Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering) *2024 25(6):470-482*
www.jzus.zju.edu.cn; www.springer.com/journal/11582
E-mail: jzus_a@zju.edu.cn

Research Article
https://doi.org/10.1631/jzus.A2300273

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 2D images by taking advantage of the GAF, which preserves the temporal correlation. The raw signals can be mapped to 2D image features such as texture and color. To integrate the 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 98.40%. When −20 dB≤signal-to-noise ratio (SNR)≤20 dB, the diagnostic accuracy of the proposed method is higher than 94.00%. The proposed method has superior diagnostic performance. Moreover, it has a certain anti-noise capability and variable-load adaptive capability.

Key words: Multi-attention mechanisms (MAM); Convolutional neural network (CNN); Gramian angular field (GAF); Image fusion; 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; Karatuğ and Arslanoğlu, 2022; Rao et al., 2022). The air-distribution mechanism operates in a harsh environment, especially the valves and valve 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 (Dhamande and Chaudhari, 2016; Nayana and Geethanjali, 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 (Cerrada 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

✉ Congyue LI, licongyue0126@163.com
Congyue LI, https://orcid.org/0000-0002-9290-5119
Yihuai HU, https://orcid.org/0000-0003-3000-7855

Received May 21, 2023; Revision accepted Sept. 18, 2023;
Crosschecked Feb. 22, 2024; Online first Apr. 3, 2024

status of diesel engines are characterized by their large volume, diversity, and low-value density (Hoang and Kang, 2019; Qian et al., 2022). 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 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 (CWTs) (Du et al., 2022) are used to transform signals into 2D images (Fu et al., 2023). They are input into a convolutional neural network (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 (Gou 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 nonlinear 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) contain 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., 2020; 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 study we propose a new valve-clearance fault-diagnosis method. The main contribution points are as follows:

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

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

3. A multi-attention mechanism (MAM) 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 Gramian angular field

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

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

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

(2) $\tilde{x}_i$ (i=1, 2, $\cdots$, n) is mapped as the angle and t_i is mapped as the radius:

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

where ϕ_i denotes the polar angle, r denotes the polar radius, and $\tilde{X}$ is the signal sequence. t_i is a time stamp and N is a constant.

(3) Two images can be obtained. Eq. (3) is the formula for the Gramian angular summation field (GASF). Eq. (4) is the formula for the Gramian angular difference field (GADF).

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

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

where $\tilde{\boldsymbol{X}}$ denotes the signal sequence matrix and $\boldsymbol{I}=[1, 1, \cdots, 1]$.

If the time 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 2D image with the GAF.

2.2 Image fusion

Frequency information was extracted in the GASF and GADF 2D matrices and represented by density histograms (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., 2021). The weighted average fusion is calculated as follows:

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

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

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. $d_1(x,y)+d_2(x,y)=W$, where W is the image width.

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.

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., 2021) and improved it for the purposes of this

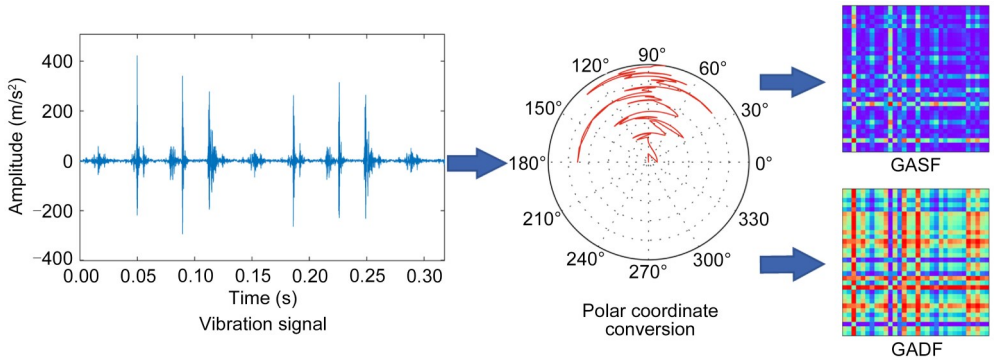

Fig. 1 GAF conversion process

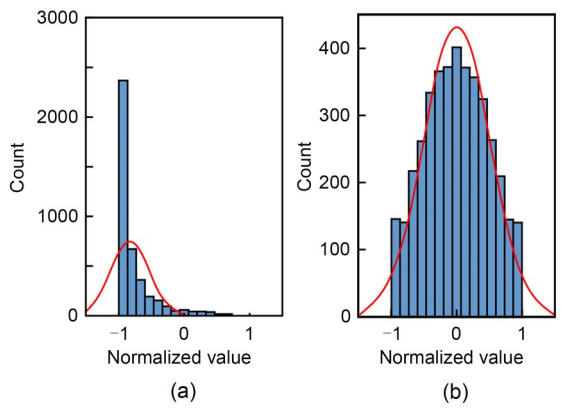

Fig. 2 Density histograms of GASF and GADF: (a) GASF matrix density histogram; (b) GADF matrix density histogram

