Research Article

Improving Rolling Bearing Fault Diagnosis by DS Evidence Theory Based Fusion Model

Xuemei Yao,¹ Shaobo Li,^{1,2} and Jianjun Hu^{2,3}

¹Key Laboratory of Advanced Manufacturing Technology, Ministry of Education, Guizhou University, Guiyang 550025, China

²School of Mechanical Engineering, Guizhou University, Guiyang 550025, China

³Department of Computer Science and Engineering, University of South Carolina, Columbia, SC 29208, USA

Correspondence should be addressed to Shaobo Li; lishaobo@gzu.edu.cn

Received 28 May 2017; Revised 30 August 2017; Accepted 17 September 2017; Published 22 October 2017

Academic Editor: Guiyun Tian

Copyright © 2017 Xuemei Yao 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.

Rolling bearing plays an important role in rotating machinery and its working condition directly affects the equipment efficiency. While dozens of methods have been proposed for real-time bearing fault diagnosis and monitoring, the fault classification accuracy of existing algorithms is still not satisfactory. This work presents a novel algorithm fusion model based on principal component analysis and Dempster-Shafer evidence theory for rolling bearing fault diagnosis. It combines the advantages of the learning vector quantization (LVQ) neural network model and the decision tree model. Experiments under three different spinning bearing speeds and two different crack sizes show that our fusion model has better performance and higher accuracy than either of the base classification models for rolling bearing fault diagnosis, which is achieved via synergic prediction from both types of models.

1. Introduction

Healthy operation of machinery systems is important in modern manufacturing enterprises, which leads to increasing attention to fault diagnosis technology that detects, identifies, and predicts abnormal states of manufacturing systems. Complex fault states and uncertain fault information bring high demand for real-time and intelligent fault diagnosis. Rolling bearing is a key component of rotating machinery, and any failure may cause equipment malfunction or catastrophic consequences. Almost every machine has at least one of these components, and their faults can be the direct cause of subsequent problems in other parts. Thus, bearing faults should be detected at the early stage [1]. Generally, rolling bearing consists of shaft, balls, inner race, outer race, cage, and housing. In principle, each component may fail. However, the inner race, outer race, and balls are the most vulnerable components due to friction and thus more prone to malfunction. Therefore, rolling bearing fault detection and diagnosis is of great significance to ensure production efficiency and equipment safety. The essence of fault diagnosis process is signal processing and pattern recognition. Signal processing functions to extract the features that characterize

the nature of the faults from complex original signals, whereas pattern recognition can classify the fault types and identify specific faults according to the input features, which can thus reduce reliance on technical personnel.

Thus far, several methods have been used for bearing fault diagnosis, and each has its case history of successes and failures. These methods can be classified according to their information source types, such as acoustic measurements, current and temperature monitoring, wear debris detection, and vibration analysis. Vibration analysis is broadly considered as the most effective monitoring technique in rotating machinery. Numerous vibration phenomena can be interpreted to be an amplitude modulation of the characteristic vibration frequency of a machine. Once the bearing fails, vibration pulses are produced. Pulse signals with smooth and less fluctuations are produced even when the bearing is normally operating. Recently, many fault diagnosis studies for rolling bearing based on vibration data have been reported. Sanz et al. [2] presented a method for detecting the states of rotating machinery with vibration analysis. Zhou and Cheng [3] proposed a fault diagnosis method based on image recognition for rolling bearing to realize fault classification under variable working conditions. Li et al. [4] presented

FIGURE 1: Structure of our fusion model. It mainly includes ① feature extraction based on wavelet transform; ② dimension reduction using PCA; ③ training of LVQ neural network and decision tree models; and ④ DS evidence theory based fusion prediction model.

a model for deep statistical feature learning from vibration measurements of rotating machinery, and the results showed that deep learning with statistical feature extraction has an essential improvement potential for diagnosing rotating machinery faults.

Initial high-dimensional features are obtained by decomposing the vibration signals with wavelet transform. However, redundant information of high-dimensional features may cause dimensionality problems to subsequent pattern analysis. Hence, principal component analysis (PCA) has been introduced to reduce dimensionality and eliminate redundant information aiming at improving the classification speed and accuracy. Taouali et al. [5] proposed a new method for fault detection using a reduced kernel PCA and obtained satisfactory results. Wodecki et al. [6] presented a multichannel processing method for local damage detection in gearbox using the combination of PCA and time frequency. Cho et al. [7] suggested a fault identification method that is especially applicable to process monitoring using PCA and achieved a high efficiency. Nguyen and Golival [8] addressed the fault detection problem in mechanical systems using the PCA method, which effectively improved the detection results. Their findings indicate that the optimal features extracted with PCA are considered accurate and efficient.

As a simple and efficient classifier, the decision tree can be used to infer the classification rules from a set of training samples, which has been extensively used in fault diagnosis. Karabadi et al. [9] discussed a new approach for fault diagnosis in rotating machines based on improved decision trees and gained ideal results. Rutkowski et al. [10] presented a new algorithm based on decision trees to determine the best attributes and obtained a high accuracy of classification with a short processing time. Amarnath et al. [11] selected the descriptive features extracted from acoustic signals using a decision tree algorithm to realize the fault bearing diagnosis. Krishnakumari et al. [12] selected the best features through a decision tree to train a fuzzy classifier for fault diagnosis.

In recent years, artificial neural networks (ANNs) are widely used in fault diagnosis because of their capability of learning highly nonlinear relationships. The back propagation (BP) neural network is extensively used, but it can easily fall into local optimum. The learning vector quantization (LVQ) neural network is a learning algorithm that trains hidden layers under the supervision state, which can overcome the shortcomings of the BP network and achieve better prediction performance. Rafiee et al. [13] presented an ANN for fault detection and identification in gearbox using features that are extracted from vibration signals. Umer and

Khiyal [14] evaluated the LVQ network for classifying text documents, and the results showed that the LVQ required less training of samples and exceeded other classification methods. Melin et al. [15] described the application of competitive neural networks using the LVQ algorithm to classify electrocardiogram signals and produced desired results.

Another strategy to improve prediction performance is to use information fusion approach. For example, the Dempster-Shafer (DS) evidence theory has been adopted to handle information fusion. Kushwah et al. [16] proposed a multisensor fusion methodology using evidence theory for indoor activity recognition, which gained ideal identification accuracy. Basir and Yuan [17] investigated the use of DS theory as a tool for modelling and fusing multisensor pieces of evidence pertinent to engine quality. Bhalla et al. [18] used DS evidence theory to integrate the results of BP neural network and fuzzy logic to overcome the conflicts of fault diagnosis.

