Research Article

Fault diagnosis of a marine power-generation diesel engine based on the Gramian angular field and a convolutional neural network

Congyue LI[1]✉, Yihuai HU[1], Jiawei JIANG[2], Dexin CUI[1]

[1]*Merchant Marine College, Shanghai Maritime University, Shanghai 201306, China*
[2]*School of Mechanical and Energy Engineering, Shanghai Technical Institute of Electronics & Information, Shanghai 201411, China*

Abstract: Marine power-generation diesel engines operate in harsh environments. Their vibration signals are highly complex and the feature information exhibits a non-linear distribution. It is difficult to extract effective feature information from the network model, resulting in low fault-diagnosis accuracy. To address this problem, we propose a fault-diagnosis method that combines the Gramian angular field (GAF) with a convolutional neural network (CNN). Firstly, the vibration signals are transformed into 2-D images by taking advantage of the GAF, which preserves temporal correlation. The raw signals can be mapped to 2-D image features such as texture and color. To integrate feature information, the images of the Gramian angular summation field (GASF) and Gramian angular difference field (GADF) are fused by the weighted-average fusion method. Secondly, the channel attention mechanism and temporal attention mechanism are introduced in the CNN model to optimize the CNN learning mechanism. Introducing the concept of residuals in the attention mechanism improves the feasibility of optimization. Finally, the weighted-average fused images are fed into the CNN for feature extraction and fault diagnosis. The validity of the proposed method is verified by experiments with abnormal valve clearance. The average diagnostic accuracy is 98.4%. When -20 < SNR (Signal-to-noise ratio) < 20 dB, the diagnostic accuracy of the proposed method is higher than 94.0%. The proposed method has superior diagnostic performance. Moreover, it has a certain anti-noise capability and variable-load adaptive capability.

Key words: Multiple attention mechanisms; Convolutional neural network; Gramian angular field; Fusion image; Marine power-generation diesel engine; Fault diagnosis

1 Introduction

The power-generation diesel engine is an essential piece of equipment for ships. In addition to providing electricity, it is widely used for electric propulsion. Recently, there has been a lot of interest in fault diagnosis of key diesel engine components (Cai et al., 2020; Çağlar and Yasin, 2022; Rao et al., 2022). The air-distribution mechanism operates in a harsh environment, especially the valves and seats. They are directly connected to the combustion chamber and often exposed to high temperatures and pressures. The valve train is one of the main moving parts. It directly affects the power and combustion economy of the diesel engine. The opening and closing of the air valves cause periodic shocks to the valve seats. In addition, particulate matter from combustion accelerates the wear and corrosion of valves and valve seats. Therefore, troubleshooting the valve mechanism is essential.

Traditional intelligent fault-diagnosis methods require human extraction of feature information, such as time-domain information, frequency-domain information, or time-frequency-domain information (Nayana and Geethanjali, 2017; Dhamande and Chaudhari, 2017; Sun et al., 2022). This feature information is fed into a classifier for fault classification. Machine-learning methods commonly used include support vector machines (Song et al., 2023) and random forests (Mariela et al., 2016). In practical engineering, feature selection and extraction depend on professional knowledge. Manually extracted fault features do not fully reflect the features of mechanical vibration signals (Peng et al., 2020). Data that can reflect the health status of diesel engines are characterized by their large volume, diversity, and low-value density (Qian et al., 2022; Hoang and Kang, 2018). This makes it challenging for traditional intelligent methods to meet the needs of fault diagnosis using big data. With the rapid development of artificial intelligence, deep learning is gaining

✉ Congyue LI, licongyue0126@163.com

🔟 Congyue LI, https://orcid.org/0000-0002-9290-5119
Yihuai HU, https://orcid.org/0000-0003-3000-7855

popularity in fault diagnosis (Zhao et al., 2019). Deep learning offers the capacity to learn complex non-linear relationships (Schmidhuber, 2015) and overcomes the disadvantages of shallow learning. The gearbox faults are successfully classified using data fusion and multiple classifiers (Senanayaka et al., 2019). Continuous wavelet transforms (CWT) (Du et al., 2022) are used to transform signals into 2-D images (Fu et al., 2023). They are input into a CNN to obtain fault diagnosis. However, CWT requires manual selection of the basis function, and its accuracy will vary depending on the choice of the basis function (Guo et al., 2020; Manarikkal et al., 2021). Alsalaet et al. (2023) used normalized feature maps as the input to a CNN to achieve bearing fault diagnosis. Hu et al. (2023) input acoustic emission signals into a CNN model to perform fault diagnosis of exhaust-valve leakage. He et al. (2022) designed a signal fusion model based on transfer learning to realize fault diagnosis of an axial piston pump. Inspired by the idea of multi-scale feature extraction, Xie et al. (2023) designed a multi-scale convolutional layer and incorporated a hybrid attention mechanism to achieve fault diagnosis of rolling bearings. Meanwhile, Xu et al. (2022) used a combination of a multi-scale CNN, feature-enhancement module, and a joint attention mechanism to perform fault diagnosis on rotating machinery.

The aforementioned scholars have conducted outstanding research on CNN-based fault diagnosis. However, diesel-engine vibration signals present non-linear characteristics and contain a large amount of background noise. Often only part of the data in the overall signal (e.g., continuous-pulse signal segments) contains key information, and data not related to faults can interfere with the learning of the network model. This has led to many CNN models that improve diagnostic accuracy through complex structures (Wen et al., 2019; Pan et al., 2021). Not only does this lead to problems such as high model computation and performance degradation, it also limits the model's ability to generalize when working conditions change. Accurate and efficient air-valve fault diagnosis can effectively improve combustion economy and ensure normal power output of diesel engines. Inspired by previous studies, in this paper we propose a new valve-clearance fault-diagnosis method. The main contribution points are as follows:

1. The vibration signals are transformed into 2-D images with temporal correlation by the GAF. In this way, 2-D images can maintain absolute correlation with time and provide different levels of information granularity.

2. The weighted average fusion algorithm is used to fuse 2-D images. This further portrays the invisible information in the signals and maximizes use of the signals' feature information.

3. A multi-attention mechanism is embedded into the CNN model to improve its ability to capture critical information and adaptively enhance the fault-feature response.

2 Background theory

2.1 The Gramian angular field (GAF)

The GAF (Cui et al., 2022) uses the idea of coordinate transformation to convert signals into 2-D images. Given a time series $X = \{x_1, x_2, ..., x_n\}$, its transformation process is as follows:

(1) Normalise the series to the interval [-1,1]. The calculation formula is:

$$\tilde{x}_i = \frac{(x_i - \min X) + (x_i - \max X)}{\max X - \min X} \tag{1}$$

(2) $\tilde{x}_i$ is mapped as angle and t_i is mapped as radius:

$$\begin{cases} \phi_i = \arccos \tilde{x}_i, -1 < \tilde{x}_i < 1, \tilde{x}_i \epsilon \tilde{X} \\ r = \dfrac{t_i}{N} \end{cases} \tag{2}$$

where t_i is a time stamp and N is a constant.

(3) Two images can be obtained. Eq. (3) is the formula for the GASF. Eq. (4) is the formula for the GADF.

$$G_{gasf} = \begin{bmatrix} \cos(\phi_1+\phi_1)\cdots\cos(\phi_1+\phi_n) \\ \cos(\phi_2+\phi_1)\cdots\cos(\phi_2+\phi_n) \\ \cdots \\ \cdots \\ \cos(\phi_n+\phi_1)\cdots\cos(\phi_n+\phi_n) \end{bmatrix} = X^T \cdot X - \sqrt{I-X^2}^{\mathrm{T}} \cdot \sqrt{I-X^2} \tag{3}$$

$$G_{gadf} = \begin{bmatrix} \sin(\phi_1-\phi_1)\cdots\sin(\phi_1-\phi_n) \\ \sin(\phi_2-\phi_1)\cdots\sin(\phi_2-\phi_n) \\ \cdots \\ \cdots \\ \sin(\phi_n-\phi_1)\cdots\sin(\phi_n-\phi_n) \end{bmatrix} = \sqrt{I-X^2}^{\mathrm{T}} \cdot X - X^{\mathrm{T}} \cdot \sqrt{I-X^2} \tag{4}$$

where $I = [1, 1, ... ,1]$.