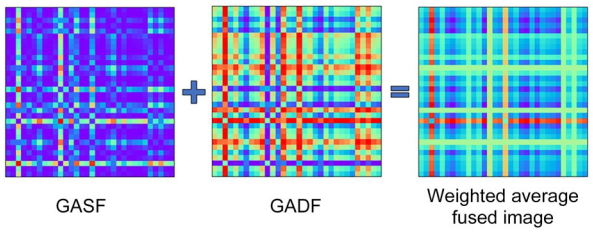

Fig. 3 Image fusion process

study. The features were first subjected to the 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 (m), we were able to discard unimportant feature information.

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

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

where σ denotes the activation function.

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 $\boldsymbol{C}'=\boldsymbol{X}+\boldsymbol{C}$. Fig. 4 shows the channel attention mechanism and S is the sigmoid function.

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 1×1×1 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 $\boldsymbol{t}\in\mathbb{R}^{1\times 1\times L}$, 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:

$$\boldsymbol{T}=\boldsymbol{C}'\sigma(\boldsymbol{t})=\boldsymbol{C}'\sigma\left(w_{i,j}^{l}{}'\times\boldsymbol{c}_{i,j}^{l}{}'+b_{j}^{l\prime}\right), \tag{8}$$

where $\boldsymbol{T}$ is the output matrix; $\boldsymbol{C}'$ is the input matrix; $w_{i,j}^{l}{}'$ and $b_{j}^{l\prime}$ are the weight and deviation generated

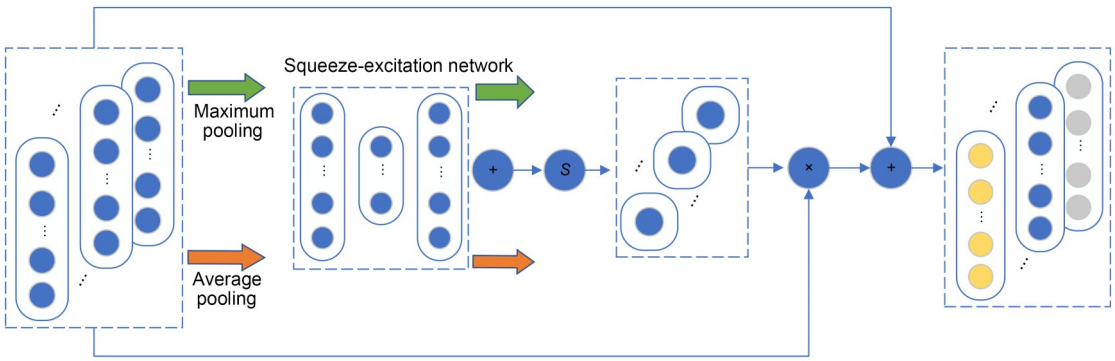

Fig. 4 Channel attention mechanism

by the convolution operation, respectively; $c_{i,j}^{l}{}'$ is the eigenvalue.

The residual idea was similarly introduced in temporal attention mechanism. The residual connection method is used, that is $\boldsymbol{T}'=\boldsymbol{C}'+\boldsymbol{T}$. Fig. 5 shows the temporal attention mechanism.

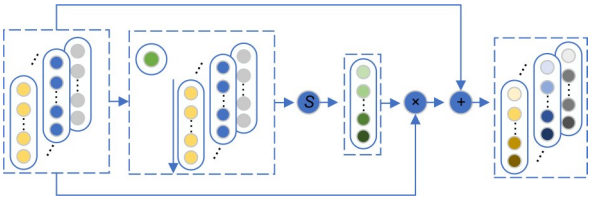

Fig. 5 Temporal attention mechanism

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.

