International Journal of Engineering Innovation & Research
Volume 13, Issue 5, ISSN: 2277 – 5668

Engine Fault Diagnosis Based on Support Vector Machine and Time-Frequency Domain Analysis

Kunming Gao
College of Automotive Engineering, Zibo Vocational Institute, Shandong, China.
Corresponding author email id: gaokm_zibovin@163.com

Date of publication (dd/mm/yyyy): 31/10/2024

***Abstract* – In order to effectively improve the accuracy of the engine connecting rod fault diagnosis algorithm and ensure more reliable operation of the engine, a fault diagnosis algorithm for engine connecting rods is proposed based on engine related vibration data and combined with support vector machines. Firstly, the vibration data is preprocessed and analyzed in time domain, frequency domain and time-frequency domain; Then, the kurtosis value in time domain, the energy value in frequency domain, the singular value of wavelet packet coefficient matrix and the energy ratio eigenvalue of each frequency band of wavelet packet are extracted, and the eigenmatrix is constructed; Secondly, the eigenvalue matrix is introduced into the linear vector machine for training, and the fault diagnosis model is obtained. Finally, the validity of the fault diagnosis model is verified by the measured sample data on the bench. The results show that the fault diagnosis accuracy of the algorithm is 99.5%, which meets the requirements of fault diagnosis, and the algorithm can accurately identify engine connecting rod faults.**

***Keywords* – Engine, Time-Frequency Analysis, Support Vector Machine, Fault Diagnosis.**

I. Introduction

The engine is the power source for driving a car. Through the crankshaft connecting rod mechanism, the reciprocating motion of the internal piston of the engine can be converted into the rotational motion of the flywheel, and then the power is transmitted to the wheels through the transmission system. The working state of the engine has a significant impact on the driving performance of the car [1-3]. The usual way to handle engine faults is through regular inspection and maintenance, which is inefficient, costly, and lacks foresight. Therefore, researching intelligent fault diagnosis methods is particularly important. Caisen Chen proposed an engine fault diagnosis model based on BA-RVM algorithm, which can predict fault types based on engine parameters [4]. Hua Liang proposes an engine fault diagnosis process based on local edge discrimination projection, which combines feature extraction methods and pattern recognition methods [5].

The time-frequency analysis method can provide more feature information that cannot be provided by time-domain signals by mapping one-dimensional vibration signals to two-dimensional time-frequency surfaces [6, 7]. Chunming Hu conducted time-frequency analysis on detonation combustion based on instantaneous heat release rate, and explored the effect of dual spark plug ignition strategy on detonation combustion in aviation kerosene engines [8]. Based on the obvious order characteristics of vibration noise caused by traditional internal combustion engine vehicles, Long Chen used short-time Fourier transform for speed estimation and combined with order tracking method to analyze the order of transmission vibration signals during acceleration conditions of automobiles [9].

In fault identification and classification, the main methods include deep learning [10], BP neural networks [11, 12], and support vector machines [13, 14], etc. Junhong Zhang proposed a diagnostic method based on an improved convolutional neural network for diesel engine valve fault diagnosis, and verified the accuracy of the method [15].

In order to further improve the accuracy of engine fault diagnosis, based on relevant research, this paper extracts time-domain and frequency-domain analysis feature values using engine vibration data and time-frequency analysis methods, and constructs feature matrices. Finally, a fault diagnosis algorithm model is obtained through support vector machine training.

II. Signal Time-Frequency Analysis and Feature Extraction

The key of fault diagnosis is to extract the characteristic parameters that can reflect the fault cause from the complex vibration signal. This paper will analyze the signal by using the time-domain and frequency-domain analysis methods and extract the fault characteristics.

