Proceedings of the 4th International Conference on Functional Materials and Civil Engineering
DOI: 10.54254/2755-2721/2026.35117

A Lightweight Physics-Guided Feature Fusion Network for Fault Waveform Classification in Three-Phase Inverter Circuits

Huiyu Gan

Changkong College, Nanjing University of Aeronautics and Astronautics, Nanjing, China
gan_hy@nuaa.edu.cn

Abstract. Three-phase inverters are widely used in renewable energy conversion, industrial drives, and energy storage systems. Their IGBTs and other power switching devices often work for long periods under high-frequency and high-current conditions, which makes fault diagnosis an important issue for reliable operation. Traditional diagnosis methods usually depend on manually designed features obtained from Fourier transform, wavelet analysis, or related signal-processing tools. Although these features are interpretable, their performance is closely tied to expert experience. Purely data-driven deep learning models can learn features from raw waveforms, but they often show limited physical interpretability and may overfit when fault samples are insufficient. This paper proposes a lightweight Multi-view Fault Feature Fusion Network (MFF-Net) for fault waveform classification in three-phase inverter circuits. The model contains a lightweight one-dimensional convolutional neural network branch for temporal waveform representation and a physics-guided branch for extracting single-phase statistics, three-phase imbalance indices, zero-sequence components, and multi-band harmonic energy ratios. A gated fusion module is then used to combine the two feature groups according to sample-specific fault characteristics. A simulated dataset with 3,200 samples is built under normal, overcurrent, phase-loss, and IGBT bridge-arm open-circuit conditions. Experimental results show that MFF-Net achieves stable training, high classification accuracy, and clear feature separation, offering a feasible solution for lightweight online fault diagnosis in power electronic systems.

Keywords: three-phase inverter, fault diagnosis, feature fusion, gating mechanism, lightweight convolutional neural network

1. Introduction

As renewable generation, industrial variable-frequency drives, and energy storage converters are connected on a wider scale, the three-phase inverter has become a common interface in power conversion systems [1]. Its health is not an isolated device issue; abnormal inverter behavior can affect current quality, equipment protection, and the safe operation of the downstream load. In practice, IGBTs and other power switching devices must work for long periods at high temperature, large current, and high-frequency switching. Ageing, thermal stress, and transient disturbance may

DOI: 10.54254/2755-2721/2026.35117

then appear as overcurrent, phase loss, bridge-arm open-circuit, or related faults [2]. IGBT open-circuit faults are particularly troublesome, because the affected phase may lose the ability to track the commanded current. If such a fault is missed at the early stage, the system can show degraded output performance and may eventually be forced into an unplanned shutdown. For this reason, fault diagnosis for inverter waveforms still deserves careful engineering attention.

At present, inverter fault diagnosis is commonly developed along two routes: traditional signal-processing methods and data-driven methods. Fourier transform, wavelet analysis, and similar signal-processing tools are often used to construct features before classification. These features are easy to relate to electrical phenomena, but the final result is strongly affected by prior knowledge, parameter choices, and expert experience [3]. Once the operating point drifts or the waveform contains several types of distortion at the same time, hand-designed features may not separate the fault classes reliably. Deep learning provides another option by learning representations directly from raw signals in an end-to-end manner. The difficulty is that the learned features are not always physically transparent, and the model may become unstable when real fault samples are limited. In addition, a large network is inconvenient for edge devices that must complete diagnosis online [4].

These observations show why neither route alone is fully satisfactory for the target application. A signal-processing scheme offers interpretability, whereas a deep model usually offers stronger representation capability. What is still missing is a compact framework that can retain physical mechanism information while also using data-driven waveform features. Another issue lies in feature fusion. Direct concatenation and fixed weighting are simple, but they assume that the contribution of each feature source is almost unchanged from one fault waveform to another. In online industrial terminals, this assumption is restrictive, and an unnecessarily complex network also increases memory and inference cost.

