Eksploatacja i Niezawodnosc - Maintenance and Reliability

Volume 29 (2027), Issue 1

https://ein.org.pl

The Polish Maintenance Society

Research on UAV rotor fault diagnosis for small sample data

Zheng Gong[1,*], Feng Lin[1]

[1] College of Electrical Engineering, Zhejiang University, China
* Corresponding Author: gongzheng23@zju.edu.cn

Received: 3 April 2026
Revised: 22 June 2026
Accepted: 25 June 2026
Online: 7 August 2026

Abstract

Unmanned Aerial Vehicles (UAVs) face significant challenges in rotor fault diagnosis under small-sample and imbalanced data conditions, which severely limit the effectiveness of existing data-driven methods. In practical applications, fault samples are often fewer than 50 per class, with imbalance ratios typically 0.3–0.5, under which conventional deep learning approaches exhibit notable performance degradation. To address these issues, this paper proposes an MSCNN-BiLSTM-Attention framework for UAV rotor fault diagnosis. The method integrates multi-scale convolution to enhance spatial feature extraction from limited vibration data, bidirectional LSTM to capture temporal dependencies, and an attention mechanism combined with Focal Loss to improve minority-class learning. Experiments demonstrate that the proposed framework consistently outperforms conventional methods under small-sample, noisy, and imbalanced conditions, achieving a maximum improvement of up to 9 percentage points, which confirms its robustness and effectiveness in challenging diagnostic scenarios.

Keywords: *UAV rotor fault diagnosis, small sample learning, multi-scale convolutional neural network, bidirectional long short-term memory network, attention mechanism*

Article citation:
Gong Z, Lin F, Research on UAV rotor fault diagnosis for small sample data, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2027: 29(1) http://doi.org/10.17531/ein/225080

Highlights

- A novel MSCNN-BiLSTM-Attention model is proposed for UAV rotor fault diagnosis.
- The MSCNN module extracts multi-scale spatial vibration features with low complexity.
- The BiLSTM module captures temporal correlations across multiple time sequences.
- Attention with Focal Loss enhances minority feature learning and classification accuracy.
- The model achieves 100% accuracy with 45 samples and 85.33% with only 15 samples per class.

1. Introduction

Today, quadcopter drones are extensively used in military, transportation, agriculture, and other sectors [1]. As the primary control mechanism within drones [2], rotors play a critical role in flight stability. However, with the widespread adoption and increasing complexity of UAVs, the rate of accidents has also risen. Among the various components susceptible to failure, rotors are particularly vital; imbalances or structural abnormalities in rotors can lead to flight instability or even catastrophic failures. Therefore, to ensure safe and reliable UAV operations, rapid and efficient fault diagnosis for rotors has become an urgent requirement.

A series of studies on rotor fault diagnosis have been conducted both domestically and internationally, and these studies can be primarily categorized into three approaches: analytical model-based analysis, signal analysis-based methods, and artificial intelligence-based methods. Analytical model-based methods typically involve the development of mathematical models for UAV systems, designing observers to estimate system states, and comparing these estimates with actual outputs to generate residuals for fault detection and diagnosis. While these methods offer accurate representations of the UAV's flight state and can achieve high fault diagnosis accuracy, the increasing complexity of UAV structures has significantly raised the difficulty of model construction, presenting challenges for practical application. In addition, these methods generally rely on accurate system modeling and sufficient prior knowledge, which limits their adaptability when fault samples are scarce or system parameters vary under complex operating conditions.

Signal analysis methods, in contrast, focus on feature analysis of physical signals such as vibration and current, which are collected during drone flight in either the time or frequency domain. Fault diagnosis is performed by comparing these data with fault feature templates. Common methods in this category include active energy Bayesian [3], frequency domain vibration spectrum analysis [4], ensemble empirical mode decomposition [5], and wavelet packet transform [6,7]. These methods can achieve high diagnostic accuracy without the need for precise mathematical models. However, the diagnostic performance of signal analysis–based methods is highly dependent on manually selected features and thresholds, making them less robust when fault samples are limited or unevenly distributed across categories.

In recent years, artificial intelligence methods have gradually gained traction in drone fault diagnosis, primarily utilizing backpropagation neural networks (BP) [8], convolutional neural networks (CNN) [9], stacked autoencoders [10,11], Long Short-Term Memory Networks (LSTM) [12,13], and Generative Adversarial Networks (GANs) [14]. These methods do not require complex mathematical modeling or manual parameter thresholding. By designing and training network models specifically tailored for fault data, they achieve high diagnostic accuracy. For instance, Zhang et al. demonstrated collaborative feature-sharing strategies and hybrid deep domain adaptation networks that improved fault classification in small-sample UAV datasets [15]. However, most existing AI-based diagnostic models implicitly assume the availability of sufficient and balanced training data, and their performance tends to degrade due to overfitting and biased learning toward majority classes under small-sample and class-imbalanced conditions.

In practical applications, drone failures occur infrequently, fault simulation is costly, and fault data collection is challenging, often resulting in an insufficient number of fault data samples. However, AI-based diagnostics rely heavily on large volumes of high-quality fault data. Therefore, there is an urgent need for fault diagnosis research under small-sample conditions. Solutions to the small-sample problem generally fall into two categories: generating and augmenting original data, and improving neural network algorithms. Examples include using Generative Adversarial Networks (GANs) for data augmentation [16] and enhancing Deep Forest models [17,18] to achieve efficient fault diagnosis. Recent research has shown that GAN-based augmentation can improve classification performance under small-sample and imbalanced fault conditions, with synthetic time-series samples leading to more balanced training and better classifier generalization in rotating machinery fault diagnosis [19]. However, despite their potential, GAN-based methods still face challenges such as training instability and difficulty capturing complex temporal dependencies in vibration signals, and imbalance-aware strategies alone cannot fully address insufficient joint spatial-temporal feature learning in small-sample UAV fault datasets [20]. So existing research faces several challenges, such as low-quality generated data and insufficient feature extraction techniques tailored to small-sample datasets, leaving significant room for improvement in diagnostic performance. In particular, many existing approaches focus primarily on data quantity expansion while overlooking the joint optimization of spatial and temporal feature representation under small-sample constraints.

Furthermore, in practical applications, the probability of rotor failure is relatively low, leading to a significant imbalance between normal and fault data samples. This imbalance causes overfitting in AI models when dealing with categories that contain a large number of samples. Approaches to addressing imbalanced data primarily focus on three levels: data, algorithms, and loss functions. For example, data augmentation is performed using convolutional models [21], multi-scale convolutional algorithms [22], and modified loss functions that prioritize scarce sample categories [23,24]. However, existing methods still struggle to effectively address the substantial imbalance between major categories in the sample data. Especially under small-sample conditions, imbalance-aware strategies often fail to fully exploit discriminative features of minority fault classes, leading to unstable generalization performance. Moreover, temporal feature extraction for small samples remains insufficient, resulting in low diagnostic accuracy.

More recently, deep learning–based methods, including CNNs, RNNs, and their hybrid variants, have demonstrated superior feature extraction capabilities and improved diagnostic accuracy. However, most existing deep models assume the availability of sufficient and balanced training data. Under small-sample and class-imbalanced conditions, these models are prone to overfitting, biased learning toward majority classes, and unstable generalization performance. In addition, many existing hybrid architectures increase model complexity without explicitly addressing minority-class learning and multi-scale feature representation simultaneously.

To address the above challenges, this study makes the following main contributions:

(1) A novel MSCNN-BiLSTM-Attention framework is proposed for UAV rotor fault diagnosis under small-sample and imbalanced data conditions, effectively integrating spatial–temporal feature learning and class-aware optimization.

(2) A multi-scale convolutional module is designed to enhance spatial feature extraction from limited vibration data, improving inter-class discriminability compared with conventional single-scale CNN–LSTM hybrid models.

(3) An attention mechanism combined with Focal Loss is introduced to emphasize informative temporal features and mitigate the adverse effects of class imbalance.

(4) Extensive experiments under small-sample, noisy, and imbalanced conditions validate the effectiveness and robustness of the proposed method.