Previous methods for fault diagnosis of rolling bearing suffer from either a single source of information or single type of models, which leads to biased prediction. To address these issues, we propose an information fusion model for bearing fault diagnosis by combining the LVQ neural network and the decision tree classifier, of which the predictions are fused using the DS evidence theory. Our algorithm is inspired by that fact that ensemble machine learning algorithms based on fusing predictions from multiple base machine learning models have been shown to be able to achieve most competitive performance [19, 20] and DS evidence theory based fusion method has been successfully applied to fault diagnosis of hydraulic pump [21], rolling bearing [22, 23], and complex electromechanical systems [24]. However, previous work of DS fusion for bearing fault diagnosis has not studied whether DS based evidence fusion can be used to combine heterogeneous base models into more accurate prediction models. The inputs of our model are statistical characteristics obtained by decomposing the vibration signals using wavelet transform. We use the PCA technique to reduce the features dimensions according to the cumulative contribution rate of the eigenvalues. We also use the LVQ neural network and decision tree to perform the initial fault prediction. Then, we calculate the basic probability assignment (BPA) of the two models through normalized operations. Finally, we fuse the results of the two methods and adopt the DS evidence theory to identify the fault type of the rolling bearing. The structure of our fusion model is shown in Figure 1.

The rest of this paper is organized as follows. Section 2 describes the experimental setup and design. Section 3

FIGURE 2: Schematic of the setup.

presents the methodology used in this study, including feature extraction, decision tree, LVQ neural network, PCA technique, and DS evidence theory. Section 4 presents the experimental results, together with some discussions. Finally, Section 5 concludes the paper.

2. Experimental Setup and Design

2.1. Test Rig. In the present study, the bearing fault diagnosis problem is to confirm whether a rolling bearing is in good or faulty condition. The rolling bearing state needs to be categorized into the three states: good, inner race fault, and outer race fault. The test rig is equipped with fixing and clamping devices for fixing the bearing outer race while the inner race rotates with the shaft. The test rig consists of an induction motor, bearing, piezoelectric accelerometer, signal processing unit, and computer.

In this study, five KBC 6203 rolling bearings are derived by the induction motor (2 HP). One is a new bearing without any defects to simulate the good condition. The other four bearings are designed to simulate the inner and outer race faults. Two bearings are used to simulate two types of cracks in each fault. The defect is created by the spark erosion technique to control the defect. The piezoelectric type accelerometer IMI Sensors 608A111 measures the vibration signals and is mounted on the base of the rolling bearing using adhesive. The output of the accelerometer is sent to the signal processing unit, where the signal goes through the charge amplifier and analogue-to-digital (A/D) converter. Vibration signals are amplified by a DACELL DNAM100 amplifier and are transformed from analogue to digital signals. The signals are then transmitted to the computer memory through the USB port. Subsequently, the signals are read and processed from the memory to extract different statistical features according to the requirements. The schematic of the setup is shown in Figure 2.

2.2. Experiment Design. Some rubber sheets are added in the legs of the test rig to avoid environmental noise and vibration, thereby obtaining more realistic data. The signal processing unit is switched on, and the first few signals are discarded purposefully to avoid initial random variation. The vibration signals with undamped natural frequency of 10 KHz are accepted after the signal becomes stable.

The vibration signals are gained from the accelerometer at the base of the rolling bearing. The sampling frequency is 24 kHz, the length of the record is 10 s, and the sample length is 1024 for all experiment cases. The highest frequency is found to be 12 kHz via experimenting. The sampling

frequency must be twice that of the highest measured frequency, according to the Nyquist sampling theorem; thus, it is set at 24 kHz. The choice of the sample length is considered arbitrary to a certain extent. However, the statistical measurement is more meaningful when the number of samples is large; meanwhile, the computing time increases with the number of samples. Generally, a sample length of approximately 1000 is selected to achieve balance. In some feature extraction techniques, the sample length is always 2^n , and the nearest 2^n to 1000 is 1024. Therefore, 1024 is selected as the sample length under normal circumstances [25].

Ninety experiments are conducted by varying the parameters under three different spinning speeds of the bearing (500, 700, and 900 rpm). First, a rolling bearing without any defects is used for the good case. Ten samples are collected under each spinning speed of the bearing. Then, 30 different cases are obtained by changing the shaft speed. Second, the outer race fault condition is conducted in the test rig. The crack of the outer race fault is created via spark erosion technique. The crack size is 0.5 mm wide and 0.7 mm deep, and the other crack is 0.3 mm wide and 0.6 mm deep. The performance characteristics of the outer race fault of the bearing are studied, as explained for the good case. Vibration signals with outer race fault are recorded in memory, keeping all other modules in good condition. A crack corresponds to three different spinning speeds of the bearing. Fifteen samples are obtained for each speed with five samples. Thirty different cases are obtained by changing the shaft speed and crack size. Third, the inner race fault is simulated by the same crack size as that of the outer race. Thirty samples are also obtained. The cumulative 90 samples, of which the dimensions are reduced by PCA, are used as input for the LVQ neural network. Similarly, 90 samples are collected again as the input for the decision tree.

3. Methodology

3.1. Feature Extraction. Generally, statistical features are good indicators of the state of machine operation. Vibration signals are obtained from different spinning speeds and fault types, and the required statistical characteristics can be extracted using time or frequency domain analysis. The most commonly used statistical feature extraction methods are fast Fourier transform (FFT) and wavelet transform. FFT converts time domain signal into frequency domain signal and is widely used in signal detection. However, the FFT method is inherently flawed in handling unstable processes. This method only acquires the frequency components of the signal as a whole but is unaware of the moment at which

the components appear. The difference between two time domain signals is large, but the spectrum may be the same; thus, FFT cannot render good performance [26]. In comparison with FFT, wavelet transform is a local transformation of time and frequency [27]. Wavelet transform has good spatial and frequency domain localization characteristics. It can effectively extract information from the signal through the expansion, translation, and other computing functions. Wavelet transform is widely applied in multiscale refinement analysis of functions or signals. It can focus on the detail of the analysis object by using fine time domain or space step in high frequency, which solves many problems that FFT cannot [28].