Based on this consideration, a lightweight physics-guided feature fusion network, MFF-Net, is developed for fault waveform classification in three-phase inverter circuits. The network keeps a dual-branch form. One branch is a lightweight 1D-CNN, which extracts temporal morphological features from raw three-phase current waveforms. The other branch calculates interpretable electrical features, including single-phase statistical features, three-phase imbalance features, zero-sequence components, and harmonic energy features. A gated adaptive fusion module is then used to regulate the two branches at the feature level according to the input sample. In this way, waveform morphology and physical prior information are used together, while the parameter size remains small. The resulting model is intended for inverter fault diagnosis where accuracy, interpretability, and deployment efficiency must be considered at the same time.

2. Related work

2.1. Traditional signal-processing-based fault diagnosis methods

Early inverter fault diagnosis was largely built on manual feature engineering. In those studies, researchers first transformed current or voltage signals into fault-sensitive representations by signal analysis [5]. Fast Fourier transform, wavelet transform, empirical mode decomposition, and related tools were used to obtain time-domain statistics, frequency-domain harmonic components, and time-frequency features [6]. These handcrafted features were then combined with manual thresholds, support vector machines, random forests, or other classifiers to distinguish operating conditions [7]. The merit of signal-processing-based diagnosis is its readable electrical interpretation. Phase imbalance, for example, naturally points to phase-loss faults, while a change in harmonic energy is often connected with waveform distortion. This readability is useful for engineers who need to

DOI: 10.54254/2755-2721/2026.35117

check the diagnosis result. The weakness is also evident: feature selection depends on domain experience, and the extraction procedure can become cumbersome. When load fluctuation, measurement noise, or IGBT bridge-arm open-circuit faults occur together, manually designed features may lose robustness. Thus, traditional methods are not easily extended to fine-grained identification across multiple operating conditions [8].

2.2. Data-driven fault diagnosis methods based on deep learning

Deep learning has recently become a common tool in industrial fault diagnosis. For inverter current waveforms, one-dimensional convolutional neural networks, recurrent neural networks, and their variants have been used to learn fault features from raw sequences [9]. Compared with manual feature engineering, these models can capture local waveform distortion, transient changes, and other complex signal patterns with less explicit feature design. Transfer learning, domain adaptation, and lightweight network structures have also been introduced to improve cross-condition diagnosis and edge deployment ability [10]. Purely data-driven deep learning, however, still leaves several practical concerns. The learned features are usually hidden in the network, so their link to the fault mechanism is not direct. This black-box property makes it harder for engineers to judge maintenance actions from the model output [11]. Data availability is another limitation. Many real faults are destructive or unsafe to reproduce, so large-scale fault samples are difficult to obtain. Under small-sample conditions, deep models may fit the training set too closely and generalize poorly [12]. Moreover, high-performance networks often carry many parameters and require considerable computation, which is unfavorable for fast online diagnosis on industrial edge terminals.

2.3. Physics-guided feature fusion diagnosis methods

Physics-guided feature fusion has therefore been used as a compromise between accuracy and interpretability [13]. In such methods, circuit mechanism features, statistical features, and deep network features describe the fault state from different views. Physical priors and data-driven representations are not redundant; they often supply complementary information for diagnostic robustness. Still, the way these features are fused remains a key issue. Simple concatenation and fixed-weight fusion are widely used, but they handle different feature sources almost identically for every input sample, even though different faults depend on different cues [14]. Phase-loss faults, for instance, can often be identified from imbalance indicators, whereas switch open-circuit faults may need more waveform morphology information. Some fusion networks also include redundant modules and are not lightweight enough for online deployment [15]. A compact fusion structure with sample-dependent flexibility is therefore needed.

3. Method

3.1. Overall architecture of MFF-Net

MFF-Net is constructed for fault waveform classification in three-phase inverter circuits. The input of the model is a segment of three-phase current waveform, and the output is the corresponding operating condition. Its design combines two viewpoints: data-driven waveform modeling and physics-guided feature modeling. The framework consists of three parts. First, a lightweight 1D-CNN branch extracts temporal morphological features from raw three-phase current waveforms.

