mathematics

MDPI

Article

Hybrid CNN-BiLSTM-MHSA Model for Accurate Fault Diagnosis of Rotor Motor Bearings

Zizhen Yang [1,2], Wei Li [3], Fang Yuan [3], Haifeng Zhi [4], Min Guo [5], Bo Xin [1,2] and Zhilong Gao [1,2,*]

[1] Key Laboratory of Engine Health Monitoring-Control and Networking (Ministry of Education), Beijing University of Chemical Technology, Beijing 100029, China; 2023200722@buct.edu.cn (Z.Y.); 2024400255@buct.edu.cn (B.X.)

[2] Beijing Key Laboratory of Health Monitoring Control and Fault Self-Recovery for High-End Machinery, Beijing University of Chemical Technology, Beijing 100029, China

[3] China Academy of Aerospace Aerodynamics, Innovation and Application Center, Beijing 100074, China; liweibuct@163.com (W.L.); yuanfang1017@163.com (F.Y.)

[4] National Key Laboratory of Vehicle Power System, China North Engine Research Institute, Tianjin 300405, China; zhihf5615@163.com

[5] Shanxi Diesel Engine Industry Co., Ltd., Datong 037003, China; guomindt@163.com

* Correspondence: buct200931212@163.com

Abstract: Rotor motor fault diagnosis in Unmanned Aerial Vehicles (UAVs) presents significant challenges under variable speeds. Recent advances in deep learning offer promising solutions. To address challenges in extracting spatial, temporal, and hierarchical features from raw vibration signals, a hybrid CNN-BiLSTM-MHSA model is developed. This model leverages Convolutional Neural Networks (CNNs) to identify spatial patterns, a Bidirectional Long Short-Term Memory (BiLSTM) network to capture long- and short-term temporal dependencies, and a Multi-Head Self-Attention (MHSA) mechanism to highlight essential diagnostic features. Experiments on raw rotor motor vibration data preprocessed with Butterworth band-stop filters were conducted under laboratory and real-world conditions. The proposed model achieves 99.33% accuracy in identifying faulty bearings, outperforming traditional models like CNN (93.33%) and LSTM (62.00%) and recent advances including CNN-LSTM (98.87%), the Attention Recurrent Autoencoder hybrid Model (ARAE) (66.00%), Lightweight Time-focused Model Network (LTFM-Net) (96.67%), and Wavelet Denoising CNN-LSTM (WDCNN-LSTM) (96.00%). The model's high accuracy and stability under varying conditions underscore its robustness, making it a reliable solution for rolling bearing fault diagnosis in rotor motors, particularly for dynamic UAV applications.

Keywords: rotor motor; fault diagnosis; convolutional neural network; bidirectional long short-term memory network; multi-head self-attention

MSC: 37M05

check for updates

Academic Editor: António Lopes

Received: 18 December 2024
Revised: 14 January 2025
Accepted: 17 January 2025
Published: 21 January 2025

Citation: Yang, Z.; Li, W.; Yuan, F.; Zhi, H.; Guo, M.; Xin, B.; Gao, Z. Hybrid CNN-BiLSTM-MHSA Model for Accurate Fault Diagnosis of Rotor Motor Bearings. *Mathematics* **2025**, *13*, 334. https://doi.org/10.3390/math13030334

1. Introduction

The rotor motor in UAVs is critical to ensuring the reliability and safety of flight operations, as its performance directly impacts the efficiency of each UAV's power system. Health monitoring is, thus, essential for the early detection of anomalies, reducing the risk of catastrophic failures, particularly in demanding environments. UAV rotor systems commonly utilize brushless outer rotor motors, with speed regulation managed by Electronic Speed Controllers (ESCs). In this configuration, the rotor—integrated with a permanent magnet—rotates with the motor casing and output shaft, while the stator houses the winding coils. Research on rotor motor fault diagnosis holds significant theoretical and practical importance, as it directly influences flight stability, operational reliability, and the broader application potential of UAVs.

However, the highly integrated and compact design of UAVs presents significant challenges for motor health monitoring. Space constraints limit the integration of multiple diagnostic measures, while the size and weight of traditional diagnostic devices often conflict with the lightweight and miniaturized design requirements of UAVs. Furthermore, sensor placement and data acquisition become increasingly complex in such confined layouts, making fault detection more difficult. Mechanical failures in UAV motors typically involve bearing, rotor, and stator faults, with bearing failures accounting for approximately 40% of all motor issues, making them the most common type of failure [1]. Effective bearing fault diagnosis relies on identifying fault characteristics, signal processing, and feature extraction techniques, all of which are crucial for ensuring the reliable operation of mechanical systems.

To address these challenges, recent advancements in Artificial Intelligence (AI) have led to the integration of sophisticated fault diagnosis models, such as neural networks and Support Vector Machines (SVMs), into bearing and rotor fault diagnostics. These models effectively automate fault identification by learning from large datasets of fault-related information, substantially increasing diagnostic accuracy and efficiency [2–4]. Machine learning, particularly Deep Learning (DL), has significantly improved fault diagnosis in bearing systems by enabling more accurate predictions and better processing of complex, nonlinear data. This development represents a paradigm shift in motor-bearing fault diagnosis, moving from traditional physical mechanism analysis toward AI-integrated diagnostic systems that utilize vibration signal analysis and DL techniques for enhanced fault feature extraction [5–7].

Several studies highlight these advancements. An et al. developed an Efficient Convolutional Neural Network (ECNN) for edge computing, enabling real-time fault diagnosis and dynamic control of electric motors [8]. Evangeline et al. proposed a hybrid deep residual-based neural network combined with multi-SVM for the diagnosis of mechanical and electrical faults in synchronous motors [9]. Fan et al. proposed a CNN model with adaptive batch normalization for diagnosis of rotor-bearing faults using gray texture images derived from vibration signals, achieving improved fault identification [10]. Additionally, Zhang et al. introduced a novel approach for diagnosing Electro-Hydraulic Steer-By-Wire (EH-SBW) systems, employing a 1DCNN-LSTM model with attention mechanisms and transfer learning, which proved highly accurate in identifying fault types and severities, even with limited data [11]. Kim et al. developed a machine learning-based approach for diagnosing faults in induction motors, with CNN and SVM models achieving superior accuracy, while XGBoost exhibited the highest computational efficiency for real-time applications [12]. Yang et al. introduced the LTFM-Net framework, which features an LTFM with an innovative Weighted Diminish Recurrent Unit (WDRU), demonstrating high accuracy, robustness, and interpretability under complex conditions [13]. Izaz Raouf et al. proposed an attention-guided Feature Aggregation Network (FAN) for detecting faults in industrial robot servo motor bearings, effectively supporting predictive maintenance [14]. Liu et al. proposed the LOODG framework, a three-stage causal feature-learning method for bearing fault diagnosis. It excels in generalizing across diverse operating conditions without needing domain labels [15]. Ma et al. introduced an adaptive-embedding Flexible Tensor Singular Spectrum Decomposition (FTSSD) method, utilizing the Trajectory Dimension Ratio (TDR) index for optimal embedding dimension selection. This approach was applied to multichannel signal fusion for fault diagnosis, demonstrating enhanced accuracy and robustness in extracting fault features compared to traditional methods such as Tensor Robust Principal Component Analysis (TRPCA) [16]. On the basis of this research, Huang et al. proposed a novel tensor decomposition method, the first-Kind Flexible Tensor Singular Value Decomposition (1K-FTSVD), which avoids tensor flattening and offers stable decompositions. They utilized this method to develop 1K-FTSSD for multi-channel data fusion, demonstrating its effectiveness through simulations and experimental validations, which showed superior performance compared to existing techniques [17]. Xu et al. proposed the Graph-Embedded Low-Rank Tensor Learning Machine (GE-LRTLM) for semi-supervised, multi-sensor data fusion in fault diagnosis, effectively handling tensor data to preserve multi-dimensional structure. By incorporating tensor nuclear norm and manifold regularization, the method achieved high accuracy of up to 97%, even with scarce labeled data, offering a robust framework for industrial applications [18].

Looking forward, continued integration of sensing, signal processing, and artificial intelligence technologies is expected to enhance both the accuracy and real-time capabilities of fault diagnosis systems, ensuring the stable operation of motors and advancing the reliability of industrial processes [19–22].

Bearing fault diagnosis methods share a common goal of ensuring reliable system operation by detecting anomalies through feature extraction and signal processing. Traditional techniques, such as vibration analysis and spectral methods, provide practical and accessible solutions. However, their effectiveness is often limited under dynamic conditions, such as varying speeds and loads, which are common in real-world scenarios. To address these limitations, modern approaches leveraging artificial intelligence have demonstrated significant advancements by automating feature extraction and improving diagnostic accuracy. These AI-driven methods enhance adaptability and precision but also face challenges, particularly in resource-constrained environments, such as UAV applications where computational power and labeled data are limited.

Building on this progress, current data-driven fault diagnosis methods for rotor motors focus on analyzing specific fault types or operating conditions, often employing standalone models such as CNN or LSTM. While these methods achieve reasonable accuracy, their ability to generalize under diverse operational conditions remains limited. Many existing approaches struggle to capture and integrate spatial, temporal, and long-range dependencies within vibration signals, which is essential for distinguishing between subtle differences in fault severity and type. These challenges highlight the need for more advanced models capable of balancing accuracy, generalization, and efficiency, particularly in lightweight and dynamic application scenarios like UAVs.

In UAV applications, where constraints on weight, size, and power consumption are critical, our model is designed to be both lightweight and efficient. Experimental results show that the model can be deployed on UAV platforms equipped with specific hardware, ensuring real-time bearing fault detection with minimal power consumption. This makes the model highly suitable for onboard monitoring systems integrated in UAVs used in industrial inspections and other practical applications.

The main contributions of this paper are summarized as follows:

1. A CNN-BiLSTM-MHSA-based fault diagnosis model was developed for rotor motors, combining the strengths of CNN, BiLSTM, and MHSA to extract spatial, temporal, and attention-based features for accurate fault detection.
2. Advanced signal processing techniques, including Fast Fourier Transform (FFT) and Butterworth band-stop filters, were applied to enhance the quality of vibration data, improving feature extraction and the robustness of the fault diagnosis model under varying operational conditions.

3. A comprehensive diagnostic framework was proposed to classify different bearing fault severities, achieving a high accuracy of 99.33% on a test set, significantly outperforming traditional CNN and LSTM models in fault classification tasks.
4. The model's performance was validated using experimental data from a rotor motor fault simulation test bench, demonstrating its effectiveness and reliability in real-world applications.

2. Theoretical Basis

2.1. CNN

CNNs are deep learning models specifically designed to automatically extract features from data through convolutional operations. A schematic representation of the CNN architecture is provided in Figure 1.

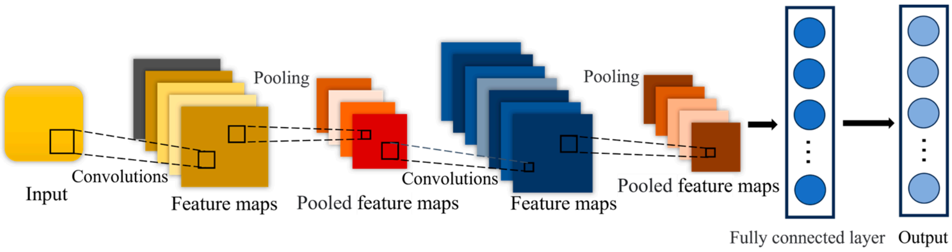


Figure 1. Schematic diagram of a convolutional neural network.

The key components of a CNN model include the following:

1. The convolutional layer applies a filter (or convolution kernel) to the input data, performing weighted summation over local regions to extract relevant features. This operation is defined by Equation (1):

$$y(i,j,k) = \sum_{c=1}^{C} \sum_{m=1}^{F_w} \sum_{n=1}^{F_h} x(i+m-1, j+n-1, c) \cdot w(m,n,c,k) + b(k) \tag{1}$$

where $y(i,j,k)$ is the position (i,j) and k channel of the output feature map; x is the input feature map; w is the convolution kernel; b is the bias; F_w and F_h are the width and height of the convolution kernel, respectively; m and n are the row and column indexes of the convolution kernel, respectively; and C is the number of channels of the input feature map.

2. The activation function introduces nonlinearity into the network, enhancing its ability to model complex patterns. The Rectified Linear Unit (ReLU) is one of the most commonly used activation functions. It is defined as shown in Equation (2):

$$f(x) = max(0, x) \tag{2}$$

3. The pooling layer reduces the spatial dimensions of the features and reduces the amount of computation. Common pooling operations are maximum pooling and average pooling. Their calculation formulas are expressed as follows in Equation (3):

$$y(i,j,k) = \max_{(m,n)\in R} x(i+m, j+n, k) \tag{3}$$

where $y(i,j,k)$ is the output value of position (i,j) after pooling operation and channel k, $x(i+m, j+n, k)$ is the pixel value of the position $(i+m, j+n)$ in the output feature map and channel k, x is the input feature map, R is the range of the pooled window, m and n represent the current position offset in the pooled window, and k is the channel index of the feature map.