3 Experimentation and analysis

3.1 Experimental setup

The experimental equipment was a 4135AC marine power-generation diesel engine (Shanghai Diesel Engine Co., Ltd., China). 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 s and 20.48 kHz, respectively. The sampling interval was 30 s. 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. The above sensors and acquisition equipment are developed by China Orient Institute of Noise & Vibration. 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.

Table 1 Main parameters of the diesel engine

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

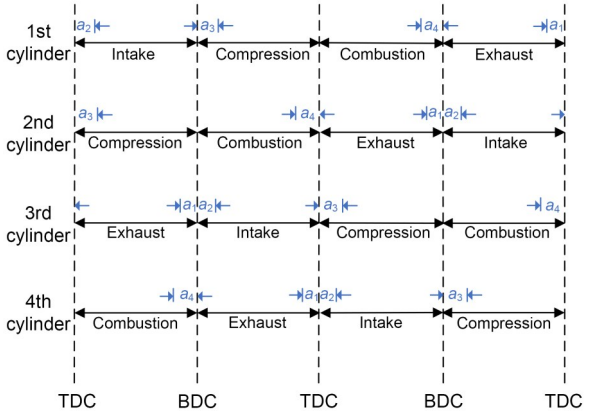

Fig. 6 Phase diagram of diesel engine. BDC: bottom dead center. a_1: intake-valve opening angle, 16° crankshaft angle (CA); a_2: exhaust-valve closing angle, 16° CA; a_3: intake-valve closing angle, 42° CA; a_4: exhaust-valve opening angle, 43° CA

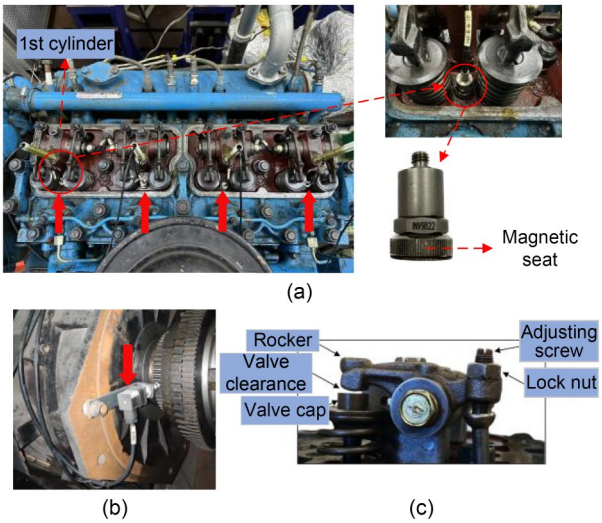

Fig. 7 Arrangement of sensors and valve clearance: (a) vibration sensors; (b) TDC sensor; (c) valve clearance

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 the rated speed of 1500 r/min and 75% load.

Table 2 Fault-simulation experimental scheme

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

3.2 Signal analysis

After collecting the vibration signals for the different health states, we used the normal state and abnormal exhaust-valve clearances as examples (R1, R6, and R7). 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. Figs. 8a and 8b show 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/s^2, 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.

To link computer vision and fault diagnosis, we transformed the original signals into 2D images. Fig. 9 shows the results of 2D image conversion for the seven

