*Article*

# A New Method of Intelligent Fault Diagnosis of Ship Dual-Fuel Engine Based on Instantaneous Rotational Speed

**Ji Gan [1], Huabiao Jin [1,2], Qianming Shang [1,3,*] and Chenxing Sheng [1,4]**

[1] School of Naval Architecture, Ocean and Energy Power Engineering, Wuhan University of Technology, Wuhan 430063, China; 331519@whut.edu.cn (J.G.); jhb@whut.edu.cn (H.J.); scx01@126.com (C.S.)
[2] Key Laboratory of Ship Power Engineering Technology and Transportation Industry, Wuhan University of Technology, Wuhan 430063, China
[3] Key Laboratory of High Performance Ship Technology, Ministry of Education, Wuhan University of Technology, Wuhan 430063, China
[4] National Key Laboratory of Waterway Traffic Control, Wuhan University of Technology, Wuhan 430063, China
[*] Correspondence: sqmte@whut.edu.cn; Tel.: +86-13986120968

**Abstract:** Ship engine misfire faults not only pose a serious threat to the safe operation of ships but may also cause major safety accidents or even lead to ship paralysis, which brings huge economic losses. Most traditional fault diagnosis methods rely on manual experience, with limited feature extraction capability, low diagnostic accuracy, and poor adaptability, which make it difficult to meet the demand for high-precision diagnosis. To this end, a fusion intelligent diagnostic model—ResNet–BiLSTM—is proposed based on a residual neural network (ResNet) and a bidirectional long short-term memory network (BiLSTM). Firstly, a multi-scale decomposition of the instantaneous rotational speed signal of a ship's engine is carried out by using the continuous wavelet transform (CWT), and features containing misfire fault information are extracted. Subsequently, the extracted features are fed into the ResNet–BiLSTM model for learning. Finally, the intelligent diagnosis of ship dual-fuel engine misfire faults is realized by the classifier. The model combines the advantages of ResNet18 in image feature extraction and the capability of BiLSTM in temporal information processing, which can efficiently capture the time-frequency features and dynamic changes in the fault signal. Through comparison experiments with fusion models AlexNet–BiLSTM, VGG–BiLSTM, and the existing AlexNet–LSTM and VGG–LSTM models, the results show that the ResNet–BiLSTM model outperforms the other models in terms of diagnostic accuracy, robustness, and generalization ability. This model provides an effective new method for intelligent diagnosis of ship dual-fuel engine misfire faults to solve the traditional diagnostic methods' limitations.

**Keywords:** ship dual-fuel engine; internal combustion engines (ICEs); instantaneous rotational speed; misfire fault diagnosis; anomaly detection; machine learning; deep learning; ResNet; BiLSTM

## 1. Introduction

As the foundation of global logistics, the shipping industry is responsible for the majority of international trade transportation. Ensuring the safe and efficient operation of ships is of paramount importance [1–3]. Ship engines, as the primary source of power for maritime vessels, are responsible for propelling ships forward. The performance and operational status of these engines directly impact the safety, economy, and reliability of ships [4–6]. In particular, in modernized, large-scale ocean-going vessels, the engine is subjected to high loads and a complex operational environment over extended periods. Any fault may result in the disruption of the ship's operations or even lead to significant safety incidents. Among the numerous types of engine faults, misfire faults are particularly prevalent and hazardous [7,8].

A misfire fault is typically observed when a cylinder within an engine fails to ignite the fuel mixture at the designated point in the combustion cycle. Cylinder misfire results in

inadequate power output and diminished fuel efficiency and may even precipitate a series of chain reactions, including increased engine vibration and elevated emissions [9]. In the event that such faults are not identified and rectified in a timely manner, they may result in additional wear or damage to engine components and potentially even lead to more severe safety incidents [10]. For instance, a power system fault resulting from an engine misfire during an ocean voyage may result in the loss of propulsion and control, thereby placing the vessel in a perilous situation. Furthermore, a misfire fault increases the operating cost of a ship, as it not only elevates fuel consumption but may also necessitate a broader range of equipment maintenance and repair.

In light of the aforementioned considerations, the diagnosis of malfunctions in marine propulsion systems, with a particular emphasis on the early identification and characterization of misfire issues, assumes paramount importance. The early detection and treatment of misfire faults can prevent the development of minor issues into significant accidents, assist ship operators in reducing operational risks, minimize unnecessary economic losses, and ensure the safety of the ship and its personnel [11,12]. Fault diagnosis technology is a method of extracting fault characteristics and determining the type and location of fault occurrence through the analysis of equipment operation data. In industrial equipment and mechanical systems, fault diagnosis technology has been widely utilized in wind turbines, aviation engines, railroad locomotives, and other fields, with the fault diagnosis of ship engines representing a significant area of interest within this field. With the advancement of marine engine technology, particularly in the context of the increased complexity of electronic control and fuel injection systems, traditional diagnostic methods have proven inadequate in addressing the increasingly complex fault modes and signal characteristics [13].

In practice, traditional methods for diagnosing marine engine misfire faults typically rely on the experience and intuition of the operator or the detection of simple engine control system alarms. These methods are clearly inadequate. First, manual diagnosis depends on the expertise of the operator, which may prove challenging in complex navigational conditions, potentially leading to delays in identifying the root cause of the fault. Secondly, the alarm system of the engine control system is typically only capable of detecting significant faults that have already occurred. It lacks the necessary sensitivity and early warning capability to detect early faults, which makes it challenging to identify potential issues in a timely manner. Furthermore, traditional methods often prove inadequate for diagnosing elusive or subtle misfire faults. The growing complexity and nonlinear characteristics of ship engine signals have rendered traditional signal processing methods, which rely on rules and statistical analysis, insufficient for modern ships that require efficient and accurate fault diagnosis [14–16]. For example, Han [17] proposed the AGap slope as a novel approach to misfire detection. By comparing the inter-cylinder slope difference between the teeth of the same cylinder in two adjacent cycles, the AGap slope can effectively eliminate the inter-cylinder slope error. Experimental results demonstrate that the method exhibits an average misfire detection rate of 90.2% across a range of test conditions. Furthermore, the detection rate can reach 93% to 98% within the 1500 to 4000 rpm range. However, the detection rate is reduced when the engine load is close to neutral or the speed exceeds 4000 rpm. Wang et al. [18] proposed a diagnostic strategy with an adaptive threshold algorithm. This algorithm is based on an angular domain identification method that determines the misfire-sensitive region in real time through relative scatter analysis. The time unit is then computed based on this analysis. The time unit change value of each operational cylinder is employed to compute a weighted average, thereby constructing a misfire feature signal as an analytical object. The results of real-vehicle validation demonstrate that the novel strategy is capable of adjusting the diagnostic threshold in real-time, enhancing the real-time diagnosis of a misfire (89% improvement), and increasing the feature signal amplitude by over 25% following cylinder filtering in the continuous misfire mode. Moreover, the method is capable of detecting various misfire types across the full spectrum of operating conditions, obviating the necessity to establish discrete thresholds for different vehicle

driving states and operational scenarios. This reduces the calibration workload and the impact of vehicle dispersion. Sharib et al. [19] proposed the use of RedLeo Pro V8 software to simulate input data for the purpose of monitoring and controlling the engine system. This method has been demonstrated to be effective in distinguishing between normal and abnormal signals, with the signals being designed through an adaptive system with the objective of reducing noise and improving diagnostic accuracy. The final results demonstrate the efficacy of the proposed method in feature extraction and selection, rendering it an effective approach for engine troubleshooting. Syta et al. [20] analyzed the vibration signals of the Rotax 912 ULS aircraft engine to detect misfires in individual cylinders. A linear metric was developed to describe the vibration level based on power amplitude spectral values at two selected frequencies. In addition, a nonlinear metric was calculated from the periodicity of engine operation. The results demonstrate that both methods are effective in detecting misfires in diverse cylinder configurations and that their combination enables the identification of faulty cylinders. Jafari et al. [21] employed an acoustic emission sensor to detect misfires in a multi-cylinder diesel engine. The angular periodic modulation (cyclic bursts) in the signal power was highlighted by squared envelope spectral processing of the acoustic emission signal. This study demonstrates the effectiveness of combining sensor technology with signal processing for misfire detection in a six-cylinder diesel engine. Kang et al. [22] proposed an efficient method for detecting and monitoring engine misfires, focusing on small speed changes on the crankshaft, simulating five engine states (one normal ignition and four misfires) in the experiment. The results show that the composition of **6f** is the largest under normal conditions, but with the occurrence of fire, the composition of **f** increases gradually. 3D FFT modes with ratios of **f**, **2f** and **3f**, **6f** show a greater distance between the misfire state and the normal state. However, it should be noted that all these methods have certain limitations.

In order to overcome the limitations of traditional signal processing methods, machine learning techniques have been introduced into the field of fault diagnosis of ship engines in a gradual and progressive manner [23]. In contrast to conventional methodologies, machine learning enables the automatic discovery of features through a data-driven approach, eliminating the necessity for manually designed feature extraction techniques. This significantly enhances the automation and precision of fault diagnosis [24,25]. For instance, Syta et al. [26] put forth a methodology for the detection and identification of misfires in aviation internal combustion engines through the analysis of vibration time series. This approach employs a machine learning classification model to discern the operational states of the engine. The findings indicated that the utilization of nonlinear metrics facilitated a high degree of accuracy. The classification accuracy was demonstrated with a reduced number of samples. Singh [27] put forth a novel approach to identifying misfires through the assessment of radiated sound quality metrics in the vicinity of the cylinder block or exhaust pipe. This method has been subjected to rigorous testing on a four-stroke, four-cylinder engine across a range of load and speed conditions. Sound quality metrics, including noise, roughness, and fluctuation intensity, were predicted by a support vector machine classifier with an accuracy of 94%. In comparison, conventional methods for vibration signals and sound pressure levels exhibited a prediction accuracy of 82% and 85%, respectively. This suggests that misfire detection based on sound quality is more accurate and independent of engine speed and torque. In contrast to conventional methods, the new method does not necessitate direct contact with engine components, is computationally rapid, has a broad range of applicability, and can be readily implemented under the hood or in close proximity to the exhaust pipe via acoustic sensors. Mulay et al. [28] employed piezoelectric accelerometers to obtain cylinder vibration signals for the purpose of detecting misfires and analyzing the specific vibration modes that occur at the time of misfire. Twelve statistical features were extracted, and useful features were filtered by the J48 decision tree algorithm and classified using regression classification and IBk classification. Subsequently, the performance of the classifiers was compared, and an effective misfire detection algorithm was proposed by integrating the classifiers through voting. While the aforementioned

machine learning techniques exhibit commendable classification capabilities in certain marine engine fault diagnosis scenarios, they are susceptible to the challenge of excessive computational complexity when confronted with voluminous data or intricate signals.

As data size and model complexity have increased, artificial neural networks (ANNs) have emerged as a prominent area of research in the field of fault diagnosis. ANNs emulate the intricate workings of neurons in the human brain, forming intricate multilayer networks that can automatically extract high-dimensional features from data and perform complex nonlinear mapping. Compared to traditional machine learning, ANNs offer enhanced flexibility in fault diagnosis and the ability to handle more complex signal features. For example, Liu et al. [29] proposed a novel misfire detection model for turbocharged diesel engines using artificial neural networks (ANNs). The model was implemented in the MATLAB/Neural Network Toolbox environment and experimentally investigated on a V6 turbocharged diesel engine. The preliminary results demonstrated that the model successfully detected misfires in the majority of cases, although some misdetections were observed, and the mean-square error was relatively high. However, by incorporating the engine speed variations within the cycle into the training data, the model ultimately achieved fully accurate detection, thereby providing a new method for accurately detecting misfires in turbocharged diesel engines. Jafarian et al. [30] investigated misfire faults in internal combustion engines, with a particular focus on the analysis of signal variations captured using different sensors. The engine faults were subjected to experimental analysis, and it was proposed that a Fast Fourier Transform (FFT) be employed for signal transformation and feature extraction, with the utilization of artificial neural networks (ANNs) in the fault classification stage. By measuring the performance metrics of the ANNs and comparing them with the results of similar studies in the related literature, the results demonstrate the efficacy of incorporating vibration signals into the analysis of internal combustion engine faults. However, traditional shallow neural networks are susceptible to converging on local optimal solutions due to their shallow network structure, exhibiting suboptimal training efficiency and poor generalization ability when confronted with voluminous and complex datasets.

To address these challenges, deep learning methodologies have been extensively employed, particularly with the advent of convolutional neural networks (CNNs) and recurrent neural networks (RNNs), along with their enhanced long short-term memory (LSTM) networks and bidirectional long short-term memory (BiLSTM) networks. These developments have considerably elevated the sophistication of diagnostic techniques for ship engine faults [31,32]. For instance, Zhang et al. [33] investigated a misfire detection methodology based on convolutional neural networks (CNNs), utilizing experimental data from a six-cylinder inline diesel engine for network training and testing to identify misfire patterns in one and two cylinders. The results demonstrate that the convolutional neural network (CNN) is capable of accurately detecting complete misfires in one or two cylinders under steady-state conditions, with a detection accuracy exceeding 96% in the case of partial misfires in one cylinder when fuel injection is reduced to half of the normal amount. Furthermore, under non-steady-state conditions, such as acceleration or deceleration, the CNN demonstrates satisfactory performance within a limited acceleration range. However, the network's efficacy declines when the absolute acceleration of the engine speed surpasses 100 r/min/s. Venkatesh et al. [34] put forth a methodology for identifying misfires in internal combustion engines through the application of transfer learning techniques. Initially, vibration signals are gathered from the upper portion of the engine and then presented as input to the deep learning algorithm. In order to identify misfire states, pre-trained networks (e.g., AlexNet, VGG-16, GoogLeNet, and ResNet-50) are employed. Furthermore, the effects of hyper-parameters (e.g., batch size, solver, learning rate, and training-to-test ratio) are investigated. Xu et al. [35] proposed a domain-adversarial wide-kernel convolutional neural network (DAWDCNN) for diesel engine misfire fault diagnosis. This was done with the aim of addressing the impact of diesel engine noise variations and stochasticity on the performance of existing diagnostic methods.

The DAWDCNN demonstrates superior generalization performance in 11 noisy domain adaptation tasks relative to the conventional staged domain adversarial training approach. The experimental outcomes indicate that the mean accuracy of the DAWDCNN on the four datasets surpasses that of random forests, long- and short-term memory networks, and other comparable techniques. Wang et al. [36] proposed a novel approach based on long short-term memory recurrent neural networks (LSTM RNNs) for the detection of diesel engine misfires. The findings indicate that the LSTM RNN-based algorithm is capable of overcoming the inherent limitations of traditional methods. The network structure, which inputs a fixed segment of raw rotational speed signals and utilizes misfire or no-fault labels as outputs, has demonstrated a notable accuracy in diagnosing misfires.

However, convolutional neural networks (CNNs) are designed for the extraction of static image features, particularly those related to spatial dimensions. They are less effective when processing time-series information due to their limited processing time [37,38]. Recurrent neural networks (RNNs) and their enhanced variants, such as long short-term memory (LSTM) and bidirectional LSTM (BiLSTM), are optimized for time-series data analysis. They excel in capturing temporal dependencies but are less adept at feature extraction from images [39,40].

In order to address the deficiencies of convolutional neural networks (CNNs) in capturing time series dependencies and the limitations of recurrent neural networks (RNNs) in extracting spatial features of images, a novel intelligent diagnostic model for marine dual-fuel engine misfire with ResNet18 in combination with BiLSTM is proposed, aiming to improve accuracy and real-time diagnosis of faults. In contrast to traditional fault diagnosis techniques, this approach employs the continuous wavelet transform (CWT) to transform the one-dimensional instantaneous rotational speed signal into a two-dimensional time-frequency image, thereby preserving the time-frequency characteristics of the signal. The two-dimensional image data are fed into a network to extract high-dimensional feature representations through a deep convolutional layer. These are then passed to a bidirectional long short-term memory (BiLSTM) network for temporal processing, which enables the capture of the dynamically changing characteristics of the signal. This method not only extracts the deep features of fault signals from the images but also processes the time-dependent information in the signals through the BiLSTM network, thereby achieving more accurate fault identification.

The principal findings of the study can be classified into three main aspects.

(1) An intelligent diagnostic framework combining continuous wavelet transform (CWT) and deep learning models is proposed. This framework utilizes the continuous wavelet transform (CWT) to convert the instantaneous rotational speed signal of a ship's engine from a one-dimensional time series to a two-dimensional time-frequency image. Additionally, it captures the time-frequency features of the misfire fault signal through multiscale decomposition. This framework effectively addresses the limitations of traditional signal processing methods in capturing non-smooth signals, providing a more comprehensive input for subsequent deep learning models.

(2) An intelligent diagnostic model (ResNet–BiLSTM) is constructed by fusing ResNet18 and BiLSTM. The ResNet18 model serves as a feature extractor, enabling comprehensive mining of local spatial features in the time-frequency image. The BiLSTM network, on the other hand, is capable of capturing temporal dependencies in the signal. The fusion model enables the dual learning of time-frequency features and timing information, thereby markedly enhancing the detection capability for misfire faults.

(3) A series of comparative experiments were conducted to evaluate the performance of the proposed ResNet–BiLSTM model in comparison with fusion models (AlexNet–BiLSTM, VGG11–BiLSTM) and existing methods (AlexNet–LSTM, VGG–LSTM). The results demonstrated that the ResNet–BiLSTM model exhibited superior comprehensive performance, outperforming the other models.

The remaining sections are organized as follows: Section 2 introduces the fundamental principles of the relevant theories. Section 3 describes the implementation process of

the intelligent fault diagnosis method based on the improved ResNet–BiLSTM fusion model. Section 4 provides a comprehensive account of the data collection process and the construction of the dataset. Section 5 presents a comparative analysis of the different models and their respective outcomes. Section 6 offers a summary of the conclusions and suggests directions for future research.

## 2. Basic Theory

### 2.1. Continuous Wavelet Transform

The continuous wavelet transform (CWT) is a powerful tool for analyzing signals at multiple scales. It provides a joint representation of a signal in time and frequency by convolving the signal with a set of wavelet functions with different scales and positions [41]. This method is particularly effective in analyzing non-stationary signals and transient phenomena and is applicable to a variety of engineering and scientific fields, including the fault diagnosis of instantaneous rotational speed signals from ship dual-fuel engines.