Time domain analysis is the most fundamental signal analysis method. Due to its simple and intuitive curve characteristics, it is often used for initial fault diagnosis and can quickly analyze the vibration situation of the system. The characteristic parameters of time-domain analysis mainly include peak value, root mean square value, root mean square amplitude, absolute mean, slope, and kurtosis. After obtaining the time-domain characteristic values of some experimental data, kurtosis is used as the characteristic value of time-domain analysis in this paper. The kurtosis formula is as follows,

$$f_1 = \frac{1}{N}\sum_{n=1}^{N} x_n^4 \tag{1}$$

Where x is the signal and N is the number of data points.

Frequency domain analysis mainly decomposes a signal into periodic sub signals for identification. Short time Fourier transform (STFT) is one of the most commonly used linear frequency domain analysis methods, which can represent the signal characteristics at a certain time through a segment of the signal within a time window. In the STFT transformation process, the time and frequency resolution of the spectrum are determined by the length of the selected window, and the formula is as follows,

$$STFT_z(t, f) = \int_{-\infty}^{+\infty} z(\tau)\eta^*(\tau - t)e^{-j2\pi ft}dt \tag{2}$$

In the formula, $z(t)$ represents the signal, $\eta(t)$ is the given window function, t and f represent time and frequency respectively, STFTZ (t, f) represents the energy distribution of signal $z(t)$ at time t and frequency f.

After performing Fourier transform on time-domain signals, the vibration of the signal in the time series is transformed into the vibration in the frequency series. Frequency domain analysis can obtain many information that is difficult to obtain in time-domain analysis. In this paper, energy is used as the characteristic value of frequency domain analysis, and the calculation formula for energy is as follows,

$$f_2 = \sum_{n=1}^{N} x_n^2 \tag{3}$$

Wavelet refers to a cluster of functions formed by $\Psi(t)$ after stretching and shifting. The formula is as follows,

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\Psi\left(\frac{t-b}{a}\right) \tag{4}$$

Where $\Psi(t)$ is a function with oscillatory attenuation and compact support set, parameter *a* is the scale factor, parameter *b* is the translation factor.

The definition of wavelet transform for any signal *x* (*t*) is given by formula (5),

$$(W_\Psi x)(a,b)=\frac{1}{\sqrt{|a|}}\int_{-\infty}^{+\infty}x(t)\Psi^*\left(\frac{t-b}{a}\right)dt \tag{5}$$

In the formula, $\Psi^*(t)$ represents the conjugate of $\Psi(t)$.

It can be proven that only when the wavelet function satisfies the permissible conditions, that is

$$C_\Psi=\int_{-\infty}^{+\infty}\frac{|\Psi(\omega)|^2}{\omega}d\omega<\infty \tag{6}$$

In the equation, $\Psi(\omega)$ represents the Fourier transform of $\Psi(t)$.

When the above equation holds, the original signal *x*(*t*) can be reconstructed using wavelet transform $(W_\Psi x)\ (a,b)$. At this point,

$$x(t)=\frac{1}{C_\Psi}\int_{-\infty}^{+\infty}\frac{1}{a^2}da\int_{-\infty}^{+\infty}(W_\Psi x)(a,b)\Psi_{a,b}(t)db \tag{7}$$

This article uses the DB10 wavelet basis to perform 3-level wavelet packet decomposition on vibration data, extracting the singular values of the wavelet packet coefficient matrix and the energy percentage of each frequency band for 8 frequency bands as feature values in time-frequency analysis.

III. Fault Identification

A. *Fault Identification Method*

Support Vector Machine (SVM) is a general machine learning method based on statistical learning theory. It has many advantages in solving small sample, nonlinear, and high-dimensional problems, and has good generalization ability. The main idea of SVM is to establish a classification hyperplane in the input space to maximize the distance between two sample sets and the classification hyperplane. Figure 1 is a schematic diagram of support vector machine classification, where the prototype and square represent two types respectively. The samples on H1 and H2 are support vectors.