In the present study, wavelet transform is used to collect time domain features of the vibration signals, which are gained from the accelerometer. Wavelet coefficients cannot be directly used as input of the diagnostic model; thus, a feature extraction preprocessing step is required to prepare the data for the model. A large number of features can be extracted from each signal, which can be divided into two categories: features with dimensions and dimensionless features. The features with dimensions such as variance, mean, and peak are more likely to be affected by working conditions. Dimensionless features such as kurtosis, crest, and pulse are less sensitive to external factors. Different features reflect different aspects of the fault information of the bearing. Effective feature extraction and selection and preprocessing are critical for successful classification [29]. The increase in the number of features will inevitably lead to redundancy and curse of dimensionality while ensuring comprehensive and complete access to the fault information. To achieve a balanced control, only 10 statistical characteristics with good sensitivity and differentiation to the fault type are selected as inputs to the model in this work as shown below.

(1) *Variance*. It is the measurement of signal dispersion degree. A larger variance indicates a greater fluctuation of data, whereas a smaller variance indicates a smaller fluctuation of data. The following formula is used to compute variance.

$$\text{Variance} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}. \quad (1)$$

(2) *Kurtosis*. It indicates the flatness or the spikiness of the signal. It is considerably low under normal condition but increases rapidly with the occurrence of faults. It is particularly effective in detecting faults in the signal. The following formula is used to solve for kurtosis:

$$\text{Kurtosis} = \frac{(1/n) \sum_{i=1}^n (x_i - \bar{x})^4}{((1/n) \sum_{i=1}^n (x_i - \bar{x})^2)^2}. \quad (2)$$

(3) *Mean*. It represents the central tendency of the amplitude variations of the waveform and can be used to describe signal stability, which is the static component of the signal. The following formula is used to obtain the mean:

$$\text{Mean} = \frac{\sum_{i=1}^n x_i}{n}. \quad (3)$$

(4) *Standard Deviation (Std)*. It is the measurement of the effective energy of the signal and reflects the discrepancy degree between individuals within the group. The following formula is used for its computation:

$$\text{Std} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}. \quad (4)$$

(5) *Skewness*. It is the measurement of the skew direction and extent of the data distribution and is a numerical feature of the degree of asymmetry of the statistical data. The following formula is used to compute skewness:

$$\text{Skewness} = \frac{(1/n) \sum_{i=1}^n (x_i - \bar{x})^3}{((1/n) \sum_{i=1}^n (x_i - \bar{x})^2)^{3/2}}. \quad (5)$$

(6) *Peak*. It refers to the instantaneous maximum value of the fault signal in a given time. The following formula is used to compute peak:

$$\text{Peak} = \max |x_i|. \quad (6)$$

(7) *Median*. It refers to the value of a variable in the middle of the array, which is sorted from small to large with all variables. The following formula is used to determine the median:

$$\text{Median} = \begin{cases} x_{(n+1)/2}, & n \text{ is odd,} \\ \frac{x_{n/2} + x_{(n/2)+1}}{2}, & n \text{ is even.} \end{cases} \quad (7)$$

(8) *Root Mean Square (RMS)*. It is an important index in determining whether the running state is normal in the mechanical fault diagnosis system. Moreover, it reflects the magnitude of the signal energy. The following formula is used to compute for RMS:

$$\text{RMS} = \sqrt{\frac{\sum_{i=1}^n (x_i)^2}{n}}. \quad (8)$$

(9) *Crest Factor (CF)*. It is the measurement of a waveform showing the ratio of peak values to the effective value. In other words, crest factor indicates the extremeness of peaks in a waveform. The following formula is used to compute it:

$$\text{CF} = \frac{x_{\text{peak}}}{x_{\text{RMS}}}. \quad (9)$$

(10) *K Factor*. It reflects the shock characteristics of vibration signals and is sensitive to abnormal pulses that are produced by bearing faults. Its normal value is 3. If it is close to or more than 4, shock vibration exists. The following formula is used to determine K factor:

$$K \text{ factor} = \frac{x_{\text{kurtosis}}}{x_{\text{RMS}}^4}. \quad (10)$$

The statistical feature matrix of some samples is shown in Table 1.

TABLE 1: Feature matrix of some samples.

	Class	Variance	Kurtosis	Mean	Std	Skewness	Peak	Median	RMS	CF	K factor
Train set	1	0.0214	16.1539	0.0011	0.0870	2.0854	0.5870	-0.0307	0.1377	8.3919	0.0873
	1	0.1115	12.4934	0.0011	0.0984	2.0647	0.6643	-0.0130	0.1063	8.7519	0.1421
	1	0.0075	12.2625	0.0011	0.1346	2.0755	0.8143	-0.0127	0.1148	8.2310	0.2199
	2	0.0340	15.0917	0.0012	0.1843	2.6764	1.2198	-0.0255	0.1843	6.6173	0.2248
	2	0.0098	10.7345	0.0009	0.0988	1.6925	0.6694	-0.0104	0.0998	6.7763	0.0661
	2	0.1396	20.1365	0.0010	0.3736	3.4306	2.6430	-0.0684	0.3736	7.0734	0.9875
	3	0.0886	28.7487	0.0007	0.3918	4.1160	2.5614	-0.0290	0.2149	6.3282	0.9030
	3	0.1843	26.8139	0.0011	0.5596	4.0948	3.2983	-0.0845	0.3269	4.9603	1.6723
	3	0.2343	33.8579	0.0010	0.3422	3.4560	2.4916	-0.0324	0.1991	3.7395	1.0920
Test set	1	0.0274	12.0830	0.0010	0.0863	2.0656	0.9590	-0.0218	0.1096	8.6386	0.1770
	1	0.0559	13.7078	0.0008	0.0662	2.4363	0.7831	-0.0092	0.0959	6.7902	0.0352
	1	0.0088	13.0394	0.0010	0.1602	2.1901	1.2309	-0.0227	0.1201	6.8103	0.0270
	2	0.0934	17.2709	0.0012	0.1513	0.5766	0.6190	-0.0632	0.2841	9.1493	0.9153
	2	0.0059	24.5557	0.0011	0.1197	0.8212	2.0096	-0.0707	0.1739	9.8313	0.7793
	2	0.1178	18.0737	0.0011	0.3140	1.0114	2.2261	-0.0400	0.1789	10.6499	0.5463
	3	0.1387	34.2424	0.0009	0.3473	4.3834	3.3007	-0.0508	0.4638	9.3689	1.4088
	3	0.3126	28.7760	0.0007	0.3842	4.1692	1.8429	-0.0832	0.3311	10.6652	0.5623
	3	0.2959	29.1322	0.0010	0.4585	3.3189	2.3107	-0.0182	0.1846	4.4879	1.2974

3.2. Dimensionality Reduction. PCA is a statistical method and is widely used in data reduction. By means of orthogonal transformation, a group of variables that may be related to one another are transformed into a set of linearly uncorrelated variables, which is called the principal components. It functions to maintain the primary information of original features while reducing the complexity of the data, which reveals the simple structure behind the complex data. PCA is a simple and nonparametric method of extracting relevant information from intricate data.

