Investigation of a multi-sensor data fusion technique for the fault diagnosis of gearboxes

Jun He [1], Shixi Yang [2], Evangelos Papatheou [3], Xin Xiong [4], Haibo Wan [5], Xiwen Gu [6],

[1,2,5,6] The State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou, 310027, China

[1,2,6] The Key Laboratory of Advanced Manufacturing Technology of Zhejiang Province, School of Mechanical Engineering, Zhejiang University, Hangzhou, 310027, China

[3] College of Engineering Mathematics and Physical Sciences, University of Exeter, Exeter EX4 4QF, UK

[4] School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, 200072, China

[5] College of Mechanical and Automotive Engineering, Zhejiang University of Water Resources and Electric Power, Hangzhou, 310058, China

Corresponding author:

Shixi Yang, The State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou, 310027, China

E-mail: yangsx@zju.edu.cn

Abstract: Gearbox is the key functional unit in a mechanical transmission system. As its operating condition being complex and the interference transmitting from diverse paths, the vibration signals collected from an individual sensor may not provide a fully accurate description on the health condition of a gearbox. For this reason, a new method for fault diagnosis of gearboxes based on multi-sensor data fusion is presented in this paper. There are three main steps in this method. First, prior to feature extraction, two signal processing methods, i.e., the energy operator and time synchronous averaging (TSA), are applied to multi-sensor vibration signals to remove interference and highlight fault characteristic information, then the statistical features are extracted from both the raw and preprocessed signals to form an original feature set. Second, a coupled feature selection scheme combining the distance evaluation technique (DET) and max-relevance and min-redundancy (mRMR) is carried out to obtain an optimal feature set. Finally, the deep belief network (DBN), a novel intelligent diagnosis method with a deep architecture, is applied to identify different gearbox health conditions. As the multi-sensor data fusion technique is utilized to provide sufficient and complementary information for fault diagnosis, this method holds the potential to overcome the shortcomings from an individual sensor that may not accurately describe the health conditions of gearboxes. Ten different gearbox health conditions are simulated to validate the performance of the proposed method. The results confirm the superiority of the proposed method in gearbox fault diagnosis.

Keywords: fault diagnosis, multi-sensor data fusion, feature selection, deep belief network, gearbox.

1 Introduction

Fault diagnosis of gearboxes is of great importance to avoid serious and even fatal accidents in various industrial applications. Vibration measurement and analysis, which is one of the most effective diagnosis methods for rotating machinery, has been widely applied to evaluate the health conditions of gearboxes.[1-4] Current researches in the fault diagnosis of gearboxes are mainly based on the vibration signals obtained from individual sensor, with the basic principal being detecting fault characteristics from the signals. However, the gearbox vibration signals consist of multiple complex components because of the unique mechanical structure, complicated operating conditions and massive amount of background noise, meaning that some useful fault characteristics contained in the vibration signals can be easily overwhelmed or distorted.[2] Thus, the extraction of fault characteristics can be a very challenging task. Moreover, it is typical to acquire the vibration signals using sensors mounted on the exterior of a gearbox case. Therefore, the vibration signals generated by multiple vibration sources in the gearbox must pass through complex transmission path before reaching the sensors. Consequently, the transfer function between the vibration sources and the sensors can easily modify the vibration signals, causing interference.[5] To overcome these shortcomings, one alternative is to use multi-sensor data fusion. This method has an advantage of multiple sensors mounted on several appropriate locations to provide sufficient and complementary information about the vibration state of a gearbox,[6] reducing the interference from transmission path effects. Furthermore, data fusion is good at screening out representative features from multi-sensor datasets, which can better characterize the health conditions of a gearbox.

In addition, effectively processing the multidimensional multi-sensor datasets and learning the complex non-linear relationships in the data is pivotal for gearbox fault diagnosis. Artificial intelligence (AI) techniques, which can effectively analyze a mass of data and automatically provide accurate diagnosis results, has good potential for this.[6-9] However, the gearbox vibration signals are usually non-linear, non-stationary and noisy. Some conventional AI techniques such as support vector machine (SVM) and back propagation neural network (BPNN) are shallow architectures, which contain no more than one non-linear transformation, do not easily learn the complex non-linear relationships in the data.[10,11] Deep learning has recently proven its superiority in feature learning and pattern recognition.[12-16] Compared to conventional intelligent techniques, this new technique utilizes unsupervised feature learning to learn a layer of feature representations from raw data adaptively.[11] Moreover, several layers of feature representations can be stacked to create deep networks, which are more capable of modeling complex structures in the data.[11,17] In view of the above advantages, we present a representative deep learning method called deep belief network (DBN) to diagnose gearbox faults in this study. The DBN contains a deep architecture with multiple stacked restricted Boltzmann machines (RBMs). Through unsupervised pre-trained in a layer-by-layer fashion and then supervised fine-tuned, the DBN is expected to have better performance than conventional intelligent diagnosis methods.

In general, fault diagnosis based on artificial intelligent techniques has three main steps: feature extraction, feature selection and fault classification. First, representative features related to the health conditions of machinery should be extracted from the signals by employing appropriate signal processing methods.[18] In this study, the statistical features associated with the time domain and frequency domain are used to characterize the health conditions of gearboxes. Besides, according to the characteristics of gearbox vibration signals, two signal processing methods, i.e., the energy operator and time synchronous averaging (TSA), are utilized to preprocess the vibration signals. The energy operator has the advantage of measuring the instantaneous energy changes of signals with a high time resolution and a good adaptability. Therefore, this method can highlight the transient features and is applicable for detecting changes in vibration signals caused by faults.[19] There have been a number of successful cases that applied this method for fault diagnosis of rotating machinery.[19-21] Additionally, the TSA is a typical preprocessing method for gearbox vibration signals, which can be used to eliminate signal components that are not synchronous with the shaft rate of rotation.[5,22] This method has the advantage of providing a direct visualization of the synchronously averaged signal, making some localized faults easily discernible. Besides, the calculation of some statistical features is also feasible to detect certain typical faults.[5,22] In this study, to acquire sufficient faulty information, we extract the statistical features from not only the raw vibration signals but also the preprocessed signals based on the energy operator and TSA. However, it is not appropriate to directly utilize all the features for subsequent classification since there still exists some irrelevant or redundant information in these extracted features, which may confuse the subsequent fault classification and decrease the accuracy. In addition, too many features may lead to higher computational cost and lower efficiency.[23] Therefore, feature selection is another indispensable step before classification. In this study, a coupled feature selection scheme combining the distance evaluation technique (DET)[24] and max-relevance and min-redundancy (mRMR)[25] is carried out. The DET is used to evaluate the ability of features in separating various health conditions, and the mRMR aims to minimize irrelevant or redundant information among these features. Consequently, an optimized feature set consists of robust features with less irrelevant or redundant information can be obtained for the subsequent fault classification.