If the times series is too long, a high-dimensional matrix is generated. This increases the amount of computation. Therefore, piecewise

aggregate approximation (PAA) (Ren et al., 2018) is used. After dividing the series into equal length sub-series and averaging them to represent the raw series, the raw series information can be retained and the sequence length is reduced. Fig. 1 shows the process of transforming the signal into a 2-D image with the GAF.
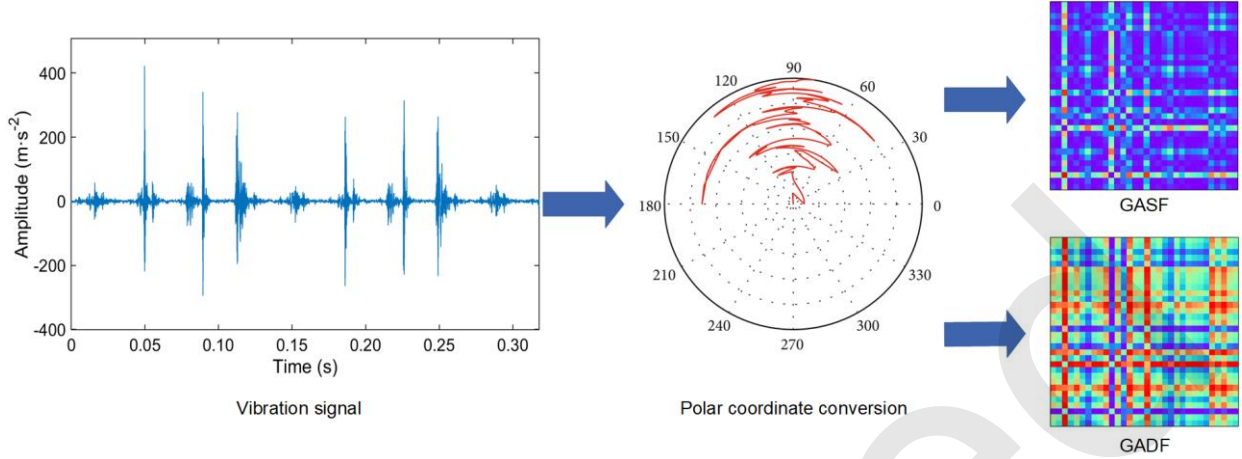

Fig. 1 GAF conversion process

2.2 Image fusion

Frequency information was extracted in the GASF and GADF 2D matrices and represented by density histograms (see Fig. 2). GASF and GADF had different distributions. To make the best use of feature information, the two images were fused using the weighted average fusion method (Ren et al., 2022). The weighted average fusion is calculated as follows:

$$I(x,y) = \begin{cases} I_1(x,y) & (x,y) \in I_1 \\ \omega_1(x,y)I_1(x,y) + \omega_2(x,y)I_2(x,y) & (x,y) \in (I_1 \cap I_2) \\ I_2(x,y) & (x,y) \in I_2 \end{cases} \quad (5)$$

$$\begin{cases} \omega_1 = \dfrac{d_1(x,y)}{W} \\ \omega_2 = \dfrac{d_2(x,y)}{W} \end{cases} \quad (6)$$

where $I_1(x,y)$ and $I_2(x,y)$ are images to be fused. $\omega_1(x,y)$ and $\omega_2(x,y)$ are the weights when adding the corresponding pixel values of images $I_1(x,y)$ and $I_2(x,y)$, respectively. $\omega_1, \omega_2 \in [0,1]$, $\omega_1 + \omega_2 = 1$. $d_1(x,y)$ and $d_2(x,y)$ indicate overlapping area boundaries. $d_1(x,y) + d_2(x,y) = W$.

Fig. 3 shows the fusion process for the two images, which complements and integrates the feature information. This method provided sufficient feature information for the network model.
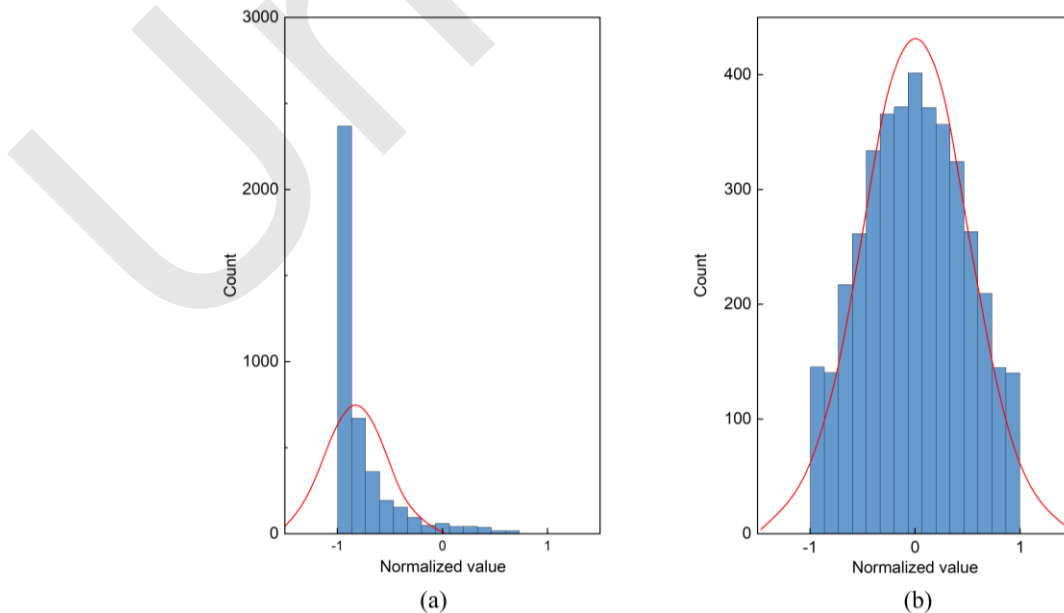

Fig. 2 The density histogram of GASF and GADF: (a) GASF matrix density histogram; (b) GADF matrix density histogram
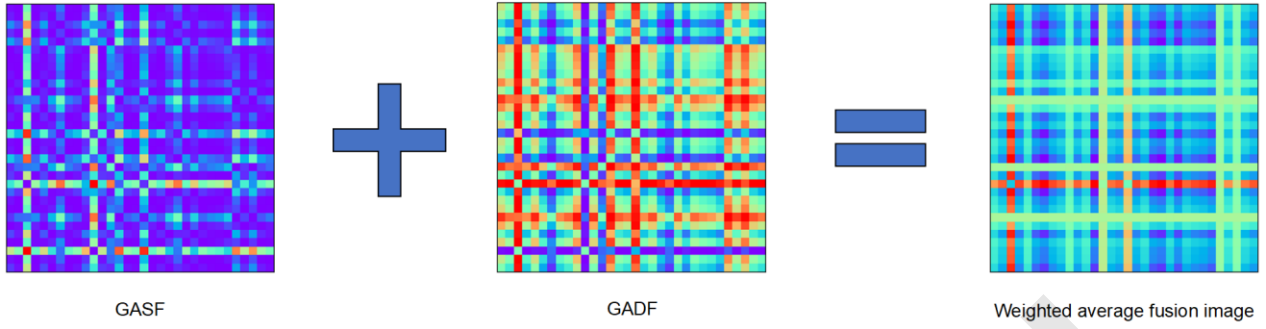
Fig. 3 Image-fusion process

2.3 Multi-attention mechanism

In CNN training, a large amount of channel information is generated. This feature information has unequal value, which affects the CNN training process and classification accuracy. Therefore, we introduced the channel attention mechanism model (Wang et al., 2020) and improved it for the purposes of this study. The features were first subjected to maximum and average pooling. The global information was compressed into the channel dimension to establish the information relationship between channels. After pooling, the feature information was fed into the squeezed-excitation network. The information for each channel was compressed into a weight that represented the global information response of the channel. By adjusting the

dimensionality reduction factor (r), we were able to discard unimportant feature information.