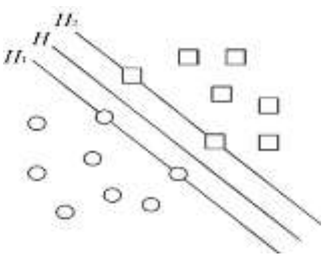

Fig. 1. Optimal parameter.

Finally, it can be converted into the quadratic programming problem shown in the solution formula, as follows,

International Journal of Engineering Innovation & Research
Volume 13, Issue 5, ISSN: 2277 – 5668

$$\begin{cases} \min \frac{1}{2}(\omega \times \omega) + C\sum_{i=1}^{n} \varepsilon_i \\ s.t. y_i\left[(\omega \times x_i) + b\right] \geq 1 - \varepsilon_i \\ \qquad \varepsilon_i \geq 0; i = 1, 2, \cdots, n \end{cases} \tag{8}$$

In the formula, C and ε_i are the penalty parameter and relaxation variable, respectively.

By introducing Lagrange multiplier $\alpha_i \left(i = 1, 2, \cdots, n\right)$, equation (8) can be transformed into the form of equation (9),

$$\begin{cases} \max \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} y_i y_j \alpha_i \alpha_j \\ s.t. \sum_{i=1}^{n} \alpha_i y_i = 0 \quad 0 \leq \alpha_i \leq C, i = 1, 2, \cdots, n \end{cases} \tag{9}$$

The decision function shown in the equation can be obtained by solving it, that is

$$\left\{ f(x) = \operatorname{sgn}\left(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b\right) \right. \tag{10}$$

In the equation, $K(x_i, x)$ represents the kernel function. Generally in the form shown below,

$$K(x_i, x) = \exp\left(-g\left\|x_i - x\right\|^2\right) \tag{11}$$

B. *Fault Identification Process*

Figure 2 shows the specific diagnostic process, which mainly includes the following steps.

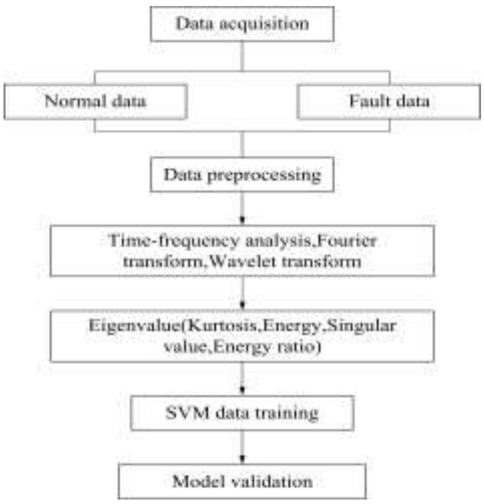

Fig. 2. Optimal parameter.

(1) Perform preliminary processing on the normal and fault data collected from the experimental bench, ensuring that the collection time and length of each set of data are consistent.

(2) Pre-process the data and perform time-domain signal reading, Fourier transform, and wavelet analysis.

(3) Extract feature values, including kurtosis values of time-domain signals, energy values of frequency-domain signals, singular values of wavelet packet coefficient matrices, and energy percentages of each frequency band.

(4) Construct a feature matrix with eigenvalues as elements.

(5) Train the signal using support vector machines, obtain a fault diagnosis model, and import the data for training to verify the accuracy of the training model.

IV. Experimental Verification

A. *Data Acquisition*

In order to verify the effectiveness of this method, this article starts with data collection and utilizes laboratory bench resources to collect vibration signals of the engine using a self-developed engine vibration signal acquisition device. Firstly, data is collected when the engine is in a normal state. Then, data is collected when the engine is in a faulty state under the same operating conditions, with the engine running at 750r/min, 0 torque, and a sampling frequency of 102400Hz. A three-axis sensor is used for data collection, and the sensor is installed on the flat part of the gear chamber. Through data analysis, it is found that the Y-axis direction of the three-axis sensor is the main direction of vibration. Collect 1200 sets of data from the experimental bench, including 1000 sets of training data and 200 sets of test data. The 1000 sets of training data consist of 500 sets of normal data and 500 sets of fault data, while the 200 sets of test data consist of 100 sets of normal data and 100 sets of fault data.

