad (13)$$

where t_j is the target of training sample x_j , $\mathcal{Y}(h; \beta) = h\beta$, and $t_j \in \{0, 1\}$. Conditioned on a hyperparameter α_j , zero-mean Gaussian prior distribution over β_i is as follows:

$$p(\beta | \alpha) = \prod_{i=1}^L \frac{\alpha_i}{\sqrt{2\pi}} \exp\left(-\frac{\alpha_i \beta_i^2}{2}\right). \quad (14)$$

The typical step of SBELM is to establish the distribution of marginal likelihood over t conditioned on α and H and

determine α by maximizing the marginal likelihood $p(t | H, \alpha)$ by Laplace approximation method:

$$\begin{aligned} & \operatorname{argmax} \ln \{p(t | \beta, H) p(\beta | \alpha)\} \\ &= \operatorname{argmax} \left[\ln \left\{ \prod_{j=1}^N y_j^{t_j} (1 - y_j)^{1-t_j} \right\} \right. \\ & \quad \left. + \ln \left\{ \prod_{i=1}^L \frac{\alpha_i}{\sqrt{2\pi}} \exp\left(-\frac{\alpha_i \beta_i^2}{2}\right) \right\} \right] \\ &= \operatorname{argmax} \left[\sum_{j=1}^N \left\{ t_j \ln y_j + (1 - t_j) \right. \right. \\ & \quad \left. \left. \cdot \ln(1 - y_j) - \frac{\beta^T A \beta}{2} + \text{const} \right\} \right], \quad (15) \end{aligned}$$

where $y_j = \sigma(\mathcal{Y}(h_j; \beta))$, $A = \text{diag}(\alpha)$, and $\text{const} = \sum_{i=1}^L \ln \alpha_i - 1/2 \ln 2\pi$. Then, make quadratic approximation for log of posterior probability:

$$\nabla \beta \nabla \beta \ln \{p(t | \beta, H) p(\beta | \alpha)\} = -(H^T K H + A), \quad (16)$$

where K is a diagonal matrix in which $k_j = y_j(1 - y_j)$ with $j = 1, \dots, N$. Therefore, the center and covariance matrix of Gauss distribution of β expressed as β' and Φ are obtained as follows:

$$\beta' = \Phi H^T K t', \quad \Phi = (H^T K H + A)^{-1}, \quad (17)$$

where $t' = H\beta + K^{-1}(t - y)$. By obtaining Gauss approximation of β , the log of marginal likelihood is represented as follows:

$$\begin{aligned} \mathcal{L}(\alpha) &= \ln p(t | \alpha, H) \\ &= -\frac{1}{2} \left[N \ln(2\pi) + \ln |C| + (t')^T C^{-1} t' \right], \quad (18) \end{aligned}$$

where $C = K + H A H^T$. By setting the differential of $\mathcal{L}(\alpha)$ with respect to α as 0, update the hyperparameter α as follows:

$$\alpha_i = \frac{1 - \alpha_i \Phi_{ii}}{(\beta_i')^2}. \quad (19)$$

The main procedure of SBELM is described as follows.

- (1) Initialize α_i and β_i randomly with $i = 1, \dots, L$.
- (2) By utilizing Laplace approximation approach, obtain approximated Gauss distribution of β and update β' and Φ by using (17).
- (3) By maximizing the marginal likelihood, utilize (19) to update hyperparameter α until reaching the termination criteria.
- (4) By tuning some β_i into 0, obtain the sparse representation of hidden layer output weight.
- (5) For an unknown sample x_u , utilize (12) to predict probability distribution $p(t | x_u, \beta')$.

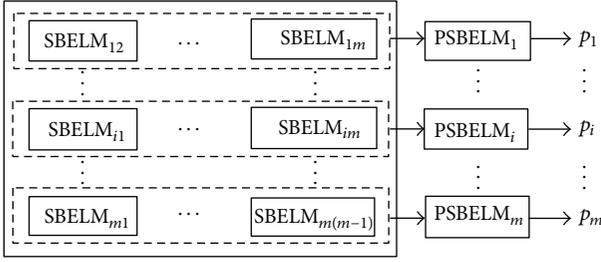


FIGURE 1: Structure of paired SBELM for simultaneous failure diagnosis.

2.3. Paired SBELM. SBELM is excellent in solving binary classification by obtaining probability distribution of each class $p(t | x, \beta')$. With the purpose of final drive simultaneous failure diagnosis in which training samples of single failure are ample while training samples of simultaneous failure are scarce, this paper combines the state-of-the-art coupling approach proposed in [20] with SBELM to construct a set of paired SBELM (PSBELM) classifiers expressed as $[\text{PSBELM}_1, \dots, \text{PSBELM}_m]$ for a m -label classification problem shown in Figure 1 and each paired classifier $\text{PSBELM}_i = [\text{SBELM}_{i1}, \dots, \text{SBELM}_{ij}, \dots, \text{SBELM}_{im}]$ in which SBELM_{ij} is trained by every pair of classes and its output is $p_{ij}(t_i | \beta, x)$ for sample x belonging to the i th against the j th class. The total number of classifiers is $m(m-1)/2$.

In simultaneous failure diagnosis, more than one failure may occur at the same time that can infer the concept: $\sum_{i=1}^m p_i \neq 1$ [21]. By estimating each probability output of binary classifiers SBELM_{ij} to measure correlation between various classes, obtain the paired probability output p_i as follows:

$$p_i = \frac{\sum_{j=1, j \neq i}^d n_{ij} p_{ij}}{\sum_{j=1, j \neq i}^d n_{ij}}, \quad i = 1, \dots, m, \quad (20)$$

where n_{ij} is the number of training sample belonging to the i th and the j th class.