The input matrix is $X \in \mathbb{R}^{H \times 1 \times L}$, H is the number of channels, and L is the length. $c \in \mathbb{R}^{H/r \times 1 \times 1}$ is the intermediate vector. The output of the squeezed-excitation network can be summed to obtain the vector $c' \in \mathbb{R}^{H \times 1 \times 1}$. The output is:

$$C = X\sigma(c') \tag{7}$$

The concept of residual was introduced in channel attention mechanism. By this method, the original information is retained to improve the feasibility of optimizing the network. Connected by residuals, the output is $C' = X + C$. Fig. 4 shows the channel attention mechanism.
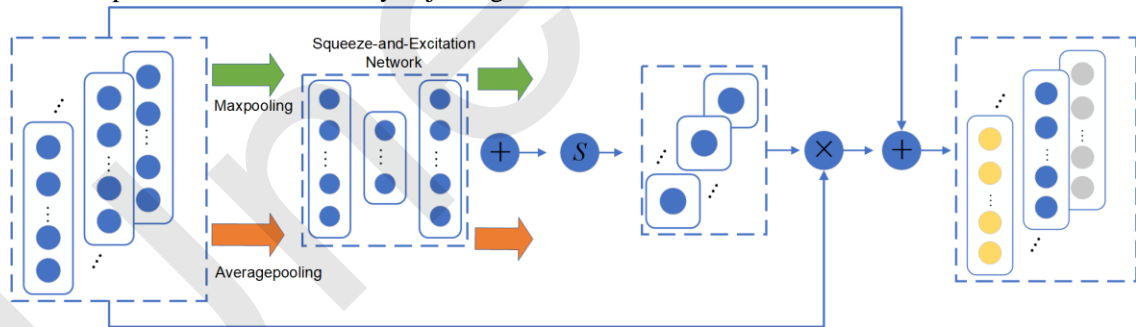

Fig. 4 Channel attention mechanism

The temporal attention mechanism (Tian et al., 2022) determines the signal segment of the fault feature information. The premise is to retain the original distribution of feature information in the temporal dimension. Pooling operations inevitably result in the loss of some feature information. Therefore, we introduced the $1 \times 1 \times 1$ convolution operation to obtain the weights of the feature information in the time dimension. Feature information can be mapped non-linearly to higher dimensions. The weight vector obtained was $t \in \mathbb{R}^{1 \times 1 \times L}$, which had been processed by the

channel attention mechanism.

Both attention mechanisms used the sigmoid function to generate modulation weights. These were multiplied by the original signal to obtain:

$$T = C'\sigma(t) = C'\sigma(w_{i,j}^{l}{}' \times c_{i,j}^{l}{}' + b_{j}^{l}{}') \tag{8}$$

where C' is the input matrix, $w_{i,j}^{l}{}'$ and $b_{j}^{l}{}'$ are the weight and deviation generated by the convolution operation.

The residual idea was similarly introduced in

temporal attention mechanism. $T' = C' + T$. Fig. 5 shows the temporal attention mechanism.
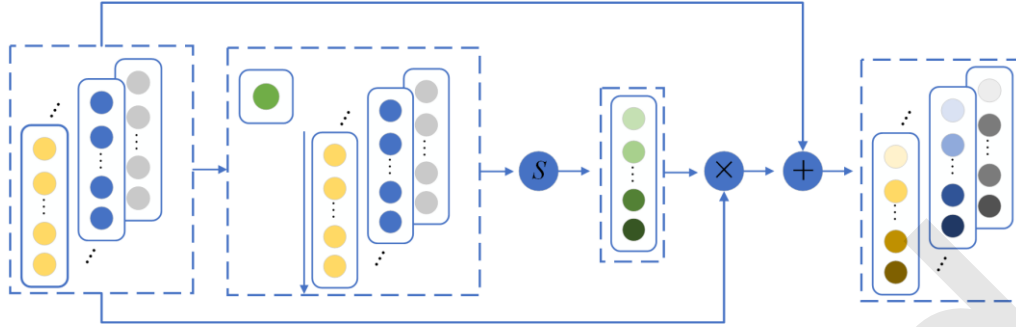
In designing the model, the channel attention mechanism and time attention mechanism were successively arranged to extract the features of key channels and key signal segments, respectively.

Fig. 5. Temporal attention mechanism

3 Experimentation and analysis

3.1 Experimental setup

The experimental equipment was a 4135AC marine power-generation diesel engine. Table 1 shows the main parameters and the diesel engine phase diagram is shown in Fig. 6. The vibration acceleration sensor model was INV9822, a general-purpose piezoelectric acceleration sensor. Due to the magnetic seat, the acceleration sensor and the diesel cylinder head could be magnetically attached. The sensitivity of the vibration was 100 mV/g and the sampling time and sampling frequency were 10 seconds and 20.48 kHz, respectively. The sampling interval was 30 seconds. The data-collector model was INV3062. The vibration sensors were mounted between the intake and exhaust valves of the diesel engine, which minimized interference from adjacent cylinders. The opening and closing moments of the intake and exhaust valves were different for different cylinders. We used a sensor (type SZB-16L) to collect the top dead center (TDC) signal. Based on the TDC signal and the phase diagram of the diesel engine, we were able to determine the opening and closing times of the intake and exhaust valves for each cylinder. The sensor arrangement and valve clearance are shown in Fig. 7. During the experiment, vibration signals were collected for each cylinder. In the data-processing and fault-diagnosis process, we used the vibration signal of the 1st cylinder as an example for analysis. Because the screw was hollow, its main function was to hold the cylinder-head cover in place. Passing the sensor through it did not damage

or interfere with the structure of the engine. The fault-simulation experimental scheme is shown in Table 2; the vibration signals of the normal and abnormal valve clearance of the diesel engine were acquired at rated speeds of 1500 r/min and 75% load, respectively.
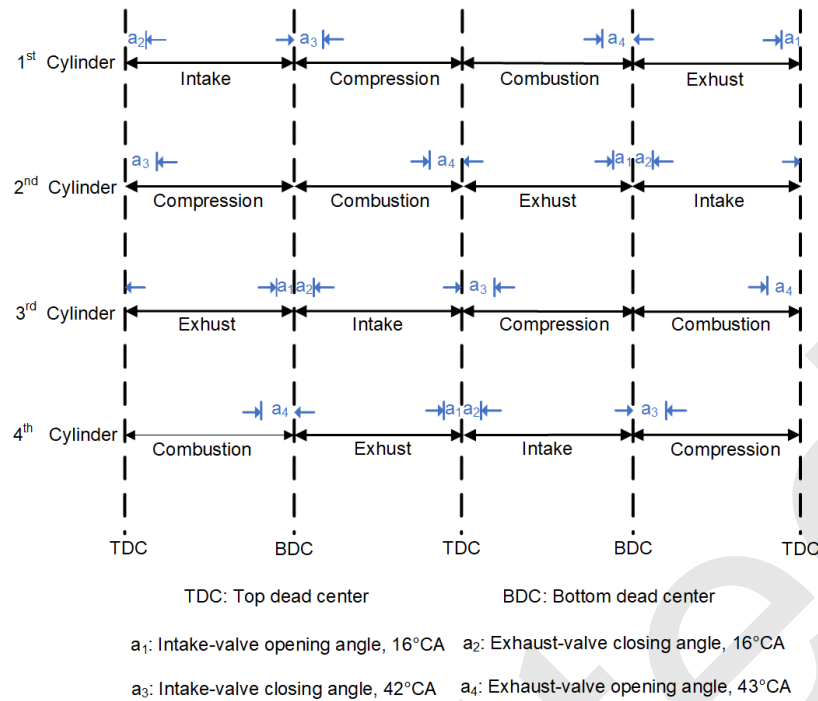
Table 1 The main parameters of the diesel engine

Machine number	A0422497
Cylinder diameter	135 mm
Intake-valve clearance	0.25 mm
Exhaust-valve clearance	0.30 mm
Stroke	150 mm
Continuous power	66.2 kW

Table 2 The fault-simulation experimental scheme

Label	Fault type	Valve clearance (mm)
R1	Normal	0.25 (intake), 0.30 (exhaust)
R2	Abnormal intake-valve clearance	0.15
R3		0.35
R4		0.65
R5	Abnormal exhaust-valve clearance	0.20
R6		0.45
R7		0.70

a_1: Intake-valve opening angle, 16°CA a_2: Exhaust-valve closing angle, 16°CA

a_3: Intake-valve closing angle, 42°CA a_4: Exhaust-valve opening angle, 43°CA

Fig. 6 Phase diagram of diesel engine
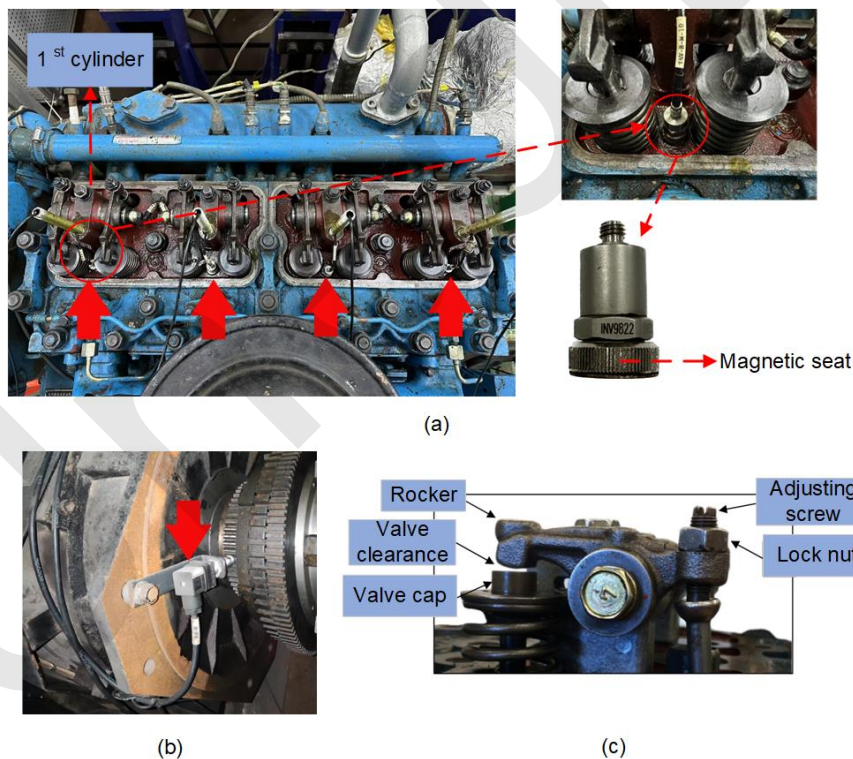
Fig. 7 Arrangement of sensors and valve clearance: (a) vibration sensors; (b) TDC sensor; (c) valve clearance

