




## Article

# Multi-Features Fusion for Fault Diagnosis of Pedal Robot Using Time-Speed Signals

Yuhao Zhu <sup>1</sup> , Zeyu Fu <sup>1</sup> , Zhuang Fu <sup>1,\*</sup>, Xi Chen <sup>2</sup> and Qi Wu <sup>3</sup> 

<sup>1</sup> State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai 200240, China; yuhaozhu@sjtu.edu.cn (Y.Z.); summer0528@sjtu.edu.cn (Z.F.)

<sup>2</sup> Shanghai Engineering Research Center of Civil Aircraft Health Monitoring, Shanghai Aircraft Customer Service Co., Ltd., Shanghai 200241, China; chenxi1@comac.cc

<sup>3</sup> School of Electronic, Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China; wuqi7812@sjtu.edu.cn

\* Correspondence: zhfu@sjtu.edu.cn; Tel.: +86-1893-076-1906

Received: 3 December 2018; Accepted: 27 December 2018; Published: 4 January 2019

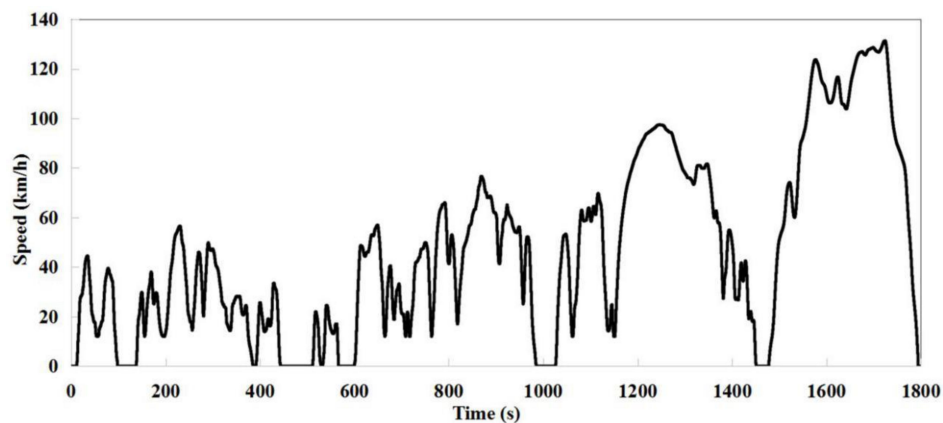


**Abstract:** In order to realize automation of the pollutant emission tests of vehicles, a pedal robot is designed instead of a human-driven vehicle. Sometimes, the actual time-speed curve of the vehicle will deviate from the upper or lower limit of the worldwide light-duty test cycle (WLTC) target curve, which will cause a fault. In this paper, a new fault diagnosis method is proposed and applied to the pedal robot. Since principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), and Autoencoder cannot extract feature information adequately when they are used alone, three types of feature components extracted by PCA, t-SNE, and Autoencoder are fused to form a nine-dimensional feature set. Then, the feature set is reduced into three-dimensional space via Treelet Transform. Finally, the fault samples are classified by Gaussian process classifier. Compared with the methods using only one algorithm to extract features, the proposed method has the minimum standard deviation, 0.0078, and almost the maximum accuracy, 98.17%. The accuracy of the proposed method is only 0.24% lower than that without Treelet Transform, but the processing time is 6.73% less than that without Treelet Transform. These indicate that the multi-features fusion model and Treelet Transform method is quite effective. Therefore, the proposed method is quite helpful for fault diagnosis of the pedal robot.

**Keywords:** pedal robot; fault diagnosis; autoencoder; Treelet Transform; multi-features fusion

## 1. Introduction

Before vehicle road tests or quality inspection, pollutant emission tests must be carried out. Over the past decade, these tests have been conducted by people who drive vehicles and follow the world light-duty test cycle (WLTC) target curve reported in Reference [1]. As shown in Figure 1, it is a time-speed curve that lasts 1800 s. In fact, when a vehicle is driven, the actual curve often deviates from the target curve. The noise in the test environment is too loud to be tolerated by operators, which may cause human error. In addition, research shows that although the deviation caused by robot control is less than that caused by human control, the number of deviations still needs to be reduced to almost zero. Generally, when the deviation between the actual curve and the target curve is greater than 2 km/h, a fault will occur. Different control parameters of pedal robot are needed for different types of faults. Therefore, how to identify fault types is still the most challenging part of robot control optimization.



**Figure 1.** Curve of the worldwide light-duty test cycle (WLTC).

Traditional fault diagnosis is analyzed by skilled engineers in time domain [2], which is difficult and time-consuming, especially in the case of large samples and/or high-dimensional samples. With the increase of data size, analysis becomes quite complex. It is necessary to reduce data size by feature extraction [3]. So far, many studies on feature extraction methods have been reported for the dynamic characteristics of fault signals [4]. One of the traditional feature extraction techniques is Fourier transform [5], but a Fourier coefficient represents a component that lasts for all time. Therefore, Fourier analysis is less suitable for nonstationary signals [6]. Other feature extraction algorithms, such as local mean decomposition, correlation coefficient, and kurtosis, can also extract features. In general, these features are directly used as input of the classifier. The calculation speed of these methods is fast, but they are not able to present adequate information of fault signals. Autoencoder, a deep learning algorithm, is used to extract more abstract features from the raw input. A major feature component is received via t-distributed stochastic neighbor embedding (t-SNE) algorithm or principal component analysis (PCA) method. These extracted feature components, or representations, are fused into a new matrix, which may contain more local major feature information. Thus, this work proposes a fusion methodology of multi-feature learning models via Autoencoder, t-SNE, and PCA.