The purpose of PCA is to reduce the dimensionality of data while preserving as many changes as possible in the original dataset. PCA transforms the data into the coordinate system; thus, the maximum variance of any projection of data lies on the first coordinate, and the second largest variance falls on the second coordinate, and so on. PCA algorithm can remove redundant information, simplify the difficulty of dealing with the problem, and improve the resistance to external interference through the processing of raw data. Therefore, PCA is used in this paper. The specific steps of PCA algorithm are shown as follows.

Step 1. Input sample matrix $D = \{x_1, x_2, \dots, x_n\}^T$. The rows of the matrix represent the number of samples, whereas the columns represent the dimensions.

Input percentage of information retention after dimension reduction (e).

Step 2. Calculate the mean by columns.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}. \quad (11)$$

Step 3. Obtain the new sample matrix M with data centralization.

$$\begin{aligned} \theta_i &= x_i - \bar{x}, \\ A &= [\theta_1, \theta_2, \dots, \theta_n], \\ M &= AA^T. \end{aligned} \quad (12)$$

Step 4. Calculate the eigenvalues and eigenvectors.

$$\begin{aligned} MU &= \lambda U \longrightarrow \\ \lambda_1 &> \lambda_2 > \dots > \lambda_n, \\ AU &= \lambda U \longrightarrow \\ U &= \{u_1, u_2, \dots, u_n\}. \end{aligned} \quad (13)$$

Step 5. Determine the final dimension k .

$$\begin{aligned} \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n \lambda_i} &\geq e \longrightarrow \\ \lambda_1, \lambda_2, \dots, \lambda_k. \end{aligned} \quad (14)$$

The cumulative contribution rate of the eigenvalues (e) is used to measure the representation of the newly generated principal components to the original data. Generally, e should be greater than or equal to 85% to extract the former k principal components as the sample features.

Step 6. Output the principal components.

$$\begin{aligned} U_k &= (u_1, u_2, \dots, u_k), \\ P &= x * U_k. \end{aligned} \quad (15)$$

Notably, the dataset is divided into training and testing sets prior to importing the model in this study. Therefore, both sets must be processed separately when PCA is used. Subtracting the mean value of the training sample and using the transformation matrix obtained from the training sample are important when the dimension of the testing sample is reduced to ensure that the training and testing samples are switched to the same sample space. In this study, the first four principal components are selected, and some of them are shown in Table 2.

3.3. Decision Tree. Decision tree is a common supervised classification method in data mining. In supervised learning, a bunch of samples is provided. Each sample has a set of attributes and a category label. These categories are determined in advance, and a classifier is then created by a learning algorithm. The topmost node of the decision tree is the root node. Decision tree classifies a sample from the root to the leaf node. Each nonleaf node represents a test of the attribute value. Each branch of the tree represents the results of the test. Each leaf node represents a category of the object. In short, the decision tree is a tree structure similar to a flow diagram.

A decision tree is built recursively following a top-down approach. It compares and tests the attribute values on its internal nodes from the root node, then determines the corresponding branch according to the attribute value of the given instance, and finally draws the conclusion in its leaf node. The process is repeated on the subtree, which is rooted with a new node. Thus far, many decision tree algorithms exist, but the most commonly used one is the C4.5 algorithm. The pseudocode of the C4.5 algorithm is shown in Pseudocode 1.

A leafy decision tree may be created due to the noises and outliers of the training data, which will result in overfitting. Many branches reflect anomalies of the data. The solution is pruning and cutting off the most unreliable branches, in which pre- and postpruning are widely used. The C4.5 algorithm adopts a pessimistic postpruning. If the error rate can be reduced through replacing a subtree with its leaf node, the subtree will be pruned.

3.4. LVQ Neural Network. LVQ is an input feed-forward neural network with supervised learning for training the hidden layers. The LVQ neural network consists of input, hidden, and output layers. The hidden layer automatically learns and classifies the input vectors. The results of classification depend only on the distance between the input vectors. If the distance between the two input vectors is particularly similar, then the hidden layer divides them into the same class and output them.

The network is completely connected between the input and hidden layers; meanwhile, the hidden layer is partially connected with the output layer. Each output layer neuron is connected to different groups of hidden layer neurons. The

TABLE 2: Principal components of some samples.

	Class	PCA1	PCA2	PCA3	PCA4
Train set	1	-1.9913	-0.2596	-1.0560	-0.2412
	1	-2.1547	-1.9631	1.9927	0.8611
	1	-3.1183	-0.0118	1.1916	0.1564
	2	-2.1922	0.6848	-0.5967	0.3052
	2	-0.8282	1.2599	-0.6597	0.6612
	2	-1.3281	-0.5063	-0.8747	0.4467
	3	3.7605	1.6074	0.2554	-0.5114
	3	7.2074	-1.3968	1.1143	-0.5616
	3	4.2621	1.8465	0.2609	0.0197
Test set	1	-1.5410	-1.0017	0.3682	0.2004
	1	-1.2102	-2.1453	0.7366	0.5607
	1	-1.6120	0.4242	0.4177	-0.5693
	2	2.2970	-1.4313	-0.9372	-2.6193
	2	-3.5065	-1.3869	-1.7524	-0.9630
	2	-0.3051	-0.8110	-0.3567	-0.7103
	3	2.8365	-0.4960	-1.1518	1.8498
	3	3.9953	-0.4302	0.3341	0.7450
	3	2.9665	-0.9305	-0.6439	0.3039

Input: an attribute set dataset D

Output: a decision tree

- (a) Tree = { }
- (b) if D is “pure” or other end conditions are met, then
- (c) terminate
- (d) end if
- (e) for each attribute $a \in D$ do
- (f) compute information gain ratio (InGR)
- (g) end for
- (h) a_{best} = attribute with the highest InGR
- (i) Tree = create a tree with only one node a_{best} in the root
- (j) D_v = generate a subset from D except a_{best}
- (k) for all D_v do
- (l) subtree = C4.5 (D_v)
- (m) set the subtree to the corresponding branch of the Tree according to the InGR
- (n) end for

PSEUDOCODE 1: Pseudocode of the C4.5 algorithm.

number of neurons in the hidden layer is always greater than that of the output layer. Each hidden layer neuron is connected with only one output layer neuron, and the connection weight is always 1. However, each output layer neuron can be connected to multiple hidden layer neurons. The values of the hidden and output layer neurons can only be 1 or 0. The weights between the input and hidden layers are gradually adjusted to the clustering center during the training. When a sample is placed into the LVQ neural network, the neurons of the hidden layer generate the winning neuron by the winner-learning rules, allowing the output to be 1 and the other to be

0. The output of the output layer neurons that are connected to the winning neuron is 1, whereas the other is 0; then, the output provides the pattern class of the current input sample. The class learned by the hidden layer becomes a subclass, and the class learned by the output layer becomes the target class [30]. The architecture of the LVQ neural network is shown in Figure 3.