Based on these studies, a novel method using multi-sensor data fusion is proposed for fault diagnosis of gearboxes in this paper. First, multi-sensor vibration signals are collected under different health conditions. Second, the energy operator and TSA are utilized to preprocess these signals. Then the statistical features are extracted from both the raw and preprocessed signals to form the original feature set. Third, a coupled feature selection scheme combining the DET and mRMR is carried out to select a more compact feature subset. Finally, these features are placed into the classifier based on DBN for the fault diagnosis of gearboxes. Ten different health conditions are simulated in a gearbox experimental system to validate the performance of the proposed method. The results demonstrate that the proposed method can obtain superior accuracy in fault diagnosis of gearboxes compared to other methods. The rest of this paper is organized as follows. In Section 2, the experimental system and vibration dataset are briefly described. The proposed method is detailed in Section 3. Section 4 presents the results of the experiment and the discussions of the results. Finally, Section 5 summarizes the conclusions.

2 Experiment description

In this work, a gearbox fault diagnosis system is designed to simulate different gearbox health conditions. As shown in Figure 1, the experimental system contains a two-stage gearbox, which is driven by an ac motor. The converter is used to control the driven speed and the data acquisition system is used to collect multi-sensors data. Three mono-axial accelerometers are used to acquire vibration signals at different locations, which are mounted on the input side of the gearbox, the output side of the gearbox and the mounting plate of the gearbox, respectively. The real speed of the input shaft is acquired by a tachometer. The input gear has 32 teeth, the idler gear has 64 teeth and the output gear has 96 teeth. Ten health conditions are tested in this work, some examples of the faulty gears are shown in Figure 2. In each experiment, the tachometer signal and the vibration signals of three accelerometers are considered simultaneously for multi-sensor data fusion. The sampling frequency is set to 25.6 KHz. The detailed descriptions on ten health conditions are displayed in Table 1.

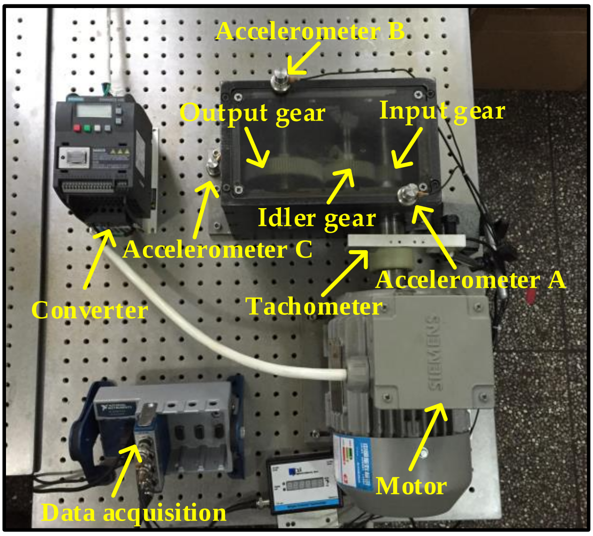

Figure 1. The experimental system.

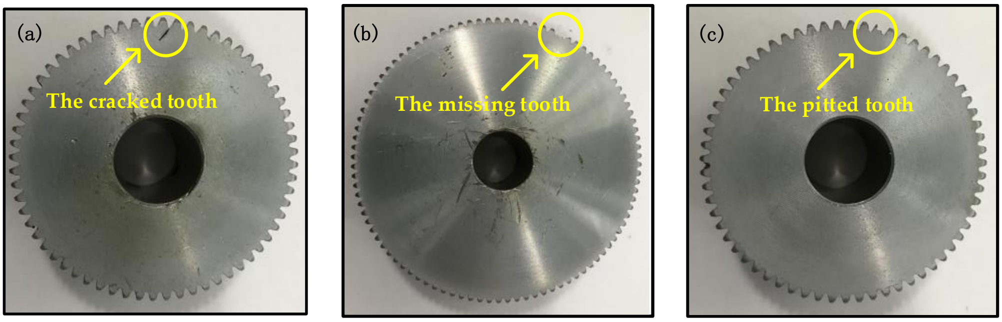

Figure 2. Examples of faulty gears: (a) the cracked tooth; (b) the missing tooth; (c) the pitted tooth.

Table 1. Description of the health conditions

Label	Condition	Driven speed (rpm)
1	a broken tooth on the input gear	2700
2	a pitted tooth on the input gear	
3	a pitted tooth on the idler gear	
4	a pitted tooth and a broken tooth on the output gear	
5	a missing tooth on the output gear	
6	a cracked tooth on the input gear	
7	a cracked tooth on the idler gear	
8	a cracked tooth on the output gear	
9	a broken tooth on the input gear and a pitted tooth on the idler gear	
10	normal	

3 Methods

In this section, a multi-sensor data fusion technique based on statistical analysis, the energy operator, the time synchronous averaging (TSA), the distance evaluation technique (DET), max-relevance and min-redundancy (mRMR) and the deep belief network (DBN) is proposed for the fault diagnosis of gearboxes. Figure 3 displays the procedure of the proposed method.