PCA generates a set of orthogonal bases that capture the direction of the largest variance in training data [7]. Dunia et al. [8] used PCA for sensor fault identification via reconstruction. The disadvantage of PCA is that it can only extract linear features, while t-SNE can solve nonlinear problems. T-SNE, as proposed in Reference [9], is a very effective nonlinear dimension reduction technology, which was widely used recently. It is most commonly used to generate two-dimensional embedding of high-dimensional data in order to simplify cluster recognition [10]. However, PCA and t-SNE need to decompose the problem into different parts and merge the results in the final stage. In addition, most of the features in machine learning need to be recognized by experts, which is very troublesome. Deep learning has no such problems mentioned above, and can obtain the representations with sensitive and valuable feature through automatically learning from the original feature set [11]. In addition, because of the deep model [12], deep learning has stronger adaptability and stability. Therefore, an Autoencoder-based learning algorithm is used to extract abstract features from the raw input. The Autoencoder model also a special kind of neural network and is able to learn richer representation of input data, as well as represent the target output [13]. The aim of this model is to reconstruct input from the learnt representation [14]. Although the Autoencoder can automatically learn features, it also has some disadvantages. For example, when the training set is small, or the training sample does not follow the same distribution as the test sample, the accuracy of feature extraction will be reduced. Clearly, PCA, t-SNE, and Autoencoder have their own advantages and shortcomings over the feature learning. In order to make full use of their advantages, this work proposes a novel additive combination feature representation respectively from PCA, t-SNE, and Autoencoder. The features provided by them will contain more dimensional information of fault signals.

After feature extraction, the features are combined into a high-dimensional feature set. In order to speed up the calculation and visualize the dimension reduction results, it is necessary to reduce the dimension

of feature set to three. Treelet Transform is a data dimension reduction technology [15]. It generates some sparse components, which is able to reveal the inherent structure of data [16]. After dimension reduction, a pattern classification method is needed. Gaussian process classifier based on Bayesian theory is one of the machine learning methods [17]. It can effectively classify nonlinear samples [18]. Gaussian process is used to learn the extracted features and present the fault detection results.

In this paper, a new fault diagnosis method based on multi-feature learning model is proposed. This method can identify fault type and help optimize the control of the pedal robot. The experimental results are compared with some typical feature extraction algorithms and the proposed model in terms of accuracy. The structure of this paper is as follows: Section 2 gives the proposed methodology, Section 3 gives experiment setup, Section 4 presents the results and discussion, and the conclusion is arranged in Section 5.

## 2. Methodology

The proposed fault diagnosis method for pedal robot mainly includes three steps: Feature extraction, dimension reduction, and fault classification. First, fault samples are extracted from the curve samples. Second, three-dimensional features are extracted from the fault samples by PCA, t-SNE, and Autoencoder, respectively. Next, since PCA, t-SNE, and Autoencoder cannot extract feature information adequately when they are used alone, and the types of features extracted by them are different, the features are combined into a nine-dimensional feature set to reduce information loss. Then, the feature set is reduced to three dimensions by Treelet Transform, thus, the calculation speed could be accelerated, and the fault samples could be visualized in three-dimensional space. Finally, after being classified by Gaussian process classifier, the accuracy and result of fault diagnosis are obtained.

### 2.1. Data Preprocessing

A fault sample contains two curves, namely, the target curve and the actual curve. Ten seconds before and after the midpoint of the deviation, the segment was extracted as a fault sample. Since the curve took a point at 0.1 s intervals, a fault sample had 200 points. In order to extract more features, data integration was necessary. The 200 points of the actual curve were combined with the 200 points of the corresponding target curve to form a 400-dimensional feature. In addition, this study also collected the time-displacement curves of the pedal robot's two end-effectors. Therefore, the 400-dimensional feature and the 200 points of each end-effector displacement at the corresponding time were combined to form an 800-dimensional feature, which could be used as a fault sample.

### 2.2. Feature Extraction

This work explores the feasibility of feature extraction by using PCA, t-SNE, and Autoencoder to extract three-dimensional features, respectively, then these three three-dimensional features are combined into a nine-dimensional feature matrix.

#### 2.2.1. Feature Extraction by PCA

The process of feature extraction by PCA consists of the following five stages:

Stage 1: Each sample is 800 dimensions, and 11430 samples constitute a sample array:

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T, i = 1, 2, \dots, 11430; p = 1, 2, \dots, 800, \quad (1)$$

then, the sample elements are normalized as follows:

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, i = 1, 2, \dots, 11430; j = 1, 2, \dots, 800, \quad (2)$$

where  $\bar{x}_j = \frac{\sum_{i=1}^{11430} x_{ij}}{11430}$ ,  $s_j^2 = \frac{\sum_{i=1}^{11430} (x_{ij} - \bar{x}_j)^2}{11430 - 1}$ , thus the normalized matrix  $\mathbf{Z}$  is found.

Stage 2: Calculate the correlation coefficient matrix of the normalized matrix  $\mathbf{Z}$ .

$$\mathbf{R} = \frac{\mathbf{Z}^T \mathbf{Z}}{11430 - 1}, \quad (3)$$

Stage 3: Solve the eigenvalue of matrix  $\mathbf{R}$ , then 800 eigenvalues are got to determine the principal components. The value of  $m$  is determined by the following equation [19]:

$$\frac{\sum_{j=1}^m \lambda_j}{\sum_{j=1}^{800} \lambda_j} \geq 0.85, \quad (4)$$

For each  $\lambda_j, j = 1, 2, \dots, m$ , solve the equations set  $\mathbf{R}\mathbf{b} = \lambda_j \mathbf{b}$  to obtain the unit eigenvector  $\mathbf{b}_j^0$ .

