processes

MDPI

Article

Diesel Engine Fault Diagnosis Method Based on Optimized VMD and Improved CNN

Xianbiao Zhan, Huajun Bai, Hao Yan, Rongcai Wang, Chiming Guo and Xisheng Jia *

Shijiazhuang Campus, Army Engineering University of PLA, Shijiazhuang 050003, China
* Correspondence: xsjia123@sina.com

Abstract: The safe operation of diesel engines performs a vital function in industrial production and life. Because diesel engines often work in harsh environmental conditions, they are prone to failure. Therefore, this paper proposes a fault analysis method based on a combination of optimized variational mode decomposition (VMD) and improved convolutional neural networks (CNN) to address the necessary need for preventive maintenance of diesel engines. The authentic vibration sign is first decomposed by using the (VMD) algorithm, then the greatest range of decomposition layers is decided by using scattering entropy and the useful components are preferentially chosen for reconstruction. The continuous wavelet transform (CWT) records preprocessing method is then delivered to radically change the noise-reduced vibration sign into a time-frequency map, which is fed into the CNN for model coaching and extraction of fault features. Finally, fault classification is realized by support vector machine (SVM) with excellent classification performance. Through preset fault experiments on diesel engines, it is established that the technique proposed in this paper can successfully identify fault states, and the classification accuracy is higher than alternative methods.

Keywords: convolutional neural network; variational mode decomposition; diesel engine fault diagnosis; continuous wavelet transform; support vector machine

Citation: Zhan, X.; Bai, H.; Yan, H.; Wang, R.; Guo, C.; Jia, X. Diesel Engine Fault Diagnosis Method Based on Optimized VMD and Improved CNN. *Processes* **2022**, *10*, 2162. https://doi.org/10.3390/pr10112162

Academic Editors: Hongyun So, Jong-Won Park and Sunghan Kim

Received: 20 September 2022
Accepted: 18 October 2022
Published: 22 October 2022

1. Introduction

Diesel engines are widely used in the fields of national defense, agriculture, petrochemicals and ships, providing the main power for mechanical equipment. Once a failure occurs, it may cause system downtime and even lead to major safety incidents [1]. Therefore, it is of great importance to find out about the fault diagnosis technology of diesel engines to ensure the protected operation of the diesel engines. The raw vibration signals cannot directly characterize engine faults, thus they need to be processed in order to extract features from different domains, which are either calculated by dimensionality reduction or fed directly into machine learning algorithms for classification. Common signal processing strategies include: empirical modal decomposition (EMD), short time Fourier transform (STFT) and wavelet transform (WT) [2]. Traditional machine learning diagnosis methods include: support vector machine (SVM), decision tree (DT) and naive Bayesian model (NBM) [3]. Kumar et al. [4] first analyzed the bearing signals using fast Fourier transform (FFT). Fault features were then extracted and fed into a particle swarm optimization SVM model for diagnosis and compared with other methods, concluding that the proposed method outperformed other existing algorithms. Mathew et al. [5] used the fundamental aspect evaluation algorithm to limit the dimension of the extracted wavelet packet seriously changed coefficient features, and then used the Bayesian optimization model for classification. The experiment proved that this approach can achieve higher diagnostic outcomes and lower testing time. Yang et al. [6] proposed a fault prognosis approach combining EMD and SVM. Firstly, the EMD algorithm was once used to decompose the sign and then input to SVM for classification, which achieved better classification results. However, EMD [7] has the problem of end-point impact and mode aliasing. Subsequent scholars

have proposed improved algorithms for the disadvantages of EMD, such as: ensemble empirical mode decomposition (EEMD) [8], local mean decomposition (LMD) [9] and other algorithms. EEMD is an improvement of EMD. Although the trouble of modal aliasing is solved to a certain extent, the calculation efficiency is reduced after adding noise. The LMD decomposition overcomes the EMD over-decomposition and other problems, but there are still end-point effects and modal aliasing. In view of this, Konstantin et al. [10] proposed the variational mode decomposition (VMD) algorithm in 2014. VMD is a non-stationary and nonlinear signal processing method. It overcomes the shortcomings of EMD, EEMD and LMD and is widely used in the fields of fault diagnosis and life prediction. Bi et al. [11] decomposed the authentic vibration signal through VMD, and then classified the characteristic parameters of the intrinsic mode function based totally on the improved fuzzy C-means clustering algorithm (KFCM). The proposed approach has benefits in phrases of accuracy and efficiency. Li et al. [12] proposed a method based on VMD and piecewise Fourier transform for the weak early signal of bearing fault that is difficult to extract, which can provide effective filtering for fault frequency and help fault diagnosis. Zheng et al. [13] proposed a fault diagnosis method based on VMD, Hilbert-Huang and generalized learning models in order to be able to accurately identify the failure modes of bearings, and obtained a high accuracy rate. The above-mentioned literature using VMD to decompose the original vibration signal are based on human experience to determine the value of the decomposition layer k. Although it is relatively simple to determine the k-value through experience, it is prone to phenomena such as over-decomposition and under-decomposition, and has certain limitations. Qiao et al. [14] proposed a bearing fault diagnosis method based on VMD and improved SVM algorithm, which mainly determines the k-value by the center frequency and then calculates the sample entropy of each IMF component as the input feature of the improved SVM algorithm. Experiments have shown that the method proposed in this paper has good diagnostic performance. In addition, some other researchers have used algorithms such as the fruit fly optimization algorithm [15], the sparrow search algorithm (SSA) [16] and whale optimization (WOA) [17] to determine the optimal number of decomposition k layers. Although better results were achieved by using various optimization algorithms to determine the k-value, the optimization algorithms need to determine many parameters, such as: batching and number of iterations, and the selection of batching and number of iterations will seriously affect the decomposition efficiency of VMD [18]. In view of this, this paper uses a VMD method based on scattering entropy improvement to determine the optimal number of decomposition k layers. The original vibration signal is decomposed by VMD, and the optimal number of decomposition k layers and useful components are determined by calculating the scatter entropy of each component, and then the useful components selected are reconstructed to extract features for classification.

The aforementioned scholars have done a lot of research on signal processing in the early stages, and good signal preprocessing has laid a good foundation for subsequent extraction of weak fault features and classification. Traditional fault feature extraction methods mainly extract the features of the time-frequency domain, but such feature extraction methods are prone to cause errors and poor generalization. In addition, traditional shallow machine learning algorithms are not effective in learning the non-linear relationships of the system [19]. In current years, with the upward trend of deep learning, the use of convolutional neural networks (CNN) and other algorithms to extract fault feature information has gradually become a research hotspot [20]. Because the CNN has terrific feature extraction ability, it can automatically extract high-dimensional information features. Through visual displays such as t-SNE, it can be viewed that with the increase of the variety of community layers, the feature expression ability and classification capacity are also improved layer by layer. Houssem et al. [21] applied the VMD algorithm to process the vibration signal, and then used CNN for classification and diagnosis. Furthermore, verified on public datasets, the results show that the algorithm can reduce redundant information, have better diagnosis results and lower diagnosis delay. Wang et al. [22] used a multi-sensor fusion approach. Features are first extracted from the vibration and sound indicators of the bearings sepa-

rately. Then they are input into CNN for fusion and classification. By examining the loss characteristic and accuracy underneath extraordinary signal-to-noise ratios, it is concluded that the fault features of this paper combining vibration and sound signals are richer and more beneficial for fault diagnosis, and its diagnosis accuracy is higher. Gong et al. [23] proposed an improved CNN algorithm for fault diagnosis of CNN networks, which are prone to overfitting and computational time-consumption. This method abandoned the preceding utterly-connected layer and replaced it with a world common pooling layer; the experimental outcomes exhibit that this approach can reap higher classification effect and reduce the risk of overfitting.

The models studied by the above scholars all use the one-dimensional CNN network model, but CNN has more advantages in processing two-dimensional data, and it is easier to extract feature information from images [24,25]. Therefore, researchers began to convert the authentic sign into a picture for fault diagnosis. Xiao et al. [26] proposed a fault diagnosis method based on the combination of improved variational mode decomposition (IVMD) and CNN, where IVMD mainly refers to the optimization of VMD by determining the optimal k-value of VMD decomposition based on the decomposition of the traditional VMD method and then using the Pearson correlation coefficient principle. The processed signal is fed into the CNN for classification by obtaining a two-dimensional time-frequency map through continuous wavelet transform (CWT). The method not only removes the redundant interference in the signal and retains the main features of the fault, but also has a high accuracy rate. Nishat et al. [27] used the EMD algorithm to decompose the accumulated vibration signals, and then performed spectrum analysis. Finally, the received spectrum was first compressed and input into the CNN classification network for coaching and classification in order to confirm the generalization and classification of the proposed method. To enhance the anti-noise ability, white noise is delivered to the unique signal. The experimental outcomes show that the proposed approach nevertheless has an excessive accuracy. Liang et al. [28] combined WT and CNN without delay to use the raw vibration sign of the gearbox for end-to-end fault diagnosis, and used two instances for experimental verification. The outcomes exhibit that the proposed hybrid approach for analysis is higher than other strategies in the literature. The method has greater precision and greater stability. Although the use of CNN models has achieved many successful applications in the diagnostic field, there are still some problems:

1. Because diesel engines frequently work under complex environmental conditions, the fault vibration sign is rather vulnerable and often blanketed through robust noise and interference signals, thus it is challenging to extract fault records from combined signals.
2. The excellent mechanical fault classification potential of CNN is primarily based on a giant number of coaching samples; however, it is tough to achieve fault samples in engineering practice.

