Article

Fault Diagnosis of Loader Gearbox Based on an ICA and SVM Algorithm

Zhongxin Chen [1,2], Feng Zhao [1,2], Jun Zhou [1,2,*], Panling Huang [1,2,*] and Xutao Zhang [1,2]

[1] School of Mechanical Engineering, Shandong University, Jinan 250061, China;
201920453@mail.sdu.edu.cn (Z.C.); zhaofeng@mail.sdu.edu.cn (F.Z.); zhangxutao@mail.sdu.edu.cn (X.Z.)

[2] Key Laboratory of High Efficiency and Clean Mechanical Manufacture, Ministry of Education,
Jinan 250061, China

* Correspondence: zhoujun@sdu.edu.cn (J.Z.); hfpl@sdu.edu.cn (P.H.)

check for
updates

Abstract: When a part of the loader's gearbox fails, this can lead to equipment failure due to the complex internal structure and the interrelationship between the parts. Therefore, it is imperative to research an efficient strategy for transmission fault diagnosis. In this study, the non-contact characteristics of noise diagnosis using sound intensity probes were used to collect noise signals generated under gear breaking conditions. The independent component analysis (ICA) technique was applied for feature extraction from the original data and to reduce the correlation between the signals. The correlation coefficient between the independent components and the source data was used as the input parameters of the support vector machine (SVM) classifier. The separation of the independent components was achieved by MATLAB simulation. The misdiagnosis rate was 5% for 40 sets of test data. A 13-point test platform for noise testing of the loader gearbox was built according to Chinese national standards. Source signals under the normal and fault conditions were analyzed separately by ICA and SVM algorithms. In this case, the misdiagnosis rate was 7.5% for the 40 sets of experimental test data. This proved that the proposed method could effectively realize fault classification and recognition.

Keywords: independent component analysis; support vector machines; feature extraction; fault diagnosis; gearbox

1. Introduction

As the most widely used type of engineering machinery, a loader is commonly utilized in construction sites such as roads, railways, and ports. Due to poor working conditions, it is particularly important to detect the failure of this construction machinery, especially through online detection. The gearbox is an important transmission component in loaders. Statistics show that the gear is the most common part causing failure: approximately sixty percent of failures are caused by the gear [1]. At present, the diagnosis methods for gearbox faults include temperature detection, optical fiber sensing, and vibration detection. These methods require that the position of the measuring instrument be changed when examining different fault locations at the same or different gearboxes. This is time-consuming and inefficient. Therefore, it is particularly important to propose a reasonable method for fault diagnosis.

Independent component analysis (ICA) is a new signal analysis technology developed in recent decades from the blind signal separation problem. It has been successfully applied in fields such as bioengineering [2], communication [3], speech recognition [4], and fault diagnosis [5]. In the field of neuroimaging, Melissant et al. used the ICA algorithm to preprocess the EEGs of patients with Alzheimer's disease, which improved the accuracy of diagnosis [6]. Gohel effectively repressed signal

leakage by using the orthogonality of the ICA to reconstruct the source signal [7]. Back was the first to introduce ICA to the field of financial analysis, and many scholars have since used ICA to analyze and process financial data [8]. Independent component analysis was applied to the financial time series data belonging to a retail chain, where it detected a few factors that affect the cash flow [9]. Independent component analysis has also been used in machine condition monitoring and fault diagnosis. For example, an approach involving dynamic independent component analysis in fault detection and diagnosis, proposed by George, is able to accurately detect and isolate the root causes for each individual fault [10]. Principal component analysis (PCA) and ICA techniques were applied to the diagnosis of turbines for fault detection, and the validity and effectiveness of the approach were verified [11]. The French scholar Badaoui used ICA in combination with a Wiener filter to separate combustion noise and piston slap in a diesel engine [12]. Many improved algorithms such as constrained independent component analysis (cICA) [13,14] and kernel independent component analysis (KICA) [15,16] have been developed in order to improve the practical range and precision of the ICA algorithm.

The application of an intelligent system for fault diagnosis has been widely researched in recent years. Support vector machines (SVMs) and neural networks have been employed to solve classification problems [17]. In the field of economics, Jae identified the optimal parameter values of the kernel function of the SVM to predict bankruptcy and obtained better performance than multiple discriminant analysis (MDA), Logit, and back-propagation neural networks (BPNs) [18]. Linear SVM classifiers were introduced for credit risk evaluation [19]. In signal processing, Salvati used an SVM to locate near-field sound sources and improved the accuracy of the beam former in a reverberation environment [20]. Haik built an optimized SVM-based probability model to select a window from a continuous data, and the effectiveness of the approach was proven [21]. Many scholars regard SVM as an intelligent classifier for machine fault classification, especially bearing faults [22,23], gear faults [24,25], wind turbine faults [26,27], and induction motors [28,29]. When applied in the current study, SVM has high computational complexity and low precision when directly classifying using raw data with redundant characteristics. Therefore, the ICA technique was applied for feature extraction from the original data and to reduce the correlation between the signals. The correlation coefficients between the feature vector and the source data were then used as the input parameters of the SVM classifier.

2. Fault Diagnosis Algorithm

2.1. Independent Component Analysis

Independent component analysis is a technique that has been demonstrated to be capable of separating an independent component from the source signal by utilizing its independence and non-Gaussian property. The process used by ICA is illustrated in Figure 1.
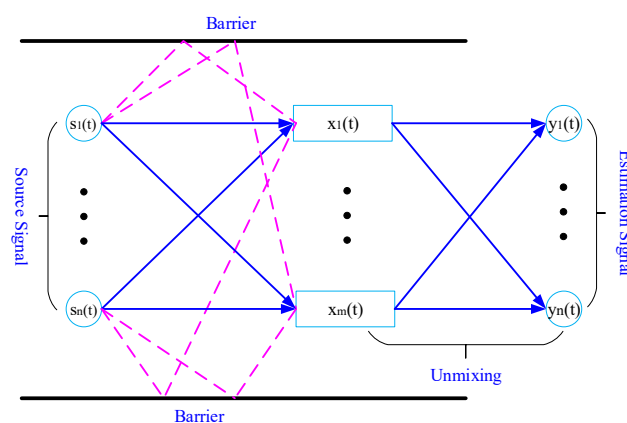

Figure 1. Linear convolution mixed model.

In Figure 1, the source signal is composed of a few independent acoustic sources, $s_n(t)$. Considering the effect of path delay in the signal transmission process, each signal can take multiple channels to reach the receivers, $x_m(t)$. The process of unmixing the signals is performed by the ICA algorithm. Assuming that n known independent source signals were transmitted by an unknown channel and eventually accepted by m receivers, there are n independent components $y_n(t)$, which constitute the estimates of the source signal. The received signal can be expressed as [30]

$$x_i(t) = \sum_{j=1}^{n} \sum_{p=0}^{P} b_{ij}(p)s_j(t-p) \tag{1}$$

where $x_i(t)$ is the mixed signal received by receiver i, b_{ij} is the response function between the signal from the ith receiver and the jth source signal, and $s_j(t-p)$ is the delay function from the jth source signal to the ith receiver. Its matrix form is

$$X(t) = B(t) * S(t) \tag{2}$$

where $B(t)$ is a mixed matrix in the time domain.

Equation (2) is subjected to Fourier transform:

$$X(f) = B(f)S(f) \tag{3}$$

where $X(f) = [x_1(f), x_2(f), L, x_m(f)]^T$ is the observed signal matrix that has been transformed in the frequency domain, $b(f)$ is the $m \times n$ hybrid matrix, and $S(f) = [s_1(f), s_2(f), L, s_n(f)]^T$ is the source signal matrix. In the processing of the mixed solution, a linear matrix W is determined to obtain the estimated signal from the source signal. Therefore, the separation of the ICA is given by

$$Y(f) = WX(f) = WAS(f) \tag{4}$$

where $Y(f)$ is the estimation of $S(f)$ in the frequency domain, and the inverse Fourier transform is used to transform the frequency to the time domain, A is an unknown mixing matrix.