4. Located at the end of the network, the fully connected layer maps the features extracted from the convolutional and pooling layers to the final output classification. The formula of the fully connected layer is expressed as follows in Equation (4):

$$y = W \cdot x + b \tag{4}$$

where y is the output vector, x is the input eigenvector, W is the weight matrix, and b is the bias.

CNNs have shown significant effectiveness in bearing fault diagnosis due to their ability to automatically learn complex patterns in bearing vibration signals. This allows for high-precision fault identification and provides a robust technical foundation for intelligent predictive maintenance in mechanical systems [23,24].

2.2. BiLSTM

LSTM networks are specialized Recurrent Neural Networks (RNNs) designed to capture long-term dependencies, making them effective for sequential data analysis. They can process vibration signals or sensor data to assess bearing health or predict failure, as illustrated in Figure 2.

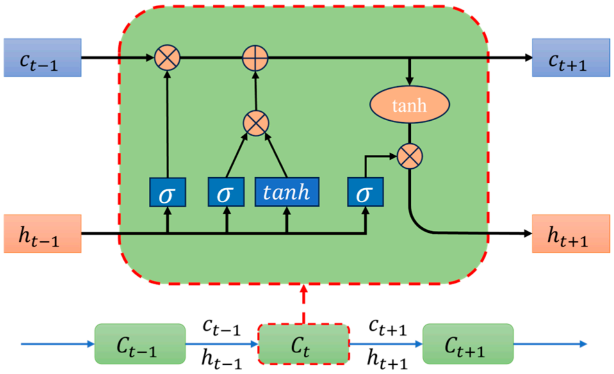


Figure 2. Schematic diagram of the LSTM network.

The LSTM network consists of the following parts [25,26]:

1. Cell state: The core component of an LSTM network is the cell state, responsible for storing and transferring information, enabling the network to retain long-term dependencies and excel in processing sequential data.
2. Forget gate: The forget gate determines which information to discard from the cell state. Using a sigmoid activation function, it evaluates each cell's state value to decide how much to retain, allowing the network to flexibly adapt its memory. The forget gate operation is defined by Equation (5):

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{5}$$

where f_t is the output of the forgetting gate, W_f is the weight matrix of the oblivion gate, h_{t-1} is the hidden state of the previous time step, x_t is the input at the current time, and b_f is the bias term of the oblivion gate.

3. Input gate: The input gate contains a sigmoid layer that selects values to update the cell state and a tanh layer that generates candidate values for addition. These functions are defined by Equations (6) and (7):

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{6}$$

$$\overset{\sim}{\mathrm{C}}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_{\mathrm{C}}) \tag{7}$$

where i_t is the output of the input gate; W_i is the weight matrix of the input gate; $\widetilde{C}_t$ is the new candidate message; W_c is the weight matrix of the new candidate message; and b_i and b_C are the bias terms of the input gate and the new candidate information, respectively.

4. Module status update: The process of updating the cell state involves multiplying the output of the forget gate by the original cell state, then adding the output of the input gate. This operation maintains the validity and continuity of the cell state. The formula for updating a cell is expressed as follows in Equation (8):

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C}_t \tag{8}$$

where C_t is the cell state at the current moment, f_t is the output of the oblivion gate, $\odot$ refers to the elemental multiplication operation, C_{t-1} is the cell state at the previous time step, i_t is the output of the input gate, and $\widetilde{C}_t$ is the candidate cell state generated by the input gate.

5. Hidden state: The hidden state is the output of the LSTM unit, containing information from the current time step, which is passed to the next time step to ensure coherence and consistency of the data.
6. Output gate: The output gate determines the value of the next hidden state, which is the actual output of the LSTM network. The output gate is adjusted so that the network output reflects the key information of the current time step. The formula for the output gate is expressed as follows in Equations (9) and (10).

$$O_t = \sigma(W_O \cdot [h_{t-1}, x_t] + b_O) \tag{9}$$

$$h_t = O_t \odot tanh(C_t) \tag{10}$$

where O_t is the output of the output gate, W_O is the weight matrix of the output gate, h_t is the hidden state of the current time step, and b_O is the bias term of the output gate.

LSTM networks are particularly effective at capturing and preserving long-term dependencies in sequential data due to their memory cells and gate mechanisms. However, LSTMs perform unidirectional propagation, which limits their ability to capture dependencies in both directions. To address this, BiLSTM was developed.

BiLSTM extends the LSTM architecture by incorporating both forward and backward propagation, enabling the model to capture dependencies in both directions of sequential data. This bidirectional mechanism shown in Figure 3 enhances feature extraction from temporal sequences, making BiLSTM a robust solution for tasks that require complex, time-sensitive pattern recognition [27]. The dual propagation in BiLSTM greatly improves the accuracy and reliability of model outputs, particularly for applications requiring comprehensive temporal feature extraction in sequence analysis.

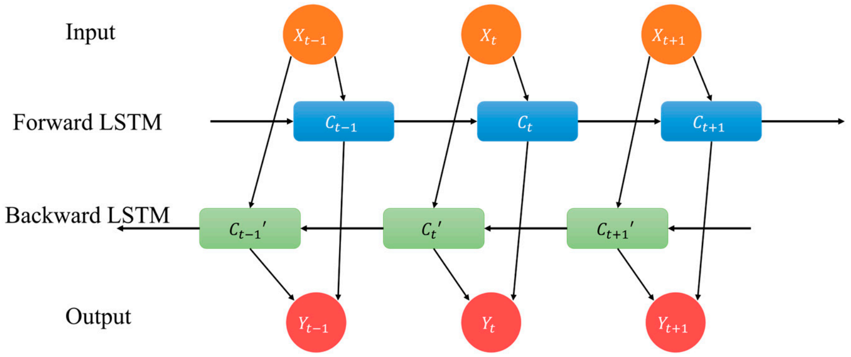


Figure 3. Schematic diagram of BiLSTM network.

The cycle calculations in both directions are shown in Equations (11) and (12):

$$\overrightarrow{h_t} = LSTM_{forward}\left(x_t, \overrightarrow{h_{t-1}}, \overrightarrow{c_{t-1}}\right) \tag{11}$$

$$\overleftarrow{h_t} = LSTM_{backward}\left(x_t, \overleftarrow{h_{t+1}}, \overleftarrow{c_{t+1}}\right) \tag{12}$$

where x_t is the input vector of the current time step (t); h_{t-1} and h_t are the hidden states of the previous time step and the current time step, respectively; and c_{t-1} and c_t are the cell states of the previous time step and the current time step, respectively.

The final output hidden state (H_t) is a splice in both directions:

$$H_t = \left[\overrightarrow{h_t}; \overleftarrow{h_t}\right] \tag{13}$$

2.3. MHSA

MHSA is a deep learning technique in deep learning models that is the state of the art in the field of Natural Language Processing (NLP) that allows models to pay more attention to relevant features and prevents overfitting. The traditional fault diagnosis workflow is simplified in the model.

Attention is a mechanism that allows a model to capture dependencies within a sequence by focusing on different locations within the sequence as it is processed. The multi-head self-attention mechanism replicates this process several times, with each "head" learning a different representation of the input data. Finally, the output vectors from all heads are stitched together and processed through a linear layer to obtain the final output. The model can learn different relationships between different locations, which makes it more flexible when dealing with long-distance dependencies [28,29]. The output matrix of the attention layer in the MHSA mechanism (query (Q), key (K), and value (V)) is shown in Equation (14):

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{14}$$

where d_k is the feature dimension of each key, which is used for weight scaling and normalized to the interval of [0,1] by softmax. The MHSA used in this paper is four attention heads, and the calculation of each attention head is shown in Equation (15):

$$head_i = Attention\left(QW_i{}^Q, KW_i{}^K, VW_i{}^V\right) \tag{15}$$

where $W_i{}^Q$, $W_i{}^K$, and $W_i{}^V$ denote the weight matrices of Q, K, and V, respectively, and $head_i$ is the ith head in the MHSA.

The MHSA mechanism combines multiple attention heads, each with its linear transformation matrix. The output is shown in Equation (16):

$$MHSA(x) = Concat(head_1, \ldots, head_4)W^O \tag{16}$$

Linear represents the linear mapping operation, W^O is the weight of linear mapping, and Concat is the splicing operation. The model structure of MHSA is shown in Figure 4.

An attention mechanism is therefore introduced to filter the features during their extraction to improve the correctness of fault diagnosis. In fault identification, certain anomalies may involve multiple locations in the sequence rather than just locally specific parts [30,31]. By introducing MHSA, the model can focus on different parts of the sequence at the same time to better capture the global dependencies, improving the perception of the overall sequence pattern. The MHSA used in this paper is a four-attention head.

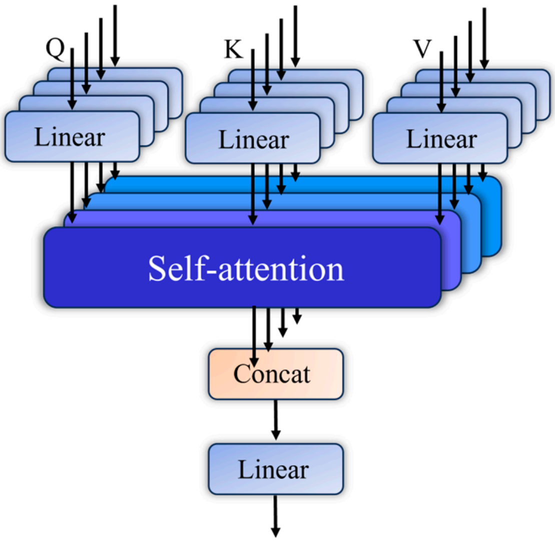


Figure 4. The structure of the MHSA mechanism.

2.4. The Proposed Diagnostic Model

Before feeding the vibration signals into the diagnostic model, a preprocessing step is applied to enhance the quality of the raw data. The preprocessing includes the use of a Butterworth band-stop filter to eliminate noise and interference within specific frequency ranges, ensuring that the signals are clean and free from distortions. Additionally, FFT is employed to show the frequency-domain characteristics of the vibration signals, providing valuable insights into fault-related features. This step ensures that the input data retain only the most relevant information, which is crucial for accurate fault diagnosis.

The proposed fault diagnosis model adopts a hybrid architecture that integrates CNN, BiLSTM, and MHSA. This design addresses the challenges of diagnosing UAV rotor motor bearing faults by comprehensively capturing spatial, temporal, and contextual features from the input data.

In this model, the CNN module extracts spatial features from the preprocessed vibration signals, identifying local patterns and anomalies that are indicative of faults. The convolutional and pooling operations reduce data dimensionality while retaining essential features, ensuring efficiency and preserving critical information for further analysis.

The BiLSTM module processes the spatial features to capture temporal dependencies in the sequential vibration data. Its bidirectional structure enables the integration of information from both past and future time steps, which is essential for recognizing fault signatures distributed over time. This temporal modeling enhances the model's ability to identify complex and subtle patterns in the data.

To refine the extracted features, the MHSA mechanism emphasizes the most relevant information by assigning higher attention weights to critical data points. This mechanism allows the model to focus on key diagnostic features, ensuring that subtle yet significant fault characteristics are highlighted. The multi-head structure further enhances robustness by enabling the model to learn diverse feature representations, improving adaptability to different fault scenarios.

The CNN-BiLSTM-MHSA model achieves a comprehensive and hierarchical representation of the input data. The final features are passed through a fully connected layer, which maps them to specific fault categories. This integrated approach not only ensures high diagnostic accuracy but also maintains robustness across diverse operating conditions, including variable rotational speeds.

Despite its complexity, the model is computationally efficient. The combination of CNN, BiLSTM, and MHSA minimizes redundancy in feature extraction and focuses computational resources on the most informative aspects of the data. This efficiency makes it suitable for real-time fault diagnosis in UAV systems, meeting the constraints of limited onboard computational power and the demand for rapid decision making.

3. Experimental Verification

3.1. Experimental Design and Data Acquisition

3.1.1. Experimental Overview

The experiment was conducted under controlled conditions with the rotor motor running without load. Deep-groove ball bearings with various types and degrees of failure were artificially induced through wire-cut machining. While this method is effective for generating clear and reproducible fault scenarios, it is important to note that real-world bearing faults in UAV applications may develop over time due to gradual wear, fatigue, and environmental influences rather than the abrupt fault simulations used in this study. Thus, while these fault types serve as useful test cases, further studies will need to incorporate more realistic fault development processes.

As depicted in Figure 5, the vibration response of the rotor, both under normal conditions and with implanted faulty bearings, was simulated across different rotational speeds.

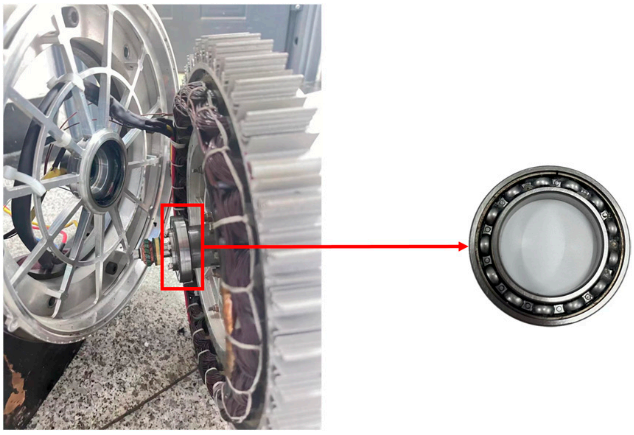


Figure 5. Schematic diagram of the rotor motor to replace the faulty bearing.