2.4. Optimization of Decision Threshold for Simultaneous Failure Mode Recognition. For a m -class classification problem, the output of classifiers based on PSBELM is a probability vector $P = [p_1, \dots, p_m]$ in which p_j represents occurring possibility of the j th failure. In order to obtain final simultaneous failure modes, an appropriate threshold value is indispensable. In general, researchers usually use 0.5 to be the threshold value [21] which is of generality but not suitable for specific application. This paper utilizes Grid Search method and an independent sample set containing both single failure and simultaneous failure to generate an optimal decision threshold ε^* between 0 and 1 which can convert probability output vector into result vector $F = [f_1, \dots, f_i, \dots, f_m]$ effectively:

$$f_i = \begin{cases} 1 & p_i \geq \varepsilon^* \\ 0 & p_i < \varepsilon^* \end{cases}, \quad i = 1, \dots, m, \quad \varepsilon^* \in (0, 1). \quad (21)$$

The simultaneous failure modes are those single failures that their corresponding f_i is equal to 1. Since the range of searching is limited, the time-consuming characteristic of global optimization compared with GA and PSO [18].

2.5. Evaluation of Performance Based on F_1 -Measure. Considering that partial matching is valid and significant in simultaneous failure diagnosis [22], utilize an independent testing set and F_1 -measure [23] which is commonly used for evaluation of information retrieval systems to evaluate diagnostic accuracy for the proposed simultaneous failure diagnostic framework. Given a data set $D = (x_i, t_i)$, $i = 1, \dots, N$, $x_i \in R^n$, $t_i \in R^m$, $t_{ij} \in \{0, 1\}$, $j = 1, \dots, m$, define two variables namely precision (P) and recall (R) among which P represents the ratio between correct identified single failure modes and the actual simultaneous failure modes and R represents the ratio between correct identified single failure modes and the predicted simultaneous failure modes:

$$P = \frac{\sum_{j=1}^m \sum_{i=1}^N f_{ij}^* t_{ij}}{\sum_{j=1}^m \sum_{i=1}^N t_{ij}}, \quad R = \frac{\sum_{j=1}^m \sum_{i=1}^N f_{ij}^* t_{ij}}{\sum_{j=1}^m \sum_{i=1}^N f_{ij}^*}, \quad (22)$$

where $f_i^* = [f_{i1}^*, \dots, f_{im}^*]$ is the predicted simultaneous failure modes by using the proposed framework and $t_i = [t_{i1}, \dots, t_{im}]$ is the actual simultaneous failure modes of x_i . The F_1 -measure value can be obtained as follows:

$$E = \frac{2 * P * R}{P + R} = \frac{2 \sum_{j=1}^m \sum_{i=1}^N f_{ij}^* y_{ij}}{\sum_{j=1}^m \sum_{i=1}^N y_{ij} + \sum_{j=1}^m \sum_{i=1}^N f_{ij}^*}. \quad (23)$$

2.6. The Proposed Framework for Final Drive Simultaneous Failure Diagnosis. The structure of the proposed framework is shown in Figure 2 and the procedure of the proposed framework is described as follows.

- (1) Divide sample set into four parts: $D_{\text{training1}}$, $D_{\text{training2}}$, $D_{\text{threshold}}$, and D_{testing} . All the sample should be preprocessed by WPT and utilize fuzzy entropy to measure the feature of oscillatory.
- (2) Utilize $D_{\text{training1}}$ containing only single failure modes to optimize parameters of WPT including number of layer L and mother wavelet in preprocessing by using failure diagnostic model of SBELM.
- (3) By using optimal parameters of WPT obtained from step 2, preprocess $D_{\text{training2}}$ containing only single failure modes and train classifiers based on paired SBELM.
- (4) The optimum diagnostic model of PSBELM generates a probability output vector $[p_1, \dots, p_m]$ in m -label classification problem and uses $D_{\text{threshold}}$ containing both single and simultaneous failure modes and Grid Search to confirm the decision threshold value ε^* which is used to obtain simultaneous failure modes.
- (5) Use D_{testing} and F_1 -measure to evaluate the diagnostic accuracy of the proposed framework.