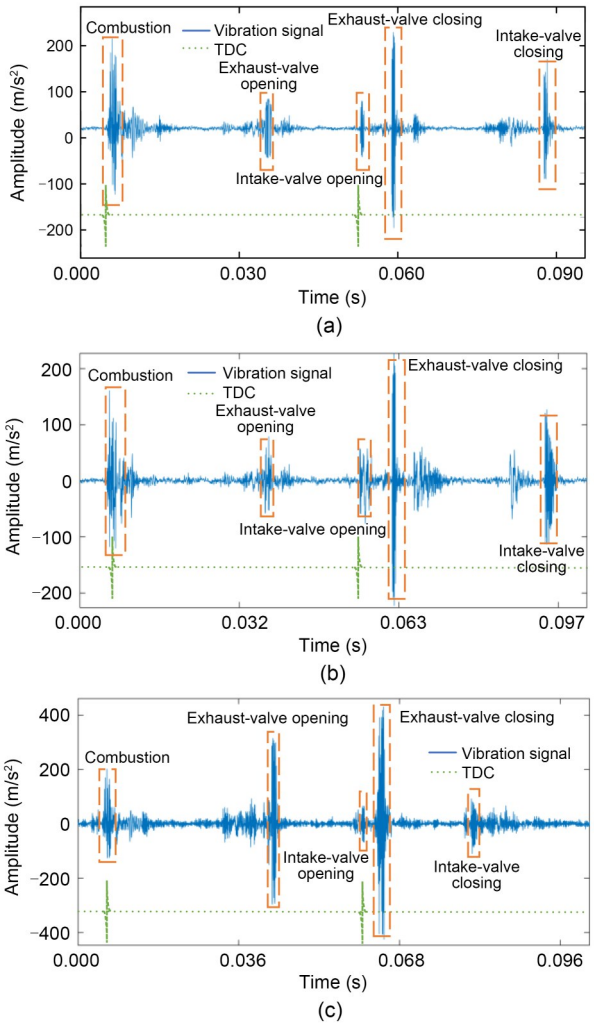

Fig. 8 Time-domain diagram of normal condition and abnormal exhaust-valve clearance: (a) R1; (b) R6; (c) R7

health states. When the valve clearance was abnormal, the texture and color features of the 2D 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. 9a, 9f, and 9g indicate normal, slightly abnormal, and severely abnormal exhaust-valve clearances, respectively. The vibration signals are converted into 2D images, and 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

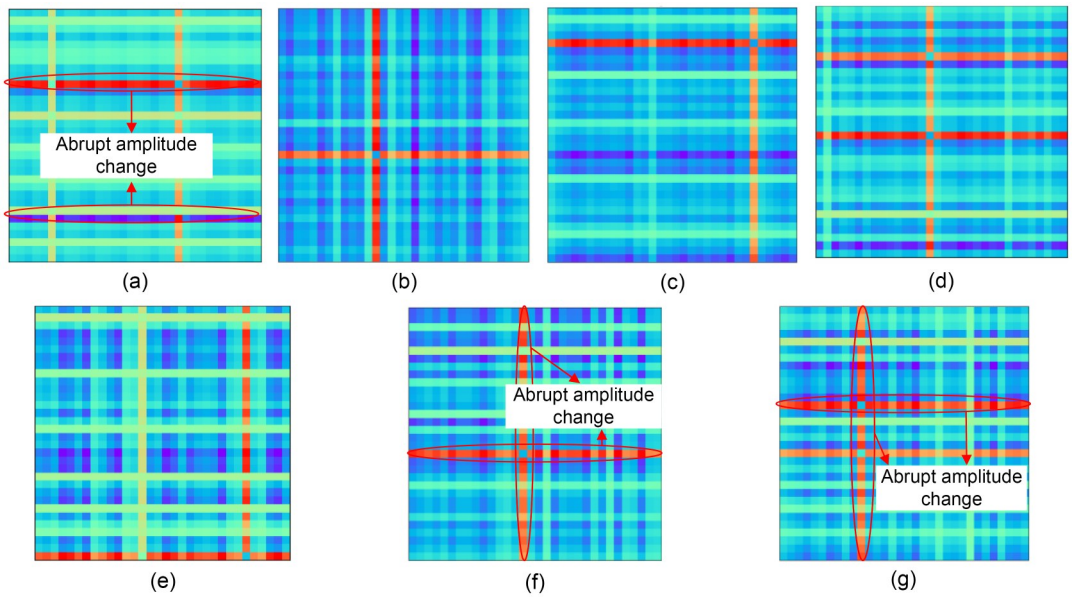

Fig. 9 Two-dimensional images of the seven health states: (a) R1; (b) R2; (c) R3; (d) R4; (e) R5; (f) R6; (g) R7

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.

3.3 CNN structure and parameter selection

The deep learning setup included an AMD Ryzen 7 5800H CPU@3.2 GHz, NVIDIA GeForce RTX 3060 Laptop GPU, and the Windows 11 operating system. In a squeeze-incentive network with a channel attention mechanism, the value of m needs to be discussed. The selectable values of m were 4, 8, 16, 32, and 64. The results of different m values are shown in Table 3. When m changed from 4 to 64, the training time dropped by 125 s and diagnostic accuracy decreased by 2.78%. Therefore, the value of m had little effect on model performance. We set the value of m to 4.

Table 3 Comparison of results for different dimensionality reduction factors

m	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 (the maximum number of iteration rounds), and other parameters in advance. We will use the learning rate and MaxEpoch as examples to show the hyperparameter setting process. The optional ranges of the learning rate were 0.0001, 0.0010, and 0.0100. The performance of the CNN is shown in Fig. 10. When the range was 0.0010, the CNN had the highest accuracy and the shortest running time, so we chose this value. The optional values of MaxEpoch were 20, 30, 40, 50, and 80. The results of the experiment are shown in Fig. 11. When MaxEpoch=20, the network model required the least 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.