During the experiment, the rotational speed of the rotor motor was gradually increased from 0 to 1600 r/min to better emulate real-world operational conditions. Five distinct speeds—200 r/min, 500 r/min, 800 r/min, 1200 r/min, and 1600 r/min—were tested, allowing for controlled data collection across a range of operating conditions. This variable-speed approach is designed to emulate real scenarios, capturing the rotor's vibrational behavior under both normal and faulty conditions.

Testing at multiple speeds enables the collection of a comprehensive dataset encompassing both normal and fault-specific conditions, forming a robust foundation for fault diagnosis and improving the reliability and accuracy of the diagnostic model.

As illustrated in Figure 6, the rotor motor was firmly affixed to an experimental table, with vibration acceleration sensors positioned along the vertical (x-axis) and horizontal (y-axis) axes. This arrangement ensures comprehensive monitoring of the rotor's vibrational dynamics, facilitating accurate data acquisition.

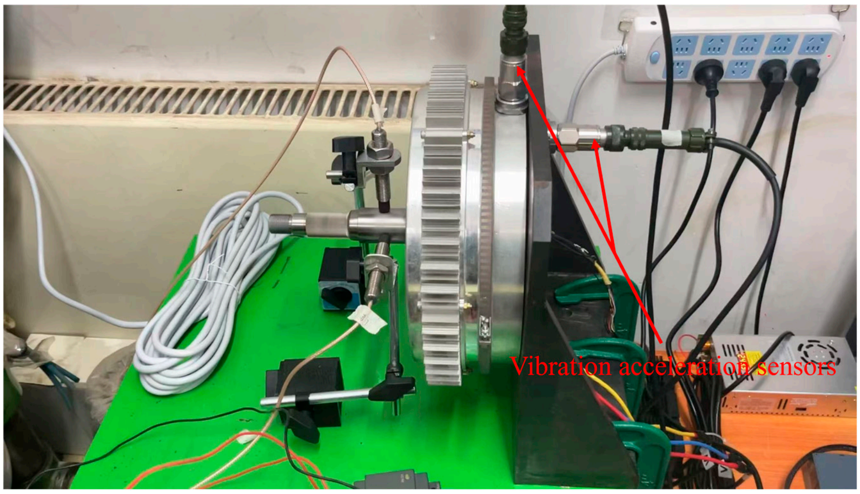


Figure 6. Layout of rotor motor entity and sensors of a certain model.

3.1.2. Laboratory Equipment and Test Instruments

The main experimental equipment and test instruments are listed in Table 1. The testbench system includes both hardware and software configurations. The hardware consists of sensors, a host computer, a data acquisition processor, and an industrial power supply. The software includes data acquisition software, data management software, middleware communication software, alarm management software, and a standalone version of the client software.

Table 1. Main equipment used in the fault simulation experiment.

Equipment	Model Number	Parameters	Brand
Acceleration sensors	BH5011	Sensitivity: 10 mV/g Amplitude range: ±500 g Frequency range: 0–13 kHz	BH
Data acquisition processor	AC5000	Resolution: 16-bit Enter path: 32 Maximum sampling rate: 102.4 KSPS	BH
Deep-groove ball bearings	6907	Rolling body diameter: 5.2 mm Section diameter: 45 mm Inside diameter: 35 mm External diameter: 55 mm Thickness: 10 mm Number of rolling bodies: 13	NSK

3.1.3. Bearing Failure Implantation

In this experiment, fault simulations were conducted on NSK6907 rolling bearings used in a specific rotor motor. Faults were introduced in the inner and outer rings through wire-cut machining, creating both mild and severe defects to simulate real bearing failures [32,33]. The sealing rings were removed, and precise cuts were made on the bearing rings, as outlined in Table 2. The wire-cut machining setup and the resulting faulty bearings are shown in Figure 7.

Table 2. Bearing machining allowance table.

Fault Severity / Fault Type	Minor (Width mm × Depth mm)	Severe (Width mm × Depth mm)
Outer ring failure	0.5 × 0.5	0.5 × 1
Inner ring failure	0.5 × 0.5	0.5 × 1

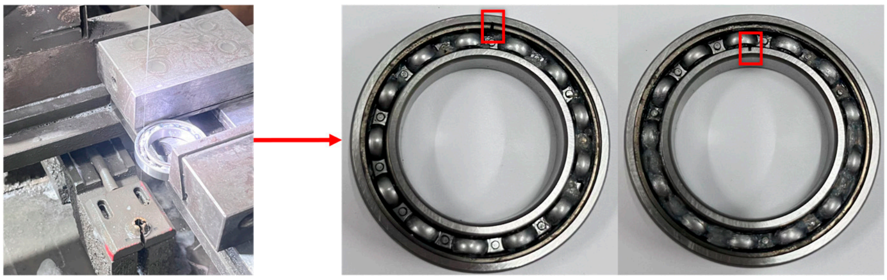


Figure 7. Wire-cut implantation fault: outer ring with a severely failed bearing and inner ring with a severely failed bearing.

3.1.4. Test Procedure

The online monitoring and diagnostic systems were used to record the vibration acceleration of the motor rotor at various rotational speeds. After completing the experiments, the motor was turned off, and the systems, along with the control interface, were shut down. Ensuring the integrity and accuracy of the recorded data was essential (Figure 8).

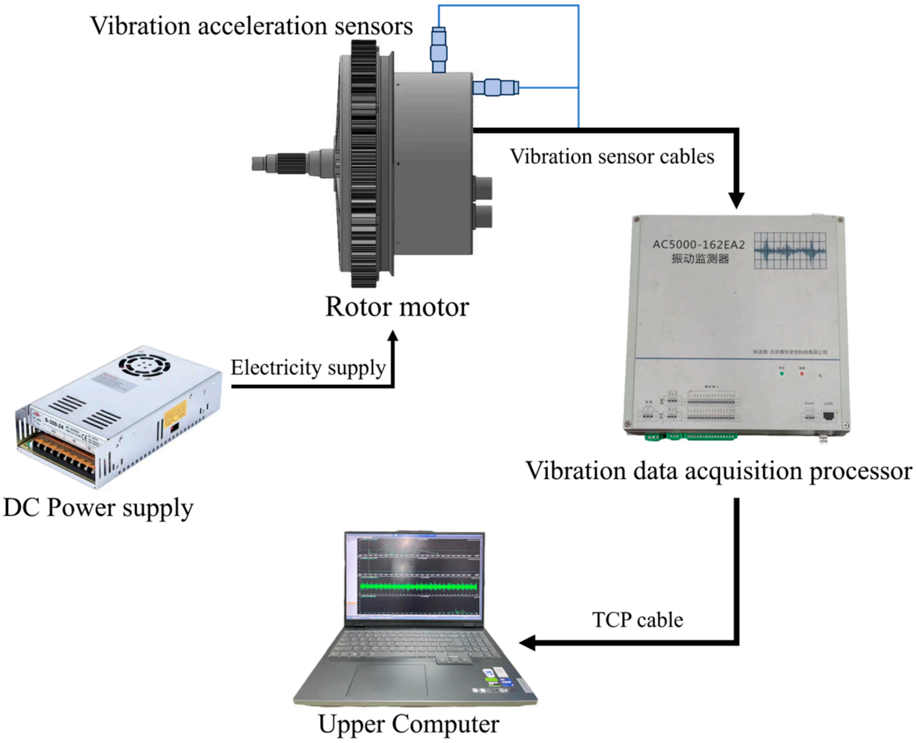


Figure 8. Diagram of the test architecture for rotor motor failure simulation.

3.2. *The Preprocessing Process of Raw Collected Data*

Signal processing of vibration data is crucial for machinery fault diagnosis, as effective preprocessing significantly enhances the accuracy of feature extraction and subsequent fault diagnosis. In this study, a band-stop filter is utilized to eliminate noise and interference within specific frequency ranges, and FFT is employed to assess the frequency-domain characteristics of the vibration signals. The raw data comprise the vibration acceleration of the rotor motor collected at various rotational speeds, with a sampling frequency of 25,600 Hz. The vibration signals are analyzed in the frequency domain to extract relevant features.

3.2.1. Band-Stop Filters

Filtering is a critical step in signal processing, aimed at removing or suppressing noise or interference within specified frequency ranges. A Butterworth band-stop filter is employed for this purpose. The Butterworth filter was chosen due to its relatively smooth frequency response and minimal ripple, making it suitable for applications requiring signal smoothing. The design is based on its transfer function, which is shown in Equation (17):

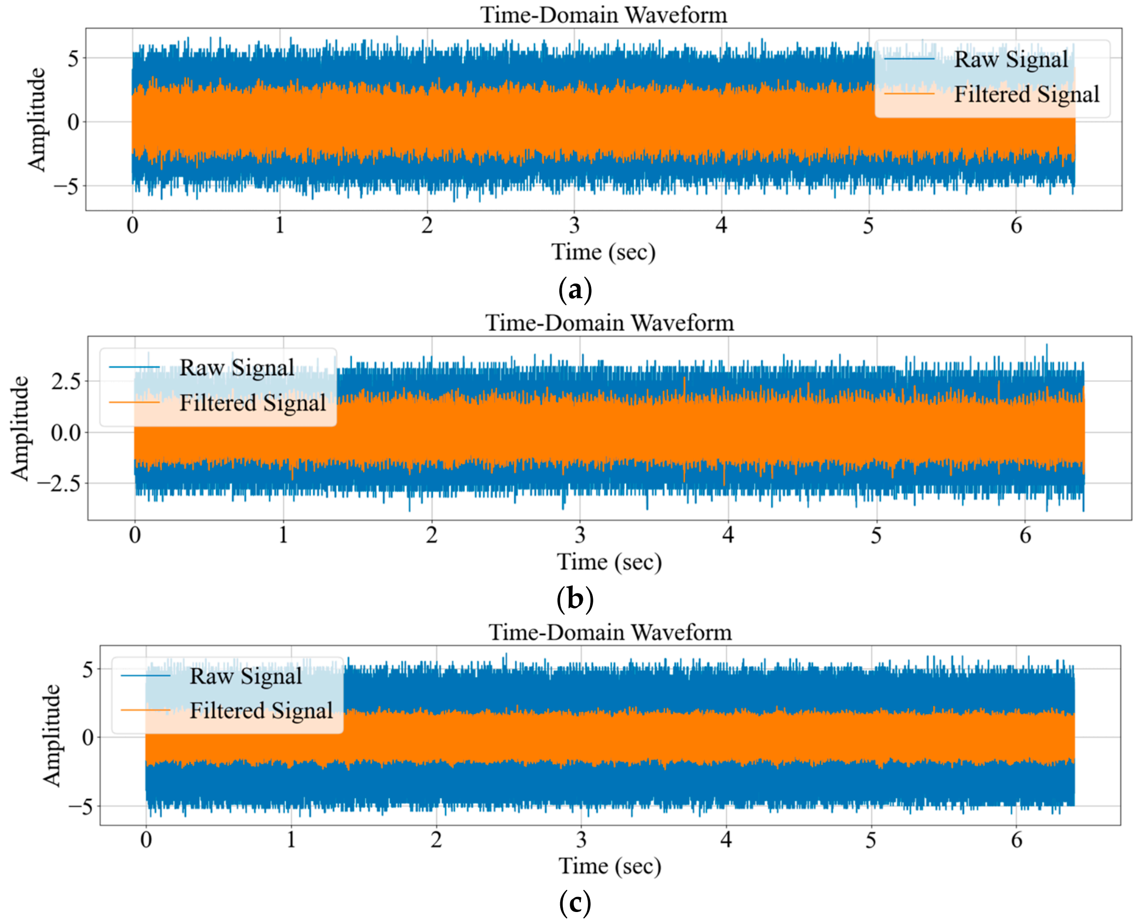
$$H(s) = \frac{1}{1 + \left(\frac{s}{\omega_c}\right)^{2N}} \tag{17}$$

where s is the complex frequency-domain variable, ω_c is the cutoff frequency, and N is the order of the filter.

The following parameters are set for the filter:

1. The filter order is set to 4. A higher filter order yields a steeper frequency response, enhancing the effectiveness of the filter in attenuating specific frequency bands.
2. The band-stop frequency range is defined from 5000 Hz to 10,000 Hz, targeting a range where noise and interference is notably significant and, thus, requires suppression.

The filtering process is executed using the scipy.signal.filtfilt function, which provides bidirectional filtering. This approach helps avoid phase shifts in the signal, ensuring that the phase of the filtered signal remains consistent—a critical aspect for time-domain signal analysis. The results of vibration signal filtering from the experimental data are presented in Figures 9 and 10. The preprocessed data are then saved for subsequent bearing diagnosis.