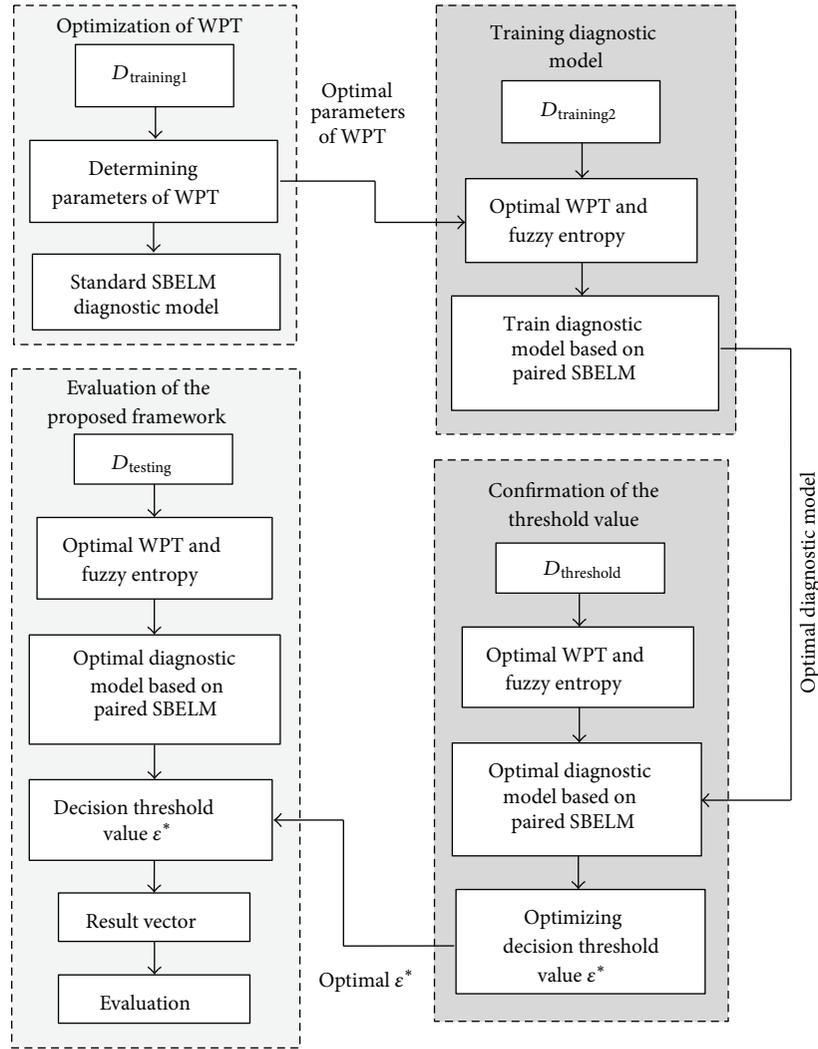


FIGURE 2: The structure of the proposed framework.

3. Experiment Setup and Preprocessing

3.1. Experiment Setup. In order to obtain sample data with representativeness for constructing diagnostic model and verify the efficiency of the proposed diagnostic platform, implement the experiments on a test bed containing a PC, two sensors, a signal amplifier, and a simulation turntable with the composition as shown in Figure 3 in quiet room to collect enough original vibration signal of final drive. Two sensors are laid on the final drive in horizontal and vertical direction as shown in Figure 4 to collect vibration signal when it is set into running state. Most failures of final drive such as gear error, gear hard point, and tooth broken occurred in gear pair which is the hard core of final drive and consisted of a driving-gear and a driven-gear. In this research simulate 9 common failures including 6 single failures and 3 simultaneous failures which are described in detail in Table 1 under the rotating speed of 1200 r/m for driving motor. As shown in Figure 5, amplitudes of simultaneous failure are obviously greater than single failure because when simultaneous failure occurs these single failures are coupled together severely. The wave profiles

between single and simultaneous failure patterns are similar, so that it is difficult to distinguish them manually but the characteristics embedded in each vibration signal can be extracted and identified by using methods afore mentioned.

Considering the universality of vibration signals which are used to construct diagnostic models, repeat simulating each failure mode for 100 times and record the most stable 2 seconds in each time with the sampling rate of 12 kHz which should be higher than the gear meshing frequency so that effective failure information may not be discarded during the sampling. Eventually, 1000 sample data are obtained and prepared to be preprocessed. All the simulations are implemented in MATLAB 7.0 which is running in a PC with CPU of 3.4 GHZ and RAM of 4.0 GB.

3.2. Feature Extraction Based on WPT and Fuzzy Entropy. In this paper, fuzzy entropy is used to reflect the change of complexity. By calculating the average fuzzy entropy of vibration signal corresponding to each failure mode which is shown in Table 2, we find out that the values of fuzzy entropy of these 10 failure modes are approximate so that it can not

TABLE 1: Simple and simultaneous failure modes description.

Failure label		Failure mode description
C1	Single failures	Normal status
C2		Gear error
C3		Gear burr
C4	Single failures	Gear hard point
C5		Misalignment
C6		Gear tooth broken
C7		Gear crack
C8		Gear tooth broken and gear crack
C9	Simultaneous failures	Gear tooth broken and gear hard point
C10		Misalignment and gear crack

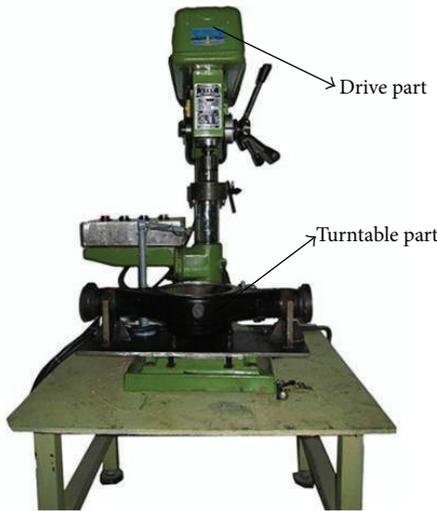


FIGURE 3: The test bed.

correctly distinguish failures of final drive. The reason is that information supplied by fuzzy entropy of original signal is limited and unable to reflect the deep-seated information of each failure situation. Therefore, we employ the frequently used mother wavelet function called Daubechies wavelet which has orthogonal characteristic and effectiveness in filtering signal of vibrating machinery to implement wavelet package transform and decompose original vibration signal into several subfrequency bands. By calculating fuzzy entropy of each frequency band obtained from L -level wavelet package transform decomposition, construct a feature vector with the dimension of 2^L which can effectively reflect the complexity and self-similarity of oscillation characteristic of failure modes occurring in final drive:

$$\text{Feature} = [\text{FuzzyEn1}, \text{FuzzyEn2}, \dots, \text{FuzzyEn}2^L]. \quad (24)$$