DOI: 10.54254/2755-2721/2026.35117

Second, a physics-interpretable branch calculates prior electrical features and maps them into an embedding whose dimension matches the CNN output. Third, a gated adaptive fusion module assigns sample-dependent weights to the two feature groups before classification. This arrangement keeps the representation ability of deep learning, but introduces physical knowledge into the diagnosis process.

3.2. Lightweight 1D-CNN waveform feature extraction branch

The waveform branch learns temporal information directly from raw three-phase current signals. A three-layer 1D convolutional structure is used so that fault patterns can be observed at several temporal scales, including switching ripple, short-time transient disturbance, waveform distortion, and global envelope variation. The first convolutional layer contains 16 kernels with a kernel size of 9, which is useful for rapid fluctuation in the input waveform. The second layer increases the channel number to 32 and adopts a kernel size of 7 to describe medium-scale waveform morphology. The third layer uses 64 channels and a kernel size of 5, focusing on more abstract fault-related features. Batch normalization and ReLU activation follow each convolutional layer. Max pooling with a stride of 2 is placed after the first two layers to shorten the temporal length and reduce computation. After convolutional extraction, global adaptive average pooling compresses the feature map into a fixed-length vector. A fully connected layer then maps it to a 64-dimensional deep temporal feature. The branch can be written as:

$$h_{cnn} = FC(GAP(Conv_3(Conv_2(Conv_1(x))))) \tag{1}$$

Here, $Conv_i$ denotes the i-th convolutional block, including batch normalization, activation, and optional pooling. GAP denotes global adaptive average pooling, and FC denotes the fully connected mapping layer. With about 18K parameters, this branch keeps the computational burden relatively low.

3.3. Physics-interpretable feature extraction branch

The physics-guided branch computes features that have explicit electrical meanings in the three-phase current waveform. Unlike the CNN branch, this part is based on deterministic analytical calculation. It does not add learnable parameters during feature extraction, which makes the resulting descriptors stable and easier to interpret.

The physical feature vector has 24 dimensions and is organized into three groups. The first group is made up of single-phase time-domain statistical features for each phase, including root mean square, peak value, peak-to-peak value, mean absolute value, crest factor, and DC offset. These indicators reflect overcurrent, transient impulse, waveform asymmetry, and related fault phenomena. The second group contains three-phase imbalance features, namely three-phase RMS imbalance and zero-sequence component RMS, which are useful for phase-loss recognition. The third group is composed of frequency-domain harmonic energy features. Fast Fourier transform is used to calculate the energy ratios of the fundamental component, second harmonic, third harmonic, and high-frequency ripple, so that spectral changes caused by half-wave loss, clipping, and other waveform distortions can be represented.

The raw physical feature vector is mapped by a two-layer fully connected network before fusion. The first layer changes the 24-dimensional input into a 48-dimensional intermediate representation, and the second layer produces a 64-dimensional feature embedding aligned with the CNN branch

DOI: 10.54254/2755-2721/2026.35117

output dimensionality. ReLU activation is inserted between the two layers, and Dropout with a rate of 0.10 is used to reduce overfitting. The branch contains only about 4.3K parameters.

$$h_{phys} = FC_2(ReLU(Dropout(FC_1(f_{phys})))) \tag{2}$$

Here, $f_{phys} \in R^{24}$ denotes the original physical feature obtained by analytical calculation; the corresponding mapped embedding is used as the physical branch representation.

3.4. Gated adaptive fusion module

Concatenating the two feature vectors directly would treat both sources as equally useful for every sample. MFF-Net therefore introduces a gated adaptive fusion module. The module estimates a dimension-wise gating vector and uses it to mix the CNN feature with the physics-guided feature according to the current waveform.

