Rolling Bearing Fault Diagnosis Based on DTCWPT and DBN

Shiyuan Liu, Cui Zhao, Fangyuan He*
College of Applied Science and Technology, Beijing Union University, Beijing, China
**Corresponding Author.*

Abstract: This paper constructs a fault analysis and identification model by employing the dual-tree complex wavelet packet transform (DTCWPT) together with the deep belief network (DBN), aiming to achieve precise fault diagnosis of rolling bearings. The DTCWPT decomposition of the vibration signals was first performed, and an initial feature set of the fault pattern was developed by extracting the features of initial vibration signals under different frequency bands. Subsequently, the Laplacian Score (LS) approach was utilized to detect the fault-sensitive characteristics within the original high-dimensional feature collection. Furthermore, leveraging depth learning techniques which are proficient in high-dimensional data manipulation and nonlinear data analysis, an adaptive exploration of fault characteristics and an intelligent discrimination of faults were carried out with the Dunn Validity Index and standard deviation ratio in the context of DBN. Four tests for different cases were performed using the previously reported bearing data. The experimental outcomes manifested that the LS approach is efficacious in extracting fault-sensitive features, and the DBN model is capable of enhancing the precision of fault recognition.

Keywords: Dual-Tree Complex Wavelet Packet; Deep Belief Network; Feature Extraction; Rolling Bearings; Fault Diagnosis

1. Introduction

Since rolling bearings operate under high variable load conditions over a prolonged period and are affected by external loads or unstable factors, they are prone to fault states [1], which can result in unexpected equipment damage and financial losses. Consequently, the fault diagnosis of rolling bearings is of significance. Researches on the vibration mechanisms of rolling bearings have indicated that the vibration signals generated during the operation of bearings encompass abundant state information, and the analysis of vibration signals represents an efficient approach for conducting fault diagnosis of rolling bearings. In recent times, owing to the progressions in signal processing, data mining, and artificial intelligence technologies, data-propelled intelligent fault diagnosis techniques have been extensively utilized in the fault diagnosis of bearings [2]. The main steps involved in bearing fault diagnosis include signal processing, feature extraction, feature selection, and identification of fault models [3, 4]. Essentially, the process of diagnosing faults in rolling bearings is tantamount to the process of recognizing fault state models [4]. Hence, the first three steps are the foundation for the fourth step.

One of the essential difficulties in extracting features of rolling bearing faults is to disassemble the fundamental characteristics that reflect fault types from the nonlinear and non-stationary vibration signals emerging during fault conditions [5]. The wavelet transform allows a refined multi-resolution analysis of vibration signals through several operations, such as scaling transform and translation [6-8]. Nevertheless, it encounters the trade-off predicament between time resolution and frequency resolution as a result of their mutual interaction and dependence. The transform coefficient and scale coefficient of DTCWPT possess translation invariance, resistance to frequency aliasing, limited redundancy, and no phase distortion, rendering it a preferable approach for extracting fault signal characteristics of rolling bearings. Qu [8] used DTCWPT for the feature extraction of vibration signals and combined multiple classifiers for fault diagnosis. Wu et al. [9] extracted singular spectrum features of diesel engines under different working conditions using DTCWPT.

The statistical features and equipment fault modes obtained by feature extraction have complex mapping relationships, and different statistical features exhibit varying sensitivities to faults. Without prior knowledge, manually selecting statistical features based on personal experience for fault diagnosis yields unsatisfactory results [9]. To address this issue, this study uses the Laplacian Score (LS) [10] feature selection method in order to evaluate the statistical features, quantitatively analyze their fault sensitivity, and select highly correlated and sensitive statistical features for the construction of the training set.

Regarding the model recognition issue of rolling bearing fault states, shallow models like the BP neural network and Support Vector Machine (SVM) are unable to effectively depict the intricate mapping relationship between the high-dimensional feature space of measured signals and rolling bearing faults, and they are liable to the problem of dimensionality. Deep learning, which is an emerging machine learning method, can automatically extract the necessary features from large datasets, which makes it suitable for the diagnostic analysis of diverse, nonlinear, and high-dimensional data in big data. The deep belief network (DBN), a representative deep learning algorithm, has been successfully utilized in information retrieval, feature dimensionality reduction as well as fault classification. Wang et al. [11] used DBN for intrusion detection, which selected features layer by layer to perform feature dimensionality reduction. Shi et al. employed DBN for transformer fault classification, and compared it with traditional classification methods in terms of feature extraction and fault tolerance. Shao et al. [1] utilized DBN for the fault diagnosis of rolling bearings with the aim of enhancing the accuracy of fault recognition. In this study, wavelet packet decomposition was first performed on collected vibration signals to extract their initial feature set. The LS method was then used to select fault-sensitive features that were used as the input space for the DBN training. Ultimately, the trained DBN was adopted for fault diagnosis. The experimental results showed that the proposed model boasts a high performance in recognizing rolling bearing faults.