3.3. Distribution Plan of Samples. There are 1000 sample data containing 100 normal samples, 600 single failure samples,

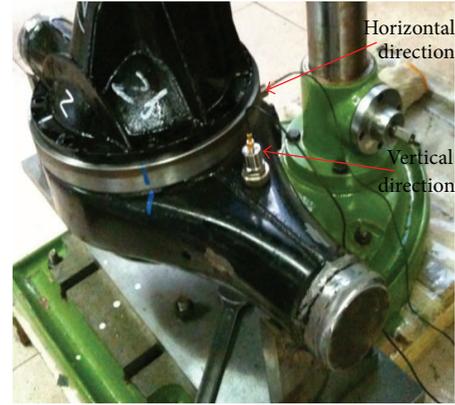


FIGURE 4: Two sensors on final drive.

and 300 simultaneous failure samples. In each trial, randomly divide the whole sample set into four parts: $D_{\text{training1}}$, $D_{\text{training2}}$, $D_{\text{threshold}}$, and D_{testing} . $D_{\text{training1}}$ and $D_{\text{training2}}$ which are consisting of only single failure modes are used for optimizing parameters of WPT and training optimal diagnostic model based on paired SBELM. $D_{\text{threshold}}$ which contains both single and simultaneous failure modes is used to generate optimal decision threshold which convert probability result of diagnostic model based on paired SBELM into final simultaneous failure modes. D_{testing} is used to test and evaluate the proposed framework by using F_1 -measure. Ensure that the whole sample set be preprocessed and the size of training samples should be more than testing samples to ensure the generalization of the proposed framework. The distribution plan is shown in Table 3.

4. Result and Discussion

4.1. Optimization of Preprocessing and Feature Extraction. In data preprocessing, optimum combination of decomposition level and mother wavelet of WPT and parameters of fuzzy entropy can achieve better performance in classification. We use $D_{\text{training1}}$ containing random 250 single samples to obtain the optimal combination of level number and mother wavelet which is suitable for preprocessing samples collected from final drive. In order to simplify experiment and on the basis of previous research result, focus on three wavelets Db3, Db4, and Db5 and three decomposition levels from 3 to 5. Three parameters of fuzzy entropy including m , n , and r are defined empirically in advance. Parameter m is usually set to be 2. Related to the boundary of fuzzy function, parameters r and n are setting as 0.2 and 2 STD where STD is the standard deviation of original data [8].

By using single failure samples contained in $D_{\text{training1}}$ and standard diagnostic model based on SBELM with failure parameters, find out appropriate parameters of WPT to achieve best performance of preprocessing. The standard diagnostic model based on SBELM is only used for selecting optimal parameters of WPT that exist in the best failure identification model in which the accuracy of classification is highest. The comparison result is shown in Figure 6 which indicates that classification accuracy of the preprocessing by

TABLE 2: Average fuzzy entropy of 10 failure modes.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Average fuzzy entropy	1.7761	1.8695	1.8855	1.7904	1.8734	1.9018	1.8080	1.8653	1.8107	1.9268

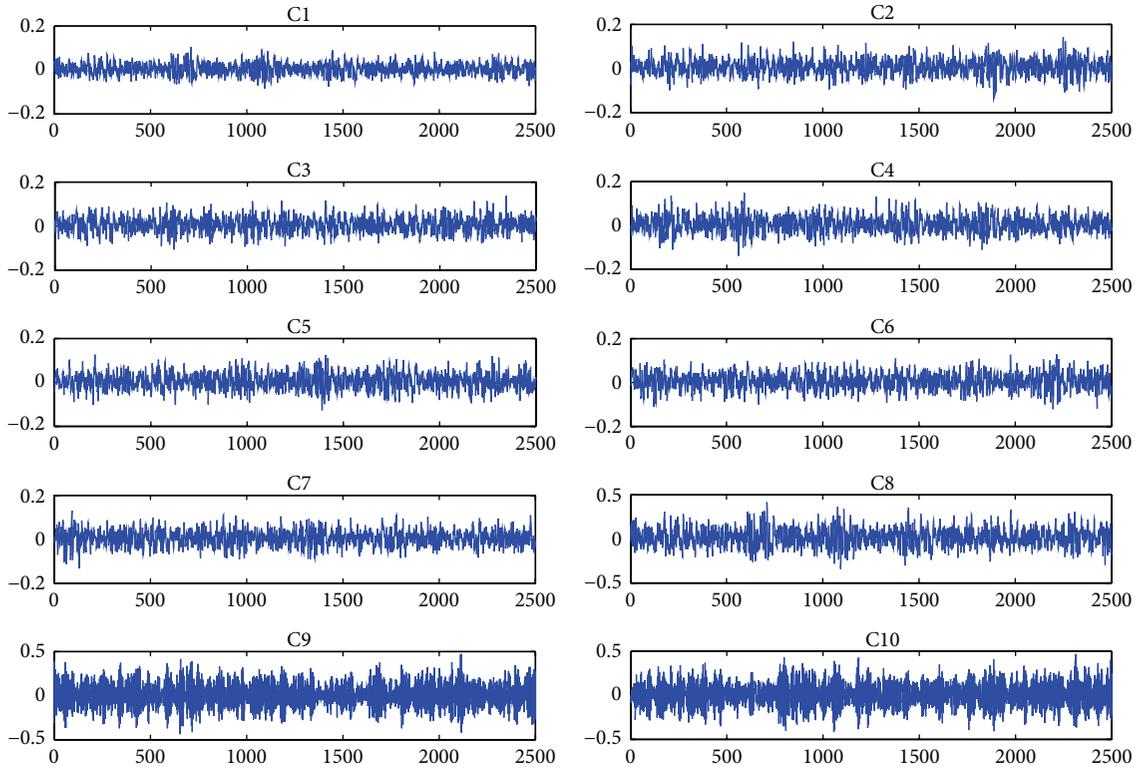


FIGURE 5: Vibration waveforms of 9 failure modes and normal status.