The two 64-dimensional feature vectors are first concatenated as a 128-dimensional joint representation. This representation is fed to a small two-layer fully connected network with Sigmoid activation, from which the gating vector is obtained:

The branch outputs are first organized as a 128-dimensional joint representation $[h_{cnn}; h_{phys}]$, which then passes through a two-layer fully connected network with Sigmoid activation. The gating vector takes values between 0 and 1 in each dimension, indicating how much that dimension depends on the CNN waveform branch. The fused feature is computed as follows:

$$g = \sigma(W_2 \cdot ReLU(W_1 \cdot [h_{cnn}; h_{phys}] + b_1) + b_2) \tag{3}$$

$$h_{fused} = g \odot h_{cnn} + (1 - g) \odot h_{phys} \tag{4}$$

Here, σ denotes the Sigmoid function, and the multiplication in Eq. (4) is element-wise. When the waveform distortion is complicated and the physical indicators cannot fully describe the fault, the gate tends to give the CNN branch a larger weight. For faults whose physical indicators are already discriminative, such as those dominated by imbalance, the physics feature branch contributes more. The module therefore adjusts the feature advantage dynamically across samples.

The fused vector is sent to a two-layer fully connected classifier (64 to 64 to 4). ReLU and Dropout with a rate of 0.15 are applied between the layers, and class logits are trained by cross-entropy loss. AdamW is used with an initial learning rate of 1×10^{-3} and weight decay of 1×10^{-4}. The complete model has approximately 39.9K parameters, a size that is suitable for lightweight online diagnosis on constrained terminals.

4. Experiments

4.1. Experimental setup and dataset construction

The three-phase inverter fault waveform dataset is generated by numerical simulation. All experiments are carried out in Python 3.10 and PyTorch on a standard Intel CPU platform. Training is performed without GPU acceleration, so the experiment can also reflect the feasibility of lightweight deployment. The simulation settings are chosen to cover common operating variation. The sampling rate is 5,000 Hz, and each waveform contains 512 sampling points, corresponding to an observation window of about 102.4 ms. The fundamental frequency is randomly set within 48-52

DOI: 10.54254/2755-2721/2026.35117

Hz to represent grid disturbance. The amplitude is sampled from 0.8 to 1.2 for different load levels. In addition, 5th and 7th harmonic components, PWM switching ripple within 900-1,800 Hz, and Gaussian measurement noise are added. A three-phase amplitude imbalance coefficient of 0.95 to 1.05 is also introduced to imitate practical operating errors. Four operating conditions are included: normal, overcurrent, phase loss, and switch open-circuit. Each class contains 800 samples; in total, the dataset has 3,200 three-phase current waveform samples. Table 1 lists the definitions and waveform characteristics.

Table 1. Operating condition definitions and waveform characteristics

Condition	Waveform Characteristics
Normal	Three-phase currents approximately balanced with minor harmonics, switching ripple, and noise
Overcurrent	Overall amplitude increased 1.6-2.4x; 50% probability of transient impulse; 25% probability of clipping
Phase loss	One random phase attenuated to 2-15% of original; remaining phases slightly increased with possible DC offset
Switch open	Half-wave of one phase attenuated to 2-20%; accompanied by DC offset and increased harmonic distortion

The dataset is divided randomly into training, validation, and test subsets. The training set includes 2,240 samples, or 70% of the data. The validation set and test set contain 480 samples each, corresponding to 15% for each subset. During preprocessing, the per-channel mean and standard deviation are calculated only from the training set. The same normalization parameters are then applied to the validation and test sets, which avoids data leakage.

4.2. Training strategy and evaluation metrics

AdamW is adopted for model training. The initial learning rate is 1×10^{-3}, the weight decay coefficient is 1×10^{-4}, the batch size is 64, and the total training epoch number is 18. Cross-entropy loss is used as the objective function. During training, loss and accuracy are recorded on both the training set and the validation set. The model checkpoint with the best validation accuracy is retained for test evaluation. The evaluation uses four indicators. Overall classification accuracy reports the general recognition performance. Per-class accuracy compares the results for different fault types. The confusion matrix shows how the remaining misclassifications are distributed. PCA visualization is used to inspect the separability of categories in the fused feature space.