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}_{m \times n} \tag{5}$$

2.2. Support Vector Machines

Given the set of training samples $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ $y_i \in \{-1, 1\}$, the most basic objective of the SVM is finding a division hyperplane from the sample space according to the training samples D to separate data that have different characteristics. The following linear equation describes the hyperplane function [31]

$$w^T x + b = 0 \tag{6}$$

where $w = (w_1, w_2, \ldots, w_d)$ is the normal vector that determines the direction of the hyperplane, and b is a constant term that decides the distance between the hyperplane and the coordinate origin. Obviously, the divided hyperplane is constrained by w and b. The distance from point x belonging to the sample space to the plane (w,b) can be written as

$$r = \frac{\left| w^T x + b \right|}{\|w\|}. \tag{7}$$

Int. J. Environ. Res. Public Health **2019**, *16*, 4868

4 of 20

Suppose the plane (w,b) can classify the training samples correctly and has the following constraints for $(x_i, y_i) \in D$.

$$\begin{cases} w^T x_i + b \geq +1, y_i = +1 \\ w^T x_i + b \leq +1, y_i = -1 \end{cases} \tag{8}$$

As shown in Figure 2, the training sample point that is nearest to the hyperplane and gives a true statement when substituted into Equations (2)–(7) is called the "support vector." The sum of the distances of the two types of heterogeneous support vectors to the hyperplane is

$$r = \frac{2}{\|w\|}. \tag{9}$$
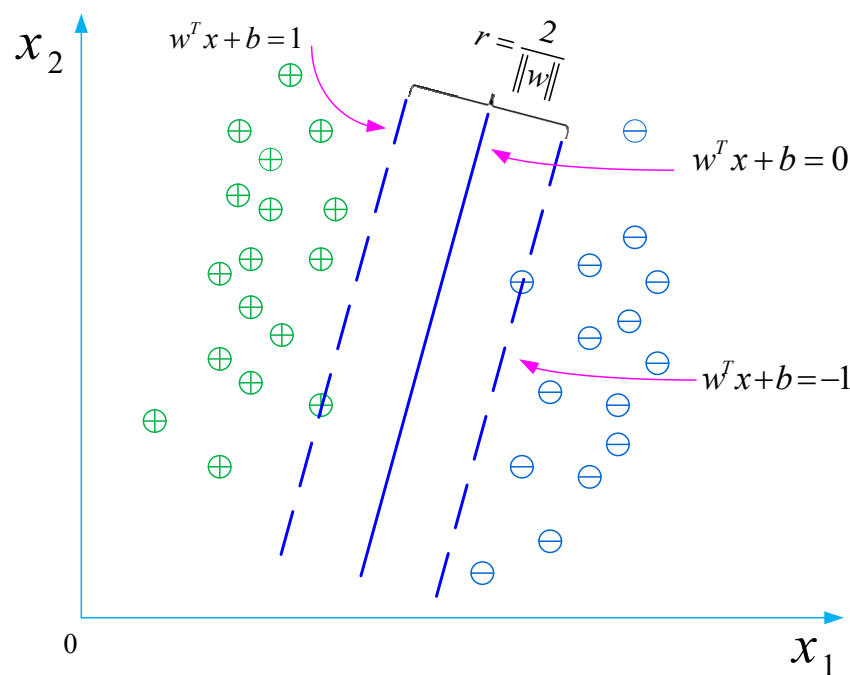
Figure 2. Principle of the support vector machine (SVM) classifier.

The plane of the "maximum margin" is equivalent to finding the parameter w and b, which must satisfy Equation (7).

$$\max_{w,b} \frac{2}{\|w\|} s.t. y_i(w^T x_i + b) \geq 1, \ i = 1,2,3,\ldots,m \tag{10}$$

Obviously, $\|w\|^{-1}$ should take the maximum value in order to maximize the interval. Therefore, Equations (2)–(9) is rewritten as

$$\min_{w,b} \frac{1}{2}\|w\|^2. \tag{11}$$

3. Algorithm Simulation

To test the classification and recognition of the fault signal through the combined algorithms of ICA and SVM, MATLAB software was used to simulate three types of vibration signals, namely, the signal from the constant meshing gear $S_1(t)$, the output gear pair signal $S_2(t)$, and the signal from a broken tooth $S_{cp}(t)$. In order to simplify the simulation process, the other vibration sources in the gearbox were not considered when simulating the gearbox conditions. There were two different working conditions, which were convolutions of the three kinds of signals. Two hundred and forty sets of data were collected separately under different conditions. Two hundred groups of data were applied to train the classifier, and forty groups of data were used to test the accuracy of the classifier.

The noise signals were replaced by vibration signals due to difficulty in the simulation of the noise signals and the correlation between noise and vibration. The description of the simulated signals is shown in Table 1.

Table 1. Description of simulation signals.

Signals	Expression
Normal	$s_1(t) * s_2(t)$
Fault	$s_1(t) * s_2(t) * s_{cp}(t)$

The parameters of the simulation are consistent with the gearbox parameters to make the simulation closer to the actual situation. In this study, the gearbox of a loader from Shandong Lingong Construction Machinery Co., Ltd., that had three gears, was used, as shown in Figure 3.
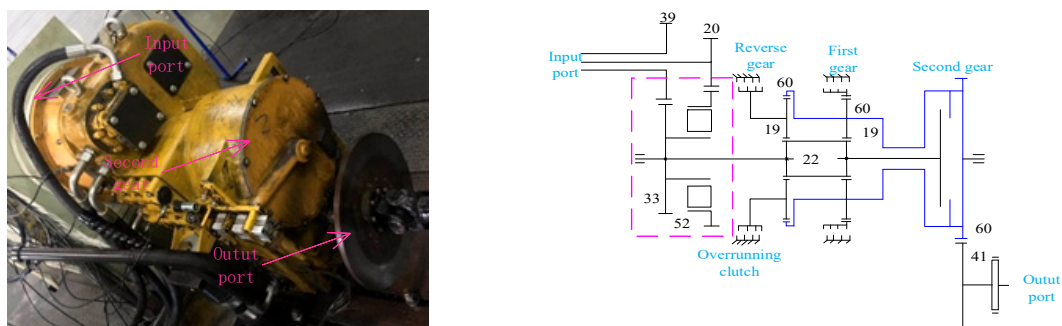
Figure 3. Transmission principle of loader gearbox.

Aiming at the noise signal of gearbox failure, this paper designs a diagnostic algorithm based on ICA and SVM. The FastICA principle is used to separate the independent sources in the signal, and the correlation coefficients of the noise signal and the independent source are used to construct the eigenvector of the noise signal. The algorithm flow is shown in Figure 4. Gaussian kernel is used. The function SVM model trains the classifier to diagnose the fault of the gearbox. The algorithm flow is shown in Figure 5.
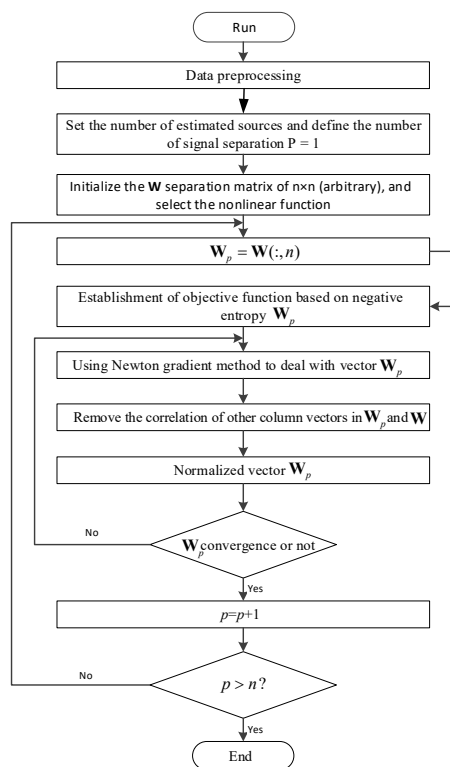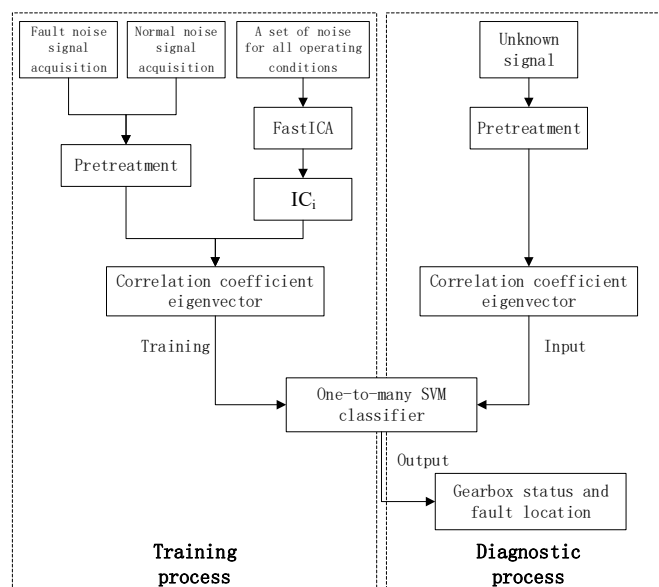
Figure 4. FastICA algorithm flow.

Figure 5. Independent component analysis (ICA)-SVM-based diagnostic algorithm flow.

3.1. Simulation of the Original Signal

The second gear was selected, and the output speed of the torque converter was 1200 rpm in this working condition. The influence of the vibration signal of the torque converter was ignored.