The core of the CWT lies in the selection of wavelet functions. In contrast to the Fourier transform, the wavelet transform employs basis functions that are localized, finitely supported waveforms—namely, wavelets—that can be adjusted in both time and frequency. The CWT can be expressed as follows [42]:

$$W(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} f(t) \cdot \psi^*(\frac{t-b}{a}) dt \tag{1}$$

In the equation, $f(t)$ represents the original input signal, $\psi(t)$ is the wavelet basis function (mother wavelet), $\psi^*(t)$ denotes the complex conjugate of the wavelet function, $a$ is the scale factor (controlling the width of the wavelet), $b$ refers to the time translation factor (controlling the position of the mother wavelet), and $\frac{1}{\sqrt{|a|}}$ is a normalization factor ensuring that the transformation maintains the same energy across different scales.

In this study, the Morse wavelet is employed as the mother wavelet function, and the Fourier transform of the generalized Morse wavelet is:

$$\Psi_{\beta,\gamma}(\omega) = U(\omega) \cdot a_{\beta,\gamma} \cdot \omega^\beta \cdot e^{-\omega^\gamma} \tag{2}$$

In this context, $U(\omega)$ represents the unit step, $a_{\beta,\gamma}$ is a normalizing constant, $\omega$ is the frequency parameter that controls the frequency of the wavelet function, $\beta$ is viewed as a decay or compactness parameter, and $\gamma$ characterizes the symmetry of the Morse wavelet, respectively.

### 2.2. Structure of the ResNet Network Model

The ResNet (residual network) is a deep convolutional neural network (CNN) architecture, initially proposed by Kaiming He and colleagues [43]. The fundamental innovation of ResNet can be attributed to the introduction of the concept of residual learning, which markedly enhances the training efficacy and performance of deep networks.

In the context of intelligent fault diagnosis for engine misfires, ResNet demonstrates the capacity to process complex time-series data and high-dimensional feature data. The deep convolutional operations effectively extract key features from the data, thereby improving diagnostic accuracy. Moreover, engine fault samples are frequently scarce in comparison to normal samples, and ResNet's residual learning mechanism is particularly adept at addressing class imbalance issues, enabling the model to effectively learn features from the limited fault samples. The residual block represents the fundamental unit of ResNet. It comprises two principal components: one or more convolutional layers and a shortcut connection.

The fundamental configuration of a residual block is illustrated in Figure 1. As illustrated in the diagram, the sole distinction between the two types of residual blocks pertains to the manner of implementing the shortcut connection. In one instance, the

shortcut connection is implemented through a convolutional layer to adjust the number of channels (as illustrated by the dashed line on the right side of Figure 1b). In contrast, in the other instance, it is directly connected without adjusting the number of channels (as illustrated by the solid line on the right side of Figure 1a).
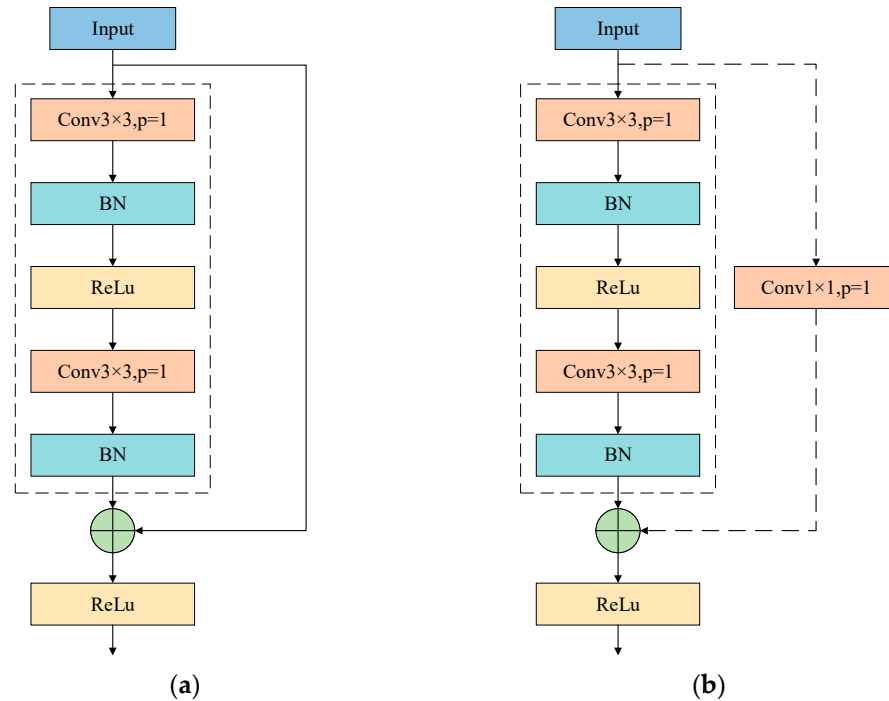


**(a)**                    **(b)**

**Figure 1.** Residual block structure. (**a**) Direct connection; (**b**) Convolutional connection.

### 2.3. Structure of the BiLSTM Network Model

LSTM (long short-term memory) network was developed to address the "vanishing gradient" and "exploding gradient" issues that are commonly encountered in traditional recurrent neural networks (RNNs). This is achieved through the introduction of specialized memory cells and three gate structures: the forget gate, the input gate, and the output gate, which regulate the flow of long short-term information. LSTM effectively selects which time-step information to retain or discard, thereby overcoming the limitations of traditional RNNs in capturing long-term dependencies [44]. The overall framework of the LSTM model is illustrated in Figure 2.
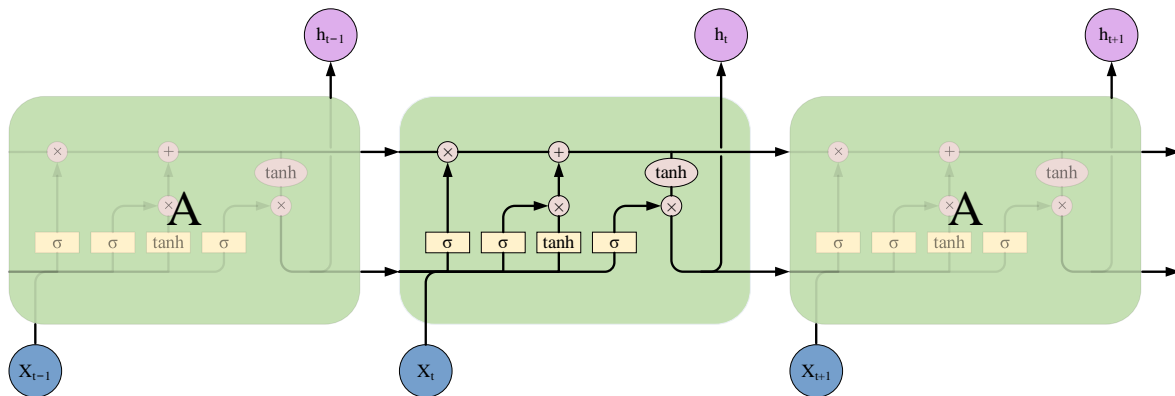


**Figure 2.** General framework diagram of the LSTM model.

The operation of the LSTM can be described as a process of filtering information within the cell state. The network discards superfluous, dated data and incorporates novel

information based on the present input and the hidden state from the preceding time step. This process enables the network to retain pertinent data for subsequent time steps. Initially, the forget gate determines which components of the cell state should be discarded based on the preceding hidden state and the present input. Subsequently, the input gate determines which novel information will be incorporated into the cell state. In conclusion, the output gate regulates which data from the present cell state will be utilized for the final output and updates the hidden state. The coordinated operation of these gates enables LSTM to efficiently retain, update, and output information at each time step, thereby ensuring that its hidden state reflects long-term dependencies.

The detailed computation process of LSTM is illustrated in Figure 3, where (a), (b), (c), and (d), respectively, show the computation processes for the forget gate $f_t$, input gate $i_t$, current cell state $C_t$, and output gate $o_t$.
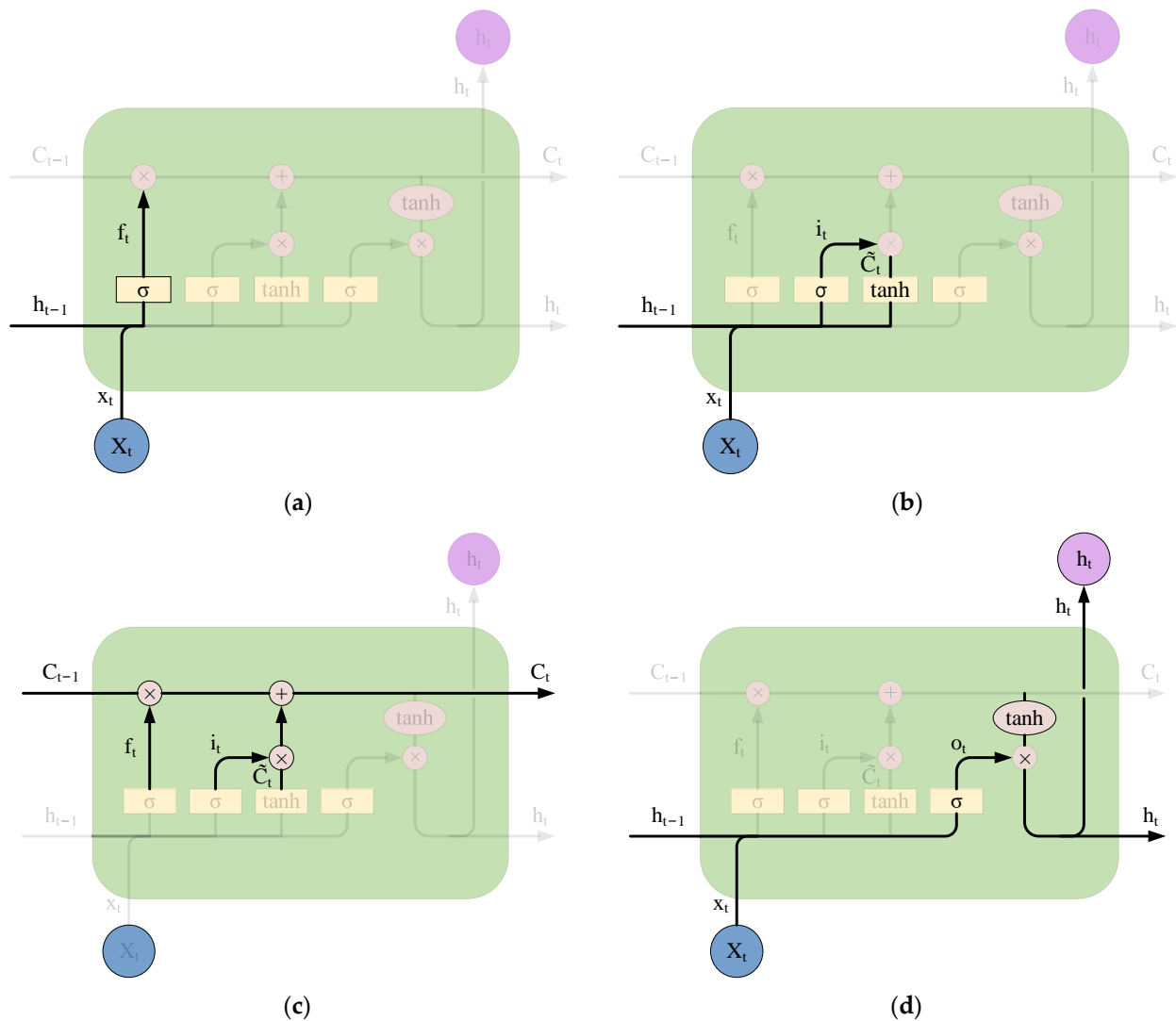


**Figure 3.** Detailed calculation process of LSTM. (**a**) The forget gate; (**b**) The input gate; (**c**) Current cell state at the moment; (**d**) The output gate.

The computation formula for the forget gate is given by

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3}$$

The input gate is calculated as

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$\widetilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{4}$$

The formula for the cell state at the current moment is

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C}_t \tag{5}$$

The output gate is given by

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t \odot tanh(C_t) \tag{6}$$

In Equations (3)–(6), $\sigma$ represents the sigmoid activation function, while $W_f$, $W_C$, and $W_o$ correspond to the weight matrices for the forget gate, input gate, and output gate, respectively. The bias terms $b_f$, $b_i$, and $b_o$ correspond to the forget gate, input gate, and output gate, respectively. The symbol $[h_{t-1}, x_t]$ denotes the concatenation of two vectors into a longer vector, while the symbol $\odot$ represents multiplication by element. Through these computational processes, the output and state updates of each layer of the LSTM can be obtained.

Bidirectional long short-term memory (BiLSTM) networks represent a variant of the LSTM. A limitation of the traditional LSTM is its inability to process time-series information in a bidirectional manner, from past to future or vice versa. In contrast, a bidirectional long short-term memory (BiLSTM) network incorporates a backward long short-term memory (LSTM) layer, enabling the simultaneous extraction of information from both the past and future directions of the time series. The dual-direction capability of BiLSTM renders it more effective at capturing long-term dependency information, thereby making it particularly suitable for tasks with strong temporal dependencies.

In the context of fault diagnosis for engine misfires in ships, the instantaneous rotational speed signal of the engine represents a typical time-series dataset that contains dynamically changing fault patterns. The bidirectional structure of BiLSTM enables the capture of changing trends in the signal both before and after, thereby enhancing sensitivity to fault features and improving diagnostic accuracy.

Figure 4 depicts the architecture of the bidirectional long short-term memory (BiLSTM) model. In contrast to the traditional unidirectional LSTM, the BiLSTM is constituted of two LSTM networks. One LSTM (the upper part) processes the input sequence in a forward direction (from time step $t_0$ to $t_2$), while the other LSTM (the lower part) processes the input sequence in a reverse direction (from time step $t_2$ to $t_0$).

As illustrated, the input vector at each time step (e.g., Vector 1, Vector 2, Vector 3) is fed simultaneously into both the forward and backward LSTMs. The forward LSTM generates the hidden states, designated as $h_{L0}$, $h_{L1}$, and $h_{L2}$, while the backward LSTM generates the hidden states, designated as $h_{R0}$, $h_{R1}$, and $h_{R2}$. At each time step, the outputs from the forward and backward LSTMs are concatenated (e.g., $h_0$, $h_1$, $h_2$) to form a complete output vector for that time step.

This structure enables BiLSTM to simultaneously utilize both past and future information in the sequence, thereby allowing it to capture complex temporal dependencies within time-series data. In the case of complex signals with time dependencies, such as the instantaneous rotational speed of a ship's dual-fuel engine, the BiLSTM model is better able to learn the dynamic variation features in the signal, thereby improving the accuracy of fault diagnosis. The combination of forward and backward LSTMs allows the model to focus on the changes in fault signals prior to the current moment while also incorporating information from future time steps, thereby facilitating more accurate fault detection and identification.
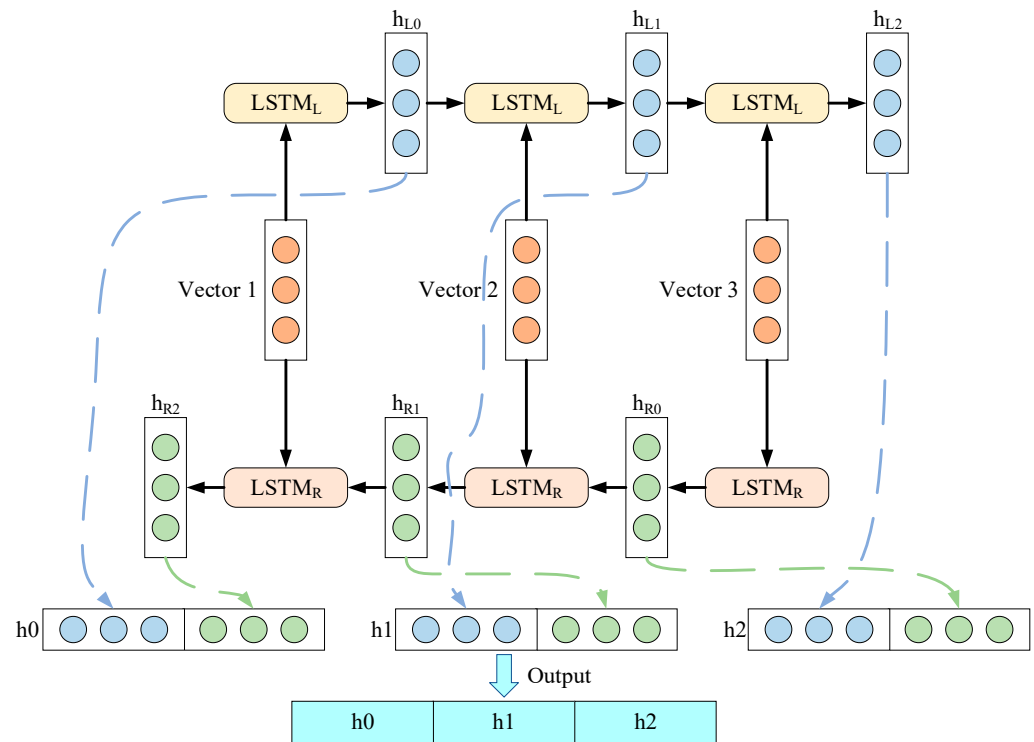
**Figure 4.** BiLSTM model structure diagram.

## 3. Intelligent Fault Diagnosis Model for Engine Misfire

In the field of engine fault diagnosis, engine misfire represents a common and critical fault mode. To enhance the precision and efficacy of diagnosing this fault, this paper proposes an intelligent diagnostic model, ResNet–BiLSTM, which integrates sensor data with the strengths of deep learning models, specifically ResNet and BiLSTM. The deep residual learning mechanism inherent to ResNet enables the effective extraction of high-level spatial features from sensor data while simultaneously addressing the vanishing gradient problem that is commonly observed in traditional deep learning models. Concurrently, the bidirectional long short-term memory network structure of BiLSTM enables the capture of both forward and backward information from time-series data, thereby enhancing the model's capacity to represent temporal dependencies and improving the accuracy of fault detection. Subsequent to the extraction of spatial features by ResNet, BiLSTM processes these features further by capturing temporal correlations, thus augmenting the model's diagnostic capability for detecting engine misfire faults.