3.2 Signal analysis

After collecting the vibration signals for the different health states, we used the normal state and abnormal exhaust-valve clearance as examples. Referring to the diesel engine phase diagram and the TDC signal, the time-domain diagram of an operating cycle for the 1st cylinder is shown in Fig. 8. Both combustion and valve opening and closing moments transmit significant shock signals to the cylinder head. The change in valve clearance affects the opening and closing time, motion speed, and acceleration of the

valve. This in turn affects the impact and energy of the valve when it is seated. Fig. 8a-8b shows that the time-domain waveform when the exhaust-valve clearance was slightly abnormal was not much different from the normal waveform. There was no significant change in the shock of the exhaust valve to the valve seat. According to the information shown in Fig. 8c, when the exhaust-valve clearance was seriously abnormal, the valves opened late and closed early. The amplitude was up to 431 m/s², two to three times higher than normal. This reflected the dramatic rise in signal shock and energy. The valve clearance may be slightly abnormal in practical engineering. It is difficult to determine whether abnormal valve clearances have occurred through the time-domain waveform. Hence, further processing of the vibration signals was necessary.
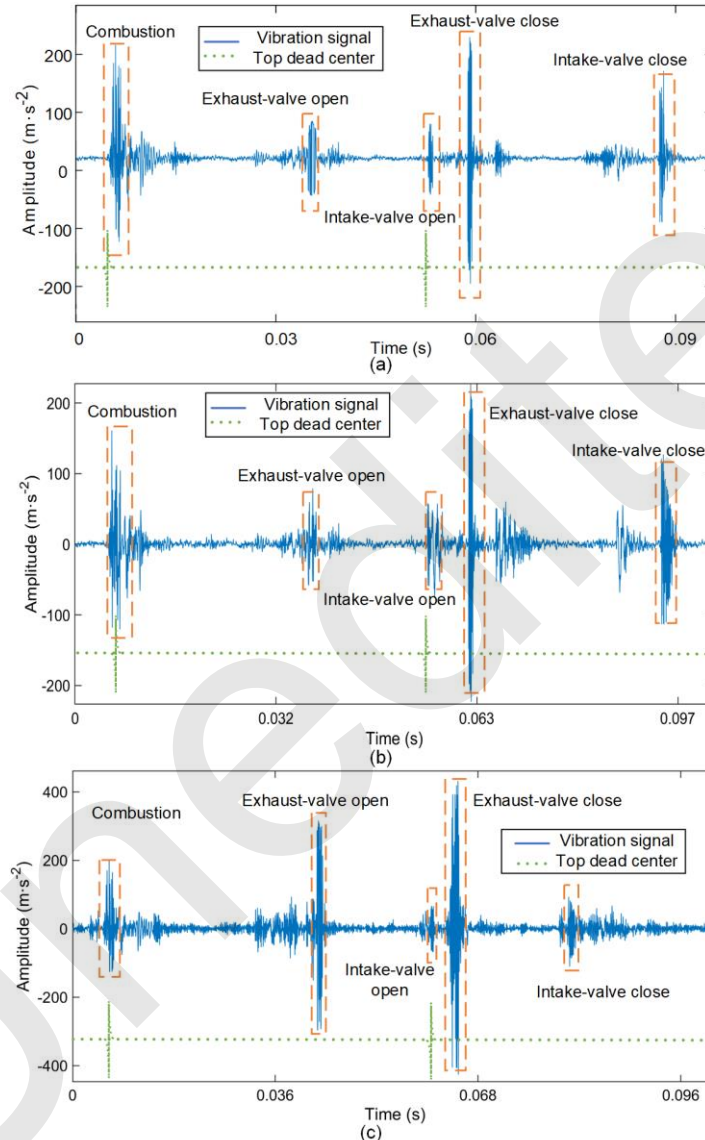
Fig. 8 Time-domain diagram of normal condition and abnormal exhaust-valve clearance: (a) R1; (b) R6; (c) R7

To link computer vision and fault diagnosis, we transformed the original signals into 2-D images. Fig. 9 shows the results of 2-D image conversion for the four health states. When the valve clearance was abnormal, the texture, and color features of the 2-D image were more complex. This was due to the increased excitation force and excitation energy on the valve seat. When the amplitude was small, the crossover feature with lighter color appeared in the feature map. The vertical and horizontal highlighting bands indicate rapid changes in amplitude. The shock from the opening and closing of the valve was well retained. Figs. 9 panels a, f, and g indicate normal, slightly abnormal, and severely abnormal exhaust-valve clearance, respectively. The vibration signals are converted into 2-D images, the vibration shock of slightly abnormal exhaust-valve clearance can also be indicated. This overcomes the defect of

the time-domain signals not being very distinguishable. With the GAF method, the time series is uniquely mapped in the coordinate transformation and the integrity of the signal is preserved. The fused images of different health states had marked differences in texture, color, and other features. This provided the basis for the CNN to extract feature information and perform fault diagnosis.
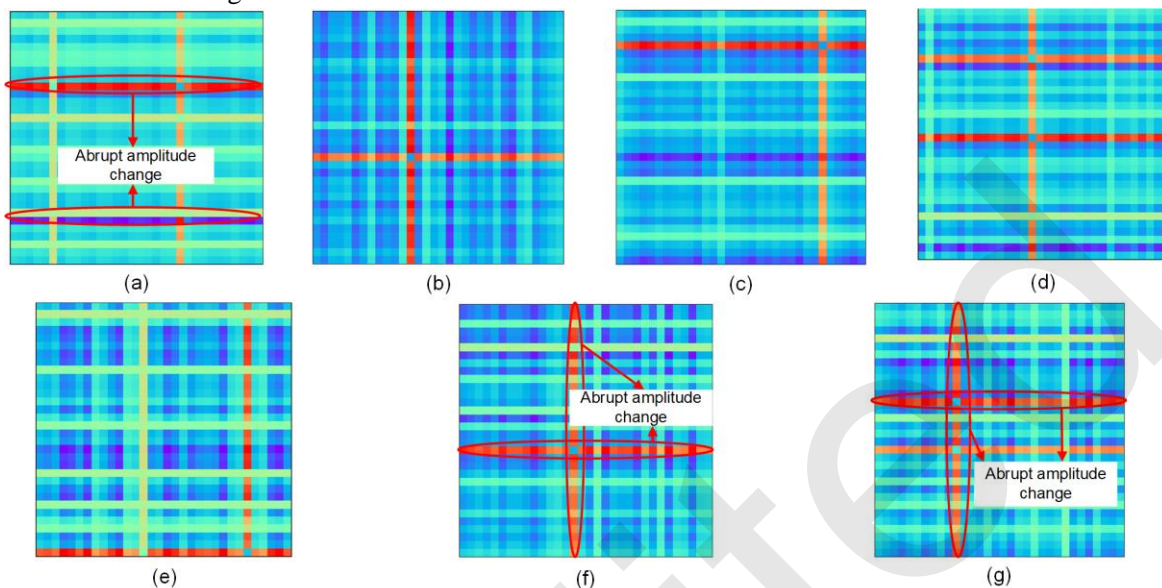

Fig. 9 Two-dimensional images of the four health states: (a) R1; (b) R2; (c) R3; (d) R4; (e) R5; (f) R6; (g) R7

3.3 CNN structure and parameter selection

The deep learning setup included an AMD Ryzen 7 5800H CPU@3.2GHz, NVIDIA GeForce RTX 3060 Laptop GPU, and the win11 operating system. In a squeeze-incentive network with a channel-attention mechanism, the value of the dimensionality reduction factor (r) needs to be discussed. The r selectable values were 4, 8, 16, 32, and 64. The results of different r values are shown in Table 3. When r changed from 4 to 64, the training time dropped by 125 seconds and diagnostic accuracy decreased by 2.78%. Therefore the value of r had little effect on model performance. We set the value of r to 4.