2. DTCWPT and DBN

2.1 DTCWPT

DTCWPT is an enhanced wavelet transform that uses two parallel real wavelet transform trees with distinct high-pass and low-pass filter banks referred to as the real tree and the imaginary tree, respectively.

DTCWT independently decomposes and reconstructs the input signals using two discrete wavelet-based functions to obtain the real and imaginary parts of the complex wavelet. The complex wavelet is expressed as:

$$\varphi(t) = \varphi_h(t) + j\varphi_g(t) \quad (1)$$

where φh(t) and φg(t) are respectively the real and imaginary parts of the complex wavelet that are both real functions.

DTCWT can be considered as the operation of two parallel real wavelets. In the process of the signal decomposition and reconstruction process, The sample points of the imaginary tree always remain positioned in the middle of the real tree. This can ensure the effective utilization of wavelet decomposition coefficients from the two trees, helps achieve information complementarity and approximate translational invariance of the real and imaginary trees, and reduces the useful information loss to a certain extent. Moreover, DTCWT uses wavelet coefficient dichotomy in each decomposition layer, which cuts down the computational redundancy and boosts the computational efficiency.

Since the discrete wavelet transform cannot perform fine-grained decomposition of high-frequency parts, the wavelet packet transform is introduced. Similarly, the DTCWPT, which retains the advantages of DTCWT and can achieve effective decomposition of high-frequency signal parts, is proposed.

2.2 Printing Area

Hinton et al. [11] put forward DBN possessing multiple hidden layers, which are formed by stacking multiple Restricted Boltzmann Machines (RBMs) and whose neural network weights are trained through a layer-wise greedy learning algorithm. The raw data are input into the bottom layer of DBN, and data features are successively extracted through each layer. This process gradually abstracts information from lower layers to higher layers, which results in the formation of features with

increasingly enhanced representational capacities. The hierarchical structure and learning process of DBN is shown in Figure 1.

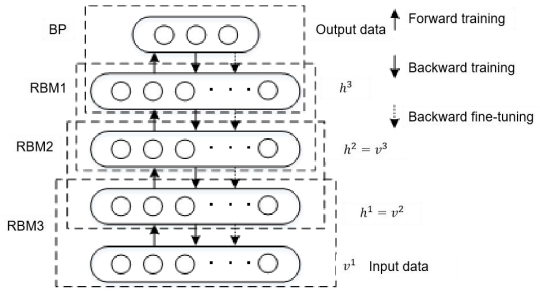


Figure 1. DBN Structure Model and Learning Process

The learning paradigm of the Deep Belief Network (DBN) can be bifurcated into two sequential phases: the unsupervised pre-training phase, which aims to initialize the network parameters in an unsupervised manner to capture the underlying data distribution, and the subsequent supervised error fine-tuning training phase, wherein the network is refined using labeled data to minimize the prediction errors and enhance its discriminative capabilities. After parameter initialization, DBN uses a greedy algorithm to perform unsupervised training of RBMs layer by layer in order to determine the initial parameters. In this procedure, the output generated by the lower layer functions as the input for the upper layer. After pretraining, DBN employs the BP algorithm for supervised fine-tuning, where the errors are propagated to each layer of RBM from top to bottom in order to perform parameter fine-tuning.

The training of RBMs is the core of the DBN learning process. Each RBM consists of visible and hidden layers, and represents a specific type of Markov random field. The visible layer receives data and transforms them into the hidden layer to complete the learning process. RBM is an energy-based model. The joint energy of visible variable $V = (v_1, v_2, \ldots, v_n)^T$ and hidden variable $H = (h_1, h_2, \ldots, h_m)^T$ is given by:

$$E(v,h;\theta) = -\sum_{ij} W_{ij} v_i h_j - \sum_i b_i v_i - \sum_j a_j h_j \quad (2)$$

where θ represents the parameter of RBM, a_i and b_j are respectively the biases of the hidden and visible units, and W is the weight connecting the two units.

the joint probability of V and H is denoted based on this energy function as:

$$P_\theta(v,h;\theta) = \frac{1}{Z(\theta)} \exp(-E(v,h;\theta)) \quad (3)$$

where Z(θ) is the normalization factor which is also referred to as the partition function.

Hence, the conditional probabilities of the visible and hidden units are given by:

There are no connections within an RBM layer. That is, given V, the hidden units are independent from each other; given H, the visible units are also independent from each other. Thus, the conditional probability distributions can be computed as:

$$P(v_i = 1 \mid h) = sigm(a_i + \sum_j w_{ij} h_j) = \frac{1}{1 + \exp(-a_i - \sum_j w_{ij} h_j)} \quad (4)$$

$$P(h_j = 1 \mid v) = sigm(b_j + \sum_i v_i w_{ij}) = \frac{1}{1 + \exp(-b_j - \sum_j v_i w_{ij})} \quad (5)$$

The training procedure of RBMs is concisely presented as follows:(1) The visible unit is set as the training sample data and the hidden unit is updated using Eq. (7); (2) v'_I is reconstructed based on the state of the hidden unit and Eq. (6), and h'_j is determined using Eq. (7). This process is repeated to update the weight w_{ij}, which can be calculated as:

$$\Delta w_{ij} = \eta(< v_i h_j > - < v'_i h'_j >) \quad (6)$$

where $\eta \in (0, 1)$ represents the learning rate and $< \cdot >$ represents the average of the training data.