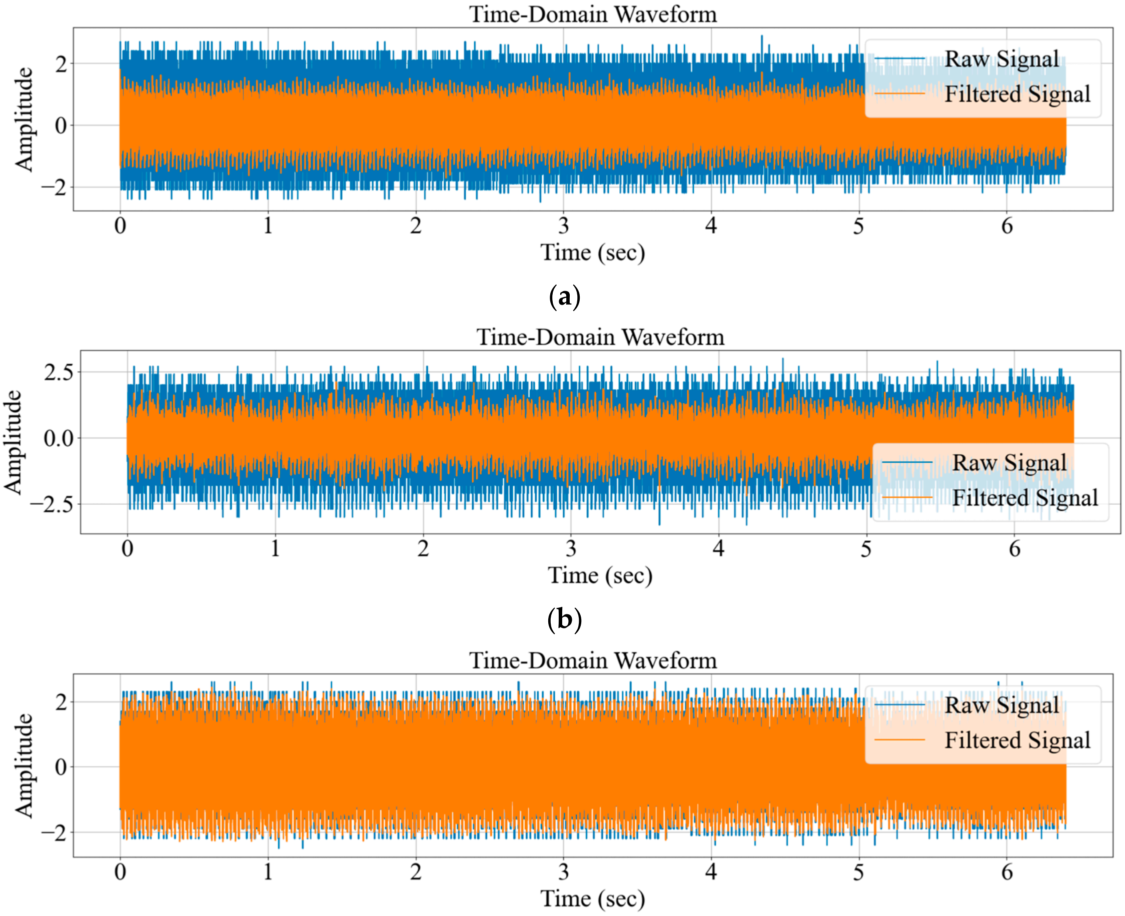


(**a**)

(**b**)

(**c**)

Figure 9. Time-domain diagram of vibration acceleration in the x direction after band-stop filtering. (**a**) Time-domain waveform of inner ring failure; (**b**) time-domain waveform of outer ring failure; (**c**) time-domain waveform of normalcy.



(**a**)

(**b**)

(**c**)

Figure 10. Time-domain diagram of vibration acceleration in the y direction after band-stop filtering. (**a**) Time-domain waveform of inner ring failure; (**b**) time-domain waveform of outer ring failure; (**c**) time-domain waveform of normalcy.

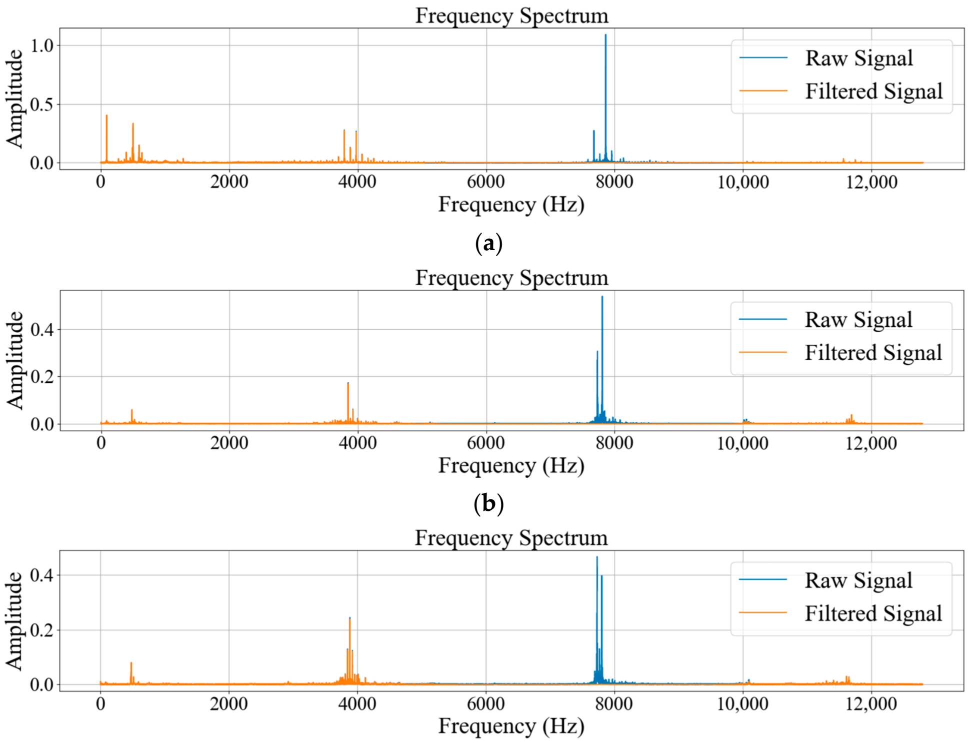
3.2.2. FFT

FFT is an efficient algorithm for Discrete Fourier Transform (DFT) that utilizes the symmetry and periodicity of the signal to drastically reduce the amount of computation. The basic formula of DFT is shown in Equation (18):

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot e^{-j\frac{2\pi}{N}kn} \tag{18}$$

where $x(n)$ is the time-domain discrete signal, $X(k)$ is the time-domain discrete signal, and N is the number of sampling points of the signal.

To eliminate the zero-frequency component from the signal, the first step involves applying a de-mean operation. Following this, the FFT computes the frequency components of the signal, retaining only the positive spectral part of the frequency vector. The data processing result is subsequently obtained by taking the absolute value of the FFT coefficients and normalizing the results [34,35]. The outcomes of the FFT applied to the experimental vibration signals are illustrated in Figures 11 and 12.

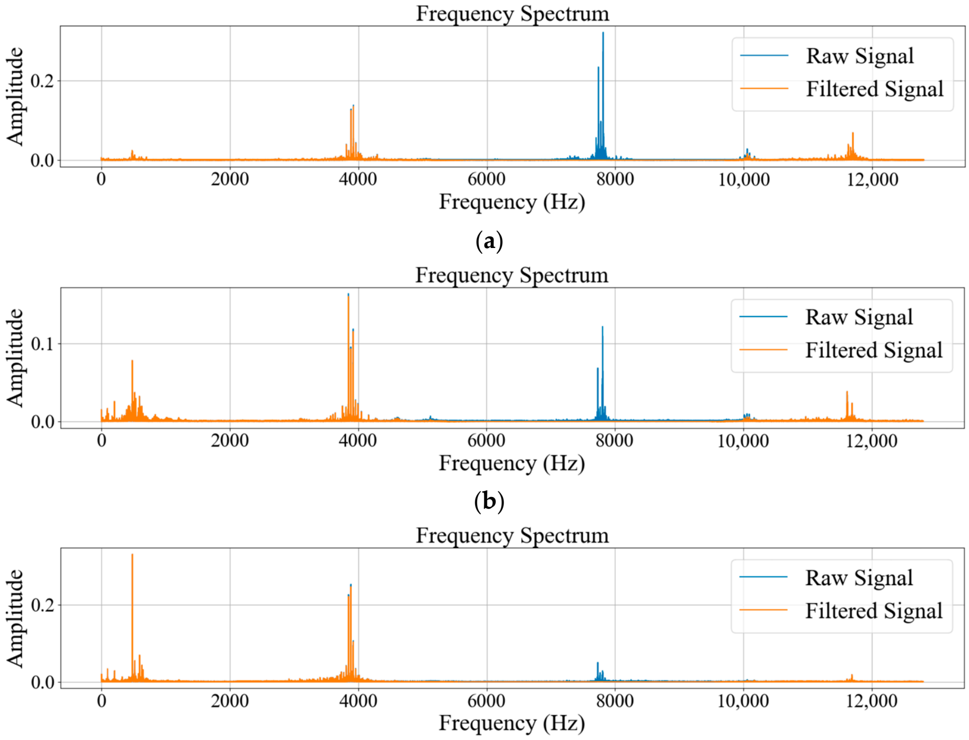


(a)

(b)

(c)

Figure 11. FFT frequency-domain plot of vibration acceleration in the x direction. (**a**) Frequency spectrum diagram of inner ring failure; (**b**) frequency spectrum diagram of outer ring failure; (**c**) frequency spectrum diagram of normalcy.



(a)

(b)

(c)

Figure 12. FFT frequency-domain plot of y-direction vibration acceleration. (**a**) Frequency-domain diagram of inner ring failure; (**b**) frequency spectrum diagram of outer ring failure; (**c**) frequency spectrum diagram of normalcy.

3.3. Model Design and Construction

The CNN-BiLSTM-MHSA model was implemented in PyTorch (version 2.4.1) under Python 3.12. Training and testing were conducted on a Windows 11 system equipped with an Intel Core i9-14900HX CPU and an NVIDIA GeForce RTX 4060 Laptop GPU. The model components were implemented using correlation functions, and the layers were assembled sequentially according to the specified network architecture.

The CNN extracts spatial features through convolutional layers, batch normalization, ReLU activation, pooling, and dropout layers. The BiLSTM captures sequential dependencies by processing features bidirectionally, while the MHSA mechanism highlights critical features for classification. Finally, a fully connected layer maps the features to the fault categories, and a softmax layer provides the classification output. Equation (19) illustrates the process of mapping fault categories from the fully connected layer to the final output.

$$Output = Softmax\left(H_{final}W_{fc} + b_{fc}\right) \tag{19}$$

where H_{final} is the feature input of the last layer, i.e., the feature vector from the MHSA mechanism; W_{fc} is the weight matrix of the fully connected layer; and b_{fc} is the bias term of the fully connected layer.

After building the network model, the training options, including the optimization algorithm, number of iterations, batch size, and validation data, are defined. The prepared training data are then fed into the model, which is trained based on these configurations. Upon completing the training process, the model's performance on the training data and its ability to generalize to unseen samples are evaluated using a separate validation set, ensuring robustness and accuracy.

The detailed network parameter settings and hyperparameter configurations are summarized in Tables 3 and 4, respectively. Figure 13 illustrates the rolling bearing diagnosis process implemented using the CNN-BiLSTM-MHSA model.

Table 3. Detailed parameter settings for network structure.

Layer Name	Input Size	Kernel Size	Stride	Number of Output Channels	Output Size
Conv1	(16, 1, 10,000, 1)	(3, 1)	(1, 1)	64	(16, 64, 9998, 1)
Pool1	(16, 64, 9998, 1)	(2, 1)	(2, 1)	—	(16, 64, 4999, 1)
Conv2	(16, 64, 4999, 1)	(3, 1)	(1, 1)	128	(16, 128, 4997, 1)
Pool2	(16, 128, 4997, 1)	(2, 1)	(2, 1)	—	(16, 128, 2498, 1)
Conv3	(16, 128, 2498, 1)	(3, 1)	(1, 1)	256	(16, 256, 2496, 1)
Pool3	(16, 256, 2496, 1)	(2, 1)	(2, 1)	—	(16, 256, 1248, 1)
Flatten	(16, 256, 1248, 1)	—	—	—	(16, 319, 488)
Dropout	(16, 319, 488)	—	—	—	(16, 319, 488)
BiLSTM	(16, 1248, 256)	—	—	256	(16, 1248, 256)
MHSA	(16, 1248, 256)	—	—	256	(16, 1248, 256)
FC	(16, 256)	—	—	5	(16, 5)

Table 4. Network training hyperparameter configuration table.

Parameter Name	Parameter Value	Description
Learning Rate	0.0005	Used to adjust the update magnitude of the model parameters.
Batch Size	16	The number of samples used in each training iteration.
Number of Epochs	50	Total training rounds (may end early due to the early stopping mechanism).
Patience	5	Early stopping mechanism parameter that stops training if there is no improvement in 5 consecutive epoch losses.
Loss Function	Cross-entropy Loss	Loss functions for multi-categorization problems.
Optimizer	Adam	Optimizers with adaptive learning capabilities.
Learning Rate Scheduler	Optimizer; step size = 10; γ = 0.1	The learning rate is updated every 10 epochs with a decay factor of 0.1.
Dropout	50%	The dropout ratio is applied in the model to minimize overfitting.