In the simulation, the signal of the constant meshing gear $s_1(t)$ is as follows:

$$s_1(t) = \sum_{i=1}^{n} A_i \cos(2\pi i f_{m1} t + \varphi_i) \tag{12}$$

Int. J. Environ. Res. Public Health **2019**, *16*, 4868

7 of 20

where A_i and φ_i are the amplitude and phases of the *i*th order vibration of the constant meshing gear pair, respectively, and f_{m1} is the gear meshing frequency.

In this paper, the frequency of the input shaft is $f_{n1} = 20Hz$. The torque converter output speed is 1200 rpm. The gear meshing frequency is:

$$f_{m1} = 20 \times f_{n1} = 400Hz. \tag{13}$$

Assuming the second-order vibration response of the gear pair is known, $A_1 = 0.5$ um, $A_2 = 1$ um, $\varphi_1 = \varphi_2 = 0$. Then,

$$s_1(t) = 0.5\cos(800pt) + \cos(1600pt). \tag{14}$$

The output gear pair signal $s_2(t)$ is as follows:

$$s_2(t) = \sum_{i=1}^{n} B_i \cos(2\pi i f_{m2} t + \theta_i) \tag{15}$$

where B_i and θ_i are the amplitude and phases of the *i*th order vibration of the output gear pair, respectively, and f_{m2} is the frequency of the output gear pair in the second gear condition.

Output shaft frequency f_{n2} is:

$$f_{n3} = f_{n1} \times \frac{20}{52} \times \frac{60}{41} = 11.3Hz. \tag{16}$$

The frequency of the output gear pair f_{m2} is:

$$f_{m2} = 41 \times f_{n1} \times \frac{20}{52} \times \frac{60}{41} = 462.5Hz. \tag{17}$$

In addition, the amplitudes $B_1 = 0.7$ um, $B_2 = 1.4$ um, and phases $\theta_1 = \theta_2 = 0$.

$$s_2(t) = 0.7\cos(925\pi t) + 1.4\cos(1850\pi t) \tag{18}$$

The faulty gear was selected from the driving wheel and the broken tooth fault was applied. A broken tooth in the gearbox will generate a local impact phenomenon, and the vibration signals show the impact given by Equations (3)–(5), where $s_{cp}(t)$ is defined as

$$s_{cp}(t) = \sum_{i=1}^{n} \sum_{j=1}^{m} A_{cji} exp\left[\frac{-2\pi\varepsilon_{dj}}{\sqrt{1 - \varepsilon_{dj}^2}} f_{dj}(t - \tau_{ci})\right] \cos\left[2\pi f_{dj}(t - \tau_{ci})\right] \tag{19}$$

where *n* is the number of shocks in the signal, *m* is the order of the natural frequency, and A_{cji} is the amplitude of the *i*th impact when the machine has the *j*th order of natural frequency. To simplify the simulation process, A_{cji} was assigned a random number from 0 to 10. The variables ε_{dj} and f_{dj} represent the *j*th order damping ratio and natural frequency, respectively; τ_{ci} is the moment of the *i*th impact, which is given by $\tau_{ci} = (i-1)T_c + \tau_{c0}$; T_c is the cycle of the gear failure; and τ_{c0} is the initial time of the impact response, i.e., $\tau_{c0} = 0.3T_c$. It was assumed that the two natural frequencies were 1100 and 3500 Hz. According to the actual working condition of the gearbox, $\varepsilon_{d1} = 0.04$, $\varepsilon_{d2} = 0.03$, $A_{c1} = A_{c2} \in (0, 10)$, $\tau_{c1} = \tau_{c0} = 0.3T_c = 0.2 \times 1/f_{n3} = 0.027$, and $A_{c1} = A_{c2} = 4$ um. The time and frequency diagrams of the three source signals are plotted in Figure 6.
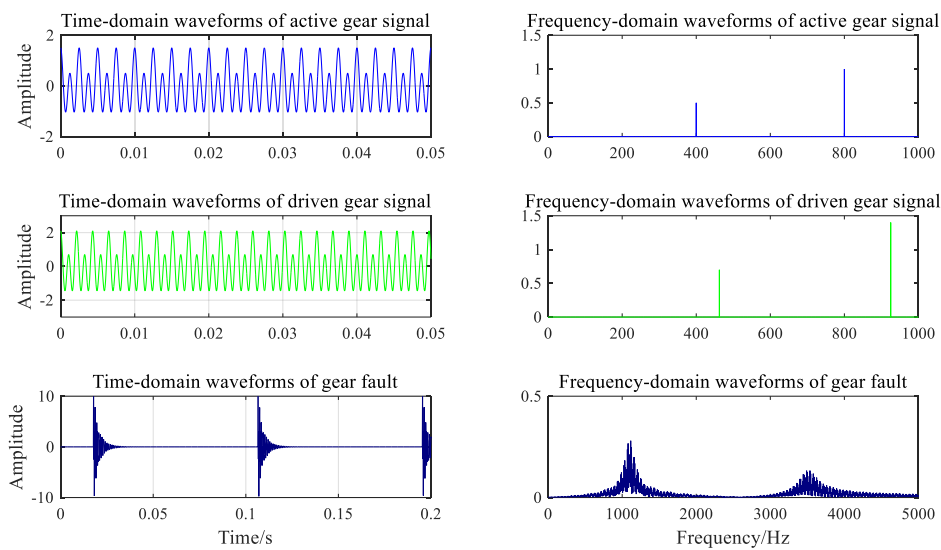
Figure 6. Time and frequency diagrams of three signals.

3.2. Simulation of Working Condition Signals

Six data collectors were used in the simulation. Suppose that each source signal from two different channels reach the same collector. Therefore, the data from each collector can be expressed as

$$x(t) = [x_1(t); \cdots; x_i(t); \cdots; x_6(t)](1 < i < 6) \tag{20}$$

where $x_i(t)$ is the data collected by the ith collector. The calculation method is as follows:

$$x(t) = b(i,:) \begin{bmatrix} s_{1-1}((st_i(1)+1):(N+st_i(1))) \\ s_{1-2}((st_i(2)+1):(N+st_i(2))) \\ s_{2-1}((st_i(3)+1):(N+st_i(3))) \\ s_{2-2}((st_i(4)+1):(N+st_i(4))) \\ s_{3-1}((st_i(5)+1):(N+st_i(5))) \\ s_{3-2}((st_i(6)+1):(N+st_i(6))) \end{bmatrix} \tag{21}$$