B. *Time Domain Analysis*

The training data is analyzed in time domain, and Fig. 3 and Fig. 4 are time domain waveform diagrams of normal data 1 and fault data 1 in the training data,

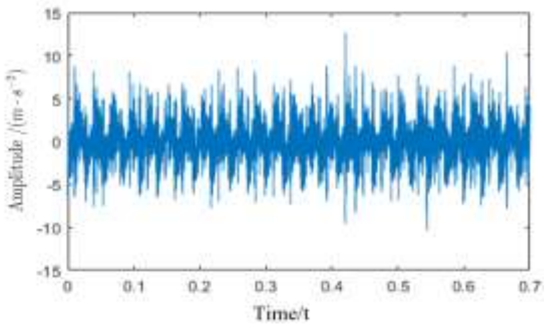

Fig. 3. Normal condition time domain graph.

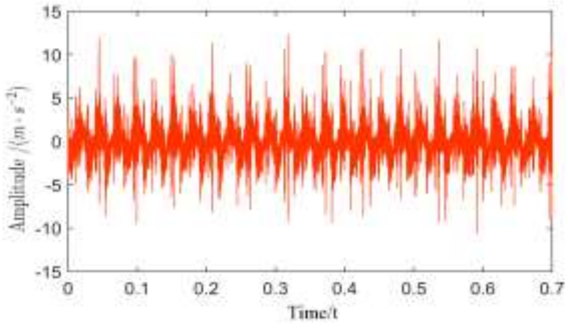

Fig. 4. Fault condition time domain graph.

From the time-domain graph, it can be seen that the vibration acceleration amplitude of engine faults exceeding 10g is significantly higher than that of normal vibration acceleration. In order to obtain more intuitive information from the time-domain graph, time-domain feature analysis is conducted. The time-domain feature value is the kurtosis value, which represents the steepness of the probability density of vibration amplitude. The more obvious the fault, the larger the kurtosis value.

The kurtosis value of the time-domain signal is solved for 1000 sets of training data, and the solution results are shown in Figure 5.

International Journal of Engineering Innovation & Research
Volume 13, Issue 5, ISSN: 2277 – 5668

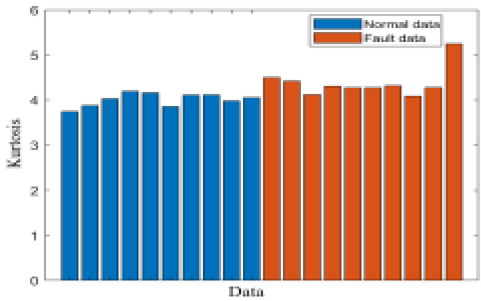

Fig. 5. Bar chart of kurtosis values.

From the graph, it can be seen that the kurtosis values of normal data are mostly smaller than those of fault data, and kurtosis values can be used as feature values to distinguish between normal data and fault data.

C. *Frequency Domain*

Frequency domain analysis decomposes the correspondence between time and amplitude into the correspondence between frequency and amplitude, achieving the transformation of the original signal from time domain to frequency domain. Information that cannot be obtained from the time domain signal can be mined from the frequency domain curve. The frequency domain curves of normal data one and fault data one are shown in Figure 6 and Figure 7,

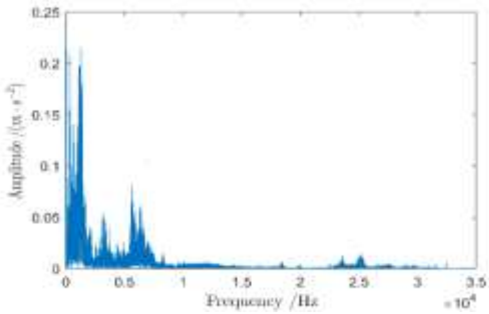

Fig. 6. Normal condition frequency domain graph.

