EAI Endorsed Transactions
on Energy Web

Research Article EAI.EU

Fine-grained fault diagnosis methods for offshore wind turbine gearboxes driven by artificial intelligence—by multi-source data analysis and deep learning

Hongfeng Chen[1], Chao Chen[1], Xingdu Li[1], Huihui Han[2], Xinfa Shi[3,*], Xin Wang[2]

[1]Datang Guoxin Binhai Offshore Wind Power Generation Co., Ltd., Nanjing 210000, China
[2]China Datang Group new energy science and Technology Research Institute Co., Ltd., East China Electric Power Test & Research Institutein, Hefei 230088, China
[3]Guangzhou Mechanical Engineering Research Institute Co., Ltd., Guangzhou 510000, China

Abstract

INTRODUCTION: Offshore wind turbine gearboxes operate under complex conditions and are highly prone to faults. Traditional single-source diagnostic methods are sensitive to noise and load fluctuations, limiting diagnostic reliability.
OBJECTIVES: This study aims to improve the diagnostic accuracy and recognition performance of gearbox fault categories.
METHODS: Multi-source monitoring data were preprocessed and fused, followed by feature optimization and construction of a deep learning-based diagnosis model for fault detection and classification.
RESULTS: The proposed model achieved accuracies of 0.951, 0.947, and 0.938 under different load conditions, with Area Under the Curve AUC values above 0.96, outperforming benchmark models.
CONCLUSION: The proposed method improves diagnostic accuracy, robustness, and engineering applicability for offshore wind turbine gearbox fault diagnosis.

Keywords: offshore wind power, gearbox, multi-source data fusion, artificial intelligence, deep learning, intelligent fault diagnosis

Received on 15 December 2025, accepted on 03 May 2026, published on 12 August 2026

doi: 10.4108/ew.14193

1. Introduction

In recent years, with the continued advancement of global energy transition toward green and low-carbon systems, offshore wind power has become a key direction in new energy development due to its abundant wind resources, high annual equivalent utilization hours, low occupation of land resources, and suitability for large-scale deployment [1, 2]. In particular, under the "dual carbon" goals and the coordinated development of the marine economy, offshore wind power is gradually shifting from nearshore demonstration projects to deep-sea, high-capacity, and clustered developments. Along with the rapid expansion of installed capacity, higher requirements have been placed on long-term operational reliability and life-cycle economic performance. In all subsystems, the transmission system plays a vital role in offshore wind turbines. Its health directly affects operation safety, energy conversion efficiency and maintenance cost control.

In the transmission system of offshore wind turbines, the gearbox is responsible for converting the low-speed and high-torque input of the rotor into the high-speed and low-torque output of the generator. It is the core component connecting pneumatic energy capture and electric energy conversion [3, 4]. Once the gearbox has some faults, such as pitting, cracking, bearing damage or lubrication degradation, it will lead to abnormal vibration, reduced transmission efficiency and power fluctuation. These failures may also lead to the chain damage of coupling, main bearing and generator, which will eventually lead to prolonged downtime or serious equipment failure [5]. Compared with onshore wind turbines, offshore units operate in an environment characterized by

*Corresponding author. Email: shixinfa136@outlook.com

high humidity, high salinity, strong winds and wave loads, temperature cycles and complex coupled operating conditions. Under these conditions, the gearbox is more susceptible to corrosion, lubrication deterioration, structural fatigue accumulation and random impact load. Therefore, their failure mechanism becomes more complicated and the degradation process is more difficult to observe [6, 7]. In addition, the maintenance activities of offshore wind farms are limited by sea conditions, weather windows, ship availability, lifting resources and personnel access costs. Therefore, compared with onshore systems, unexpected failures usually lead to a significant increase in maintenance costs and a greater energy loss [8]. Therefore, gearbox fault diagnosis is a condition monitoring problem in offshore wind system. It is also directly related to operation reliability, maintenance strategy optimization and return on investment.

At present, the research on condition monitoring and fault diagnosis of offshore wind turbine gearbox can be roughly divided into three categories. The first category includes traditional methods based on a single physical signal. These methods mainly rely on vibration analysis, oil condition monitoring, temperature and pressure monitoring and SCADA data analysis [9, 10]. Vibration analysis is sensitive to abnormal gear engagement, bearing impact and local damage. Therefore, it is suitable for identifying typical mechanical faults. Oil analysis is more effective in reflecting lubrication degradation, wear particles and long-term deterioration trend. Temperature, pressure and SCADA data are easy to obtain and provide long-term coverage. They are more suitable for operation state evaluation and early abnormal warning. However, these methods rely heavily on a single information source. They have limited adaptability under strong noise and highly variable offshore operating conditions. When wind speed, wind direction, turbulence intensity and control strategy change frequently, operational factors and fault characteristics are often coupled in the signal. This may lead to the overlap between different fault modes in the feature space, and increase the risk of misdiagnosis and missed detection [11].

The second category consists of data-driven diagnosis methods based on machine learning or deep learning. These methods usually perform feature extraction and pattern recognition on vibration, SCADA or electrical signals. They use models such as support vector machine, random forest, convolutional neural network or time network for fault classification and state identification [12]. Compared with the traditional expert-based methods, these methods provide stronger capabilities in automatic feature learning, complex pattern recognition and nonlinear mapping. They can significantly improve the accuracy and automation level of gearbox fault diagnosis. However, most existing researches still rely on single-source input or simple data connection and voting strategies. They lack in-depth modeling of the correlation, complementarity and operational adaptability between heterogeneous data sources. In the offshore wind scene, early damage, lubrication degradation and compound failure usually do not show enough obvious characteristics in any single sensor channel. Therefore, the single channel or shallow fusion strategy is still not enough to achieve stable and high-precision fault identification.

Third, the research scope focusing on fine-grained diagnostic requirements is expanding. In addition to binary fault detection, these studies further regard fault type, fault location, fault severity and degradation trajectory as more detailed tasks [13]. This direction is more in line with the actual offshore wind operation and maintenance requirements. Offshore wind farms need fault alarms, but also diagnostic information that can support predictive maintenance, spare parts scheduling, maintenance window planning and maintenance activity priority sequencing. However, the existing research in this field is still limited. On the one hand, many methods perform well in controlled experimental settings or single operating conditions, but their cross-condition generalization ability in complex offshore loading environment has not been fully verified. On the other hand, most existing models emphasize the accuracy of overall classification. They have not paid enough attention to the confusing relationship between fault categories, category imbalance and engineering deployment feasibility. Therefore, there is still a gap between the current research and the real-world offshore wind application.