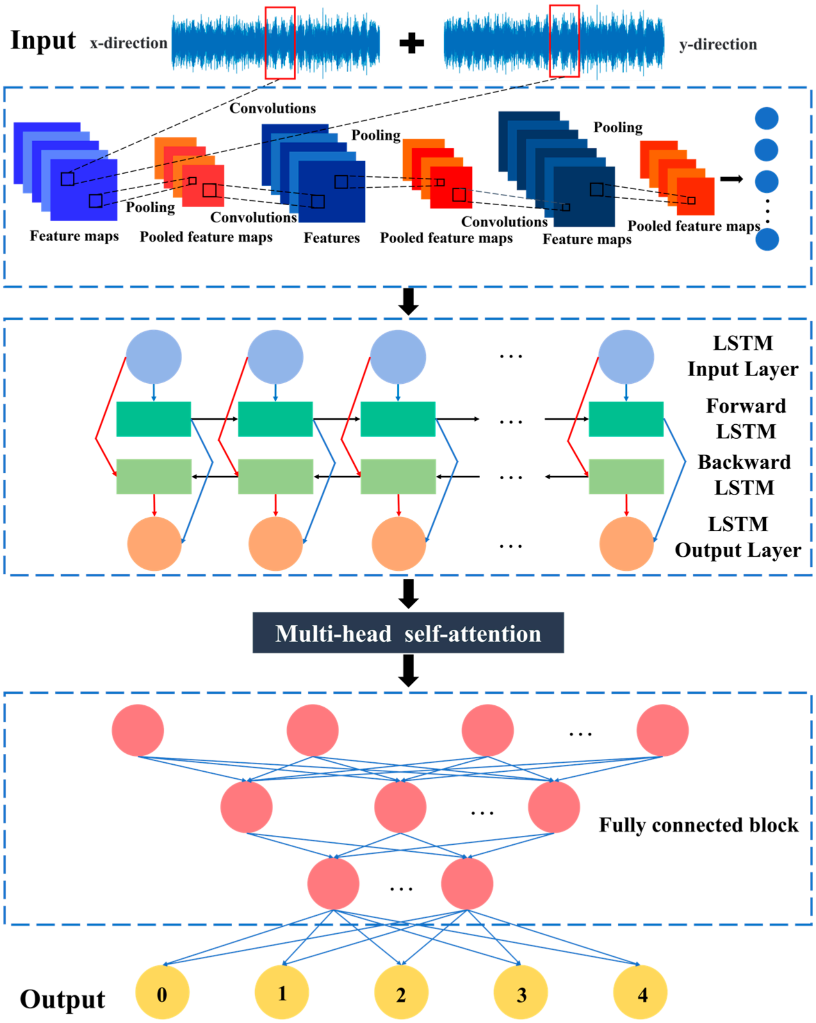


Figure 13. Diagram of rolling bearing diagnosis process based on CNN-BiLSTM-MHSA.

4. Model Verification and Performance Analysis

4.1. Efficient Data Selection for Enhanced Model Training and Generalization

A selective data interception method is applied to isolate vibration data corresponding to steady-state rotational speeds, constructing a refined dataset for input into the diagnostic network. This method reduces the dataset size by focusing on key samples, which minimizes storage requirements and improves data management efficiency. By emphasizing critical data, the method accelerates model training, reduces computational demands, and supports faster convergence toward optimal solutions. This approach is particularly advantageous in resource-constrained environments.

4.1.1. Advantages of Selective Data Extraction

The selective data extraction method optimizes the dataset size, which enhances computational efficiency and reduces training time. By focusing on vibration data from stable operating conditions, the quality of the input data is improved, minimizing the influence of noise and transient interference. This targeted approach increases the model's ability to generalize to new data, effectively lowering the risk of overfitting. Additionally, the use of a refined dataset simplifies the interpretability of the model by ensuring that training is based on meaningful samples, thereby improving trust in the model's predictions.

4.1.2. Dataset Design

The dataset is constructed following fault simulation experiments conducted at various rotational speeds. After acquiring the vibration signals in both the x and y directions, the data undergo filtering and preprocessing to remove noise and other irrelevant components. The dataset is categorized into five conditions—minor inner ring faults, severe inner ring faults, minor outer ring faults, severe outer ring faults, and normal operation—with 160 samples for each condition. Each sample consists of 10,000 data points—5000 from the x direction and 5000 from the y direction. These 5000 data points for each direction are extracted from the preprocessed signals recorded at different rotational speeds, ensuring a comprehensive representation of fault behaviors under varying operating conditions.

The dataset contains a total of 800 samples, with each of the five conditions equally represented. After preprocessing, the dataset is randomly shuffled and divided into a training set containing 650 samples and a test set containing 150 samples, with no overlap between the two sets. The labels are encoded as integers ranging from 0 to 4, corresponding to the five fault conditions. The data are flattened to fit the input format for the model.

This dataset is designed to ensure a complete representation of fault scenarios, covering different severities and fault locations, which is essential for effective model training. The high-dimensional vibration data capture rich temporal information, facilitating robust feature extraction and fault detection. Additionally, the dataset is carefully balanced, with each fault condition represented equally, preventing issues with class imbalance and improving the model's robustness and diagnostic accuracy.

The combination of well-structured data extraction, high-quality preprocessing, and balanced dataset organization ensures efficient model training, enhanced fault detection performance, and improved generalization across different fault scenarios.

4.2. Training

A training loop is implemented to train deep learning models. The training progress of bearing fault diagnosis based on CNN-BiLSTM-MHSA is shown in Figure 14.

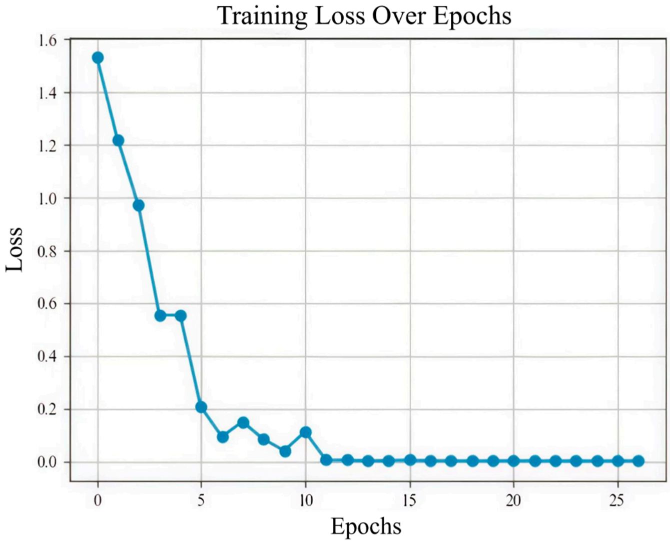


Figure 14. Progress chart of network model training.

4.2.1. Training Mechanism and Model Optimization

The training process of the deep learning model is designed to maximize performance while minimizing overfitting. A total of 50 training rounds (epochs) are defined, during which the loss value of each epoch is recorded to monitor performance and facilitate subsequent analysis and visualization. The initial "best loss" is set to a very high value (positive infinity), providing a benchmark for tracking improvements in model performance over time.

To prevent overfitting and computational inefficiency, an early stopping mechanism is implemented. The patience parameter, set at 5 epochs, defines the tolerance threshold for consecutive epochs without improvement. A counter variable tracks the number of non-improving epochs; if this count reaches the patience threshold, training halts early. This mechanism ensures the retention of the best-performing model parameters throughout training, improving both stability and resource efficiency.

Each epoch begins with the model in training mode, activating dropout and batch normalization layers to enhance generalization. Training data are fed in small batches to optimize GPU usage, particularly for large datasets. During each batch iteration,

1. The optimizer is reset to avoid gradient accumulation;
2. The model generates predictions for the input data, computes the loss by comparing predictions with actual labels, and performs backpropagation;
3. The computed gradients are used by the optimizer to update the model weights, improving its performance.

At the end of each epoch, the current loss is compared to the best historical loss. If the current loss improves, the best loss value is updated, and the counter is reset. If there is no improvement, the counter increments. When the counter reaches the patience threshold, the early stopping mechanism halts training. This iterative process, as shown in Figure 14, demonstrates that the model minimized loss to 0.003, prompting early termination.

4.2.2. Adaptive Learning Rate Adjustment

To accelerate convergence and improve stability, a dynamic learning rate adjustment strategy is employed. The learning rate scheduler modifies the learning rate as described by Equation (20):

$$\eta_{t+1} = \eta_t \cdot \gamma^{t/step\_size} \tag{20}$$

where η_t is the current learning rate, η_{t+1} is the updated learning rate, γ is the decay factor, and t is the iteration step.

The model utilizes the cross-entropy loss function shown in Equation (21) to measure the divergence between predicted probabilities and actual labels:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c} log(\widehat{y}_{i,c}) \tag{21}$$

where L is the average cross-entropy loss; N is the number of samples in the batch (i.e., batch size); C is the number of categories; and $\widehat{y}_{i,c}$ is the predicted probability of the model for sample i in category c, calculated by the softmax function, which satisfies Equation (22):

$$\widehat{y}_{i,c} = \frac{\exp(z_{i,c})}{\sum_{j=1}^{C}\exp(z_{i,j})} \tag{22}$$

where $z_{i,c}$ is the un-normalized output of the model.

The Adam optimizer updates the weights according to the rule described in Equation (22):

$$\theta_{t+1} = \theta_t - \eta \cdot \nabla_{\theta_t} L \tag{23}$$

where θ_t is the model parameter before updating, θ_{t+1} is the model parameter after updating, and ∇_{θ_t} is the loss-function (L) gradient with respect to parameter θ_t.

4.2.3. Key Optimization Strategies and Regularization

Key optimization strategies include early stopping, dynamic learning rate adjustment, and small-batch processing. Small batches efficiently utilize memory, reducing computational overhead while maintaining training stability.

Dropout regularization, as described in Equation (24), is employed to prevent overfitting:

$$x_{drop} = Dropout(x_{flat}, p) \tag{24}$$

where x_{flat} is the output of the final pooling layer being flattened and p is the dropout probability.

By deactivating a random subset of neurons in each iteration, dropout encourages the model to learn robust and generalized features, thereby mitigating overfitting.

4.3. Comparative Analysis of Diagnosis Results and Algorithms

The selective interception method improves both data management and computational efficiency, ensuring data quality and enhancing model generalization and interpretability. This approach is especially valuable in real-world applications where resources may be limited, providing an efficient and reliable solution for diagnosing faults in mechanical equipment.

The CNN-BiLSTM-MHSA model's predictive accuracy on the test set is illustrated in Figure 15 following training on a dataset of 650 samples derived from a rotor motor fault simulation, and Figure 16 presents the confusion matrix and t-SNE plot for the test set, where the horizontal axis represents the predicted category and the vertical axis represents the actual category.

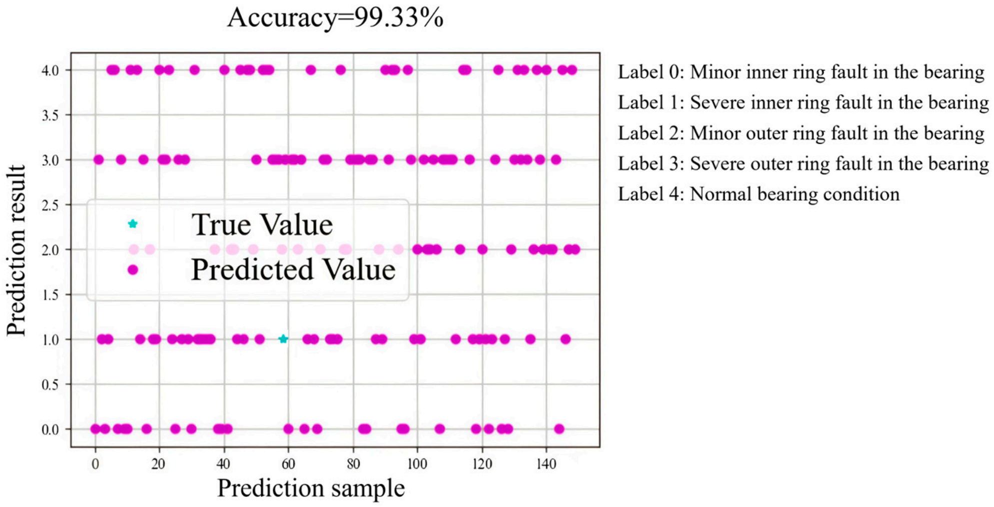


Figure 15. Comparison of test-set results.

As shown in Figures 15 and 16, the model achieves a prediction accuracy of 99.33% on the test set. In the confusion matrix, the horizontal axis represents the predicted labels, while the vertical axis represents the true labels. The matrix reveals clear distinctions among fault types. The t-SNE visualization demonstrates strong clustering of samples, with distinct separations among fault categories. The model achieves high accuracy in identifying normal, trouble-free conditions and both serious and minor faults in the outer and inner circles. However, slight misclassifications occur in distinguishing faults of a certain degree in the inner circle, with one sample misidentified.