4.3. Training process analysis

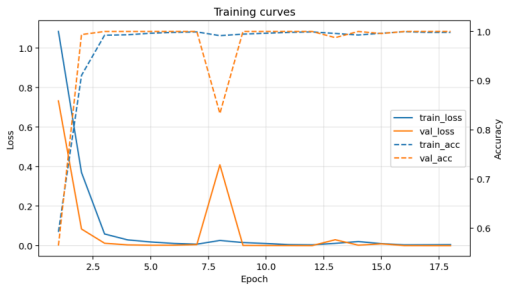

Figure 1. Training loss and accuracy curves of MFF-Net

DOI: 10.54254/2755-2721/2026.35117

The training loss and accuracy curves of MFF-Net are shown in Figure 1.

The curves indicate a fast but smooth convergence process. During the first several epochs, the training loss falls from about 1.1 to a low level, and the training accuracy increases quickly. The validation loss follows a similar downward trend. A minor fluctuation is visible near epochs 7-8, after which the curve becomes stable. Around the tenth epoch, the model has basically reached a steady training state.

No loss divergence or persistent decrease in validation accuracy is observed. The gap between training performance and validation performance also remains small. This suggests that the dual-branch design is not simply increasing model complexity; the physics-guided branch supplies useful prior information, while the lightweight CNN branch keeps the number of trainable parameters under control. Dropout and weight decay further limit overfitting. Because the model contains only about 39.9K parameters, its training cost is low enough for edge-oriented fault diagnosis studies.

4.4. Classification results and confusion matrix analysis

The test results of MFF-Net are summarized in Table 2. On 480 test samples, the model achieves an overall accuracy of 99.79%, and only one sample is classified incorrectly.

Table 2. MFF-Net test set classification accuracy

Class	Test Samples	Accuracy
normal	123	100.00%
overcurrent	118	99.15%
phase_loss	131	100.00%
switch_open	108	100.00%
Overall	480	99.79%

Normal samples, phase-loss faults, and switch open-circuit faults are all classified correctly. For the normal condition, the three-phase currents are nearly balanced, so the physical feature distribution is relatively stable. In phase-loss cases, RMS imbalance and zero-sequence component RMS increase clearly, making this class separable from the others. For switch open-circuit faults, the half-wave distortion is visible in the waveform, and the CNN branch can describe this morphology. The results show the complementary role of the physics-guided branch and the CNN branch.

The overcurrent class reaches an accuracy of 99.15%. The only error occurs on a boundary sample and is labeled as normal. Its amplitude increase is relatively small, so part of its physical feature distribution overlaps with high-load normal operation. Even in this borderline situation, the overall recognition performance remains high.

4.5. Physical feature separability analysis

The box plots in Figure 2 compare four key physical features: mean RMS, RMS imbalance, zero-sequence RMS, and high-frequency energy ratio.

DOI: 10.54254/2755-2721/2026.35117

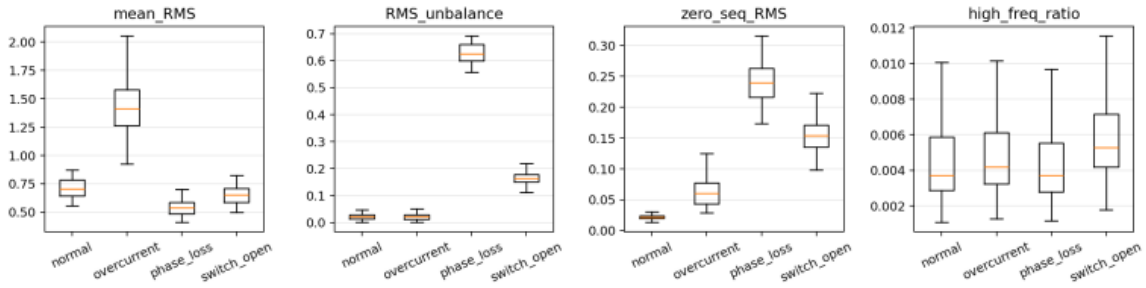

Figure 2. Box plot distributions of four operating conditions across key physical feature dimensions

Mean RMS is most helpful for overcurrent recognition, since overcurrent samples have higher amplitude than the other classes. Phase-loss samples instead show a lower mean RMS than normal samples because one phase is strongly attenuated. RMS imbalance and zero-sequence RMS are more closely related to phase-loss recognition. Under phase-loss faults, their values are much larger than those obtained under normal, overcurrent, and switch open-circuit conditions.