Stage 4: Convert the normalized indicator variables into principal components:

$$\mathbf{U}_{ij} = \mathbf{z}_i^T \mathbf{b}_j^0, j = 1, 2, \dots, m, \quad (5)$$

where  $\mathbf{U}_1$  is the first principal component,  $\mathbf{U}_2$  is the second principal component and  $\mathbf{U}_p$  is the  $p^{th}$  principal component.

Stage 5: Comprehensive evaluation of  $m$  principal components.

In this paper, the total eigenvalues contribution of the first three eigenvalues was 99.37%. It indicates that the first three dimensions already contain enough information of the original sample. Thus, they were extracted as features of the high-dimensional fault samples.

### 2.2.2. Feature Extraction by t-SNE

The  $d$ -dimensional input dataset is denoted by  $\chi = \{x_1, x_2, \dots, x_n\} \subset C^d$ . Before using t-SNE algorithm, PCA is used to initialize data. t-SNE computes a  $s$ -dimensional embedding of the points in  $\chi$ , denoted by  $\mathbf{y} = \{y_1, y_2, \dots, y_n\} \subset C^s$ , where  $s < d$ . The joint probability  $p_{ij}$  measuring the similarity between  $x_i$  and  $x_j$  is computed as:

$$p_{i|j} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma_i^2)}, \text{ and } p_{ij} = \frac{p_{i|j} + p_{j|i}}{2n}. \quad (6)$$

The bandwidth of the Gaussian kernel,  $\sigma_i$ , is often chosen such that the perplexity of  $P_i$  matches a user defined value, where  $P_i$  is the conditional distribution across all data point given  $x_i$ . The similarity between  $y_i$  and  $y_j$  in the low dimensional embedding is defined as:

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq i} (1 + \|y_k - y_i\|^2)^{-1}}. \quad (7)$$

t-SNE finds the points  $\{y_1, y_2, \dots, y_n\}$  which minimize the Kullback-Leibler divergence between the joint distribution  $\mathbf{P}$  of points in the input space and the joint distribution  $\mathbf{Q}$  of points in the embedding space:

$$C(\mathbf{y}) = KL(\mathbf{P} \parallel \mathbf{Q}) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}}. \quad (8)$$

The point  $\mathbf{y}$  is initialized randomly, and the cost function  $C(\mathbf{y})$  is minimized using gradient descent. In this paper,  $d = 800$ , and  $s = 3$ . Thus, three-dimensional feature was extracted from the high-dimensional feature set.

### 2.2.3. Feature Extraction by Autoencoder

An Autoencoder learns a feed-forward, hidden representation  $h(\mathbf{x})$  of the input  $\mathbf{x}$ , and a reconstruction  $\hat{\mathbf{x}}$  could be got, which is as close as possible to  $\mathbf{x}$ .

$$h(\mathbf{x}) = g(b + \omega \mathbf{x}), \quad (9)$$

$$\hat{\mathbf{x}} = \text{sigm}(c + v h(\mathbf{x})), \quad (10)$$

where  $\omega$  and  $v$  are matrices,  $b$  and  $c$  are vectors,  $g$  is a nonlinear activation function and  $\text{sigm}(a) = 1/(1 + \exp(-a))$ . The loss function used in this paper is cross-entropy loss:

$$l(\mathbf{x}) = \sum_{d=1}^D = x_d \log \hat{x}_d - (1 - x_d) \log(1 - \hat{x}_d), \quad (11)$$

Training the Autoencoder corresponds to optimizing the parameters to reduce the average loss on the training examples. In this paper, the hidden layer  $h(\mathbf{x})$  has three neurons, which represent the features extracted from the input.

### 2.3. Dimension Reduction by Treelet Transform

After using PCA, t-SNE, and Autoencoder to extract three-dimensional features, respectively, the features are combined into a nine-dimensional feature set. Thus, a  $11430 \times 9$  feature set matrix is extracted from the  $11430 \times 800$  fault samples matrix. Then, this feature set is reduced to three-dimensional feature set by Treelet Transform.

Stage 1: Normalize the features extracted by PCA, t-SNE, and Autoencoder, respectively.

Stage 2: Define the clustering levels  $l = 0, 1, 2, \dots, L$  ( $L_{\max} = p - 1$ ), where  $p$  is the dimension 9. At level  $l = 0$ , each signal  $\mathbf{x}_k$  is represented by the original variables  $\mathbf{x}^{(0)} = [s_{0,1}, s_{0,2}, \dots, s_{0,p}]$  where  $s_{0,k} = x_k$ . Initialize similarity matrix and covariance matrix by the following Equation:

$$\rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ij} \Sigma_{ij}}}, \quad (12)$$

with  $\Sigma_{ij} = E[(s_i - E s_i)(s_j - E s_j)]^T$ ,

where  $\rho_{ij}$  is the correlation coefficients of the similarity matrix  $\mathbf{M}_{ij}$ , and  $\mathbf{M}_{ij} = |\rho_{ij}| + \lambda |\Sigma_{ij}|$ . Initialize the basis matrix  $\mathbf{B}_0 = [\phi_{01}, \phi_{02}, \dots, \phi_{0p}]$  as an identity matrix whose size is  $p \times p$ . Initialize  $\delta = \{1, 2, \dots, p\}$ . For the  $L$ -level Treelet Transform, let  $l = 1, 2, \dots, L$ , repeat the following stages:

Stage 3: At each level of the tree, the most similar variables are obtained according to the similarity matrix  $\mathbf{M}^{(l-1)}$ , denote the two variables which have the maximum correlation coefficients as  $\alpha$  and  $\beta$ :

$$(\alpha, \beta) = \underset{i,j \in \delta}{\operatorname{argmax}} \mathbf{M}_{ij}^{(l-1)}, \quad (13)$$