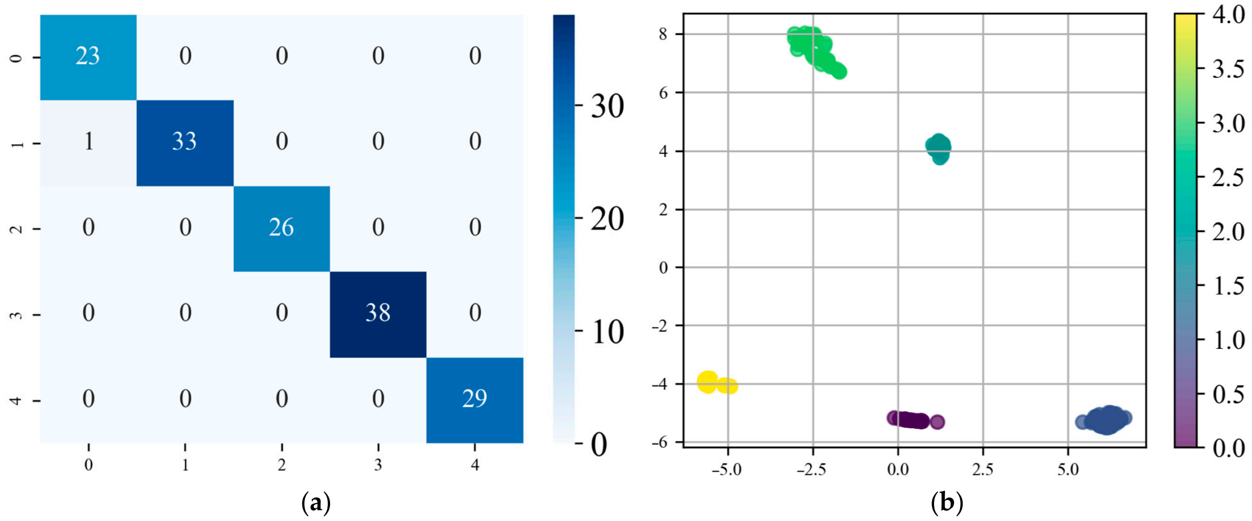


Figure 16. (**a**) Model test-set confusion matrix; (**b**) test-set t-SNE plot.

Precision (Prec), Recall (Rec), and F1 score are key metrics for evaluating the performance of classification models. The following metrics are used to assess the fault identification model:

1. Precision measures the proportion of samples predicted to be positive that are actually positive. It is calculated using Equation (25):

$$Prec = \frac{TP}{TP + FP} \tag{25}$$

2. Recall evaluates the proportion of true-positive samples correctly identified by the model. It is calculated using Equation (26):

$$Rec = \frac{TP}{TP + FN} \tag{26}$$

Here, TP (True Positive) is the number of correctly predicted positive samples, FP (False Positive) represents negative samples incorrectly classified as positive, and FN (False Negative) represents positive samples incorrectly classified as negative.

3. The F1 score is the harmonic mean of precision and recall, balancing the two metrics. It is computed using Equation (27):

$$F1 = 2 \times \frac{Rec \times Prec}{Rec + Prec} \tag{27}$$

Table 5 presents the recall, precision, and F1 score for the CNN-BiLSTM-MHSA-based bearing diagnosis model. These results underscore the model's ability to achieve high accuracy and reliability in practical fault diagnosis applications.

Table 5. Calculation of recall rate, precision rate, and F1 scores.

Indicator Fault Type	Minor Failure of the Inner Ring	Severe Failure of the Inner Ring	Minor Failure of the Outer Ring	Severe Failure of the Outer Ring	Normalcy
Precision (%)	100	97.0588	100	100	100
Recall (%)	95.8333	100	100	100	100
F1 score (%)	97.8723	98.5075	100	100	100

To quantitatively evaluate the overall performance of the fault diagnosis methods, several error metrics, including Root Mean Square Error (RMSE), Mean Absolute Error

(MAE), Mean Square Error (MSE), and Mean Absolute Percentage Error (MAPE), are used. These are supplemented by the calculation of precision, recall, and F1 scores.

1. RMSE represents the square root of the average squared differences between predicted and true values. A lower RMSE indicates a better fit between the predicted and actual values. RMSE is calculated using Equation (28):

$$RMSE = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}(y_i - \widehat{y}_i)^2} \tag{28}$$

where n is the sample size, y_i is the true value, and $\widehat{y}_i$ is the predicted value.

2. MSE is the mean of the squared deviations between the predicted and true values, reflecting the degree of model error accumulation. MSE is calculated using Equation (29):

$$MSE = \frac{1}{n}\sum\nolimits_{i=1}^{n}(y_i - \widehat{y}_i)^2 \tag{29}$$

3. MAE is the mean of the absolute differences between predicted and true values. A smaller MAE signifies a smaller deviation between predicted and true values. MAE is calculated using Equation (30):

$$MAE = \frac{1}{n}\sum\nolimits_{i=1}^{n}|y_i - \widehat{y}_i| \tag{30}$$

4. MAPE calculates the percentage difference between the predicted and true values. MAPE is calculated using Equation (31):

$$MAPE = \frac{100}{n}\sum\nolimits_{i=1}^{n}\frac{|y_i - \widehat{y}_i|}{y_i} \tag{31}$$

To validate the model's diagnostic ability, traditional CNN and LSTM models were also trained and tested using the same dataset. Four recent studies using motor bearing fault diagnosis models based on LTFM-Net [13], ARAE [36], CNN-LSTM [37], and WDCNN-LSTM [38] serve as the basis for comparison. The confusion matrices and t-SNE plots for the test set are shown in Figures 17–22. Figure 23 compares the performance metrics, and the results for each metric on the test set are presented in Table 6.

In Figures 17–22, the fault types are categorized as follows: label 0 represents a minor failure of the inner bearing ring, label 1 denotes a severe failure of the inner bearing ring, label 2 indicates a minor failure of the outer bearing ring, label 3 corresponds to a severe failure of the outer bearing ring, and label 4 represents a normal bearing.

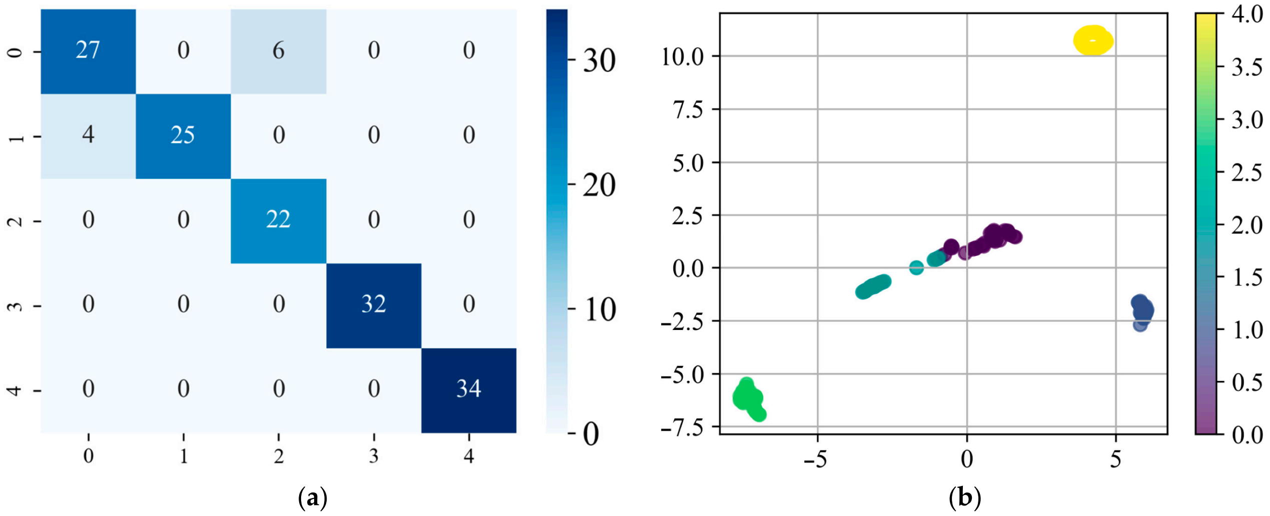
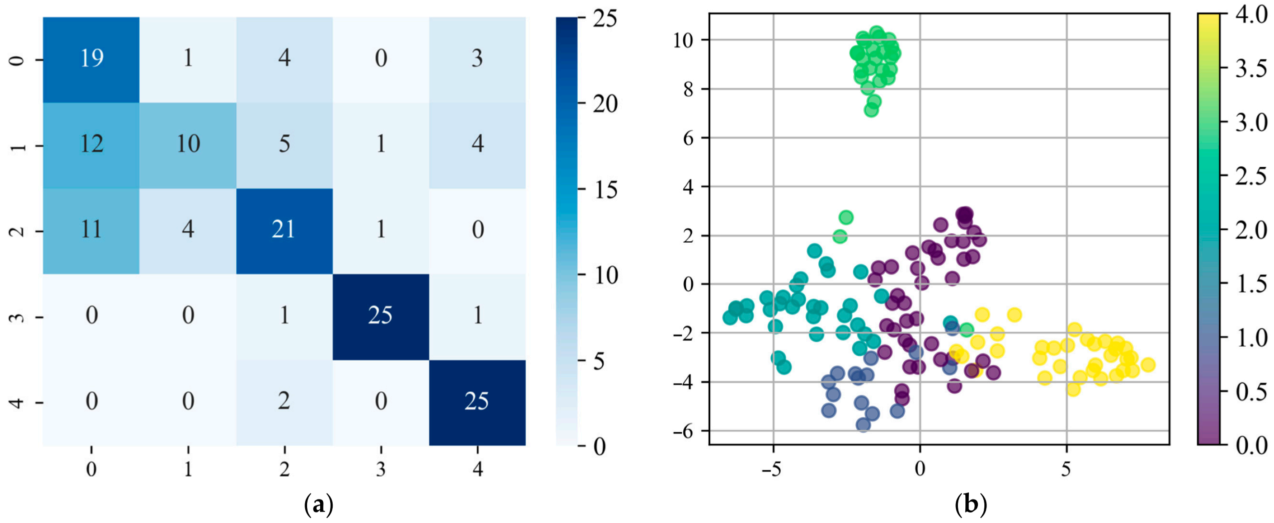
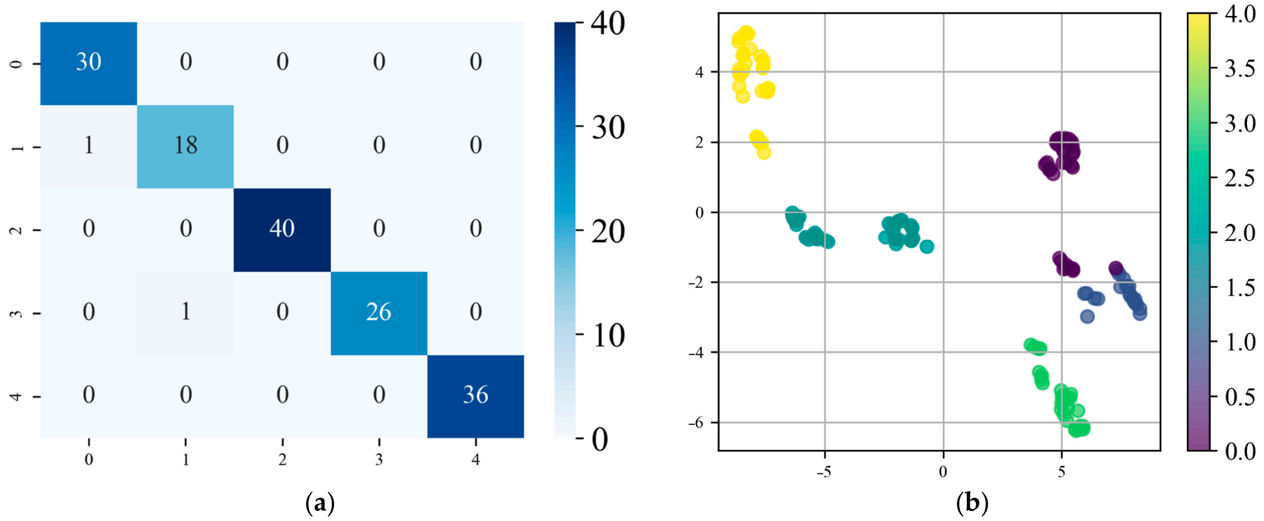


Figure 17. CNN model (**a**) test-set confusion matrix and (**b**) test-set t-SNE plot.