As illustrated in Table 1, the ResNet–BiLSTM model structure delineates the construction process. The input to the model is a $3 \times 224 \times 224$ continuous wavelet transform (CWT) image, where "3" represents the number of channels (i.e., the three channels of an RGB image), and $224 \times 224$ denotes the image's height and width. The input data is initially processed through the ResNet network for the purpose of feature extraction. The residual learning structure of ResNet enables the effective extraction of multi-level image features. Following processing through ResNet's convolutional layers and residual blocks, the output feature map has a size of $512 \times 7 \times 7$ (where 512 is the number of channels and $7 \times 7$ is the spatial dimension of the feature map). Subsequently, the model employs a pooling layer to preserve the spatial dimension of the feature map. To facilitate the processing of these spatial features by a bidirectional long short-term memory (BiLSTM) network, the feature map is first permuted using the "Permute" method, which reorders the dimensions of the feature map and is transformed into a shape of $7 \times 7 \times 512$. Subsequently, the "Reshape" method (which reshapes the feature map) further transforms it into $7 \times 3584$, where 3584 is the result of $7 \times 512$. At this juncture, the features are in a format suitable for time-series models, where "7" represents the number of time steps and 3584 is the feature

dimension at each time step. Subsequently, the features are fed into the BiLSTM layer. The bidirectional long short-term memory (BiLSTM) processes the forward and backward time-sequence information separately, generating corresponding outputs. In this model, the final temporal feature output is derived from the hidden state of the BiLSTM at the final time step, with a size of $1 \times 256$. This temporal feature represents the extracted features from the bidirectional LSTM. Subsequently, this temporal feature is passed through a fully connected layer and classified using the Softmax activation function. The model's output is a probability distribution across seven fault categories, representing different types of ship engine misfire faults.

**Table 1.** Structural parameters of ResNet–BiLSTM network model.

| Network Layer | Output Dimension | Layer Configuration |
|---|---|---|
| Input | $3 \times 224 \times 224$ | - |
| Stage 0 | $64 \times 112 \times 112$<br>$64 \times 56 \times 56$ | Conv2d k = 7 s = 2 p = 3<br>BatchNorm2d ReLu<br>MaxPool2d k = 3 s = 2 p = 1 |
| Stage 1 | $64 \times 56 \times 56$ | Residual Block $\times$ 2 |
| Stage 2 | $128 \times 28 \times 28$<br>$128 \times 28 \times 28$ | Residual Block $\times$ 2 |
| Stage 3 | $256 \times 14 \times 14$<br>$256 \times 14 \times 14$ | Residual Block $\times$ 2 |
| Stage 4 | $512 \times 7 \times 7$<br>$512 \times 7 \times 7$ | Residual Block $\times$ 2 |
| Stage 5 | $512 \times 7 \times 7$ | AdaptiveAvgPool2d |
| Stage 6 | $7 \times 7 \times 512$<br>$1 \times 3584$<br>$1 \times 256$ | Permute<br>Reshape<br>BiLSTM |
| Stage 7 | $1 \times 256$ | Linear(256,256)<br>ReLu Dropout(0.5) |
| Stage 8 | $1 \times 7$ | Linear(256,7) |

Figure 5 depicts the structural and diagnostic process of the ResNet–BiLSTM model. Initially, the instantaneous rotational speed is recorded via sensors affixed to the ship's dual-fuel engine. Then, the collected data is preprocessed. The data is transformed using a continuous wavelet transform (CWT) to convert the one-dimensional signal into a two-dimensional time-frequency image, thereby facilitating the extraction of rich time-frequency features. Subsequently, the generated time-frequency images are fed into the ResNet18 network. The ResNet18 network employs a series of convolutional layers to progressively extract salient image features, thereby ensuring the retention of critical information. Subsequently, the extracted feature maps are conveyed to the BiLSTM network. The bidirectional long short-term memory structure of the BiLSTM captures both forward and backward temporal information from the time-series data, thereby enabling a more comprehensive understanding of temporal dependencies and further optimizing feature representation. The output from the BiLSTM is passed through a fully connected layer, generating a $1 \times 7$ vector that represents seven distinct fault categories (including normal operation and six types of misfire faults). The Softmax activation function is then employed to convert this vector into a probability distribution, thereby predicting the likelihood of each fault category. Ultimately, the model utilizes cross-entropy loss to compute the discrepancy between the predicted output and the actual labels. This error is then reduced through

backpropagation, enabling the model to iteratively update its parameters and enhance its performance. Through this process, the model is able to effectively diagnose misfire faults in the ship's dual-fuel engine by analyzing the collected rotational speed signals.
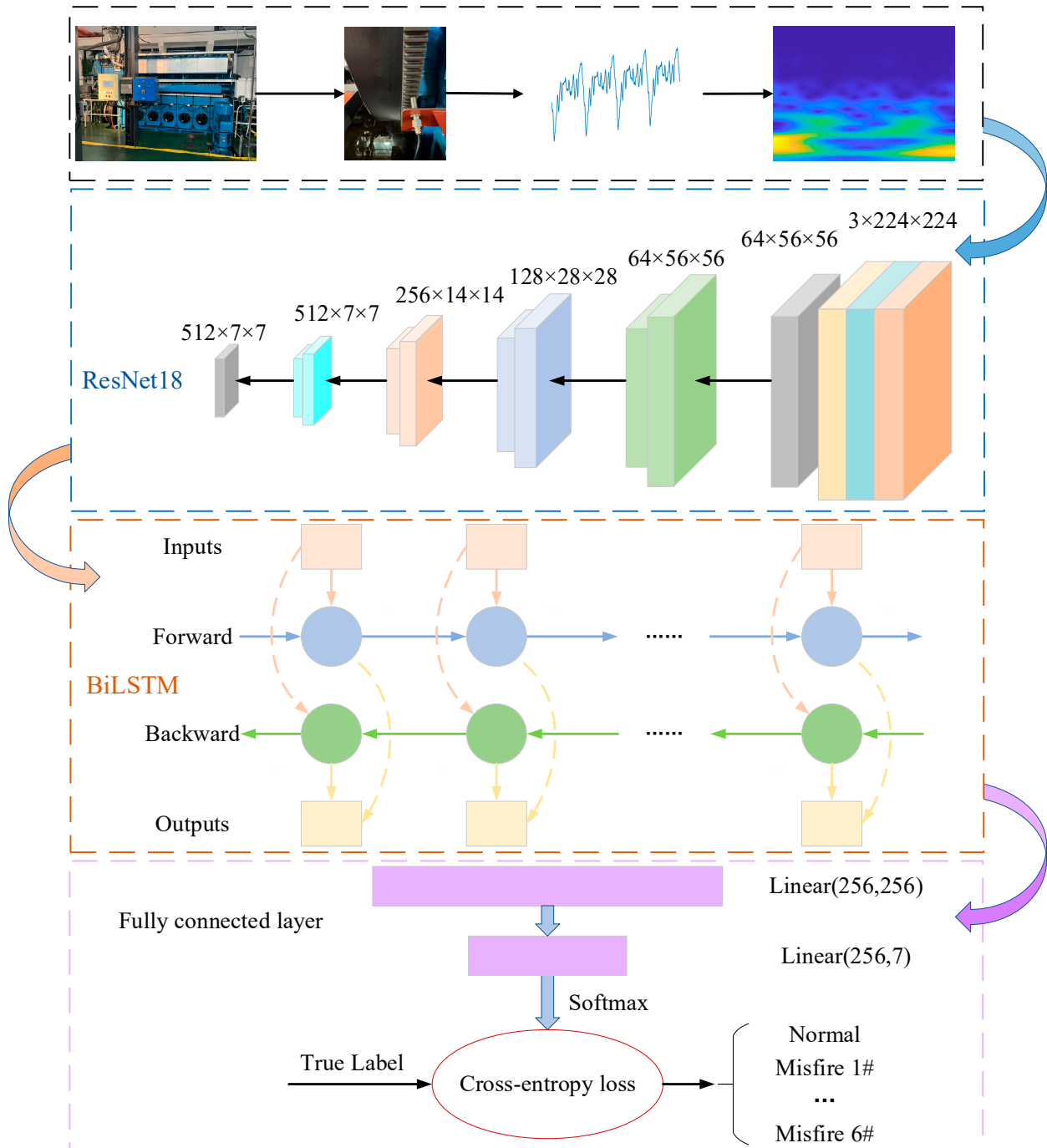


**Figure 5.** ResNet–BiLSTM model structure diagram.

## 4. Experimental Data Collection and Construction of Datasets

### 4.1. Collection of Experimental Data

The rotational speed data presented in this study were obtained from a dual-fuel engine situated within the nacelle of a smart ship, located at a university in Wuhan, China. The data were collected using a magnetoelectric speed sensor installed on the flywheel end of the engine, as illustrated in Figure 6. Figure 6a presents a photograph of the engine setup in a real-world context, Table 2 shows the specific parameters of the engine, while Figure 6b

depicts the installation location of the rotational speed sensor. The specific parameters of the dual-fuel engine are shown in the following table.
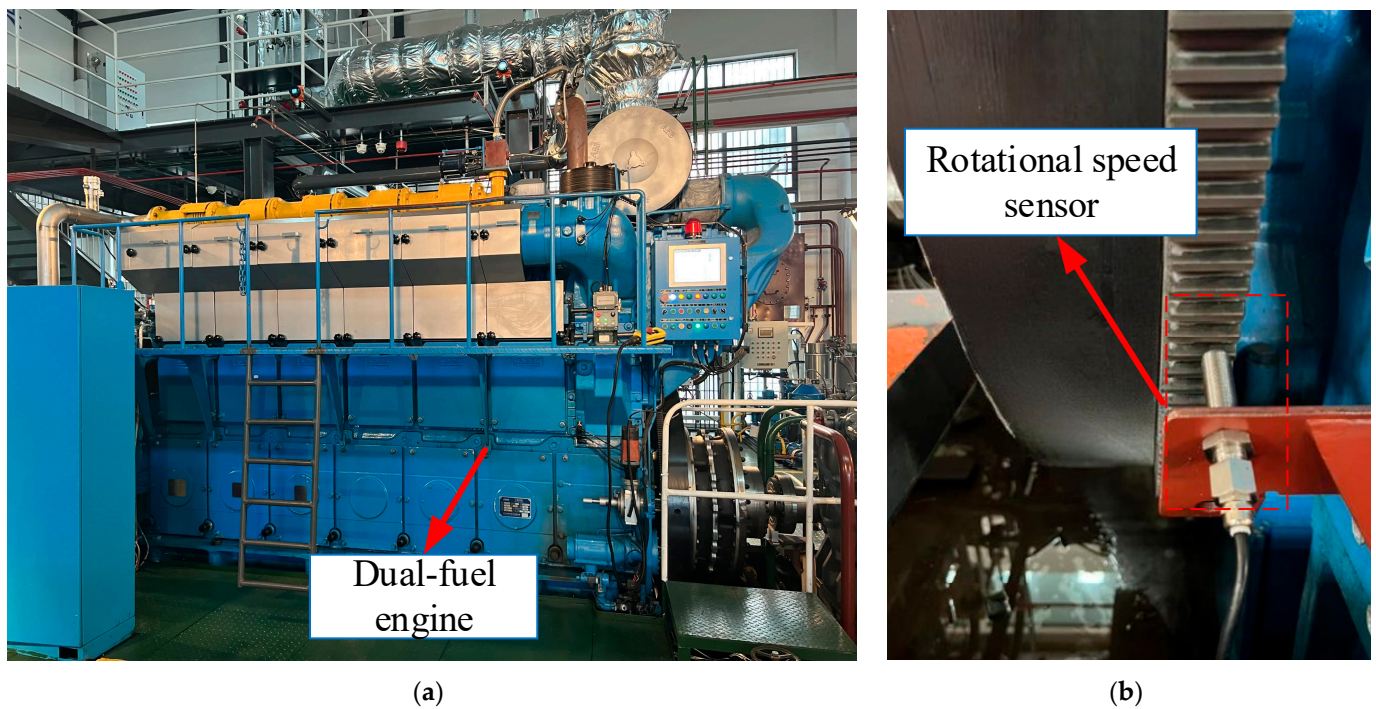


**Figure 6.** The experimental platform. (**a**) The dual-fuel engine; (**b**) The rotational speed sensor.

**Table 2.** Dual-fuel engine parameters.

| Parameters | Contents |
|---|---|
| Machine type | 6ACD320DF |
| Engine type | four stroke, in-line single action, barrel piston, main gas manifold electronic control injection, micro-injection ignition, irreversible Marine dual fuel (diesel–natgas) engine |
| Number of cylinders | 6 |
| Cylinder diameter × stroke | 320 mm × 420 mm |
| Mean effective pressure (MEP) | 1.92 MPa |
| Maximum burst pressure | ≤22 MPa |
| Specific fuel consumption (SFC) | 7423 kJ/kWh (natgas) |
|  | 185 g/kWh (diesel) |
| Volume | 5769 mm × 2370 mm × 3697 mm |
| Data of manufacture | June 2023 |

During the experiment, the sensor recorded the engine's instantaneous rotational speed at a sampling frequency of 2500 Hz. The sampling frequency satisfies the Nyquist sampling theorem, thereby ensuring the accuracy of the signal acquisition. The data were collected under different engine speed conditions, capturing both normal operation and various misfire fault conditions. This dataset provides a solid foundation for subsequent fault diagnosis and analysis.

### 4.2. Construction of Datasets

#### 4.2.1. Data Preprocessing

In order to augment the training set with a greater number of samples, this study employs a sliding window technique to enhance the instantaneous rotational speed data collected by the sensor. As illustrated in Figures 7a and 7b depict the data processing procedures for both normal and fault conditions, respectively. Specifically, the window size is set to 512 data points, with a sliding step of 320 points, which helps maintain data continuity while effectively increasing the sample size. Through this method, the original time-series data are divided into multiple overlapping subsequences, each containing 512 data points. This approach not only enriches the training set but also captures local features and temporal relationships in the data, thereby facilitating model training and generalization. The specific steps are as follows: the data are segmented into subsequences of 512 points, with a step size of 320 points, generating each subsequence as an independent sample. Collectively, these subsequences form the augmented training set. The utilization of the sliding window technique results in a notable increase in the sample size of the training set, thereby enhancing the model's accuracy and robustness while effectively reducing the risk of overfitting.
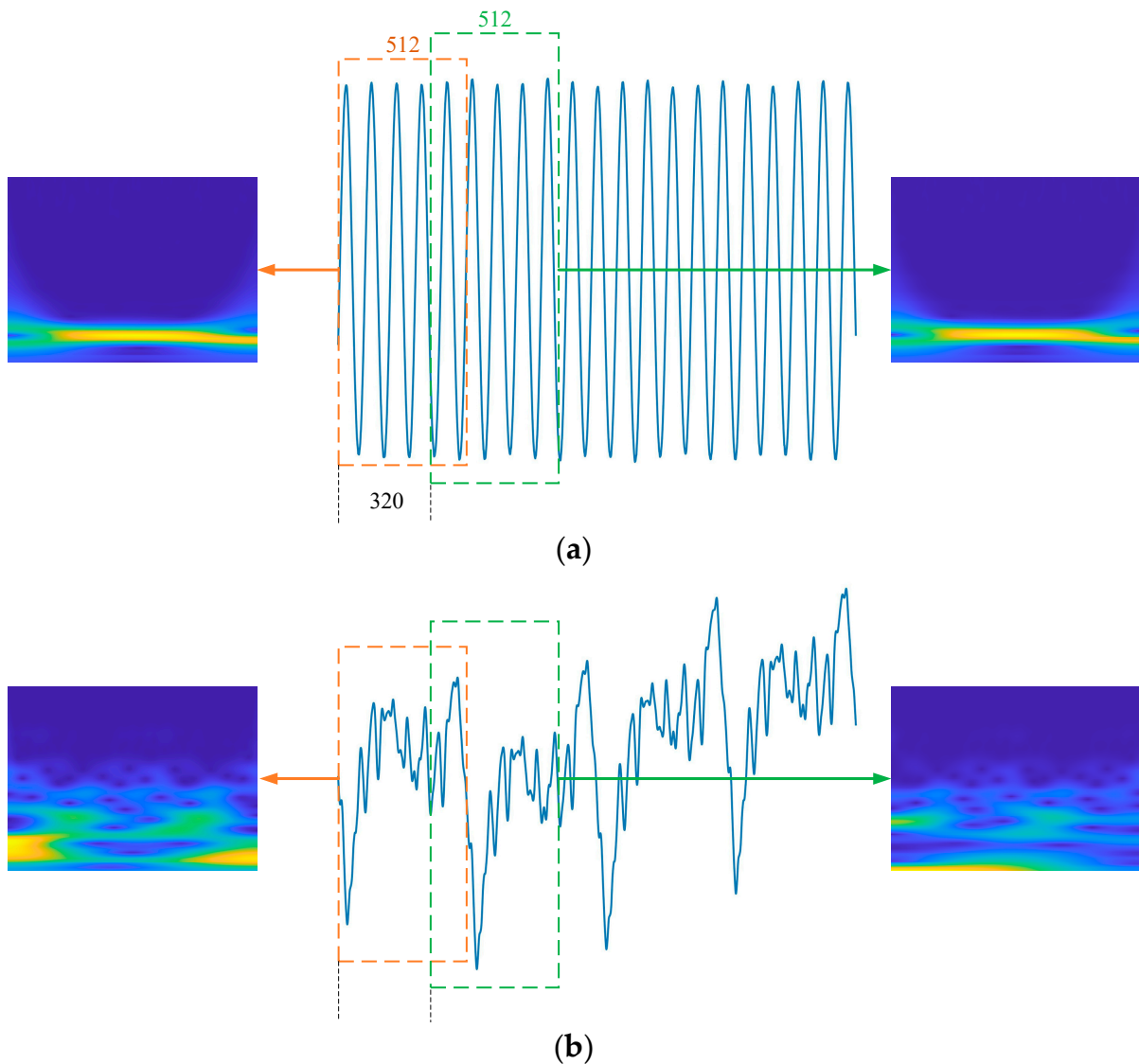


(a)



(b)

**Figure 7.** Data enhancement-sliding window method. (**a**) Normal state; (**b**) Misfire state.

In terms of fault classification, this experiment divides faults into two categories: normal state and misfire fault. The misfire fault is subdivided into six-cylinder faults because the cylinder blocks of the ship engine work independently, and the misfire of any cylinder block will have an impact on the overall performance. Therefore, subdividing cylinder block faults can more accurately locate the source of the problem and provide an accurate diagnostic basis for repair and maintenance in practical applications.