The training steps of the LVQ algorithm are as follows.

Step 1. The learning rate η ($\eta > 0$) and the weights W_{ij} between the input and hidden layers are initialized.

FIGURE 3: Architecture of the LVQ neural network.

Step 2. The input vector $x = (x_1, x_2, \dots, x_n)^T$ is fed to the input layer, and the distance (d_i) between the hidden layer neuron and the input vector is calculated.

$$d_i = \sqrt{\sum_{j=1}^n (x_j - w_{ij})^2}. \quad (16)$$

Step 3. Select the hidden layer neuron with the smallest distance from the input vector. If d_i is the minimum, then the output layer neuron connected to it is labeled with c_i .

Step 4. The input vector is labeled with c_x . If $c_i = c_x$, the weights are adjusted as follows:

$$w_{ij\text{-new}} = w_{ij\text{-old}} + \eta (x - w_{ij\text{-old}}). \quad (17)$$

Otherwise, the weights are updated as follows:

$$w_{ij\text{-new}} = w_{ij\text{-old}} - \eta (x - w_{ij\text{-old}}). \quad (18)$$

Step 5. Determine whether the maximum number of iterations is satisfied. The algorithm ends if it is satisfied. Otherwise, return to Step 2 and continue the next round of learning.

3.5. Evidence Theory. Data fusion is a method of obtaining the best decision from different sources of data. In recent years, it has attracted significant attention for its wide application in fault diagnosis. Generally, data fusion can be conducted at three levels. The first is data-level fusion [31]. Raw data from different sources are directly fused to produce more information than the original data. At this level, the fusion exhibits small loss and high precision but is time-consuming and unstable and has weak anti-interference capabilities. The second is feature-level fusion [32]. At this level, statistical features are extracted separately using signal processing technique. All features are fused to find an optimal subset, which is then fed to a classifier for better accuracy. At this level, information compression for transmission is achieved but with poor integration accuracy. The third is decision-level fusion [33]. This fusion is the highest level

FIGURE 4: Recognition framework of the DS theory.

of integration, which influences decision making, and it is the ultimate result of three-level integration. The decision-level fusion exhibits strong anti-interference ability and small amount of communication but suffers from large amount of data loss and high cost of pretreatment. In this paper, we focus on decision-level fusion, which is known as the DS evidence theory.

The DS evidence theory has been originally established in 1967 by Dempster and developed later in 1976 by Shafer, who is a student of Dempster. Evidence theory is an extension of Bayesian method. In the Bayesian method, the probability must satisfy the additivity, which is not the case for the evidence theory. The DS evidence theory can express uncertainty, leaving the rest of the trust to the recognition framework. This theory involves the following mathematical definitions.

Definition 1 (recognition framework). Define $\Omega = \{\theta_1, \theta_2, \dots, \theta_n\}$ as a set, where Ω is a finite set of possible values and θ_i is the conclusion of the model. The set is called the recognition framework. 2^Ω is a power set composed of all the subsets. The recognition framework with capacity of four layers and the relationship between the subsets is shown in Figure 4. a, b, c , and d are the elements of the framework.

Definition 2 (BPA). BPA is a primitive function in DS evidence theory. Assume Ω as the recognition framework; then, m is a mapping from 2^Ω to $[0, 1]$, and A is the subset of Ω . m is called the BPA when it meets the following equation:

$$\begin{aligned} m(\emptyset) &= 0, \\ \sum_{A \subseteq \Omega} m(A) &= 1. \end{aligned} \quad (19)$$

Definition 3 (combination rules). For $\forall A \subseteq \Omega$, a finite number of m functions ($m_1, m_2, \dots, m_n$) exist on the recognition framework. The combination rules are as follows:

$$\begin{aligned} m(\emptyset) &= 0, \\ m(A) &= \frac{1}{k} \sum_{A_1 \cap A_2 \cap \dots \cap A_n = A} m_1(A_1) * m_2(A_2) \cdots m_n(A_n), \end{aligned} \quad (20)$$

FIGURE 5: Pruning of the decision tree.

where $k = \sum_{A_1 \cap A_2 \cap \dots \cap A_n \neq \emptyset} m_1(A_1) * m_2(A_2) \cdots m_n(A_n)$ or $k = 1 - \sum_{A_1 \cap A_2 \cap \dots \cap A_n = \emptyset} m_1(A_1) * m_2(A_2) \cdots m_n(A_n)$, which reflects the conflict degree between evidences.

4. Results and Discussions

The experiments are conducted to predict good and outer and inner race fault conditions of the rolling bearing, as discussed in Section 2. The diagnosis model in this article should undergo three steps, whether it is a neural network or decision tree. First, the relevant model is created with the training set. Then, the testing set is imported to simulate results. Finally, simulation and actual results are compared to obtain the fault diagnosis accuracy. Hence, each group of experimental data, which are extracted from the vibration signals, is separated into two parts. Sixty samples are randomly selected for training and the remaining 30 samples are used for testing.

4.1. Results of the Tree-PCA. Sixty samples in different cases of fault severity have been fed into the C4.5 algorithm. The algorithm creates a leafy decision tree, and the sample classification accuracy is usually high with the training set. However, the leafy decision tree is often overfitted or overtrained; thus, such a decision tree does not guarantee an approximate classification accuracy for the independent testing set, which may be lower. Therefore, pruning is required to obtain a decision tree with relatively simple structure (i.e., less bifurcation and fewer leaf nodes). Pruning the decision tree reduces the classification accuracy of the training set but improves that of the testing set. The re-substitution and cross-validation errors are a good evidence of the change. Sixty samples, including 10 statistical features extracted from the vibration signals, are used as input of the algorithm, and the output is the pruning of the decision tree, as shown in Figure 5.