Table 3 Comparison of results for different dimensionality-reduction coefficients

r	Training time (s)	Number of model parameters ($\times 10^5$)	Accuracy (%)
4	351	0.94	66.35
8	327	0.71	64.51
16	268	0.59	64.26
32	243	0.53	64.18
64	226	0.51	63.57

Hyperparameters can affect CNN performance, so one should set the learning rate, MaxEpoch, and other parameters in advance. We will use the learning rate and MaxEpoch as examples to show the hyperparameter setting process. The optional range of the learning rate was 0.0001, 0.001, 0.01. The performance of the CNN is shown in Fig. 10. When the range was 0.001, the CNN had the highest accuracy and shortest running time, so we chose this value. The optional value of MaxEpoch is 20, 30, 40, 50, 80. The results of the experiment are shown in Fig. 11. When MaxEpoch = 20, the network model required less time and had high diagnostic accuracy. Therefore, we chose this value for the experiment.

The other parameters were set with a similar process. Table 4 shows the hyperparameters. The CNN model had 17 layers and its structure is shown in Table 5.

Table 4 Optimal values for the main hyperparameter settings

L2Regularization	0.004
LearnRateDropFactor	0.05
LearnRateDropPeriod	10
InitialLearnRate	0.001
Slover	Adam
MaxEpochs	20
MiniBatchSize	10

Table 5 The CNN structure and main parameters

Name	Structure parameters	Output size
Image input	$64 \times 64 \times 3$	$64 \times 64 \times 3$
Conv_1	32@2×2, Stride= [1], Padding= [2]	$67 \times 67 \times 32$
Batchnorm_1	-	$67 \times 67 \times 32$
Maxpool_1	3×3, Stride= [2], padding= [0]	$33 \times 33 \times 32$
Multi-attention mechanism_1	64@$1 \times 1 \times 1$ (Time attention mechanism)	$33 \times 33 \times 32$
Relu_1	ReLu	$33 \times 33 \times 32$
Conv_2	32@3×3, Stride= [1], Padding= [2]	$35 \times 35 \times 32$
Batchnorm_2	-	$35 \times 35 \times 32$
Maxpool_2	3×3, Stride= [2], Padding= [0]	$17 \times 17 \times 32$
Multi-attention mechanism_2	128@$1 \times 1 \times 1$ (Time attention mechanism)	$17 \times 17 \times 32$
Relu_2	ReLu	$17 \times 17 \times 32$
Fc_1	-	$1 \times 1 \times 100$
Relu_3	ReLu	$1 \times 1 \times 100$
Dropout	0.5	$1 \times 1 \times 100$
Fc_2	-	$1 \times 1 \times 7$
Softmax	-	$1 \times 1 \times 7$
Class output	-	$1 \times 1 \times 7$
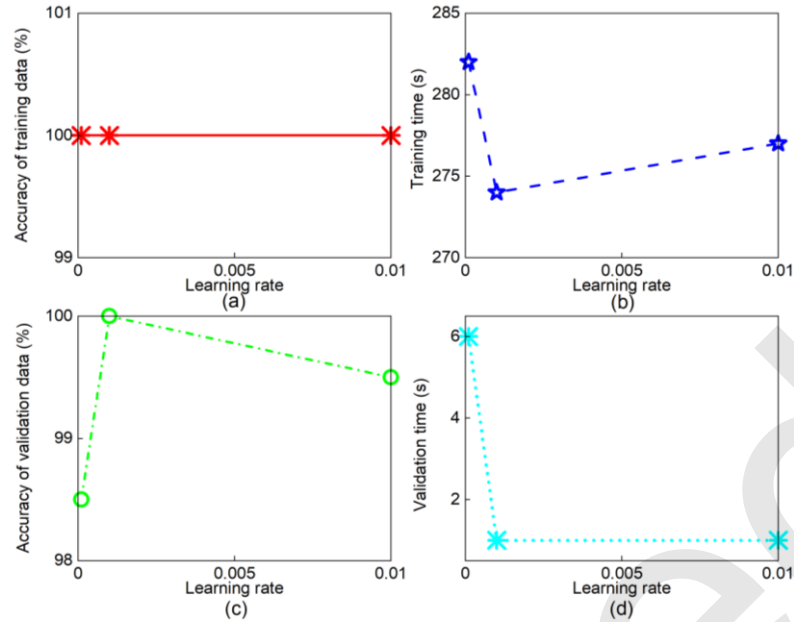
Fig. 10 Experimental results of different learning rates: (a) training accuracy; (b) training time; (c) validation accuracy; (d) validation time
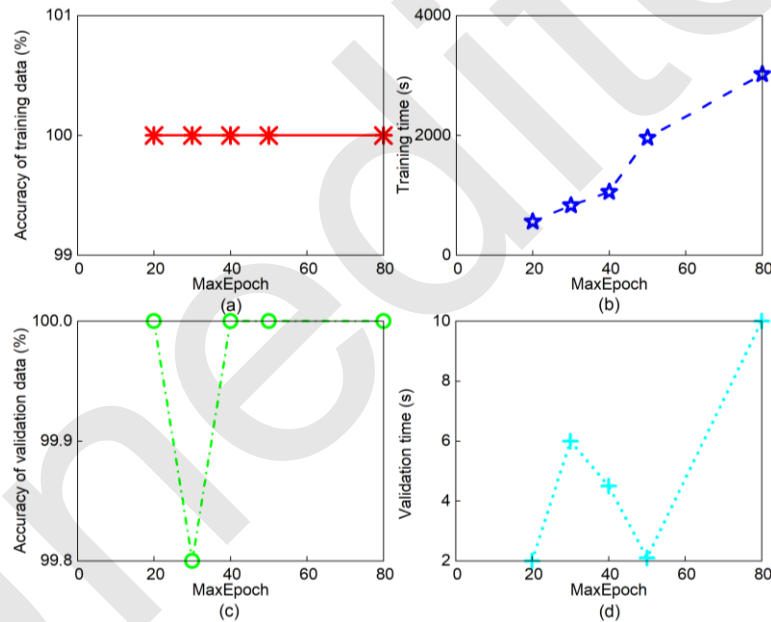
Fig. 11 Experimental results with different MaxEpoch values: (a) training accuracy; (b) training time; (c) validation accuracy; (d) validation time

3.4 CNN performance

500 data sets were collected for each health state. We randomly selected 288 sets for training, and 12 for validation. The rest were used as test data. The model was iterated 98 times in each round, a total of 1960 times. Fig. 12 shows the network training and validation process. As the number of iterations increased, the accuracy of the training and validation data increased to 100%. The loss function curve was gradually smooth. The loss function values for the training and validation data were 0.0250 and 0.0292,

respectively, and their values were close to 0. The training time was about 211 seconds, and the validation time was about three seconds. No over-fitting occurred during the model run.

Fig. 13 shows the confusion matrix of the testing data. Green represents the label with the correct prediction. A total of twenty-two samples were misclassified. The classification accuracy for the seven states was 100%, 98.5%, 97.5%, 100%, 95.5%, 98.0%, and 100%. The average accuracy was 98.4%.
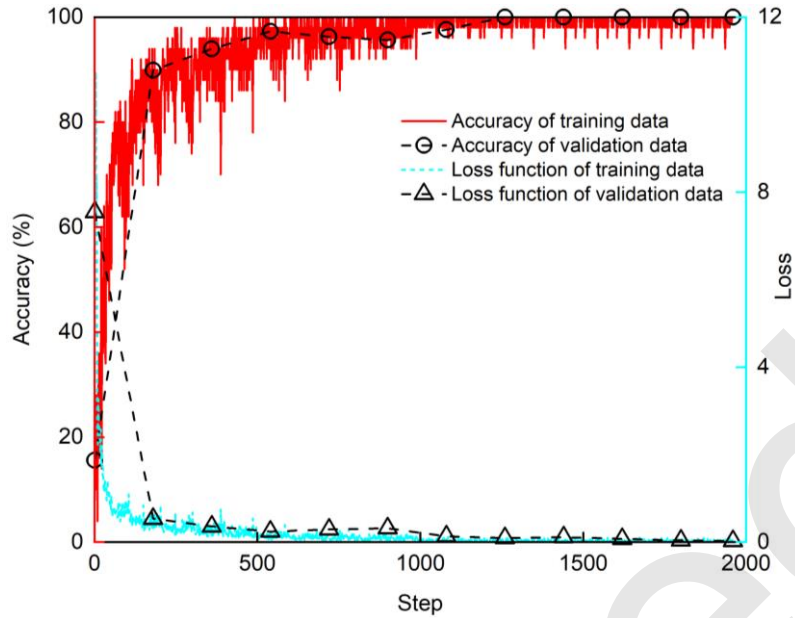
Fig. 12 The CNN model training and validation process

Fig. 13 The confusion matrix for test data (Green represents the label with the correct prediction).