The high-frequency energy ratio has a wider overlap among classes, so its standalone discriminative ability is limited. Even so, overcurrent and switch open-circuit samples still show slightly stronger high-frequency components in many cases, which helps describe waveform distortion. The comparison indicates that the designed physical features are effective for overcurrent and phase-loss faults. Switch open-circuit faults, however, still depend on additional waveform morphology information, which supports the need for the dual-branch fusion structure.

4.6. Fused feature embedding visualization

The 64-dimensional fused features from the gated fusion layer are projected into two dimensions by PCA, and the result is presented in Figure 3.

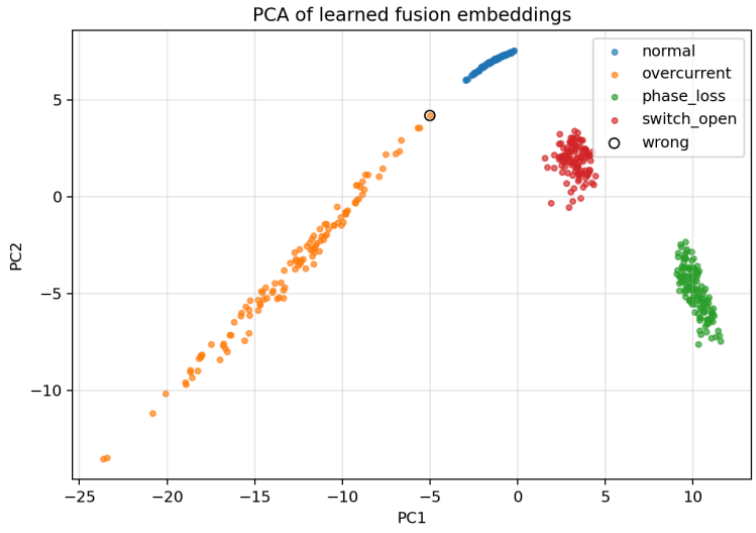

Figure 3. PCA two-dimensional visualization of test set fused feature embeddings

In the fused feature space, the four operating conditions form compact clusters with clear separation. Normal samples are the most concentrated, which agrees with their stable waveform and physical characteristics. Overcurrent samples extend along one principal component in a strip-like

DOI: 10.54254/2755-2721/2026.35117

distribution, reflecting the continuous change of overcurrent severity. Phase-loss and switch open-circuit samples form their own clusters and stay apart from the other categories.

The single misclassified sample is located near the boundary between the normal and overcurrent clusters, matching the confusion matrix result. The visualization also shows that the gated fusion mechanism combines deep temporal features with physics-guided prior features in a useful way. After fusion, the representation has strong discrimination among fault categories and supports high-accuracy fault waveform classification.

5. Conclusion

This study presents MFF-Net, a lightweight physics-guided feature fusion network for fault waveform classification in three-phase inverter circuits. The method is built around two practical concerns in inverter fault diagnosis: traditional methods rely heavily on expert-designed features, whereas purely data-driven deep learning models may lack physical interpretability and may be inconvenient for lightweight online systems. MFF-Net adopts a dual-branch architecture. The lightweight 1D-CNN branch learns temporal waveform features, such as waveform distortion, half-wave loss, and high-frequency ripple, from raw three-phase current signals. The physics-guided branch extracts 24-dimensional electrical features, including statistical features, imbalance indicators, zero-sequence components, and harmonic energy ratios. A gated adaptive fusion module then combines the two branches in a sample-dependent manner, allowing data-driven representation and physical prior knowledge to work together. Experiments on the simulated dataset with 3,200 samples show that the model obtains 99.79% classification accuracy on the test set. The training process remains stable, and no obvious overfitting is observed. The confusion matrix and PCA visualization further indicate that the fused feature space separates the four operating conditions clearly. With about 39.9K parameters, the model is suitable for further study in edge deployment and online fault diagnosis. This work is still mainly based on simulated data. Future research will add real operating condition data, cover more complex compound fault types, and analyze the interpretability of the gated fusion process in greater detail.