where b is an arbitrary 6×6 mixed matrix, and N is the number of FFT calculations. A 0.2-s signal was collected at each collector to show the difference between the data collected by each collector. The vibration signal of the gearbox under the normal condition is shown in Figure 7. The fault signal corresponding to the broken gear from six data collectors is shown in Figure 8.
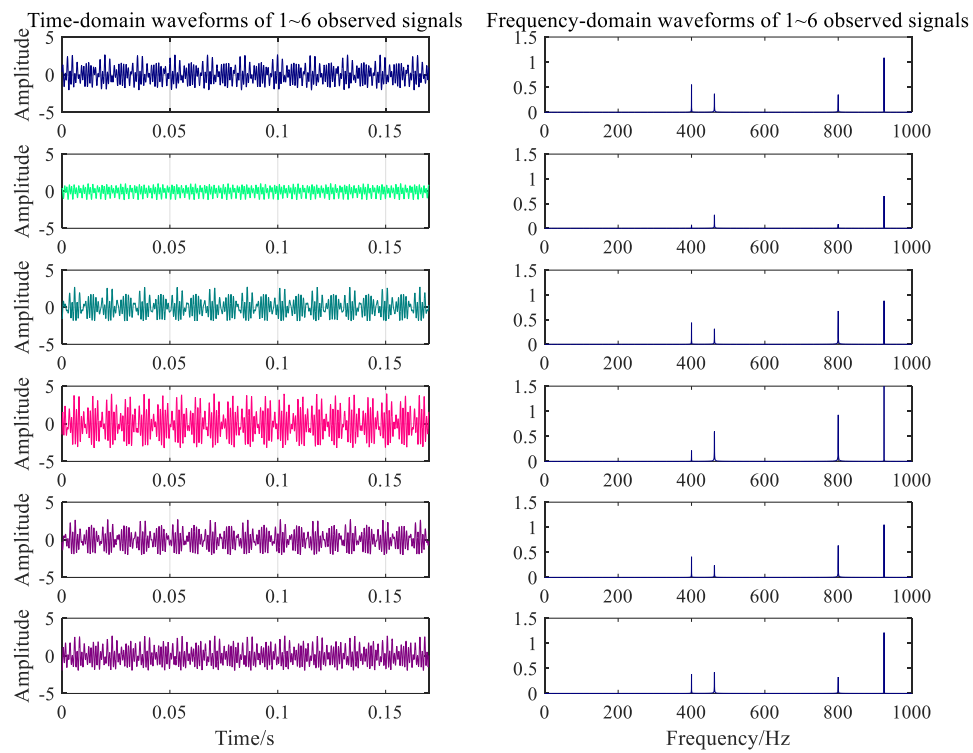
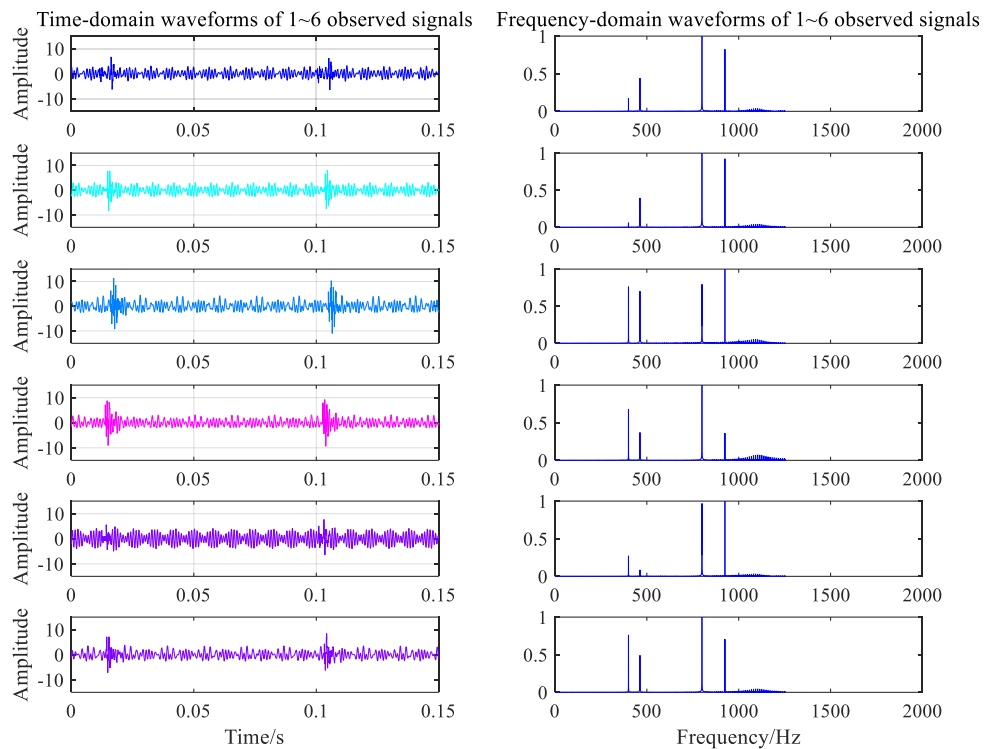**Figure 7.** Time-frequency waveforms of signals under normal conditions.

Figure 8. Time-frequency waveforms of signals under fault condition.

3.3. Feature Extraction

Due to the source signal emitted under the fault condition being generated by the convolution of three independent source signals in the simulation process, three independent components were selected to estimate the source signal. The FastICA algorithm was applied to separate the data for

different conditions. First, the independent matrix was calculated. Second, the results of the signal were estimated by the matrix in the frequency domain, and the amplitude of the signal was corrected. Finally, the independent source was obtained through the iterative feedback tuning (IFT) criterion.

The results separated by FastICA are shown in Figure 9, and the three separated independent source signals have obvious periodic characteristics. By comparing the independent source signals in Figure 9 and the source signals in Figure 6, it can be seen that the separated signals have a certain degree of reduction of the normally meshing gear pair signal $s_1(t)$ and the output gear pair signal $s_2(t)$. Numerous differences in the sequence, amplitude, and shape can be found when comparing the results of the separation of the source signals. This can be attributed to the correlations between the signals used for the convolution. However, the independent components, once separated by FastICA, have obvious periodic characteristics and similarities when compared with the mixed source signals. When the fault signal was decomposed, the first signal contained the basic features of the fault signal components. The signals separated by FastICA are different and can be used to characterize the mixed signals under any condition.
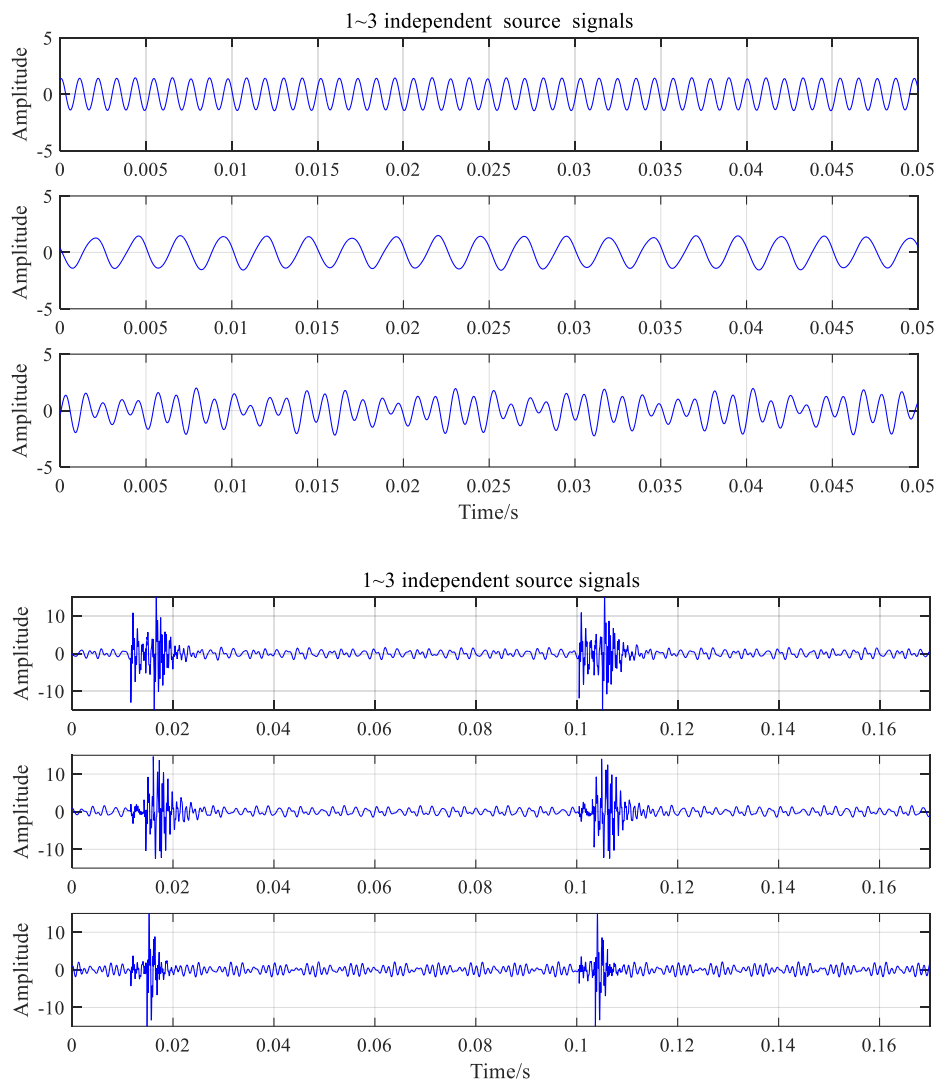

Figure 9. Separation results by FastICA.

3.4. Fault Identification

To improve the training precision, the correlation coefficients between the independent components and source data were used as the input parameters of the SVM classifier. The NIC*n* is the *nth*

independent component extracted under normal conditions(n = 1,2,3), and the FICn is the *nth* independent component extracted under fault conditions (n = 1,2,3). The procedure for the calculation of the correlation coefficient can be summarized as follows:

Step 1: Represent the different conditions by applying S_1, S_2, S_3 according to the classification of the conditions. The FastICA algorithm is used to extract m independent components, which are recorded as IC_i.

Step 2: Calculate the correlation coefficient between the first independent component of IC_i and each data channel of the signal, and take the absolute maximum. The first feature value F_1 is the sum of the *3* numbers.

Step 3: Repeat the second step to obtain the feature value F_2, F_3. The feature vectors are column vectors from F_1 to F_3.

Step 4: Repeat the second and third steps for the training signals, and obtain the feature vector of all signals.

The correlation coefficients of the NIC, FIC, and gearbox data collected under different conditions are shown in Table 2. It is worth noting that for n different conditions, only repeat step 2 and step 3 to get the eigenvectors of all signals.

Table 2. Correlation coefficients of the independent component and measured source signal.