3.5 Noise immunity of the model

The diesel engine operates in a harsh environment, so it is essential to study the adaptability of the proposed model to noise disturbances. We added noise with signal-to-noise ratios (SNRs) of -20, 0, 10, and 20 dB to the raw signals. The training data were from 80 groups, and the testing data from 60 groups. The experiment was repeated five times for each SNR. The average diagnostic accuracy and average loss-function values are shown in Table 6. In all four cases, the average

accuracy rate was above 94%.

The SNR equation is as follows:

$$\mathrm{SNR_{dB}} = 10\log_{10}\frac{P_{\mathrm{Signal}}}{P_{\mathrm{Noise}}} \quad (9)$$

where P_{Signal} is the power of the original signal and P_{Noise} is the power of added noise.

From the equation, it is evident that the smaller the SNR, the greater the intensity of the noise in the signal.

To verify the improvement of the noise-immunity performance of the CNN model

effected by the multi-attention mechanism, we fed the data (SNR=-20 dB) into two network models, one with a no-attention mechanism (NAM), and one with a multi-attention mechanism (MAM). All other parameters were the same for both models. The experiment was repeated five times. The average accuracy for different health states is shown in Fig. 14. The fault-diagnosis accuracy of the model with the attention mechanism was 14.8% higher than that of the model without it. Thus, the attentional mechanism appears to improve the noise resistance of the model.

SNR (dB)	Average training accuracy (%)	Average testing accuracy (%)	Average loss
-20	98.8	94.6	0.0581
0	99.1	96.6	0.0474
10	99.3	98.1	0.0367
20	99.6	98.3	0.0304

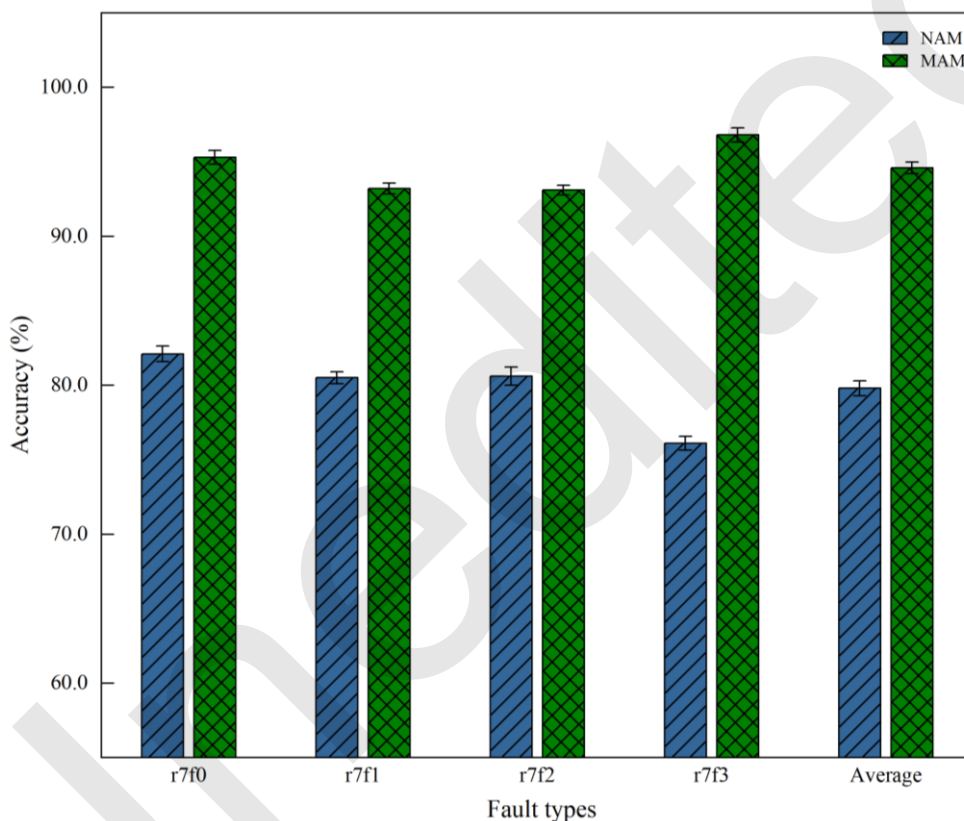**Table 6 CNN model performance with different SNRs**

Fig. 14 Influence of an attention mechanism on CNN performance

3.6 Evaluation of different input methods

To verify the effectiveness of the fused image method, we input the fused images, GASF images, and GADF images into the CNN model. The same model is used for the different input methods.

Fig. 15 shows the accuracy of different input methods. The comparison shows that the image-fusion method has the highest fault-diagnosis accuracy. The overall diagnostic accuracy using GASF images and GADF images methods is 94.0% and 84.9%, respectively. The image fusion method has better signal visualization, and can also provide more comprehensive feature information for the model, which is conducive to improving fault-diagnosis accuracy.
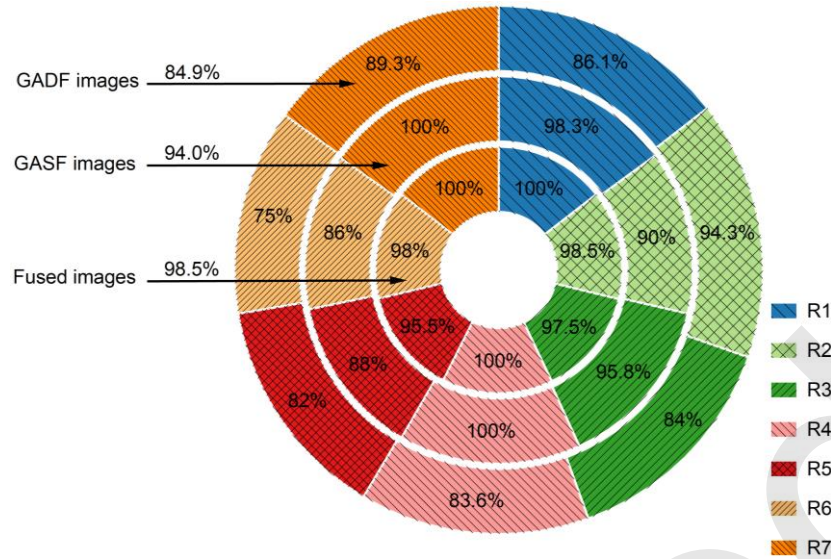
Fig. 15 Accuracy of different input methods

3.7 Adaptability to variable loads

In practical engineering, collecting data under all loads is difficult. Therefore, a model trained under a specific data load ought to be able to perform well under other loads. Based on the original 75% load data set, we collected the vibration signals under 25%, 50%, and 100% load . The data from the three loads were used as training data, and the data from one load was used as testing data (Peng et al., 2020). In this way, we were able to improve the generalization ability of the network model.
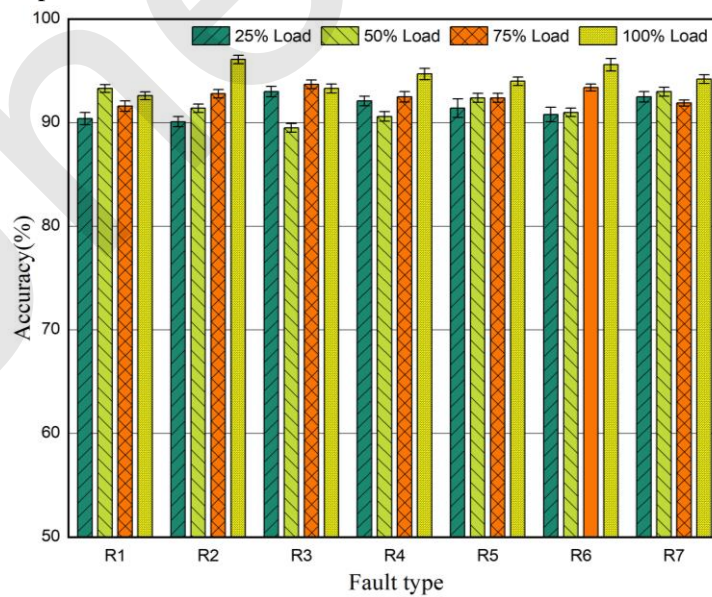
The experiment was repeated five times. The average diagnostic accuracy is shown in Fig. 16. Diagnostic accuracy of more than 89% was achieved for all the health states. The CNN model is quite adaptable to variable loads. Image fusion can provide more comprehensive feature information to the CNN model. The multi-attention mechanism was introduced to extract key and potential feature information in the channel and time dimensions. In summary, the proposed method can adaptively learn feature information from data and has a strong capability for generalization.