Stage 4: Find the Jacobi rotation matrix:

$$J(\alpha, \beta, \theta_1) = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & c \end{bmatrix} \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & c & \cdots & -s & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & s & \cdots & c & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix}, \quad (14)$$

where  $c = \cos(\theta_1)$  and  $s = \sin(\theta_1)$ , and the rotation angle  $\theta_1$  can be obtained by:

$$|\theta_l| \leq \frac{\pi}{4}, \sum^{(l)} = J^T \sum^{(l-1)} J, \sum_{\alpha\beta}^{(l)} = \sum_{\beta\alpha}^{(l)} = 0, \quad (15)$$

This transformation corresponds to a change of basis  $\mathbf{B}_1 = \mathbf{B}_{1-1}\mathbf{J}$ , and new coordinates  $\mathbf{x}^{(l)} = \mathbf{J}^T \mathbf{x}^{(l-1)}$ . Update the similarity matrix  $\mathbf{M}^{(l)} = \mathbf{J}^T \mathbf{M}^{(l-1)} \mathbf{J}$  accordingly.

Stage 5: Multi-resolution analysis. If  $\sum_{\alpha\alpha}^{(l)} \geq \sum_{\beta\beta}^{(l)}$ , define the difference and sum as  $d = x_\beta^l$  and  $s_1 = x_\alpha^{(l)}$ . Define the scaling and detail functions  $\phi_1$  and  $\psi_1$  as columns  $\alpha$  and  $\beta$  of the basis matrix  $\mathbf{B}_1$ .

Stage 6: Repeat stage 2 to stage 4 until reach the highest level  $l = L$ . Then, the final orthogonal basis matrix  $\mathbf{B}$  can be obtained as follows:

$$\mathbf{B} = [\phi_1, \psi_1, \psi_2, \dots, \psi_l]^T, \quad (16)$$

where  $\phi_1$  is the scale of the highest level, and  $\psi_1$  is the detail vector of the  $1^{th}$  level. Then, the orthonormal treelet decomposition at level  $l$  can be obtained:

$$x = \sum_{i=1}^{p-1} s_{l,i} \phi_{l,i} + \sum_{i=1}^l d_i \psi_i, \quad (17)$$

Stage 7: Normalize the obtained eigenvectors after Treelet Transform.

In this paper, the input was the  $11430 \times 9$  feature set matrix, and the maximum level  $L = 6$ . The output was three-dimensional normalized eigenvectors. The score of remaining three dimensions was 94.61%, which indicated that almost all the information had been extracted from the feature set matrix. So, the remaining three dimensions were input into the classifier as a representative of the high-dimensional feature set matrix.

#### 2.4. Fault Classification by Gaussian Process Classifier

After feature extraction and dimension reduction, a  $11430 \times 3$  matrix of fault samples is obtained. Then, these fault samples are trained and formed into four classes by Gaussian process classifier.

Assume the dataset is  $\mathbf{D} = \{(x_i, y_i) | i = 1, \dots, N\}$ , where  $\mathbf{x}$  is the input vector of dimension  $d$  and  $\mathbf{y}$  is the class labels  $+1/-1$ . The input  $N \times d$  matrix is denoted as  $\mathbf{X}$ . Predictions for new inputs  $\mathbf{x}^*$  are made out of this given training data using the GP model. And GP binary classifier is done by first calculating the distribution over the latent function  $f$  corresponding to the test case,

$$p(f_* | \mathbf{X}, \mathbf{y}, \mathbf{x}^*) = \int p(f_* | \mathbf{X}, \mathbf{x}_*, f) p(f | \mathbf{X}, \mathbf{y}) df, \quad (18)$$

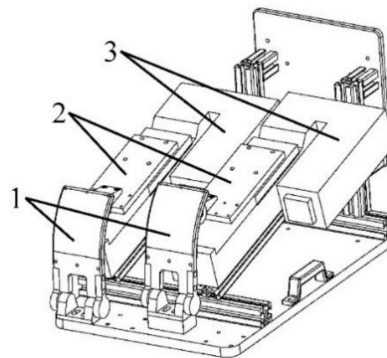
where  $p(f | \mathbf{X}, \mathbf{y}) = p(\mathbf{y} | f) p(f | \mathbf{X}) / p(\mathbf{y} | \mathbf{X})$  is the latent variable posterior. So, the probabilistic prediction is made by

$$\bar{\pi}_* = p(\mathbf{y}_* = +1 | \mathbf{X}, \mathbf{y}, \mathbf{x}_*) = \int \sigma(f_*) p(f_* | \mathbf{X}, \mathbf{y}, \mathbf{x}_*) df_*, \quad (19)$$

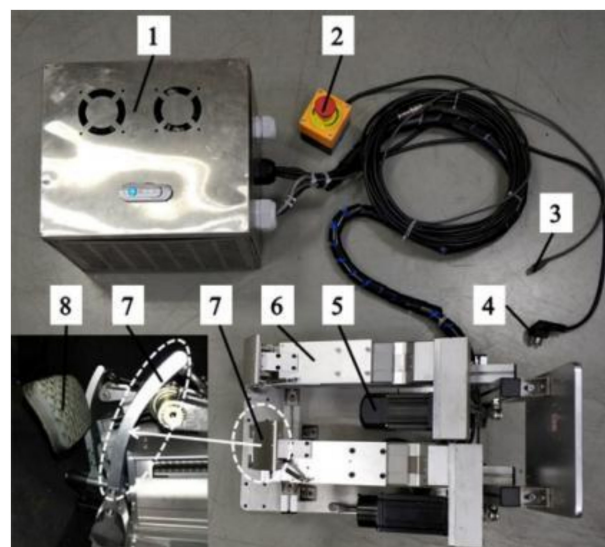
In this work,  $N = 11430$  and  $d = 3$ . Since the outputs are divided into four classes, a multi-classification method is presented as: The first class consists of larger samples, and other three sample sets form the second class. This completes one classification and presents the classification probability of larger samples. Next, this step is repeated at the other three sample sets. Finally, one probability of every sample at each classification is obtained. The maximum probability is endowed to this sample. The accuracy is computed by comparing the predictive labels with the expert experience-based labels.