4.2.2. Data Sample Analysis

Figure 8 depicts the time domain diagram, spectrogram, and CWT image of the engine operating at an average speed of 650 rpm, encompassing both the normal state and the misfire state. In this instance, the horizontal and vertical coordinates of the CWT image correspond to time and frequency, respectively. The color and luminance represent the energy intensity of the signal at the corresponding time and frequency. A bright region (yellow) indicates a high energy level, while a dark region (blue) indicates a low energy level. A higher luminance indicates a stronger energy level. The time domain and spectrograms are plotted based on the original instantaneous rotational speed, whereas the CWT image represents the result of the original data after data enhancement, where 0 represents the normal state and 1–6 represent the misfire state of cylinders 1–6 respectively. From the time domain graphs, it is evident that under normal conditions, the instantaneous rotational speed curve approximates a regular sine wave, indicating smooth engine operation. However, in the misfire fault states, the speed curves of each cylinder display considerable irregular fluctuations. To illustrate, when cylinder 1 misfires, the curve demonstrates a periodic decline, accompanied by an increase in fluctuation amplitude. This suggests that the interruption in power output has a considerable impact on stability. In the case of cylinder 2 misfires, the fluctuations become even more irregular, with a pronounced sharp decline that significantly impacts performance. A misfire in cylinder 3 results in prolonged decreases in speed and more violent fluctuations. Additionally, cylinder 5 demonstrates greater randomness in its fluctuations, making it more challenging to predict and indicating more pronounced instability in power output. From the spectrograms, it can be observed that under normal conditions, the spectrum exhibits a distinct peak in the low-frequency range (approximately 30 Hz), with frequency components concentrated and consistent with stable operational characteristics. In the misfire states, the spectrum gradually becomes more dispersed, with smaller amplitudes appearing in the high-frequency range (greater than 100 Hz). The low-frequency primary peak still exists but with reduced amplitude, indicating that the signal's frequency becomes more complex and energy extends into higher frequency bands after a misfire. The CWT images demonstrate a smooth and uniform presentation under normal operating conditions, with the spectrum predominantly concentrated in the low-frequency range, indicative of a stable engine operation and relatively singular frequency characteristics. Conversely, in instances of misfire, the CWT images demonstrate notable abnormalities, with the emergence of more intricate and pronounced high-frequency components. The high-frequency components are indicative of the severe vibrations and instability that are caused by the misfire. To illustrate, in the case of cylinder 1 misfire, the presence of multiple high-frequency signals in the time-frequency image indicates an uneven power output. Similarly, the other cylinders also display comparable high-frequency characteristics, although there are subtle differences among their time-frequency images. For instance, the misfire spectrum of cylinders 2 and 3 is relatively concentrated, while the spectrum for cylinders 4 and 5 exhibits a broader distribution range, indicating that the misfire in these cylinders has a more pronounced impact on the overall vibration of the engine.
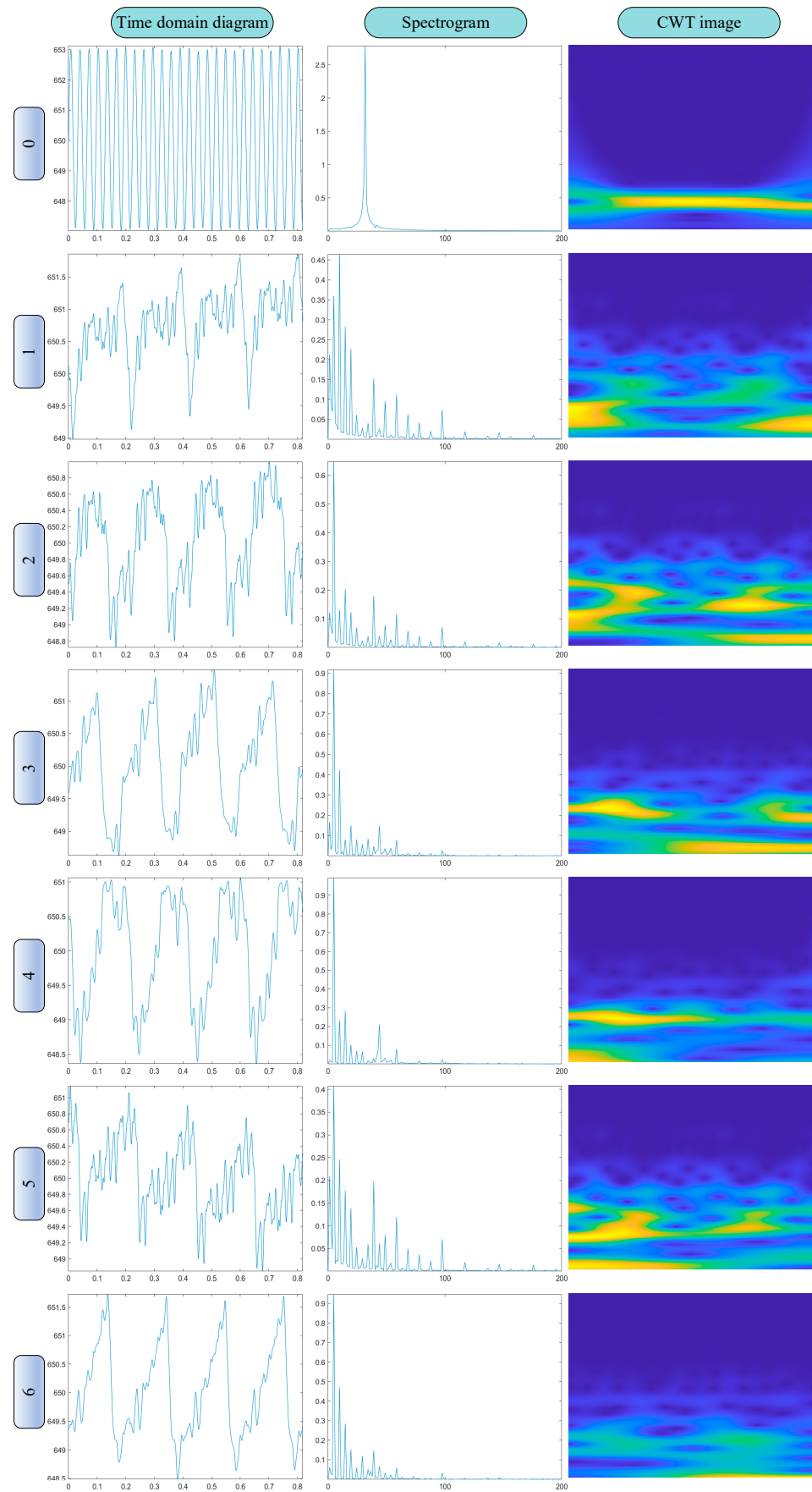
**Figure 8.** Time domain diagram, spectrogram, and CWT image under different working conditions.

In light of the aforementioned analysis of the engine's instantaneous rotational speed, the following conclusions can be drawn.

(1) Firstly, under normal operating conditions, the instantaneous rotational speed curve of the engine displays a regular sinusoidal waveform, indicating stable operation with relatively simple vibration characteristics. In contrast, during misfire fault conditions, the rotational speed signal exhibits clear anomalies, with increased vibration intensity accompanied by significant fluctuations and abrupt changes. While these anomalous signals may preliminarily indicate an issue with the engine in the time domain, they are insufficient for accurately distinguishing between misfire faults in different cylinders.

(2) Secondly, although time domain and frequency domain analyses can observe the instantaneous rotational speed variations caused by faults, the small differences in signals generated by misfires in different cylinders make it challenging to effectively differentiate between them based on time domain features. However, the time-frequency images generated by the continuous wavelet transform (CWT) can decompose the instantaneous rotational speed signals into different frequency ranges, thereby revealing the complexity and variation of frequency characteristics under fault conditions. In the misfire state, the CWT time-frequency images clearly demonstrate the presence of high-frequency components and vibration instability caused by misfires in different cylinders. Furthermore, subtle differences in the corresponding time-frequency images for each cylinder provide compelling evidence for fault classification.

(3) Consequently, by integrating time domain and time-frequency analyses, particularly through the utilization of CWT images, it is feasible to not only differentiate between normal and fault states effectively but also to accurately identify misfire fault types in different engine cylinders. For complex systems like marine dual-fuel engines, this multidimensional analysis based on instantaneous rotational speed contributes to enhanced accuracy and robustness in fault diagnosis.

### 4.2.3. Data Sample Construction

The preceding analysis demonstrates that continuous wavelet transform (CWT) images can effectively differentiate between various operating states of the engine, including normal conditions and different types of cylinder misfire faults. In light of the aforementioned conclusion, this study proceeded to augment the dataset comprising the instantaneous rotational speed signals collected from the engine. This was followed by a CWT time-frequency transformation, which yielded the corresponding CWT images. As illustrated in Table 3, the sample size for each state (normal, cylinder 1 misfire, cylinder 2 misfire, cylinder 3 misfire, cylinder 4 misfire, cylinder 5 misfire, cylinder 6 misfire) was augmented to 4000 samples per class, resulting in a total of 24,000 samples. Subsequently, the data were partitioned into training, validation, and test sets in a ratio of 7:2:1, resulting in 2800 samples for the training set, 800 samples for the validation set, and 400 samples for the test set. The dataset encompasses three distinct operating speeds. It was further enhanced in terms of diversity and representativeness by including samples at 550 rpm, 650 rpm, and 750 rpm. This approach not only ensures a reasonable division of the data but also improves the model's training effectiveness and generalization ability. The labels for each fault type were assigned values from 0 to 6, which facilitated the subsequent fault classification task. By adopting this data partitioning strategy, the model was able to learn the characteristics of various faults at different rotational speeds while evaluating Its performance on the validation and test sets. This ensured that the final model was both accurate and robust under diverse operating conditions.

**Table 3.** Division of dataset.

| Fault Type | Sample Number | | | Label |
|---|---|---|---|---|
| | **Train** | **Validation** | **Test** | |
| Normal | 2800 | 800 | 400 | 0 |
| Misfire 1# | 2800 | 800 | 400 | 1 |
| Misfire 2# | 2800 | 800 | 400 | 2 |
| Misfire 3# | 2800 | 800 | 400 | 3 |
| Misfire 4# | 2800 | 800 | 400 | 4 |
| Misfire 5# | 2800 | 800 | 400 | 5 |
| Misfire 6# | 2800 | 800 | 400 | 6 |

## 5. Results and Discussion

To validate the effectiveness of the proposed model, this study, cylinder misfire experiments were conducted by doing cylinder misfire experiments on the dual-fuel engine mentioned in Section 4.1 as well as by utilizing the samples constructed in Section 4.2.3 as inputs to the model and conducting comparative experiments with a variety of models. To comprehensively assess the superiority of the combined model, several models were selected for comparative analysis. These models include both single models and fusion models. The single models consist of classic CNN architectures (LeNet-5, AlexNet, VGG11, ResNet18) as well as BiLSTM. The fusion models include AlexNet–BiLSTM, VGG11–BiLSTM, and existing methods such as AlexNet–LSTM and VGG–LSTM.

The model parameters are set as follows: The batch size is 32, the number of epochs is 100, SGD is selected as the optimizer, cross-entropy is used as the loss function, the learning rate is set to 0.001, the deep learning framework used is Pytorch (version 2.4.0), and the programming language is Python. The hardware configurations are as follows: The central processing unit is a 12th-generation Intel Core i5-12400F, while the graphical processing unit is a NVIDIA GeForce RTX 3060 Ti, accompanied by 8 GB of RAM.

In particular, SGD is selected as the optimizer for the model for several reasons.

(1) From the convergence stability perspective, SGD provides a more stable convergence path during model training, especially at smaller learning rates. In comparison to adaptive methods (e.g., Adam, RMSprop), SGD enables the model to gradually approach the local optimal solution without over-tuning, thus avoiding unnecessary oscillations on complex datasets.

(2) Regarding the ability to generalize, SGD is often considered to have superior generalization ability. Adaptive optimizers (e.g., Adam, etc.) dynamically adjust the learning rate, which may result in the model overfitting the training data. In contrast, SGD can more effectively control the generalization effect of the model and perform more robustly on the validation and test sets.

(3) In regard to the efficacy of the training program, SGD combined with a smaller learning rate (e.g., 0.001) is suitable for long-term training (100 epochs) and can gradually approach the global optimum in the process of continuous updating. Furthermore, for image classification tasks, SGD is typically able to effectively utilize the feature extraction capabilities of deep networks during training, thereby assisting the network in learning features more effectively at different levels.

(4) In terms of resource requirements, In a given hardware configuration (e.g., an NVIDIA GeForce RTX 3060 Ti graphics card and 8 GB RAM), SGD is capable of achieving superior performance with constrained computational resources, while minimizing computational overhead, making it an optimal choice for training deep learning models. In contrast, while the Adam optimizer can facilitate convergence in certain instances, it typically requires more memory and computational resources, particularly when dealing with larger datasets. Consequently, in resource-constrained environments, SGD outperforms Adam due to its enhanced computational efficiency and reduced memory consumption, making it particularly well-suited for long or large-scale model training tasks.

In summary, SGD, with its stable convergence and excellent generalization ability, is more suitable for the long-term training requirements of this study.

### 5.1. Fault Diagnosis Based on Classic CNN Model

In order to identify the most appropriate model for comparative analysis, this study employed a number of classic convolutional neural network (CNN) models, including LeNet-5, AlexNet, VGG11, and ResNet18, in order to conduct fault diagnosis experiments on the same dataset. At the outset, the training and validation samples were input into each network model for training purposes. Following iterative optimization, the test samples were employed to evaluate the model's performance. This approach allows for an effective comparison of the performance of different CNN models in the fault diagnosis task, thereby facilitating the selection of the optimal model. The training results of the four classic CNN models are presented in Figure 9, which illustrates the training accuracy in Figure 9a, training loss in Figure 9b, validation accuracy in Figure 9c, validation loss in Figure 9d. The graphs for the visual observation of the performance of each model during training and validation, thus enabling an assessment of their effectiveness in the fault diagnosis task.
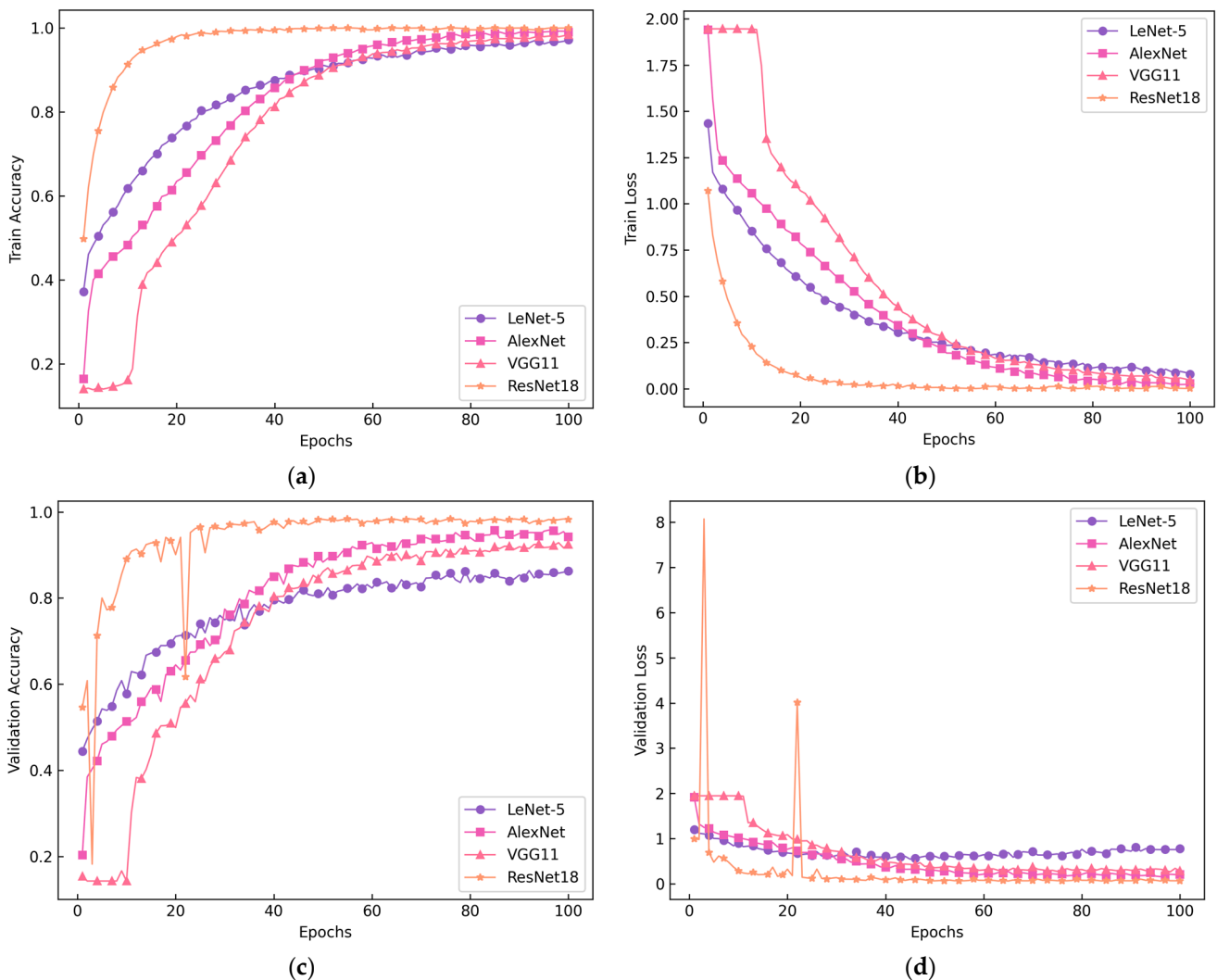


**Figure 9.** Training results of four classic CNN models. (**a**) Train Accuracy; (**b**) Train Loss; (**c**) Validation Accuracy; (**d**) Validation Loss.