References

[1] Villa-Ávila E, Ochoa-Correa D, Arévalo P. Advancements in power converter technologies for integrated energy storage systems: optimizing renewable energy storage and grid integration [J]. Processes, 2025, 13(6): 1819.
[2] Li J, Liu Y, Wu X, et al. Fault diagnosis in open circuit of inverters on electrical discharge milling machines using adaptive Gaussian wavelet convolutional network [J]. Measurement, 2025, 248: 116856.M. S. Priyadarshini, M. Bajaj, L. Prokop, et al., "Perception of power quality disturbances using Fourier, Short-Time Fourier, continuous and discrete wavelet transforms, " Scientific Reports, vol. 14, no. 1, p. 3443, 2024.
[3] Priyadarshini M S, Bajaj M, Prokop L, et al. Perception of power quality disturbances using Fourier, Short-Time Fourier, continuous and discrete wavelet transforms [J]. scientific reports, 2024, 14(1): 3443.
[4] Said N, Mansouri M, Al Hmouz R, et al. Deep learning techniques for fault diagnosis in interconnected systems: A comprehensive review and future directions [J]. Applied Sciences, 2025, 15(11): 6263.
[5] Zhang X, Hu Y, Deng J, et al. Feature engineering and artificial intelligence-supported approaches used for electric powertrain fault diagnosis: A review [J]. IEEE Access, 2022, 10: 29069-29088.
[6] Duc M L, Bilik P, Martinek R. Harmonics signal feature extraction techniques: A review [J]. Mathematics, 2023, 11(8): 1877.
[7] Al Kharusi K, El Haffar A, Mesbah M. Fault detection and classification in transmission lines connected to inverter-based generators using machine learning [J]. Energies, 2022, 15(15): 5475.
[8] Liu K, Zhou J, Jin C, et al. Artificial intelligence-based open-circuit fault diagnosis for power electronic converters: Recent advances and future prospects [J]. IEEE Transactions on Power Electronics, 2025.

DOI: 10.54254/2755-2721/2026.35117

[9] Cui B, Zhang S, Su J, et al. Fault diagnosis for inverter-fed motor drives using one dimensional complex-valued convolutional neural network [J]. IEEE Access, 2024, 12: 117678-117690.
[10] He C, Meng Q, Xu X, et al. An improved lightweight residual network model deployed on the edge device for the unsupervised cross-domain fault diagnosis [J]. Expert Systems with Applications, 2025: 129106.
[11] Mao C, Jin Y. Uncertainty quantification study of the physics-informed machine learning models for critical heat flux prediction [J]. Progress in Nuclear Energy, 2024, 170: 105097.
[12] Lei L, Li W, Zhang S, et al. Research progress on data-driven industrial fault diagnosis methods [J]. Sensors, 2025, 25(9): 2952.
[13] Xie T, Xu Q, Jiang C, et al. The fault frequency priors fusion deep learning framework with application to fault diagnosis of offshore wind turbines [J]. Renewable Energy, 2023, 202: 143-153.
[14] Sun X, Yang Y, Chen C, et al. A multi-branch convolution and dynamic weighting method for bearing fault diagnosis based on acoustic–vibration information fusion [C]//Actuators. MDPI, 2025, 14(1): 17.
[15] Nie Q, Geng J, Liu C. A Review of Fault Diagnosis Methods: From Traditional Machine Learning to Large Language Model Fusion Paradigm [J]. Sensors, 2026, 26(2): 702.