Working Condition	NIC1	NIC2	NIC3	FIC1	FIC2	FIC3
Normal1	0.5464	0.8076	−0.7904	0.1378	−0.2303	−0.4007
Normal2	0.5008	0.5905	−0.8645	0.3234	−0.3088	−0.2871
Fault1	0.4303	0.3140	−0.5269	−0.5054	0.7256	0.2977
Fault2	0.3467	−0.3878	−0.4345	−0.5601	−0.5010	−0.4469

The feature vector of each group is the sum of the correlation coefficients. Table 3 shows the feature vector under the two normal and two fault working conditions. It can be seen from the table that the data collected under normal working conditions are highly correlated with the characteristic signal extracted from normal signals by the FastICA algorithm, and the data collected under the faulty working conditions are highly correlated with the characteristic signal extracted from fault signals. This clear correlation indirectly proves that the FastICA algorithm successfully extracts the characteristic components of different signals.

Table 3. Feature vector of each group.

Normal1	Normal2	Fault1	Fault2
$\begin{pmatrix} 2.1444 \\ 0.7688 \end{pmatrix}$	$\begin{pmatrix} 1.9558 \\ 0.9193 \end{pmatrix}$	$\begin{pmatrix} 1.2712 \\ 1.5287 \end{pmatrix}$	$\begin{pmatrix} 1.1690 \\ 1.5080 \end{pmatrix}$

Only one classifier can be trained, as there are only two types of working conditions in the set. The SVM was trained with approximately 160 groups of data from the two working conditions, and 40 samples were used to test the classifier performance. Each sample was a 2×1 matrix that could be expressed in two-dimensional coordinates. As shown in Figure 10, the two working conditions were clearly distinguished by a dividing hyperplane.
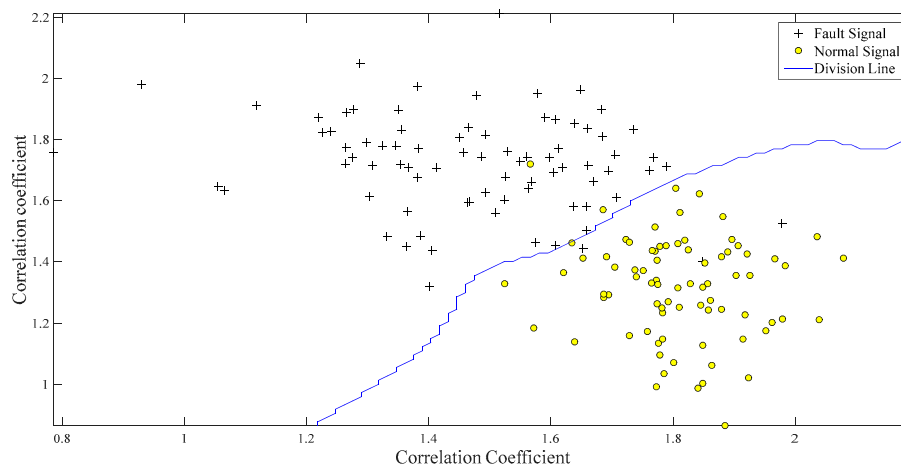
Figure 10. Characteristic distributions of samples under normal and fault conditions.

The normal and fault signals have obvious differences while calculating the feature vector. The normal signals are mainly underneath the division line because of the higher correlation coefficient between NIC and the normal conditions than between NIC and the fault conditions. The same reasoning can be applied to fault signals on the other side of the dividing line. Forty samples were tested, and there were only two false positives under the two abovementioned conditions. Therefore, the misdiagnosis rate was $2/40 \times 100\% = 5\%$.

4. Noise Diagnosis Experiment of Gearbox Failure

The fault situation during the process of gearbox operation is more complex than in the simulation not only because of the noise of a single component, but also because all kinds of aliasing noise and background noise are present. To verify the feasibility of the proposed algorithm in the noise diagnosis of the gearbox fault, the noise measurement platform was built to collect the data, and the algorithm was used to diagnose the failure of the gearbox.

4.1. Construction of Test Platform

Testing was performed in the gearbox laboratory of Shandong Lingong Construction Machinery Co., Ltd. A constant current source preamplifier (26CA) and free-field acoustic intensity sensor (46AE) from GRAS were applied. Before testing, the sound intensity sensor was first calibrated, then installed on the preamplifier and connected to the high-performance multichannel data acquisition system (LMS SCADAS) by BNC connector. Gear wear is the specific fault type considered in this paper, the length and depth of which was defined as 55 mm and 0.5 mm, respectively. The experiment was conducted using a test platform comprised of a motor, gearbox, speed sensor, constant current source preamplifier, free field acoustic intensity sensor, calibrator, and LMS SCADAS Mobile SCM05, as shown in Figure 11.
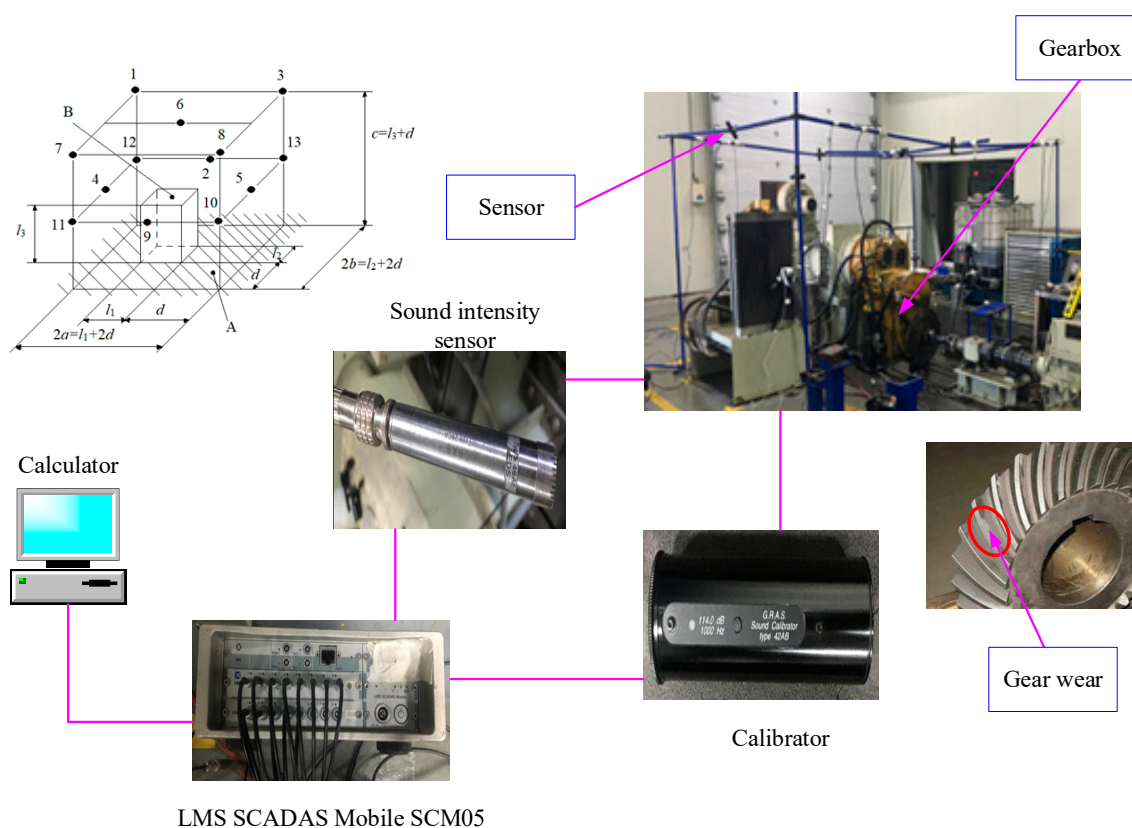
Figure 11. Construction of signal acquisition platform. Black solid point represents the sound intensity sensor, A represents reflector, B represents datum body, 2a represents measuring face length, 2b represents measuring face width, c represents height of measuring face, d represent measuring distance (80 cm), l_1 represents datum length (80 cm), l_2 represents datum width (60 cm), l_3 represents height of datum (120 cm).

The Chinese national standard GB/T 3767-2016/ISO3744:2010 "Acoustics-Determination of sound power levels and sound energy levels of noise sources using sound pressure-engineering methods for an essentially free field over a reflecting plane" was referred to arrange the positions of 13 sound intensity sensors, as shown in Figure 10. First, the speed of the test rig was adjusted to 1200 r/min, and noise data under the normal conditions were collected when the speed sensor had measured the constant speed of the out shaft. The sampling frequency was 40,960 Hz, and the sampling time was 10 min. Next, the gearbox was stopped, the fault gear ring was replaced, and the signals corresponding to the fault conditions were collected.