3.1 Feature extraction

The representative features related to the health conditions of gearboxes should be extracted using appropriate signal processing methods. In this study, fifteen features are utilized to characterize the health conditions of the gearbox. As shown in Table 2, these features may reflect the energy and distribution of vibration signals in the time domain or frequency domain, which can be used to detect changes in the signals caused by faults.[23,26] Furthermore, to eliminate background noise and highlight fault characteristic, two signal processing methods, i.e., the energy operator and TSA, are utilized to preprocess the vibration signals. Finally, the statistical features of both the raw and preprocessed signals are set together to form the original feature set.

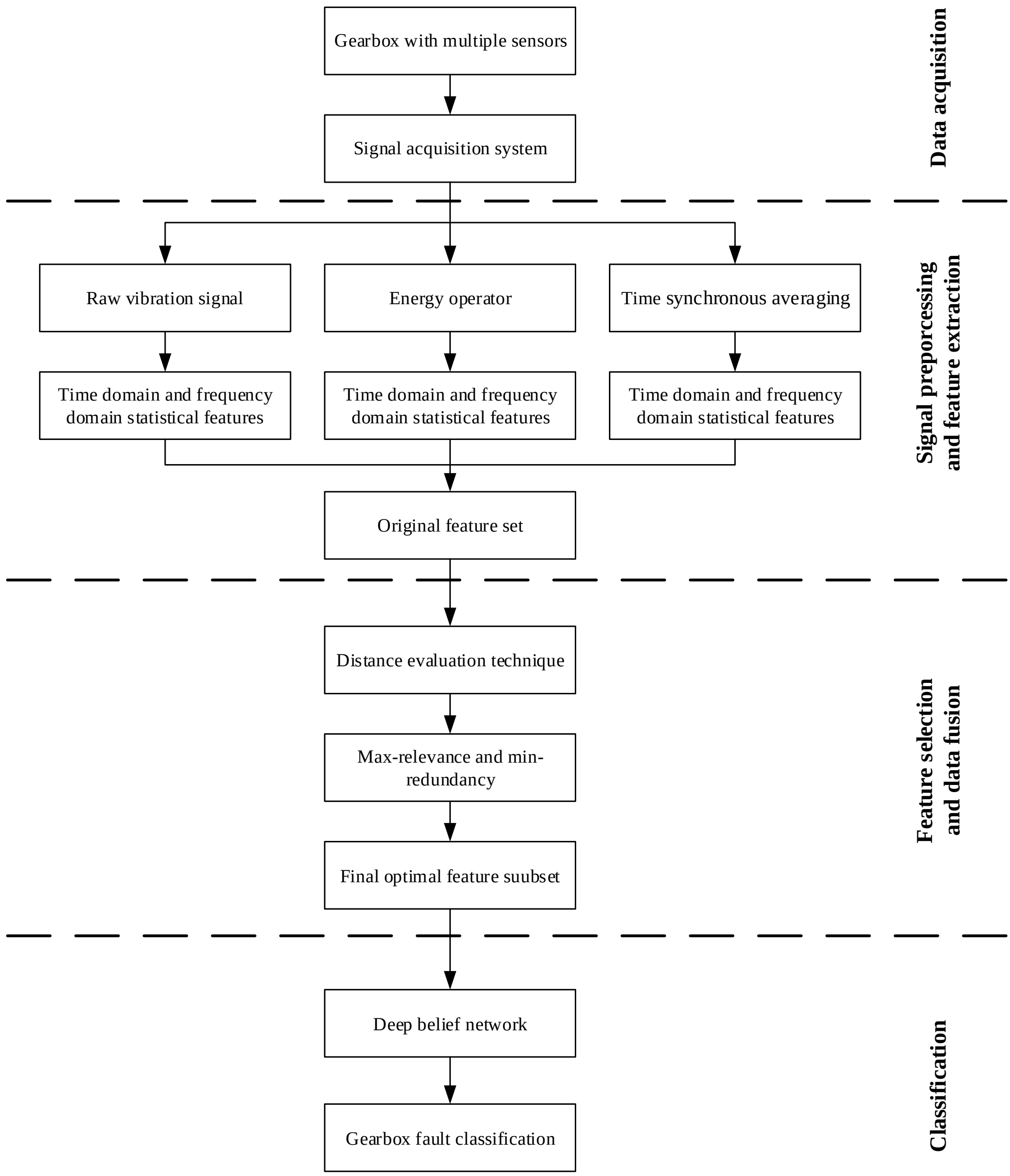

Figure 3. The flowchart of the proposed method

Table 2. The statistical features

Feature	Equation	Feature	Equation
Mean	$x_1 = \frac{1}{N}\sum_{i=1}^{N} x_i$	Rectified mean	$x_2 = \frac{1}{N}\sum_{i=1}^{N} \lvert x_i \rvert$
Peak to peak value	$x_3 = \max(x_i) - \min(x_i)$	Root mean square	$x_4 = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$

Standard deviation	$x_5=\sqrt{\frac{\sum_{i=1}^{N}\left(x_i-x_1\right)^2}{N-1}}$	Skewness	$x_6=\frac{\sum_{i=1}^{N}\left(x_i-x_1\right)^3}{\left(N-1\right)x_5^{\ 3}}$
Kurtosis	$x_7=\frac{\sum_{i=1}^{N}\left(x_i-x_1\right)^4}{\left(N-1\right)x_5^{\ 4}}$	Impulse factor	$x_8=\frac{\max\left(x_i\right)}{x_2}$
Shape factor	$x_9=\frac{x_4}{x_2}$	Crest factor	$x_{10}=\frac{\max\left(x_i\right)}{x_4}$
Coefficient of variation	$x_{11}=\frac{x_5}{x_1}$	Mean frequency	$x_{12}=\frac{\sum_{K=1}^{K}s\left(k\right)}{K}$
Frequency center	$x_{13}=\frac{\sum_{K=1}^{K}f_k s\left(k\right)}{\sum_{K=1}^{K}s\left(k\right)}$	Root mean square frequency	$x_{14}=\sqrt{\frac{\sum_{K=1}^{K}f_k^{\ 2}s\left(k\right)}{\sum_{K=1}^{K}s\left(k\right)}}$
Standard deviation frequency	$x_{15}=\sqrt{\frac{\sum_{K=1}^{K}\left(f_k-x_{13}\right)^2 s\left(k\right)}{\sum_{K=1}^{K}s\left(k\right)}}$		

Note: x_i is the i th value of time series x. N is the length of time series x. $s\left(k\right)$ is the spectrum value of the k th spectrum line. K is the number of spectrum lines. f_k is the frequency value of the k th spectrum line.