3. LS-based Feature Selection Method

The LS is a feature selection method. Based on the Laplacian Eigenmap (LE) and Local Preserving Projection (LPP), this method assesses features by gauging their capacity to maintain the local geometric structure of data.It calculates the LS of the feature set with the aim of extracting the inherent information structure and transforms the intricate high-dimensional feature space into a comparatively uncomplicated low-dimensional space [9]., and thus reduces the feature space redundancy. It selects the features having smaller LS values in the feature space, which allows to retain most of the intrinsic geometric structure information in the fault signal feature set, improving the discriminative performance of the reduced low-dimensional feature set.

Let L_r the LS of the r^{th} feature and f_{ri} the r^{th} feature of the i^{th} sample, where i = 1, 2, …, m, The LS algorithm can be broken down into the

following steps:

The nearest neighbor graph G with m nodes is constructed, where x_i and x_j correspond to the i^{th} and j^{th} nodes, respectively. If x_i and x_j are neighbors, they are connected by an edge, that is, x_i is the k-nearest neighbor of x_j or x_j is the k-nearest neighbor of x_i.In the presence of label information, an edge is also established between two nodes that possess the same label. If nodes i and j are connected, then:

$$S_{ij} = \exp(-\| x_i - x_j \|^2 / t) \quad (7)$$

where t is an appropriate constant and $\|\cdot\|$ is the Euclidean distance between two nodes.

Otherwise, $\boldsymbol{S_{ij}}$ is set to 0. The weighted matrix S represents a similarity matrix corresponding to the nearest neighbor graph G. It serves the purpose of emulating the inherent local geometric configuration within the data space and gauging the similarity among neighboring sample points. A greater value of $\boldsymbol{S_{ij}}$ indicates a stronger similarity between the two samples and a higher likelihood that the two samples belong to the same class. The following is defined for the r^{th} feature:

$$\begin{cases} f_r = (f_{r1}, f_{r2}, \cdots, f_{rm})^{\mathrm{T}} \\ D = \mathrm{diag}(SI) \\ I = (1,1,\cdots,1)^{\mathrm{T}} \\ L = D - S \end{cases} \quad (8)$$

Matrix L is often referred to as the Laplace operator. In order to avoid generating a non-zero constant vector f_r due to the excessive data variance in some dimensions, which leads to the generation of a meaningless value of L_r in the next step, mean centering is performed for each feature as:

$$\tilde{f}_r = f_r - \frac{f^{\mathrm{T}} DI}{I^{\mathrm{T}} DI} I \quad (9)$$

The LS of the th feature is calculated as:

$$L_r = \frac{\tilde{f}_r^T L \tilde{f}_r}{\tilde{f}_r^T D \tilde{f}_r} = \frac{\sum_{ij} (f_{ri} - f_{rj})^2 S_{ij}}{Var(f_r)} \quad (10)$$

where $\mathrm{Var}(f_r)$ is the estimation variance of the r^{th} feature.

According to the above steps, for a specific feature, when the value of $\boldsymbol{S_{ij}}$ increases, the value of $(f_{ri} - f_{rj})^2$ decreases. That is, the smaller the numerator, the higher the similarity among the samples in this feature, and the larger the value of $\mathrm{Var}(f_r)$, the higher the degree of distinction among the samples. Therefore, the LS and feature sensitivity are inversely proportional. In other terms, a smaller LS value implies a greater sensitivity of the feature to bearing faults. Via computation, features are ordered based on the LS value in an ascending manner, and several top-ranked features are chosen to form a set of sensitive features.

4. Process of Fault Diagnosis

4.1 Feature Extraction

The dmey wavelet was employed to conduct a four-layer Discrete Ternary Complex Wavelet Packet Transform (DTCWPT) decomposition on the vibration signal samples. Subsequently, 16 terminal nodes as well as the corresponding wavelet packet coefficients were obtained. The Hilbert Envelope Spectrum (HES) was calculated by using the 16 reconstructed signals obtained through performing single-branch wavelet packet reconstruction on each node coefficient in the fourth layer of the wavelet packet tree. Furthermore, by utilizing the 16 single-branch reconstructed signals and their Hilbert Envelope Spectrum (with a total of 32 sequences),11 statistical features were calculated. Upon the construction of the initial feature set, each fault signal sample was associated with 352 statistical features. Table 1 presents the 11 statistical features along with their calculation equations.