Fig. 16 Model diagnosis results under variable load

3.8 Comparative analysis of the performance of different methods

Different fault-diagnosis methods and CNN models were compared to validate the superiority of the proposed method. We also analyzed vibration

signals with SNRs of -20 ~ 20 dB for validation. The contrasted analysis methods included GoogleNet (Grover et al., 2022), AlexNet (Hajnayeb, 2021), Modified complete ensemble empirical mode decomposition with adaptive noise (MCEEMDAN)-CNN (Hou et al., 2022), and Continuous wavelet transforms (CWT)-CNN (Fu et al., 2023). It should be noted that both GoogleNet and AlexNet needed to be pre-trained. The image sizes input to GoogleNet and AlexNet were 224×224×3 and 227×227×3, respectively. MCEEMDAN (Hou et al., 2022) can realize signal noise reduction, eliminate modal aliasing, and lay the foundation for subsequent CNN feature extraction. CWT (Fu et al., 2023) can efficiently process non-smooth signals and characterize their local features; it helps a CNN to extract feature information. The above network models or diagnostic methods have yielded good results in machinery fault diagnosis. The fault-diagnosis accuracy of the four methods and our proposed method is shown in Fig. 17.
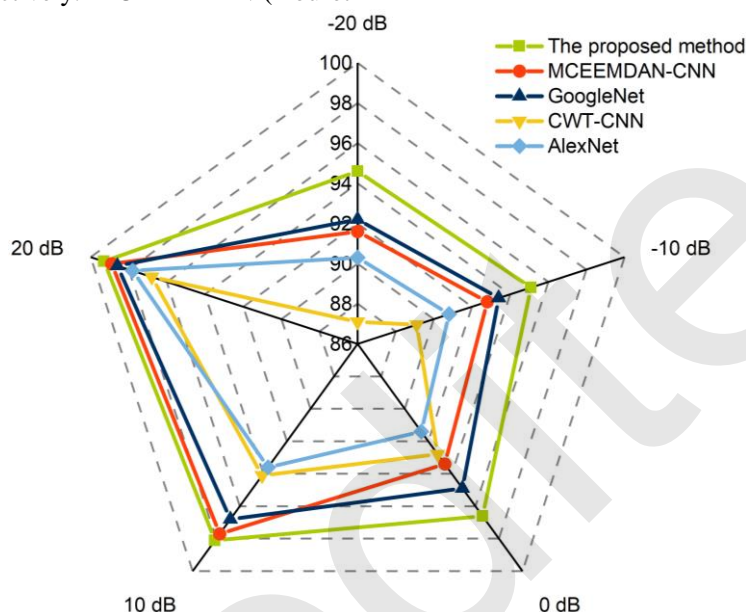

Fig. 17 Fault-diagnosis accuracy of different methods

The performance of the different methods decreased with the SNR. When SNR=-20 dB, the fault-diagnosis accuracies of GoogleNet, MCEEMDAN-CNN, AlexNet, and CWT-CNN were 92.2%, 91.6%, 90.3%, and 87.1%, respectively. The accuracy of the proposed method is 94.6%, which is higher than that of the other four methods. This further validates the finding that the proposed method offers good stability and noise immunity.

4 Discussion

During the experiment, there were certain limitations and assumptions, primarily:

1. According to the diesel-engine manual, adjustment of the valve clearance needs to be carried out when the diesel engine is in a completely cold state. However, in the actual experiment, due to the existence of temperature difference, there was a certain amount of error in adjusting the valve clearance.

2. In the variable-load experiment, a hydraulic dynamometer was used to change the load on the diesel engine. This method was used to simulate the variable load of power-generation diesel engines in actual ship navigation. The dynamic response was slightly different from actual ship conditions.

Although some progress was made in this study, there are several remaining problems.

1. The proposed method was validated only for the case of abnormal valve clearance. In practical engineering, diesel-engine fault states show diversity, with single and compound faults co-existing. Whether the proposed method applies to other fault types and other marine machinery remains to be verified.

2. Introducing the multi-attention mechanism increased the network-model computation, and the network-model training required more time. This puts higher demands on the computer hardware configuration.

3. In practical engineering, the problems of difficult data collection and costly sample labelling

may result in a small amount of training data, or the amount of data may be large but of insufficient quality. The proposed method may not be able to effectively mine equipment health-state feature information from low-quality data.

Therefore, the focus of subsequent study will be as follows:

1. Experimental validation should be extended to include more fault types for fault diagnosis of marine power-generation diesel engines. In follow-up work, we will focus on the key component faults and compound faults of diesel engines to verify the generalizability of the proposed method. In addition, it is necessary to further increase the diversity of samples in the database and explore the validity of the proposed method for other marine machinery, as well as its adaptability to harsh environments.

2. The network model should be kept as lightweight as possible while ensuring that the model performs well.

3. To support specific equipment and monitoring of specific physical quantities, relevant data monitoring, transmission, and storage standards should be formulated to consolidate the data foundation for deep-learning-based marine machinery fault-diagnosis technology. Furthermore, from the perspective of data augmentation, a diffusion model, transfer learning, a generative adversarial network, and other methods could be utilized to address the problem of designing diagnostic frameworks for network models in the case of small samples.

All in all, deep-learning-based intelligent fault-diagnosis methods for ships still have a long way to go in order to move from the lab to real ship applications.

5 Conclusions

In this work, we designed and tested a valve-clearance fault-diagnosis model based on the GAF and a CNN. The conclusions are as follows:

1. Using a multi-attention mechanism, interference information such as noise can be suppressed, and the CNN feature-extraction capability can be optimized. With strong noise, the diagnostic accuracy of the model with a multi-attention mechanism is 14.8% better than the model without the mechanism.

2. The average accuracy is above 94% with all SNRs. The proposed method has better noise immunity and stability than the four other methods.

3. Fused images provide more adequate feature information for the neural network, which improves fault-identification accuracy.

4. The variable-load experiments illustrate that the accuracy of the proposed model can be maintained above 89%, proving that it has good stability.

Author contributions

Congyue LI and Yihuai HU designed the research. Congyue LI wrote the first draft of the manuscript. Jiawei JIANG and Dexin CUI helped to organize the manuscript. Congyue LI and Yihuai HU revised and edited the final version.

Conflict of interest

Congyue LI, Yihuai HU, Jiawei JIANG, and Dexin CUI declare that they have no conflict of interest.

References

Alsalaet JK, Hajnayeb A, Bahedh AS, 2023. Bearing fault diagnosis using normalized diagnostic feature-gram and convolutional neural network. *Measurement Science and Technology*, 34(4):5901.
https://doi.org/10.1088/1361-6501/ACAD1F

Cai BP, Sun XT, Wang JX, et al., 2020. Fault detection and diagnostic method of diesel engine by combining rule-based algorithm and BNs/BPNNs. *Journal of Manufacturing Systems*, 57(2020):148-157.
https://doi.org/10.1016/j.jmsy.2020.09.001

Çağlar K, Yasin A, 2022. Development of condition-based maintenance strategy for fault diagnosis for ship engine systems. *Ocean Engineering*, 256(2022):111515.
https://doi.org/10.1016/J.OCEANENG.2022.111515

Cui JL, Zhong QW, Zheng SB, et al., 2022. A lightweight model for bearing fault diagnosis based on Gramian angular field and coordinate attention. *Machines*, 10(4): 282.
https://doi.org/10.3390/MACHINES10040282

Dhamande LS, Chaudhari MB, 2016. Bearing fault diagnosis based on statistical feature extraction in time and frequency domain and neural network international. *Journal of Vehicle Structures and Systems*, 8(4):229-240.
https://doi.org/10.4273/ijvss.8.4.09

Du JF, Li XY, Gao LP, 2022. Integrated gradient-based continuous wavelet transform for bearing fault diagnosis. *Sensors*, 22(22):8760.
https://doi.org/10.3390/S22228760

Fu WL, Jiang XH, Li BL, et al., 2023. Rolling bearing fault diagnosis based on 2D time-frequency images and data augmentation technique. *Measurement Science and Technology*, 34(4):045005.
https://doi.org/10.1088/1361-6501/ACABDB