2. Theoretical design and model construction

2.1. Longitudinal spatial feature extraction module

Longitudinal spatial feature extraction focuses on multi-scale feature mining of UAV rotor vibration signals at a single time instant, aiming to capture discriminative spatial characteristics under small-sample and imbalanced conditions. In this study, a convolutional neural network (CNN) is employed as the basic feature extractor due to its ability to model local correlations and share weights efficiently [25]. The convolution operation can be expressed as:

$$x_j^l = f\left(\sum_{i=1}^{m} x_i^{l-1} \cdot k_{ij}^l + b_j^l\right), \quad j = 1,2,\cdots,n \tag{1}$$

where: x_j^l represents the feature plane of the jth output UAV rotor signal, with l denoting the index layer number in the convolutional network; x_i^{l-1} denotes the feature plane of the ith input UAV rotor signal; k_{ij}^l is the parameter of the jth convolutional kernel connected to the ith input signal feature plane; b_j^l is the bias of the jth convolutional kernel; f is the activation function. Convolution essentially acts as a filter for the signal. By setting the parameters of the convolutional kernel, the CNN network can effectively extract fault features from the rotor signal.

To improve training stability and feature representation, batch normalization and nonlinear activation are applied after each convolutional operation. Batch normalization normalizes intermediate features to accelerate convergence and alleviate gradient instability, as described by:

$$L_N(x) = \frac{x - E(x)}{\sqrt{V_{ar}(x) + \epsilon}} \times \alpha + \beta \tag{2}$$

In Equation (2), x represents the rotor signal processed by the convolutional layer; $E(x)$ denotes the mean of the input rotor signal; $V_{ar}(x)$ indicates the variance; to prevent division by zero, ϵ is typically set to 10^{-5}; α and β are trainable parameters. The ReLU activation function is then adopted to enhance nonlinearity and maintain efficient gradient propagation:

$$\mathrm{ReLU}(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{3}$$

Compared to other activation functions, ReLU simplifies differentiation, thereby accelerating network training during backpropagation. Additionally, its non-saturating nature prevents issues such as gradient vanishing, even with extremely large or small values. These advantages contribute to faster and more accurate fault diagnosis.

A global average pooling layer is subsequently introduced to aggregate spatial information across each channel while significantly reducing the number of trainable parameters. This design effectively mitigates overfitting risks under small-sample conditions and improves model generalization. The extracted features are then flattened into one-dimensional vectors to facilitate subsequent temporal modeling.

The key to achieving accurate fault diagnosis under small-sample and imbalanced conditions lies in the ability of the extracted features to capture the differences between various rotor signals. The aforementioned single-scale CNN layer can only capture specific portions of the signal, making it challenging to extract sufficient inter-class discriminative features under small-sample and imbalanced conditions. To address this, this paper proposes a longitudinal spatial feature extraction module based on Multi-Scale Convolutional Neural Network. By employing a CNN network that combines two convolutional kernels of different sizes, it enables thorough feature mining for small-sample and imbalanced data, facilitating effective fault diagnosis by subsequent layers. Unlike single-scale CNNs that capture limited local patterns, the proposed multi-scale design is explicitly tailored to enhance inter-class feature separability when training samples are scarce and class distributions are imbalanced.

As shown in Figure 1, preprocessed UAV rotor signals are fed into two parallel convolutional modules with different kernel sizes. These modules extract features from the input signals, generating feature maps of varying dimensions. Pooling layers are then applied to reduce the dimensionality of these feature maps, minimizing network parameters and helping to prevent overfitting. Finally, feature fusion combines the two-scale feature maps into a unified dimension.

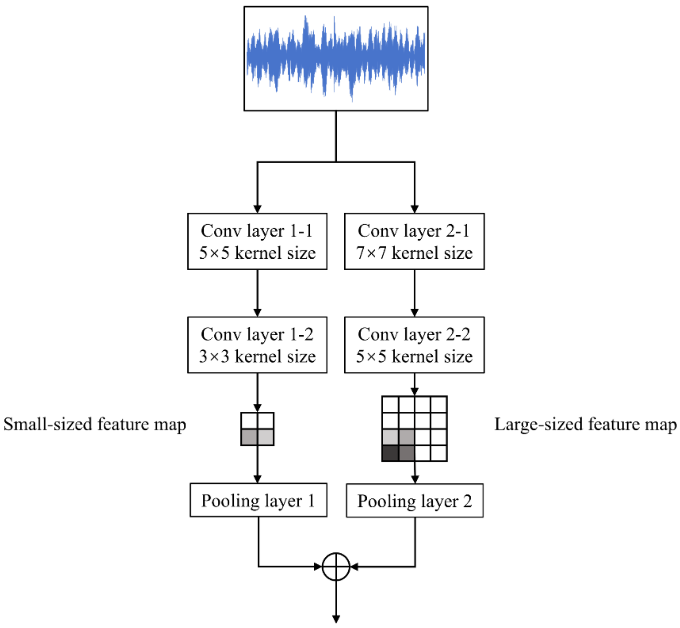

Figure 1. Structure diagram of the multi-scale feature extraction module.

The iterative calculation formula for the receptive field of different convolutional kernel layers is as follows:

$$R\_i=K\_i\ (R\_(i+1)-1)+D \tag{4}$$

where: D denotes the convolution kernel size;K represents the stride; R indicates the receptive field; and the subscript i signifies the convolution layer number.

When the stride and iteration count are fixed, the convolution kernel size determines the receptive field size. Smaller convolution kernels focus on local feature correlations within the drone rotor data, capturing critical information from the vibration signals. Larger kernels, on the other hand, enable the extraction of global, holistic signal characteristics. By applying kernels of different sizes in parallel to the same signal, the network can learn core features across multiple scales. This approach allows the extraction of distinctive vibration features from diverse scales, thereby enhancing the network's fault diagnosis capability.

Moreover, combining convolutional kernels of varying sizes allows the model to adapt to complex and variable input signals, enhancing its generalization capabilities.

The multi-scale feature extraction module adjusts the receptive field for input rotor signals, reducing the need for empirical selection of kernel sizes while extracting robust multi-scale features. Compared to single-scale features, multi-scale features better capture inter-class differences in fault data, allowing for the full utilization of small and imbalanced samples and improving the low accuracy of traditional diagnostic methods.

2.2. Lateral temporal feature extraction module

Lateral temporal feature extraction primarily identifies connections between rotor signal features across multiple time points, extracting lateral features that reflect the temporal relationships between signals. This module is built on a bidirectional long short-term memory (BiLSTM) network. The BiLSTM network is a variant of the Long Short-Term Memory (LSTM) network, combining a forward LSTM with a backward LSTM [26]. The LSTM network is an extension of the Recurrent Neural Network (RNN), addressing the short-term memory limitation of RNNs. By incorporating three specialized gate structures and a memory cell, it overcomes the issue of gradient vanishing, where distant information has little influence on the current time step [27].The structure of the LSTM unit is illustrated in Figure 2.

In Figure 2:x_t represents the input at time t;h_t denotes the hidden layer output at time t;C_t signifies the current memory cell at time t;$\tilde{C}_t$ represents the temporary memory cell;$\otimes$ indicates element-wise multiplication;$\oplus$ signifies element-wise addition; the forget gate determines which information is added to the current memory cell C_t, while the output gate controls the output result of memory cell C_t. The computational process is as follows:

$$\begin{cases} f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = tanh(W_{\tilde{C}} \cdot [h_{t-1}, x_t] + b_{\tilde{C}}) \\ C_t = f_t \otimes C_{t-1} \oplus i_t \otimes \tilde{C}_t \\ o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t = o_t \otimes tanh\, C_t \end{cases} \tag{4}$$

Where:W_f ,W_i ,$W_{\tilde{C}}$,W_o represent the corresponding weight matrices for each module; b_f ,b_i ,$b_{\tilde{C}}$,b_o denote the bias terms; σ is the sigmoid activation function; tanh is the hyperbolic tangent activation function.