Although the existing research provides an important foundation for fault diagnosis of wind turbine gearbox, there are still three key challenges. First of all, the single source monitoring method is difficult to adapt to the high noise, strong variability and multi-disturbance coupling environment of offshore wind system. Second, many data-driven methods do not make full use of multi-source heterogeneous information. They can't use the complementary characteristics of different monitoring signals to represent faults. Thirdly, most researches still focus on coarse-grained fault identification. They can't meet the requirements of predictive maintenance and fine operation decision-making in offshore wind farms. Therefore, it is necessary to develop a gearbox fault diagnosis method suitable for offshore wind power scenarios. This method should consider the ability of multi-source fusion, cross-condition robustness and fine-grained diagnosis performance.

With the continuous development of digital and intelligent operation of offshore wind power, a wide range of multi-source heterogeneous monitoring data can be obtained during the operation of wind turbines, including vibration, SCADA, oil condition, electrical and environmental data. These data reflect the dynamic response, operating conditions, lubrication state and degradation trend of gearbox from different angles. They provide a powerful data base for fine fault diagnosis in offshore wind power applications. However, these multi-source data have significant differences in sampling frequency, scale, time synchronization, signal-to-noise ratio and sensitivity to operating conditions. Without effective preprocessing, feature optimization and fusion modeling strategies, they may introduce redundant information and diagnostic bias. Therefore, it has become an important research direction to develop a multi-source data fusion framework that balances accuracy, robustness and engineering applicability under offshore wind operation conditions.

Based on these considerations, this study follows the technical route of multi-source data preprocessing, multi-source feature extraction optimization and fault diagnosis model construction. It solves the fine-grained fault diagnosis requirements of offshore wind turbine gearbox under complex operating conditions. The proposed method focuses on reducing the limited anti-noise ability of single-source method, improving the granularity of fault identification and enhancing the cross-condition generalization ability. It aims to provide a practical method framework and theoretical support for state monitoring, fault early warning and intelligent maintenance decision-making of offshore wind power generation system.

2. Literature Review

In recent years, the research on fault diagnosis of offshore wind turbine gearbox has gradually developed from single-source data analysis to multi-source data fusion, from coarse-grained judgment to fine identification. Dubaish and Jaber studied a gear defect diagnosis method based on vibration signal. Their approach employed envelope demodulation and wavelet packet energy distribution to identify gear pitting and tooth breakage faults, and it achieved high accuracy on a laboratory test rig [14]. However, this method relied heavily on high signal-to-noise ratio vibration signals. Its adaptability to the strong noise conditions of offshore wind farms was limited, which resulted in a relatively high false alarm rate under complex operating conditions. At present, single-source data analysis methods retain advantages such as simple structure and low implementation cost. Nevertheless, they are highly sensitive to operating condition fluctuations and environmental noise, and they fail to comprehensively reflect the dynamic characteristics of gearboxes [15, 16]. Under the complex operating conditions of offshore wind power, reliance on a single data source easily leads to missed detections and misdiagnoses. As a result, such methods struggle to meet the requirements of refined fault diagnosis. Lin et al. developed a decision-level fusion method based on heterogeneous multi-sensor data. Their approach separately constructed vibration-based and oil-condition-based models, and it integrated the results through weighted voting at the decision level [17]. This method improved robustness to a certain extent. However, due to the lack of deep feature level fusion, the correlation and complementarity between different data sources are not fully utilized, which limits its ability to locate complex faults. The research of multi-source data fusion shows the advantages of multi-modal information in gearbox fault diagnosis. However, the existing research mainly focuses on simple feature stitching or weight-based fusion. Data heterogeneity, different sampling frequencies and different operating conditions of system modeling are still insufficient, which limits the high-precision state identification of gearbox in complex offshore environment [18, 19]. Liu et al. proposed a time series diagnosis framework combining long-term and short-term memory (LSTM) networks. By capturing the change of long-term operating conditions, this method realizes the prediction of gearbox degradation trend [20]. Although this method shows advantages in trend prediction, its model structure is complex. Its ability to fuse multi-source data is limited, and it is difficult to support integrated multi-task diagnosis covering detection, classification, location and severity assessment.

The intelligent diagnosis model improves the automation and recognition performance. However, problems such as limited generalization ability, strong dependence on training data and insufficient accurate diagnosis ability remain unsolved [21, 22]. These limitations are particularly obvious in the highly variable operating conditions of offshore wind power generation, where there is still a lack of systematic methods that can handle multi-source data and multi-level diagnostic tasks.

In response to these limitations, the main contributions of this study are summarized as follows:

(1) In order to solve the problem of insufficient consideration of data heterogeneity and differences in operating conditions in previous studies, a multi-source data preprocessing method suitable for complex offshore wind power operating conditions is proposed.

(2) A multi-source feature extraction and optimization framework was constructed, which mitigated the issues of feature redundancy and inefficient utilization of informative features in traditional approaches.

(3) A multi-source data fusion model for refined gearbox fault diagnosis was developed, which overcame the limitation of existing studies that mainly achieved coarse-grained diagnostic results.

3. Research Methodology

3.1. Multi-Source data preprocessing

To ensure that multi-source data from Supervisory Control and Data Acquisition (SCADA) systems, vibration sensors, oil condition monitoring, electrical measurements, and environmental monitoring serve as unified, stable, and comparable inputs, the raw data are systematically preprocessed in this study. The preprocessing aims to improve data quality, remove abnormal records, and synchronize data collected at different sampling frequencies. It also mitigates the influence of operating condition fluctuations on feature values. Finally, all features are scaled and normalized to construct a standardized feature space suitable for fusion modeling.

Data collected from different sensors exhibit substantial differences in sampling frequency, noise characteristics, physical meaning, and recording cycles [23]. SCADA data mainly reflect low-frequency operating condition characteristics. Vibration and electrical signals capture high-frequency dynamic responses. Oil condition monitoring data shows discrete sampling behavior, while environmental data records external changes that affect the turbine [24, 25]. Without systematic preprocessing, these differences will lead to inconsistent feature scales, uneven noise levels and misaligned timestamps, thus reducing the accuracy and robustness of the diagnosis model [26].

To address these issues, a multi-source data preprocessing pipeline was designed from four aspects: data cleaning, time alignment, operating condition partitioning, and standardization or normalization. The detailed procedures are summarized in Table 1.

Table 1. Multi-source data preprocessing pipeline

Data type	Main issues	Preprocessing operations	Purpose
SCADA data	Missing values, abnormal operating points, inconsistent sampling intervals	Missing value imputation (linear or nearest-neighbor), outlier removal (3σ and threshold-based methods), aggregation using a unified time window	Ensure continuity and stability of operating condition features, and improve the quality of subsequent operating condition partitioning
Vibration signals	Strong high-frequency noise, possible sensor saturation or drift	Abnormal segment removal, time-window slicing, extraction of statistical and frequency-domain features for time alignment	Reduce the influence of noise and map the original waveform into features suitable for unified fusion.
Oil condition data	Discrete sampling, extreme fluctuations in particle counts	Outlier removal, trend smoothing, time interpolation and alignment with a unified time window	Save actual wear trends and synchronize oil data with other sources.
Electrical signals	Harmonic mutations, short-term disturbances, varying sampling rates	Noise segment filtering, resampling by time window, extraction of harmonic and power quality features	Extract reliable electrical representation and enhance the correlation with vibration and SCADA data.
Environmental data	Inconsistent sampling periods, possible missing values	Missing value completion, interpolation to a unified time window, smoothing of key variables such as temperature and wind speed	Construct stable environmental characteristics to support the analysis of changes in operating conditions.
Operating condition labels	Rapid operating condition changes affecting feature distributions	Operating condition partitioning based on clustering of wind speed, rotational speed, and power	Reduce interference from differences in operating conditions and enhance sensitivity to inherent fault characteristics.
All feature data	Large numerical scale differences and inconsistent units	Z-score standardization combined with Min–Max normalization	The multi-source features are mapped to a unified scale to improve the stability of the fusion model.