### 3. Experiment

In order to realize the automation of pollutant emission tests without changing and disassembling the structure of vehicle products, a pedal robot instead of a human-driven vehicle is designed. As shown in Figure 2, the main components of the pedal robot are servo motors, sliders, and end-effectors. The servo motor pushes the end effector by controlling the slider. In order to improve the stability of control, the end-effector is designed to imitate the shape of human foot. As shown in Figure 3, the robot system consists of three parts: Control box, manipulator, and end effector.



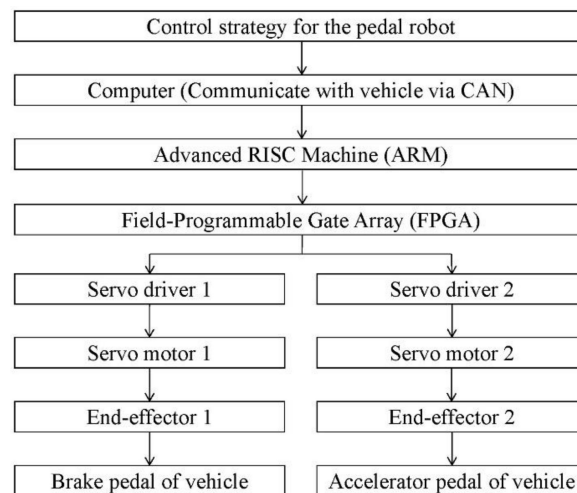
**Figure 2.** The axonometric drawing of pedal robot designed in this work. The numbers denote the following: 1—end-effectors of the robot, 2—sliders, and 3—servo motors.



**Figure 3.** The robot system designed in this work. The numbers denote the following: 1—control box, 2—emergency stop button, 3—interface to the computer, 4—interface to the power supply, 5—servo motor, 6—slider, 7—end-effectors of the robot, and 8—brake pedal of the vehicle.

As shown in Figure 4, the operation process is as follows: First, the control strategy of the robot is transmitted to the advanced RISC machine (ARM) by computer. Next, ARM sends control signals to Field Programmable Gate Array (FPGA). Then, the servo motor is controlled by the servo driver. Finally, the servo motor pushes the end-effector to drive the brake pedal or acceleration pedal of the vehicle. During emission testing, the pedal robot will control the vehicle and make the time-speed curve of the vehicle follow the WLTC curve. The actual speed of the vehicle will be fed back to the computer every 0.1 s by Controller Area Network (CAN). Therefore, the faults can be analyzed according to the feedback information.



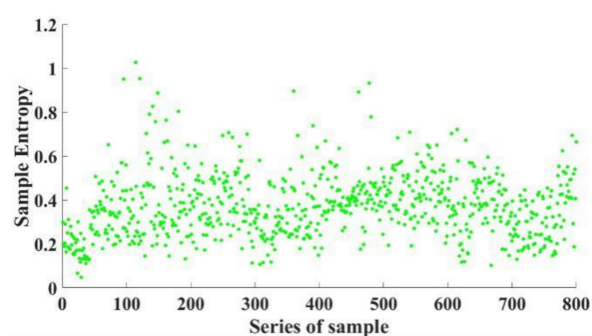


**Figure 4.** Procedures of robot control.

In this work, 50 different vehicles are used. After each vehicle test, the parameters of the robot control are adjusted according to the type and number of the faults, and the test is done again. On average, each vehicle undergoes about seven optimization processes, and finally, the number of faults in one test is reduced to almost zero. The fault appearance of each vehicle is not the same. By testing 50 vehicles and testing each vehicle several times, the usability of our method is improved, and the accuracy of fault prediction is also improved.

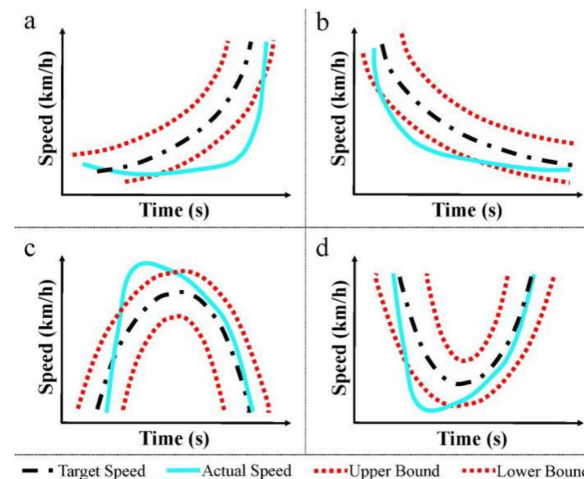
#### 4. Results and Discussion

In this work, 346 pollutant emission tests are carried out, and 11,430 fault samples are obtained. Figure 5 shows the sample entropy of 800 samples. Based on expert experience, the fault samples are classified into four types, and each fault sample has a type label. The expert experience-based labels is the opinion of human operators. The operators are experienced, and by comparing the actual curve with the target curve, the type of fault can be correctly identified. Figure 6 shows the schematic diagram of four types of faults. According to the time-speed curve, faults can be classified into four types: Acceleration fault, deceleration fault, convex fault, and concave fault.