A review of the performance of the four classic CNN models during the training process, as illustrated in the accompanying figure, reveals the following observations:

(1) The training accuracy is as follows: As the number of training epochs increases, the training accuracy of all models demonstrates a gradual improvement. The ResNet18 model exhibits the highest accuracy, approaching 1.0, suggesting that it performs optimally on the training set. Additionally, VGG11 and AlexNet demonstrate robust performance with high accuracy, though slightly below that of ResNet18. LeNet-5 exhibits comparatively suboptimal performance, with a gradual increase in accuracy that does not reach a high level by the conclusion of the experiment.

(2) Similarly, as the number of epochs increases, the training loss of all models decreases, which aligns with the trend of rising accuracy. The decrease in loss for ResNet18 is the most rapid, with the loss value dropping and stabilizing at an early stage of the process, indicating a smooth optimization. VGG11 and AlexNet exhibit a slower decline in loss, with the final values slightly above that of ResNet18. LeNet-5's loss value decreases rapidly in the initial stages but remains at a higher level by the end.

(3) With regard to the validation set accuracy, the ResNet18 model demonstrates the most optimal performance on the validation set, with its accuracy approaching 1.0 at an early stage, thereby exhibiting excellent generalization capabilities. Subsequently, VGG11 and AlexNet demonstrate a gradual increase in validation accuracy, stabilizing in the later epochs but still lower than ResNet18. In contrast, LeNet-5 exhibits relatively low validation accuracy, reaching a peak of approximately 0.8, which suggests limited generalization capability.

(4) From the validation set loss, ResNet18 exhibits the most stable validation loss, reaching a minimum at an early stage, which reflects its robust optimization performance. VGG11 and AlexNet display some fluctuations in the initial stages, but their losses stabilize as the training progresses. LeNet-5's validation loss initially decreases at a gradual rate, followed by an increase, and remains relatively high, indicating the potential for underfitting on the validation set.

To validate the generalization capability of the aforementioned classic CNN models and eliminate the influence of randomness on the results, each model was subjected to 10 independent repeat experiments. The specific steps were as follows: while maintaining the model hyperparameters constant, 50% of each fault class was randomly selected as a new test set for the independent evaluation of the models. The mean, standard deviation, and measurement time for each model over 10 trials are presented in Table 4, with the results illustrated in Figure 10. The results demonstrate that ResNet18 not only exhibits the shortest measurement time and the smallest standard deviation in comparison to other classic CNN models, but it also displays superior accuracy performance.

**Table 4.** Independent repetition test time of classical CNN models.

| Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Mean | Standard Deviation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LeNet-5 | 14.85 | 14.85 | 14.91 | 14.95 | 15.31 | 14.83 | 14.82 | 15.38 | 15.24 | 15.35 | 15.05 | 0.23 |
| AlexNet | 14.93 | 14.86 | 14.79 | 14.97 | 15.16 | 14.97 | 15.01 | 15.09 | 14.98 | 14.87 | 14.96 | 0.10 |
| VGG11 | 17.81 | 17.54 | 17.35 | 17.71 | 17.63 | 17.54 | 17.54 | 17.35 | 17.42 | 17.68 | 17.56 | 0.15 |
| ResNet18 | 13.76 | 13.75 | 13.81 | 13.69 | 13.82 | 13.76 | 13.77 | 13.81 | 13.77 | 13.78 | 13.77 | 0.04 |

Moreover, Figure 11 depicts the four most optimal training outcomes for the ResNet18 model. Figure 11a–d illustrate the training accuracy, training loss, validation accuracy, and validation loss, respectively. The presented graphs offer a more detailed representation of the model's performance, showcasing the consistency and efficacy of ResNet18 across various training processes.
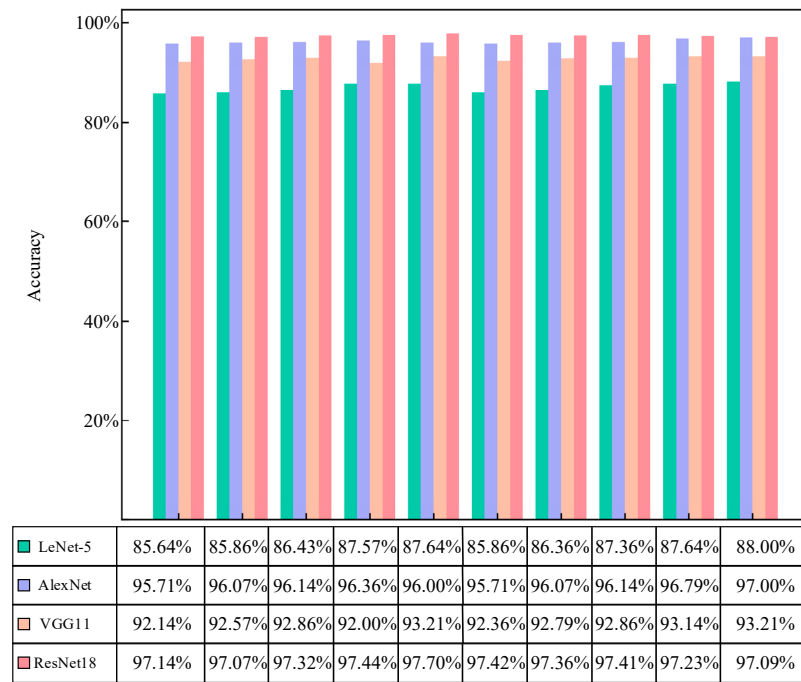
| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ■ LeNet-5 | 85.64% | 85.86% | 86.43% | 87.57% | 87.64% | 85.86% | 86.36% | 87.36% | 87.64% | 88.00% |
| ■ AlexNet | 95.71% | 96.07% | 96.14% | 96.36% | 96.00% | 95.71% | 96.07% | 96.14% | 96.79% | 97.00% |
| ■ VGG11 | 92.14% | 92.57% | 92.86% | 92.00% | 93.21% | 92.36% | 92.79% | 92.86% | 93.14% | 93.21% |
| ■ ResNet18 | 97.14% | 97.07% | 97.32% | 97.44% | 97.70% | 97.42% | 97.36% | 97.41% | 97.23% | 97.09% |

**Figure 10.** Repeat test results of classical CNN models.
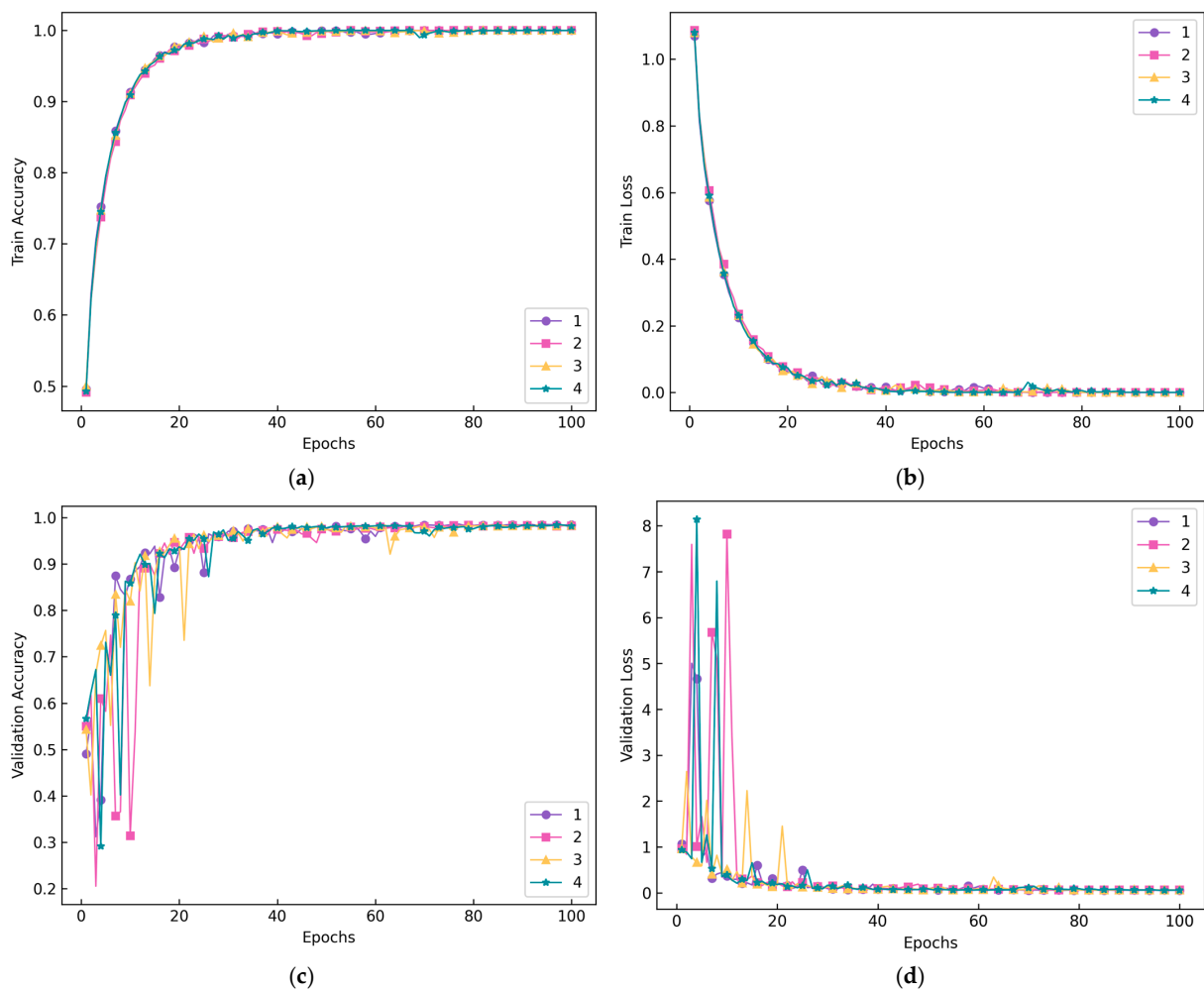


**Figure 11.** Four training results of the ResNet18 model. (**a**) Train Accuracy; (**b**) Train Loss; (**c**) Validation Accuracy; (**d**) Validation Loss.

As illustrated in Figure 11, while ResNet18 demonstrates remarkable proficiency on the training set, it displays considerable variability on the validation set, particularly during the initial stages of training. This suggests that ResNet18 is unable to fully account for the temporal dependencies inherent in the rotational speed data. The model is unable to effectively process time-series information, which has a detrimental impact on its ability to generalize on the validation set.

### 5.2. Fault Diagnosis Based on BiLSTM Model

To assess the efficacy of models designed to process time-series data, fault diagnosis experiments were conducted utilizing BiLSTM models (single-layer, double-layer, and triple-layer) on the identical dataset. In this experiment, the single-layer BiLSTM model is represented by BiLSTM1, the double-layer BiLSTM model is represented by BiLSTM2, and the triple-layer BiLSTM model is represented by BiLSTM3. The number of hidden nodes for all three BiLSTM models was set to 128. The training results of these models are illustrated in Figure 12, where plots (a), (b), (c), and (d) represent the training accuracy, training loss, validation accuracy, and validation loss, respectively. A summary of the diagnosis results is provided in Table 5.
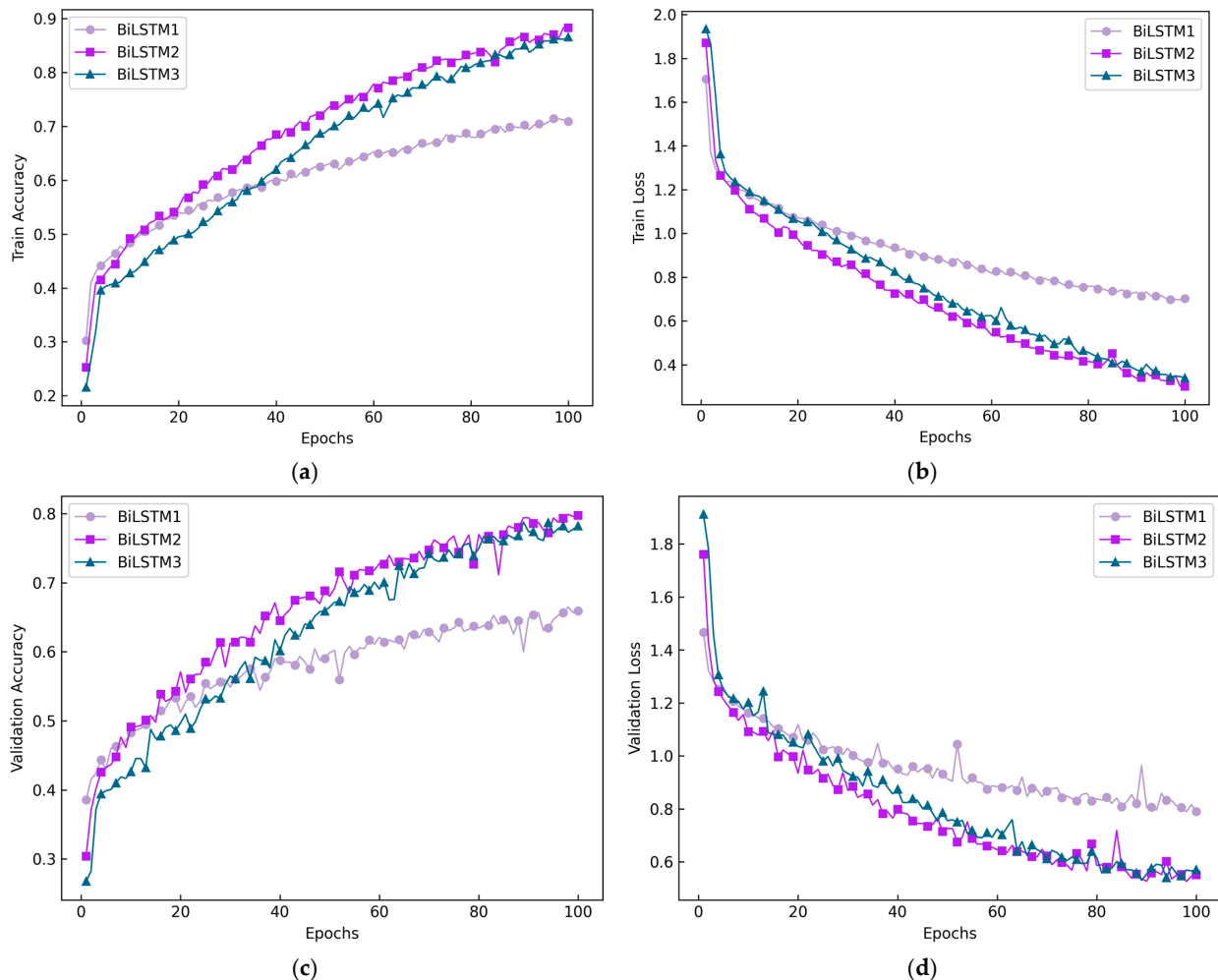


**Figure 12.** Comparison results of different BiLSTM models. (**a**) Train Accuracy; (**b**) Train Loss; (**c**) Validation Accuracy; (**d**) Validation Loss.

**Table 5.** Diagnostic results of different BiLSTM models.

| Model | Accuracy of Training/% | Training Loss | Accuracy of Verification/% | Validation Loss | Times/h |
|---|---|---|---|---|---|
| BiLSTM1 | 71.47 | 0.6971 | 66.48 | 0.7891 | 0.784 |
| BiLSTM2 | 88.35 | 0.3006 | 79.89 | 0.5259 | 0.801 |
| BiLSTM3 | 86.60 | 0.3395 | 78.77 | 0.5308 | 0.835 |

As illustrated in Figure 12 and Table 5, the double-layer BiLSTM model exhibits superior performance in comparison to the single-layer and triple-layer models. The double-layer BiLSTM model exhibits superior training accuracy, reaching 88.35%, and validation accuracy, at 79.89%. These values exceed those of the other two models. Furthermore, the loss values of the double-layer BiLSTM are markedly lower than those of the single-layer and triple-layer models, suggesting that it more effectively captures the data features during training. Furthermore, the training time remains within a reasonable range. The double-layer structure allows for deeper extraction of temporal information, thereby providing stronger modeling capability compared to the single-layer model. At the same time, it avoids the potential overfitting issues observed in the triple-layer model, as its moderate complexity strikes a balance between feature learning and model generalization.