Through these preprocessing steps, the original multi-source data with different sampling frequencies, recording periods and physical sizes are converted into time-aligned, high-quality and consistent operating conditions data sets with uniform feature scales. This data set provides reliable fusion input for subsequent feature extraction and accurate fault diagnosis modeling.

3.2. Optimization of multi-source feature extraction methods

After the preprocessing of multi-source data, this study further extracts and optimizes the characteristics of vibration signal, SCADA and operating condition data, oil state and electrical data. The goal is to preserve the fault-sensitive information related to the gearbox state in a limited feature dimension, while suppressing the influence of redundancy and noise.

For offshore wind turbine gearbox, vibration signal is the most direct information source to reflect the meshing conditions and the health status of rotating parts. Considering the influence of strong environmental noise and fluctuation of operating conditions in the sea scene, this study selects statistical indicators that are more sensitive to pulse and non-stationary changes based on traditional time domain characteristics. These indicators are estimated in a robust manner using a sliding time window. In each time window, the root mean square (RMS) characteristic is defined as:

$$\mathrm{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{1}$$

where RMS reflects the vibration energy level, N denotes the number of sampling points, and x_i represents the vibration acceleration or velocity at the i-th sampling point. In order to improve the sensitivity to local impact and early defects, kurtosis is used to characterize the pulse of vibration signals in each sampling window:

$$\text{Kurt} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^4}{\sigma^4} \tag{2}$$

In this expression, μ represents the average value of vibration signals and σ represents the standard deviation. Under normal operating conditions, the vibration signal approximately follows Gaussian distribution, and the kurtosis remains close to a constant value. Abnormal gear meshing, pitting corrosion or cracks will introduce impact components, which leads to a significant increase in kurtosis. In practice, the kurtosis of multiple time windows is statistically analyzed to reduce the misjudgment caused by transient noise. In order to further reduce the influence of the change of operating conditions, the characteristics of standardized power deviation and temperature gradient of operating conditions are constructed based on the original variables. These characteristics are used to describe whether the gearbox shows abnormal behavior under comparable operating conditions. Normalized power deviation index is defined as:

$$\Delta P_{\text{norm}} = \frac{P-\hat{P}}{\hat{P}} \tag{3}$$

where ΔP_{norm} represents the normalized power deviation, P represents the measured power, and p represents the theoretical power obtained at a given wind speed or rotational speed. When the internal fault of gear box leads to the reduction of transmission efficiency or frequent control adjustment, ΔP_{norm} often appears continuous deviation or fluctuation increase. After constructing multi-source features from vibration, SCADA, oil condition and electrical data, the obtained feature space may become high-dimensional. In addition, different features show different sensitivities to different failure modes. Therefore, feature standardization and feature importance evaluation are applied to screen and optimize feature sets. This process ensures that the model input retains enough information content, while avoiding redundancy and noise interference.

Let the j-th original feature after preprocessing and construction be f_j, the average value on the sample set is μ_j, and the standard deviation is σ_j. Standardized features are defined as:

$$z_j = \frac{f_j-\mu_j}{\sigma_j} \tag{4}$$

where z_j represents the standardized feature mapped into a zero-mean, unit-variance space. By applying this transformation to all elements, the influence of different physical units and numerical ranges on model training is reduced. This enables the subsequent fusion model to pay more attention to the internal relationship between features and fault labels.

3.3. Refined gearbox fault diagnosis model based on multi-source data fusion

After the preprocessing of multi-source data (Section 3.1) and the optimization of feature extraction method (Section 3.2), a fine fault diagnosis model of offshore wind turbine gearbox based on multi-source data fusion is further developed. The proposed diagnostic model consists of four main modules:

(1) Multi-source input and feature integration module: This module receives standardized feature vectors from vibration, SCADA and operating condition data, oil condition monitoring, electrical measurement and environmental monitoring, and it uses a unified time window to align these features.

(2) Shared feature encoding module: This module employed deep neural networks to perform nonlinear mapping and deep fusion of multi-source features, thereby generating low-dimensional representations that reflected the comprehensive health state of the gearbox.

(3) Refined multi-task diagnosis module: Based on the shared representations, four task-specific heads were constructed for fault detection, fault type identification, fault location localization, and fault severity assessment, which enabled multi-level diagnosis.

(4) Diagnostic decision and uncertainty assessment module: This module output prediction results and confidence levels for each task, and it supported operating-condition-partitioned diagnostic decisions using operating condition labels.

The detailed structure of the refined fault diagnosis model is summarized in Table 2.

Table 2. Structure of the refined fault diagnosis model

Module	Input	Core methods/structure	Output (diagnostic level)	Main function
Multi-source input and feature integration module	Vibration features, SCADA and operating condition features, oil condition features, electrical features, environmental features	Time-window alignment, multi-source feature concatenation	Unified multi-source feature vector	Achieve unified representation of different data sources in time and dimension
Shared feature encoding module	Unified multi-source feature vector	Multi-layer fully connected network or attention-based fusion network	Low-dimensional shared health state representation	Perform nonlinear mapping and deep fusion of multi-source features
Fault detection task head	Shared feature representation	Classification layer	Fault existence (healthy or faulty)	Perform preliminary detection of gearbox faults

Fault type identification task head	Shared feature representation	Multi-class classification layer	Fault types	Distinguish gearbox faults with different mechanisms
Fault location localization task head	Shared feature representation	Multi-class classification layer	Fault location	Achieve refined localization of internal gearbox faults
Fault severity assessment task head	Shared feature representation	Ordinal classification or regression layer	Fault severity level or continuous degradation index	Support maintenance strategy decisions and prioritization
Decision and assessment module	Outputs and confidence levels from all tasks	Weighted strategies and operating-condition-partitioned statistics	Comprehensive diagnostic report	Enable scenario-oriented and interpretable diagnostic decision output

Through this model architecture design and the proposed multi-source data fusion strategy, the refined gearbox fault diagnosis model achieved multi-source perception, deep feature fusion, and multi-level fault identification under complex offshore operating conditions.