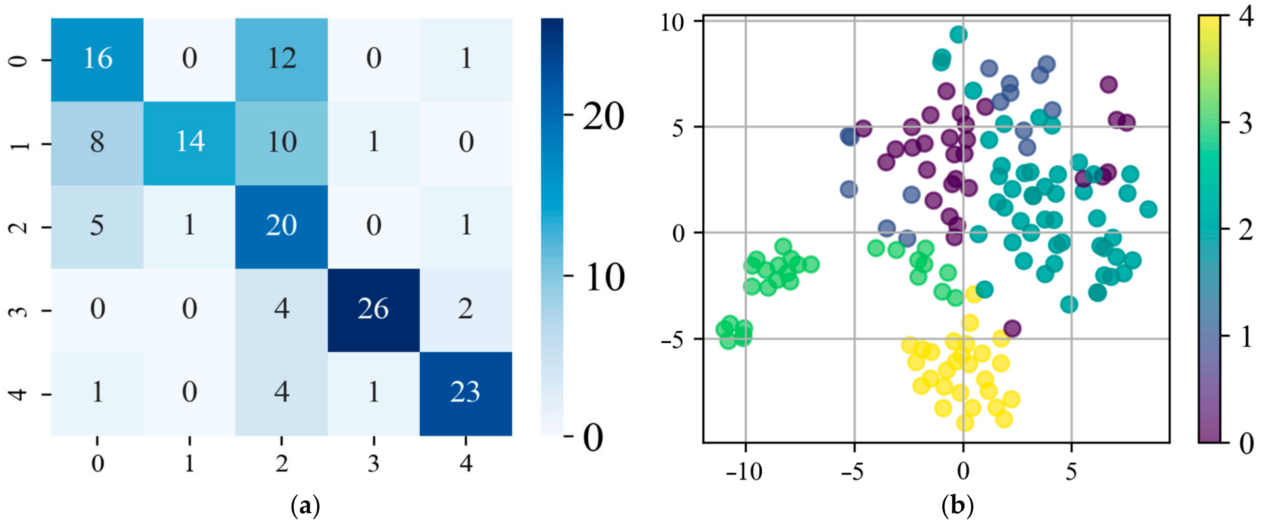
(**a**) (**b**)

Figure 18. LSTM model (**a**) test-set confusion matrix and (**b**) test-set t-SNE plot.



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

Figure 19. CNN-LSTM model (**a**) test-set confusion matrix and (**b**) test-set t-SNE plot.



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

Figure 20. ARAE model (**a**) test-set confusion matrix (**b**) and test-set t-SNE plot.

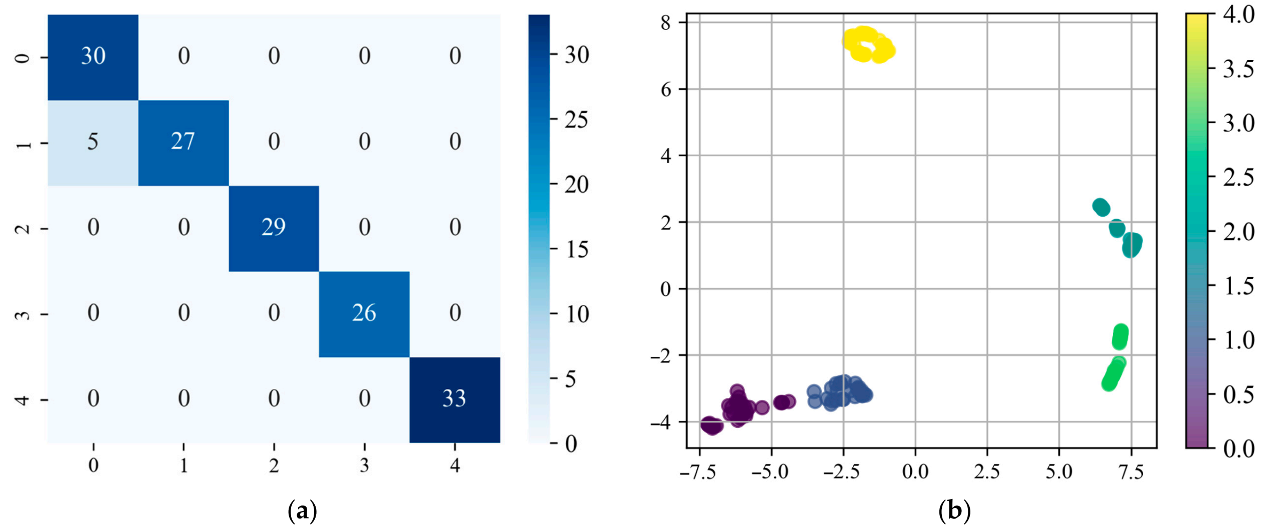


Figure 21. LTFM-Net model (**a**) test-set confusion matrix and (**b**) test-set t-SNE plot.

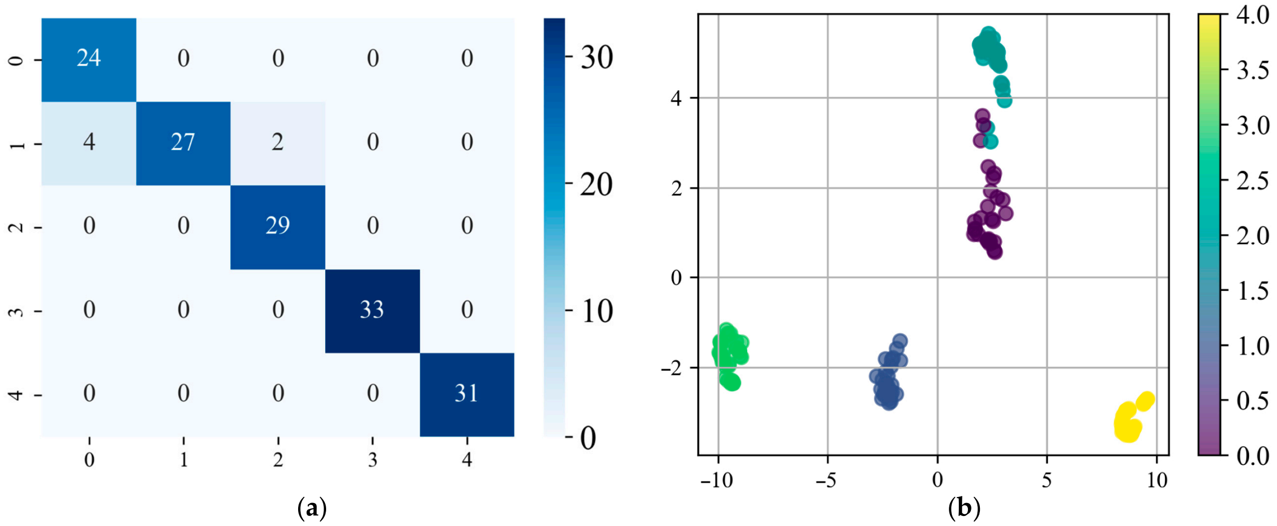


Figure 22. WDCNN-LSTM model (**a**) test-set confusion matrix and (**b**) test-set t-SNE plot.

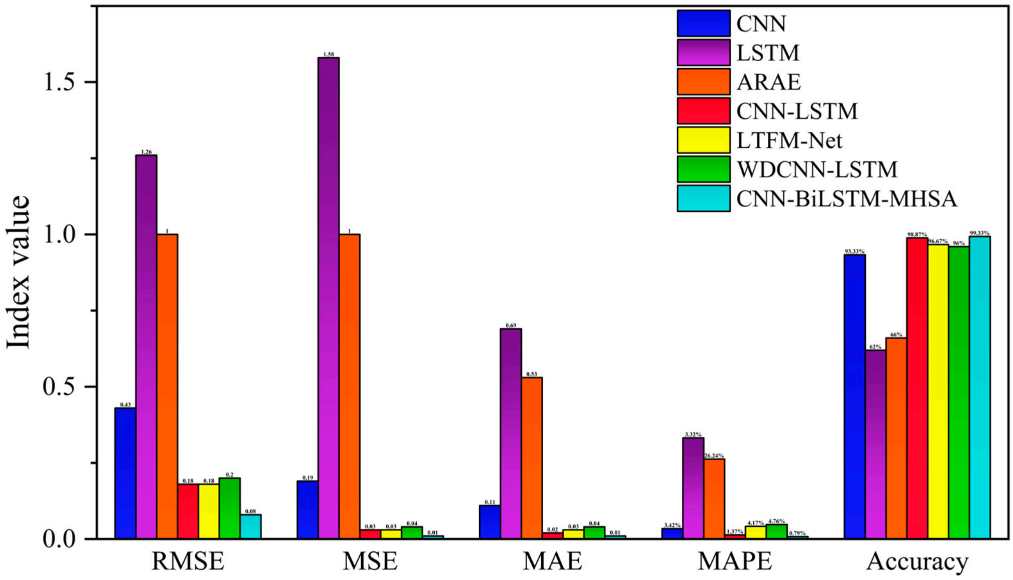


Figure 23. Comparison of performance indicators.

Table 6. Calculation results of test-set performance metrics.

Model	RMSE	MSE	MAE	MAPE	Accuracy
CNN	0.43	0.19	0.11	3.42%	93.33%
LSTM	1.26	1.58	0.69	33.20%	62.00%
ARAE	1.00	1.00	0.53	26.24%	66.00%
CNN-LSTM	0.18	0.03	0.02	1.37%	98.87%
LTFM-Net	0.18	0.03	0.03	4.17%	96.67%
WDCNN-LSTM	0.20	0.04	0.04	4.76%	96.00%
CNN-BiLSTM-MHSA	0.08	0.01	0.01	0.79%	99.33%

A comparative evaluation of various models for rolling bearing fault diagnosis was conducted, highlighting differing levels of performance across methods. The LSTM and ARAE models exhibited limited fault identification capabilities. Specifically, their t-SNE diagrams showed chaotic clustering, with insufficient separation between fault categories. While LSTM captures temporal dependencies, it struggles with multi-scale datasets due to its fixed time scales, leading to suboptimal performance in varying operational conditions. Similarly, ARAE, as a self-supervised learning model, extracts low-dimensional features but fails to capture complex, non-linear fault characteristics, limiting its sensitivity and diagnostic accuracy.

In contrast, the CNN, CNN-LSTM, and WDCNN-LSTM models performed better, achieving accuracies of 93.33%, 98.87%, and 96.00%, respectively. CNN's robust feature extraction abilities contribute to its relatively high accuracy, while incorporating LSTM in the CNN-LSTM model improves its ability to capture temporal features, leading to higher accuracy. The t-SNE diagrams for these models show clearer fault separations, though some misclassifications persist, particularly for minor inner ring faults.

The proposed CNN-BiLSTM-MHSA model further enhances diagnostic performance. By integrating BiLSTM, it captures global temporal dependencies and better handles non-stationary fault features, surpassing traditional LSTM. The addition of the MHSA mechanism enables the model to focus on critical features, thereby enhancing its sensitivity to key diagnostic characteristics. As a result, the CNN-BiLSTM-MHSA model achieves a remarkable fault diagnosis accuracy of 99.33%, outperforming the CNN (93.33%), LSTM (62.00%), ARAE (66.00%), CNN-LSTM (98.87%), LTFM-Net (96.67%), and WDCNN-LSTM (96.00%) models. Its t-SNE diagrams show clear fault separations, further reinforcing its superior diagnostic reliability.

Evaluation metrics for the CNN-BiLSTM-MHSA model, such as RMSE (0.08), MSE (0.01), MAE (0.01), and MAPE (0.79%), demonstrate significantly lower values compared to other models, emphasizing its superior predictive accuracy and robust data fit. The model also achieves an exceptional precision (100%) and F1 score (99.83%) in diagnosing normal bearings and outer ring faults. However, the recall rate for inner ring faults, particularly severe ones, is slightly lower than expected, suggesting challenges in capturing subtle variations in fault severity. This is reflected in the confusion matrix, where minor misclassifications between mild and severe inner ring faults occasionally occur. These discrepancies indicate that while the model performs well across most fault types, its sensitivity to nuanced inner ring faults requires improvement.

For outer ring faults, the CNN-BiLSTM-MHSA model outperforms the CNN and LSTM models, achieving a higher accuracy (99.33%) and F1 score (99.34%) while maintaining lower error metrics. To further enhance its robustness, especially in real-world rotor motor applications with varying conditions, expanding the training dataset to include more instances of severe outer ring faults could improve diagnostic accuracy and adaptability.

5. Conclusions

This study introduces the CNN-BiLSTM-MHSA model, a hybrid deep learning architecture combining convolutional neural networks, bidirectional long short-term memory networks, and multi-head self-attention mechanisms, tailored for rotor motor bearing fault diagnosis. The model effectively extracts spatial, temporal, and attention-based features from vibration signals, achieving a remarkable diagnostic accuracy of 99.33%. Comparative analysis reveals that it outperforms traditional models such as CNN (93.33%), LSTM (62.00%), ARAE (66.00%), CNN-LSTM (98.87%), LTFM-Net (96.67%), and WDCNN-LSTM (96.00%) across multiple fault categories. Metrics like RMSE, MSE, MAE, and MAPE further highlight its reliability, and its high precision, recall, and F1 scores demonstrate superior fault classification performance. These results affirm the model's robustness under diverse operational conditions, making it suitable for dynamic environments like UAV systems and industrial machinery.