To assess the generalization capacity of the double-layer BiLSTM model, performance evaluations were conducted using the test set. The specific procedure was as follows: the trained model was subjected to 10 independent experiments using the test set. For each test, 50% of the samples from each category were randomly selected to form a new test set, ensuring category balance. The resulting test results are presented in Figure 13 and Table 6. The content of the bar graph represents accuracy and the content of the line graph represents test time. The average accuracy across the ten independent experiments was 79.99%, with a standard error of 1.01, indicating that the model's performance remains relatively stable across different combinations of test samples.
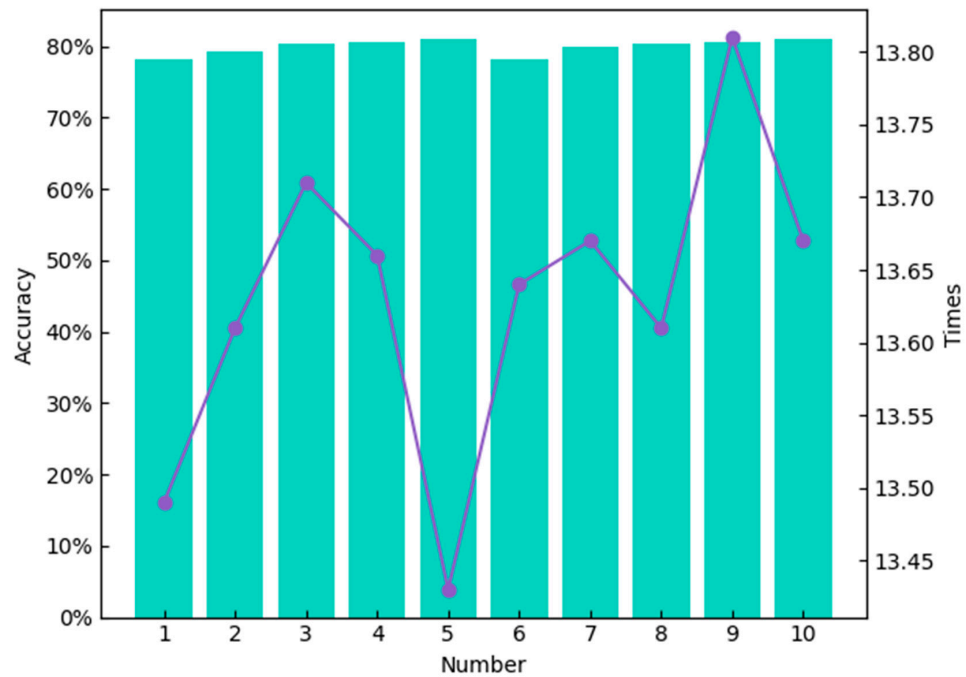


**Figure 13.** Repeat test results of the double-layer BiLSTM model.

**Table 6.** Repeated test accuracy and time of the double-layer LSTM model.

| Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Mean | Standard Deviation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Test Accuracy/% | 78.21 | 79.36 | 80.29 | 80.64 | 81.07 | 78.21 | 79.93 | 80.43 | 80.71 | 81.07 | 79.99 | 1.01 |
| Test Times/s | 13.49 | 13.61 | 13.71 | 13.66 | 13.43 | 13.64 | 13.67 | 13.61 | 13.81 | 13.67 | 13.63 | 0.10 |

The analysis of BiLSTM models with varying numbers of layers revealed that, while the double-layer BiLSTM model exhibited relatively superior performance, the training process still encounters convergence issues. This suggests that the model's capacity to extract features remains constrained, particularly when confronted with intricate time-series data. While the double-layer BiLSTM structure offers improvements in capturing temporal dependencies, it may still encounter difficulties in optimizing when faced with more intricate signal patterns, potentially resulting in suboptimal performance.

### 5.3. Fault Diagnosis Based on Fusion Model

In light of the preceding analysis, this study has selected AlexNet, VGG11, and ResNet18—three classic convolutional neural network (CNN) models that have demonstrated robust performance—and has combined each with a double-layer bidirectional long short-term memory (BiLSTM) unit to construct three fusion models. These convolutional neural network (CNN) models are capable of extracting powerful features from the input data, effectively capturing spatial characteristics. Bidirectional long short-term memory (BiLSTM) networks are particularly adept at handling sequential data and capturing long-range dependencies within sequences. Therefore, the combination of convolutional neural networks (CNNs) and bidirectional long short-term memory (BiLSTM) networks can fully leverage the advantages of CNNs in feature extraction while utilizing BiLSTM to manage dynamic changes in time series. In the AlexNet–BiLSTM and VGG11–BiLSTM models, the feature extraction components utilize the AlexNet and VGG11 networks, respectively. The extracted features are adjusted using the Permute and Reshape methods to ensure compatibility with the input format of the BiLSTM. The training results of the three fusion models are illustrated in Figure 14, where graphs (a), (b), (c), and (d) represent training set accuracy, training set loss, validation set accuracy, and validation set loss, respectively. The diagnostic results are shown in Table 7.
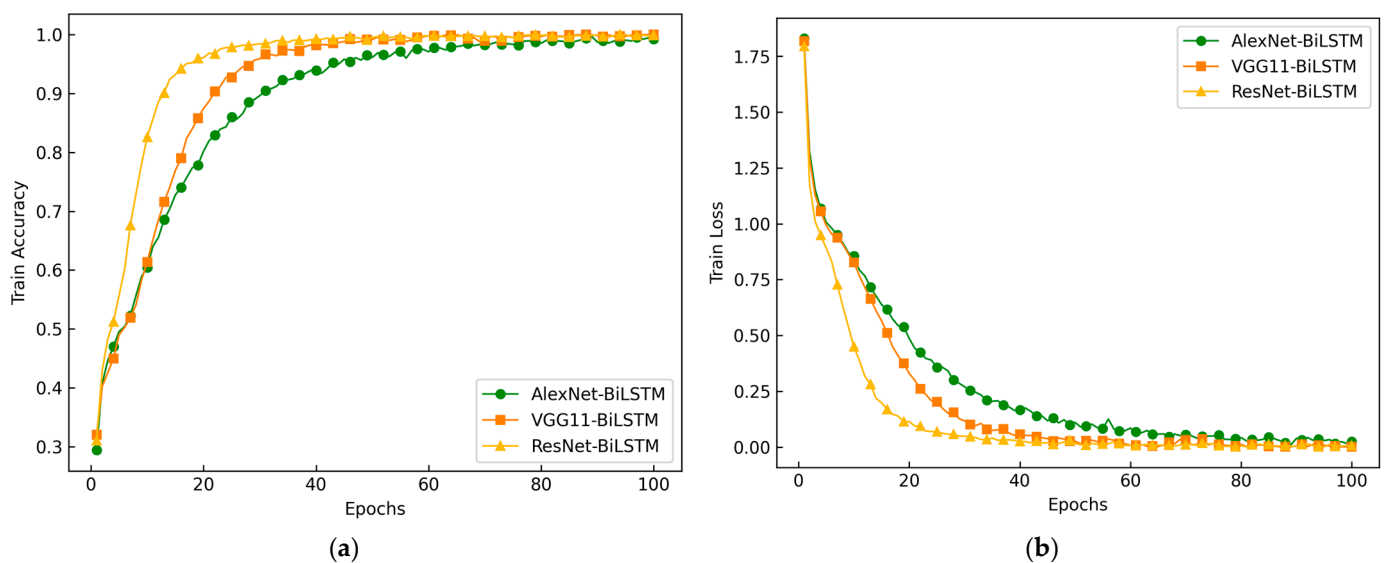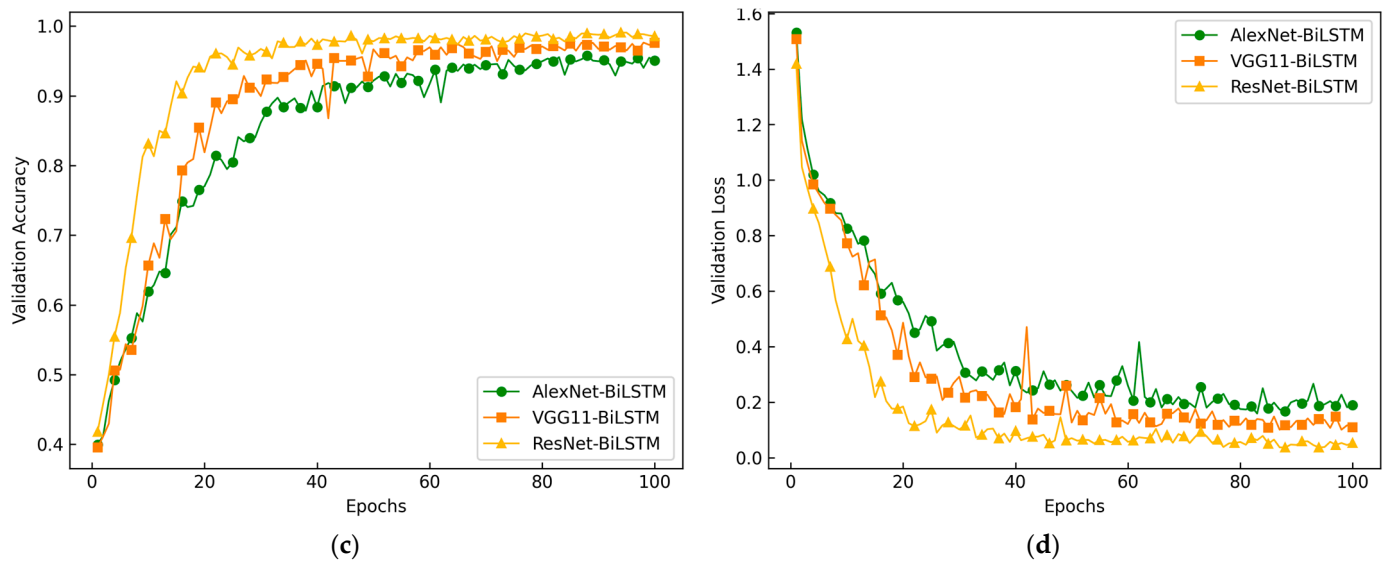


(a)

(b)

**Figure 14.** *Cont.*

**Figure 14.** Training results of three fusion models. (**a**) Train Accuracy; (**b**) Train Loss; (**c**) Validation Accuracy; (**d**) Validation Loss.

**Table 7.** Diagnostic results of three fusion models.

| Model | Accuracy of Training/% | Training Loss | Accuracy of Verification/% | Validation Loss | Times/h |
|---|---|---|---|---|---|
| AlexNet–BiLSTM | 99.66 | 0.0116 | 95.78 | 0.1589 | 1.521 |
| VGG11–BiLSTM | 99.95 | 0.0018 | 97.64 | 0.1007 | 2.818 |
| ResNet–BiLSTM | 99.97 | 0.0016 | 99.08 | 0.0364 | 2.949 |

The figure above illustrates the performance of the three different fusion models during the training process, thereby revealing the following conclusions.

(1) The accuracy of the training set is as follows: As the number of epochs increases, the ResNet–BiLSTM model demonstrates a rapid increase in training accuracy, reaching a value of approximately 1.0 after approximately 40 epochs, with a relatively stable curve. Additionally, the VGG11–BiLSTM model rapidly attains an accuracy approaching 1.0, exhibiting a trajectory closely aligned with that of the ResNet–BiLSTM model, but the convergence rate was insufficient. In contrast, the AlexNet–BiLSTM model demonstrates a more gradual improvement in accuracy. While it ultimately approaches 1.0, its overall growth rate is not as rapid as that of the first two models, and it experiences notable delays during the mid-training phase.

(2) With the increase in epochs, the loss curves for ResNet–BiLSTM and VGG11–BiLSTM decrease rapidly, approaching zero after 50 epochs, indicating that these models experience a swift reduction in loss during training. In contrast, the loss curves for AlexNet–BiLSTM display a slower decline, with significantly higher values in the initial stages compared to the other two models. While it also approaches zero after 100 epochs, this reduction occurs at a slower pace.

(3) The validation accuracy curve for ResNet–BiLSTM exhibits the most optimal performance, displaying a rapid increase and stabilizing near 1.0 after 30 epochs, with a relatively smooth curve and minimal fluctuations. Similarly, the VGG11–BiLSTM model also exhibits a rapid increase in accuracy, approaching 1.0. However, it displays greater fluctuations between certain epochs, indicating slightly lower stability. In contrast, the performance of AlexNet–BiLSTM on the validation set is inferior, with a curve that fluctuates significantly and maintains a relatively low accuracy.

(4) Regarding the validation set loss, ResNet–BiLSTM demonstrates a rapid decrease, maintaining a low level throughout the latter stages of training with minimal fluctuations.

VGG11–BiLSTM also exhibits a rapid decrease in validation loss, but it experiences considerable fluctuations between some epochs. AlexNet–BiLSTM presents the poorest performance in terms of validation loss, with relatively high values and substantial volatility.

The table above presents the diagnostic results of three distinct fusion models. As evidenced by the data presented in the table, ResNet–BiLSTM exhibits the most optimal performance on both the training and validation sets. It achieves the lowest training loss and validation loss, at 0.0016 and 0.0364, respectively, while also attaining the highest validation accuracy of 99.08%. Although ResNet–BiLSTM necessitates a more extended training period due to its deeper network structure as a feature extractor, in addition to the supplementary computational overhead of BiLSTM when processing temporal features, this augmented complexity results in elevated time costs. In contrast, other models, such as AlexNet–BiLSTM and VGG11–BiLSTM, while demonstrating reduced training times, exhibit inferior accuracy and loss performance. This suggests that although ResNet–BiLSTM has greater computational resource and time demands, its notable performance enhancement makes it a valuable compromise for the high accuracy demands of practical applications.

In conclusion, the ResNet–BiLSTM model demonstrates the optimal integration of the residual network architecture of ResNet18 with the temporal dependency capture capabilities of BiLSTM, thereby achieving the most comprehensive performance among the three models. The model demonstrates a high level of accuracy during training, approaching 1.0 with remarkable swiftness. Furthermore, it exhibits superior performance on the validation set, attaining a validation accuracy of 99.08% and the lowest validation loss with minimal fluctuations. This illustrates the model's robust generalization capacity and stability. Furthermore, the training time for ResNet–BiLSTM is only 2.949 h, indicating that the model maintains high precision while also exhibiting considerable computational efficiency, rendering it well-suited for applications where training time and model performance are of paramount importance. This makes it the optimal model selected for this study. In contrast, while VGG11–BiLSTM also achieves a validation accuracy of 97.64% and performs well on the training set, its validation loss and accuracy exhibit significant fluctuations during certain epochs, indicating slightly lower stability compared to ResNet–BiLSTM. Although the training time for VGG11–BiLSTM is marginally shorter than that of ResNet–BiLSTM, this time advantage does not translate into a significant performance improvement. In contrast, the larger fluctuations in its validation loss suggest that its generalization ability and stability are somewhat inferior to those of ResNet–BiLSTM. AlexNet–BiLSTM, while ultimately achieving a high training accuracy, shows a slower improvement rate and comparatively poor performance on the validation set, with a validation accuracy of only 95.78%. The higher validation loss and noticeable fluctuations indicate an insufficient generalization ability. Despite the shortest training time at just 1.521 h, the performance falls significantly short of the other two models.

To assess the model's capacity for generalization, a new test set was constructed by randomly selecting 50% of the samples from each class, thus ensuring a fair and unbiased evaluation. Subsequently, the model's performance was evaluated through 10 independent experiments, with the objective of ensuring the reliability of the results and eliminating the effects of randomness. The results of the experiment are presented in Figure 15 and Table 8. The mean test accuracy was 99.30%, with a standard deviation of 0.08, indicating that the model exhibits remarkable stability across diverse test sets. The average test time was 13.56 s, with a standard deviation of 0.07 s, thereby demonstrating consistency and efficiency in the model's operational performance. These results indicate that the proposed model not only achieves high diagnostic accuracy but also exhibits excellent stability and consistency in terms of testing time and performance fluctuations, thereby further confirming the model's generalization ability and reliability in practical applications.
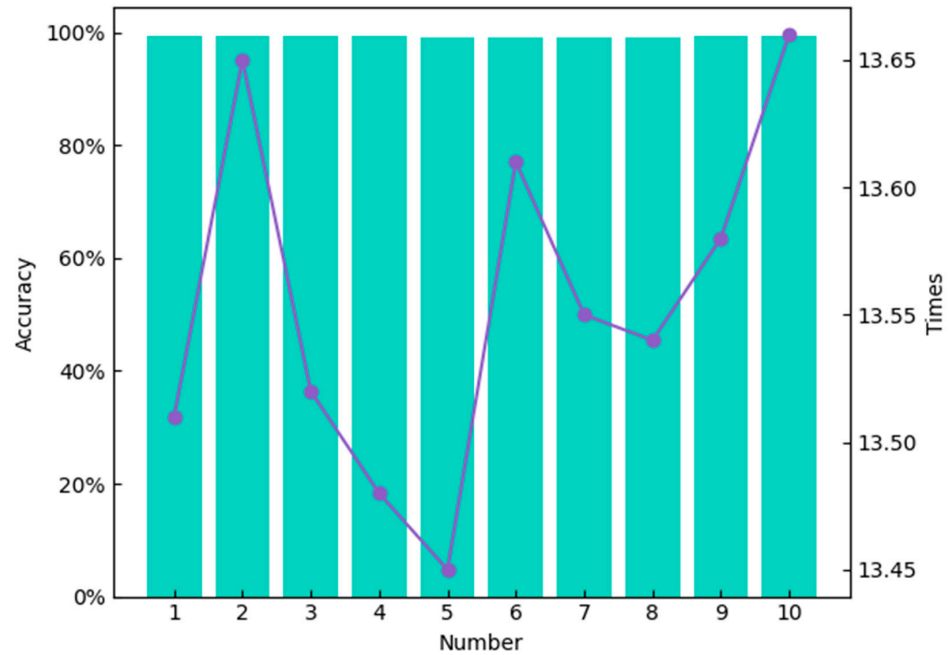
**Figure 15.** Repeat test results of ResNet–BiLSTM fusion model.

**Table 8.** Repeat test time and accuracy of ResNet–BiLSTM fusion model.

| Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Mean | Standard Deviation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Test Accuracy/% | 99.29 | 99.32 | 99.36 | 99.43 | 99.21 | 99.19 | 99.21 | 99.23 | 99.33 | 99.41 | 99.30 | 0.08 |
| Test Times/s | 13.51 | 13.65 | 13.52 | 13.48 | 13.45 | 13.61 | 13.55 | 13.54 | 13.58 | 13.66 | 13.56 | 0.07 |

*5.4. Comparative Experimental Analysis of Different Models*

To enhance the reliability of the proposed model's performance assessment, this study selected the AlexNet–LSTM model referenced in [45] and the VGG–LSTM model referenced in [46] for diagnosing engine misfire faults. The aforementioned models were then compared with the proposed ResNet–BiLSTM model based on a number of criteria, including accuracy, loss values, training time, and the number of parameters. The training results are illustrated in Figure 16, wherein figures (a), (b), (c), and (d) represent the training set accuracy, training set loss, validation set accuracy, and validation set loss, respectively. The specific diagnostic parameter results are presented in Table 9 for the reader's convenience. It is noteworthy that the diagnostic outcomes presented herein encompass comparative results for all models discussed in this study.

As illustrated in Figure 16, the performance of three distinct models during the training phase is depicted. From the figure, the following observations can be made.