Table 1. 11 Statistical Features of the Signals (x Denotes a Sequence Having a Length of n)

Feature	Equation
Amplitude	$T_1 = \max(\lvert x(i)\rvert) - \min(\lvert x(i)\rvert)$
Mean	$T_2 = (1/n)\sum_{i=1}^{n} x(i)$
Standard deviation	$T_3 = \sqrt{(1/(n-1))\sum_{i=1}^{n}(x(i)-T_1)^2}$
Peakness	$T_4 = \sum_{i=1}^{n}(x(i)-T_1)^3 / ((n-1)T_2^3)$
Kurtosis	$T_5 = \sum_{i=1}^{n}(x(i)-T_1)^3 / ((n-1)T_2^3)$
Energy	$T_6 = \sum_{i=1}^{n}\lvert sp(k)\rvert^2$
Energy entropy	$T_7 = -\sum_{i=1}^{n} P_i \log P_i, P_i = \frac{\lvert x(i)\rvert^2}{Energy}$
Crest factor	$T_8 = \max\lvert x(i)\rvert / \sqrt{(1/n)\sum_{i=1}^{n} x(i)^2}$
Impact factor	$T_9 = \max(\lvert x(i)\rvert) / \sqrt{(1/n)\sum_{i=1}^{n}\lvert x(i)\rvert}$
Shape factor	$T_{10} = \sqrt{(1/n)\sum_{i=1}^{n} x(i)^2} / (1/n)\sum_{i=1}^{n}\lvert x(i)\rvert$
Latitude factor	$T_{11} = \max(\lvert x(i)\rvert) / (1/n)\sum_{i=1}^{n}\lvert x(i)\rvert$

4.2 Feature Selection

The least squares method (LS method) was utilized to analyze 352 features in the initial feature set. The corresponding LS values of the features were computed and then ranked in descending order. Fault-sensitive features were then selected based on the ranking result.

4.3 Model Recognition

The sensitive features selected by the LS method were used as input vectors to train the DBN model. Subsequently, the trained DBN model was applied to conduct rolling bearing fault diagnosis. Figure 2 shows the flowchart of rolling bearing fault diagnosis based on Dual-Tree Complex Wavelet Packet Transform (DTCWPT) and Deep Belief Network (DBN).

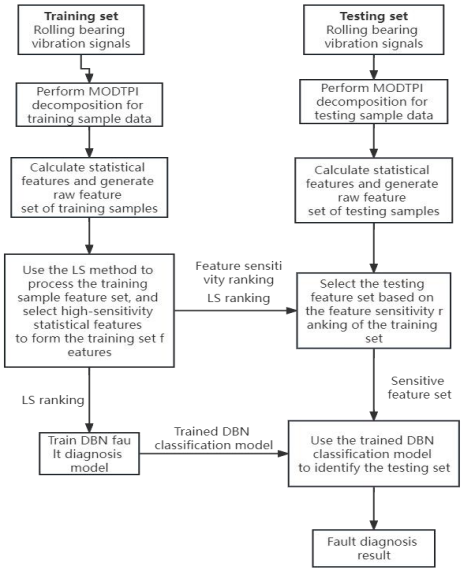


Figure 2. Flowchart of the DBN-based Intelligent Fault Diagnosis

5. Results and Analysis

The experimental data employed in this research were gathered from the rolling bearing fault experiments carried out by the Electrical Engineering Laboratory at Case Western Reserve University [3]. The bearing test setup is shown in Figure 3. A three-phase motor having a power of 1491.4 W is located on the left side. A dynamometer, which generates rated power, is located on the right side. The defective bearing, being a deep groove ball bearing (with the model of SKF-6205-2RS), is mounted at the motor drive terminal. Acceleration sensors, for collecting bearing vibration data, are installed above the motor drive end and fan end. To simulate the bearing fault conditions, single-point defects are created on the inner race, outer race, and roller by means of electrical discharge machining. The sampling frequency is configured at 12 kilohertz.

The bearing located at the drive end exhibits three kinds of faults, namely those in the roller, inner race, and outer race, with the fault sizes being 0.007, 0.014, 0.021, and 0.028 inches respectively. The motor operates under loads ranging between 0 hp and 3 hp. Table 2 shows 12 bearing states, comprising one normal state, four inner race fault states, four roller fault states, and three outer race fault states. For each bearing fault state, 2000 consecutive data points were collected as one sample, which resulted in a total of 60 vibration signal samples. In this research, four experimental scenarios (Case 1-4) were devised to assess the efficacy and flexibility of the proposed method within both constant and variable load circumstances. Cases 1-2 are form a group of comparative experiments, employing the data samples corresponding to a motor load of 2 hp as the training set and take the data samples with loads of 2 hp and 3 hp as the testing set. Cases 3 and 4 are also a group of comparative experiment, which uses the data samples under a motor load of 3 hp as the training set and those under loads of 3 hp and 2 hp as the testing set.