Gou LF, Li HH, Zheng H, et al., 2020. Aeroengine control system sensor fault diagnosis based on CWT and CNN. *Mathematical Problems in Engineering*, 2020(2020):1-12. https://doi.org/10.1155/2020/5357146

Grover C, Turk N, 2022. A novel fault diagnostic system for rolling element bearings using deep transfer learning on bispectrum contour maps. *Engineering Science and Technology, an International Journal*, 31(2022):101049 https://doi.org/10.1016/J.JESTCH.2021.08.006

Hoang DT, Kang HJ, 2018. A survey on deep learning based bearing fault diagnosis. *Neurocomputing*, 335(2018):327-335. https://doi.org/10.1016/j.neucom.2018.06.078

Hu J, Yu YH, Yang JG et al., 2023. Research on the generalisation method of diesel engine exhaust valve leakage fault diagnosis based on acoustic emission. *Measurement*, 210(2023):112560. https://doi.org/10.1016/J.MEASUREMENT.2023.112560

He Y, Tang HS, Ren Y, et al., 2022. A deep multi-signal fusion adversarial model based transfer learning and residual network for axial piston pump fault diagnosis. *Measurement*, 192(2022):110889. https://doi.org/10.1016/J.MEASUREMENT.2022.110889

Hajnayeb A, Sun Q, 2021. Cavitation analysis in centrifugal pumps based on Vibration Bispectrum and Transfer Learning. *Shock and Vibration*, 2021(2021):1-8. https://doi.org/10.1155/2021/6988949

Hou SZ, Guo W, Wang ZQ, et al., 2022. Deep-learning-based fault type identification using modified CEEMDAN and image augmentation in distribution power grid. *IEEE Sensors Journal*, 22(2):1583-1596. https://doi.org/10.1109/JSEN.2021.3133352

Mariela C, Grover Z, Diego C, et al., 2016. Fault diagnosis in spur gears based on genetic algorithm and random forest. *Mechanical Systems and Signal Processing*, 70:87-103. https://doi.org/10.1016/j.ymssp.2015.08.030

Manarikkal I, Elasha F, Mba D, 2021. Diagnostics and prognostics of planetary gearbox using CWT, auto regression (AR) and K-means algorithm. *Applied Acoustics*, 184(2021):108314. https://doi.org/10.1016/J.APACOUST.2021.108314

Nayana BR, Geethanjali P, 2017. Analysis of statistical time-domain features effectiveness in identification of bearing faults from vibration signal. *IEEE Sensors Journal*, 17(17):5618-5625. https://doi.org/10.1109/jsen.2017.2727638

Peng DD, Wang H, Liu ZL, et al., 2020. Multibranch and multiscale CNN for fault diagnosis of wheelset bearings under strong noise and variable load condition. *IEEE Transactions on Industrial Informatics*,16(7):4949-4960. https://doi.org/10.1109/tii.2020.2967557

Pan JH, Qu LL, Peng KX, 2021. Sensor and actuator fault diagnosis for robot joint based on deep CNN. *Entropy*, 23(6):751. https://doi.org/10.3390/E23060751

Qian CH, Zhu JJ, Shen YH, et al., 2022. Deep transfer learning in mechanical intelligent fault diagnosis: application and challenge. *Neural Processing Letters*, 54(3):1-23. https://doi.org/10.1007/S11063-021-10719-Z

Rao X, Sheng CX, Guo ZW, et al., 2022. A review of online condition monitoring and maintenance strategy for cylinder liner-piston rings of diesel engines. *Mechanical Systems and Signal Processing*, 165(2022):108385. https://doi.org/10.1016/J.YMSSP.2021.108385

Ren HR, Liao XJ, Li ZW, et al., 2018. Anomaly detection using piecewise aggregate approximation in the amplitude domain. *Applied Intelligence*, 48(5):1097-1110. https://doi.org/10.1007/s10489-017-1017-x

Ren K, Zhang DW, Wan MJ, et al., 2022. An infrared and visible image fusion method based on improved DenseNet and mRMR-ZCA. *Infrared Physics & Technology*, 115:103707. https://doi.org/10.1016/J.INFRARED.2021.103707

Sun F, Xu H, Zhao Y, et al., 2022. Data-driven fault diagnosis of control valve with missing data based on modeling and deep residual shrinkage network. *Journal of Zhejiang University-SCIENCE A*. 23(4): 303–313. https://doi.org/10.1631/jzus.A2100598

Song RW, Yu BQ, Shi H, et al., 2023. Support vector machine fault diagnosis based on sparse scaling convex hull. *Measurement Science and Technology*, 34(3): 035101. https://doi.org/10.1088/1361-6501/ACA217

Schmidhuber J, 2015. Deep learning in neural networks: an overview. *Neural Networks*, 61(2015):85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Senanayaka JSL, Khang HV, Robbersmyr KG, 2019. Multiple classifiers and data fusion for robust diagnosis of gearbox mixed fault. *IEEE Trans. Industrial Informatics*, 15(8):4569-4579. https://doi.org/10.1109/TII.2018.2883357

Tian HX, Li RJ, Yang LZ, 2022. Operation status monitoring of reciprocating compressors based on the fusion of spatio-temporal multiple information. *Measurement*, 204(2022):112087. https://doi.org/10.1016/J.MEASUREMENT.2022.112087

Wen L, Li XY, Gao L, 2019. A transfer convolutional neural network for fault diagnosis based on ResNet-50. *Neural Computing and Applications*, 32(10):1-14. https://doi.org/10.1007/s00521-019-04097-w

Wang B, Lei YG, Li NP, et al., 2020. Multi-scale convolutional attention network for predicting remaining useful life of machinery. *IEEE Transactions on Industrial Electronics*, 68(8):1. https://doi.org/10.1109/tie.2020.3003649

Xu YD, Yan XA, Feng K, et al., 2022. Attention-based multiscale denoising residual convolutional neural networks for fault diagnosis of rotating machinery. *Reliability Engineering and System Safety*, 226(2022):108714. https://doi.org/10.1016/J.RESS.2022.108714

Xie JS, Lin MQ, Yang BY, et al., 2023. A novel bearing fault

diagnosis method under small samples using time-frequency multi-scale convolution layer and hybrid attention mechanism module. *Measurement Science and Technology*, 34(9):5121.
https://doi.org/10.1088/1361-6501/ACDC45

Zhao R, Yan RQ, Chen ZH, et al., 2019. Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing*, 115(2019):213-237.
https://doi.org/10.1016/j.ymssp.2018.05.050

中文概要

题　　目：基于格拉姆角场和卷积神经网络的船用发电柴油机故障诊断

作　　者：李从跃[1]，胡以怀[1]，蒋佳炜[2]，崔德馨[1]

机　　构：[1]上海海事大学, 商船学院, 中国上海, 201306；[2]上海电子信息职业技术学院，机械与能源工程学院，中国上海，201411

目　　的：船用发电柴油机工作环境恶劣，在内外多激励源的干扰下，振动信号呈现非线性非平稳性特点。本文旨在如何对船舶发电柴油机的振动信号进行有效特征提取并准确识别故障类型。研究所提方法的有效性，以提高船舶发电柴油机的故障诊断精度。

创新点：1. 一维振动信号可通过格拉姆角场转换为二维图像。一维振动信号可以映射到 2D 图像的颜色、点、线和其他特征。为充分利用故障特征信息，将格拉姆角和场和格拉姆角差场获得的图像进行加权平均融合。2. 利用多注意力机制来优化卷积神经网络学习机制，使网络有选择地提取信号中的关键特征信息。

方　　法：1. 对船舶发电柴油机的气阀间隙进行故障预设，采集柴油机不同健康状态振动信号。2. 将振动信号转化为二维图像，将格拉姆角和场和格拉姆角差场获得的图像进行加权平均融合，充分利用原信号中的故障特征信息。 3. 将融合后的图像输入到卷积神经网络中，进行自适应特征提取和故障识别。

结　　论：1. 所提方法可准确识别故障类型，诊断精度可达 98.4%，与其他方法相比，所提方法具有更高的故障诊断精度。2. 在不同信噪比下，所提方法与无注意力机制方法相比，准确精度可提高 14.8%。3. 融合后的图像可为神经网络提供更充足的特征信息，可提高故障识别精度。4. 变负荷实验中，所提方法的准确率均保持在 89%以上，这进一步验证了所提方法的稳定性。

关键词：多注意力机制；卷积神经网路；格拉姆角场；图像融合；船舶发电柴油机；故障诊断