(1) The training accuracy is as follows: As the number of training epochs increases, the training accuracy of the ResNet–BiLSTM model remains at a consistently high level throughout the training process, ultimately approaching 1.0. This demonstrates the model's capacity for effective fitting and robust learning with respect to the training set. Furthermore, the increasing trend is notably smooth, exhibiting minimal fluctuations, which indicates rapid and stable convergence of the model. While the remaining two models also demonstrated a training set accuracy of approximately 1.0, neither exhibited the same rapid convergence as the proposed model.

(2) As the number of training epochs increases, the training loss of the ResNet–BiLSTM model rapidly decreases in the initial stages, eventually stabilizing at a value close to zero. This trend indicates the model's effective fitting to the training data, reflecting a strong

learning capability with a low loss value. The overall trend is characterized by a smooth trajectory with minimal fluctuations, which indicates excellent convergence. In contrast, the AlexNet–LSTM model demonstrates a relatively modest reduction in training loss, ultimately reaching a value of approximately 0.017. This indicates that, despite achieving a reasonable level of accuracy on the training set, the model's fitting ability is not as strong as that of the ResNet–BiLSTM model. The final loss of the VGG–LSTM model is numerically lower than that of AlexNet–LSTM, but its convergence rate is slow.

(3) In terms of the validation accuracy. The ResNet–BiLSTM model demonstrates a rapid increase in validation accuracy during the initial stages of training (the first 30 epochs), subsequently stabilizing and approaching a level approximating 1.0. This suggests that the model is highly effective in capturing data features, thereby enabling it to learn effective representations rapidly. Throughout the majority of the training process, the validation accuracy remains at a high level, thereby demonstrating the model's robust capacity for generalization. In contrast, the AlexNet–LSTM model exhibits a gradual increase in validation accuracy, ultimately reaching a value of approximately 0.96. While this final accuracy is relatively high, the improvement process is slower, indicating that the learning speed is not as rapid as that of ResNet–BiLSTM. During specific training phases (such as the initial 30 epochs), the increase in accuracy is constrained, indicating a potential limitation in feature extraction capabilities. The VGG–LSTM model demonstrates the lowest validation accuracy throughout the training process, reaching a maximum of approximately 0.93. In the initial 40 epochs, the accuracy exhibits significant fluctuations, suggesting an unstable learning process that may be influenced by overfitting or underfitting.

(4) The ResNet–BiLSTM model exhibited the lowest validation loss, ultimately converging to approximately 0.036. This indicates that the model performs with great stability and efficacy on the validation set, exhibiting minimal fluctuations throughout the training process. This demonstrates the model's capacity for adaptability to the validation data. In contrast, the AlexNet–LSTM model exhibits a validation loss of approximately 0.116, which is considerably higher than that observed in the ResNet–BiLSTM model. This indicates that the model's performance on the validation set is suboptimal and exhibits some degree of overfitting. Although the fluctuations in validation loss are minimal, they remain higher than those of the ResNet–BiLSTM model, indicating a deficiency in generalization capability compared to ResNet–BiLSTM. The VGG–LSTM model exhibits the highest final validation loss, approximately 0.204, and experiences considerable fluctuations in the early stages of training, suggesting an unstable performance on the validation set. Although the training loss is relatively low, the validation loss indicates that the model is unable to generalize effectively, suggesting that overfitting may be a risk.

As illustrated in Table 9, the diagnostic outcomes of the various models are delineated. As can be seen from the table, the following observations can be made:

The ResNet–BiLSTM model demonstrated the most optimal overall performance, attaining the highest accuracy on both the training and validation sets (99.97% and 99.08%, respectively) and the lowest loss (0.0016 and 0.0364, respectively). Moreover, it demonstrates remarkable stability and generalization capability. Despite the increased computational complexity associated with a larger number of parameters and longer training time, the model's superior performance justifies the additional computational resources required.

In comparison, among the single models, ResNet18 demonstrates the most optimal performance, while the other models exhibit overall performance inferior to ResNet18. Nevertheless, the accuracy and loss of ResNet18 on the validation set suggest that it has limited capacity for handling sequential data, particularly in dynamic environments where the model's adaptability may be constrained. Moreover, the relatively extended training period does not confer a notable advantage in efficiency over the ResNet–BiLSTM.
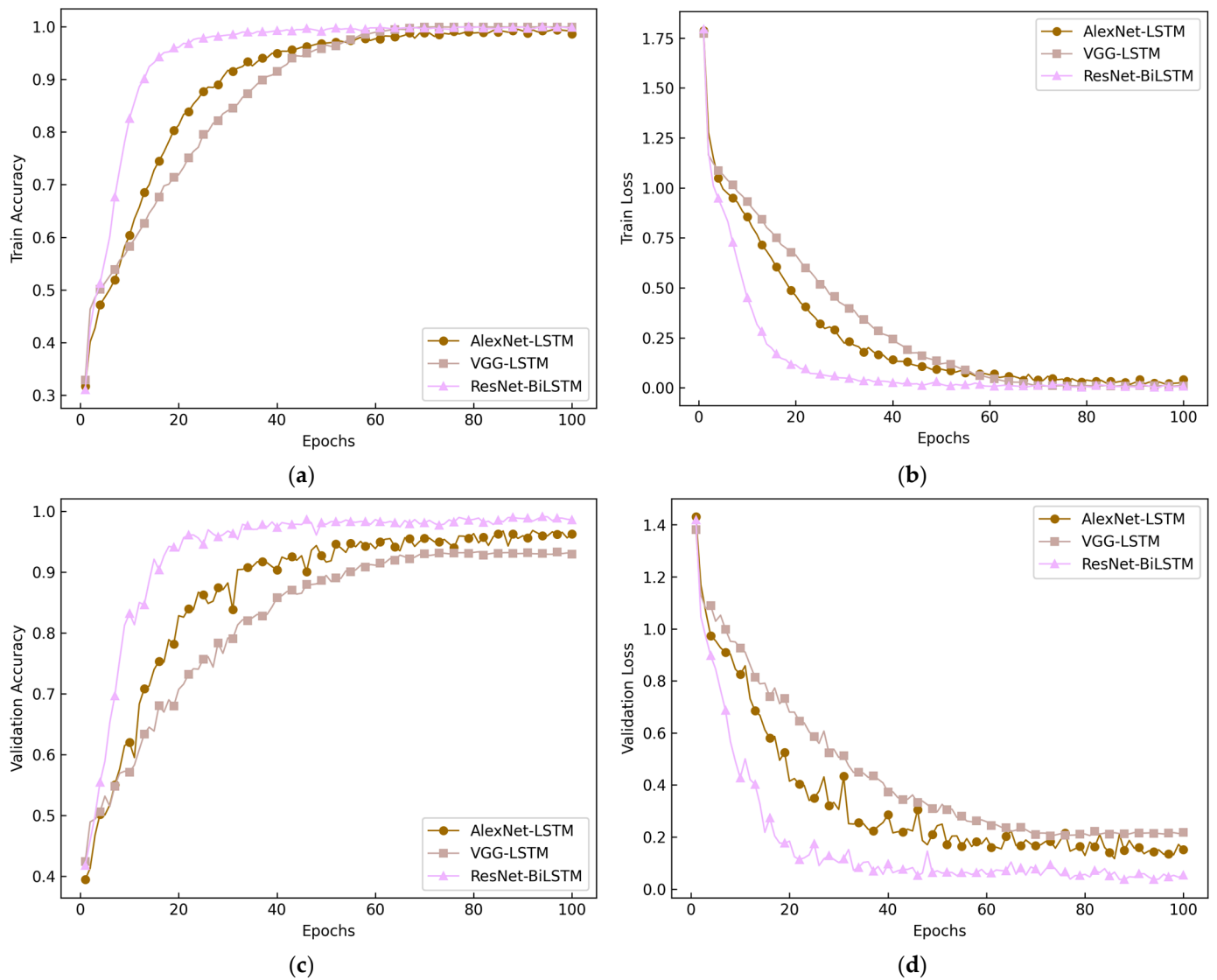
**Figure 16.** Comparison results of different models. (**a**) Train Accuracy; (**b**) Train Loss; (**c**) Validation Accuracy; (**d**) Validation Loss.

**Table 9.** Results of comparison of multiple models.

| Model | Accuracy of Training/% | Training Loss | Accuracy of Verification/% | Validation Loss | Accuracy of Testing/% | Parameter/M | Times/h |
|---|---|---|---|---|---|---|---|
| LeNet-5 | 97.17 | 0.0785 | 86.37 | 0.5446 | 86.62 | 5.407 | 1.952 |
| AlexNet | 99.23 | 0.0224 | 95.76 | 0.1573 | 96.25 | 46.776 | 3.878 |
| VGG11 | 98.41 | 0.0459 | 92.83 | 0.2626 | 92.55 | 132.863 | 4.601 |
| ResNet18 | 99.95 | 0.0023 | 98.35 | 0.0601 | 98.04 | 11.183 | 2.485 |
| BiLSTM1 | 71.47 | 0.6971 | 66.48 | 0.7891 | 65.72 | 0.823 | 0.784 |
| BiLSTM2 | 88.35 | 0.3006 | 79.89 | 0.5259 | 80.11 | 1.218 | 0.801 |
| BiLSTM3 | 86.60 | 0.3395 | 78.77 | 0.5308 | 77.64 | 1.614 | 0.835 |
| AlexNet–BiLSTM | 99.66 | 0.0116 | 95.78 | 0.1589 | 95.75 | 3.328 | 1.521 |
| VGG11–BiLSTM | 99.95 | 0.0018 | 97.64 | 0.1007 | 97.78 | 13.487 | 2.818 |
| AlexNet–LSTM | 99.48 | 0.0174 | 96.82 | 0.1168 | 96.54 | 2.835 | 1.208 |
| VGG– LSTM | 99.94 | 0.0082 | 93.32 | 0.2048 | 93.25 | 11.185 | 2.492 |
| ResNet–BiLSTM | 99.97 | 0.0016 | 99.08 | 0.0364 | 99.11 | 15.443 | 2.949 |

Among the fusion models, AlexNet–LSTM exhibits relatively good overall performance, effectively combining the feature extraction ability of AlexNet with the time-series

learning ability of LSTM. However, AlexNet–LSTM demonstrates significantly lower accuracy than ResNet–BiLSTM in both the validation and test sets, indicating that it may encounter greater challenges in addressing complex time-series data, particularly in terms of the model's generalization ability and robustness.

In conclusion, the ResNet–BiLSTM model, which demonstrated superior performance in training and validation, as well as notable advantages in generalization capability and model stability, was identified as the optimal model in this study.

As illustrated in Figure 17, the confusion matrices of the four models—ResNet18, AlexNet–LSTM, VGG–LSTM, and ResNet–BiLSTM—further substantiate the exceptional performance of the ResNet–BiLSTM model. The model exhibits high accuracy in identifying diverse fault types in the classification task. It is noteworthy that the classification accuracy for the normal state reached 100%, which is indicative of exceptional performance. The classification accuracy for cylinder 6 misfire is 99.50%, while the accuracies for cylinders 2, 4, and 5 misfires are all 99.25%. The accuracy for cylinder 3 misfire is 99.00%. In comparison, the classification accuracy for cylinder 1 misfire is slightly lower, at 98.25%. These findings suggest that the ResNet–BiLSTM model exhibits a high degree of discernibility and reliability in addressing complex fault types associated with diverse cylinder misfires.
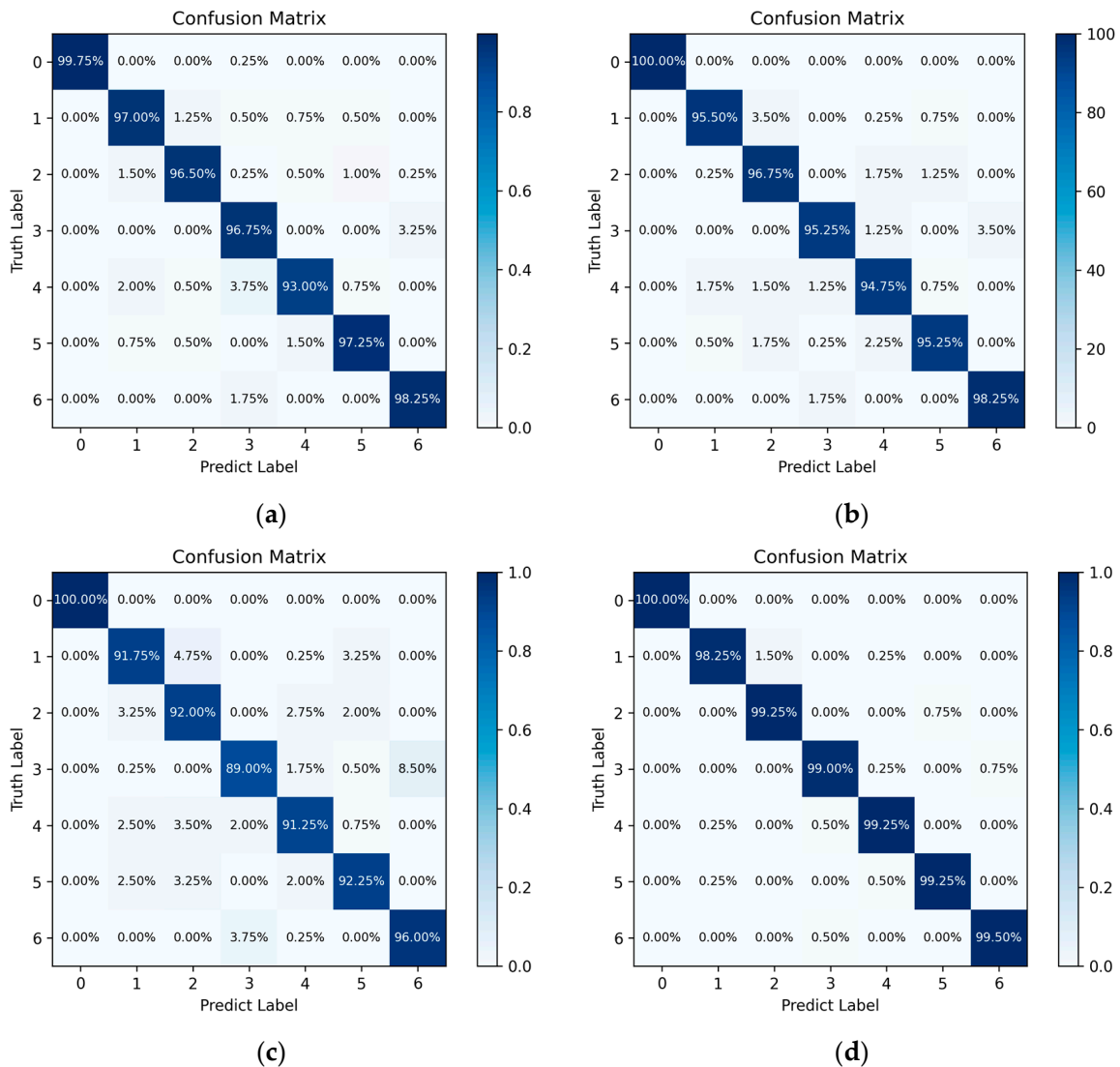


**Figure 17.** Comparison of confusion matrices among four models. (**a**) ResNet18; (**b**) AlexNet–LSTM; (**c**) VGG–LSTM; (**d**) ResNet–BiLSTM.

In comparison, while ResNet18 performs optimally as a standalone model, its accuracy in diagnosing cylinder 4 misfire is notably inferior to that of the ResNet–BiLSTM ensemble model, thereby underscoring its limitations in certain specific classification tasks. Furthermore, the diagnostic accuracy of ResNet18 is inferior to that of ResNet–BiLSTM when applied to other cylinder misfires, thus reinforcing the superiority of the ensemble model. While the classification performance of the AlexNet–LSTM model is superior to that of VGG–LSTM, it nevertheless falls short of the proposed model in terms of overall performance.

In summary, the ResNet–BiLSTM model, which integrates the exceptional feature extraction capabilities of ResNet18 with the advantages of BiLSTM in processing sequential information, not only outperforms the standalone ResNet18 model and the compared models in classification accuracy but also demonstrates superior generalization ability. This is particularly evident in the context of intelligent fault diagnosis for engine misfires, where it demonstrates superior stability and efficiency in diagnostic performance.

The preceding analysis demonstrates that the ResNet–BiLSTM model exhibits superior overall performance compared to the other models. This paper will conduct a more detailed evaluation of the ResNet–BiLSTM model to gain further insights into its performance in practical applications. Specifically, this section provides an in-depth analysis of the model based on precision, recall, and F1-score, as mentioned in reference [47]. The formulas for these metrics are as follows:

$$
\begin{aligned}
Precision &= \frac{TP}{TP+FP} \\
Recall &= \frac{TP}{TP+FN} \\
F1 - Score &= \frac{Precision \times Recall}{Precision + Recall}
\end{aligned}
\tag{7}
$$

where $TP$ represents the true positives, which is the number of samples that the model correctly predicts as positive; $FP$ denotes the false positives, which is the number of samples that the model incorrectly predicts as positive; and $FN$ signifies the false negatives, which is the number of samples that the model incorrectly predicts as negative.

Table 10 presents results that are consistent with those shown in Figure 17, indicating that the model performs well across all categories. While there are minor shortcomings in precision and recall for specific categories (such as cylinders 1 and 2), the overall performance remains satisfactory. This provides evidence that the model is effective and accurate in classification tasks.

**Table 10.** Evaluation results of ResNet–BiLSTM model.

| Label | Precision | Recall | F1-Score |
|---|---|---|---|
| 0 | 1 | 1 | 1 |
| 1 | 0.995 | 0.983 | 0.989 |
| 2 | 0.985 | 0.993 | 0.989 |
| 3 | 0.990 | 0.990 | 0.990 |
| 4 | 0.990 | 0.993 | 0.991 |
| 5 | 0.993 | 0.993 | 0.993 |
| 6 | 0.993 | 0.995 | 0.994 |

## 6. Conclusions

In order to achieve intelligent diagnosis of ship dual-fuel engine misfire faults, this paper proposes a ResNet–BiLSTM model that integrates ResNet with BiLSTM. This model fuses the robust local feature extraction capabilities of deep residual networks (ResNets) with the benefits of bidirectional long short-term memory (BiLSTM) networks in processing time series data, markedly enhancing the precision of identifying intricate fault patterns and augmenting diagnostic efficacy. The principal conclusions are as follows:

(1) By employing sensor-collected instantaneous rotational speed data from the engine, this study utilized a sliding window technique for data augmentation, which not only markedly increased the sample size but also simulated the operating states of the engine at disparate moments, thereby enhancing the model's adaptability to various operating

conditions. Subsequently, a continuous wavelet transform (CWT) was applied to convert the one-dimensional time series data into two-dimensional graphical data. This approach permits the concurrent examination of signals in both the time and frequency domains, thereby disclosing spectral characteristics obscured within the time series. Furthermore, the incorporation of image data augmented the diversity of data representation, enabling the model to comprehend and learn the characteristics of engine misfires from multiple scales and perspectives, thereby achieving exemplary performance in fault diagnosis tasks.