4. Experimental design

The experiments in this study were conducted using the *Wind Turbine Gearbox Condition Monitoring Vibration Analysis Benchmarking Datasets* publicly released by the National Renewable Energy Laboratory (NREL), USA (DOI: 10.25984/1844194). The dataset was collected from a full-scale wind turbine gearbox test platform established under the Gearbox Reliability Collaborative program. The test object was a 750 kW, three-bladed, stall-controlled wind turbine gearbox. The dataset included experimental data from a healthy gearbox of a given design and a faulty gearbox of the same design, in which internal bearings and gear components were damaged following a loss-of-lubrication incident. On the drivetrain test rig, twelve accelerometers and high-speed shaft rotational speed and instantaneous angular velocity measurement points were installed. Multiple sets of vibration signals with a duration of 60 s were recorded under various rotational speed and load conditions at a sampling frequency of 40 kHz. In addition, the *Wind Turbine SCADA Data for Early Fault Detection* dataset (DOI: 10.5281/zenodo.10958775) is used as an external validation dataset representing real offshore wind farm scenarios. This dataset was released by Fraunhofer IEE and contains SCADA time-series data collected from three operational wind farms and 36 wind turbines, organized into 95 subsets. The total recording duration is approximately 89 turbine-years. It also provides 44 abnormal event intervals associated with turbine faults and 51 normal operating time series. Wind farms B and C are located in Germany and correspond to real offshore wind installations. They are therefore used as publicly available sources for validating model performance under offshore operating conditions. The dataset includes multi-dimensional SCADA variables for each wind farm. Wind farm B contains 257 variables, while wind farm C contains 957 variables. The sampling interval is 10 minutes. Each timestamp is annotated with turbine operating status labels. The dataset also provides partial fault descriptions as well as the start and end times of abnormal events. It is suitable for research on anomaly detection, fault early warning, and cross-farm generalization in wind turbine systems. Since this dataset originates from real offshore wind farms and provides explicit anomaly annotations and operational labels, it is used as a supplementary validation dataset in this study. However, the primary performance comparison is conducted on the NREL gearbox benchmark dataset to ensure consistency with existing gearbox diagnosis studies. In contrast, fault detection and fault type recognition experiments further incorporate this offshore wind farm dataset to assess the generalization capability and practical applicability of the proposed method under real offshore operating conditions.

All experiments were conducted on a local server using a deep learning framework. The hardware and software configurations are as follows:

(1) The server is Dell PowerEdge R740, equipped with two Intel Xeon Silver 4214R processors, and the main frequency is 2.4 GHz.

(2) The graphics processing unit is NVIDIA RTX 3090 with 24 GB memory, which is used to speed up model training and reasoning.

(3) The system is equipped with 128 GB DDR 4 memory, which supports batch loading and caching of multi-source data.

(4) The operating system is Ubuntu 20.04, which provides a stable Linux operating environment.

(5) The deep learning framework is PyTorch 1.11.0, and CUDA 11.3 is adopted for calculation acceleration.

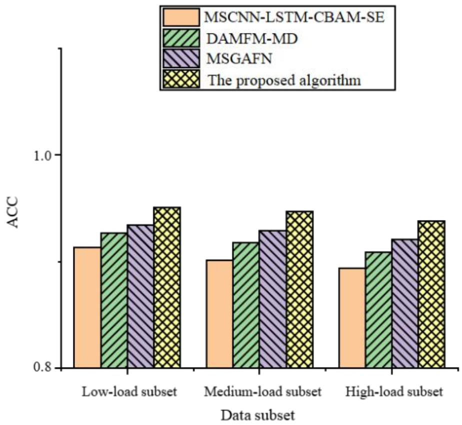
(6) Matlab R2021b is mainly used for preprocessing and feature extraction of vibration signals, including envelope demodulation and wavelet transform.

During the model training, the parameters of the optimized model are systematically configured to ensure the stability and convergence between different tasks in the multi-source feature fusion network. In the feature level fusion module, the number of hidden neurons is set to 256, 128 and 64 in decreasing structure. ReLU activation function is used to enhance nonlinear mapping ability. In the shared feature coding stage, the learning rate is set to 1×10^{-3}, and Adam optimizer is used to improve the convergence efficiency of high-dimensional multi-source feature space. To avoid gradient instability, a weight decay term of 1×10^{-5} was introduced during training. Dropout with a rate of 0.3 was applied between fully connected layers to reduce overfitting. The benchmark models used for comparison included Multi-Scale Convolutional Neural Network–Long Short-Term Memory–Convolutional Block Attention Module–Squeeze-and-Excitation (MSCNN-LSTM-CBAM-SE), Dual-Attentive Multimodal Fusion Method with Multiscale Dilated Convolutional Neural Network (DAMFM-MD), and Multi-Scale Graph Attention Fusion Network (MSGAFN).

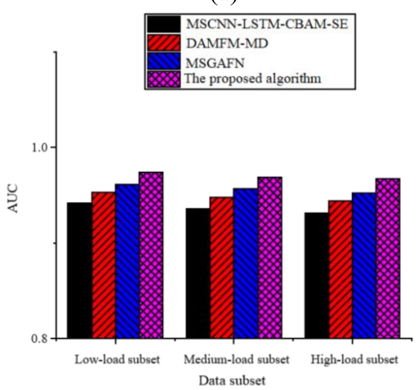
5. Experimental comparison and analysis

5.1. Performance comparison experiments

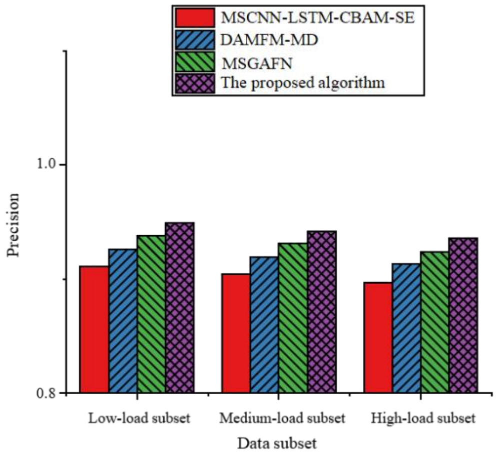
To validate the effectiveness of the proposed optimized model for refined gearbox fault diagnosis under multi-source data conditions, comparative experiments were conducted from two perspectives: classification performance, and model robustness and engineering applicability. The comparison results for classification performance are presented in Figure 1.