3.4 CNN performance

For each health state, 500 data sets were collected. We randomly selected 288 sets for training and 12 sets 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 accuracies of the training and validation data increased to 100.00%. The loss function curve was

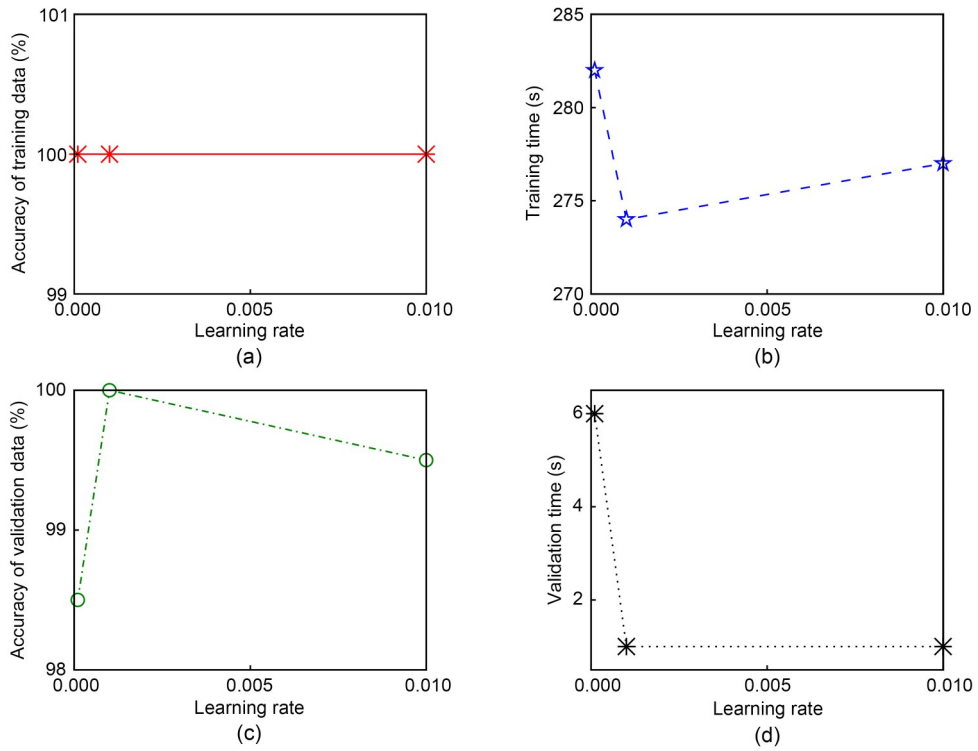

Fig. 10 Experimental results of different learning rates: (a) training accuracy; (b) training time; (c) validation accuracy; (d) validation time