Through the above research analysis, this discovery proposes a diesel engine fault analysis method for small pattern training. Firstly, the authentic vibration signal is decomposed using VMD, the optimal quantity of decomposition layers is decided by using scattering entropy and the beneficial components are preferentially selected for reconstruction, then the noise reduced vibration signal is converted into a time-frequency map using CWT and finally the CNN is used for feature extraction of the time-frequency map, but this technique discards the Softmax layer and uses SVM, which has better classification performance, as the classifier to construct an improved CNN fault diagnosis method for diesel engines. Experimental outcomes exhibit that the proposed method in this paper can extract engine fault points quickly and accurately with a high accuracy rate. The main contributions of this paper are as follows.

1. The study uses scattering entropy to decide the most fulfilling quantity of layers and to choose useful elements for VMD decomposition, and the optimized VMD excludes the noise and interference indicators from the unique vibration signal, for that reason efficaciously decreasing noise.

2. The reconstructed data after noise discounting is transformed into a two-dimensional time-frequency photograph using CWT, which can correctly signify the non-smoothness of the vibration signal and incorporates richer fault features, which is more conducive to function extraction with the aid of CNN.
3. This lookup method not only utilizes the exceptional characteristic studying capability of CNN, but also the wonderful classification functionality of SVM, which avoids the errors caused by way of guide characteristic extraction and improves the accuracy.

The rest of this paper is organized as follows: Section 2 describes the denoising method of VMD decomposition, the theoretical knowledge of scatter entropy, the transformation method of time-frequency map using CWT and the theoretical knowledge of CNN and SVM algorithms. The fault diagnosis method of diesel engines is described in Section 3; Section 4 presents the experimental analysis and method comparison; and Section 5 is the conclusion.

2. Theory Backgrounds

2.1. Variational Mode Decomposition

The vibration signal of diesel engines has the characteristics of non-stationarity and nonlinearity. VMD decomposition is a processing method for non-stationary and nonlinear signals. It solves modal aliasing, end effect and sensitivity to noise of EMD, EEMD and LMD. It is extensively used in the discipline of vibration signal processing and can decompose the signal adaptively. The authentic time collection $f(t)$ can be decomposed into exclusive factors $u_k(t)$ with restricted bandwidth by way of an iterative search variational model, and the corresponding middle frequency is ω_k. The bandwidth of every component can be estimated by way of the following steps.

1. Performing Hilbert seriously changes every vibration element to reap a unilateral spectrum.
2. Adjust the exponent of every estimated middle frequency, and switch the spectrum of each vibration issue to the baseband region.
3. Calculate the L^2-norm of the demodulated sign gradient to achieve the bandwidth of every vibration component.

Then the vibration output signal VMD of the diesel engine can be expressed by the following formula:

$$\begin{cases} \min\left\{ \sum_{k=1}^{k} \left\| \partial_t((\delta(t) + \frac{j}{\pi t}) * u_k(t)) \exp(-j\omega_k t) \right\|_2^2 \right\} \\ s.t. \sum_{k=1}^{k} u_k = f(t) \end{cases} \tag{1}$$

In the formula: ∂_t represents the partial derivative; $\delta(t)$ is the Dirac distribution function; "*" represents the convolution operation, k is the total number of components; and $f(t)$ is the original planetary gearbox output signal.

The above restricted extremum issue is transformed into an unconstrained problem to be cleared up by the use of Lagrangian multiplier λ and quadratic penalty term a, as shown in Equation (2):

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_{k=1}^{k} \left\| \partial_t((\delta(t) + \frac{j}{\pi t}) * u_k(t)) \exp(-j\omega_k t) \right\|_2^2 + \left\| f(t) - \sum_{k=1}^{k} u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^{k} u_k(t) \right\rangle \tag{2}$$

Each factor u_k and the corresponding central frequency ω_k can be optimally solved with the aid of the alternating direction multiplier approach, which is updated as follows.

$$\begin{cases} \hat{u}_k^{n+1}(\omega) = \dfrac{\hat{f}(\omega) - \sum\limits_{i \neq k} \hat{u}_k^n(\omega) + \dfrac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \\ \omega_k^{n+1} = \dfrac{\int_0^\infty \omega \left|\hat{u}_k^{n+1}(\omega)\right|^2 d\omega}{\int_0^\infty \left|\hat{u}_k^{n+1}(\omega)\right|^2 d\omega} \end{cases} \tag{3}$$

In the formula: $\hat{f}(\omega)$, $\hat{u}_k^n(\omega)$ and $\hat{\lambda}(\omega)$ are the Fourier transform of $f(t)$, $u_k(t)$ and $\lambda(t)$, respectively; ω is the frequency; and n is the number of iterations.

It can be seen from the above VMD algorithm that VMD decomposes the original vibration signal into k components. The traditional selection of k values is mainly based on human experience, which has certain limitations. If the k-value is too large, it will produce over decomposition, and if the k-value is too small, it will not be able to decompose the useful signal. In view of this, an improved VMD algorithm based on dispersion entropy is proposed.

2.2. Dispersion Entropy

The complexity of the time sequence can be measured in phrases of statistics entropy. Common information entropies are: pattern entropy [29], permutation entropy [30] and scattering entropy [31]. Among them, pattern entropy is much less environment friendly to calculate, and although permutation entropy displays the complexity of the signal and the non-linear characteristics of the signal, permutation entropy ignores the distinction characteristics of the amplitude. In view of this, Rostaghi and Azami proposed the scattering entropy algorithm in 2016, which is a new algorithm to measure the complexity of time series, with a quicker calculation velocity compared to sample entropy and better stability through thinking about the suggested cost of amplitude and the variability between amplitudes, with the following calculation procedure.

Step 1: Use the normal distribution function:

$$\mathrm{y}_j = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_j} e^{\frac{-(t-\mu)^2}{2\sigma^2}} \tag{4}$$

Map the time series x to $y = \{y_1, y_2, \cdots, y_N\}, y_j \in (0,1)$, where μ and σ^2 are the expectation and variance of the time series, respectively.

Step 2: Mapping y to an integer in the range $[1, c]$ yields:

$$z_j^c = \mathrm{int}(cy_j + 0.5) \tag{5}$$

In the formula, c is the number of categories, and int is the rounding.

Step 3: Scatter pattern $w_{v_0 v_1 \cdots v_{m-1}} (v = 1, 2, \cdots, c)$, and calculate the probability of all scatter patterns:

$$p(w_{v_0 v_1 \cdots v_{m-1}}) = \frac{num(w_{v_0 v_1 \cdots v_{m-1}})}{N - (m-1)d} \tag{6}$$

where $z_i^c = v_0, z_{i+d}^c = v_1, \cdots, z_{i+(m-1)d}^c = v_{m-1}$; $num(w_{v_0 v_1 \cdots v_{m-1}})$ is the number of $z_i^{m,c}$ mappings to scatter patterns, m is the embedding dimension and d is the experimental delay.

Step 4: According to the definition of information entropy, DE of signal x is defined as:

$$DE(x, m, c, d) = -\sum_{w=1}^{c^m} p(w_{v_0 v_1 \cdots v_{m-1}}) In(p(w_{v_0 v_1 \cdots v_{m-1}})) \tag{7}$$

According to the above scattering entropy formula, the original vibration signal is decomposed in VMD to derive each IMF component, and then the scattering entropy value of each IMF component is calculated using the scattering entropy formula, and the optimal

number of decomposition layers is determined when the scattering entropy value takes a turn, thus achieving the optimal selection of k values.

2.3. Continuous Wavelet Transform

The image can comprise richer fault function information, thus it is vital to seriously change the one-dimensional vibration sign into a time-frequency map. Short-time Fourier transform (STFT) can convert the original fault vibration sign into a time-frequency map [32], however the window feature in STFT has a constant size and the choice of the window battles to satisfy both high frequency decision and excessive time resolution. Therefore, in the image classification, it is challenging to attain a higher classification effect using the STFT method. The emergence of the wavelet solves the hassle of constant window function, because the wavelet results in no longer using the concept of a window, but directly replaces the trigonometric feature foundation with the wavelet basis. This resolves the conflict between time and frequency resolution. The continuous wavelet transform is:

$$T_{cw,x}(u,v) = \int_{-\infty}^{+\infty} x(t)\varphi_{u,v}(t)dt = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} x(t)\varphi(\frac{t-\tau}{a})dt \quad (8)$$

Among them, the place a represents the scale parameter, τ represents the time parameter, $x(t)$ represents the authentic signal and $\varphi(t)$ is the wavelet basis function. The wavelet base selected in this paper is Morse wavelet, because many commonly used analytical wavelets are special cases of the generalized Morse wavelet. The Morse wavelet can obtain analytical wavelets with different properties and behaviors by adjusting the time-bandwidth product and symmetric parameters. Its expression is:

$$\varphi_{u,v}(t) = \frac{1}{\sqrt{a}}\varphi(\frac{t-\tau}{a}), v > 0, u \in \mathrm{R} \quad (9)$$

The reason why wavelet transform can solve the conflict between time and frequency resolution is mainly because the scale parameter a and translation parameter τ in wavelet transform can be automatically adjusted according to the properties of the signal.