Figure 5 shows that the decision tree has leaf nodes, which stand for class labels (namely, 1 as good, 2 as outer race fault, and 3 as inner race fault), and decision nodes, which stand for the capability of discriminating (namely, x_5 as skewness, x_2 as kurtosis, x_1 as variance, and x_8 as RMS).

TABLE 3: Error values before and after pruning.

	Before pruning	After pruning
re-sub-err	0.01	0.07
cross-val-err	0.09	0.08
Average accuracy	82.98%	84.09%

Not every statistical feature can be a decision node, which depends on the contribution of the entropy and information gain. Attributes that meet certain thresholds appear in the decision tree; otherwise, they are discarded intentionally. The contribution of 10 features is not the same, and the importance is not consistent. Only four features appear in the tree. The importance of the decision nodes decreases from top to bottom. The top node is the best node for classification. The most dominant features suggested by Figure 5 are kurtosis, RMS, mean, and variance.

Re-substitution error refers to the difference between the actual and predicted classification accuracy, which is obtained by importing the training set into the model again after creating the decision tree using the training set. The cross-validation error is an error value of prediction model in practical application by cross-validation. Both are used to evaluate the generalization capability of the prediction model. In this study, the re-substitution error is expressed by “re-sub-err,” the cross-validation error is expressed by “cross-val-err,” and the average classification accuracy rate is expressed by “average accuracy.” In the experiment, we can obtain the results shown in Table 3.

Table 3 shows that the cross-val-err is approximately equal ($0.08 \approx 0.09$) and the re-sub-err after pruning is greater than before ($0.07 > 0.01$), but the average accuracy of the fault of the testing set after pruning significantly improves ($84.09\% > 82.98\%$).

At the same time, the PCA technique is used to reduce the dimension of statistical features. The first four principal components are extracted to create the decision tree according to the principle that the cumulative contribution rate of eigenvalues is more than 85%. Thus far, the dimension of the

FIGURE 6: Decision tree after dimension reduction.

TABLE 4: Classification errors before and after dimensionality reduction.

	Before PCA	After PCA
re-sub-err	0.07	0.03
cross-val-err	0.08	0.08
Average accuracy	84.09%	86.56%
Time (s)	2.02	1.33

statistical feature is reduced from 10 to 4, and the amount of data is significantly reduced. The decision tree is constructed with the first four main components, as shown in Figure 6.

Figure 6 shows that the testing set can be classified depending on the first (x_1) and second (x_2) principal components. The remaining two principal components do not appear in the decision tree because their contribution value does not reach the thresholds. When comparing Figure 5 with Figure 6, the decision tree after dimension reduction is simpler and has fewer decision nodes than before. Furthermore, the cross-val-err is equal, and the average accuracy is not low. Table 4 shows the experimental results.

Table 4 shows that the cross-val-err is equal ($0.08 = 0.08$), the re-sub-err of the decision tree after dimension reduction is lower ($0.03 < 0.07$), and the average accuracy is slightly higher ($86.56\% > 84.09\%$); however, the running time of the program is considerably lower ($1.33 \text{ seconds} < 2.02 \text{ seconds}$). Therefore, dimension reduction is necessary and effective for constructing the decision tree, especially with many statistical attribute values.

4.2. Results of the LVQ-PCA. The LVQ neural network belongs to the feed-forward supervised neural network; it is one of the most widely used methods in fault diagnosis. Thus, LVQ neural network is used in this study to distinguish the different fault states of the rolling bearing. The training samples are imported into the LVQ neural network. The input layer contains 10 statistical characteristics, which are extracted from the vibration signals. The output layer is the classification accuracy of the fault, including three types of fault, namely, good, outer race fault, and inner race fault.

FIGURE 7: LVQ neural network training error convergence diagram.

Meanwhile, the design of the hidden layer is important in the LVQ neural network; thus, it is defined by K cross-validation method. The initial sample is divided into K sub-samples. A subsample is retained as the validation data, and the other $K - 1$ subsamples are used for training. The cross-validation process is repeated K times. Each subsample is validated once, and a single estimate is gained by considering the average value of K times. This method can avoid overlearning or underlearning state, and the results are more convincing. In this study, the optimal number of neurons in the hidden layer is 11 through 10-fold cross-validation, which is most commonly used.

Sixty samples are used for training, and the remaining 30 samples are used for testing in the LVQ neural network. The network structure of 10-11-3 is used in the experiment. The network parameters are set as follows:

Maximum number of training steps: 1000

Minimum target training error: 0.1

Learning rate: 0.1.

The LVQ neural network is created to obtain the error curve, as shown in Figure 7. To highlight the superiority of LVQ, a BP neural network is also created using the same parameter settings, and the error curve is shown in Figure 8.

We used the mean squared error (MSE) as the evaluation measure, which calculates the average squared difference between outputs and targets. A lower value is better, and zero means no error. When comparing Figure 7 with Figure 8, the BP neural network has less training time and smaller MSE. Evidently, the BP neural network is superior to LVQ in this case. However, the BP neural network algorithm is essentially a gradient descent method. It is a way of local optimization and can easily fall into the local optimal solution, which is demonstrated by the results in Table 5. The

TABLE 5: Comparisons of classification performances of different models.

Bearing condition	Classification accuracy						
	LVQ	BP	LVQ-PCA	Decision tree	Tree pruning	Tree-PCA	DS fusion
Good, F_1 (%)	70.0	9.1	87.5	85.5	86.2	82.4	90.4
Outer race fault, F_2 (%)	66.7	11.1	78.9	72.7	74.0	97.9	96.6
Inner race fault, F_3 (%)	100.0	90.0	100.0	90.7	92.1	79.4	94.9
Average accuracy (%)	78.9	36.7	88.8	83.0	84.1	86.6	94.0
Time (s)	2.3	1.6	1.5	2.6	2.0	1.3	4.2

FIGURE 8: BP neural network training error convergence diagram.

maximum classification accuracy of BP neural network is 90.0% and the minimum is 9.1% under the same data and network parameters. The gap is significantly large, which leads to the low average accuracy of approximately 36.7%. The classification accuracies of the LVQ neural network are not much different from each other, and the average accuracy is 78.9%. This phenomenon indicates that the performance of the LVQ neural network is better than that of BP neural networks.

Likewise, better performance can be achieved by combining PCA and LVQ. The original 10 feature attributes are replaced by the four principal components, and the other network parameters are unchanged. Figures 9 and 10 are obtained from the experiment.