TABLE 3: Division of the whole sample dataset.

	Single failure	Simultaneous failure	Total number
$D_{\text{training1}}$	250		250
$D_{\text{training2}}$	250		250
$D_{\text{threshold}}$	100	200	300
D_{testing}	100	100	200
Total	700	300	1000

using 3 level decomposition and Db4 as mother wavelet and standard diagnostic model based on SBELM is highest with the accuracy of 95.2%. This parameter combination of WPT is suitable for preprocessing the dataset in this application.

After decomposing vibration signal by using three-level wavelet package decomposition, calculate the corresponding value of fuzzy entropy as shown in Figure 7. In Figure 7, horizontal ordinate represents eight subfrequency bands of three-level wavelet package decomposition, and longitudinal coordinate represents the fuzzy entropy value. The FuzzyEn of the oscillation from final drive with simultaneous failures is larger than that of single failures and normal status. When simultaneous failures occur under rotation of gear pair, different failure points are coupling together to make the

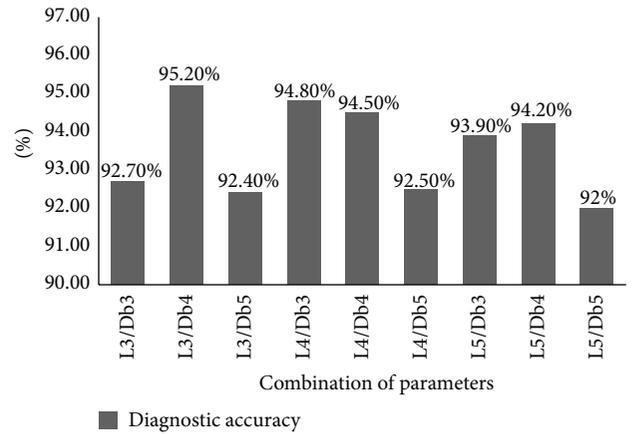


FIGURE 6: The diagnostic accuracy of different parameters.

oscillation complex and stronger. Furthermore, the values of fuzzy entropy vary from one failure pattern to another. This characteristic denotes fuzzy entropy can be used as feature of failure diagnosis.

By calculating fuzzy entropy of each frequency band obtained from three-level wavelet package decomposition, we

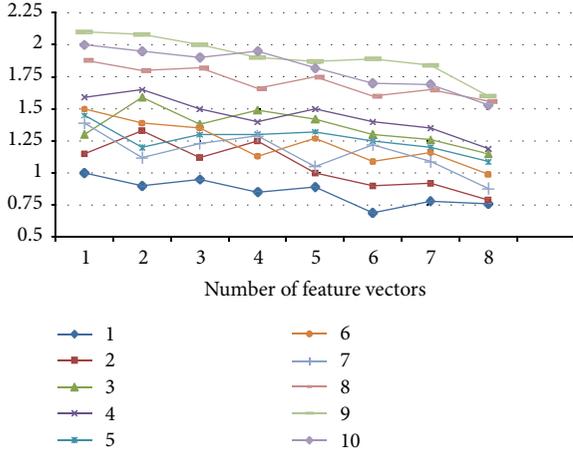


FIGURE 7: Mean value of fuzzy entropy for failure modes.

construct a feature vector with the dimension of 8 which can effectively reflect the failure modes of final drive:

$$\text{Feature} = [\text{FuzzyEn1}, \text{FuzzyEn2}, \dots, \text{FuzzyEn8}]. \quad (25)$$

4.2. Effectiveness of Optimal Decision Threshold. After constructing optimal diagnostic model based on paired SBELM with optimal parameters of WPT in preprocessing by using only single failure modes, generation of optimal decision threshold is the pivotal point which affects final diagnostic accuracy of simultaneous failure. Traditional machine learning methods usually adopt 0.5 as general threshold value (GT) [24]. This research uses $D_{\text{threshold}}$ containing 100 single failure modes and 200 simultaneous failure modes and Grid Search method with interval of 0.01 to search final decision threshold ε^* in range of 0 to 1. Although Grid Search is time consuming, it can obtain global optimum.

With the purpose of verifying effectiveness of optimal decision threshold, utilize 5-fold cross validation method to implement a set of experiments by using $D_{\text{threshold}}$ for both single and simultaneous failure modes recognition. Results are shown in Figure 8.

After optimizing threshold, the accuracy of diagnostic model improves by an average of 6%. Fixed General threshold is generated by experience so that it has generalization but without optimization [25]. Even using the same diagnostic model to diagnose different sample set would require different threshold. Therefore, this research uses an independent sample set to generate optimal decision threshold.