2.4. Convolutional Neural Network

The CNN model uses multiple processing layers to process the input statistics [33], and has emerged as a frequent approach for feature extraction in deep studies [34]. CNN can automatically extract and integrate signal features, particularly in photo classification, which has been widely used. The common simple shape of the CNN is proven in Figure 1.

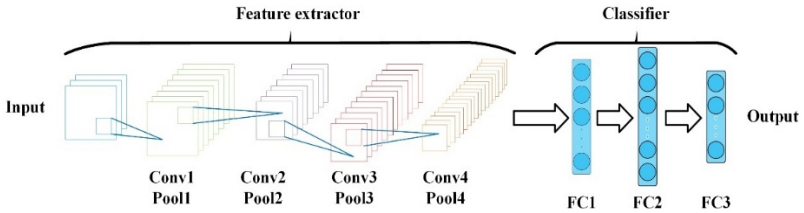

Figure 1. Structural diagram of a convolutional neural network.

2.4.1. Convolutional Layer

In the convolution layer, the top layer characteristic map is first convolved with the aid of the convolution kernel and then the next layer function map is obtained by means of the activation function. By extracting local features from the entered data, the range of community parameters is decreased and the complexity of the model is reduced. The convolution method can be described as follows:

$$x_j^l = f(\sum_{i \in M_j} x_i^{l-1} k_{ij}^l + b_j^l) \quad (10)$$

Among them, l is the l layer convolutional layer; i, j represents the feature map number in the l layer and the $l-1$ layer respectively; x_j^l is the output of the l layer; M_j is the feature set of the $l-1$ layer, k_{ij}^l is the weight matrix, b_j^l is the bias and $f(\,)$ is the activation function.

2.4.2. Pooling Layer

In the pooling layer, downsampling is carried out by using the potential of pooling kernels, which serve to reduce the computational complexity of the community and extract useful fault features. This can be expressed in terms of the following equation:

$$x_j^l = f(\beta_j^l down(x_j^{l-1}) + b_j^l) \tag{11}$$

where $down(\,)$ is the downsampling function and β is the weight of the network.

2.4.3. Fully Connected Layer

The image features are entered into the entirely linked layer at the end of the CNN after multi-layer convolution and pooling. In the fully linked layer, the neurons in the post-layer layer and the neurons in the front layer are connected one by one to distinguish the deep feature facts and to construct the extracted elements and labels of the mapping relationship between them. This can be expressed in terms of the following equation:

$$y^k = f(w^k x^{k-1} + b^k) \tag{12}$$

Among them, x^{k-1} is the input of the fully connected layer, k is the network of the k layer, y^k is the output of the fully connected layer, w^k is the weight coefficient and b^k is the bias.

2.5. Support Vector Machines

SVMs were proposed by Vapnik [35] and are commonly used in small-sample and nonlinear problems [36,37]. The schematic diagram of SVM classification is shown in Figure 2. The fundamental model of SVM is to locate or create a hyperplane that maximizes the distance between the nearest sample factors on both sides of the hyperplane. In the pattern space, the partitioned hyperplane can be described as:

$$f(x) = w^T x + b \tag{13}$$

where $w = (w_1, w_2, \cdots, w_d)$ is the normal vector of the hyperplane and b is the distance between the hyperplane and the far point.

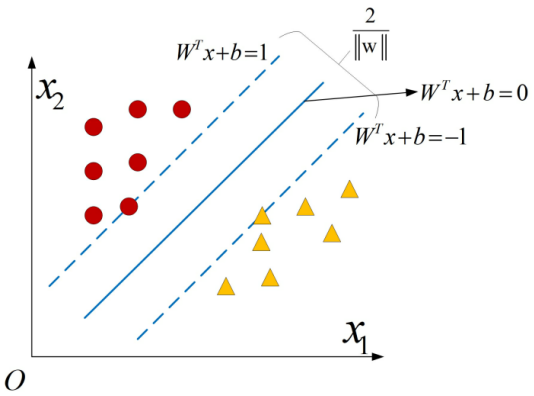

Figure 2. Schematic diagram of SVM classification.

This linearly separable optimal classification surface is:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \tag{14}$$

The constraints are:

$$y_i(w \cdot x_i + b) \geq 1, i = 1, \ldots, n \tag{15}$$

In order to solve the above problems, the Lagrangian function is introduced:

$$L(w, b, a) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} a_i y_i (w \cdot x_i + b) + \sum_{i=1}^{n} a_i \tag{16}$$

where a_i is the Lagrange multiplier and $a_i \geq 0$, b is the classification threshold. The optimal classification is as follows:

Samples in the unique house are generally non-linear and indivisible. Therefore, a kernel characteristic is used to map the non-linear samples into the greater dimensional space, and a most reliable hyperplane is observed in the greater dimensional space to make the samples linearly divisible. Finally, the primary model of the SVM can be obtained as follows:

$$f(x) = sign(\sum_{i=1}^{n} a_i y_i K(x_i \cdot x) + b) \tag{17}$$

where K denotes the kernel function. The Gaussian radial basis characteristic with its sturdy localization and knowledge gaining capability is chosen. The most suitable values of the width coefficients and penalty factors of the kernel characteristic are acquired with the aid of a cross-validation method.

3. Optimizing VMD and Improving CNN for Diagnosis

Aiming at the small amount of on-hand fault records of diesel engines and the complicated characteristic extraction of normal wise analysis methods, this paper proposes a small-sample fault detection approach for diesel engines. The approach combines the tremendous signal processing functionality of VMD, the computerized characteristic extraction functionality of CNN and the processing capability and generalization capability of SVM for small samples. Its flow is shown in Figure 3.

The specific troubleshooting procedure is as follows:

Step 1: Carry out a preset fault test on the diesel engine and acquire the fault facts of the diesel engine. The authentic vibration sign is decomposed with the aid of the VMD algorithm, the most effective variety of decomposed layers is decided by means of the spread entropy and the beneficial IMF factors are screened out and subsequently the noise reduction signal is obtained by means of superposition and reconstruction.

Step 2: Use CWT to convert the noise discount vibration signal into a time-frequency photo with a dimension of $227 \times 227 \times 3$. Every 5000 vibration signal factors are converted into a sample. The samples of every kind of fault are divided into education samples and take the form of samples in a ratio of 8:2.

Step 3: Train the CNN network, and enter the education samples into the CNN community for characteristic extraction.

Step 4: The CNN model education is completed and the features from the utterly connected layer are fed into the SVM for classification.

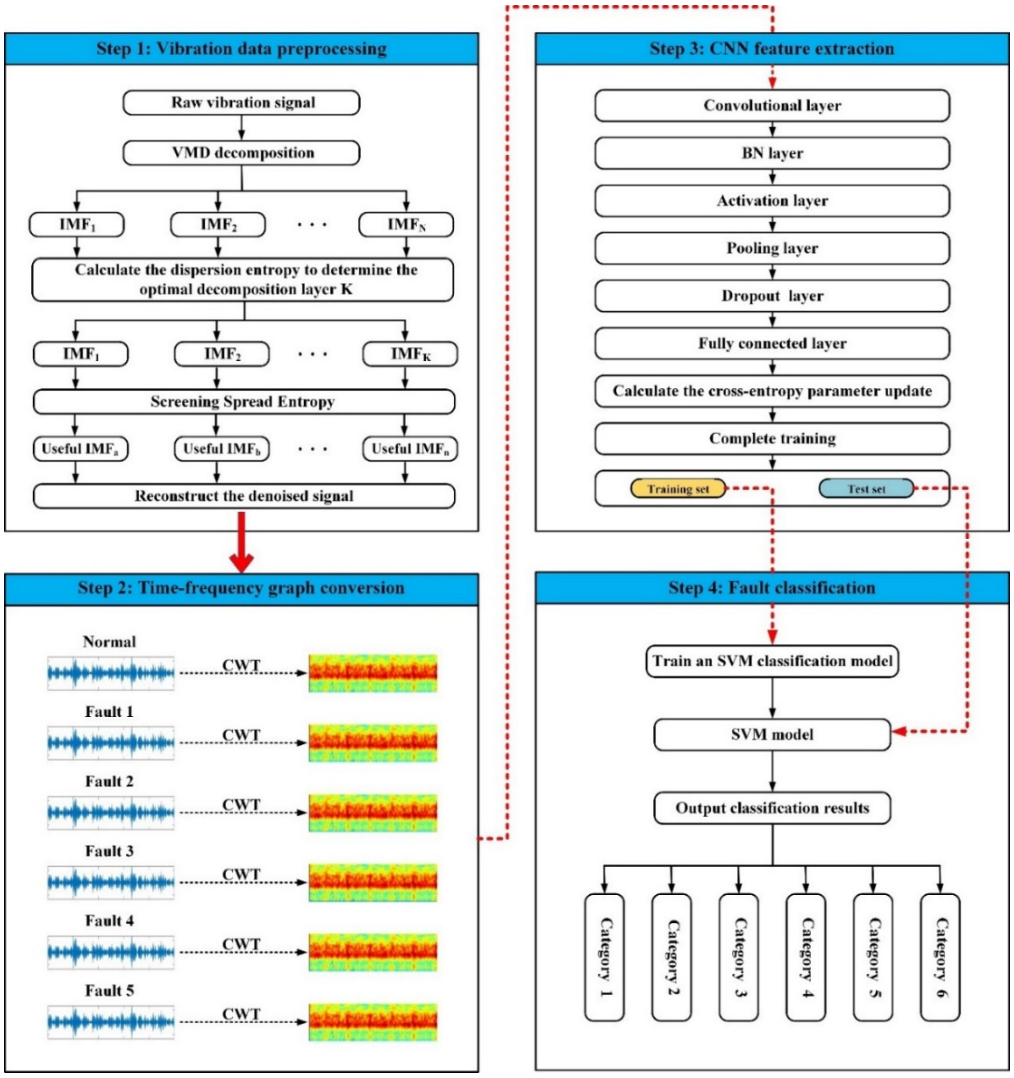

Figure 3. Fault diagnosis flow chart.