(2) By employing image data, a series of convolutional neural network (CNN) and recurrent neural network (RNN) models were developed, encompassing LeNet-5, AlexNet, VGG11, ResNet18, and bidirectional long short-term memory (BiLSTM) networks. The integration and comparative analysis of different CNNs with a double-layer BiLSTM model revealed that the ResNet–BiLSTM model, which combines ResNet with BiLSTM, demonstrated superior performance across various performance metrics. In particular, the ResNet–BiLSTM model demonstrates significantly lower loss values on both the training and validation sets in comparison to other fusion models, which indicates its superior capacity for data fitting. Moreover, this model demonstrates superior classification accuracy and exceptional generalization ability, outperforming other fusion models in these respects.

(3) A comprehensive performance analysis was conducted on the proposed ResNet–BiLSTM model in comparison to the existing AlexNet–LSTM and VGG–LSTM models, with a particular emphasis on key metrics, including accuracy, loss value, training time, and parameter count. The findings demonstrate that despite the ResNet–BiLSTM model exhibiting a greater number of parameters and a longer training period in comparison to the other two models, it attains a more rapid convergence, higher accuracy, and lower loss values, thereby exhibiting markedly superior overall performance. From a practical standpoint, the ResNet–BiLSTM model is particularly well-suited to tasks that necessitate high precision and model performance, given its exceptional accuracy and stability.

Moreover, a comprehensive assessment of the model was conducted using pivotal metrics, including the confusion matrix, precision, recall, and F1-score. The findings demonstrate that the ResNet–BiLSTM model markedly outperforms existing techniques in terms of fault diagnosis accuracy. Even in instances where fault categories are difficult to differentiate, the ResNet–BiLSTM model demonstrates an exceptional capacity for classification. The model exhibits remarkable precision in the majority of categories, underscoring its robust capacity to accurately identify positive samples.

Furthermore, the preliminary results demonstrate that the methodology proposed in this paper is not only applicable to the diagnosis of misfires in marine dual-fuel engines but also has the potential for extension to fault diagnosis tasks in engines with varying cylinder numbers (12 or 16) and various models, including diesel and gas engines. Other similar diagnostic tasks can be realized by appropriately adjusting the model parameters and structure. This method demonstrates robust fault recognition capabilities across diverse internal combustion engine types, showcasing remarkable generality and adaptability. It offers novel insights into fault detection in other internal combustion engines, further enhancing the practical applicability of the research.

Although the ResNet–BiLSTM model has been shown to perform well in intelligent fault diagnosis of ship dual-fuel engine misfires, there is still scope for further improvement. Further optimization opportunities may be identified in the following areas:

(1) It is recommended that the dataset be expanded and diversified. While the current data preprocessing and augmentation methods have effectively enhanced model performance, the scale and diversity of the dataset remain limited. Expanding the dataset, particularly by incorporating data from a greater variety of operating conditions and fault types, could enhance the model's generalization ability and robustness.

(2) The model structure may be optimized as follows: Although the ResNet–BiLSTM model effectively combines the strengths of ResNet and BiLSTM, there is still scope for further optimization of its structure. It would be beneficial to enhance the model's feature

extraction capabilities and classification performance, thereby improving its ability to recognize complex fault patterns.

(3) The integration of data from diverse sensors through multimodal learning approaches can facilitate the consolidation of information from disparate data sources, thereby enhancing the accuracy and reliability of fault diagnosis and reducing diagnostic errors attributable to the limitations of a single data source.

**Author Contributions:** Methodology, H.J. and Q.S.; Validation, J.G.; Investigation, J.G.; Writing—Original Draft Preparation, H.J. and J.G.; Writing—Review and Editing, Q.S.; Funding Acquisition, Q.S and C.S. All authors have read and agreed to the published version of the manuscript.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study. Meanwhile, data will be made available upon request.

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

## Nomenclature

| Abbreviation | Full Name |
| --- | --- |
| ResNet | Residual network |
| BiLSTM | Bidirectional long short-term memory |
| CWT | Continuous wavelet transform |
| VGG | Visual geometry group network |
| LSTM | Long short-term memory |
| ICE | Internal combustion engine |
| FFT | Fast fourier transform |
| ANNs | Artificial neural networks |
| CNNs | Convolutional neural networks |
| RNNs | Recurrent neural networks |
| SGD | Stochastic gradient descent |
| BiLSTM1 | Single-layer Bidirectional long short-term memory |
| BiLSTM2 | Double-layer Bidirectional long short-term memory |
| BiLSTM3 | Triple-layer Bidirectional long short-term memory |

## References

1. Su, Y.; Gan, H.; Ji, Z. Research on Multi-Parameter Fault Early Warning for Marine Diesel Engine Based on PCA-CNN-BiLSTM. *J. Mar. Sci. Eng.* **2024**, *12*, 965. [CrossRef]
2. Sagin, S.; Kuropyatnyk, O.; Matieiko, O.; Razinkin, R.; Stoliaryk, T.; Volkov, O. Ensuring Operational Performance and Environmental Sustainability of Marine Diesel Engines through the Use of Biodiesel Fuel. *J. Mar. Sci. Eng.* **2024**, *12*, 1440. [CrossRef]
3. Pająk, M.; Kluczyk, M.; Muślewski, Ł.; Lisjak, D.; Kolar, D. Ship Diesel Engine Fault Diagnosis Using Data Science and Machine Learning. *Electronics* **2023**, *12*, 3860. [CrossRef]
4. Ma, L.; Chen, S.; Jia, S.; Zhang, Y.; Du, H. Multi-Dimensional Global Temporal Predictive Model for Multi-State Prediction of Marine Diesel Engines. *J. Mar. Sci. Eng.* **2024**, *12*, 1370. [CrossRef]
5. Wiaterek, D.; Chybowski, L. Assessing the topicality of the problem related to the explosion of crankcases in marine main propulsion engines (1972–2018). *Sci. J. Marit. Univ. Szczec.* **2022**, *71*, 1–8. [CrossRef]
6. Gharib, H.; Kovács, G. A Review of Prognostic and Health Management (PHM) Methods and Limitations for Marine Diesel Engines: New Research Directions. *Machines* **2023**, *11*, 695. [CrossRef]
7. Je-Gal, H.; Park, Y.-S.; Park, S.-H.; Kim, J.-U.; Yang, J.-H.; Kim, S.; Lee, H.-S. Time-Series Explanatory Fault Prediction Framework for Marine Main Engine Using Explainable Artificial Intelligence. *J. Mar. Sci. Eng.* **2024**, *12*, 1296. [CrossRef]

8. Zhou, Z.; Bao, T.; Ding, J.; Chen, Y.; Wang, F.; Zhang, B. Diesel Engine Monitoring and Diagnostics Based on Artificial Neural Networks. In Proceedings of the 2024 13th International Conference on Communications, Circuits and Systems (ICCCAS), Xiamen, China, 10–12 May 2024.

9. Zhang, P.; Gao, W.; Song, Q.; Li, Y.; Wei, L.; Wei, Z. Real-Time Angular Velocity-Based Misfire Detection Using Artificial Neural Networks. *J. Eng. Gas Turbines Power* **2019**, *141*, 061008. [CrossRef]

10. Ghazaly, N.M.; Abdel-Fattah, M.; Makrahy, M.M. Determination of Engine Misfire Location Using Artificial Neural Networks. *Int. J. Vehicle Struct. Syst.* **2020**, *11*, 407–412. [CrossRef]

11. Arockia Dhanraj, J.; Muthiya, S.J.; Subramaniam, M.; Chaurasiya, P.K.; Akshayraj, N.; Selvaraju, N.; Ramanathan, K.C. Implementation of K∗ Classifier for Identifying Misfire Prediction on Spark Ignition Four-Stroke Engine through Vibration Data. In Proceedings of the SAE 2021 International Conference on Advances in Design, Materials, Manufacturing and Surface Engineering for Mobility, ADMMS 2021, Online, 8–9 October 2021; Hindustan Institute of Technology and Science, Dayananda Sagar College of Engineering, Kumaraguru College of Technology, Sagar Institute of Science and Technology: Bangalore, India, 2021.

12. Szabo, J.Z.; Bakucz, P. Real-Time Misfire Detection of Large Gas Engine Using Big Data Analytics. In Proceedings of the 2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY 2018), Subotica, Serbia, 13–15 September 2018.

13. Youssef, A.; Noura, H.; Amrani, A.E.; Adel, E.M.E.; Ouladsine, M. A Survey on Data-Driven Fault Diagnostic Techniques for Marine Diesel Engines. *IFAC-PapersOnLine* **2024**, *58*, 55–60. [CrossRef]

14. Lv, Y.; Yang, X.; Li, Y.; Liu, J.; Li, S. Fault Detection and Diagnosis of Marine Diesel Engines: A Systematic Review. *Ocean Eng.* **2024**, *294*, 116798. [CrossRef]

15. Cui, Y.; Liu, H.; Wen, M.; Ming, Z.; Zheng, Z.; Han, Y.; Cheng, S.; Yao, M. Comprehending flame development and misfire at advanced engine conditions: Detailed experimental characterizations and machine learning-assisted kinetic analyses. *Phys. Fluids* **2024**, *36*, 055161. [CrossRef]

16. Li, C.; Hu, Y.; Jiang, J.; Cui, D. Fault Diagnosis of a Marine Power-Generation Diesel Engine Based on the Gramian Angular Field and a Convolutional Neural Network. *J. Zhejiang Univ. Sci. A* **2024**, *25*, 470–482. [CrossRef]

17. Han, P. Tooth Time-Based Engine Misfire Detection Index for Multicylinder Engines of Vehicles Not Affected by Various Deviations between Cylinders. *SAE Int. J. Engines* **2022**, *15*, 413–426. [CrossRef]

18. Wang, P.; Xu, C.; Wang, F.; Shen, C.; Fan, L. Function Development of Real-Time Detection for Engine Misfire Fault Based on Adaptive Threshold Diagnosis. *J. Harbin Eng. Univ.* **2024**, *45*, 622–632. [CrossRef]

19. Sharib, M.K.N.b.M.; Bakar, E.b.A.; Hawary, A.F.b.; Akhtar, M.N.; Abdullah, M.N. Adaptive system on engine misfire and knocking features for IC engine. *Mater. Today Proc.* **2023**, *in press*. [CrossRef]

20. Syta, A.; Czarnigowski, J.A.; Jaklinski, P. Detection of cylinder misfire in an aircraft engine using linear and non-linear signal analysis. *Measurement* **2021**, *174*, 108982. [CrossRef]

21. Jafari, M.; Borghesani, P.; Verma, P.; Eslaminejad, A.; Ristovski, Z.; Brown, R. Detection of Misfire in a Six-Cylinder Diesel Engine Using Acoustic Emission Signals. In Proceedings of the ASME 2018 International Mechanical Engineering Congress and Exposition, Brisbane, Australia, 9–15 November 2018.

22. Kang, H.H.; Ahn, J.H.; Kim, H.Y. Monitoring of Misfiring Status of Ship Engines Using Minute Speed Changes in the Crankshaft. *J. Sensor Sci. Technol.* **2022**, *31*, 51–56. [CrossRef]

23. Mohan, B.; Pal, P.; Badra, J.; Pei, Y.; Som, S. Chapter 1—Introduction. In *Artificial Intelligence and Data Driven Optimization of Internal Combustion Engines*; Badra, J., Pal, P., Pei, Y., Som, S., Eds.; Elsevier: Amsterdam, The Netherlands, 2022; pp. 1–23.

24. Lei, Y.; Yang, B.; Jiang, X.; Jia, F.; Li, N.; Nandi, A.K. Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mech. Syst. Signal Process.* **2020**, *138*, 106587. [CrossRef]

25. Tsaganos, G.; Nikitakos, N.; Dalaklis, D.; Ölcer, A.I.; Papachristos, D. Machine learning algorithms in shipping: Improving engine fault detection and diagnosis via ensemble methods. *WMU J. Marit. Aff.* **2020**, *19*, 51–72. [CrossRef]

26. Syta, A.; Czarnigowski, J.; Jakliński, P.; Marwan, N. Detection and identification of cylinder misfire in small aircraft engine in different operating conditions by linear and non-linear properties of frequency components. *Measurement* **2023**, *223*, 113763. [CrossRef]

27. Singh, S. An improved method of detecting engine misfire by sound quality metrics of radiated sound. *Proc. Inst. Mech. Eng. D.* **2019**, *233*, 3112–3124. [CrossRef]

28. Mulay, S.; Sugumaran, V.; Devasenapati, S.B. Misfire detection in I.C. engine through ARMA features using machine learning approach. *PIE* **2018**, *12*, 93–111. [CrossRef]

29. Liu, B.; Zhao, C.; Zhang, F.; Cui, T.; Su, J. Misfire Detection of a Turbocharged Diesel Engine by Using Artificial Neural Networks. *Appl. Therm. Eng.* **2013**, *55*, 26–32. [CrossRef]

30. Jafarian, K.; Mobin, M. Misfire Fault Detection in the Internal Combustion Engine using the Artificial Neural Networks (ANNs). In Proceedings of the Industrial and Systems Engineering Conference, Pittsburgh, PA, USA, 20–23 May 2017.

31. Kim, J.-y.; Lee, T.-h.; Lee, S.-h.; Lee, J.-j.; Lee, W.-k.; Kim, Y.-j.; Park, J.-w. A Study on Deep Learning-Based Fault Diagnosis and Classification for Marine Engine System Auxiliary Equipment. *Processes* **2022**, *10*, 1345. [CrossRef]

32. He, M.; He, D. Deep Learning Based Approach for Bearing Fault Diagnosis. *IEEE Trans. Ind. Appl.* **2017**, *53*, 3057–3065. [CrossRef]

33. Zhang, P.; Gao, W.; Li, Y.; Wang, Y. Misfire Detection of Diesel Engine Based on Convolutional Neural Networks. *Proc. Inst. Mech. Eng. D* **2021**, *235*, 2148–2165. [CrossRef]

34. Naveen Venkatesh, S.; Chakrapani, G.; Babudeva Senapti, S.; Annamalai, K.; Elangovan, M.; Indira, V.; Sugumaran, V.; Mahamuni, V.S. Misfire Detection in Spark Ignition Engine Using Transfer Learning. *Comput. Intell. Neurosci.* **2022**, *2022*, 7606896. [CrossRef]

35. Xu, S.; Lei, J.; Qin, C.; Zhang, Z.; Tao, J.; Liu, C. A Domain-Adversarial Wide-Kernel Convolutional Neural Network for Noisy Domain Adaptive Diesel Engine Misfire Diagnosis. *IEEE Trans. Instrum. Meas.* **2023**, *73*, 3506819. [CrossRef]

36. Wang, X.; Zhang, P.; Gao, W.; Li, Y.; Wang, Y.; Pang, H. Misfire Detection Using Crank Speed and Long Short-Term Memory Recurrent Neural Network. *Energies* **2022**, *15*, 300. [CrossRef]

37. Huang, Y.; Chen, C.-H.; Huang, C.-J. Motor Fault Detection and Feature Extraction Using RNN-Based Variational Autoencoder. *IEEE Access* **2019**, *7*, 139086–139096. [CrossRef]

38. Chui, K.T.; Gupta, B.B.; Vasant, P. A Genetic Algorithm Optimized RNN-LSTM Model for Remaining Useful Life Prediction of Turbofan Engine. *Electronics* **2021**, *10*, 285. [CrossRef]

39. Ravikumar, K.N.; Yadav, A.; Kumar, H.; Gangadharan, K.V.; Narasimhadhan, A.V. Gearbox Fault Diagnosis Based on Multi-Scale Deep Residual Learning and Stacked LSTM Model. *Measurement* **2021**, *186*, 110099. [CrossRef]

40. Zhang, X.; Hua, X.; Zhu, J.; Ma, M. Intelligent Fault Diagnosis of Liquid Rocket Engine via Interpretable LSTM with Multisensory Data. *Sensors* **2023**, *23*, 5636. [CrossRef] [PubMed]

41. Akram, S.; Akhter, G.; Ge, Y.; Azeem, T. ResNet and CWT Fusion: A New Paradigm for Optimized Heterogeneous Thin Reservoir Evaluation. *ACS Omega* **2024**, *9*, 4775–4791. [CrossRef] [PubMed]

42. Mou, J.; Xiong, W. Fault Detection Based on Sliding Window and Multiblock Convolutional Autoencoders. *J. Syst. Simul.* **2024**, *36*, 423–435. [CrossRef]

43. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

44. Sherstinsky, A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Phys. D* **2022**, *404*, 132306. [CrossRef]

45. Goyal, P.; Rani, R.; Singh, K. A Multilayered Framework for Diagnosis and Classification of Alzheimer's Disease Using Transfer Learned AlexNet and LSTM. *Neural Comput. Appl.* **2023**, *36*, 3777–3801. [CrossRef]

46. Zhu, Y.; Su, H.; Tang, S.; Zhang, S.; Zhou, T.; Wang, J. A Novel Fault Diagnosis Method Based on SWT and VGG-LSTM Model for Hydraulic Axial Piston Pump. *J. Mar. Sci. Eng.* **2023**, *11*, 594. [CrossRef]

47. Shang, Q.; Jin, T.; Chen, M. A New Cross-Domain Motor Fault Diagnosis Method Based on Bimodal Inputs. *J. Mar. Sci. Eng.* **2024**, *12*, 1304. [CrossRef]