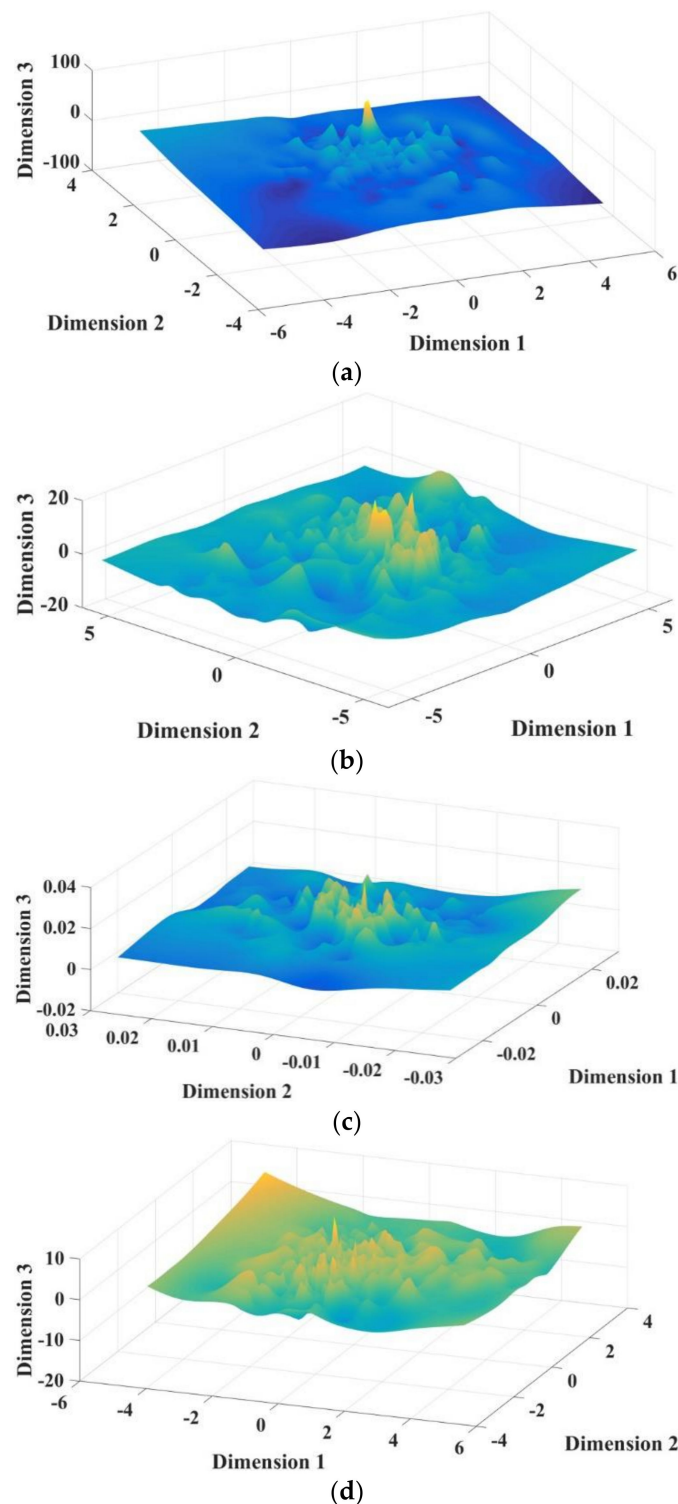
**Figure 5.** The sample entropy of 800 samples.





**Figure 6.** The schematic diagram of four types of faults: (a) Acceleration fault; (b) Deceleration fault; (c) Convex fault; and (d) Concave fault.

Gaussian process classification is quite time-consuming. If the dimension of the extracted feature is high, the total processing time will be long. But, if the dimension is low, the information will be lost. After many tests, it is found that the first three dimensions already contain enough information. It can not only reduce the dimension of data and accelerate the calculation speed, but also retain the feature information and the classification accuracy. Furthermore, 3D is good for visualization. After feature extraction, three-dimensional features are extracted from the fault samples by PCA, t-SNE, and Autoencoder, respectively. After dimension reduction by Treelet Transform, the feature set is reduced to a three-dimensional feature set matrix. Figure 7 is the 3D surface graph of three-dimensional features extracted by PCA, t-SNE, Autoencoder, and Treelet Transform, respectively. The values of  $x$ ,  $y$ , and  $z$  coordinates are extracted three vectors.



**Figure 7.** 3D surface graph of three-dimensional features extracted by each algorithm. (a) Three-dimensional features extracted by principal component analysis (PCA); (b) Three-dimensional features extracted by t-distributed stochastic neighbor embedding (t-SNE); (c) Three-dimensional features extracted by Autoencoder; (d) Three-dimensional features extracted by Treelet Transform.

The fault samples are classified into four types by Gaussian process classifier. Since ten-fold cross-validation has good performance [20], it is used in this paper. The average accuracy and standard deviation comparison between different methods is shown in Table 1. Method 4 is the method proposed in this paper, which has the minimum standard deviation and almost the maximum accuracy.