4. Experiments and Analysis

4.1. Experimental Equipment

In this experiment, a CA6DF3-20E3 diesel engine was used, and its rotational velocity was 800 rpm. The specific indicators are shown in Table 1. The sketch of the experimental setup and sensors is shown in Figure 4. The diesel engine control panel commonly controls the starting and stopping of the diesel engine and monitors the status of the diesel engine. The records acquisition system selects the acquisition card of the PXI-3342 model and the pc of PXI-9082. The sensor model is a B&W14100 acceleration sensor, and the sampling frequency is 20 KHz. The sensor is positioned on the cylinder head of the diesel engine, and the experiment below records six working conditions. As shown in Table 2. Specific experimental test steps are: (1) install the vibration sensor and connect the collection equipment; (2) start the diesel engine and set the speed to 800 rpm; (3) wait for the engine to run steadily for 3 min and start collecting data; (4) preset five kinds of faults, repeat steps (2) and (3), respectively, collect 10 sets of data for each fault, and collect 12 s for each set of data. Among them, for the injection pump fault and broken supply pipe fault experiments, the approach used was to replace the faulty injection pump and broken supply pipe fittings; for one cylinder misfiring and six cylinder misfiring, the approach taken was to disconnect its corresponding injector power cable; and for the air filter blockage fault

experiments, the air intake cover was added. The specific preset fault test method is shown in Figure 5.

Because of the complex structure of diesel engines and the difficulty of acquiring fault sample data, it is necessary to analyze the uncertainty analysis of diesel engine experiments. Engine uncertainty is mainly manifested in the following ways: (1) in the acquisition of fault data, there is noise uncertainty in the vibration signal, thus this paper uses a VMD to decompose the original vibration signal, reduce the impact of noise on fault identification and extract the weak features of the fault; (2) the uncertainty caused by the sensor, a single sensor will result in the fault information acquisition not being comprehensive, thus in this paper, the B&W14100 piezoelectric acceleration sensor with high accuracy is selected and the sensor is rigidly linked to the engine cylinder head. Six sensors were arranged on the diesel engine to collect the diesel engine fault information more comprehensively.

Table 1. Diesel engine specific detailed specifications.

Item	Indicators	Item	Indicators
Type	Four-stroke, high pressure common rail	Ignition	1-5-3-6-2-4
Model	Tin Diesel CA6DF3-20E3	Bore*stroke	107 mm × 125 mm
Size	1330 × 970 × 1005 mm	Common rail systems	BOSCH Electronically controlled common rail
Net power	147 KW	Power rating	155 KW
Maximum torque	760 NM	Maximum horsepower	200 horsepower

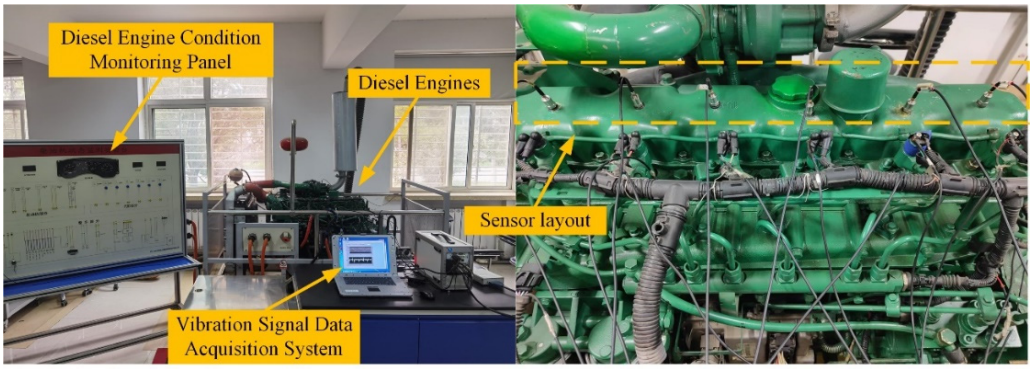

Figure 4. Diesel engine test platform.

Table 2. Diesel engine operating conditions.

Fault Tags	Fault Tags	Fault Tags	Type of Fault
01	normal	04	Six cylinder misfire
02	Insufficient oil supply to high pressure pump	05	Air filter clogged
03	Cylinder misfire	06	Oil supply pipe is damaged and dripping oil

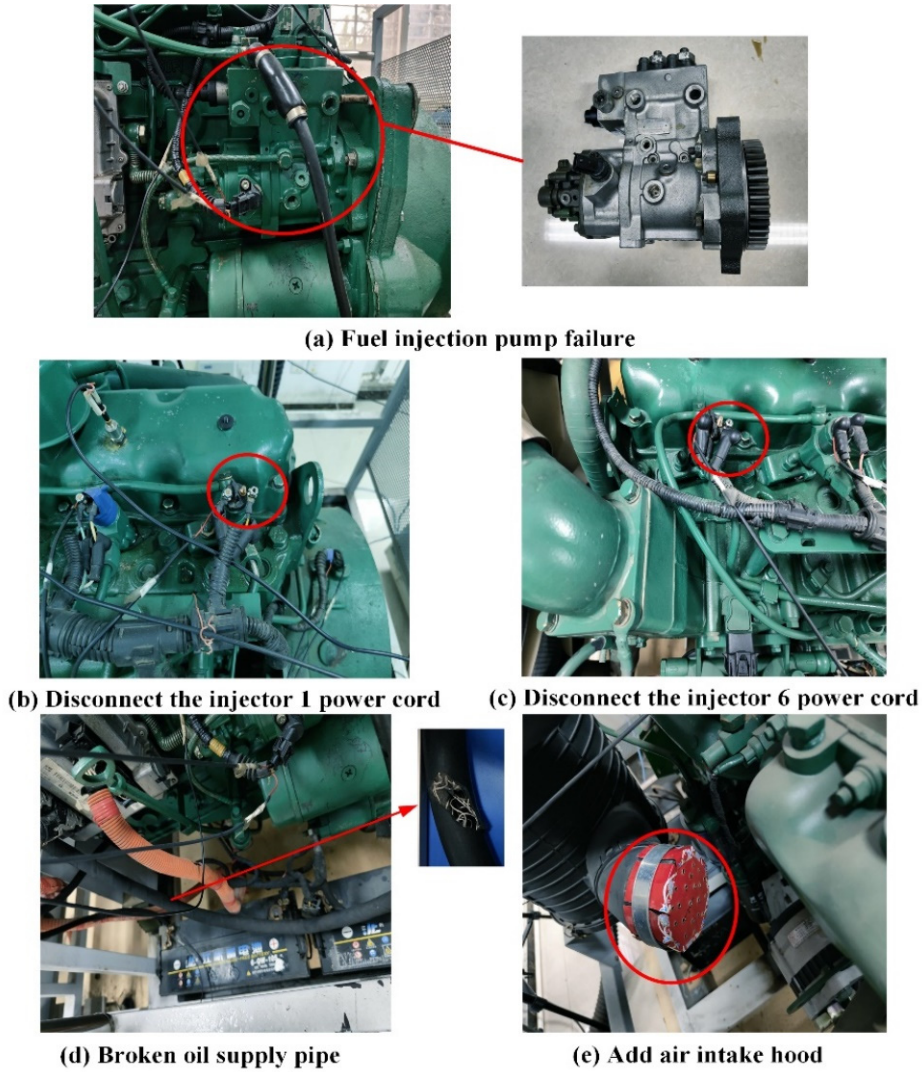

Figure 5. Preset experimental fault method diagram.

4.2. Preprocessing and Analysis of Experimental Data

The unique vibration signal was collected, the diesel engine speed was set to 800 rpm, the sensor sampling frequency was 20 KHz, a set of information was collected every 30 s, each set of information collection time every 12 s, and 10 sets of records were collected for every condition. The particular experimental facts are shown in Table 3.

Table 3. Experimental dataset.

Fault Tags	Rotational Speed	Frequency	Number of Sensors	Collection Time	Sample Collection
01	800 rpm	20 KHz	6	12 s	10
02	800 rpm	20 KHz	6	12 s	10
03	800 rpm	20 KHz	6	12 s	10
04	800 rpm	20 KHz	6	12 s	10
05	800 rpm	20 KHz	6	12 s	10
06	800 rpm	20 KHz	6	12 s	10

The experimental information preprocessing was carried out with only one cylinder misfire and the 4th sensor acquisition sign for instant analysis. The authentic vibration sign was decomposed using the VMD to calculate the scattering entropy values of every

IMF aspect to determine the most suitable wide variety of decomposition layers and the most treasured IMF components. For calculating the scattering entropy, the embedding dimension of m = 2, the quantity of categories of c = 8 and the time delay of d = 1 were selected. The scattering entropy effects of every aspect of the special decomposition layers are displayed below in Table 4. At k = 5 the scattering entropy cost starts off evolved to show a clear turnaround at IMF2. This means that the noise issue and the useful factor are already present at this point, and the scattering entropy price of IMF2 is the largest, indicating the biggest sign complexity. Therefore, the best range for decomposition layers of k = 5 is determined.