4.2. Noise Spectrum Analysis

The two types of working noise data collected by the data acquisition system were divided into 120 groups with 5 s as the signal sampling period. There were 240 groups of samples, 200 groups of which were the training samples, and 40 samples were used to test the classification performance of the classifier. Signals were randomly intercepted to obtain 0.5 s signal fragments, as shown in Figures 12 and 13.
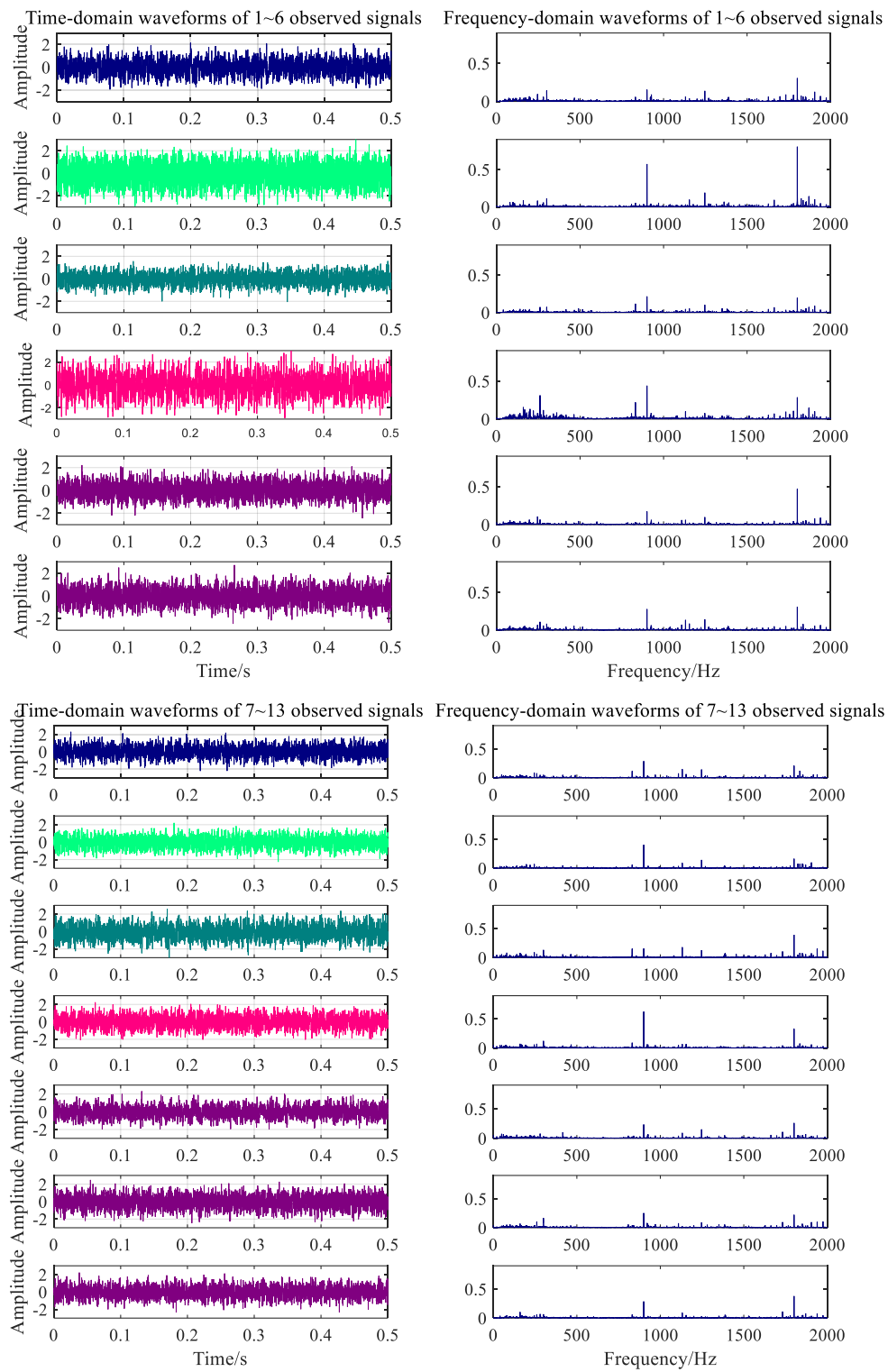
Figure 12. Time-frequency waveforms of measured signal under normal condition.
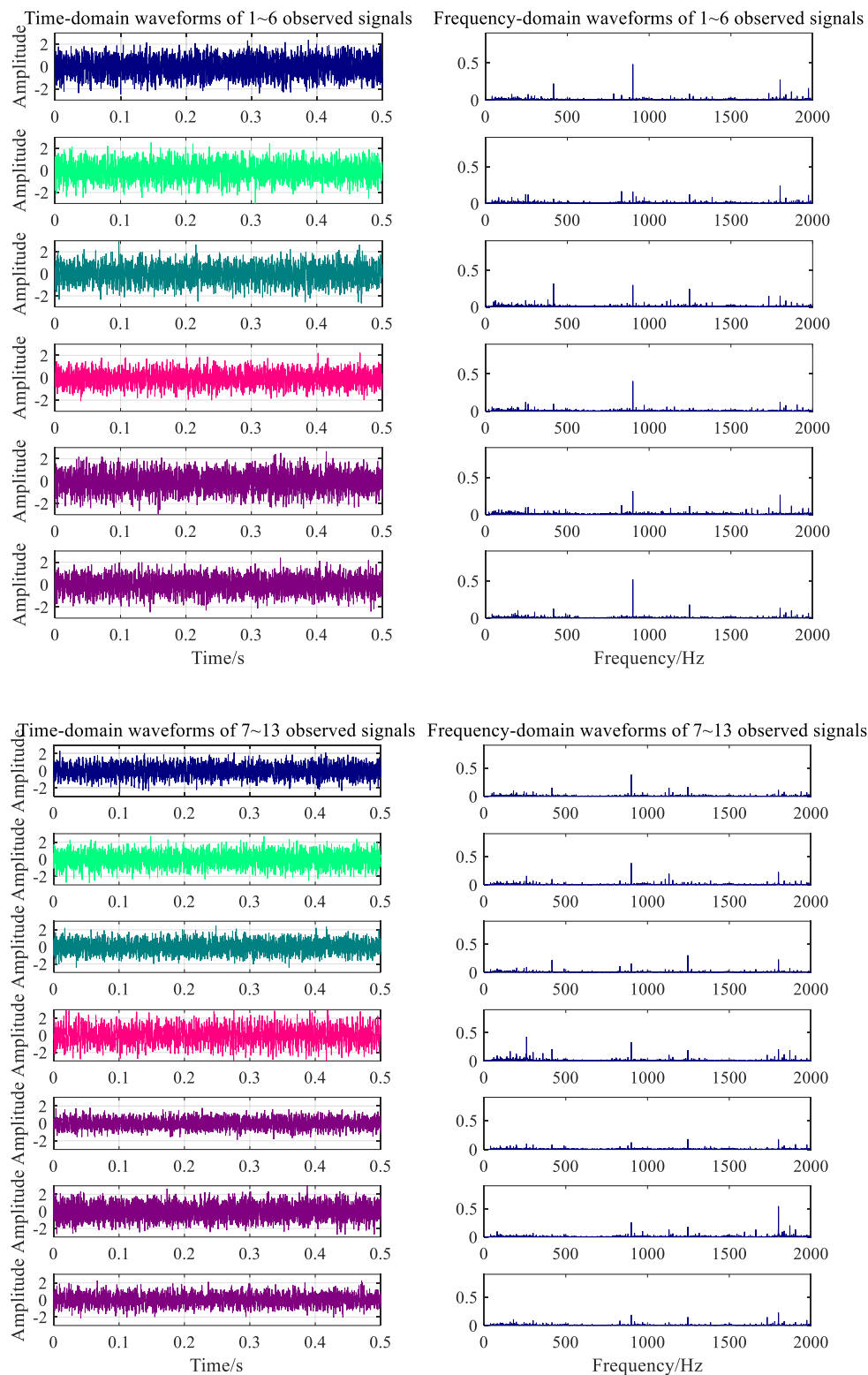
Figure 13. Time-frequency waveforms of measured signal under fault condition.

Frequency spectrum analysis was performed for noise data from 13 measuring points under different working conditions, as shown in Figures 10 and 11. The main frequency components of the noise collected under the normal working condition were 300.2, 900.5, 1247, and 1801 Hz. The speed measured by the velocimeter was 1317 r/min when the velocity was constant. Using the transmission ratio, the meshing frequency of the gear with 22 teeth was calculated to be approximately 300 Hz, the

Int. J. Environ. Res. Public Health **2019**, 16, 4868

16 of 20

meshing frequency of the driving gear when in the second gear was calculated to be 900 Hz, and the meshing noise of the second-order fitting and of the engine were calculated to be 1800 and 1247 Hz, respectively. The results of the calculation are consistent with the spectral analysis components of the experimental data. From Figure 11, the main frequency components of noise acquisition under gearbox fault conditions are 300.5, 415, 900.5, 1247, and 1801 Hz. Among these, the frequency components of 300.5, 900.5, 1247, and 1801 Hz are the same as normal working conditions, so the noise at 415 Hz might be the frequency of the fault ring.