(a)



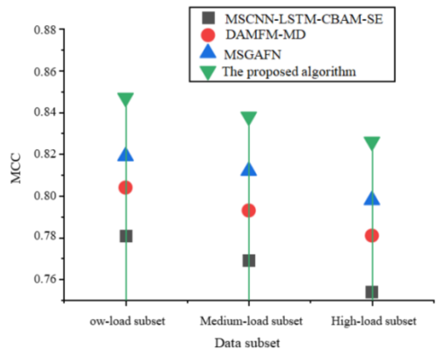
(b)



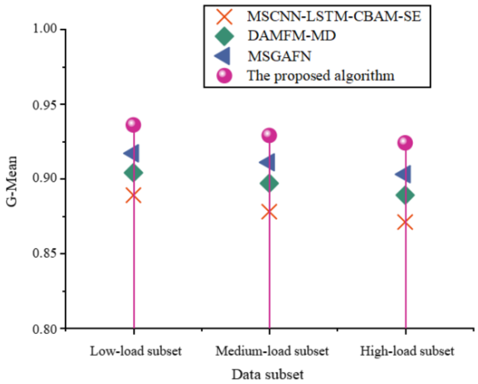
(c)

Figure 1. Classification performance. (a) Accuracy (ACC) (b) Area Under the Curve (AUC) (c) Precision

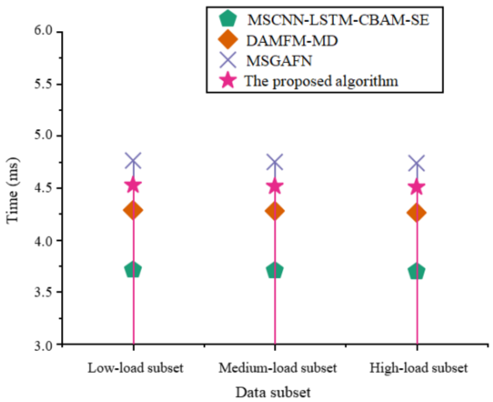
Figure 1 show that the proposed optimized model achieved high overall accuracy across all three data subsets. The ACC values under low-load, medium-load, and high-load conditions were 0.951, 0.947, and 0.938, respectively. Although a slight decrease was observed as the load increased, the accuracy remained at a high level. By contrast, MSCNN-LSTM-CBAM-SE achieved an accuracy of approximately 0.894 on the high-load subset. This indicated that the performance of traditional convolutional–temporal architectures degraded more noticeably under stronger operating condition fluctuations and lower signal-to-noise ratios. After fusing multi-source features, the proposed model demonstrated better adaptability to complex operating conditions. In terms of AUC, the optimized model achieved values of 0.974, 0.969, and 0.967 on the low-load, medium-load, and high-load subsets, respectively. These results indicated that the model consistently distinguished faulty samples from healthy samples across different decision thresholds. In comparison, DAMFM-MD achieved an AUC of approximately 0.948 under medium-load conditions. Although its overall performance exceeds the traditional model, it is still slightly inferior to the proposed method. This discovery shows that multi-source fusion leads to a clearer decision boundary, especially in unbalanced sample distribution. For accuracy measurement, the optimized model achieves values of 0.949, 0.942 and 0.936 on three subsets. These results show that most of the samples identified as faults are real faults under different operating conditions, which effectively suppresses false alarms. For example, under high load conditions, MSGAFN achieves an accuracy of about 0.924, which is lower than 0.936 obtained by the proposed model. This difference means that the proposed model provides a more conservative and reliable fault classification in the case of mixed fault types and increased noise.



(a)



(b)



(c)

Figure 2. Dimensions of model robustness and engineering applicability. (a) Matthews Correlation Coefficient (MCC); (b) Geometric Mean (G-Mean); (c) Inference Time

As shown in Figure 2, the MCC values of 0.847, 0.838 and 0.826 are achieved by the optimized model under the conditions of low load, medium load and high load respectively. All values exceed 0.8, which reflects the balance performance when considering true positive, true negative, false positive and false negative at the same time. Under the condition of medium load, DAMFM-MD achieves an MCC of about 0.793, while the proposed model reaches 0.838. The results show that the proposed method maintains more stable multi-class classification performance under unbalanced samples and different operating conditions, and will not be biased towards specific classes. G-means measurement further verifies the balance of recognition performance between different categories. The optimization model achieves G-means of 0.936, 0.929 and 0.924 on low load, medium load and high load subsets respectively. These values are obviously higher than about 0.871 obtained by MSCNN-LSTM-CBAM-SE under high load. This discovery shows that, while maintaining the high recognition rate of main fault types, the proposed model also retains the strong recognition ability of a few faults and boundary samples, which is helpful to reduce the missed detection of weak fault categories. As for the reasoning time, the average single-sample reasoning time of 4.528 ms, 4.517 ms and 4.509 ms was achieved in the optimized model under three operating conditions. These values are slightly higher than the more compact MSCNN-LSTM-CBAM-SE, which takes about 3.7 milliseconds, but they are lower than the more complex MSGAFN, which runs at the level of about 4.7 milliseconds. Generally speaking, the model maintains high diagnostic accuracy and robustness while keeping the reasoning time within an acceptable range, which meets the requirements of online or near-real-time fault diagnosis of offshore wind turbines.

To further verify whether the performance improvements of the proposed optimized model over the three benchmark models were statistically significant across different evaluation metrics, paired *t*-tests were conducted for each metric based on the three data subsets (low-load, medium-load, and high-load). The significance analysis results are summarized in Table 3.

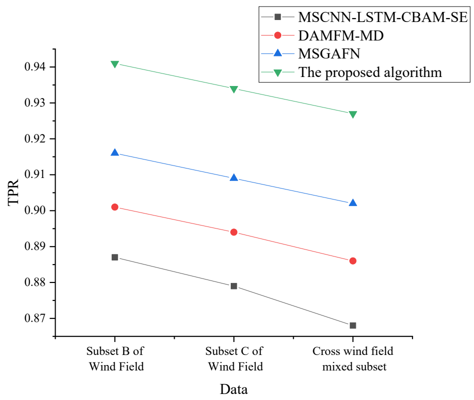
Table 3. Significance analysis

Metric	Proposed model vs. MSCNN-LSTM-CBAM-SE	Proposed model vs. DAMFM-MD	Proposed model vs. MSGAFN
ACC	0.013 (*)	0.019 (*)	0.028 (*)
AUC	0.017 (*)	0.023 (*)	0.034 (*)
Precision	0.021 (*)	0.029 (*)	0.041 (*)
MCC	0.011 (**)	0.018 (*)	0.027 (*)
G-Mean	0.014 (*)	0.022 (*)	0.033 (*)
Inference Time	0.018 (*)	0.039 (*)	0.271 (ns)

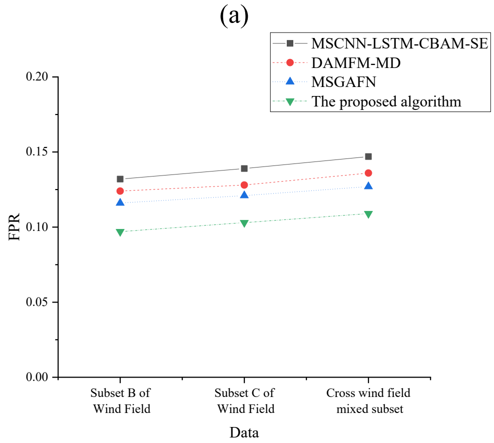
As shown in Table 3, for the five performance indicators of ACC, AUC, Precision, MCC and G-Mean, most differences between the proposed optimization model and the three benchmark models reached statistical significance at the level of 0.05. Especially, for MCC measurement, the difference between the proposed model and MSCNN-LSTM-CBAM-SE reached a highly significant level of 0.01. This result shows that multi-source fusion and fine modeling provide more stable advantages in overall classification balance. With respect to inference time, the differences between the proposed model and MSCNN-LSTM-CBAM-SE, as well as DAMFM-MD, were statistically significant at the 0.05 level. However, the difference between the proposed model and MSGAFN did not reach statistical significance. This finding suggested that, while maintaining high diagnostic performance, the inference efficiency of the proposed model was comparable to that of complex graph-attention-based models, thereby demonstrating good engineering applicability.