Table 4. Scattering entropy values for every aspect of the VMD decomposition.

IMF	Scattering Entropy Values for Each IMF Issue at One of a Kind K				
	3	4	5	6	7
IMF1	3.9627	3.9711	3.8240	3.8330	3.8339
IMF2	3.5992	3.6195	3.9498	3.9653	3.9703
IMF3	3.1888	3.3924	3.6125	3.7180	3.8439
IMF4		3.0777	3.3808	3.5933	3.6937
IMF5			3.0669	3.3597	3.5810
IMF6				3.0567	3.3384
IMF7					3.0575

When k = 5, the spectrum plot of every decomposed IMF factor is calculated, as shown in Figure 6. It can be seen in Figure 6 that each IMF1 and IMF2 are excessive frequency factors with a relatively extensive bandwidth and a complex spectrum, and there can also be a large variety of interfering alerts and complicated signals. Therefore, the IMF1 and IMF2 aspects are discarded and the final three factors are chosen for reconstruction. The reconstructed signal after noise reduction is transformed into a CWT time-frequency map, and every 5000 samples are modified into a two-dimensional time-frequency map. Figure 7 indicates an example CWT plot for each fault state (the color spectrum represents the sign power intensity stage in dB). It can be seen in Figure 7 that although the time-frequency distribution of the fault circumstance is close to that of the regular condition, and the foremost energy distribution of the engine tends to be 1.5–4 Khz, the time-frequency plan can visualize the distribution of power in distinct frequency ranges. In addition, the time-frequency area points developed with the aid of CWT can efficiently represent the non-smooth characteristics of the vibration signal, which is extra beneficial for CNN to extract features. As a result, the CWT time-frequency maps include richer fault information than time or frequency area alerts and are more beneficial for fault diagnosis.

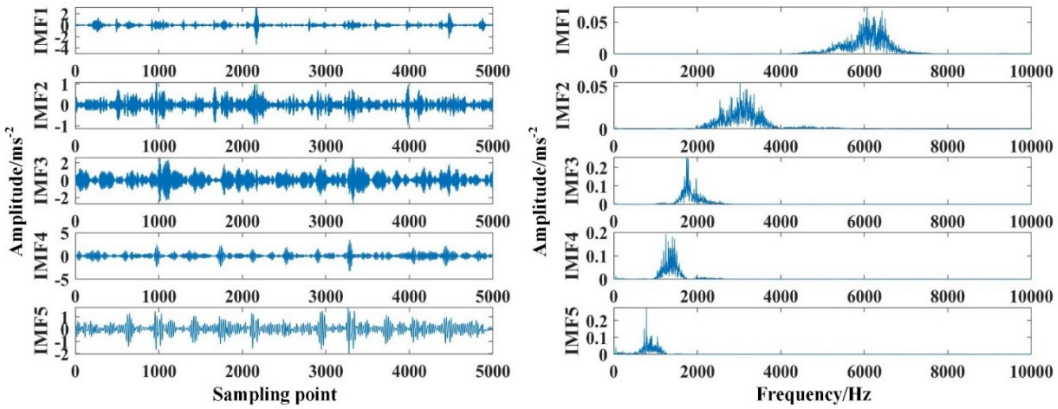

Figure 6. IMF component and IMF component spectrograms.

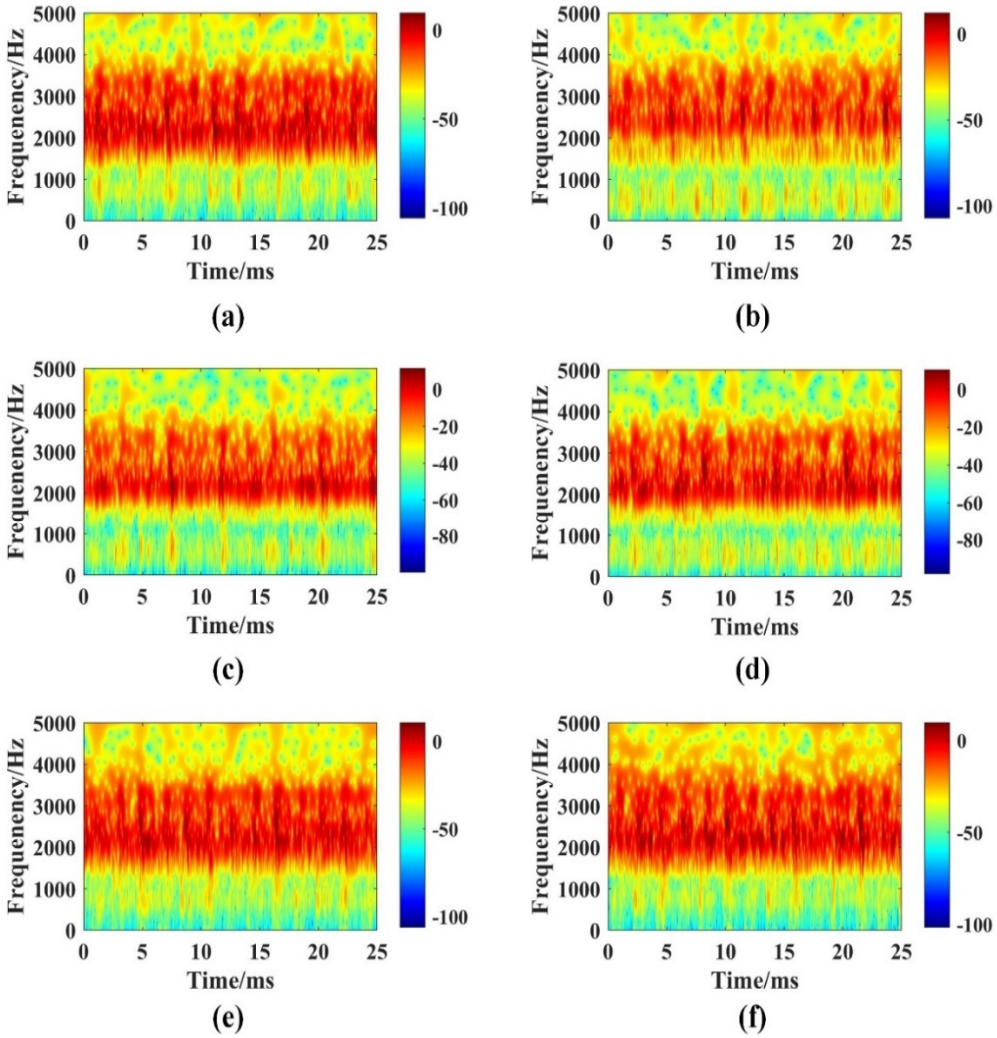

Figure 7. CWT time-frequency diagram: (**a**) 01 status; (**b**) 02 status; (**c**) 03 status; (**d**) 04 status; (**e**) 05 status; (**f**) 06 status.

4.3. CNN Model Structure

An excellent community model without delay influences the great of function extraction. In order to obtain excellent extraction of feature information, the CNN model used in this paper has 24 layers, consisting of an input layer, a convolutional layer, a normalization layer, a ReLU layer, a pooling layer, an utterly-linked layer, a Softmax layer and an output layer, and its shape is shown in Table 5. After many experiments, considering factors such as classification effect and time cost, the initial learning rate is set to 0.01; the activation function selects the ReLU function; the pooling layer with maximum pooling, which not only reduces the input data and training parameters of the next layer, and greatly preserve the largest local features in the feature map; the model training optimization uses the Adam algorithm; and the minimum batch size is set to 20. The convolution method adopts a zero-padding method, which is beneficial to control the shape of the output. In order to reduce overfitting and improve model convergence speed, dropout and batch normalization (BN) layers are introduced.

Table 5. CNN model structure.

Name	Activations	Learnables
Image input	227 × 227 × 3	-
Conv_1	227 × 227 × 8	Weights 5 × 5 × 3 × 8, Bias1 × 1 × 8
Batchnorm_1	227 × 227 × 8	Offset 1 × 1 × 8, Scale1 × 1 × 8
Relu_1	227 × 227 × 8	-
Maxpool_1	113 × 113 × 8	-
Conv_2	113 × 113 × 12	Weights3 × 3 × 8 × 12, Bias1 × 1 × 12
Batchnorm_2	113 × 113 × 12	Offset 1 × 1 × 12, Scale1 × 1 × 12
Relu_2	113 × 113 × 12	-
Maxpool_2	56 × 56 × 12	-
Conv_3	56 × 56 × 16	Weights3 × 3 × 12 × 16, Bias1 × 1 × 16
Batchnorm_3	56 × 56 × 16	Offset 1 × 1 × 16, Scale1 × 1 × 16
Relu_3	56 × 56 × 16	-
Maxpool_3	28 × 28 × 16	-
Conv_4	28 × 28 × 20	Weights 3 × 3 × 16 × 20, Bias1 × 1 × 20
Batchnorm_4	28 × 28 × 20	Offset 1 × 1 × 20, Scale1 × 1 × 20
Relu_4	28 × 28 × 20	-
Maxpool_4	14 × 14 × 20	-
Fc1	1 × 1 × 32	Weights 32 × 3920, Bias32 × 1
Dropout_1	1 × 1 × 32	-
Fc2	1 × 1 × 64	Weights 64 × 32, Bias64 × 1
Dropout_2	1 × 1 × 64	-
Fc3	1 × 1 × 6	Weights 6 × 64, Bias6 × 1
Softmax	1 × 1 × 6	-
Classification Output	-	-

Dropout is when some of the hidden nodes are set to zero, which reduces the training of weights and effectively reduces the overfitting of the model; dropout is set to 0.2. The BN layer normalizes the data and speeds up the training of the model.