4.3. Independent Source Separation

The MDL criterion was used to calculate the noise of the two working conditions, using the seven independent sources. The FastICA algorithm in MATLAB was applied to separate the data. First, the matrices that were independent of each other were calculated. Second, the results of the signal were estimated using the matrix in the frequency domain, and the amplitude of the signal was corrected. Finally, the independent component of the time domain was obtained through the IFT criterion.

The time-frequency diagrams of the seven independent component signals separated from the normal working noise data are shown in Figure 14. All types of noise sources under the normal conditions were successfully separated and extracted by the FastICA algorithm. These mainly included the four above-mentioned principal frequencies: 300.5, 900.5, 1801, and 1247 Hz, as well as the approximate white Gaussian noise. In Figure 15, the algorithm separates four frequency components that correspond to normal conditions and the frequency of the gear ring fault.
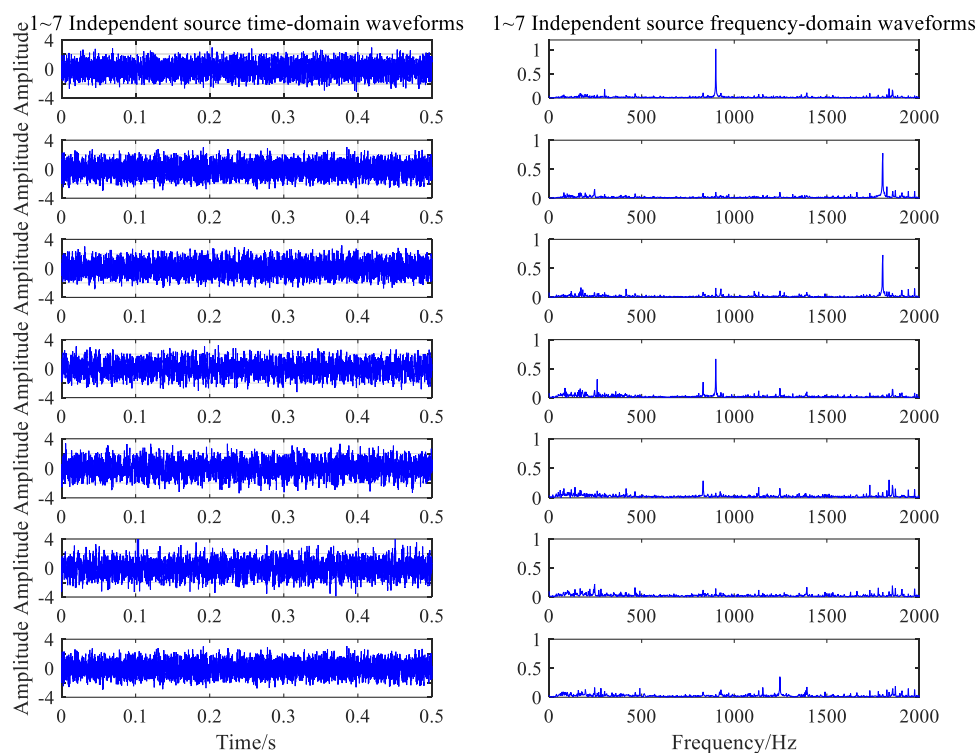

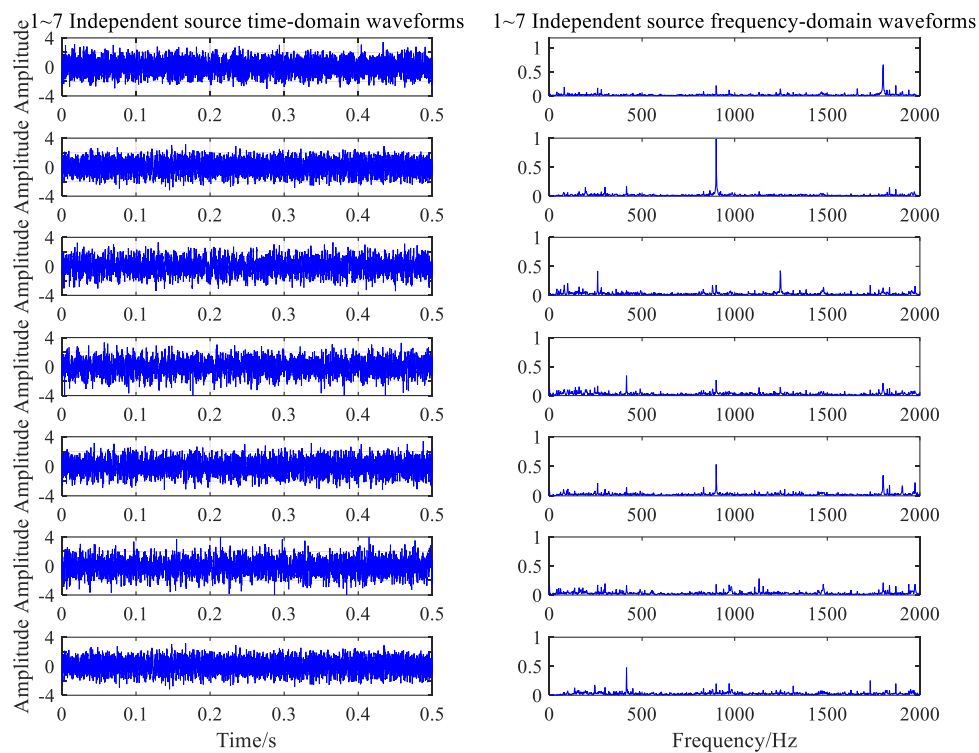**Figure 14.** Independent components of the normal condition.

Figure 15. Independent components of the fault condition.

4.4. Parameter Calculation of SVM

Two samples were extracted from the training set under different conditions. Calculation of the correlation coefficients was performed in the manner described in Section 3. The results are shown in Tables 4 and 5.

Table 4. Correlation coefficients of independent components from normal condition and source signals.

	NIC1	NIC2	NIC3	NIC4	NIC5	NIC6	NIC7
Normal1	−0.4457	0.3786	0.3969	0.2652	−0.0838	0.1085	−0.1083
Normal2	0.4512	0.3218	−0.3612	−0.2755	−0.0655	−0.0609	−0.1400
Fault1	0.4594	0.2638	0.2603	−0.2019	0.0870	0.0615	−0.0734
Fault2	−0.3279	−0.3823	−0.2488	−0.2905	−0.0593	0.0569	−0.0725

Table 5. Correlation coefficients of independent components from fault condition and source signals.

	FIC1	FIC2	FIC3	FIC4	FIC5	FIC6	FIC7
Normal1	0.3504	−0.441	0.1837	0.0978	−0.2111	−0.1088	−0.0693
Normal2	0.3603	0.2635	−0.1372	−0.1918	0.3860	0.0649	−0.0848
Fault1	0.3942	0.3173	−0.1187	0.1528	0.2563	0.0946	−0.0937
Fault2	−0.3380	−0.4281	0.1407	0.1862	−0.3218	0.1344	0.1187

The correlation coefficients between the independent components and the source signals from each working condition are obviously smaller than the results of the simulation because the signal has been separated into seven independent components. Clearly, the first five independent components provide the main characteristics of the source signal, while the remaining components can be regarded as the high-frequency noise of the source signal. This conclusion is made according to the calculation of the input parameters based on the SVM algorithm and the sum of the correlation coefficients for each row. There are no obvious gaps between the feature vectors, which may be due to the difference in gear wear under the two conditions and the masking of the vibration effect by random noise.

Int. J. Environ. Res. Public Health **2019**, *16*, 4868

18 of 20

One hundred groups of normal data samples and 100 groups of fault data samples were selected to train the SVM classifier. The division hyperplane of the SVM with the Gauss kernel function is plotted in Figure 16, in which it can be observed that the clusters corresponding to the samples for the two conditions are well separated. Forty groups of normal and fault signals were applied to verify the performance of the classifier, and there were three false positives. Therefore, the error probability was $3/40 \times 100\% = 7.5\%$.
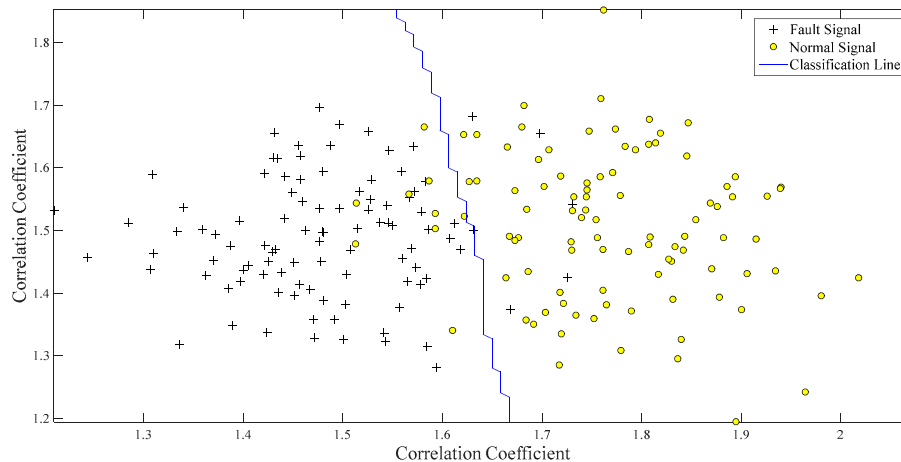
Figure 16. Classification of two classes using SVM.