5.2. Analysis of fault detection and fault type identification results

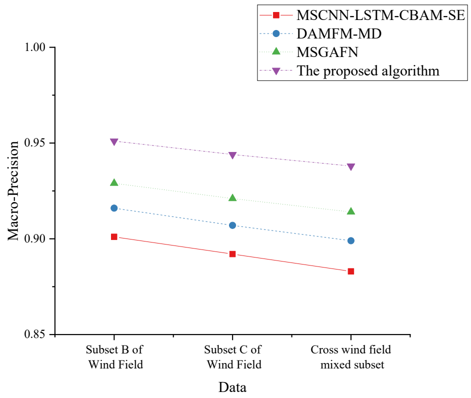
After completing the overall comparison of classification performance, this section further evaluated the detection and identification capabilities of each model from two perspectives: whether a fault was successfully detected, and which specific fault type it belonged to. Three experimental scenarios are designed, including a subset of wind farm B, a subset of wind farm C and a hybrid subset across wind farms. These settings are used to comprehensively evaluate the stability and generalization performance of the model under real offshore wind operating conditions. The corresponding results are shown in Figure 3.



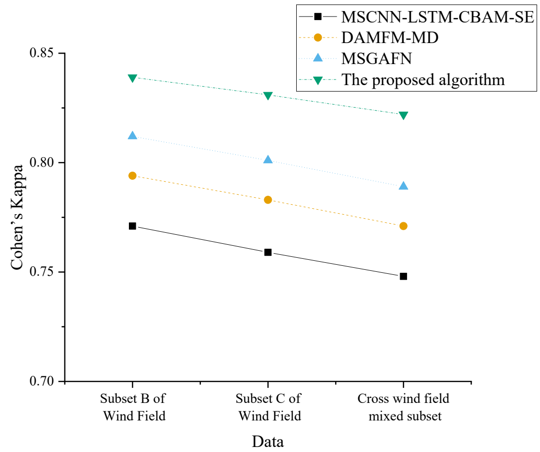
(a)



(b)



(c)



(d)

Figure 3. Fault detection performance. (a) True Positive Rate (TPR); (b) False Positive Rate (FPR); (c) Macro-Precision; (d) Kappa coefficient

The results in Figure 3 show that the proposed optimization model achieves TPR values of 0.941, 0.934 and 0.927 on three test subsets respectively. These values are always the highest among all comparison models, indicating that the proposed method can reliably detect abnormal samples under real offshore wind operating conditions. It is worth noting that the model still achieves a TPR of 0.927 on the cross-farm mixed subset. In contrast, MSCNN-LSTM-CBAM-SE achieved a TPR of 0.868 on the same subset. This shows that the traditional convolution-time architecture often suffers from performance degradation in cross-domain transmission scenarios, while the proposed method shows stronger generalization ability among different wind farms. In the aspect of false positive rate, the proposed model has achieved 0.097, 0.103 and 0.109 FPR values on wind farm B subset, wind farm C subset and cross-wind farm mixed subset respectively. These values remain low in all cases. In contrast, DAMFM-MD achieved an FPR of 0.128 on subset C of wind farms, while MSGAFN achieved an FPR of 0.127 on hybrid subset across wind farms. Both of them are higher than the proposed model. These results show that the proposed method not only maintains a high detection rate, but also effectively reduces the situation of misclassification of normal samples as faults. This makes it more suitable for online early warning application of offshore wind farms.

As shown in fig. 3(c), the proposed model achieves macro-precision values of 0.951, 0.944 and 0.938 on the subset of wind farm B, the subset of wind farm C and the hybrid subset across wind farms, respectively. This proves the balanced multi-class fault identification performance in different maritime scenarios. Taking subset C of wind farm as an example, the macro accuracy of MSGAFN reaches 0.921, while the proposed model reaches 0.944. This shows that the proposed method provides better identification performance for a few fault types under more complex operating conditions and unbalanced fault distribution in offshore wind environment. In terms of the Kappa coefficient, the proposed model obtains values of 0.839, 0.831, and 0.822 across the three subsets, all of which are higher than those of the compared models. On the cross-farm mixed subset, DAMFM-MD achieves a Kappa value of 0.771, while the proposed model reaches 0.822. This suggests that the proposed method maintains higher consistency and stability under cross-farm conditions. Since the Kappa coefficient effectively reflects agreement between predicted results and ground truth under class imbalance conditions, these results further confirm the reliability of the proposed model for fine-grained multi-class fault diagnosis.

To further analyze the fine-grained fault identification performance of the proposed model, representative samples from the cross-farm mixed subset were selected. Based on maintenance alarm logs and fault descriptions, samples were categorized into five classes: normal condition, gear wear, bearing fault, lubrication degradation, and compound transmission fault. A confusion matrix was then constructed, as shown in Table 4.

Table 4. Confusion matrix for the medium-load data subset

True \ Predicted	Normal	Gear wear	Bearing fault	Lubrication degradation	Compound fault	Total
Normal	81	3	2	2	2	90
Gear wear	4	76	3	2	3	88
Bearing fault	3	4	69	2	2	80
Lubrication degradation	2	2	3	63	2	72
Compound fault	2	3	2	3	60	70
Total	92	88	79	72	69	400

Table 4 shows that the proposed optimized model achieves satisfactory classification performance across all five operating states in the cross-farm mixed subset. For the normal condition, there are 90 ground-truth samples, of which 81 are correctly classified. Only a small number are misclassified as gear wear, bearing fault, lubrication degradation, or compound transmission fault. This shows that the model establishes a relatively clear decision boundary between health and fault state.