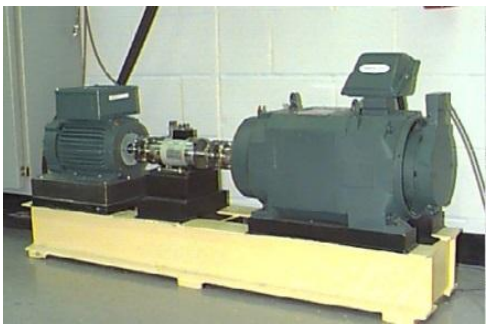


Figure 3. Setup for the CWRU Bearing Fault Test

Maximum discrete wavelet packet analysis was first performed for the vibration signal samples. 352 statistical features were extracted to form the initial feature set space. Since different statistical features had varying sensitivities to fault types, The Least Squares (LS) method was applied in the feature

selection process to measure the sensitivity of the statistical features with respect to the fault types.

Table 2. Experimental Data

Bearing state	fault size (inch)	Case 1, Case 2			Case 3, Case 4			Type
		training set	testing set Case 1	testing set Case 2	training set	testing set Case 3	testing set Case 4	
		2 hp		3 hp	3 hp		2 hp	
Health condition	0	20	40	40	20	40	40	1
Roller fault	0.007	20	40	40	20	40	40	2
	0.014	20	40	40	20	40	40	3
	0.021	20	40	40	20	40	40	4
	0.028	20	40	40	20	40	40	5
Inner race fault	0.007	20	40	40	20	40	40	6
	0.014	20	40	40	20	40	40	7
	0.021	20	40	40	20	40	40	8
	0.028	20	40	40	20	40	40	9
Outer race fault	0.007	20	40	40	20	40	40	10
	0.014	20	40	40	20	40	40	11
	0.021	20	40	40	20	40	40	12
Number of samples		240	480	480	240	480	480	

5.1 LS-free Fault Sensitivity Analysis

In order to verify the effectiveness and suitability of the proposed approach, two sets of comparative experiments were carried out.

n the first experimental group, the Least Squares (LS) method was not incorporated. Instead, experimental analyses were conducted on the OFS-SVM, OFS-PCA-SVM, and OFS-DBN models. The Original Feature Set (OFS) comprises 352 initial statistical features. The OFS-SVM represents an SVM-centered fault analysis model, wherein the OFS is directly input into the SVM for the purposes of training and fault recognition. Its fault recognition accuracy is shown in Table 3. It can be observed that Cases 1 and 3 had significantly higher recognition accuracy compared with Cases 2 and 4. In the OFS-PCA-SVM model, OFS having a dimension reduced by PCA, was used as the input of SVM. The recognition accuracies in the four cases under varying numbers of Principal Component Analysis (PCA) are presented in Table 4. When the number of PCA increases, the recognition accuracy increases in all the cases. However, the peak recognition accuracy of the OFS-PCA-SVM model in Cases 2 and 4 were 77.71% and 74.38%, respectively. OFS-PCA-SVM showed a slight improvement compared with OFS-SVM. However, this improvement was not significant. For the OFS-DBN model, the quantity of input features is 352.the numbers of output nodes at the first two layers of RMB are respectively 100 and 50, and the number of types at the final output layer is 12. The fault recognition accuracies of the OFS-DBN model are shown in Table 5. It is evident that the recognition accuracy has been remarkably enhanced in all cases. In addition, the recognition accuracies in Cases 2 and 4 reached 98.33% and 98.75%, respectively. The achieved results illustrated that DBN is endowed with highly proficient feature analysis and fault recognition capabilities.

Table 3. Fault Recognition Accuracy (%) Of The OFS-SVM Model in Cases 1-4

Case 1	Case 2	Case 3	Case 4
92.92	75.83	92.08	74.38

Table 4. Fault Recognition Accuracy (%) of the OFS-PCA-SVM Model in Cases 1-4

Number of PCAs *n*	Case 1	Case 2	Case 3	Case 4
5	80.83	66.46	85.42	66.46
10	85.00	68.96	87.71	69.38
15	89.17	69.17	91.67	72.50
20	93.33	74.58	93.54	73.33
25	93.96	77.71	94.17	74.38
30	91.84	73.15	93.87	72.68
35	89.38	71.59	93.15	71.73
40	88.12	68.76	91.57	69.89