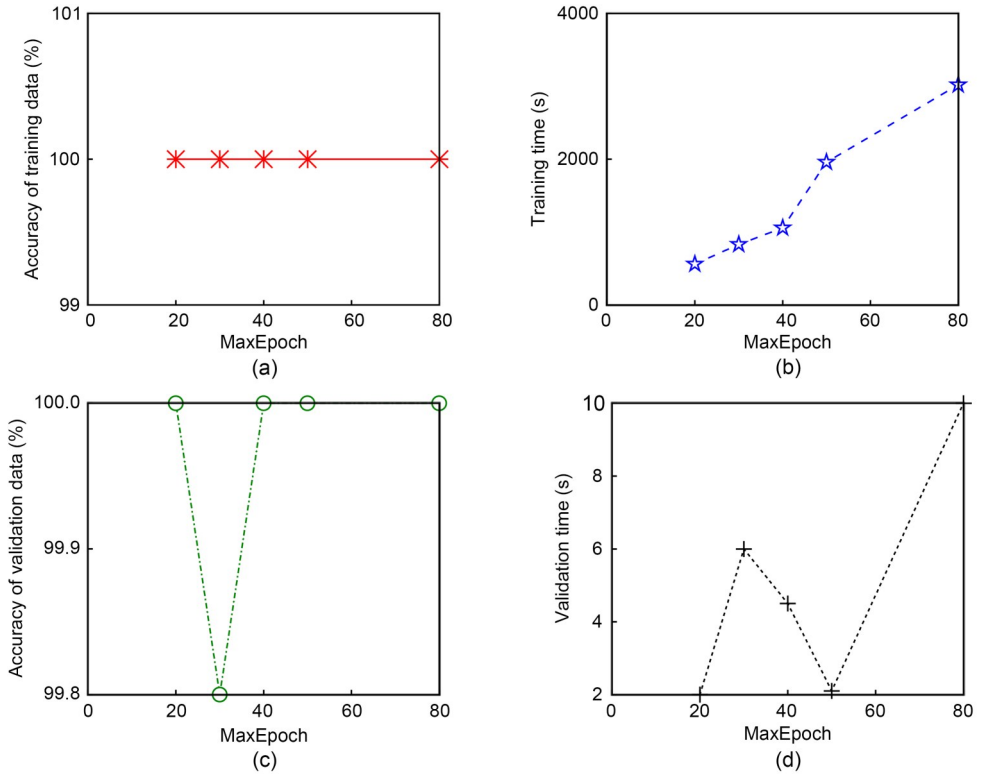

Fig. 11 Experimental results with different MaxEpoch values: (a) training accuracy; (b) training time; (c) validation accuracy; (d) validation time

Table 4 Optimal values for the main hyperparameter settings

Hyperparameter	Description
L2Regularization	0.004
LearnRateDropFactor	0.05
LearnRateDropPeriod	10
InitialLearnRate	0.001
Slover	Adam
MaxEpoch	20
MiniBatchSize	10

Table 5 CNN structure and main parameters

Name	Structure parameter	Output size
Image input	64×64×3	64×64×3
Conv_1	32@2×2, stride=[1], padding=[2]	67×67×32
Batchnorm_1	–	67×67×32
Maxpool_1	3×3, stride=[2], padding=[0]	33×33×32
MAM_1	64@1×1×1 (time attention mechanism)	33×33×32
ReLu_1	ReLu	33×33×32
Conv_2	32@3×3, stride=[1], padding=[2]	35×35×32
Batchnorm_2	–	35×35×32
Maxpool_2	3×3, stride=[2], padding=[0]	17×17×32
MAM_2	128@1×1×1 (time attention mechanism)	17×17×32
ReLu_2	ReLu	17×17×32
Fc_1	–	1×1×100
ReLu_3	ReLu	1×1×100
Dropout	0.5	1×1×100
Fc_2	–	1×1×7
Softmax	–	1×1×7
Class output	–	1×1×7

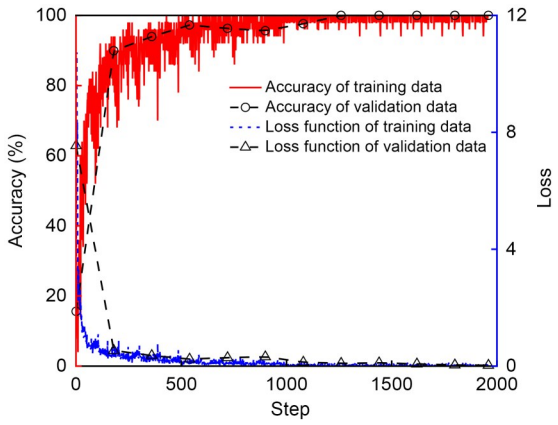

Fig. 12 CNN model training and validation process

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 s and the validation time was about 3 s. No over-fitting occurred during the model run.

Fig. 13 shows the confusion matrix of the testing data. The data in main diagonal cells represent the number of the correctly classified samples and their percentages of the number of total samples, and the data in other cells represent the number of incorrectly classified samples and their percentages of the number of total samples. The data in green represent the accuracy rates for every row and column, the data in red represent the error rates for every row and column, and the data in the lower right corner represent the overall accuracy and error rate. A total of 22 samples were misclassified. The classification accuracies for the seven states were 100.00%, 98.50%, 97.50%, 100.00%, 95.50%, 98.00%, and 100.00%, respectively. The average accuracy was 98.40%.