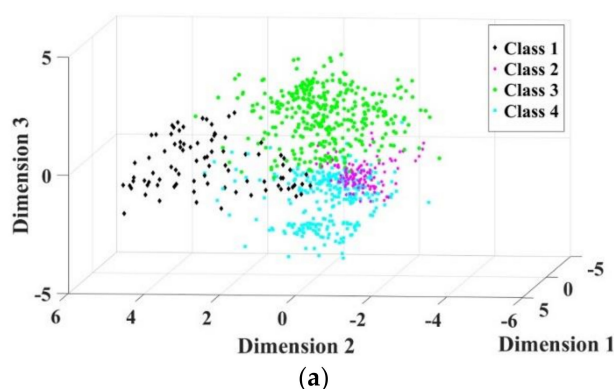
As shown in Table 1, when PCA, t-SNE, and Autoencoder are used to extract features as the input to Gaussian process classifier, respectively, the accuracy of Autoencoder is higher than that of t-SNE and PCA. This shows that the deep learning method is more effective than other two feature extraction methods. Since PCA, t-SNE, and Autoencoder cannot extract feature information adequately when they are used alone, and the feature types extracted by them are different, these features are combined into a nine-dimensional feature set as the input of Gaussian process classifier. And the accuracy of this method (Method 5) is 98.41%, which is much higher than that of PCA, t-SNE, and Autoencoder, respectively. This shows that the information loss can be effectively prevented by multi-features fusion.

In addition, as shown in Table 1, the processing time of Method 4 is far less than that of Method 5. Because the higher the dimension, the longer the time required for Gaussian process classification. The accuracy of Method 4 is only 0.24% lower than that of Method 5, but the processing time of one sample in Method 4 is 6.73% less than that of Method 5. These clearly prove that Treelet Transform can not only reduce the dimension of data and accelerate the calculation speed, but also retain the feature information and the classification accuracy. Therefore, the hybrid multi-models and Treelet Transform methodology is more effective than other methods.

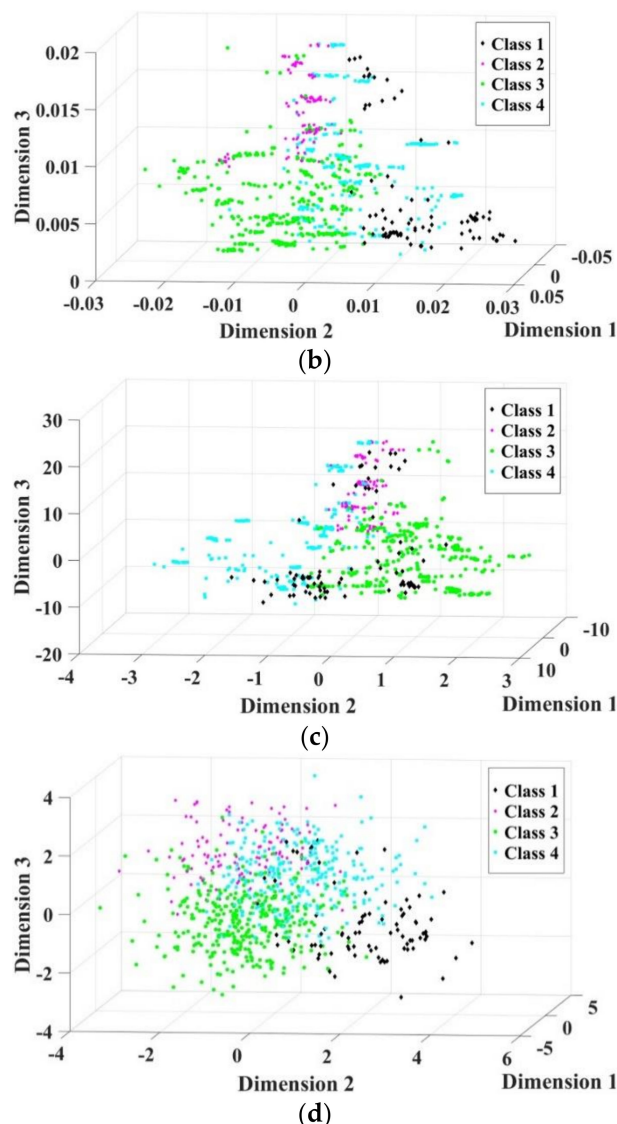
**Table 1.** Comparison among different methods.

Method	Feature Extraction Algorithm	Dimension Reduction Algorithm	Accuracy	Standard Deviation	Processing Time of One Sample
1	PCA	None	87.60%	0.0197	9 ms
2	t-SNE	None	91.53%	0.0112	72 ms
3	Autoencoder	None	93.28%	0.0136	23 ms
4	PCA + t-SNE + Autoencoder	Treelet Transform	98.17%	0.0078	97 ms
5	PCA + t-SNE + Autoencoder	None	98.41%	0.0094	104 ms

Furthermore, since the feature set is reduced to three dimensions, the fault samples could be visualized in three-dimensional space. This paper randomly selects some samples to make the figure clearer. As shown in Figure 8, the same position of three-dimensional axes are imposed in all the figures. The dots of Method 4 have more obvious interfaces than those of Method 1, Method 2, and Method 3. This also indicates that the method proposed in this paper produces great performance.



**Figure 8.** Cont.



**Figure 8.** Distribution of 4 classes of faults. Class 1—convex fault; Class 2—concave fault; Class 3—acceleration fault; and Class 4—deceleration fault. (a) Distribution of Method 4; (b) Distribution of Method 3; (c) Distribution of Method 2; (d) Distribution of Method 1.

## 5. Conclusions

In order to realize the automation of pollutant emission tests, a pedal robot is designed instead of a human-driven vehicle. The actual time-speed curve of the vehicle should follow the WLTC target curve. However, sometimes the actual curve will deviate from the upper or lower limit of the target curve, which will cause a fault. In this paper, a new fault diagnosis method is proposed and applied to the fault diagnosis of pedal robot. There were 346 pollutant emission tests carried out, and 11430 fault samples obtained for fault diagnosis. To find more effective feature information from the fault samples, a deep learning algorithm, Autoencoder, is used for feature extraction. The results show that it is better than other feature extraction algorithms. Since PCA, t-SNE, and Autoencoder cannot extract feature information adequately when they are used alone, and the feature types they extract are different, three types of feature components extracted by PCA, t-SNE, and Autoencoder are fused to form a nine-dimensional feature set. Then, the feature set is reduced into three-dimensional space via Treelet Transform. Finally, the fault samples are classified into four types by using Gaussian process classifier. The experimental results show that the proposed method has the minimum standard deviation, 0.0078, and almost the maximum accuracy, 98.17%. Compared with the method using only one algorithm to

extract features, the proposed method has much higher classification accuracy. This shows that the information loss can be effectively prevented by multi-features fusion.