4.4. Experimental Analysis

4.4.1. CNN Feature Extraction Effect

To show the impact of CNN characteristic extraction, exclusive fault types are represented by the t-SNE [38] method. t-SNE has superb visualization functionality and can intuitively show the statistics feature distribution functionality of unique community layers. t-SNE is used in this paper to change multi-dimensional statistics into 2-dimensional data. When the variety of statistics samples for every circumstance is 240, the function distribution of unique layers is proven in Figure 8. It can be seen in Figure 8 that the fault state of 3 is easy to distinguish. The data features in the Conv1 layer are scattered and disordered, different fault states overlap each other, and the classification effect is poor. In the Conv2 layer, two failure modes of 03 and 05 can be clearly distinguished. With the increase of the number of network layers, in the Conv4 layer, only a few features are indistinguishable, and the 04 and 02 states still partially overlap. In the FC3 layer, all fault states are fully distinguishable and clustering is better. The experimental evaluation suggests that the technique has accurate function mastering capability and can improve the classification accuracy. Table 6 presents the feature extraction and diagnosis results of the CNN model for each failure mode.

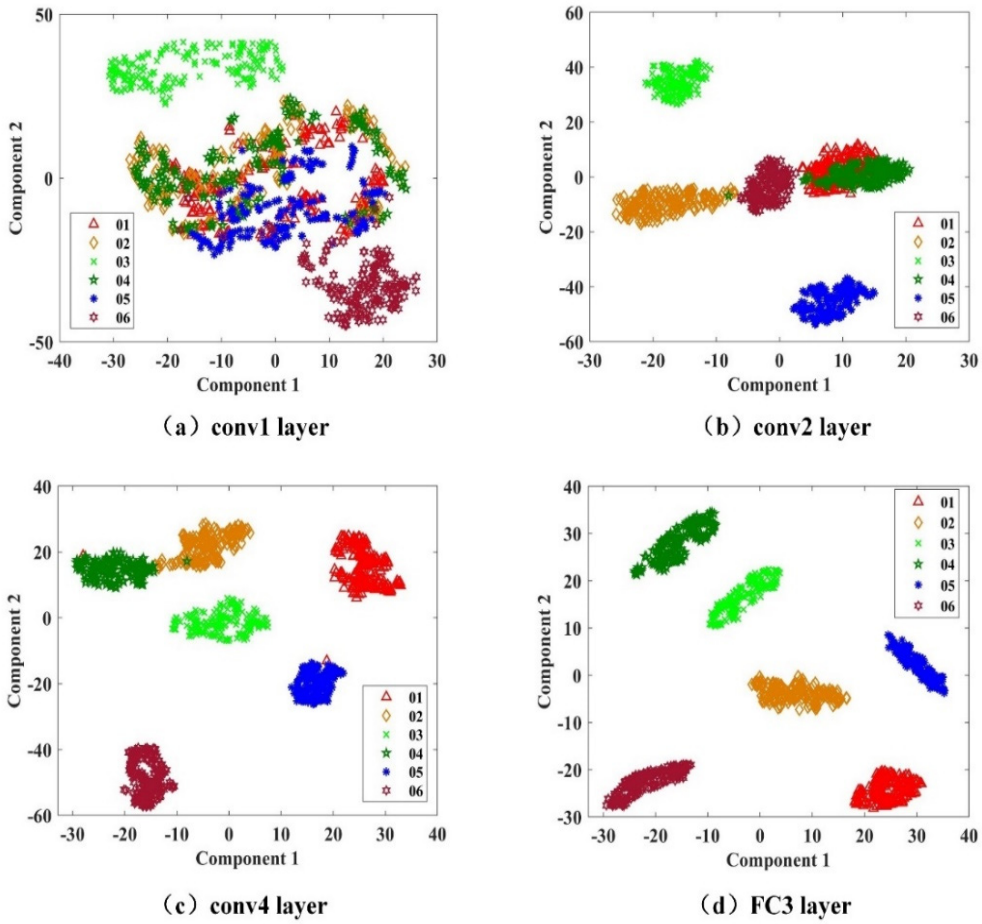

Figure 8. Feature data distribution of each layer.

Table 6. CNN model feature extraction diagnosis results.

Fault State	Total Sample	Accuracy
01	240	100.0%
02	240	100.0%
03	240	100.0%
04	240	100.0%
05	240	100.0%
06	240	100.0%

4.4.2. Comparative Analysis of Multi-Mode Decomposition Diagnosis Results

As can be seen from Figure 8, CNN has excellent feature extraction ability. In order to prove that the optimized VMD signal is helpful to improve the classification accuracy, different signal processing algorithms are selected, and then CWT is used to convert the noise reduction signal into a time-frequency map, which is input into the CNN-SVM network. The experimental outcomes are shown in Table 7. The EMD-CNN-SVM technique has the shortest education time, however the classification accuracy is low due to modal blending. The EEMD-CNN-SVM technique gave the worst classification outcomes and time for this dataset, whilst the CEEMD-CNN-SVM approach multiplied the accuracy to some extent, however it took longer. The comprehensive analysis shows that the optimized VMD algorithm has excellent noise reduction ability, and the calculation efficiency is higher than other algorithms, which helps to improve the classification accuracy. The specific classification and diagnosis diagram is shown in Figure 9.

Table 7. Comparative results of the different models.

Network Model	Total Sample	Training Sample	Test Sample	Accuracy	Time (s)
VMD-CNN-SVM	240	192	48	100.0%	200.3
EMD-CNN-SVM	240	192	48	95.1%	194.6
EEMD-CNN-SVM	240	192	48	92.3%	364.1
CEEMD-CNN-SVM	240	192	48	97.6%	356.4

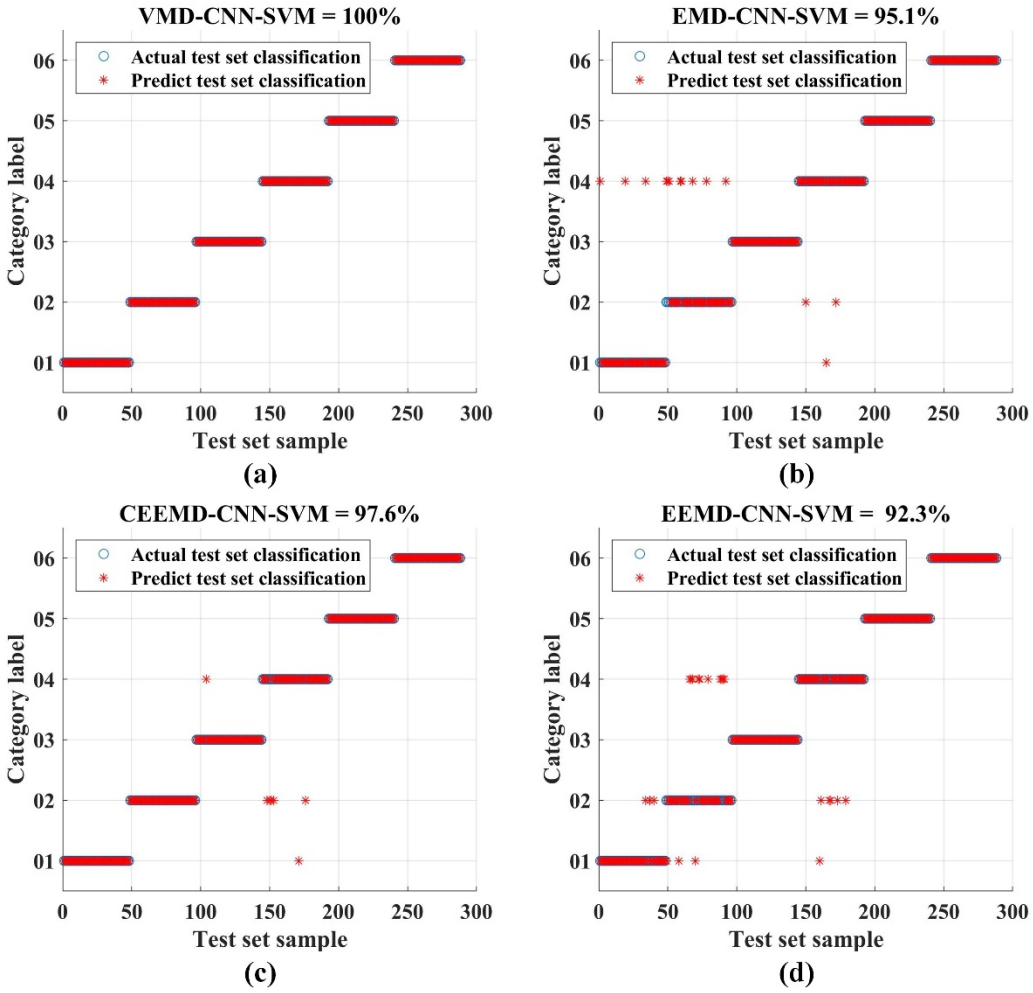

Figure 9. Comparison of the results of different modalities: (**a**) VMD-CNN-SVM = 100%; (**b**) EMD-CNN-SVM = 95.1%; (**c**) CEEMD-CNN-SVM = 92.3%; (**d**) EEMD-CNN-SVM = 97.6%.