3.1.1 Energy operator

For a discrete time series $x\left(n\right)$, the energy operator $\psi\left[\cdot\right]$ is defined as:[19-21]

$$\psi\left[x\left(n\right)\right]=\left[x\left(n\right)\right]^2-x\left(n+1\right)x\left(n-1\right) \tag{1}$$

Because only three samples are required for computation at each time instant, the energy operator is convenient for operation and is nearly instantaneous, which holds the potential to capture the instantaneous energy fluctuations in the signals.[19] The amplitude envelope (AM signal) $a\left(n\right)$ and the instantaneous frequency $f\left(n\right)$ of the discrete time series can be estimated as follows:[27]

$$y\left(n\right)=x\left(n\right)-x\left(n-1\right) \tag{2}$$

$$\left|a\left(n\right)\right|=\sqrt{\frac{\psi\left[x\left(n\right)\right]}{1-\left(1-\frac{\psi\left[y\left(n\right)\right]+\psi\left[y\left(n+1\right)\right]}{4\psi\left[x\left(n\right)\right]}\right)^2}} \tag{3}$$

$$f(n)=\arccos\left(1-\frac{\psi[y(n)]+\psi[y(n+1)]}{4\psi[x(n)]}\right) \tag{4}$$

To prove the superiority of this method in extracting characteristic frequencies of vibration signal, we present the analysis of vibration signal of the fifth condition (a missing tooth on the output gear). Figure 4 shows the frequency spectrum of the raw signal acquired by accelerometer A and the corresponding AM signal calculated via the energy operator. Although the gear meshing frequency (GMF) and its sidebands can be observed, the rotating frequencies of each gear are not recognizable in the frequency spectrum of the raw signal because of the low signal noise ratio. Therefore, it is difficult to detect the missing tooth fault by using Fourier transform only. By contrast, the frequency spectrum of AM signal clearly exhibits the rotating frequencies of the input and output gears (RF1 and RF3), which is conductive to identify missing tooth fault. Obviously, this method can effectively highlight the characteristic frequency of the vibration signal and remove the noise.

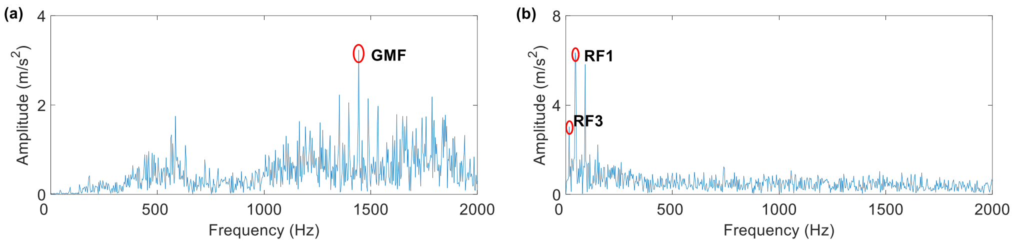

Figure 4. Frequency spectrum of signal of the fifth condition: (a) raw vibration signal; (b) AM signal

3.1.2 Time synchronous averaging

The time synchronous averaging (TSA) is a typical processing method for vibration signals of gearboxes, which can be used to remove components that are not synchronous with the gear in question.[5,22,28] Generally, the synchronously averaged signal is obtained from the measured accelerometer and tachometer signals. The tachometer signal is used to determine the zero crossing corresponding to one revolution of the gear. Then the vibration data between each of these zero crossings are interpolated by using cubic spline interpolation. Finally, the synchronously averaged signal can be obtained by dividing the sum of the interpolated data by the number of synchronous averages. The number of synchronous average should be determined according to the practical applications. In the time domain, the synchronously averaged signal shows the pattern of the tooth meshing vibration, including any modulation or distortion over one revolution, while the frequency domain gives the tooth meshing components and all the modulation sidebands at the shaft rotation frequency. Therefore, a simple visual inspection of the synchronously averaged signal or a calculation of some statistical features is feasible to detect certain typical faults.[5,22,28]

3.2 Feature selection

As described in section 3.1, we can obtain sufficient features from multi-sensor signals via appropriate signal processing methods. It should be noticed that different features have different degrees of importance when identifying different faults. For one specific fault diagnosis task, some features may be sensitive to specific faults while the others are not. Besides, too many features may lead to higher computational cost and lower efficiency. Therefore, feature selection is critical before fault classification. In this study, a coupled feature selection scheme is carried out to obtain an optimal feature set that can provide the most discrimination among the various faults and reduce the feature dimensionality.

3.2.1 Distance evaluate technique

In this subsection, the distance evaluate technique (DET) is applied to evaluate the sensitivity of each feature. Suppose that an original feature set of C conditions is

$$\left\{q_{m,c,j};\quad m=1,2,\cdots,M_c;\quad c=1,2,\cdots,C;\quad j=1,2,\cdots,J\right\} \tag{5}$$

where $q_{m,c,j}$ is the j^{th} feature of the m^{th} sample under the c^{th} condition, M_c is the sample number of the c^{th} condition, and J is the feature number of each sample.[26] There are three main steps.