For a specific fault category, the model correctly identifies 76 of 88 gear wear samples, 69 of 80 bearing fault samples, 63 of 72 lubrication degradation samples and 60 of 70 composite transmission fault samples. Generally speaking, the value along the main diagonal is significantly higher than that at the off-diagonal position. This shows that the model can effectively extract the distinguishing features of different gearbox fault types under real offshore wind operation conditions. As far as the wrong classification mode is concerned, there is some confusion between gear wear and bearing failure. Limited overlap is also found between lubrication degradation and compound transmission failure. This shows that different mechanical faults may show some similar characteristics in the vibration response, power fluctuation and operation index of real offshore data sets. This kind of overlap is more likely to occur in the early fault stage or in the case of severe load change, when the category boundary becomes inconspicuous. However, these error classifications are mainly concentrated in physically related fault types, and their overall proportion is still limited. They do not significantly affect the overall classification performance.

Based on the results of confusion matrix, the proposed model shows strong fine-grained fault classification ability under real offshore wind conditions. It can effectively distinguish normal and fault states, and accurately identify typical gearbox faults, including gear wear, bearing faults, lubrication degradation and compound transmission faults. These results further confirm the practical potential of the proposed method for the condition monitoring and maintenance of offshore wind turbines.

5.3. Ablation Study and Hyperparameter Analysis

In order to evaluate the effectiveness of key components in the proposed multi-source data fusion framework and model architecture, the following ablation settings were constructed:

- Baseline: Single source vibration characteristics with basic classification network.
- Baseline+Multi-source: Multi-source data fusion without functional optimization
- Baseline+multi-source+characteristic optimization: the characteristic optimization module is added.
- Complete model (suggestion): A complete model with multi-source integration, function optimization and multi-task architecture.

Table 5 summarizes the ablation results.

Table 5. Ablation study results

Model structure	Low-load	Medium-load	High-load
Baseline	0.891	0.884	0.873
+ Multi-source	0.914	0.907	0.895
+ Feature Optimization	0.932	0.926	0.914
Full Model	0.951	0.947	0.938

As shown in Table 5, with the introduction of multi-source fusion and feature optimization module, the model performance is gradually improved. The baseline model using only a single source data achieves an accuracy of 0.873 under high load conditions. After the introduction of multi-source fusion, the accuracy is improved to 0.895, which shows that the additional information from multiple sources enhances the feature representation. When the feature optimization module is further combined, the performance is improved to 0.914, which shows that feature selection and reconstruction are helpful to reduce noise interference. Finally, the accuracy of the complete model is 0.951, 0.947 and 0.938 under low, medium and high load conditions, respectively. This proves the complementary effect of multi-task learning and depth feature fusion.

For nonparametric sensitivity analysis, the learning rate and fusion weight coefficient are selected, and the results are shown in Table 6.

Table 6. Sensitivity analysis results

Learning Rate	ACC	λ	ACC
0.0005	0.928	0.2	0.933
0.0013	0.947	0.4	0.941
0.0021	0.939	0.6	0.947
0.0057	0.921	0.8	0.936

The results show that the model is sensitive to the learning rate. When the learning rate is too low, the convergence becomes very slow, while too high a value will lead to unstable training. The best performance is achieved at about 0.0013, indicating the balance between convergence speed and training stability. For the fusion weight λ, the best performance is obtained at 0.6. This shows that multi-source characteristics need a balanced contribution, because overemphasizing any single data source may reduce the overall performance.

6. Discussion

The overall experimental results show that the proposed optimization model is superior to the three comparison models in all evaluation indexes. In addition, its performance remains relatively stable under different load conditions, which shows that the proposed multi-source feature fusion strategy and working condition division scheme can effectively deal with the common load fluctuation and signal instability problems in the operation of offshore wind power gearbox. In offshore wind farms, turbines rarely operate under ideal steady-state conditions. On the contrary, they are constantly affected by wind speed changes, turbulence, wave-induced loads and dynamic control adjustments. Therefore, under a single operating condition, cross-condition stability is usually more important than high accuracy. The consistent performance of the proposed model under low, medium and high loads shows that it has great practical application potential in complex marine environment. From the point of view of fault detection and classification, the model keeps a high detection rate and a low false alarm rate. It also achieves strong performance in macro-average accuracy and Kappa coefficient across multiple fault categories. These results show that the model can not only detect the potential gearbox faults in time, but also keep the balance between different fault types in the case of unbalanced categories. This characteristic is particularly important in offshore wind power applications. In practice, there are far more normal samples than fault samples. In addition, the maintenance cost and shutdown loss related to offshore faults are significantly higher than those of onshore systems. A high false positive rate may lead to unnecessary maintenance actions and ship deployment. High false negative rate may lead to missing maintenance window and minor defects developing into serious transmission system failure. The proposed model achieves a favorable balance between these two aspects, making it suitable as the core component of offshore wind condition monitoring and early warning system. The confusion matrix analysis further shows that most of the misclassification occurs between physically similar fault categories, such as gear pitting, tooth root cracking and bearing failure. This level of confusion is usually acceptable. It also reflects the actual characteristics of offshore wind gear box diagnosis, in which multiple mechanical faults can show similar vibration response under the coupled operation interference. This effect is especially obvious in the early degradation stage, when the fault boundary is not obvious. Although the proposed multi-source fusion strategy has alleviated this problem to some extent, further improvement may require stronger physical information constraints, richer multi-modal sensing information and a secondary classification mechanism specially designed for closely related fault categories.

7. Conclusions

This study solves the challenges of complex operating conditions, high maintenance costs and strict reliability requirements of offshore wind turbines, with the focus on fine-grained fault diagnosis of wind turbine gearboxes. A diagnosis framework based on multi-source data fusion is developed, and the system is evaluated by using common gearbox data set and multiple baseline models. Compared with onshore wind farms, offshore wind turbines operate in worse environmental conditions, with more limited maintenance accessibility and more serious failure consequences. Therefore, it is of great engineering significance to develop a high-precision and high-robustness diagnostic method suitable for marine environment. From the perspective of methodology, a fine-grained fault diagnosis model based on multi-source data fusion is proposed. By adopting the shared feature coding strategy and multi-task learning architecture, the model realizes the integrated learning of fault detection, fault type classification, fault location identification and fault severity evaluation. The depth feature level fusion of multi-source information enables the model to make full use of vibration signals, operating conditions and long-term degradation modes. This design improves the adaptability to offshore wind turbines operating under variable wind speed, fluctuating load and strong environmental interference. The experimental results show that the proposed method is always superior to the comparison model in many evaluation indexes, while maintaining stable diagnostic performance under different load conditions. These results confirm that the proposed method effectively meets the requirements of offshore wind condition monitoring in terms of accuracy, robustness and cross-condition generalization ability.

From the application point of view, the proposed framework not only improves the accuracy of gearbox fault identification, but also enhances the ability of the model to deal with complex fault modes and unbalanced data distribution. This is of practical significance to the predictive maintenance of offshore wind farms, the optimization of maintenance plans and the reduction of unexpected downtime. Through more precise diagnosis output, maintainers can identify potential risks earlier and implement targeted maintenance strategies within a limited weather-related maintenance window, thus improving the operational reliability and economic efficiency of wind farms.