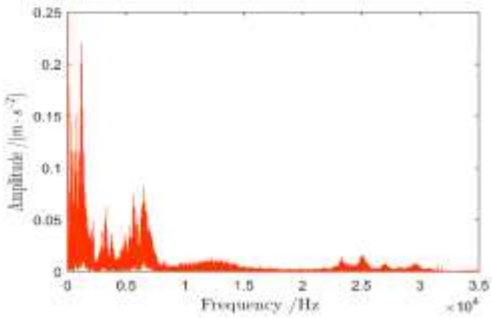

Fig. 7. Fault condition frequency domain graph.

From the frequency domain graphs of normal and fault data, it can be seen that the vibration amplitude of fault data is significantly greater than that of normal data between frequency values of 8000Hz and 35000Hz. The characteristic value of frequency domain analysis is the energy value, which characterizes the overall vibration amplitude. Generally, the more obvious the fault, the larger the energy value. The energy values of frequency domain signals between 8000Hz and 35000Hz were calculated using 1000 sets of training data, and the results are shown in Figure 8,

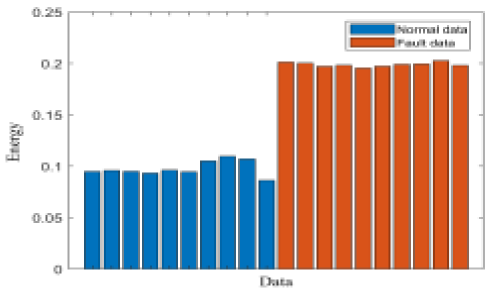

Fig. 8. Bar chart of energy values.

International Journal of Engineering Innovation & Research
Volume 13, Issue 5, ISSN: 2277 – 5668

From Figure 8, it can be seen that the root mean square value of normal data is significantly smaller than that of fault data, and the root mean square value can be used as a characteristic value to distinguish between normal data and fault data.

D. *Time-Frequency Analysis*

Time frequency analysis can analyze non-stationary signals that are difficult to analyze in both time and frequency domains. DB10 wavelet basis is used to perform 3-level wavelet packet decomposition on vibration data, extracting singular values of wavelet packet coefficient matrix and energy percentages for 8 frequency bands. The calculation results of singular values and energy percentages of wavelet packet coefficient matrix are shown in the following figure,

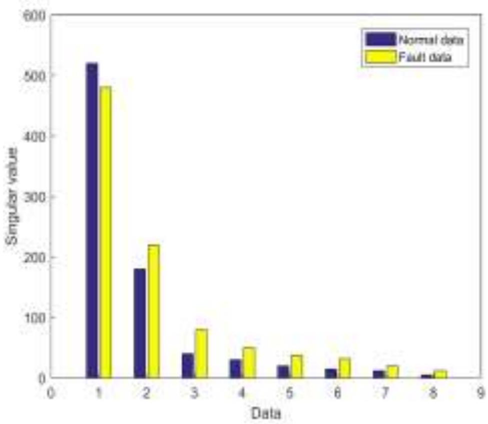

Fig. 9. Singular value of wavelet packet coefficient matrix.

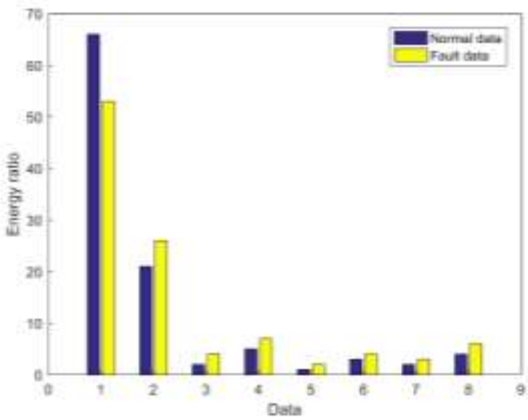

Fig. 10. Energy ratio of each frequency band of wavelet packet.