(1) Calculating the average distance of each feature in the same condition $d_j^{(w)}$

$$d_j^{(w)} = \frac{1}{C\times M_c\times(M_c-1)}\sum_{c=1}^{C}\sum_{m,n=1}^{M_c}\left|q_{m,c,j}-q_{n,c,j}\right|,\quad (m\neq n) \tag{6}$$

(2) Calculating the average distance of each feature between different conditions

$$d_j^{(b)} = \frac{1}{C\times(C-1)}\sum_{c,e=1}^{C}\left|u_{e,j}-u_{c,j}\right|,\quad (c\neq e) \tag{7}$$

where $u_{c,j}$ is the average feature value of all samples under the same condition as follows:

$$u_{c,j} = \frac{1}{M_c}\sum_{m=1}^{M_c} q_{m,c,j} \tag{8}$$

(3) Calculating the distance evaluation criteria $\alpha_j = \dfrac{d_j^{(b)}}{d_j^{(w)}}$. It is clear that the features with larger α_j are better at separating different conditions.[24] Therefore, the sensitive features related to the faults can be selected according to the larger α_j.

3.2.2 Max-relevance and min-redundancy

Although the sensitive features can be selected via the DET, the redundancy between these selected features has not been given adequate attention. Several research has shown that the simple combination of some sensitive individual features does not necessarily lead to the satisfactory classification results.[23,25,29,30] Therefore, to minimize the irrelevant or redundant information for improving diagnosis performance, the max-relevance and min-redundancy (mRMR) is applied.

The main purpose of max-relevance is to find a feature set S containing m features $\{x_i\}$, which has the maximum relevance to the target class c. This relevance is evaluated by the mutual information $I(x_i,c)$ as follows:

$$I(x_i,c)=\iint p(x_i,c)\log\frac{p(x_i,c)}{p(x_i)p(c)}dx_i\,dc \tag{9}$$

where $p(x_i)$, $p(c)$ and $p(x_i,c)$ mean their probabilistic density functions, respectively.

Consequently, the m best individual features can be selected by the max-relevance criterion as follows:

$$\max D(S,c) \quad D=\frac{1}{|S|}\sum_{x\in S} I(x_i,c) \tag{10}$$

where $|S|$ denotes the number of features contained by S.[23,31]

However, it is possible that the features selected based on max-relevance may have rich redundancy.[32] Therefore, the min-redundancy criterion should be added to select mutually exclusive features to a certain extent, which is represented by

$$\min R(S), \quad R=\frac{1}{|S|^2}\sum_{x_i,x_j\in S} I(x_i,x_j), \quad (i\neq j) \tag{11}$$

and the operator $\phi(D,R)$ is defined to optimize D and R simultaneously:[23,31]

$$\max\phi(D,R) \quad \phi=D-R \tag{12}$$

Finally, combining with the advantages of DET and mRMR, the features that are sensitive to the faults and contain less irrelevant or redundant information can be selected.

3.3 Classification based on the Deep Belief Network

After determining the feature subset based on DET and mRMR, the deep belief network (DBN) is applied as the classifier. The DBN contains a deep architecture with multiple stacked restricted Boltzmann machines (RBMs). As shown in Figure 5, each RBM contains the visible units (input units) $\mathbf{v}$ and the hidden units $\mathbf{h}$, which are connected by a weight matix $\mathbf{w}$ and has bias vectors $\mathbf{c}$ and $\mathbf{b}$, respectively.[7,33]

The energy function is defined by the given visible and hidden units as

$$E(\mathbf{v},\mathbf{h}) = -\sum_{i=1}^{V}\sum_{j=1}^{H} w_{ij} v_i h_j - \sum_{i=1}^{V} c_i v_i - \sum_{j=1}^{H} b_j h_j \tag{13}$$

where v_i and h_j are the binary states of visible unit i and hidden unit j; c_i and b_j are their biases and w_{ij} is the weight between them; V and H are the number of visible and hidden units.[11,19] The joint distribution for the visible and hidden units is defined via the energy function as

$$P(\mathbf{v},\mathbf{h}) = \frac{1}{Z} e^{-E(\mathbf{v},\mathbf{h})} \tag{14}$$

where Z is the partition function that ensures the normalization of the joint distribution. For binary units, the conditional probabilities that v_i and h_j are mutually activated, are given by

$$P(h_j|\mathbf{v}) = \frac{1}{1+\exp\left(-b_j - \sum_i w_{ij} v_i\right)} \tag{15}$$

$$P(v_i|\mathbf{h}) = \frac{1}{1+\exp\left(-c_i - \sum_j w_{ij} h_j\right)} \tag{16}$$

respectively. The parameters $\mathbf{w}$, $\mathbf{c}$ and $\mathbf{b}$ are trained simultaneously to minimize the reconstruction error.[15,19,34] The gradient of the logarithmic probability of the training data with respect to a weight is defined as:

$$\frac{\partial \log P(\mathbf{v})}{\partial w_{ij}} = \left\langle v_i h_j \right\rangle_{data} - \left\langle v_i h_j \right\rangle_{model} \tag{17}$$

where $\langle \cdot \rangle_{data}$ and $\langle \cdot \rangle_{model}$ denotes the expectation under the distribution of the data and the model.[7,34,35]

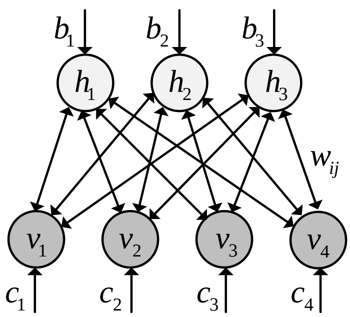

Figure 5. RBM architecture

A certain number of RBMs can be stacked to build a DBN. Given the input data, first the parameters of the lowest-layer RBM are continuously optimized by the contrastive divergence algorithm.[34-36] Then the hidden layer of this RBM becomes the input layer of the second RBM, and the second-layer RBM is trained in the same way. Finally, all the layers of the DBN can be optimized through this layer-by-layer unsupervised training process. For classification tasks, after the unsupervised training process is finished, all layers are fine-tuned under supervision with the back propagation (BP) algorithm. This fine-tuning process further reduces the training error by using the label information of data.

4 Data analysis and discussion

The vibration signals used in this section are collected from the gearbox fault diagnosis system displayed in section 2. For each health condition, 1200 samples can be acquired in the experiments. Therefore, a total of 12000 samples corresponding to ten health conditions are collected for data analysis. Each sample contains three channels of vibration signals and one channel of tachometer signal, with 6400 data points of each channel. To provide sufficient information on the health conditions of the gearbox, various features are extracted from the multi-sensor signals.