4.3. Sensitivity Analysis of SBELM. For diagnosis based on ELM, diagnostic accuracy and training speed are sensitive to the initial number of hidden nodes. To analyze the sensitivity of SBELM on the number of hidden nodes in this application, use 500 single failure samples in $D_{\text{training1}}$ and $D_{\text{training2}}$ to train classifier based on SBELM and the best average accuracy along with the increase of hidden nodes is shown in Figure 9. As shown in Figure 9, the average accuracies of ELM with increment of hidden nodes are in larger variation. The reason for this fluctuation is that ELM

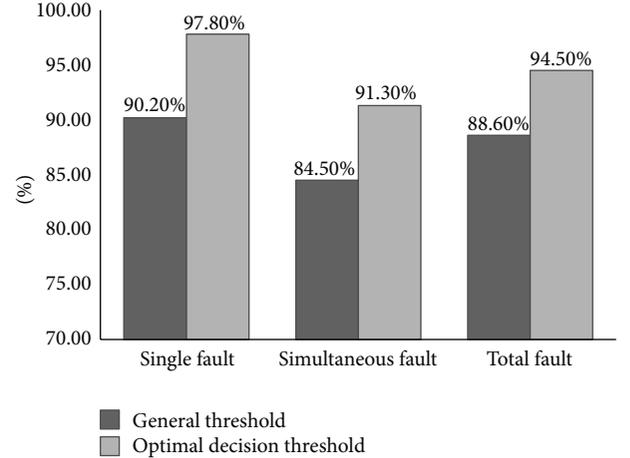


FIGURE 8: Diagnostic accuracies of models with general threshold and optimal decision threshold.

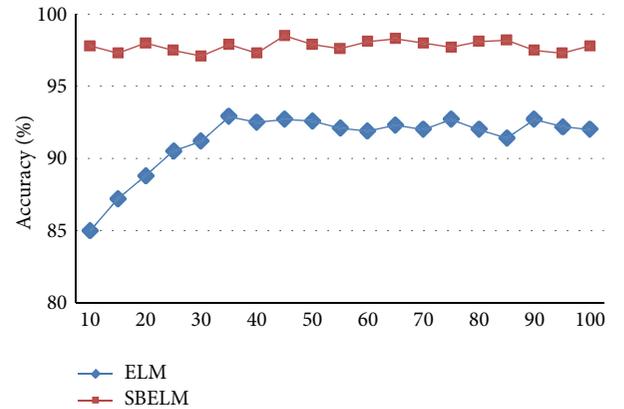


FIGURE 9: Variation of accuracy of ELM and SBELM.

is in poor generalization because of data overfitting [17]. However, the average accuracies are stable and are obviously higher than ELM. The result verifies that SBELM is relatively insensitive to the initial number of hidden nodes. Moreover, SBELM can obtain an excellent accuracy with a small hidden layer which reduces the computational cost effectively.

4.4. Evaluation of the Proposed Framework. In order to effectively confirm the availability of the proposed simultaneous failure diagnosis framework, we use D_{testing} containing 100 single failure modes and 100 simultaneous failure modes and F_1 -measure method to measure performance of the proposed framework and diagnostic model based on PNN and SVM in diagnostic accuracy and diagnostic speed. Firstly, use sample set which are consisting of $D_{\text{training1}}$ and $D_{\text{training2}}$ to construct and tune parameters of diagnostic model based on PNN and SVM separately and then use $D_{\text{threshold}}$ to generate optimal threshold value. Since SVM is essentially used for binary-class classification [26], with the purpose of simultaneous failure diagnosis we combine SVM with multiclass classification strategy to construct a set of classifiers in which each classifier

TABLE 4: Comparison of paired strategy and one-to-all strategy.

Classifier	Decision threshold $\in [0, 1]$	Multiclass classification strategy	Accuracy (%)		
			Single failures	Simultaneous failures	Entire sample
PNN	0.69	One-to-all	91.54 (± 1.02)	88.22 (± 1.55)	89.42 (± 1.67)
		paired strategy	93.21 (± 1.25)	89.14 (± 1.76)	92.33 (± 2.05)
SVM	0.69	One-to-all	92.02 (± 2.35)	80.90 (± 1.62)	84.70 (± 1.87)
		paired strategy	94.84 (± 1.75)	84.32 (± 1.35)	88.92 (± 2.14)
SBELM	0.72	One-to-all	95.13 (± 1.22)	90.54 (± 2.05)	92.94 (± 1.73)
		paired strategy	98.42 (± 1.41)	92.81 (± 2.35)	96.23 (± 1.59)

is only focusing on two failure modes. Trying to ensure the excellent performance of classifiers based on SVM, set the value of regularization parameter C of SVM to be 10^α where α is between 0 and 2. Radial basis kernel function is employed in SVM with $C = 10$ and $r = 2$ which show the best accuracy of classification. As a probability classifier, the crucial hyperparameter of PNN is spread s . In this research, the value of s is chosen from 1 to 3 with interval of 0.5 according to conclusion of references. Finally, the best hyperparameters s and threshold value ε for PNN are 1 and 0.69.

To verify the effectiveness of the paired strategy in the proposed framework, implement a set of experiments with one-to-all strategy. The experimental results are shown in Table 4. Comparing different classifiers with one-to-all strategy and paired strategy, the accuracies of classifiers with paired strategy are generally 2% to 4% higher than that of classifiers with one-to-all strategy. The primary reason is that paired strategy which is used in the proposed framework fully considers the correlation between each single failure. However, one-to-all strategy may cause some indecision regions between different classes. The indecision region is prone to sinking into misclassification.