From the singular value histogram 9 of the wavelet packet coefficient matrix, it can be seen that the singular values of the wavelet packet coefficients in the 8 frequency bands of normal data and fault data are greater in the first frequency band than in the fault data. The singular values of the wavelet packet coefficients in the normal data from the second to the eighth frequency band are smaller than in the fault data, showing a certain regularity. Therefore, the singular values of the wavelet packet coefficient matrix can be used as characteristic values to distinguish between normal data and fault data.

The regularity observed from the energy ratio bar chart 10 of each frequency band in the wavelet packet is consistent with the singular value bar chart of the wavelet packet coefficient matrix. Therefore, the energy ratio of each frequency band in the wavelet packet can be used as a characteristic value to distinguish between normal data and fault data.

E. *Diagnostic Model and Validation*

The support vector machine fault diagnosis model of engine vibration data is established by taking 1000 sets of training data selected from the diesel engine vibration experiment under steady state conditions as training samples. The eigenvalues of normal data 1 and fault data 1 are shown in Table 1 and Table 2 respectively. The trained engine vibration data is used to support the vector machine fault diagnosis model for fault diagnosis of the test data. The verification results of the fault diagnosis model are shown in Figure 11.

International Journal of Engineering Innovation & Research
Volume 13, Issue 5, ISSN: 2277 – 5668

Table 1. Normal data eigenvalue.

Kurtosis Value	Energy Value	Singular Values of Coefficient Matrix	Energy Ratio of Each Frequency Band
3.863	0.095	523.29	65.97
		184.66	21.50
		63.21	3.39
		48.78	7.12
		30.74	1.08
		30.33	3.72
		16.51	1.91
		9.54	5.32

Table 2. Fault data eigenvalue.

Kurtosis Value	Energy Value	Singular Values of Coefficient Matrix	Energy Ratio of Each Frequency Band
4.575	0.208	473.93	52.20
		225.92	25.24
		64.42	3.41
		48.52	7.22
		30.87	1.06
		30.47	3.55
		16.59	1.91
		9.43	5.40

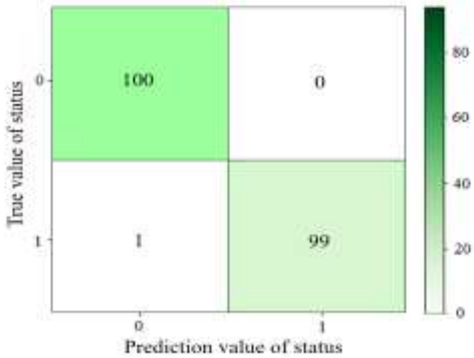

Fig. 11. Energy ratio of each frequency band of wavelet packet.

The gradient color coordinate values in Figure 11 represent the size of the test data, with 0 representing the normal state and 1 representing the fault state. The (0, 0) coordinate value represents that the true value of the state is normal, and the predicted value of the state is also normal. The (1, 1) coordinate value represents the true value of the state as a fault, and the predicted value of the state is also a fault, so the coordinates (0, 0) and (1, 1) represent accurate fault diagnosis. (0, 1) The coordinate value represents that the true value of the state is a fault, but the predicted value of the state is normal, which is a missed alarm; The (1, 0) coordinate value represents that the true value of the state is normal, and the predicted value of the state is a fault, which is a false alarm.

From the verification results of the fault diagnosis model, it can be seen that the diagnostic accuracy of the support vector machine fault diagnosis model obtained through training data in this paper is 99.5%.