Table 5. Fault Recognition Accuracy (%) of the OFS-DBN Model in Cases 1-4

Case 1	Case 2	Case 3	Case 4
99.38	98.33	100.00	98.75

5.2 LS-based Fault Sensitivity Analysis

In the second experimental group, a feature selection approach grounded on the Least Squares (LS) was incorporated. and experimental investigations were conducted on the OFS-LS-SVM, LS-PCA-SVM, and OFS-LS-DBN models. In the OFS-LS-SVM model, the LS method was utilized to extract fault-sensitive features from the Original Feature Set (OFS). The selected features were subsequently employed to train the Support Vector Machine (SVM), and the data from the testing set were utilized to validate the efficacy of the OFS-LS-SVM model. The recognition accuracies of the OFS-LS-SVM model are presented in Table 6, wherein "sfn" denotes the quantity of features selected via the LS method. Regarding the OFS-LS-SVM model, the peak fault recognition accuracies for Cases 2 and 4 were 99.79% and 99.79% respectively. When compared with the results shown in Table 3, it is evident that the recognition accuracies have been substantially enhanced. Figure 4 illustrates the variation of the recognition accuracy of the OFS-LS-SVM in relation to the number of features selected by the LS method (sfn). It is observable that, when sfn is approximately 100, the accuracies in the four cases exhibit a significant increase. The recognition accuracies of the LS-PCA-SVM model are presented in Table 7. Figure 5 depicts the variation of the recognition accuracy in relation to sfn. It can be noted that, as the number of Principal Component Analysis (PCA) components increases, the recognition accuracy is marginally enhanced. Nevertheless, the final outcome is comparable to that of the OFS-LS-SVM model. For the OFS-LS-DBN model, with the exception that the number of Input - layer units was modified to sfn, all other parameters remained analogous to those in the first experimental group. The recognition accuracies of this model are exhibited in Table 8. It can be discerned that the maximum recognition accuracies in Cases 1 - 4 attain 100%, 99.79%, 100%, and 99.79% respectively. Subsequent to the introduction of LS, the fault recognition accuracies for all cases are conspicuously enhanced, as demonstrated in Table 5. Figure 6 portrays the variation of the the precision rate of fault identification of the OFS-LS-DBN model with respect to sfn.

Table 6. Fault Recognition Accuracy (%) of the OFS-LS-SVM Model in Cases 1-4

sfn	Case 1	Case 2	Case 3	Case 4
10	86.67	81.04	83.96	69.38
40	81.04	65.47	81.04	81.88
70	93.54	95.83	96.04	90.00
100	99.79	99.79	100.00	99.79
130	99.79	93.96	100.00	85.42
160	98.96	97.29	99.79	90.21
190	98.96	77.92	100.00	89.17
220	97.92	86.04	98.96	91.04
250	98.96	93.75	99.17	96.67
280	99.58	92.50	99.17	98.12
310	98.33	88.75	98.12	78.33

Table 7. Fault recognition accuracy (%) of the LS-PCA-SVM model in Cases 1-4 with a number of PCA of 20

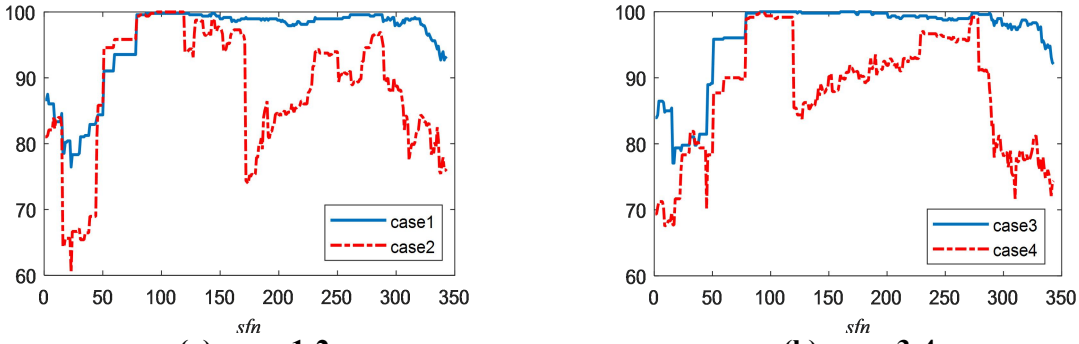
sfn	Case 1	Case 2	Case 3	Case 4
20	91.88	84.17	92.50	83.75
40	92.50	90.00	85.83	86.88
70	99.58	98.96	99.58	88.33
100	100.00	99.17	99.79	99.38
130	100.00	88.96	99.79	93.33
160	100.00	74.79	100.00	68.54
190	99.58	85.63	99.17	72.50
220	98.54	89.79	98.13	84.38
250	98.13	86.88	99.38	92.08
280	98.75	87.71	99.58	92.08
310	99.17	87.08	99.58	90.00

Table 8. Fault Recognition Accuracy (%) of the OFS-LS-DBN Model in Cases 1-4

sfn	Case 1	Case 2	Case 3	Case 4
10	8.33	8.33	8.33	8.33
40	100.00	77.08	99.79	98.75
70	100.00	79.38	100.00	99.58
100	99.79	92.92	100.00	99.79
130	99.79	98.75	100.00	98.96
160	99.17	99.17	100.00	99.38
190	98.58	96.46	100.00	99.79
220	99.58	97.71	100.00	99.58
250	99.79	97.50	100.00	99.38
280	99.58	99.38	100.00	98.96
310	99.58	99.79	100.00	98.13