Despite these advantages, there are still some limitations. First of all, the current multi-source feature set mainly focuses on vibration signals and operating condition variables, while oil-related, electrical and environmental information has not been fully utilized. This limits the ability of the model to distinguish subtle failure modes, such as lubrication degradation and structural damage. Considering the long-term exposure of offshore wind turbines to salt spray corrosion, humidity-temperature cycle and complex load conditions, the future work should include oil particle analysis, torsional vibration signal, acoustic emission, structural strain measurement and environmental coupling characteristics, so as to establish a more comprehensive multi-mode diagnosis system suitable for offshore scenes. Secondly, although the current model achieves acceptable reasoning efficiency for near-real-time diagnosis, challenges still exist for large-scale deployment of edge devices or resource-limited platforms in offshore wind farms. Future research may explore model compression techniques, such as network pruning, knowledge extraction and low-level quantization, to significantly reduce the calculation cost and model size, while maintaining the accuracy of diagnosis. This will better support the requirements of distributed, online and long-term continuous monitoring in offshore wind turbine systems.

References

[1] Yousaf MZ, Singh AR, Khalid S, Khalid R, Uddin M, Khurshaid T. Enhancing HVDC transmission line fault detection using disjoint bagging and bayesian optimization with artificial neural networks and scientometric insights. Sci Rep. 2024;14(1):23610.

[2] Abubakar A, Jibril MM, Almeida CFM, Šljivac D, Dumnić B, Kamenko I. A novel hybrid optimization approach for fault detection in photovoltaic arrays and inverters using AI and statistical learning techniques: a focus on sustainable environment. Processes. 2023;11(9):2549.

[3] Teta A, Korich B, Bakria D, Mharakurwa E, Chola C. Fault detection and diagnosis of grid-connected photovoltaic systems using energy valley optimizer based lightweight CNN and wavelet transform. Sci Rep. 2024;14(1):18907.

[4] Najafzadeh M, Pouladi J, Daghigh A, Mohammadi K, Tatar A, Aghsami A. Fault detection, classification and localization along the power grid line using optimized machine learning algorithms. Int J Comput Intell Syst. 2024;17(1):49.

[5] Yousaf MZ, Singh AR, Khalid S, Khalid R, Uddin M, Khurshaid T. Bayesian-optimized LSTM-DWT approach for reliable fault detection in MMC-based HVDC systems. Sci Rep. 2024;14(1):17968.

[6] Raamesh L, Jothi S, Radhika S. Enhancing software reliability and fault detection using hybrid brainstorm optimization-based LSTM model. IETE J Res. 2023;69(12):8789-8803.

[7] Quiles-Cucarella E, Sánchez-Roca P, Agustí-Mercader I. Performance Optimization of Machine-Learning Algorithms for Fault Detection and Diagnosis in PV Systems. Electronics. 2025;14(9):1709.

[8] Fan F, Chu SC, Pan JS, Zheng WJ, Chen D. An optimized machine learning technology scheme and its application in fault detection in wireless sensor networks. J Appl Stat. 2023;50(3):592-609.

[9] Rojas L, Peña Á, Garcia J. AI-driven predictive maintenance in mining: a systematic literature review on fault detection, digital twins, and intelligent asset management. Appl Sci. 2025;15(6):3337.

[10] Imane M, Rahmoune C, Zair M, Boutasseta N, Bouakkaz MS. Bearing fault detection under time-varying speed based on empirical wavelet transform, cultural clan-based optimization algorithm, and random forest classifier. J Vib Control. 2023;29(1-2):286-297.

[11] Wang Z, Tao Y, Du Y, Wang S, Tao T. Optimization of gearbox fault detection method based on deep residual neural network algorithm. Sensors. 2023;23(17):7573.

[12] Yao Y, Liang T, Tan J, Xu J. A bearing fault detection and remaining useful life prediction method based on a multi-branch residual feature fusion mechanism and optimized weight allocation. Meas Sci Technol. 2023;35(2):025906.

[13] Konda SK. AI Integration in Building Data Platforms: Enabling Proactive Fault Detection and Energy Conservation. Int J Adv Res Comput Sci Technol. 2024;7(3):10327-10338.

[14] Dubaish AA, Jaber AA. State-of-the-art review into signal processing and artificial intelligence-based approaches applied in gearbox defect diagnosis. Eng Technol J. 2023;10(9):1-16.

[15] Kosuru VSR, Kavasseri Venkitaraman A. A smart battery management system for electric vehicles using deep learning-based sensor fault detection. World Electr Veh J. 2023;14(4):101.

[16] Rokani V, Kaminaris SD, Karaisas P, Ioannidis D, Tsoumas IP, Vokas G. Power transformer fault diagnosis using neural network optimization techniques. Mathematics. 2023;11(22):4693.

[17] Lin Q, Zhou X, Hong W, Yu Z. A Comprehensive Survey of Multi-View Intelligent Fault Diagnosis Tailored to the Sensor, Machinery Equipment, and Industrial System Faults. J Vib Eng Technol. 2025;13(5):335.

[18] Hodavand F, Ramaji IJ, Sadeghi N. Digital twin for fault detection and diagnosis of building operations: a systematic review. Buildings. 2023;13(6):1426.

[19] Ahilan T, Sinha S, Nekrasov S, Kanagaraj S, Selvaraj S, Ramana Murthy OV. Wind Power Turbine Fault Detection using Hybridized Convolutional Neural Network and Autoencoder. J Circuits Syst Comput. 2025;34(18):2550378.

[20] Liu Z, Tan C, Liu Y, Dong F, Wang H. A study of a domain-adaptive LSTM-DNN-based method for remaining useful life prediction of planetary gearbox. Processes. 2023;11(7):2002.

[21] Zhang BC, Sun SQ, Yin XJ, Yuan H, Liu C. Research on gearbox fault diagnosis method based on VMD and optimized LSTM. Appl Sci. 2023;13(21):11637.

[22] Jawad RS, Abid H. HVDC fault detection and classification with artificial neural network based on ACO-DWT method. Energies. 2023;16(3):1064.

[23] Bounemeur A, Chemachema M. Optimal adaptive fuzzy fault-tolerant control applied on a quadrotor attitude stabilization based on particle swarm optimization. Proc Inst Mech Eng I J Syst Control Eng. 2024;238(4):704-719.

[24] Muttashar HS, Shakir AM. Enhancing PV fault detection using machine learning: Insights from a simulated PV system. Iraqi J Electr Electron Eng. 2025;21(1):126-133.

[25] Zayed SM, Attiya G, El-Sayed A, Elhosseini MA. An efficient fault diagnosis framework for digital twins using optimized machine learning models in smart industrial control systems. Int J Comput Intell Syst. 2023;16(1):69.

[26] Duranay ZB. Fault detection in solar energy systems: A deep learning approach. Electronics. 2023;12(21):4397.