First, 15 statistical features which are listed in Table 2 can be extracted from each channel of vibration signals. Hence, 15×3 features are obtained for each sample.

Second, the energy operator is performed on the vibration signals to calculate the corresponding AM signals, which can effectively remove the interference and highlight the fault characteristic information. Then, 15 statistical features can be extracted from each channel of AM signals and a total of 15×3 features are obtained for each sample.

Third, vibration signals of the input and output sides of the gearbox are collected by accelerometers A and B, respectively. The TSA technique is then performed on these two channels to remove noise and components which are nonsynchronous with the gear in question. Then, each synchronously averaged signal is further processed to extract 15 statistical features. Therefore, another 15×2 features are obtained for each sample.

In this way, an original feature set containing 120 features is obtained for each sample. It is clear that different features have different degrees of importance when identifying different faults. Some of the above features are sensitive to the health conditions of the gearbox while the others are not. Practically, representative features are selected for one specific task by researchers based on diagnostic techniques and field expertise. In this study, we prefer to select robust features automatically using appropriate feature selection methods. Therefore, as various features have been extracted, a coupled feature selection scheme is then carried out to obtain an optimal feature set to provide the most discrimination among the various faults and reduce the feature dimensionality. The DET is utilized to evaluate the sensitiveness of the features related to the faults, and the mRMR is applied to further consider the redundancy between features based on mutual information. Figure 6 shows the priorities of all features estimated based on these two methods. The horizontal and ordinate axes demonstrate the priorities of all features estimated by the DET and mRMR, respectively. Two threshold lines are set to divide the features into four quadrants. It is clear that the features located in the third quadrant are better at separating the various health conditions with less irrelevant or redundant information, as these features get higher rankings when two methods are applied to evaluate them. The features located in the second quadrant are also sensitive to the faults. Nevertheless, these features may have rich redundancy, meaning that the respective discriminative ability does not change significantly even if one of them is removed. Consequently, only the features located in the third quadrant are selected for subsequent fault classification in this study. Finally, 55 features can be selected from the original feature set when two thresholds are both set to 60, among which 22, 20 and 13 features are extracted from the vibration signal acquired by accelerometer A, B and C, respectively.

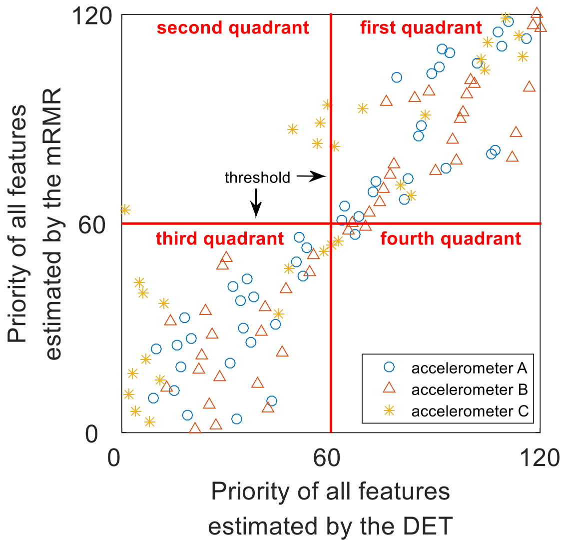

Figure 6. Priorities of all features

For comparative analysis, five datasets, A-E, are used in this study. Datasets A, B and C only contain the features extracted from the individual sensor A, B and C, respectively. Dataset D contains the 55 features selected by the coupled feature selection scheme. Dataset E contains all the features extracted from three channels of vibration signals. The k-fold validation method is used to partition these samples, where k is chosen as five. Therefore, in each trail, the training set contains 9600 samples and the testing set contains 2400 samples. Then the average accuracy and the standard deviation over five trails are calculated.

The designed DBN has four layers, in which the unit number of the input layer is equal to the dimension of the input data, and the unit number of the other three layers are 100, 50 and 10 respectively. During both the process of unsupervised pre-training and supervised fine-tuning, the maximum training epoch is set to 100. Besides, the learning rate is set to 0.001 and the momentum is set to 0.5. In addition, the BPNN is a typical AI method and has been successfully used in fault diagnosis of gearboxes. Therefore, this method is tested using the same datasets to perform a comparative study. The architecture of the BPNN is as the same as the DBN. The classification accuracies obtained from these two methods are shown in Table 3. Generally, DBN provides higher classification accuracies for all the five feature sets in comparison with BPNN, confirming that DBN is superior at achieving accurate fault classification of the gearbox compared to shallow neural network models. In detail, DBN achieves 99.04% classification accuracy which is higher than that of BPNN (97.97%) when using dataset D. For dataset E, the average accuracy of DBN is 97.88% whereas BPNN is 94.07%. Compared to dataset D, the accuracies of both classifiers are relatively lower when using dataset E. This is because there are still irrelevance or redundancy in these extracted features, which would confuse the classification process and lead to lower accuracy. In contrast, dataset D contains the robust features obtained by the coupled feature selection scheme. These features are sensitive to the faults with less irrelevance and redundancy, thus can better discriminate various health conditions and improve the classification accuracy. Additionally, the performance of these two classifiers when using dataset A~C are comparatively poor. The accuracies of DBN range from 76.42% to 93.55%, whereas the accuracies of BPNN range from 69.09% to 88.08%. These results show that it is difficult to accurately describe the health conditions of the gearbox only utilizing the diagnosis information from an individual sensor. Alternatively, multiple sensors mounted on several appropriate locations can provide sufficient and complementary diagnostic information, which is conductive to improve the classification accuracy.