Figure 9 shows the ROC curve, which is a plot of the true positive rate (sensitivity) versus the false positive rate ($1 - \text{specificity}$) as the thresholds vary. A perfect test would show points in the upper-left corner with 100% sensitivity and 100% specificity. Figure 10 shows the training regression plot. Regression value (R) measures the correlation between outputs and targets; 1 means a close relationship, and 0 means a random relationship. All ROC curves are in the upper-left corner, and the value of R is 0.96982, which is approximately

FIGURE 9: Receiver operating characteristic (ROC) plot.

equal to 1. Therefore, the network performs well. The classification accuracy of LVQ-PCA further illustrated it as shown in Table 5.

4.3. Results of the Fusion Model. The decision tree and the LVQ neural network are widely used in fault diagnosis due to their simplicity and good generalization performance. However, their classification accuracy is still dependent on the datasets and may get unsatisfactory performance. To solve this problem, the DS evidence theory is introduced in this study. The target recognition frame U is established by considering the three states of the bearing: good (F_1), outer race fault (F_2), and inner race fault (F_3). Each fault sample belongs to only one of the three failure modes and is independent. The outputs of the LVQ neural network are used as evidence 1 and those of decision tree are used as evidence 2; then data fusion is performed based on the aforementioned DS method. The experiment has been run 20 times to reduce the occurrence of extreme values and to obtain reliable results. The classification accuracies on

FIGURE 10: Regression plot.

FIGURE 11: Boxplot of DS fusion results.

the training and testing sets for each run are recorded and the final performance comparisons are plotted as a boxplot (Figure 11).

Figure 11 shows that the accuracy on the training sets of three types of faults fluctuates slightly around 98%. The accuracy on F_1 -train is low and has an outlier. The accuracy on F_2 -train is on the higher side, with a maximum of up

to 100%. The accuracy on F_3 -train is between those of F_1 -train and F_2 -train with small variation. The small variations of the accuracy on the training sets indicate that the prediction models are stable. The accuracy on the testing sets is relatively scattered. The accuracy on F_1 -test is concentrated on 90% and has an exception of up to 97%. The accuracy on F_2 -test is concentrated on 97% while the accuracy on F_3 -test is near 94%. The average value of the 20 experimental results is considered as the final result of data fusion to reduce the error, as shown in Table 5.

Table 5 presents the results of all the seven algorithms used in the present study. Each of the algorithms has been run 20 times with different partition of the training and test datasets, and the average prediction accuracy for three fault types is recorded. First, it was found that the BP neural network falls into local optimal solutions as its average accuracy is only 36.7% (see second column of Table 5). Therefore, we conclude it as a failing method in our experiments. The average accuracy of the LVQ neural network has increased from 78.9% to 88.8% after applying dimension reduction. The performance of the decision tree improves slightly using pruning (from 83.0% to 84.1%) but increases to 86.6% through combining PCA and decision tree. The results indicate that dimensionality reduction is an effective means to improve prediction performance for both base classification models. The DS fusion model proposed in this study achieved an average accuracy of 94.0% by fusing predictions of LVQ-PCA and Tree-PCA, which is the best compared to all other 6 base methods. This demonstrates the capability of DS fusion to take advantage of complementary prediction performance of LVQ and decision tree classifiers. This can be clearly seen from the second row and the third row, which show the performance of the algorithms on predicting outer race fault and inner race fault. In the prior case, the Tree-PCA achieves a higher performance (97.9%) compared to LVQ-PCA (78.9%) while, in the latter case, LVQ-PCA achieved an accuracy of 100.0% compared to 79.4% of Tree-PCA. Through the DS fusion algorithm, the prediction performances are 96.6% and 94.9%, respectively, which avoids the weakness of either of the base models in predicting some specific fault types.

5. Conclusion

We have developed a DS evidence theory based algorithm fusion model for diagnosing the fault states of rolling bearings. It combines the advantages of the LVQ neural network and decision tree. This model uses vibration signals collected by the accelerometer to identify bearing failures. Ten statistical features are extracted from the vibration signals as the input of the model for training and testing. To improve the classification accuracy and reduce the input redundancy, the PCA technique is used to reduce 10 statistical features to 4 principal components.

We compared different methods in terms of their fault classification performance using the same dataset. Experimental results show that PCA can improve the classification accuracy of LVQ neural network in most cases but not always for the decision tree. Both LVQ neural network and decision

tree do not achieve good performance for some classes. The proposed DS evidence theory based fusion model fully utilizes the advantages of the LVQ neural network, decision tree, PCA, and evidence theory and obtains the best accuracy compared with other signal models. Our results show that the DS evidence theory can be used not only for information fusion, but also for model fusion in fault diagnosis.

The accuracy of the prediction models is important in bearing fault diagnosis while the convergence speed and the running time of the algorithms also need special attention especially in the case of large number of samples. The results in Table 5 show that the fusion model has the highest classification accuracy but takes the longest time to run. Therefore, our future research is not only to ensure the accuracy but also to speed up the convergence and reduce the running time.

Conflicts of Interest

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

Acknowledgments

This work is supported by National Natural Science Foundation of China (51475097), Ministry of Industry and Intelligent Manufacturing Demonstration Project (Ministry of Industry [2016]213), and Program of Guizhou Province of China (nos. JZ[2014]2001, [2015]02, and [2016]5103).