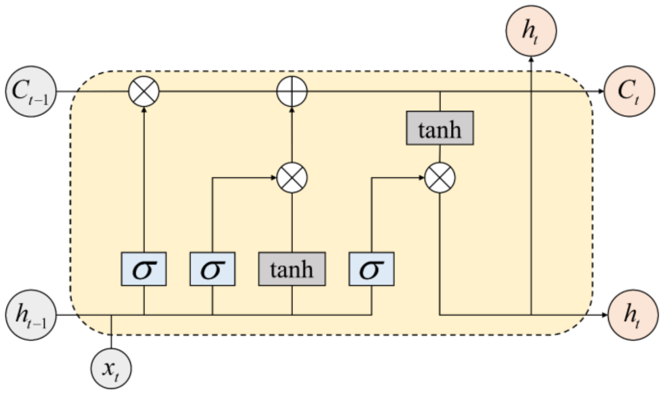

Figure 2. Unit structure of LSTM.

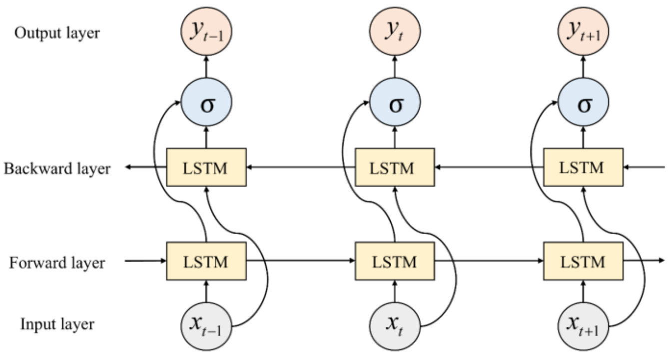

Figure 3. Structure of BiLSTM.

Unlike standard LSTM networks that process sequences only in a forward direction, BiLSTM networks simultaneously extract temporal features in both forward and backward directions. This dual-direction modeling allows the network to capture intrinsic correlations between current signal states and their surrounding temporal context, thereby improving temporal feature completeness. The overall structure of the BiLSTM network is shown in Figure 3.

For the BiLSTM network:

$$\begin{cases} \vec{h}_t = \mathrm{LSTM}(x_t, \vec{h}_{t-1}) \\ \overleftarrow{h}_t = \mathrm{LSTM}(x_t, \overleftarrow{h}_{t-1}) \\ y_t = \sigma(W_y \cdot [\vec{h}_t, \overleftarrow{h}] + b_y) \end{cases} \tag{5}$$

Where: the LSTM unit represents the traditional LSTM network computation process; $\vec{h}_t$ denotes the forward hidden layer state at time step t; $\overleftarrow{h}_t$ denotes the backward hidden layer state; W_y and b_y represent the weight matrix and bias term, respectively.

By fusing forward and backward hidden states, the BiLSTM network effectively captures lateral temporal characteristics of rotor vibration signals, enhancing temporal feature discrimination and robustness under limited and imbalanced training data.

2.3. Attention mechanism

The attention mechanism is introduced to adaptively reweight temporal features extracted by the BiLSTM network, enabling the model to emphasize informative signal segments that are most relevant to rotor fault diagnosis. For UAV rotor vibration signals, discriminative fault-related information is often concentrated in specific temporal regions rather than uniformly distributed across the entire sequence.

Under small-sample conditions, directly modeling all time steps equally may dilute critical fault features. By incorporating an attention mechanism, the proposed framework selectively highlights key temporal features while suppressing redundant information, thereby improving diagnostic robustness with limited training data.

The structure of the attention mechanism is shown in Figure 4. Here,$\vec{h}_1$ to $\vec{h}_t$ represent the forward short-term rotor feature information at each time step in the first hidden layer of the LSTM network, while $\overleftarrow{h}_1$ to $\overleftarrow{h}_t$ represent the backward long-term rotor feature information at each time step in the second hidden layer. The feature information from both layers is summed to obtain the vector set $H = [h_1, h_{2,} \cdots, h_t]$. The attention calculation formula and output results are shown in (6) -(7).

$$M = tanh(H) \tag{6}$$

$$\mathrm{Result} = \mathrm{softmax}(\mathrm{w}^T M) \tag{7}$$

M represents the computed attention weight matrix, which preserves the model's focus levels on different parts of the vibration signal. This weight matrix is multiplied by the model's internal feature parameters w^T, and the result is processed through a Softmax layer to obtain the fault diagnosis outcome.

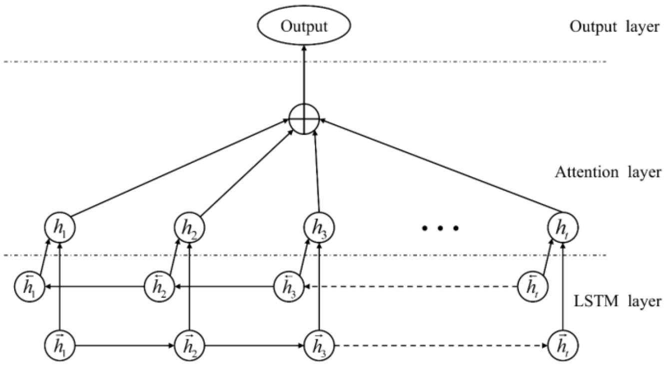

Figure 4. Structure diagram of the attention mechanism.

2.4. Focal loss function

When sample data across different categories is significantly imbalanced, the choice of loss function can greatly affect fault diagnosis performance. Traditional networks use cross-entropy loss (CE Loss) to calculate the difference between the true distribution and the predicted distribution, yielding the training loss value:

$$C_E = -\sum_{m=1}^{n} y_m^t \, lg(y_m^p) \tag{8}$$

where: $y_m^p \in [0,1]$ represents the predicted probability for each class; y_m^t is 0 or 1 indicating whether the sample belongs to the true class.

It is clear that using CE Loss to address sample imbalance has limitations: CE Loss treats the classification cost of each sample category equally, and the total loss during training is the sum of the CE Loss for all samples. However, in the presence of sample imbalance, the loss values of the majority class dominate the total loss. During loss minimization, since the CE Loss for the minority class contributes little to the overall loss, the total loss reaches a low threshold once a large number of majority class samples are correctly classified. This results in insufficient training of the minority class data, leading to low diagnostic accuracy for the network.

To address this issue, an improved Focal Loss is adopted as the training loss function:

$$F_L = -\sum_{m=1}^{n} \alpha_m (1 - y_m^p) y_m^t \, lg(y_m^p) \tag{9}$$

Compared to CE Loss, Focal Loss introduces the following improvements:

1) Introduces the weight factor α_m, which adjusts the weighting between different sample sizes through adaptive assignment, granting higher weight to small-sample data;
2) Introduces a scaling factor $(1 - y_m^p)$ to adaptively adjust the loss of classification samples. This reduces the proportion of loss from large samples while increasing the loss contribution from small samples, enabling the network to focus more on feature learning from small-sample data.

2.5. Basic approach to fault diagnosis

This paper proposes a fault diagnosis model for UAV rotorcraft that integrates a vertical spatial feature extraction module, a horizontal temporal feature extraction module, and an attention mechanism. The model structure is illustrated in Figure 5.

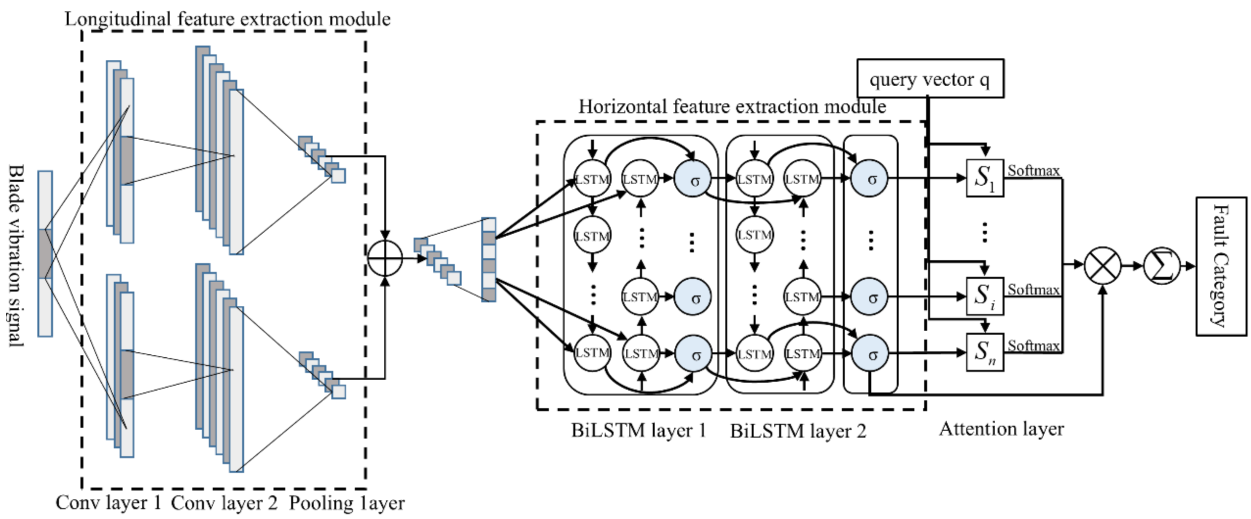

Figure 5. Network structure of CNN-BiLSTM-attention model.