4.4.3. Comparative Analysis of Different Training Samples

From the above analysis, it can be seen that by using different signal processing methods, the denoised sign is entered into the CNN-SVM model after continuous wavelet transform, and the accuracy rate can reach 90%. It can be seen that the CNN has excellent extraction ability characteristics, but the excellent mechanical fault classification ability of CNN is based on massive training samples. This approach can solve the small sample issue well with the exceptional characteristic extraction capacity of CNN and the higher classification potential of SVM. The accuracy rates of typical CNN-Softmax and CNN-SVM under different amounts of data are compared. The 5-fold cross-validation test was used. The CNN-Softmax and CNN-SVM models were trained using the stipulations of whole statistics volumes of 30, 60, 120 and 240. It can be seen from Table 8 that the accuracy of CNN-Softmax in the test set drops rapidly with the decrease of test data, and fluctuates greatly. However, the accuracy of the CNN-SVM model is less affected by the amount of

data. In the case of a small amount of data, CNN-SVM still has good performance, and the accuracy rate can still remain above 90%, which verifies the method proposed in this paper. It is still effective in solving small sample problems.

Table 8. Accuracy values of CNN-Softmax and CNN-SVM with different data volumes.

Network Model	Accuracy			
	30	60	120	240
CNN-Softmax training set	93.2%	96.9%	98.0%	100.0%
CNN-Softmax test set	60.1%	68.3%	85.6%	97.6%
CNN-SVM training set	94.4%	96.1%	98.9%	100.0%
CNN-SVM test set	92.3%	95.6%	98.2%	100.0%

4.4.4. Comparative Analysis of Diagnostic Results of Different Network Models

To confirm the effectiveness of the proposed method, the proposed approach is in contrast with VMD-CWT-CNN-RF, VMD-CWT-CNN-ELM, VMD-CWT-CNN-LSTM, VMD-CWT-CNN-DBN and VMD-CWT-CNN-Softmax strategies. Each method is trained five times, and its training time and accuracy are the averages of the five times. The diagnosis results are shown in Table 9, and the test set classification effect is shown in Figure 10. The evaluation leads to the use of scatter entropy to determine the beneficial elements of the VMD decomposition, which is then modified into a CWT time-frequency map and fed into the CNN to extract features. Using different machine learning and deep learning models as classifiers, the accuracy rate can reach more than 90%. It fully shows that the VMD algorithm has excellent signal processing ability and can obtain better noise reduction effect. The time-frequency map after CWT contains richer fault feature information, which is helpful for the CNN to extract fault features. In this paper, we use SVM, which has exact classification performance, as the classifier. In both the training and check sets, an accuracy of 100% can be attained, which is higher than other methods, verifying the effectiveness of the proposed approach in this paper. From Table 9, it can be concluded that the classifier uses machine learning methods (SVM, random forest (RF) and extreme learning machine (ELM)), and its time is better than deep learning (long short-term memory network (LSTM), deep belief network (DBN) and traditional CNN classifiers) methods. The hyperparameters, learning rate and number of iterations in deep learning need to be determined, thus the training time is long. It can also be observed through Figure 10 that among the different network models, fault state 03 and fault state 06 are the best for classification, and fault state 02 and fault state 04 are harder to distinguish. This is consistent with the use of t-SNE to demonstrate the effect of CNN extracted features, indirectly indicating that a good fault feature extractor directly affects the subsequent classification effect, while CWT adaptivity and multi-resolution are better than STFT, thus the time-frequency map extracted by CWT contains richer feature information and is more conducive to CNN extracted features.

Table 9. Diagnostic results of different models.

Network Model	Total Sample	Training Set	Test Set	Time (s)
VMD-CWT-CNN-SVM	240	100.0%	100.0%	200.3
VMD-CWT-CNN-RF	240	95.4%	94.1%	216.9
VMD-CWT-CNN-ELM	240	100.0%	98.6%	233.7
VMD-CWT-CNN-LSTM	240	98.2%	96.9%	346.7
VMD-CWT-CNN-DBN	240	97.7%	95.8%	335.4
VMD-CWT-CNN-Softmax	240	100.0%	97.6%	276.6

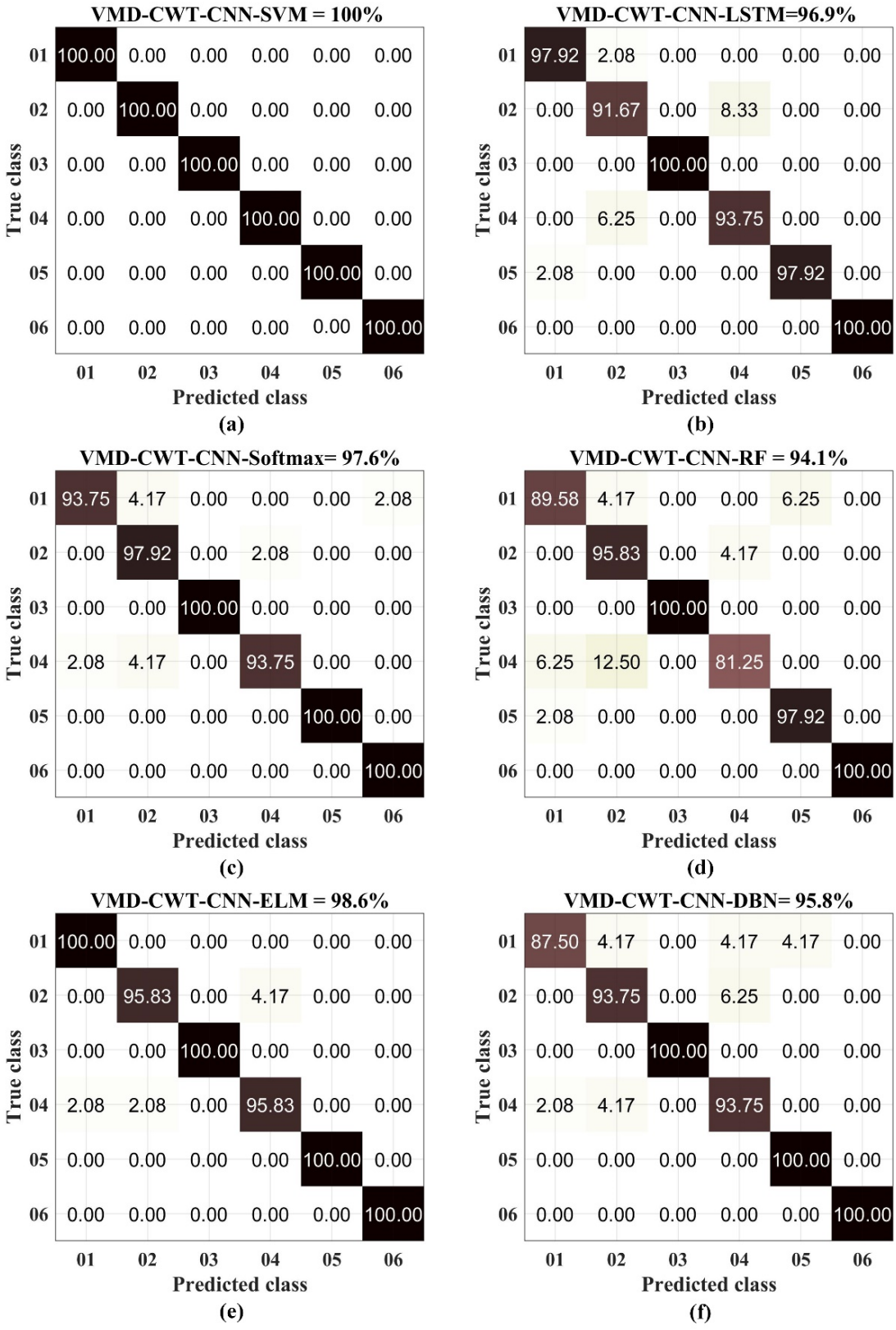

Figure 10. Fault diagnosis results of different methods: (**a**) VMD-CWT-CNN-SVM = 100.0%; (**b**) VMD-CWT-CNN-LSTM = 96.9%; (**c**) VMD-CWT-CNN-Softmax = 97.6%; (**d**) VMD-CWT-CNN-RF = 94.1%; (**e**) VMD-CWT-CNN-ELM = 98.6%; (**f**) VMD-CWT-CNN-DBN = 95.8%.

5. Conclusions

In this paper, a diesel engine fault diagnosis technique with optimized VMD and improved CNN is proposed for diesel engines. The essential findings are as follows:

1. Based on the optimized VMD method to decompose the authentic vibration signal, the gold standard range of decomposition layers and beneficial components are acquired through scattering entropy value, and a higher noise reduction effect is obtained.
2. The diesel engine fault diagnosis method primarily based on CWT time-frequency map and extended CNN network model is viable and effective. The continuous wavelet transform picture incorporates wealthy fault features, which improves the fault identification rate.
3. The visualization impact of t-SNE verifies that the method in this paper can successfully extract the fault points of diesel engines and enhance the fault consciousness cost of diesel engines. Replacing the Softmax layer in CNN with SVM can correctly solve the small pattern issue of diesel engines. Without manual function extraction and selection, the analysis method proposed in this paper avoids the errors induced by means of guide function extraction, and its training time is shorter and the classification impact is better.