References

- [1] P. N. Saavedra and C. G. Rodriguez, "Accurate assessment of computed order tracking," *Shock Vibration*, vol. 13, no. 1, pp. 13–32, 2006.
- [2] J. Sanz, R. Perera, and C. Huerta, "Fault diagnosis of rotating machinery based on auto-associative neural networks and wavelet transforms," *Journal of Sound and Vibration*, vol. 302, no. 4–5, pp. 981–999, 2007.
- [3] B. Zhou and Y. Cheng, "Fault diagnosis for rolling bearing under variable conditions based on image recognition," *Shock and Vibration*, vol. 2016, Article ID 1948029, 2016.
- [4] C. Li, R.-V. Sánchez, G. Zurita, M. Cerrada, and D. Cabrera, "Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning," *Sensors*, vol. 16, no. 6, article no. 895, 2016.
- [5] O. Taouali, I. Jaffel, H. Lahdhiri, M. F. Harkat, and H. Messaoud, "New fault detection method based on reduced kernel principal component analysis (RKPCA)," *The International Journal of Advanced Manufacturing Technology*, vol. 85, no. 5–8, pp. 1547–1552, 2016.
- [6] J. Wodecki, P. Stefaniak, J. Obuchowski, A. Wylomanska, and R. Zimroz, "Combination of principal component analysis and time-frequency representations of multichannel vibration data for gearbox fault detection," *Journal of Vibroengineering*, vol. 18, no. 4, pp. 2167–2175, 2016.
- [7] J.-H. Cho, J.-M. Lee, S. W. Choi, D. Lee, and I.-B. Lee, "Fault identification for process monitoring using kernel principal component analysis," *Chemical Engineering Science*, vol. 60, no. 1, pp. 279–288, 2005.
- [8] V. H. Nguyen and J. C. Golinval, "Fault detection based on Kernel Principal Component Analysis," *Engineering Structures*, vol. 32, no. 11, pp. 3683–3691, 2010.
- [9] N. E. I. Karabadi, H. Seridi, I. Khelf, N. Azizi, and R. Boulkroune, "Improved decision tree construction based on attribute selection and data sampling for fault diagnosis in rotating machines," *Engineering Applications of Artificial Intelligence*, vol. 35, pp. 71–83, 2014.
- [10] L. Rutkowski, M. Jaworski, L. Pietruczuk, and P. Duda, "The CART decision tree for mining data streams," *Information Sciences*, vol. 266, pp. 1–15, 2014.
- [11] M. Amarnath, V. Sugumaran, and H. Kumar, "Exploiting sound signals for fault diagnosis of bearings using decision tree," *Measurement*, vol. 46, no. 3, pp. 1250–1256, 2013.
- [12] A. Krishnakumari, A. Elayaperumal, M. Saravanan, and C. Arvindan, "Fault diagnostics of spur gear using decision tree and fuzzy classifier," *The International Journal of Advanced Manufacturing Technology*, vol. 89, no. 9–12, pp. 3487–3494, 2017.
- [13] J. Rafiee, F. Arvani, A. Harifi, and M. H. Sadeghi, "Intelligent condition monitoring of a gearbox using artificial neural network," *Mechanical Systems & Signal Processing*, vol. 21, no. 4, pp. 1746–1754, 2007.
- [14] M. F. Umer and M. S. H. Khiyal, "Classification of textual documents using learning vector quantization," *Information Technology Journal*, vol. 6, no. 1, pp. 154–159, 2007.
- [15] P. Melin, J. Amezcua, F. Valdez, and O. Castillo, "A new neural network model based on the LVQ algorithm for multi-class classification of arrhythmias," *Information Sciences*, vol. 279, pp. 483–497, 2014.
- [16] A. Kushwah, S. Kumar, and R. M. Hegde, "Multi-sensor data fusion methods for indoor activity recognition using temporal evidence theory," *Pervasive and Mobile Computing*, vol. 21, pp. 19–29, 2015.
- [17] O. Basir and X. H. Yuan, "Engine fault diagnosis based on multi-sensor information fusion using Dempster-Shafer evidence theory," *Information Fusion*, vol. 8, no. 4, pp. 379–386, 2007.
- [18] D. Bhalla, R. K. Bansal, and H. O. Gupta, "Integrating AI based DGA fault diagnosis using dempster-shafer theory," *International Journal of Electrical Power & Energy Systems*, vol. 48, no. 1, pp. 31–38, 2013.
- [19] A. Feuerverger, Y. He, and S. Khatri, "Statistical significance of the Netflix challenge," *Statistical Science. A Review Journal of the Institute of Mathematical Statistics*, vol. 27, no. 2, pp. 202–231, 2012.
- [20] C. Delimitrou and C. Kozyrakis, "The netflix challenge: Data-center edition," *IEEE Computer Architecture Letters*, vol. 12, no. 1, pp. 29–32, 2013.
- [21] X. Hu, "The fault diagnosis of hydraulic pump based on the data fusion of D-S evidence theory," in *Proceedings of the 2012 2nd International Conference on Consumer Electronics, Communications and Networks, CECNet 2012*, pp. 2982–2984, IEEE, Yichang, China, April 2012.
- [22] X. Sun, J. Tan, Y. Wen, and C. Feng, "Rolling bearing fault diagnosis method based on data-driven random fuzzy evidence acquisition and Dempster-Shafer evidence theory," *Advances in Mechanical Engineering*, vol. 8, no. 1, 2016.
- [23] K. H. Hui, M. H. Lim, M. S. Leong, and S. M. Al-Obaidi, "Dempster-Shafer evidence theory for multi-bearing faults diagnosis," *Engineering Applications of Artificial Intelligence*, vol. 57, pp. 160–170, 2017.

- [24] H. Jiang, R. Wang, J. Gao, Z. Gao, and X. Gao, "Evidence fusion-based framework for condition evaluation of complex electromechanical system in process industry," *Knowledge-Based Systems*, vol. 124, pp. 176–187, 2017.
- [25] N. R. Sakthivel, V. Sugumaran, and S. Babudevasenapati, "Vibration based fault diagnosis of monoblock centrifugal pump using decision tree," *Expert Systems with Applications*, vol. 37, no. 6, pp. 4040–4049, 2010.
- [26] H. Talhaoui, A. Menacer, A. Kessal, and R. Kechida, "Fast Fourier and discrete wavelet transforms applied to sensorless vector control induction motor for rotor bar faults diagnosis," *ISA Transactions*, vol. 53, no. 5, pp. 1639–1649, 2014.
- [27] A. Rai and S. H. Upadhyay, "A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings," *Tribology International*, vol. 96, pp. 289–306, 2016.
- [28] Z. K. Peng and F. L. Chu, "Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography," *Mechanical Systems & Signal Processing*, vol. 18, no. 2, pp. 199–221, 2004.
- [29] T. L. Chen, G. Y. Tian, A. Sophian, and P. W. Que, "Feature extraction and selection for defect classification of pulsed eddy current NDT," *Ndt & E International*, vol. 41, no. 6, pp. 467–476, 2008.
- [30] D. Nova and P. A. Estévez, "A review of learning vector quantization classifiers," *Neural Computing and Applications*, vol. 25, no. 3, pp. 511–524, 2014.
- [31] O. Kreibich, J. Neuzil, and R. Smid, "Quality-based multiple-sensor fusion in an industrial wireless sensor network for MCM," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 9, pp. 4903–4911, 2014.
- [32] P. Kumari and A. Vaish, "Feature-level fusion of mental task's brain signal for an efficient identification system," *Neural Computing and Applications*, vol. 27, no. 3, pp. 659–669, 2016.
- [33] K. Gupta, S. N. Merchant, and U. B. Desai, "A novel multistage decision fusion for cognitive sensor networks using AND and OR rules," *Digital Signal Processing*, vol. 42, pp. 27–34, 2015.