1) Data Preprocessing.

To enable more effective learning from the data, preprocessing operations are performed on the UAV rotor vibration signal data collected from the data acquisition platform. This includes normalization, ratio segmentation, and labeling to construct the UAV rotor vibration dataset.

2) Feature Extraction Module.

To address the issue of insufficient real-time performance in fault diagnosis for drone rotor signals, a longitudinal spatial feature extraction module based on multi-scale features was developed to enable rapid extraction of signal fault characteristics. Building on this foundation, and considering the challenge of inadequate diagnostic accuracy, a lateral temporal feature extraction module based on LSTM was proposed. This module facilitates the comprehensive extraction of fault features in addition to the longitudinal feature extraction module. Furthermore, an attention mechanism was introduced to supplement the model, significantly enhancing diagnostic accuracy during the classification stage.

Specifically, after normalizing the raw drone rotor vibration signals and dividing them into training and test sets, the data is fed into a longitudinal spatial feature extraction module based on multi-scale features. This module performs comprehensive feature extraction for the vibration signals at each time point. The MSCNN consists of a cascade of dual-branch parallel CNN modules, with the core of the CNN network comprising convolutional layers, normalization layers, and global average pooling layers. The convolutional layer applies convolutional filtering to extract spatial features from the single-point signals. The normalization layer follows the convolution operations to prevent gradient explosion and vanishing gradients, significantly reducing the model's error rate. The global average pooling layer then reduces the number of network parameters, accelerating model convergence.

Since the horizontal temporal feature extraction module requires one-dimensional time series input, but the convolution operation produces multi-channel deep spatial features, a flattening layer is needed to convert these features into a one-dimensional form.

The signal then enters the lateral temporal feature extraction module, which is based on a BiLSTM, to capture correlation information across all time points in the vibration signal sequence. An attention mechanism directs the network's focus toward key feature components. The Focal Loss function, incorporating weight and scaling factors, calculates the loss for backpropagation. This enables the model to adaptively assign weights, balancing the differences between sample quantities while amplifying the loss contributions from ambiguous samples, thereby enhancing diagnostic accuracy.

3) Model Diagnostic Testing.

During diagnosis, import the model parameters saved in Step 2, input the preprocessed test set, perform the diagnosis using the model, and output the fault diagnosis results.

3. Case study analysis

The dataset used in this paper is the publicly available UAV rotor failure dataset from the Figshare platform [28]. The data acquisition system consists of a drone platform, an accelerometer, a data acquisition unit, and a computer, as shown in Figure 6.

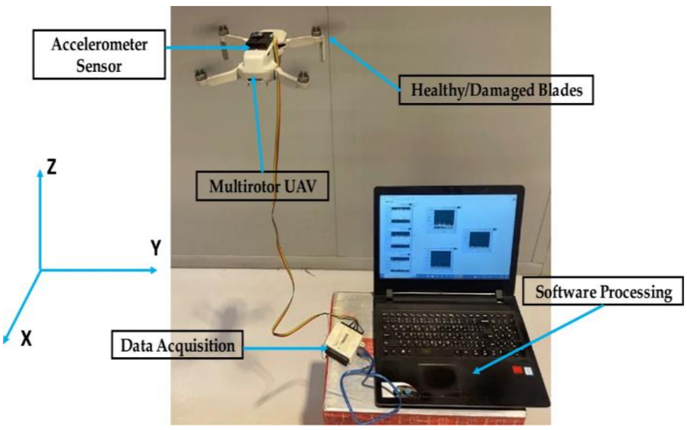

Figure 6. Drone rotor wing data acquisition platform.

In this study, a DJI Mini 2 quadcopter was used as the aerial platform for vibration data acquisition. This compact drone offers reliable stability in hover mode, making it well-suited for reproducible laboratory experiments. Key specifications of the drone are provided in Table 2. To capture vibration signals, an ADXL335 triaxial analog accelerometer was mounted at the drone's geometric center, specifically at the intersection point of the four motor arms. This placement ensures equal sensitivity to vibration along all axes caused by blade imbalance. The specifications of the ADXL335 are detailed in Table 3. The accelerometer interfaces with an NI DAQ-6008 USB data acquisition device, which transmits signals to a PC. The specifications of the DAQ system are listed in Table 4.

Table 1. Specifications of the DJI Mini 2 combo UAV.

Parameter	Value
Takeoff Weight	249 g
Max Flight Time	31 minutes
Maximum Hover Time	29 minutes
Maximum Flight Speed	16 m/s
GNSS	GPS + GLONASS + Galileo
Max Altitude	4000 m above sea level
Dimensions (Folded)	138 × 81 × 58 mm
Rotor Configuration	4 rotors (quadcopter)

Table 2. Specifications of the three-axis accelerometer ADXL335.

Parameter	Value
Axes	X, Y, Z
Output Type	Analog
Measurement Range	±3 g
Sensitivity	300 mV/g (typical)
Power Supply	1.8 V to 3.6 V
Operating Temperature	−40°C to +85°C
Bandwidth (typical)	1600 Hz

Table 3. Specifications of DAQ system.

Parameter	Value
Analog Inputs	8 channels (12-bit)
Sample Rate	10 kS/s (max)
Analog Output	2 channels (8-bit)
Digital I/O	12 lines
Communication	USB 2.0
Software Compatibility	LabVIEW, MATLAB, etc.

Table 4. Label of the rotor fault samples.

Sample Size	Sample Length	Fault Category	Label
50	2048	Healthy	1
50	2048	Damaged Bottom Right Rotor	2
50	2048	Damaged Top Right Rotor	3
50	2048	Unbalanced Bottom Right Rotor	4
50	2048	Unbalanced Top Right Rotor	5

The experimental platform acquires vibration signals under five rotor conditions: normal, right lower damage, right upper damage, right lower imbalance, and right upper imbalance. Each fault is independently applied to a single rotor. The fault configuration and rotor numbering scheme are shown in Figure 9. During signal acquisition, the sampling frequency was set to 2048 Hz with a 1-second sampling duration. Fifty experimental sets were conducted for each condition (normal, right lower damage, right upper damage, right lower imbalance, and right upper imbalance), resulting in a total of 250 data sets. The correspondence between fault categories and their labels is shown in Table 5.

Under normal operating conditions, the expected vibration value of the UAV rotor is 0 m/s², and the collected vibration signals therefore fluctuate around 0 m/s². To eliminate the influence of scale differences in the raw data on the fault diagnosis model, Z-score normalization is applied as a preprocessing step, transforming the data to follow a standard normal distribution. The normalization process is defined as:

$$x^* = (x - \mu)/\sigma \tag{10}$$

Where: x denotes the input data, μ represents the mean of the data, σ denotes the standard deviation, and x^* is the normalized output.

Conduct performance comparisons of different methods, model testing under noisy conditions, and diagnostic analysis experiments on unbalanced datasets using this small-sample dataset.

3.1. Fault diagnosis experiment for UAV rotors under small-sample conditions

To validate the advantages of the proposed model in diagnosing faults in UAV rotor signals, comparative experiments and analyses were conducted between the proposed model and commonly used models. Among these, the CNN and CNN-BiLSTM models share the same CNN structure as the proposed model, while the CNN-BiLSTM, BiLSTM, LSTM, and DNN models have the same hidden layer node counts and structures as the proposed model.