Moreover, Treelet Transform is used to reduce dimension of feature set, so that the fault samples can be visualized in three-dimensional space. The accuracy of the proposed method is only 0.24% lower than that without Treelet Transform, but the processing time of one sample is 6.73% less than that without Treelet Transform. These clearly prove that Treelet Transform can not only reduce the dimension of data and accelerate the calculation speed, but also retain the feature information and the classification accuracy. It indicates that the hybrid multi-models and Treelet Transform method is quite effective. Therefore, the proposed method can efficiently identify the type of the faults, which can help optimize the control of the pedal robot in pollutant emission test.

**Author Contributions:** Y.Z. designed the experiments and the methodology; Z.F. performed the experiments and analyzed the data; X.C. and Q.W. validated the established model in this work and performed writing—review and editing; Z.F. contributed the project administration and the funding acquisition; Y.Z. wrote the paper.

**Funding:** This work was partially supported by the National Natural Science Foundation of China (Grant nos. 61473192, U1401240 and 61075086); and the Project of Science and Technology Commission of Shanghai Municipality (Grant no. 17441901000).

**Acknowledgments:** We want to gratefully acknowledge Yao Wang for his help and feedback provided on this work.

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

## References

1. Limits and Measurement Methods for Emissions from Light-Duty Vehicles. GB18352.6–2016. 2016, pp. 57–67. Available online: <https://www.chinesestandard.net/PDF/English.aspx/GB18352.6-2016> (accessed on 2 January 2019).
2. Chen, S.; Qi, Z.; Chen, D.H.; Guo, L.F.; Peng, W.W. Investigation of Bayesian network for reliability analysis and fault diagnosis of complex systems with real case applications. *Adv. Mech. Eng.* **2017**, *9*. [CrossRef]
3. Chang, S.; Leng, M.; Wu, H.; Thompson, J. Aircraft ice accretion prediction using neural network and wavelet packet transform. *Aircr. Eng. Aerosp. Technol.* **2016**, *88*, 128–136. [CrossRef]
4. Cheng, Y.; Yuan, H.; Liu, H.; Lu, C. Fault diagnosis for rolling bearing based on sift-kpca and svm. *Eng. Comput.* **2017**, *34*, 53–65. [CrossRef]
5. Thatoi, D.N.; Das, H.C.; Parhi, D.R. Review of techniques for fault diagnosis in damaged structure and engineering system. *Adv. Mech. Eng.* **2012**, *3*, 181–197. [CrossRef]
6. Zhang, M.; Li, K.; Hu, Y. Classification of power quality disturbances using wavelet packet energy and multiclass support vector machine. *Int. J. Comput. Math. Electr. Electron. Eng.* **2013**, *31*, 424–442. [CrossRef]
7. Turk, M.; Pentland, A. Eigenfaces for recognition. *J. Cognit. Neurosci.* **1991**, *3*, 71. [CrossRef] [PubMed]
8. Dunia, R.; Qin, S.J.; Edgar, T.F.; Mcavoy, T.J. Identification of faulty sensors using principal component analysis. *AIChE J.* **2010**, *42*, 2797–2812. [CrossRef]
9. Van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
10. Linderman, G.C.; Steinerberger, S. Clustering with t-SNE, Provably. *arXiv* **2017**.
11. Lecun, Y.; Bengio, Y.; Hinton, G. Review: Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
12. Li, H. Deep learning for natural language processing: Advantages and challenges. *Natl. Sci. Rev.* **2018**, *5*, 24–26. [CrossRef]
13. Vincent, P.; Larochelle, H.; Lajoie, I.; Bengio, Y. Manzagol, stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Mach. Learn. Res.* **2010**, *11*, 3371–3408.
14. Akshay, S.; Maneet, S.; Richa, S.; Mayank, V. Residual codean autoencoder for facial attribute analysis. *Pattern Recognit. Lett.* **2018**, in press.
15. Zhao, C.; Zhang, W.; Yang, Y.; Yao, S. Treelet-based clustered compressive data aggregation for wireless sensor networks. *IEEE Trans. Veh. Technol.* **2015**, *64*, 4257–4267. [CrossRef]
16. Assi, N.; Moskal, A.; Slimani, N.; Viallon, V.; Chajes, V.; Freisling, H. A treelet transform analysis to relate nutrient patterns to the risk of hormonal receptor-defined breast cancer in the european prospective investigation into cancer and nutrition (epic). *Public Health Nutr.* **2016**, *19*, 1–13. [CrossRef] [PubMed]

17. Hassan, T.; Asghari, O.; Emery, X. Application of plurigaussian simulation to delineate the layout of alteration domains in Sungun copper deposit. *Cent. Eur. J. Geosci.* **2013**, *5*, 514–522.
18. Ghahramani, Z. Probabilistic machine learning and artificial intelligence. *Nature* **2015**, *521*, 452–459. [[CrossRef](#)] [[PubMed](#)]
19. Yuan, L.; Bai, Y.S. KNN fault detection based on principle component research. *J. Shenyang Univ. Chem. Technol.* **2018**, *32*, 366–371.
20. Kohavi, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In Proceedings of the International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 20–25 August 1995.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).