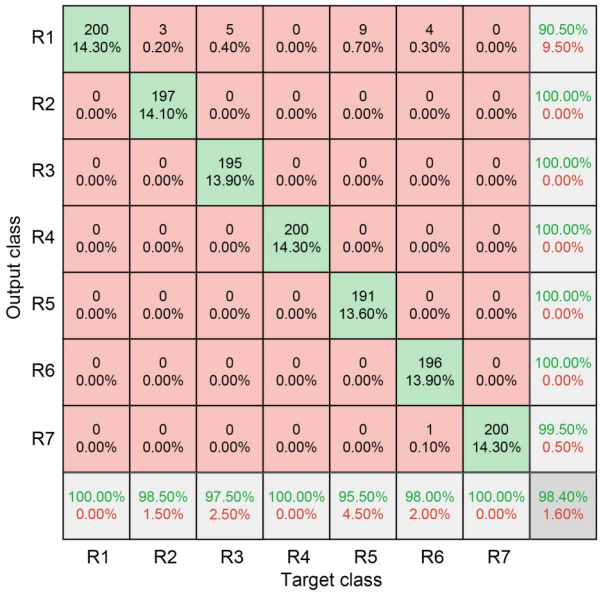

Fig. 13 Confusion matrix for test data. References to color refer to the online version of this figure

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 were from 60 groups. The experiment was repeated five times for each SNR.

The average diagnostic accuracies and average loss-function values are shown in Table 6. In all four cases, the average accuracy rate was above 94.00%.

Table 6 CNN model performance with different SNRs

SNR (dB)	Average training accuracy (%)	Average testing accuracy (%)	Average loss
−20	98.80	94.60	0.0581
0	99.10	96.60	0.0474
10	99.30	98.10	0.0367
20	99.60	98.30	0.0304

The SNR equation is as follows:

$$R_{\mathrm{SN}}=10\lg\frac{P_{\mathrm{Signal}}}{P_{\mathrm{Noise}}}, \tag{9}$$

where R_{SN} is the SNR, 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 MAM, we fed the data (SNR=−20 dB) into two network models, one with a no-attention mechanism (NAM) and one with an MAM. All other parameters were the same for both models. The experiment was repeated five times. The accuracy of the first four health states is shown in Fig. 14. The fault-diagnosis accuracy of the model with the attention mechanism was 14.80% higher than that of the model without it. Thus, the attentional mechanism appears to improve the noise resistance of the model.

3.6 Evaluation of different input methods

To verify the effectiveness of the image fusion 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 is 94.00% and 84.90%, 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.

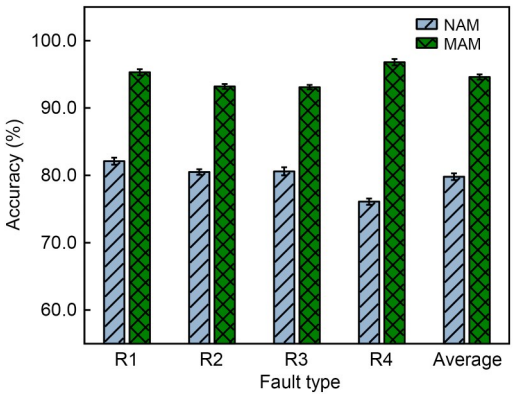

Fig. 14 Influence of two mechanisms on CNN performance

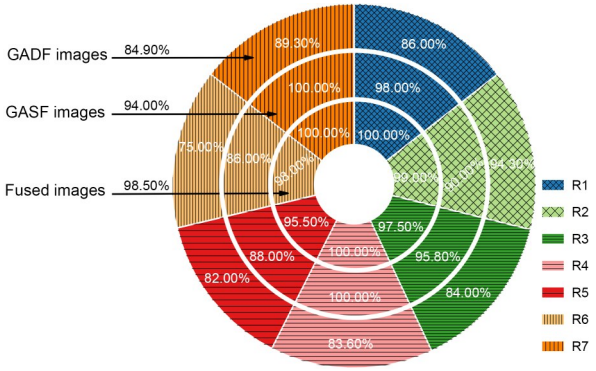

Fig. 15 Accuracies of different input methods

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% loads. The data from any three loads were used as training data, and the data from the other one load were 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 89.00% 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 MAM 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.