To validate that the proposed model maintains robust fault diagnosis performance on small-sample datasets, we selected 45, 35, 25, 15, and 5 samples for each of the five fault types (each with 50 samples) to form the training set, with the remaining samples serving as the validation set. During model training, 5-fold cross-validation was employed: the dataset was evenly divided into five portions, with four portions used as training data and one portion as validation data in each iteration. The average accuracy across the five training iterations was then calculated. To mitigate random errors, each experiment was repeated 50 times. The test results are shown in Figure 7 and Table 5.

As shown in Figure 7, the proposed model achieves the highest fault diagnosis accuracy across varying training sample sizes, demonstrating a stable upward trend as the sample size increases. When the training sample size reaches 35, the proposed model exhibits saturated diagnostic performance, significantly outperforming comparison methods such as CNN-BiLSTM and MSCNN. This indicates that the proposed framework can effectively exploit additional training data while maintaining stable generalization behavior under limited-sample conditions.

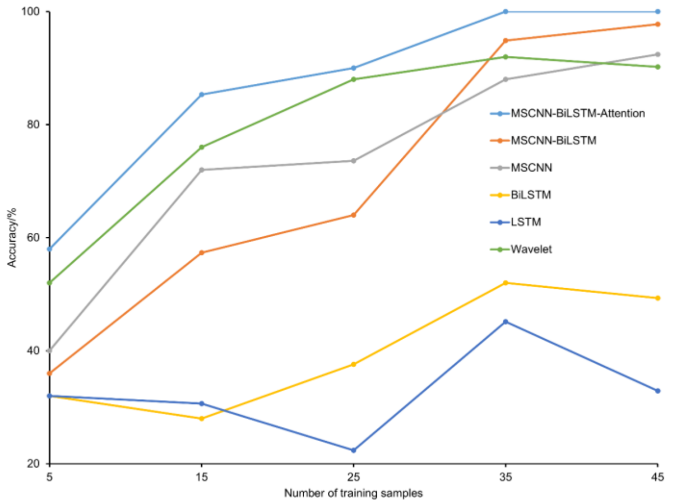

Figure 7. Diagnostic accuracy of different methods under different quantities of samples.

Table 5. Diagnostic accuracy under different methods (45 training samples).

Model	Maxa_{cc}/%	Mina_{cc}/%	Avga_{cc}/%
MSCNN-BiLSTM-Attention	100	100	100
MSCNN-BiLSTM	100	93.33	97.78
MSCNN	97.78	80	92.44
Wavelet	95.56	84.44	90.22
BiLSTM	60	42.22	49.33
LSTM	40	22.22	32.89

As shown in Table 5, the MSCNN-based longitudinal spatial feature extraction module demonstrates relatively strong performance. This advantage mainly stems from the ability of convolutional structures to capture localized and discriminative vibration patterns with relatively few parameters, which is particularly suitable for small-sample learning. Compared to the LSTM-based lateral temporal feature extraction module, its diagnostic accuracy improved from 32.89% to 92.44%. This improvement can be attributed to the strong feature extraction capability and relatively low model complexity of the CNN-based longitudinal spatial feature extraction module. The MSCNN further enhances the longitudinal spatial feature extraction performance with only a marginal increase in model complexity. In contrast, the LSTM module faces challenges in feature extraction from raw temporal data, and longer signal sequences significantly increase computational time for its hidden layer units. This is because raw vibration signals contain strong noise and redundant temporal information, making it difficult for recurrent structures to directly learn discriminative features without prior spatial abstraction. By integrating both modules, the model benefits from complementary advantages in diagnostic real-time performance and accuracy. Even with small sample sizes, the diagnostic accuracy exhibits low variance, demonstrating the model's excellent stability and robustness in feature learning.

A comparison of the first two rows in Table 5 reveals that the attention mechanism enhances the distinguishability of key fault features under limited sample conditions. The WA method, based on signal analysis, achieves approximately 50% accuracy with a small number of training samples. As the number of training samples increases to 35, the accuracy reaches about 92%, but its rate of improvement is significantly slower than that of deep learning methods. Additionally, this method requires substantial expertise, making it difficult to determine optimal threshold parameters in experiments. As a result, its performance is noticeably inferior to that of deep learning methods when ample training samples are available, indicating inherent limitations.

Comparative analysis reveals that the longitudinal spatial extraction module offers a simple model with strong feature extraction capabilities using few parameters, but it lacks sufficient ability to analyze the root causes of faults. The lateral temporal feature extraction module effectively captures historical temporal features of fault occurrences, but it struggles when applied to raw rotor vibration signals. The proposed model utilizes the longitudinal spatial extraction module for feature extraction, followed by input into the lateral temporal feature extraction module to supplement the extracted features. The attention mechanism further enhances the importance of key features, significantly reducing the dimensionality of signal features after MSCNN processing. By selectively emphasizing fault-sensitive temporal segments, the attention mechanism suppresses redundant information and alleviates the computational burden of the subsequent BiLSTM module. This alleviates the computational burden on the BiLSTM while emphasizing fault features and suppressing redundant information through attention weights, thereby improving the fault diagnosis performance of the proposed model. These observations demonstrate that the improved diagnostic performance is not solely due to model complexity, but results from the coordinated design of spatial–temporal feature extraction and attention-based feature refinement.

The features from the output layer of the proposed model are visualized using t-Distributed Stochastic Neighbor Embedding (t-SNE), with the results shown in Figure 8.

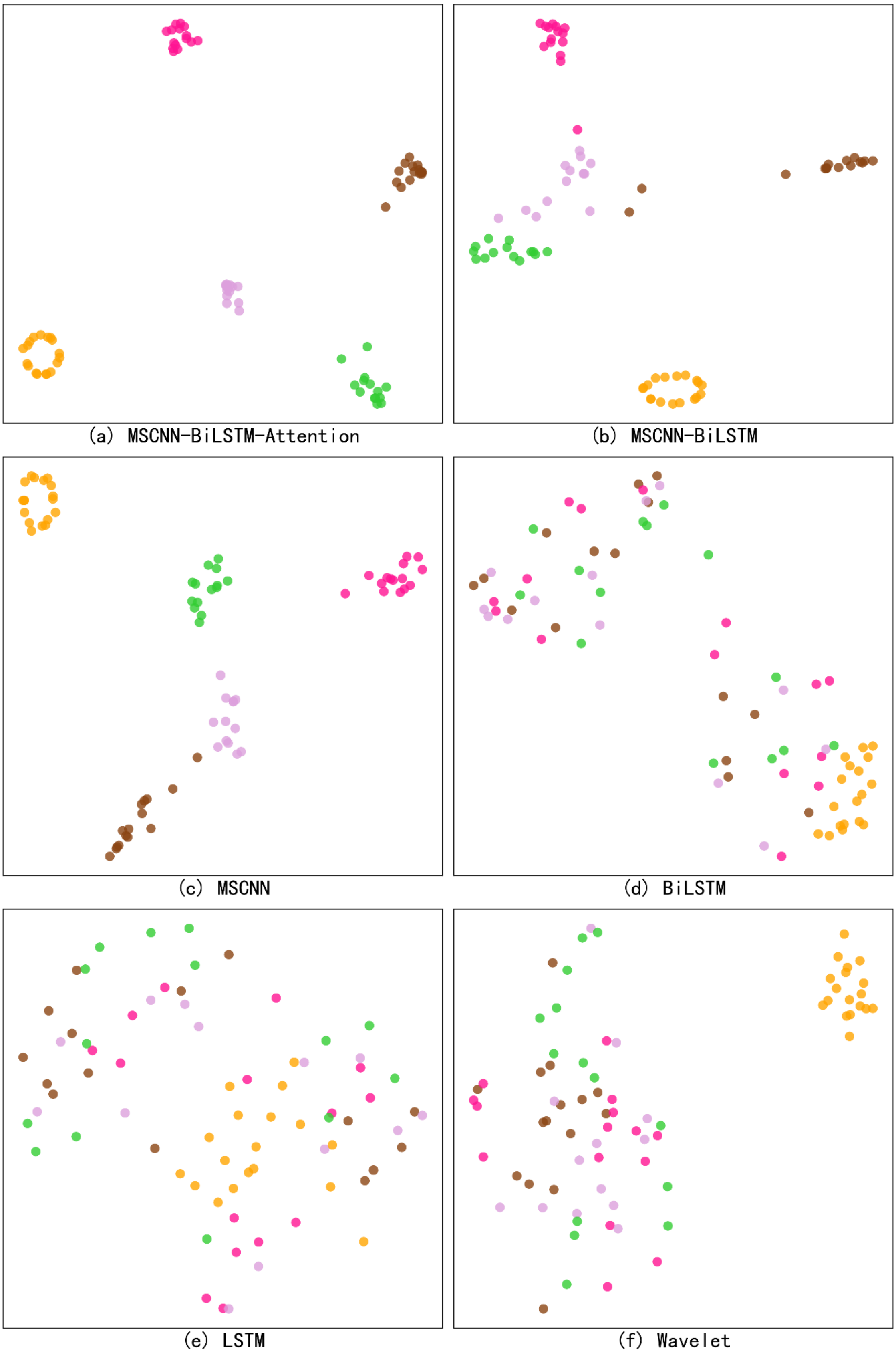

Figure 8. Diagnostic accuracy of different methods under different quantities of samples.