To verify the performance of the proposed framework, implement a set of experiments about different classifiers with the same testing set and best parameters. The decision threshold values, training time, testing time, and testing accuracy of diagnostic models based on paired SBELM, SVM, and PNN are shown in Table 5. The diagnostic accuracy of paired SBELM for single failure, simultaneous failure, and entire sample is 98.4%, 92.8%, and 96.2% which are higher than that of the SVM and PNN. The reason is that SBELM estimates the probability distribution of output values instead of fitting data to improve generalization [17]. Moreover, the training time and testing time of paired SBELM are 145.4 ms and 48.7 ms that are much fewer than SVMs. The reason for this disparity is that even though paired SBELM builds a set of binary classifiers, the sparse characteristic of SBELM reduces the computational cost. Consequently, the disparity will become more obvious if the size of sample is big.

In practical application of auto manufacturer, representative and valid samples are continuously collected and added to the training sample database to improve training accuracy. Based on this, learning speed becomes a crucial factor for evaluating the efficiency of diagnostic platform. In general, considering both diagnostic accuracy and diagnostic efficiency, the proposed platform is superior in simultaneous

TABLE 5: Performance of three classifiers.

	PNN	SVM	SBELM
Decision threshold [0, 1]	0.69	0.69	0.76
Accuracy of single failure (%)	93.24 (± 1.86)	94.81 (± 2.19)	98.42 (± 1.55)
Accuracy of simultaneous failure (%)	89.12 (± 2.41)	87.34 (± 1.96)	92.81 (± 1.85)
Accuracy of entire sample (%)	92.30 (± 2.55)	91.43 (± 2.32)	96.23 (± 2.06)
Training time (ms)	268.2	493	145.4
Testing time (ms)	86.5	119.4	48.7

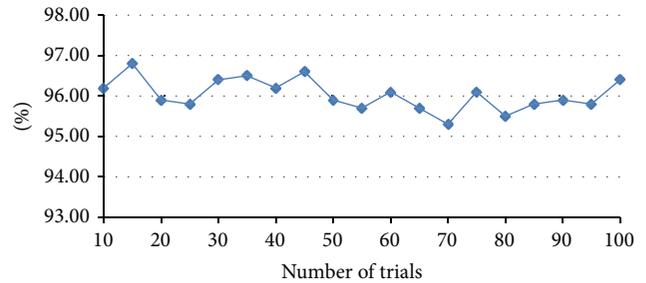


FIGURE 10: Testing result of 100 trials.

failure diagnosis and it is not only suitable for final drive of car but also it can be porting to other research fields.

In order to verify the stability of the proposed diagnostic framework based on paired SBELM, implement 100 trials and in each trial the whole sample data is reshuffled and randomly distributed into D_{testing} afresh and make sure there are enough single failure samples and simultaneous failure samples in D_{testing} . The testing result is shown in Figure 10 in which the testing accuracy is stable in the range between 95% and 97% and there is no dramatic variation in 100 simulation trials.

5. Conclusion

This paper proposes a novel framework based on SBELM and fuzzy entropy for simultaneous failure diagnosis of final drive which is hardcore to affect the performance

and safety of car. The proposed framework contains four sections: preprocessing and feature extraction based on WPT and fuzzy entropy, construction of diagnostic model based on paired SBELM, generation of decision threshold value, and recognition of simultaneous failure modes. By using single failure samples, obtain optimal parameters of WPT which are perfectly adequate for the data in this application. Diagnostic model based on paired SBELM in which each binary classifier is trained by only single failure samples. With an independent sample subset containing both single and simultaneous failure samples, use Grid Search method to generate optimal decision threshold by which probability result obtained from diagnostic model can be converted into final result of simultaneous failure modes. Compared with frequently used diagnostic model based on SVM and PNN, there are three superiorities of the proposed framework. (1) The proposed framework based on SBELM inherits the advantages of ELM (efficient approximation and learning speed) and sparse Bayesian learning (high sparsity and generalization). (2) Fully considering the difficulty and impossibility of assembling all possible simultaneous failure modes, the proposed framework trains paired classifiers based on SBELM by using only single failure samples, and moreover the paired strategy can effectively avoid indecision regions between different classes which can result in misclassification. (3) With the average testing accuracy of 96.2% and testing time of 48.7 ms, the proposed framework outperforms other diagnostic models in diagnostic accuracy and learning speed. The proposed framework is general and transplantable for simultaneous failure diagnosis, so it can be applied to other applications in industrial area in which accuracy and time cost of failure identification are key factors.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

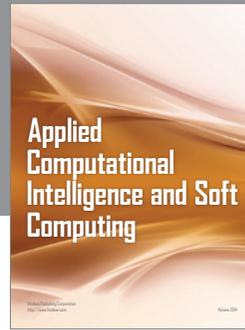
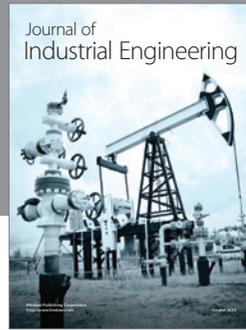
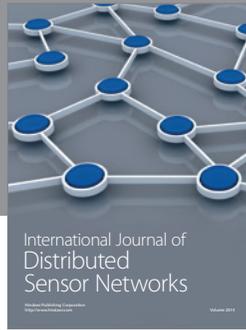
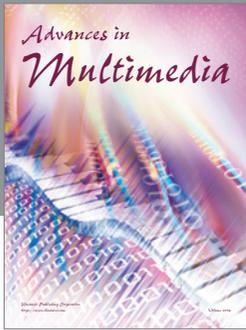
Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant no. 70701013, the Scientific Research and Technology Development Plan Project of Guangxi Province under Grant no. 2013F020202, and the Research Project of Liuzhou GM-Wuling Limited Liability Company under Grant no. 20132h0261. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the paper.