Despite these significant achievements, the study identifies notable limitations. First, the model's focus on vibration signals under variable speeds leaves a gap in evaluating performance under fluctuating load conditions and real-world fault signals. Second, while the model achieves high accuracy in diagnosing inner ring faults, its sensitivity to subtle differences between minor and severe inner ring faults is limited, likely due to imbalanced feature representation in the training dataset. Addressing these issues is essential to further enhance the model's diagnostic accuracy and generalization ability.

Future work will focus on expanding the dataset to include real-world fault scenarios characterized by gradual fault progression, fluctuating loads, and diverse environmental influences. Addressing these complexities will bridge the gap between controlled simulations and practical applications. Enhancing feature extraction techniques and incorporating advanced attention mechanisms will improve the model's ability to capture subtle fault characteristics. Additionally, refining the model architecture and training strategies will ensure adaptability and robustness in varied and dynamic operational conditions, such as those encountered in UAVs. These efforts will solidify the model's applicability in real-world environments while maintaining its high diagnostic accuracy.

In conclusion, the CNN-BiLSTM-MHSA model represents a significant advancement in rotor motor bearing fault diagnosis, demonstrating high accuracy, robustness, and reliability. By addressing its limitations, the model can further solidify its role as a robust solution for deployment in complex and dynamic environments, such as UAVs and industrial systems.

Author Contributions: Conceptualization, Z.Y.; methodology, Z.Y.; validation, W.L. and Z.Y.; formal analysis, Z.Y., W.L., F.Y. and Z.G.; writing—original draft preparation, Z.Y. and Z.G.; writing—review and editing, Z.G. and B.X.; project administration, H.Z. and M.G.; funding acquisition, M.G. and B.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the basic scientific research institute stability support project "WDZC-2023-ZNKZ-03".

Data Availability Statement: The original data cannot be released for the time being due to the confidentiality involved in the initial stage of the research.

Conflicts of Interest: Author Min Guo was employed by the Shanxi Diesel Engine Industry Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

UAV	Unmanned Aerial Vehicle
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
BiLSTM	Bidirectional Long Short-Term Memory
MHSA	Multi-Head Self-Attention
ARAE	Attention Recurrent Autoencoder
WD	Wavelet Denoising
ESC	Electronic Speed Controller
SVM	Support Vector Machine
DL	Deep Learning
AI	Artificial Intelligence
ECNN	Efficient Convolutional Neural Network
EH-SBW	Electro-Hydraulic Steer-By-Wire
1DCNN	One-Dimensional Convolutional Neural Network
LTFM	Lightweight Time-Focused Model
WDRU	Weighted Diminish Recurrent Unit
FAN	Feature Aggregation Network
FTSVD	Flexible Tensor Singular Value Decomposition
TDR	Trajectory Dimension Ratio
TRPCA	Tensor Robust Principal Component Analysis
GE-LRTLM	Graph-Embedded Low-Rank Tensor Learning Machine
RNN	Recurrent Neural Network
NLP	Natural Language Processing
FFT	Fast Fourier Transform
ReLU	Rectified Linear Unit
DFT	Discrete Fourier Transform
CPU	Central Processing Unit
GPU	Graphics Processing Unit
FC	Fully Connected
t-SNE	t-distributed Stochastic Neighbor Embedding
Prec	Precision
Rec	Recall
TP	True Positive
FP	False Positive
FN	False Negative
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
MSE	Mean Square Error
MAPE	Mean Absolute Percentage Error

References

1. Pusca, R.; Sbaa, S.; Bessous, N.; Romary, R.; Bousseksou, R. Mechanical Failure Detection in Induction Motors Using Stator Current and Stray Flux Analysis Techniques. *Eng. Proc.* **2022**, *14*, 19. [CrossRef]
2. Rajeev, K.; Anand, R.S. Bearing Fault Diagnosis Using Multiple Feature Selection Algorithms with SVM. *Prog. Artif. Intell.* **2024**, *13*, 119–133.
3. Qiu, W.; Wang, B.; Hu, X. Rolling Bearing Fault Diagnosis Based on RQA with STD and WOA-SVM. *Heliyon* **2024**, *10*, e26141. [CrossRef] [PubMed]
4. Wang, X.; Li, Y.; Zhang, J.; Liu, F. The Research on Fault Diagnosis of Rolling Bearing Based on Current Signal CNN-SVM. *Meas. Sci. Technol.* **2023**, *34*, 125021. [CrossRef]
5. Qu, J.; Xu, Z.; Li, C.; Zhang, M. Fault Diagnosis of Bearings Using Wavelet Packet Energy Spectrum and SSA-DBN. *Processes* **2023**, *11*, 1875. [CrossRef]

6. Xue, L.; Yang, F.; Chen, Z.; Gao, H. An AVMD-DBN-ELM Model for Bearing Fault Diagnosis. *Sensors* **2022**, *22*, 9369. [CrossRef]
7. Ni, Z.; Sun, J.; Liu, Q.; Wu, T.; Zhang, S. Enhanced Bearing Fault Diagnosis in NC Machine Tools Using Dual-Stream CNN with Vibration Signal Analysis. *Processes* **2024**, *12*, 1951. [CrossRef]
8. An, K.; Lu, J.; Zhu, Q.; Wang, X.; De Silva, C.W.; Xia, M.; Lu, S. Edge solution for real-time motor fault diagnosis based on efficient convolutional neural network. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 3516912. [CrossRef]
9. Evangeline, I.S.; Darwin, S.; Raj, I.F.E. A deep residual neural network model for synchronous motor fault diagnostics. *Appl. Soft Comput.* **2024**, *160*, 111683. [CrossRef]
10. Fan, H.; Ren, Z.; Zhang, X.; Cao, X.; Ma, H.; Huang, J. A gray texture image data-driven intelligent fault diagnosis method of induction motor rotor-bearing system under variable load conditions. *Measurement* **2024**, *233*, 114742. [CrossRef]
11. Zhang, S.; Liang, W.; Zhao, W.; Luan, Z.; Wang, C.; Xu, K. Electro-hydraulic SBW fault diagnosis method based on novel 1DCNN-LSTM with attention mechanisms and transfer learning. *Mech. Syst. Signal Process.* **2024**, *220*, 111644. [CrossRef]
12. Kim, M.C.; Lee, J.H.; Wang, D.H. Induction Motor Fault Diagnosis Using Support Vector Machine, Neural Networks, and Boosting Methods. *Sensors* **2023**, *23*, 2585. [CrossRef] [PubMed]
13. Yang, T.; Jiang, L.; Guo, Y.; Han, Q.; Li, X. LTFM-net framework: Advanced intelligent diagnostics and interpretability of insulated bearing faults in offshore wind turbines under complex operational conditions. *Ocean Eng.* **2024**, *309 Pt 2*, 118533. [CrossRef]
14. Raouf, I.; Kumar, P.; Kim, H.S. Deep learning-based fault diagnosis of servo motor bearing using the attention-guided feature aggregation network. *Expert Syst. Appl.* **2024**, *258*, 125–137. [CrossRef]
15. Cheng, L.; Kong, X.; Zhang, Y.; Zhu, Y.; Qi, H.; Zhang, J. A novel causal feature learning-based domain generalization framework for bearing fault diagnosis with a mixture of data from multiple working conditions and machines. *Adv. Eng. Inform.* **2024**, *15 Pt A*, 102622. [CrossRef]
16. Ma, H.; Li, J.; Huang, J.; Wang, R.; Ge, R.; Zhang, F. Adaptive Embedded Flexible Tensor Singular Spectrum Decomposition. *Electronics* **2024**, *14*, 21. [CrossRef]
17. Huang, J.; Zhang, F.; Coombs, T.; Chu, F. The first-kind flexible tensor SVD: Innovations in multi-sensor data fusion processing. *Nonlinear Dyn.* **2024**. [CrossRef]
18. Xu, H.; Wang, X.; Huang, J.; Zhang, F.; Chu, F. Semi-supervised multi-sensor information fusion tailored graph embedded low-rank tensor learning machine under extremely low labeled rate. *Inf. Fusion* **2024**, *105*, 102222. [CrossRef]
19. Abid, A.; Khan, M.T.; Iqbal, J. A review on fault detection and diagnosis techniques: Basics and beyond. *Artif. Intell. Rev.* **2021**, *54*, 3639–3664. [CrossRef]
20. Dubaish, A.A.; Jaber, A.A. State-of-the-art review into signal processing and artificial intelligence-based approaches applied in gearbox defect diagnosis. *Eng. Technol. J.* **2024**, *42*, 157–172. [CrossRef]
21. Barai, V.; Ramteke, S.M.; Dhanalkotwar, V.; Nagmote, Y.; Shende, S.; Deshmukh, D. Bearing fault diagnosis using signal processing and machine learning techniques: A review. *IOP Conf. Ser. Mater. Sci. Eng.* **2022**, *1259*, 012034.
22. Xu, X.; Cao, D.; Zhou, Y.; Gao, J. Application of neural network algorithm in fault diagnosis of mechanical intelligence. *Mech. Syst. Signal Process.* **2020**, *141*, 106625. [CrossRef]
23. Guo, Y.; Zhou, J.; Dong, Z.; She, H.; Xu, W. Research on bearing fault diagnosis based on novel MRSVD-CWT and improved CNN-LSTM. *Meas. Sci. Technol.* **2024**, *35*, 095003. [CrossRef]
24. Zhang, Q.; Wei, X.; Wang, Y.; Hou, C. Convolutional Neural Network with Attention Mechanism and Visual Vibration Signal Analysis for Bearing Fault Diagnosis. *Sensors* **2024**, *24*, 1831. [CrossRef] [PubMed]
25. Kang, J.; Zhu, X.; Shen, L.; Li, M. Fault diagnosis of a wave energy converter gearbox based on an Adam optimized CNN-LSTM algorithm. *Renew. Energy* **2024**, *121*, 1022. [CrossRef]
26. Li, Z.; Jiang, Z.; Gao, Z.; Zhang, W. A state estimation method based on CNN-LSTM for ball screw. *Meas. Control* **2024**, *57*, 1417–1434. [CrossRef]
27. Liu, X.; Chen, G.; Wang, H.; Wei, X. A Siamese CNN-BiLSTM-based method for unbalance few-shot fault diagnosis of rolling bearings. *Meas. Control* **2024**, *57*, 551–565. [CrossRef]
28. Lu, S.; Liu, M.; Yin, L.; Yin, Z.; Liu, X.; Zheng, W. The multi-modal fusion in visual question answering: A review of attention mechanisms. *PeerJ Comput. Sci.* **2023**, *9*, e1400. [CrossRef]
29. Li, S.; Xu, Y.; Jiang, W.; Zhao, K.; Liu, W. A modular fault diagnosis method for rolling bearing based on mask kernel and multi-head self-attention mechanism. *Trans. Inst. Meas. Control* **2024**, *46*, 899–912. [CrossRef]
30. Gao, H.; Ma, J.; Zhang, Z.; Cai, C. Bearing Fault Diagnosis Method Based on Attention Mechanism and Multi-Channel Feature Fusion. *IEEE Access* **2024**, *12*, 45011–45025. [CrossRef]
31. Chu, S.; Zhang, J.; Liu, F.; Kong, X.; Jiang, Z.; Mao, Z. Fault identification model of diesel engine based on mixed attention: Single-cylinder fault data driven whole-cylinder diagnosis. *Expert Syst. Appl.* **2024**, *255*, 124769. [CrossRef]
32. Wang, P.; Zhao, X.; Yang, Y.; Ma, H.; Han, Q.; Luo, Z.; Wen, B. Dynamic modeling and analysis of two-span rotor-pedestal system with bearing tilt and extended defect: Simulation and experiment. *Appl. Math. Model.* **2024**, *125*, 1–28. [CrossRef]

33. He, X.; Ding, J.; Wang, X.; Zhang, Q.; Zhao, W.; Wang, K. Adaptive extraction of characteristic ridges from time-frequency representation for wheelset bearings failure diagnosis under time-varying speed. *Measurement* **2024**, *242*, 115987. [CrossRef]
34. Xu, M.; Yu, Q.; Chen, S.; Lin, J. Rolling Bearing Fault Diagnosis Based on CNN-LSTM with FFT and SVD. *Information* **2024**, *15*, 399. [CrossRef]
35. Fang, X.; Zheng, J.; Jiang, B. A rolling bearing fault diagnosis method based on vibro-acoustic data fusion and fast Fourier transform (FFT). *Int. J. Data Sci. Anal.* **2024**, 1–10. [CrossRef]
36. Kong, X.; Li, X.; Zhou, Q.; Hu, Z.; Shi, C. Attention recurrent autoencoder hybrid model for early fault diagnosis of rotating machinery. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 2505110. [CrossRef]
37. Zhou, Q.; Tang, J. An Interpretable Parallel Spatial CNN-LSTM Architecture for Fault Diagnosis in Rotating Machinery. *IEEE Internet Things J.* **2024**, *11*, 31730–31744. [CrossRef]
38. Gao, Y.; Kim, C.H.; Kim, J.M. A novel hybrid deep learning method for fault diagnosis of rotating machinery based on extended WDCNN and long short-term memory. *Sensors* **2021**, *21*, 6614. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.