V. Conclusion

This article is based on engine vibration data under idle conditions, and combines time-domain analysis method, Fourier frequency domain analysis method, and wavelet transform time-frequency analysis method to preprocess the experimental data. Ten feature values are extracted, including time-domain kurtosis value, frequency-domain energy value, wavelet packet coefficient matrix singular value, and wavelet packet energy ratio of each frequency band. The feature matrix is constructed and imported into a linear support vector machine for training to obtain a fault diagnosis model. The verification results through test data show that the fault diagnosis model can fully recognize fault data, and the work in this paper can provide some reference for the design of fault diagnosis methods.

References

[1] Yaohua Li, Zhang Cheng. Engine performance degradation prediction based on adaptive change point [J]. Journal of Transportation Engineering, 2023, 23 (05): 143-151.
[2] Taoming Wan, Chen Guixin, He Guanzhang, et al. Advanced machine learning methods for engine performance prediction [J]. Journal of Guangxi University (Natural Science Edition), 2023, 48 (03): 594-604.
[3] Hao Duan, Chen Hui, Zhai Zhaoyang, et al. Performance prediction of hydrogen mixed natural gas engine based on support vector machine [J]. Journal of Ordnance Engineering, 2022, 43 (05): 1002-1011.
[4] Caisen Chen, Hu Hairong, Cheng Zhiwei, et al. Research on engine fault diagnosis technology based on BA-RVM Algorithm [J]. Computer Engineering and Science, 2023, 45 (02): 332-337.
[5] Hua Liang, Lv Liping, Wang Chengyong. Engine Fault Diagnosis Method Based on Local Edge Discriminant Projection [J]. Noise and Vibration Control, 2023, 43 (03): 90-94+109.
[6] Shaobo Ji, Jiang Ying, Yin Wei, et al. Correlation analysis between time-frequency characteristics of body vibration signals and combustion pressure [J/OL]. Journal of Harbin Engineering University: 1-7.
[7] Yang Liu, Liu Xiaobo, Liang Shan. Fault diagnosis of aircraft engine rotors based on Fourier decomposition method [J]. China Mechanical Engineering, 2019, 30 (18): 2156-2163.
[8] Chunming Hu, Zhao Yingbo, Liu Na, et al. Investigating the effect of ignition on kerosene engine detonation based on time frequency analysis [J]. Journal of Aerodynamics, 2019, 34 (12): 2635-2641.
[9] Long Chen, Shi Wenku, Zhang Shuguang, et al. analysis of transmission vibration order ratio based on improved peak search method [J]. Vibration, Testing and Diagnosis, 2020, 40 (06): 1071-1076+1228.
[10] Ping Yu, Cao Jie. Application of deep learning in fault diagnosis and prediction [J]. Computer Engineering and Applications, 2020, 56 (03): 1-18.
[11] Xinfa Shi, Guangxiao Xing, Shizhong He, etc. Identification of wear faults in marine diesel engines based on entropy theory and BP Neural Network [J] Lubrication and Sealing, 2022, 47 (07): 54-58.
[12] Haoqian Pang, Pan Zhang, Wen Wang, etc. Engine partial misfire fault diagnosis based on denoising residual neural network [J] Internal Combustion Engine Engineering, 2023, 44 (03): 91-100
[13] Chunli Xie, Xuezhi Sun, Shenghan Gao. Fault diagnosis of electronic control engine based on support vector machine [J] Journal of Hubei University of Automotive Industry, 2021, 35 (02): 48-52+57.
[14] Chao Wu, Lei Chen, Yuan Liu, etc Fault diagnosis of aircraft engine air path based on feature optimization and support vector machine [J] Aircraft Engine, 2024, 50 (04): 30-37.
[15] Junhong Zhang, Sun Shiyue, Zhu Xiaolong, et al. Research on diesel engine fault diagnosis method based on improved convolutional neural network [J]. Vibration and Shock, 2022, 41 (06): 139-146.

Author's Profile

Kunming Gao, Master of Engineering, Male, Assistant, College of Automotive Engineering, Zibo Vocational Institute, Shandong, China, He has published 10 articles in recognized journals.