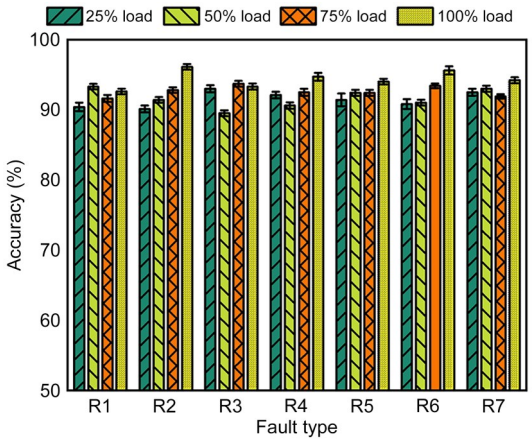

Fig. 16 Model diagnosis results under variable loads

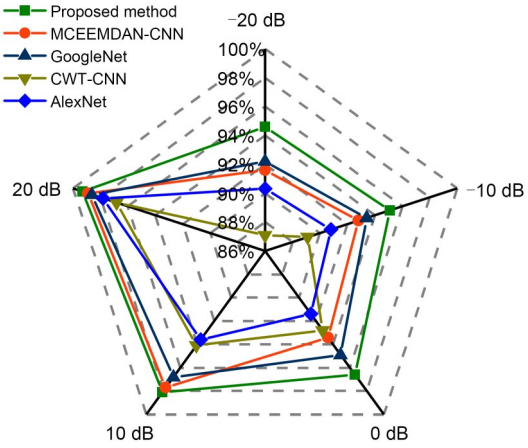

Fig. 17 Fault-diagnosis accuracy of different methods

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 and Turk, 2022), AlexNet (Hajnayeb, 2021), modified complete ensemble empirical mode decomposition with adaptive noise (MCEEMDAN)-CNN (Hou et al., 2022), and 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.

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 92.20%, 91.60%, 90.30%, and 87.10%, respectively. The accuracy of the proposed method is 94.60%, which is higher than those 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 loads 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 MAM 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, diffusion model, transfer learning, generative adversarial network, and other methods could be used 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 study, we designed and tested a valve-clearance fault-diagnosis model based on the GAF and a CNN. The conclusions are as follows:

1. Using an MAM, 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 an MAM is 14.80% better than the model without the mechanism.

2. The average accuracy is above 94.00% with all SNRs. The proposed method has better noise immunity and stability than the other four 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 89.00%, proving that it has a good stability.

Acknowledgments

This work is supported by the Project of Shanghai Engineering Research Center for Intelligent Operation and Maintenance and Energy Efficiency Monitoring of Ships (No. 20DZ2252300), China. We sincerely thank Dexin CUI (the superintendent of Ningbo Ocean Shipping Co., Ltd., China) for providing experimental guidance.

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):045901. 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*, 57:148-157. https://doi.org/10.1016/j.jmsy.2020.09.001

Cerrada M, Zurita G, Cabrera D, et al., 2016. Fault diagnosis in spur gears based on genetic algorithm and random forest. *Mechanical Systems and Signal Processing*, 70-71: 87-103. https://doi.org/10.1016/j.ymssp.2015.08.030

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 YP, et al., 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*, 2020:5357146. 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*, 31:101049. https://doi.org/10.1016/J.JESTCH.2021.08.006

Hajnayeb A, 2021. Cavitation analysis in centrifugal pumps based on vibration bispectrum and transfer learning. *Shock and Vibration*, 2021:6988949. https://doi.org/10.1155/2021/6988949

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*, 192:110889. https://doi.org/10.1016/J.MEASUREMENT.2022.110889

Hoang DT, Kang HJ, 2019. A survey on deep learning based bearing fault diagnosis. *Neurocomputing*, 335:327-335. https://doi.org/10.1016/j.neucom.2018.06.078

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

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*, 210:112560. https://doi.org/10.1016/J.MEASUREMENT.2023.112560

Karatuğ Ç, Arslanoğlu Y, 2022. Development of condition-based maintenance strategy for fault diagnosis for ship engine systems. *Ocean Engineering*, 256:111515. https://doi.org/10.1016/J.OCEANENG.2022.111515

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: 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

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

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

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):2509-2531. 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*, 165:108385. 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., 2021. 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

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

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

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

Sun F, Xu H, Zhao YH, 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 (Applied Physics & Engineering)*, 23(4):303-313. https://doi.org/10.1631/jzus.A2100598

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

Wang B, Lei YG, Li NP, et al., 2021. Multiscale convolutional attention network for predicting remaining useful life of machinery. *IEEE Transactions on Industrial Electronics*, 68(8):7496-7504. https://doi.org/10.1109/tie.2020.3003649

Wen L, Li XY, Gao L, 2020. A transfer convolutional neural network for fault diagnosis based on ResNet-50. *Neural Computing and Applications*, 32(10):6111-6124. https://doi.org/10.1007/s00521-019-04097-w

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):095121. https://doi.org/10.1088/1361-6501/ACDC45

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 & System Safety*, 226:108714. https://doi.org/10.1016/J.RESS.2022.108714

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