References

- [1] X. Chiementin, B. Kilundu, L. Rasolofondraibe, S. Crequy, and B. Pottier, "Performance of wavelet denoising in vibration analysis: highlighting," *Journal of Vibration and Control*, vol. 18, no. 6, pp. 850–858, 2012.
- [2] J. Rafiee, M. A. Rafiee, and P. W. Tse, "Application of mother wavelet functions for automatic gear and bearing fault diagnosis," *Expert Systems with Applications*, vol. 37, no. 6, pp. 4568–4579, 2010.
- [3] J.-D. Wu and J.-J. Chan, "Faulted gear identification of a rotating machinery based on wavelet Transform and artificial neural network," *Expert Systems with Applications*, vol. 36, no. 5, pp. 8862–8875, 2009.
- [4] M. Vannucci and V. Colla, "Novel classification method for sensitive problems and uneven datasets based on neural networks and fuzzy logic," *Applied Soft Computing Journal*, vol. 11, no. 2, pp. 2383–2390, 2011.
- [5] A. Janušauskas, V. Marozas, and A. Lukoševičius, "Ensemble empirical mode decomposition based feature enhancement of cardio signals," *Medical Engineering and Physics*, vol. 35, no. 8, pp. 1059–1069, 2013.
- [6] W. T. Chen, Z. Z. Wang, H. B. Xie, and W. Yu, "Characterization of surface EMG signal based on fuzzy entropy," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 15, no. 2, pp. 266–272, 2007.
- [7] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate and sample entropy," *The American Journal of Physiology—Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.
- [8] J. Zheng, J. Cheng, and Y. Yang, "A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy," *Mechanism and Machine Theory*, vol. 70, pp. 441–453, 2013.
- [9] G. L. Xiong, L. Zhang, H. S. Liu, H. J. Zou, and W.-Z. Guo, "A comparative study on ApEn, SampEn and their fuzzy counterparts in a multiscale framework for feature extraction," *Journal of Zhejiang University: Science A*, vol. 11, no. 4, pp. 270–279, 2010.
- [10] S. Deng, S. Y. Lin, and W. L. Chang, "Application of multiclass support vector machines for fault diagnosis of field air defense gun," *Expert Systems with Applications*, vol. 38, no. 5, pp. 6007–6013, 2011.
- [11] W. Jatmiko, W. P. Nulad, M. I. Elly, I. M. A. Setiawan, and P. Mursanto, "Heart beat classification using wavelet feature based on neural network," *WSEAS Transactions on Systems*, vol. 10, no. 1, pp. 17–26, 2011.
- [12] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.
- [13] E. Cambria, G.-B. Huang, L. L. C. Kasun et al., "Extreme learning machines," *IEEE Intelligent Systems*, vol. 28, no. 6, pp. 30–59, 2013.
- [14] Q. Yuan, W. Zhou, S. Li, and D. Cai, "Epileptic EEG classification based on extreme learning machine and nonlinear features," *Epilepsy Research*, vol. 96, no. 1–2, pp. 29–38, 2011.
- [15] 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.
- [16] E. Soria-Olivas and J. Gómez-Sanchis, "BELM: bayesian extreme learning machine," *IEEE Transaction on Neural Networks*, vol. 22, no. 3, pp. 505–509, 2011.
- [17] J. Luo, C.-M. Vong, and P.-K. Wong, "Sparse bayesian extreme learning machine for multi-classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 4, pp. 836–843, 2014.
- [18] Z. Yang, P. K. Wong, C. M. Vong, J. Zhong, and J. Liang, "Simultaneous-fault diagnosis of gas turbine generator systems using a pairwise-coupled probabilistic classifier," *Mathematical Problems in Engineering*, vol. 2013, Article ID 827128, 14 pages, 2013.

- [19] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: artificial bee colony(ABC)algorithm," *Journal of Global Optimization*, vol. 39, no. 3, pp. 459–471, 2007.
- [20] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *Journal of Machine Learning Research*, vol. 5, pp. 975–1005, 2004.
- [21] F. Schwenker, "Hierarchical support vector machines for multi-class pattern recognition," in *Proceedings of the 4th International Conference on Knowledge-Based Intelligent Engineering Systems & Allied Technologies*, pp. 561–565, Brighton, UK, September 2000.
- [22] T. Yingthawornsuk, "Classification of cardiac arrhythmia via SVM," in *Proceedings of the 2nd International Conference on Biomedical Engineering and Technology*, vol. 34, IPCBEE, 2012.
- [23] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, ACM Press, Addison-Wesley, Wokingham, UK, 1999.
- [24] J. Cheng, K. Zhang, and Y. Yang, "An order tracking technique for the gear fault diagnosis using local mean decomposition method," *Mechanism and Machine Theory*, vol. 55, pp. 67–76, 2012.
- [25] D. Karaboga, B. Akay, and C. Ozturk, "Artificial bee colony (ABC) optimization algorithm for training feed-forward neural networks," *Modeling Decisions for Artificial Intelligence*, vol. 4617, pp. 318–329, 2007.
- [26] C. M. Vong, P. K. Wong, and W. F. Ip, "A new framework of simultaneous-fault diagnosis using pairwise probabilistic multi-label classification for time-dependent patterns," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3372–3385, 2013.



Hindawi

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