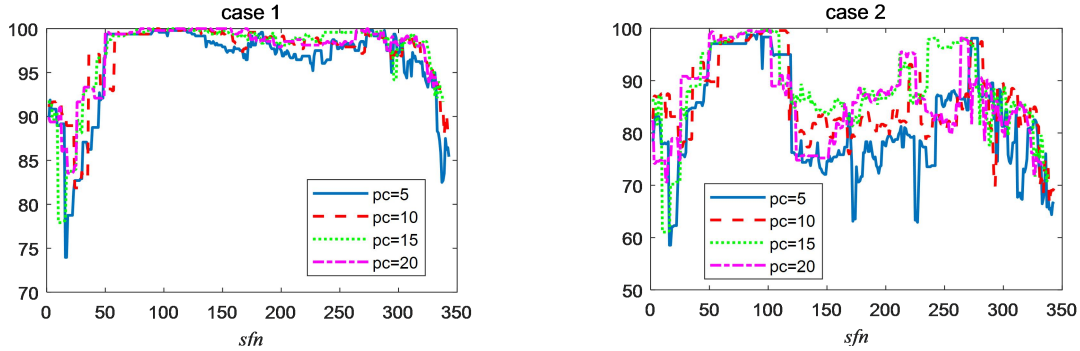
Figure 7 presents a comparative analysis of the variations in the recognition accuracy of the OFS-LS-SVM, LS-PCA-SVM, and OFS-LS-DBN fault diagnosis models with respect to sfn, under the condition where the number of Principal Component Analysis (PCA) components is set at 30. It can be discerned

that the OFS-LS-DBN model demonstrates more prominent performance with respect to fault recognition accuracy in contrast to the other two models and exhibits the lowest sensitivity to the variation of sfn.



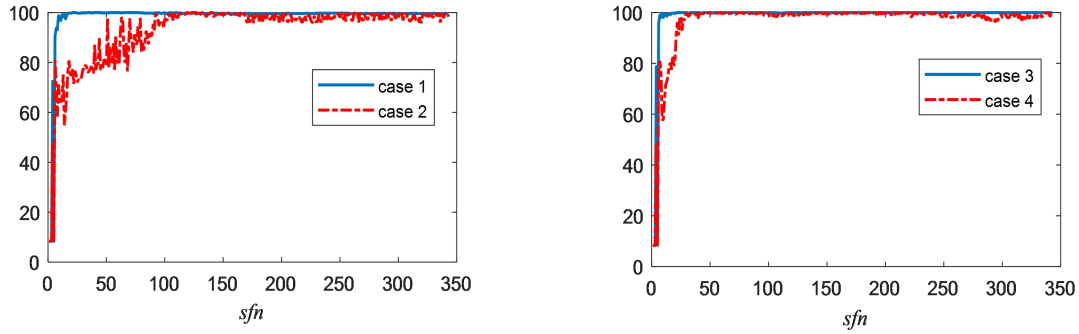
(a) cases 1-2 **(b) cases 3-4**

Figure 4. Fault Diagnosis Accuracy (%) of the OFS-LS-SVM Model with Respect to Sfn in Cases 1-4



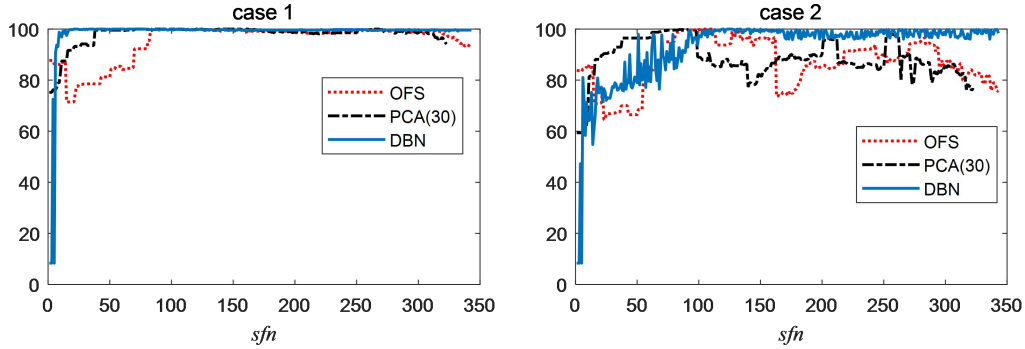
(a) cases 1 **(b) cases 2**

Figure 5. Fault Diagnosis Accuracy (%) of the LS-PCA-SVM Model with Respect to sfn in Cases 1-4



(a) cases 1-2 **(b) cases 3-4**

Figure 6. Fault Diagnosis Accuracy (%) of the OFS-LS-DBN Model with Respect to sfn in Cases 1-4