Comparing the two-dimensional spatial distributions of output layer features across different methods in Figure 8 reveals that our model exhibits clear separations between output classes. The clustering effect of the extracted features outperforms traditional WA signal analysis methods and surpasses network models such as MSCNN and BiLSTM. As a result, the multi-scale features extracted by the proposed model more accurately capture the differences between various fault signals, significantly enhancing fault diagnosis accuracy.

3.2. Noise resistance analysis under small samples

During UAV operations, flight environments are often complex and harsh, causing sensor signals to be contaminated by noise, which negatively impacts the extraction of fault features. To simulate real-world noise interference under small-sample conditions, this study selected 35 data points per fault category for the training set and 15 data points for the test set. Gaussian white noise with signal-to-noise ratios (SNR) ranging from 30 dB to 0 dB was added to both sets. The proposed model was then applied for fault detection. Experimental results under different SNR levels are shown in Table 1 and Figure 9.

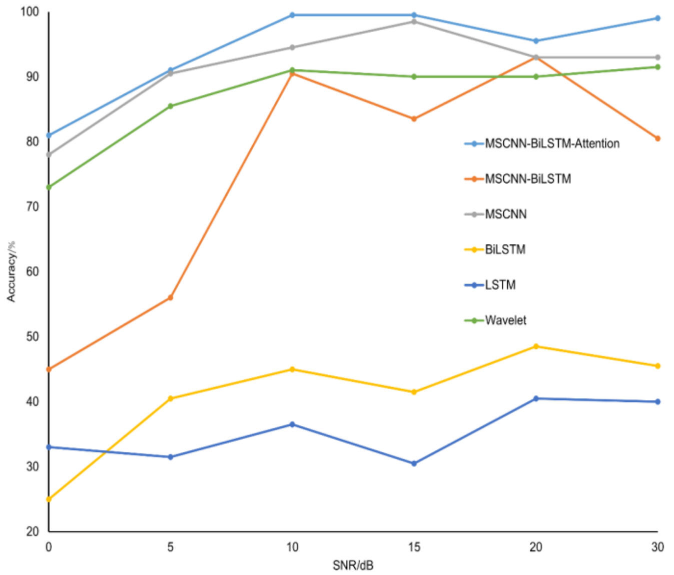

Figure 9. Diagnostic accuracy of different methods under different single-noise ratios of samples.

As shown in Figure 9, the higher the signal-to-noise ratio, the greater the diagnostic accuracy for fault data. When the signal-to-noise ratio exceeds 10 dB, the diagnostic accuracy of the proposed model exceeds 95%, demonstrating robust noise resistance under small sample conditions.

Table 6. Diagnostic accuracy under different single-noise ratios.

SNR/dB	a_{cc}/%	Time/s
30	99.00	12.7
20	95.50	12.7
15	99.50	12.4
10	99.50	12.5
5	91.00	12.6
0	81.50	12.5

3.3. Diagnostic analysis under different imbalance levels

In real-world flight environments, certain rotor failures in UAVs occur infrequently, with low probability, resulting in a limited number of recorded samples. Additionally, during data collection, sensors may fail to capture information accurately or in a timely manner when certain faults occur, and the difficulty of capturing data for different types of faults varies. Therefore, fault diagnosis analysis under imbalanced sample conditions is necessary to validate the practical applicability of this model.

This paper defines imbalance as follows:N_"normal" represents the number of normal data samples, and N_"fault" denotes the total number of fault sample data in the training set. The imbalance ratio U is calculated as shown in Equation (10).

$$U = \frac{N_{fault}}{N_{normal}} \tag{11}$$

In this study, for the balanced fault diagnosis experiments, 40 samples from each class were selected as the training set. For the imbalanced fault diagnosis experiments, only 5 samples from each class were used for training. In all experimental settings, 10 samples per class were fixed as the test set.

Considering that sample data from normal operating conditions is the most abundant, this paper classifies imbalanced sample types as follows: Type 0 includes a scenario representing standard balanced sample data with an imbalance ratio of 0; Type 1 includes four groups with one imbalanced fault data type, yielding an imbalance ratio of 0.10; Type 2 includes six groups with two imbalanced fault data types, resulting in an imbalance ratio of 0.20; Type 3 includes four groups with three types of imbalanced fault data, with an imbalance ratio of 0.30; Type 4 includes one group with four types of imbalanced fault data, with an imbalance ratio of 0.40, as shown in Table 7. A higher imbalance ratio indicates greater disparity among the data and places higher demands on the fault diagnosis model.

For each imbalanced data type, separate training and testing were conducted. For the five imbalance levels (0, 0.1, 0.2, 0.3, and 0.4), a total of 160, 125, 90, and 55 groups were proportionally sampled from the five fault categories for training, with 10 groups selected as test data. The proposed model was then applied for fault detection, and the experimental results are shown in Figure 10.

Table 7. Diagnostic accuracy under imbalanced samples.

No.	U	Groups	Avga_{cc}/%
0	0.00	1	100.00
1	0.10	4	99.50
2	0.20	6	99.33
3	0.30	4	96.50
4	0.40	1	86.00

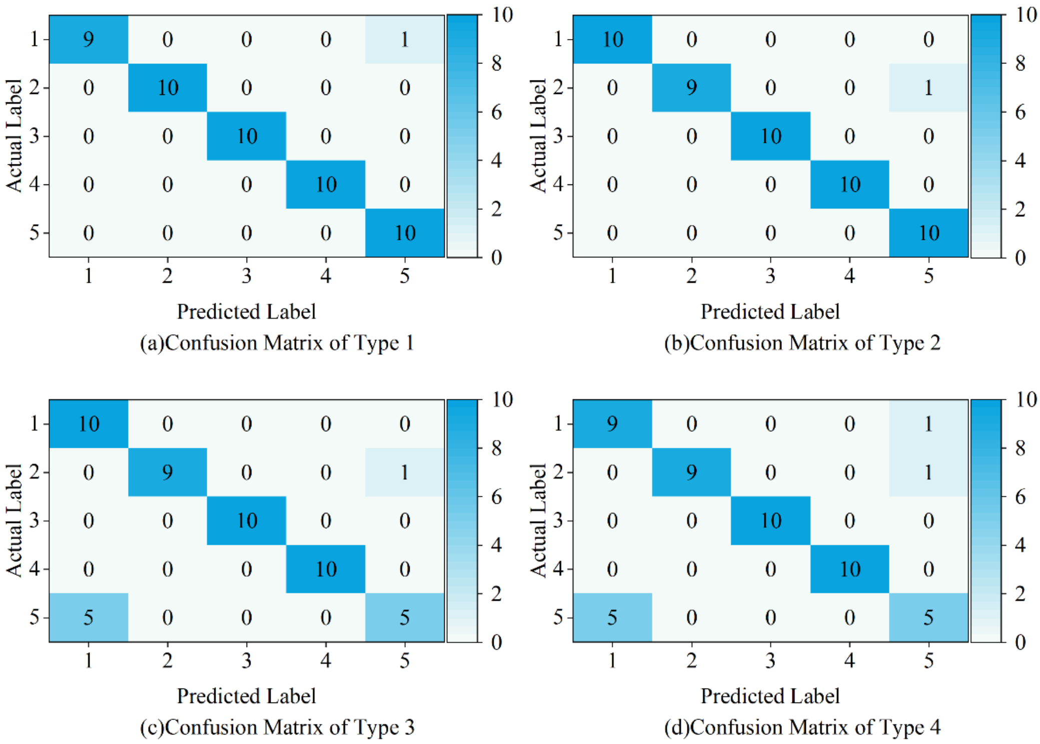

Figure 10. Example of confusion matrix for diagnosis results under different types of imbalance.