Table 3. Diagnosis results of the gearbox

Datasets	DBN		BPNN	
	Average accuracy (%)	Standard deviation (%)	Average accuracy (%)	Standard deviation (%)
Dataset A	80.82	1.09	76.83	1.07
Dataset B	93.55	0.75	88.08	5.66
Dataset C	76.42	5.52	69.09	4.33
Dataset D	99.04	0.25	97.97	1.18
Dataset E	97.88	1.71	94.07	2.18

Figure 7 shows the confusion matrix produced by the proposed method in one trail when using dataset D. The ordinate and horizontal axis of the confusion matrix refer to the actual label and the predict label of classification, respectively. It shows that the classification accuracies of all conditions are higher than 96% and some even reach 100%. Besides, the principal component analysis (PCA) is used for visualizing the separation among the feature clusters of different health conditions. Figure 8 shows the first three principal components (PCs) of the features in datasets A~D. In Figure 8(a) ~ (c), we can find that many samples are mixed together even though they represent different health conditions. On the contrary, in Figure 8(d), most samples with the same health condition are clustered and most samples with different health conditions are separated. Note that there are some overlaps between different clusters in Figure 8(d) due to the rotation angle of the figure. These results reveal that the features selected by the proposed method can characterize the health conditions of the gearbox effectively.

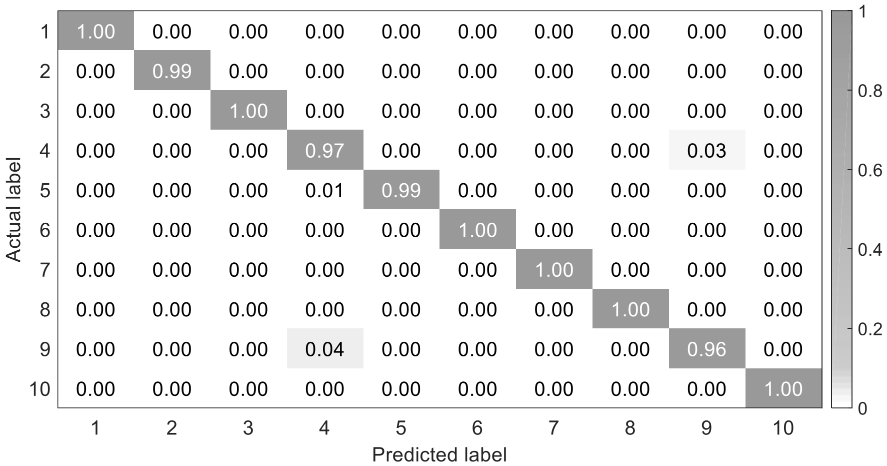

Figure 7. The confusion matrix produced by the proposed method in one trail

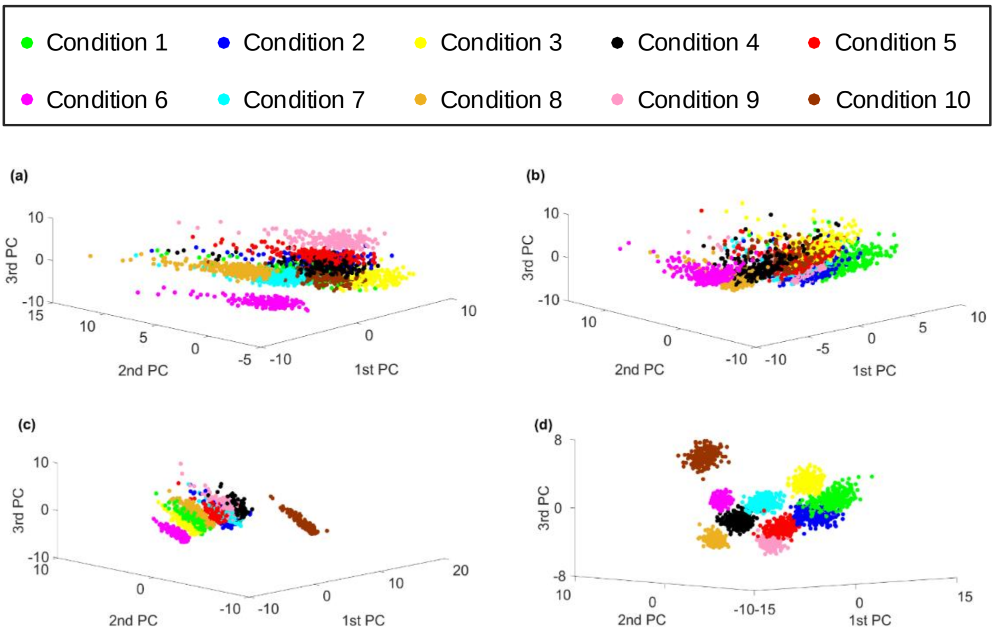

Figure 8. Scatter plots of principal components for the features: (a) dataset A; (b) dataset B; (c) dataset C; (d) dataset D

5 Conclusions

This paper presents a novel multi-sensor fault diagnosis method based on statistical analysis, the energy operator, the time synchronous averaging (TSA), the distance evaluation technique (DET), max-relevance and min-redundancy (mRMR) and the deep belief network (DBN). In this method, multi-sensor data are simultaneously applied to provide sufficient and complementary information for fault diagnosis, thus overcomes the shortcomings that individual sensor data may not accurately describe the health conditions of gearboxes. Besides, to remove interference components and highlight fault characteristic information, two signal processing methods, i.e., the energy operator and TSA, are utilized to preprocess these signals. Then, the statistical features are extracted from both the raw and preprocessed signal to form the original feature set. Furthermore, in order to select robust features related to the faults and minimize irrelevant or redundant information, a coupled feature selection scheme combining the DET and mRMR is carried out to obtain an optimal feature set. Finally, these selected features are applied as the input of the DBN to achieve fault classification of gearboxes.

Ten different gearbox health conditions are tested in a gearbox experimental system to evaluate the effectiveness of the proposed method. For comparison, other four datasets, three of which only contain features extracted from each individual sensor data and the other one contains all features without feature selection are also implemented. Moreover, a comparative study of the performance of DBN and BPNN is also carried out. The fault classification accuracy is 99.04% when using the proposed method, which is much higher than the other methods. These results confirm the superiority of the proposed method in the fault diagnosis of gearboxes.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This study was supported by the National Natural Science Foundation of China under Grant No. 51575497 and 51375434.