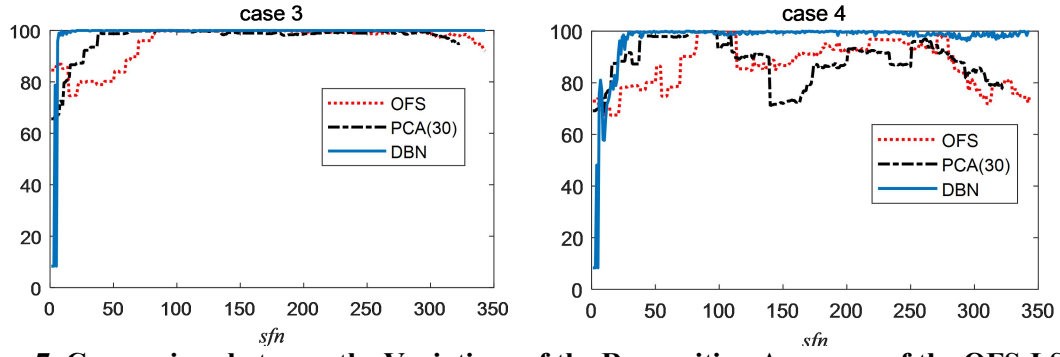


Figure 7. Comparison between the Variations of the Recognition Accuracy of the OFS-LS-SVM, LS-PCA-SVM, and OFS-LS-DBN Fault Diagnosis Models with Respect to SFN

6. Conclusion

This paper proposes the LS method, which represents a fault-sensitive feature extraction technique predicated on the maximum discrete wavelet packet analysis of vibration signals. A DBN (Deep Belief Network)-oriented intelligent diagnosis model was additionally constructed for rolling bearings. The ensuing conclusions can be derived:

(1) The statistical features obtained after the time-frequency analysis of vibration signals had varying impacts on the rolling bearing fault recognition. The LS (Least Squares) method can be employed to gauge the sensitivity of these features with respect to fault types. In addition, the features that are more favorable for fault recognition can be filtered out by ranking the feature sensitivity;

(2) Compared with the SVM classifiers and PCA dimensionality reduction methods, the DBN-based diagnosis model can effectively enhance the feature space analysis and fault recognition capacity;

(3) The fault diagnosis model which combines DTCWPT, LS, and DBN can significantly improve the fault recognition accuracy, and it has high adaptability for bearing fault diagnosis under varying load conditions.

Acknowledgement

This paper is supported by the R&D Program of Beijing Municipal Education Commission (KM202211417006、 KM202211417005) and the Academic Research Projects of Beijing Union University (ZK90202105, ZK90202106 and ZK20202204).

References

[1] Shao H, Jiang H, Wang F, et al. Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet. Isa Transactions, 2017: 187-201.

[2] Li H, Xiao D. Survey on data driven fault diagnosis methods. Control and Decision, 2011, (01): 16-25.

[3] Shiyuan Liu, Xu Qian, Hong Wan, Zongbin Ye, Shoupeng Wu, and Xiaohong Ren. NPC Three-Level Inverter Open-Circuit Fault Diagnosis Based on Adaptive Electrical Period Partition and Random Forest. Journal of Sensors. 2020. 1-18.

[4] Gao Z, Cecati C, Ding S X. A Survey of Fault Diagnosis and Fault-Tolerant Techniques Part I: Fault Diagnosis with Model-Based and Signal-Based Approaches. IEEE Transactions on Industrial Electronics, 2015, 62(6): 3757-3767.

[5] Gao Z, Cecati C, Ding S X. A Survey of Fault Diagnosis and Fault-Tolerant Techniques-Part II: Fault Diagnosis with Knowledge-Based and Hybrid/Active Approaches. IEEE Transactions on Industrial Electronics, 2015, 62(6): 3768-3774.

[6] Liu C, Wu Y, Zhen C. Rolling Bearing Fault Diagnosis Based on Variational Mode Decomposition and Fuzzy C Means Clustering. Proceedings of the CSEE, 2015, (13): 3358-3365.

[7] Yu X, Dig E, Chen C, et al. Rolling bearing fault recognition method based on HHT and Supervised Sparse Coding. Journal of China Coal Society, 2015, (11): 2587-2595.

[8] Hu Guangshu. Modern signal processing tutorial. Tsinghua University Press.2015

[9] Shi M, Luo R, Fu Y. Fault diagnosis of rotating machinery based on wavelet and energy feature extraction. Journal of

Electronic Measurement and Instrumentation, 2015, (08): 1114-1120.

[10] Jaber A, Bicker R. Industrial Robot Backlash Fault Diagnosis Based on Discrete Wavelet Transform and Artificial Neural Network. American Journal of Mechanical Engineering, 2016, 4(1): 21-31.

[11] Ou L, Yu D. Rolling Bearing Fault Diagnosis Based on Supervised Laplacian Score and Principal Component Analysis. Journal of Mechanical Engineering, 2014, (05): 88-94.