Figures 10(a)-(d) present the confusion matrices for the diagnostic results under imbalance types 1 to 4, respectively. Taking Figure 10(a) as an example, imbalance type 1 is specifically configured with 15 sets of right-lower-quadrant damaged fault samples, while normal data and the other three fault types each consist of 50 sets. The imbalance ratio in this scenario is calculated as 0.10 using Equation (11). The fault diagnosis performance of the model shown in Figure 10 is depicted, where the horizontal axis represents the model's diagnostic results for the test data, and the vertical axis indicates the actual fault type of the test data. Diagonal elements represent accurate fault diagnosis by the model. The darker the color of an element in the matrix, the more test data results fall within that region. It can be observed that the right-lower damaged fault, right-upper damaged fault, right-lower unbalanced fault, and right-upper unbalanced fault are all correctly diagnosed. Among the 10 sets of normal test data, only one set is misclassified as a right-upper unbalanced fault.

As shown in Figure 10, the proposed model achieves high diagnostic accuracy across various test conditions, although accuracy decreases as the imbalance degree increases. Misclassifications at higher imbalance levels primarily occur in fault category 5, while diagnoses for other fault categories remain nearly flawless. This demonstrates the model's effectiveness in diagnosing rotor faults in UAVs.

Table 7 shows that unbalanced datasets reduce the model's fault diagnosis accuracy. Under balanced conditions, the accuracy is 100%. At an imbalance ratio of 0.2, the model's accuracy decreases by approximately 1%, maintaining over 99% accuracy. As the imbalance ratio increases, the model's accuracy continues to decline. However, even at the highest imbalance ratio of 0.4, the model still achieves 86% diagnostic accuracy. This demonstrates that the model remains accurate and effective in diagnosing rotor failures in UAVs, even under imbalanced data conditions.

3.4. Analysis of noise resistance under imbalanced samples

For imbalanced data, this study selected 50 sets of normal data, 20 sets of right-lower-quadrant damage faults, 40 sets of right-upper-quadrant damage faults, 10 sets of right-lower-quadrant imbalance faults, and 30 sets of right-upper-quadrant imbalance faults, totaling 150 sets of data as the overall dataset. The data was proportionally divided into an 80% training set and a 20% test set. Gaussian white noise with signal-to-noise ratios (SNR) ranging from 30 dB to 0 dB was added to both sets. Subsequently, the proposed model was applied for fault detection. Experimental results under different SNRs are presented in Table 8.

Table 8. Diagnostic accuracy under imbalanced samples with different signal-to-noise ratios.

SNR/dB	a_{cc}/%	Time/s
30	96.00	12.7
20	100.00	12.7
15	96.00	12.4
10	98.00	12.5
5	86.00	12.6
0	70.50	12.5

Table 8 shows that when the SNR ranges from 10 to 30 dB, the model's diagnostic accuracy remains stable between 96% and 100%, demonstrating its exceptional fault identification capability under conditions of good signal quality. At an SNR of 5 dB, the accuracy drops to 86%, but it still maintains a relatively high level, confirming the model's robust noise resistance under imbalanced sample conditions.

3.5. Experimental analysis of model generalization ability

To further evaluate the generalization capability of the proposed model, supplementary experiments were conducted using a UAV rotor fault dataset provided by Beihang University. This dataset was collected by the Reliable Flight Control Group (Rfly) from both the RflySim simulation platform and real flight tests. During the real-flight experiments, three multirotor UAVs with different diagonal dimensions were employed to enhance the dataset's generality: the X200 (200 mm diagonal, 1.054 kg), the Longzhu X450 (450 mm diagonal, 2.084 kg), and the Longyan X680 (680 mm diagonal, 4.068 kg).

In the Software-in-the-Loop (SIL) and Hardware-in-the-Loop (HIL) simulations, the dynamic model parameters of the multirotor UAVs—including mass, moment of inertia, and motion characteristics—were based on the Droneyee X450 platform. The dataset encompasses 11 common fault types under six flight conditions to capture diverse motion characteristics of multirotor aircraft under fault scenarios.

In this study, the rotor defect condition (1–4) was selected, where the indices (1–4) indicate the range of faulty rotors. From the vibration signal sequences, 50 samples were extracted for each fault category, with each sample sequence containing 2048 data points, resulting in 200 samples used for training. Another 200 samples were selected under the same configuration for testing.

To further investigate the model's diagnostic capability under varying training sample sizes, the number of training samples was set to 200, 160, 120, and 80, respectively. Fault diagnosis experiments were then conducted under each configuration, and the results are presented in Table 9.

As shown in Table 9, when the training set size reached 160 samples or more, the proposed model achieved a diagnostic accuracy exceeding 90%. Even when each fault category contained only 20 training samples, the diagnostic accuracy remained above 85%. These results demonstrate the strong generalization capability and robustness of the proposed model across different sample scales.

Table 9. Diagnosis accuracy and elapsed time of different training samples.

The amount of sample data in the training set	a_{cc}/%	Time/s
200	92.14	15.3
160	90.56	14.4
120	88.33	13.3
80	85.74	12.8

4. Conclusion

To address the challenges of small-sample and class-imbalanced conditions in UAV rotor fault diagnosis, this study proposed an MSCNN-BiLSTM-Attention-based diagnostic framework. The proposed approach integrated multi-scale spatial feature extraction, bidirectional temporal dependency modeling, and attention-based feature weighting to enhance diagnostic robustness under complex conditions. Key innovations include:

1) Constructing a longitudinal spatial feature extraction module based on MSCNN, which enables efficient feature extraction from single-instance vibration signals while significantly reducing the number of network parameters;
2) A BiLSTM-based horizontal temporal feature extraction module is designed to fully exploit temporal correlation

features across multiple time steps, complementing the spatial features through integrated fusion;

3) An attention mechanism combined with multi-scale feature extraction is incorporated to enhance core feature learning from limited datasets, thereby maximizing the representation of inter-class distinctions;
4) The Focal Loss function is employed to adaptively adjust sample weights, enabling the network to focus more effectively on rare and easily misclassified samples during imbalanced data training.

Experimental results demonstrate that under conventional small-sample conditions (45 training samples per class), the proposed method achieves a diagnostic accuracy of 100%, representing a 9-percentage-point improvement over traditional approaches. Even under extremely limited sample conditions (15 training samples per class), the model maintains an accuracy of 85.33%. Furthermore, in noisy environments with signal-to-noise ratios as low as 5 dB, the diagnostic accuracy remains above 86%. Under severe data imbalance with an imbalance ratio of 0.4, the model still achieves an 86% accuracy rate. These findings comprehensively verify the proposed method's robustness and stability in UAV rotor fault diagnosis, demonstrating its superior adaptability and diagnostic performance under challenging conditions such as small samples, strong noise, and data imbalance. However, this study still has several limitations:

1) The experimental data were derived exclusively from UAV rotor fault scenarios. The model's generalization ability to other types of UAV faults, such as motor or sensor malfunctions, requires further validation;
2) The current research focuses primarily on single-rotor fault diagnosis, whereas the diagnosis of simultaneous multi-rotor faults remains to be explored.