References

1. Li C and Liang M. Time-frequency signal analysis for gearbox fault diagnosis using a generalized synchrosqueezing transform. *Mech Syst Sig Process* 2012; 26: 205-217.
2. Feng ZP and Liang M. Complex signal analysis for wind turbine planetary gearbox fault diagnosis via iterative atomic decomposition thresholding. *J Sound Vib* 2014; 333: 5196-5211.
3. Xiong X, Yang SX and Gan CB. A new procedure for extracting fault feature of multi-frequency signal from rotating machinery. *Mech Syst Sig Process* 2012; 32: 306-319.
4. Feng Y, Lu BC and Zhang DF. Multiscale morphological manifold for rolling bearing fault diagnosis. *Proc IMechE, Part C: J Mechanical Engineering Science* 2017; 231: 3516-3529.
5. McFadden PD. Detecting fatigue cracks in gears by amplitude and phase demodulation of the meshing vibration. *J Vib Acoust* 1986; 108: 165-170.
6. Lei YG, Lin J, He ZJ, et al. A method based on multi-sensor data fusion for fault detection of planetary gearboxes. *Sens* 2012; 12: 2005.
7. He J, Yang SX and Gan CB. Unsupervised fault diagnosis of a gear transmission chain using a deep belief network. *Sens* 2017; 17: 1564.
8. Jia F, Lei YG, Lin J, et al. Deep neural networks: A promising tool for fault characteristic mining and

intelligent diagnosis of rotating machinery with massive data. *Mech Syst Sig Process* 2016; 72-73: 303-315.

9. Zhu KH and Li H. A rolling element bearing fault diagnosis approach based on hierarchical fuzzy entropy and support vector machine. *Proc IMechE, Part C: J Mechanical Engineering Science* 2016; 230: 2314-2322.
10. Shao HD, Jiang HK, Zhao HW, et al. A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. *Mech Syst Sig Process* 2017; 95: 187-204.
11. Längkvist M, Karlsson L and Loutfi A. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognit Lett* 2014; 42: 11-24.
12. LeCun Y, Bengio Y and Hinton G. Deep learning. *Nature* 2015; 521: 436-444.
13. Schmidhuber J. Deep learning in neural networks: An overview. *Neural Networks* 2015; 61: 85-117.
14. Hinton G, Deng L, Yu D, et al**.** Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Process Mag* 2012; 29: 82-97.
15. Längkvist M, Karlsson L and Loutfi A. Sleep stage classification using unsupervised feature learning. *Adv Artif Neural Syst* 2012; 2012: 5-5.
16. Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vision* 2015; 115: 211-252.
17. Li C, Sánchez RV, Zurita G, et al. Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning. *Sens* 2016; 16: 895.
18. Lei YG, Lin J, Han D, et al. An enhanced stochastic resonance method for weak feature extraction from vibration signals in bearing fault detection. *Proc IMechE, Part C: J Mechanical Engineering Science* 2014; 228: 815-827.
19. Tran VT, AlThobiani F and Ball A. An approach to fault diagnosis of reciprocating compressor valves using Teager-Kaiser energy operator and deep belief networks. *Expert Syst Appl* 2014; 41: 4113-4122.
20. Cexus JC and Boudraa AO. Teager-Huang analysis applied to sonar target recognition. In: *Proceedings of World Academy of Science Engineering and Technology* 2005; pp. 108-111.
21. Kaiser JF. Some useful properties of Teager's energy operators. In: *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing* 1993, pp.149-152.
22. Randall RB. A new method of modeling gear faults. *J Mech Des* 1982; 104: 259-267.
23. Li B, Zhang PL, Tian H, et al. A new feature extraction and selection scheme for hybrid fault diagnosis of gearbox. *Expert Syst Appl* 2011; 38: 10000-10009.
24. Hu Q, He ZJ, Zhang ZS, et al. Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble. *Mech Syst Sig Process* 2007; 21: 688-705.
25. Peng HC, Long FH and Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell* 2005; 27: 1226-1238.
26. Lei YG, He ZJ, Zi YY, et al. New clustering algorithm-based fault diagnosis using compensation distance evaluation technique. *Mech Syst Sig Process* 2008; 22: 419-435.
27. Henriquez RP, Alonso JB, Ferrer MA, et al. Application of the Teager-Kaiser energy operator in bearing

fault diagnosis. *ISA Trans* 2013; 52: 278-284.

28. Zakrajsek JJ, Handschuh RF and Decker HJ. *Application of fault detection techniques to spiral bevel gear fatigue data.* Technical Report, No. AD-A289970, NASA Lewis Research Center, Cleveland, OH, United States, 1994.
29. Kohavi R, John GH. Wrappers for feature subset selection. *Artif Intell* 1997; 97: 273-324.
30. Cover TM. The best two independent measurements are not the two best. *IEEE Trans. Syst. Man Cybern* 1974; SMC-4: 116-117.
31. Li B, Zhang PL, Liang SB, et al. Feature extraction and selection for fault diagnosis of gear using wavelet entropy and mutual information. In: *9th International Conference on Signal Processing*, Beijing, China, 26-29 Oct 2008, pp.2846-2850.
32. Ren JF, Jiang XD, Yuan JS, et al. Optimizing LBP structure for visual recognition using binary quadratic programming. *IEEE Signal Process Lett* 2014; 21: 1346-1350.
33. Hinton GE, Osindero S and Teh YW. A fast learning algorithm for deep belief nets. *Neural Comput* 2006; 18: 1527-1554.
34. Hinton GE. A practical guide to training restricted Boltzmann machines. In: *Montavon G., Orr G.B., Müller KR. (eds) Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 2012, Vol.7700.
35. Hinton GE. Training products of experts by minimizing contrastive Divergence. *Neural Comput* 2002; 14: 1771-1800.
36. Tamilselvan P and Wang PF. Failure diagnosis using deep belief learning based health state classification. *Reliab Eng Syst Saf* 2013; 115: 124-135.