5. Conclusions

(1) A fault diagnosis algorithm based on ICA and SVM is proposed in this paper. The gearbox of a loader was taken as the research object, and source signals were used to correctly diagnose gearbox failure by applying feature extraction and classification recognition as the processing technologies. The signals collected from the gearbox, which contained the principal feature information of the fault, were extracted by the ICA algorithm. The correlation coefficient was used as the characteristic parameter of fault diagnosis, as it can improve the robustness of detection systems due to the high correlations between the independent components and working conditions.

(2) Three independent source signals in the gearbox were simulated using MATLAB, and the convolution hybrid model was used to generate the six channel signals under two working conditions. Two hundred groups of simulation data were collected. Two groups of samples were selected from each condition to display the entire process of signal processing. The training of the SVM was implemented using 160 groups of samples, and 40 groups were used to test the accuracy of the classifier. There were two false positives under the two conditions; therefore, the misdiagnosis rate was 5%.

(3) A noise testing platform with 13 measurement points was built on the gearbox test rig according to the Chinese national standards, and two gear ring wear conditions were defined. One hundred and twenty groups of data under each condition were collected for the research. The algorithm established to calculate the correlation coefficient between the independent component and the source signal was applied to train the classifier and test its accuracy. There were three false positives; therefore, the error probability was 7.5%.

(4) The experimental results show that the method proposed in this paper can effectively extract various mechanical fault features from signal data, providing significant engineering value. Furthermore, this research provides an innovative approach for the feature extraction of mechanical faults from signal data.

Int. J. Environ. Res. Public Health **2019**, *16*, 4868

19 of 20

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ding, K.; Zhu, X.; Chen, Y. The vibration characteristics of typical gearbox faults and its diagnosis plan. *J. Vib. Shock* **2001**, *20*, 7–12.

2. Barros, A.K.; Cichocki, A. Extraction of Specific Signals with Temporal Structure. *Neural Comput.* **2001**, *13*, 1995–2003. [CrossRef] [PubMed]

3. Puntonet, C.G.; Lang, E.W. Blind source separation and independent component analysis. *Neurocomputing* **2006**, *69*, 1413. [CrossRef]

4. Barros, A.K.; Carlos Príncipe, J.; Erdogmus, D. Independent Component Analysis and Blind Source Separation. *Signal Process.* **2007**, *87*, 1817–1818. [CrossRef]

5. Žvokelj, M.; Zupan, S.; Prebil, I. EEMD-based multiscale ICA method for slewing bearing fault detection and diagnosis. *J. Sound Vib.* **2016**, *370*, 394–423. [CrossRef]

6. Melissant, C.; Ypma, A.; Frietman, E.E.; Stam, C.J. A method for detection of Alzheimer's disease using ICA-enhanced EEG measurements. *Artif. Intell. Med.* **2005**, *33*, 209–222. [CrossRef]

7. Gohel, B.; Lee, P.; Kim, M.-Y.; Kim, K.; Jeong, Y. MEG Based Functional Connectivity: Application of ICA to Alleviate Signal Leakage. *IRBM* **2017**, *38*, 127–137. [CrossRef]

8. Back, A.; Weigend, A.S. A First Application of Independent Component Analysis to Extracting Structure from Stock Returns. *Int. J. Neural Syst.* **1997**, *8*, 473–484. [CrossRef]

9. Oja, E.; Kiviluoto, K.; Malaroiu, S. Independent component analysis for financial time series. In Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373), Lake Louise, AB, Canada, 4 October 2000.

10. Stefatos, G.; Hamza, A.B. Dynamic independent component analysis approach for fault detection and diagnosis. *Expert Syst. Appl.* **2010**, *37*, 8606–8617. [CrossRef]

11. Ajami, A.; Daneshvar, M. Data driven approach for fault detection and diagnosis of turbine in thermal power plant using Independent Component Analysis (ICA). *Int. J. Electr. Power Energy Syst.* **2012**, *43*, 728–735. [CrossRef]

12. Badaoui, M.E.; Danière, J.; Guillet, F.; Servière, C. Separation of combustion noise and piston-slap in diesel engine—Part I: Separation of combustion noise and piston-slap in diesel engine by cyclic Wiener filtering. *Mech. Syst. Signal Process.* **2005**, *19*, 1209–1217. [CrossRef]

13. Wang, Z.; Chen, J.; Dong, G.; Zhou, Y. Constrained independent component analysis and its application to machine fault diagnosis. *Mech. Syst. Signal Process.* **2011**, *25*, 2501–2512. [CrossRef]

14. Sun, Z.; Shang, L. An improved constrained ICA with reference based unmixing matrix initialization. *Neurocomputing* **2010**, *73*, 1013–1017. [CrossRef]

15. Feng, L.; Di, T.; Zhang, Y. HSIC-based kernel independent component analysis for fault monitoring. *Chemom. Intell. Lab. Syst.* **2018**, *178*, 47–55. [CrossRef]

16. Zhang, Y.; Ma, C. Decentralized fault diagnosis using multiblock kernel independent component analysis. *Chem. Eng. Res. Des.* **2012**, *90*, 667–676. [CrossRef]

17. Widodo, A.; Yang, B.; Han, T. Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors. *Expert Syst. Appl.* **2007**, *32*, 299–312. [CrossRef]

18. Min, J.H.; Lee, Y. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Syst. Appl.* **2005**, *28*, 603–614. [CrossRef]

19. Danenas, P.; Garsva, G. Selection of Support Vector Machines based classifiers for credit risk domain. *Expert Syst. Appl.* **2015**, *42*, 3194–3204. [CrossRef]

20. Salvati, D.; Drioli, C.; Foresti, G.L. A weighted MVDR beamformer based on SVM learning for sound source localization. *Pattern Recognit. Lett.* **2016**, *84*, 15–21. [CrossRef]

21. Kalantarian, H.; Mortazavi, B.; Pourhomayoun, M.; Alshurafa, N.; Sarrafzadeh, M. Probabilistic Segmentation of Time-Series Audio Signals using Support Vector Machines. *Microprocess. Microsyst.* **2016**, *46*, 96–104. [CrossRef]

22. Manjurul Islam, M.M.; Kim, J. Reliable multiple combined fault diagnosis of bearings using heterogeneous feature models and multiclass support vector Machines. *Reliab. Eng. Syst. Saf.* **2018**. [CrossRef]

23. Li, Y.; Yang, Y.; Wang, X.; Liu, B.; Liang, X. Early fault diagnosis of rolling bearings based on hierarchical symbol dynamic entropy and binary tree support vector machine. *J. Sound Vib.* **2018**, *428*, 72–86. [CrossRef]

24. Samanta, B. Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. *Mech. Syst. Signal Process.* **2004**, *18*, 625–644. [CrossRef]

25. Yang, D.; Liu, Y.; Li, S.; Li, X.; Ma, L. Gear fault diagnosis based on support vector machine optimized by artificial bee colony algorithm. *Mech. Mach. Theory* **2015**, *90*, 219–229. [CrossRef]

26. Saari, J.; Strömbergsson, D.; Lundberg, J.; Thomson, A. Detection and identification of windmill bearing faults using a one-class support vector machine (SVM). *Measurement* **2019**. [CrossRef]

27. Zhang, Y.; Wang, P.; Ni, T.; Cheng, P.; Lei, S. Wind Power Prediction Based on LS-SVM Model with Error Correction. *Adv. Electr. Comput. Eng.* **2017**, *17*, 3–8. [CrossRef]

28. Glowacz, A. Recognition of acoustic signals of induction motor using FFT, SMOFS-10 and LSVM. *Eksploat. I Niezawodn. Maint. Reliab.* **2015**, *17*, 569–574. [CrossRef]

29. Glowacz, A. Recognition of Acoustic Signals of Commutator Motors. *Appl. Sci.* **2018**, *8*, 2630. [CrossRef]

30. He, G.; Ding, K.; Lin, H. Gearbox coupling modulation separation method based on match pursuit and correlation filtering. *Mech. Syst. Signal Process.* **2016**, *66–67*, 597–611. [CrossRef]

31. Zhou, L. *Machine Learning*; Tsinghua University Press: Beijing, China, 2016.