Future work will focus on developing practical and targeted strategies to extend the applicability of the proposed diagnostic framework. Specifically:

1) Expanding to diverse UAV fault types: The model will be adapted to diagnose additional fault categories, such as motor failures, sensor malfunctions, and battery anomalies, by collecting representative datasets and designing fault-specific feature extraction modules. Transfer learning and data augmentation techniques, including WGAN-based synthetic data generation, will be employed to overcome small-sample limitations;
2) Multi-rotor fault diagnosis: The framework will be extended to handle simultaneous faults in multiple rotors by designing multi-branch network architectures and integrating inter-rotor correlation modeling. Techniques such as attention-based feature fusion and sequence alignment across rotors will be explored to enhance diagnostic accuracy under coupled fault scenarios;
3) Robustness under complex flight conditions: Experiments will be conducted under varying operating environments, including different flight speeds, payloads, and noisy conditions, to evaluate and improve model generalization. Adaptive loss functions and noise-resilient feature extraction methods will be further optimized.

References

1. Hua W, Chen Q. A survey of small object detection based on deep learning in aerial images. *Artificial Intelligence Review* 2025; 58(6): 162. https://doi.org/10.1007/s10462-025-11150-9.
2. Guang HE, Huaping H, Xiangke W, Li YU. Transition Mode Flight Control Method of TRUAVs Based on Hybrid Control of Model Calibration MPC and PID. *ROBOT* 2025; 47(3): 394-404. https://doi.org/10.13973/j.cnki.robot.250001.
3. Bartoszewski B, Jarzyna K, Baranowski J. Diagnosis of Mechanical Rotor Faults in Drones Using Functional Gaussian Mixture Classifier. *Aerospace* 2024; 11(9): 743. https://doi.org/10.3390/aerospace11090743.
4. Al-Haddad L, Jaber A, Neranon P, Al-Haddad S. Investigation of Frequency-Domain-Based Vibration Signal Analysis for UAV Unbalance Fault Classification. *Engineering and Technology Journal* 2023; 41: 1-9. https://doi.org/10.30684/etj.2023.137412.1348.
5. Shang X, Li W, Yuan F, Zhi H, Gao Z, Guo M, Xin B. Research on Fault Diagnosis of UAV Rotor Motor Bearings Based on WPT-CEEMD-CNN-LSTM. *Machines* 2025; 13(4): 287. https://doi.org/10.3390/machines13040287.
6. Ai S, Song J, Cai G, Zhao K. Active Fault-Tolerant Control for Quadrotor UAV against Sensor Fault Diagnosed by the Auto Sequential Random Forest. *Aerospace* 2022; 9(9): 518. https://doi.org/10.3390/aerospace9090518.

7. Moloi K, Yusuff AA. A Wavelet-Neural Network-Based Technique for Fault Diagnostics in Power System. In: *2020 7th International Conference on Soft Computing & Machine Intelligence(ISCMI);* Stockholm, Sweden. 2020. p. 131-5.
8. Zhang Y, Li S, He Q, Zhang A, Li C, Liao Z. An Intelligent Fault Detection Framework for FW-UAV Based on Hybrid Deep Domain Adaptation Networks and the Hampel Filter. *International Journal of Intelligent Systems* 2023; 2023(1): 6608967. https://doi.org/10.1155/2023/6608967.
9. Zhang Y, Li S, Zhang A, An X. FW-UAV fault diagnosis based on knowledge complementary network under small sample. *Mechanical Systems and Signal Processing* 2024; 215: 111418. https://doi.org/10.1016/j.ymssp.2024.111418.
10. Zhou S, He Z, Chen X, Chang W. An Anomaly Detection Method for UAV Based on Wavelet Decomposition and Stacked Denoising Autoencoder. *Aerospace* 2024; 11(5): 393. https://doi.org/10.3390/aerospace11050393.
11. Wen R, Lu G, Mi Y, Wu B, Zhang Y, Yang L. Autoencoder-Based Deep Learning for UAV RF Signal Recognition and Classification. In: *Proceedings of the 2024 7th International Conference on Artificial Intelligence and Pattern Recognition*. New York, USA. Association for Computing Machinery; 2025. p. 343-50.
12. Zhou H, Zhou J, Li Y, Cai F. Research on the condition monitoring method of unmanned aerial vehicle based on improved multivariate state estimation technique. *Scientific Reports* 2025; 15(1): 9511. https://doi.org/10.1038/s41598-025-93343-4.
13. Fajar M, Hakim TMI, Wirawan A, Suseno PAP, Soemaryanto AR, Pramutadi AM. Comparative Analysis of LSTM Variants for Fault Detection and Classification in Aircraft Control Surfaces. *International Journal of Prognostics and Health Management* 2025; 16(1). https://doi.org/10.36001/ijphm.2025.v16i1.4231.
14. Han B, Jia S, Liu G, Wang J. Imbalanced Fault Classification of Bearing via Wasserstein Generative Adversarial Networks with Gradient Penalty. *Shock and Vibration* 2020; 2020(1): 8836477. https://doi.org/10.1155/2020/8836477.
15. Zhang Y, Li S, Gu Y, He Q, Zhou P, Zhang A. UAV fault diagnosis based on collaborative sharing of generic and task-oriented features. *Expert Systems with Applications* 2026; 296: 128940. https://doi.org/10.1016/j.eswa.2025.128940.
16. Langenkämper D, van Kevelaer R, Möller T, Nattkemper TW. Gan-Based Synthesis of Deep Learning Training Data for Uav Monitoring. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2020 Aug 1;43B1:465-9. https://doi.org/10.5194/isprs-archives-xliii-b1-2020-465-2020.
17. T. T, Low KH, Ng BF. Actuator fault detection and isolation on multi-rotor UAV using extreme learning neuro-fuzzy systems. *ISA Transactions* 2023; 138: 168-85. https://doi.org/10.1016/j.isatra.2023.02.026.
18. Song J, Shang W, Ai S, Zhao K. Model and Data-Driven Combination: A Fault Diagnosis and Localization Method for Unknown Fault Size of Quadrotor UAV Actuator Based on Extended State Observer and Deep Forest. *Sensors* 2022; 22(19): 7355. https://doi.org/10.3390/s22197355.
19. Chen H, Wei J, Huang H, Yuan Y, Wang J. Review of imbalanced fault diagnosis technology based on generative adversarial networks. *Journal of Computational Design and Engineering* 2024; 11(5): 99-124. https://doi.org/10.1093/jcde/qwae075.
20. Liu L, Zhao Y, Hu Y, Ma Y, Guo Z. Lightweight mechanical equipment fault diagnosis framework based on GCGAN-MDSCNN-ICA model. *Scientific Reports* 2025; 15(1): 4911. https://doi.org/10.1038/s41598-025-89576-y.
21. Yang P, Wen C, Geng H, Liu P. Intelligent Fault Diagnosis Method for Blade Damage of Quad-Rotor UAV Based on Stacked Pruning Sparse Denoising Autoencoder and Convolutional Neural Network. *Machines* 2021; 9(12): 360. https://doi.org/10.3390/machines9120360
22. Tao W, Li X, Liu J, Li Z. Multi-scale attention network (MSAN) for track circuits fault diagnosis. *Scientific Reports* 2024; 14(1): 8886. https://doi.org/10.1038/s41598-024-59711-2.
23. Zhang J, Zou J, Su Z, Tang J, Kang Y, Xu H, Liu Z, Fan S. A class-aware supervised contrastive learning framework for imbalanced fault diagnosis. *Knowledge-Based Systems* 2022; 252: 109437. https://doi.org/10.1016/j.knosys.2022.109437.
24. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal Loss for Dense Object Detection.In: *2017 IEEE International Conference on Computer Vision (ICCV)*; Venice, Italy. 2017. p. 2999-3007.
25. Huang G, Liu Z, Maaten L van der, Weinberger KQ. Densely Connected Convolutional Networks. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; Hawaii, USA. 2017. p. 4700-4708.
26. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks* 2005; 18(5): 602-10. https://doi.org/10.1016/j.neunet.2005.06.042.

27. Hochreiter S, Schmidhuber J. Long Short-Term Memory. *Neural Computation* 1997; 9(8): 1735-80. https://doi.org/10.1162/neco.1997.9.8.1735.
28. Al-Haddad LA, Jaber AA, Hamzah MN, Kraiem H, Al-Karkhi MI, Flah A. Multiaxial vibration data for blade fault diagnosis in multirotor unmanned aerial vehicles. *Scientific Data* 2025; 12(1): 1383. https://doi.org/10.1038/s41597-025-05692-4.