The approach proposed in this paper ordinarily conducts preset fault experiments for diesel engines under equal operating conditions, however the tool's actual operation regularly results in exceptional working prerequisites due to unique speed and load, thus how to solve the fault analysis of variable working prerequisites will be the next step to be carried out.

Author Contributions: Conceptualization, X.Z., H.B., H.Y. and R.W.; methodology, X.Z. and H.B.; software, X.Z. and H.B.; data curation, X.Z. and H.Y.; formal analysis, X.Z.; resources, X.Z., C.G. and X.J.; validation, X.Z. and R.W.; visualization, X.Z.; investigation, X.Z., R.W. and H.Y.; writing—original draft preparation, X.Z.; writing—review and editing. X.Z., H.B., C.G. and X.J.; funding acquisition, C.G. and X.J.; project administration, C.G. and X.J.; supervision, C.G. and X.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Natural Science Foundation of China (grant No. 71871220). The support is gratefully acknowledged. The authors would also like to thank the reviewers for their valuable suggestions and comments.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data involved in this article have been presented in the article.

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

References

1. Li, Z.; Jiang, Y.; Duan, Z.; Peng, Z. A new swarm intelligence optimized multiclass multi-kernel relevant vector machine: An experimental analysis in failure diagnostics of diesel engines. *Struct. Health Monit.* **2018**, *17*, 1503–1519. [CrossRef]
2. Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mech. Syst. Signal Process.* **2018**, *108*, 33–47. [CrossRef]
3. Goyal, D.; Pabla, B.S. The vibration monitoring methods and signal processing techniques for structural health monitoring: A review. *Arch. Comput. Methods Eng.* **2016**, *23*, 585–594. [CrossRef]
4. Saha, D.K.; Hoque, M.E.; Badihi, H. Development of Intelligent Fault Diagnosis Technique of Rotary Machine Element Bearing: A Machine Learning Approach. *Sensors* **2022**, *22*, 1073. [CrossRef] [PubMed]
5. Mathew, S.K.; Zhang, Y. Acoustic-based engine fault diagnosis using WPT, PCA and Bayesian optimization. *Appl. Sci.* **2020**, *10*, 6890. [CrossRef]
6. Yang, Y.; Pan, H.; Ma, L.; Cheng, J. A fault diagnosis approach for roller bearing based on improved intrinsic timescale decomposition de-noising and kriging-variable predictive model-based class discriminate. *J. Vib. Control* **2014**, *22*, 1431–1446. [CrossRef]
7. Sun, Y.; Li, S.; Wang, X. Bearing fault diagnosis based on EMD and improved Chebyshev distance in SDP image. *Measurement* **2021**, *176*, 109100. [CrossRef]

8. Amarouayache, I.I.; Saadi, M.N.; Guersi, N.; Boutasseta, N. Bearing fault diagnostics using EEMD processing and convolutional neural network methods. *Int. J. Adv. Manuf. Technol.* **2020**, *107*, 4077–4095. [CrossRef]
9. Han, M.; Pan, J. A fault diagnosis method combined with LMD, sample entropy and energy ratio for roller bearings. *Measurement* **2015**, *76*, 7–19. [CrossRef]
10. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2013**, *62*, 531–544. [CrossRef]
11. Bi, X.; Lin, J.; Tang, D.; Bi, F.; Li, X.; Yang, X.; Ma, T.; Shen, P. VMD-KFCM algorithm for the fault diagnosis of diesel engine vibration signals. *Energies* **2020**, *13*, 228. [CrossRef]
12. Li, X.; Ma, Z.; Kang, D.; Li, X. Fault diagnosis for rolling bearing based on VMD-FRFT. *Measurement* **2020**, *155*, 107554. [CrossRef]
13. Zheng, J.; Yuan, Y.; Zou, L.; Deng, W.; Guo, C.; Zhao, H. Study on a novel fault diagnosis method based on VMD and BLM. *Symmetry* **2019**, *11*, 747. [CrossRef]
14. Qiao, M.Y.; Tang, X.X.; Liu, Y.X.; Yan, S. Fault diagnosis method of rolling bearings based on VMD and MDSVM. *Multimed. Tools Appl.* **2021**, *80*, 14521–14544. [CrossRef]
15. Zhou, J.; Guo, X.; Wang, Z.; Du, W.; Han, X.; He, G.; Xue, H.; Kou, Y. Research on Fault Extraction Method of Variational Mode Decomposition Based on Immunized Fruit Fly Optimization Algorithm. *Entropy* **2019**, *21*, 400. [CrossRef] [PubMed]
16. Li, Z.; Li, S.; Mao, J.; Li, J.; Wang, Q.; Zhang, Y. A Novel Lidar Signal-Denoising Algorithm Based on Sparrow Search Algorithm for Optimal Variational Modal Decomposition. *Remote Sens.* **2022**, *14*, 4960. [CrossRef]
17. Jin, Z.; Chen, G.; Yang, Z. Rolling Bearing Fault Diagnosis Based on WOA-VMD-MPE and MPSO-LSSVM. *Entropy* **2022**, *24*, 927. [CrossRef]
18. Jiang, X.; Shen, C.; Shi, J.; Zhu, Z. Initial center frequency-guided VMD for fault diagnosis of rotating machines. *J. Sound Vib.* **2018**, *435*, 36–55. [CrossRef]
19. Hoang, D.T.; Kang, H.J. A survey on deep learning based bearing fault diagnosis. *Neurocomputing* **2019**, *335*, 327–335. [CrossRef]
20. Chen, L.; Li, S.; Bai, Q.; Yang, J.; Jiang, S.; Miao, Y. Review of Image Classification Algorithms Based on Convolutional Neural Networks. *Remote Sens.* **2021**, *13*, 4712. [CrossRef]
21. Habbouche, H.; Amirat, Y.; Benkedjouh, T.; Benbouzid, M. Bearing Fault Event-Triggered Diagnosis using a Variational Mode Decomposition-based Machine Learning Approach. *IEEE Trans. Energy Convers.* **2021**, *37*, 466–474. [CrossRef]
22. Wang, X.; Mao, D.; Li, X. Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network. *Measurement* **2021**, *173*, 108518. [CrossRef]
23. Gong, W.; Wang, Y.; Zhang, M.; Mihankhah, E.; Chen, H.; Wang, D. A Fast Anomaly Diagnosis Approach Based on Modified CNN and Multi-Sensor Data Fusion. *IEEE Trans. Ind. Electron.* **2021**, *69*, 13636–13646. [CrossRef]
24. Ding, X.; He, Q. Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 1926–1935. [CrossRef]
25. Huo, C.; Jiang, Q.; Shen, Y.; Qian, C.; Zhang, Q. New transfer learning fault diagnosis method of rolling bearing based on ADC-CNN and LATL under variable conditions. *Measurement* **2021**, *188*, 110587. [CrossRef]
26. Xiao, Q.; Li, S.; Zhou, L.; Shi, W. Improved Variational Mode Decomposition and CNN for Intelligent Rotating Machinery Fault Diagnosis. *Entropy* **2022**, *24*, 908. [CrossRef]
27. Nishat Toma, R.; Kim, C.H.; Kim, J.M. Bearing fault classification using ensemble empirical mode decomposition and convolutional neural network. *Electron* **2021**, *10*, 1248. [CrossRef]
28. Liang, P.; Deng, C.; Wu, J.; Yang, Z.; Zhu, J.; Zhang, Z. Compound Fault Diagnosis of Gearboxes via Multi-label Convolutional Neural Network and Wavelet Transform. *Comput. Ind.* **2019**, *113*, 103132. [CrossRef]
29. Richman, J.S.; Lake, D.E.; Moorman, J.R. *Sample Entropy. Methods in Enzymology*; Academic Press: Cambridge, MA, USA, 2004; Volume 384, pp. 172–184.
30. Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [CrossRef]
31. Rostaghi, M.; Azami, H. Dispersion entropy: A measure for time-series analysis. *IEEE Signal Process. Lett.* **2016**, *23*, 610–614. [CrossRef]
32. Bai, H.; Zhan, X.; Yan, H.; Wen, L.; Jia, X. Combination of Optimized Variational Mode Decomposition and Deep Transfer Learning: A Better Fault Diagnosis Approach for Diesel Engines. *Electronics* **2022**, *11*, 1969. [CrossRef]
33. Huang, T.; Zhang, Q.; Tang, X.; Zhao, S.; Lu, X. A novel fault diagnosis method based on CNN and LSTM and its application in fault diagnosis for complex systems. *Artif. Intell. Rev.* **2022**, *55*, 1289–1315. [CrossRef]
34. Chen, X.; Zhang, B.; Gao, D. Bearing fault diagnosis base on multi-scale CNN and LSTM model. *J. Intell. Manuf.* **2021**, *32*, 971–987. [CrossRef]
35. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
36. Huo, W.; Li, W.; Sun, C.; Ren, Q.; Gong, G. Research on Fuel Cell Fault Diagnosis Based on Genetic Algorithm Optimization of Support Vector Machine. *Energies* **2022**, *15*, 2294. [CrossRef]
37. Bai, H.; Zhan, X.; Yan, H.; Wen, L.; Yan, Y.; Jia, X. Research on Diesel Engine Fault Diagnosis Method Based on Stacked Sparse Autoencoder and Support Vector Machine. *Electronics* **2022**, *11*, 2249. [CrossRef]
38. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.