EAI Endorsed Transactions
on Energy Web
Research Article EAI.EU

Lightweight Real-Time power quality disturbance recognition using Time-Frequency fusion with Cross-Attention mechanism

Fei Yu, Haoyang Xue *, Jiaming Qi, Xiaoqian Yue, Xinrui Pei

Naval University of Engineering, Wuhan, 430033, Hubei, China

Abstract

INTRODUCTION: Accurate power quality disturbances (PQDs) classification is critical for maintaining grid stability and reliability in modern power systems. However, existing deep learning methods predominantly rely on single-domain feature extraction, limiting their discriminative capability for complex composite disturbances under noisy conditions. This study addresses these limitations by proposing a dual-pathway architecture that synergistically integrates time-domain and frequency-domain representations through cross-attention fusion.
OBJECTIVES: This work aims to develop a robust PQDs classification framework capable of accurately identifying 24 disturbance classes, including complex composite types, while maintaining high noise immunity and computational efficiency for real-time monitoring applications.
METHODS: A dual-pathway deep learning architecture is proposed, comprising parallel CNN-BiLSTM branches for time-domain temporal modeling and FFT-based frequency-domain spectral analysis. A cross-attention mechanism dynamically fuses complementary features from both pathways. The model is trained and evaluated on a comprehensive dataset containing 24 PQDs classes under multiple noise levels.
RESULTS: The proposed model achieves 99.73% accuracy on the validation set and maintains 98.94% accuracy under 30dB noise conditions. Ablation studies confirm the dual-pathway structure improves accuracy by 6.51 percentage points over single-branch variants, while the cross-attention mechanism contributes an additional 2.08 percentage points. The model converges within 43 epochs with inference latency of 251μs per sample, satisfying real-time requirements.
CONCLUSION: The proposed dual-pathway cross-attention architecture demonstrates superior performance in PQDs classification, effectively balancing accuracy, noise robustness, and computational efficiency. This approach provides a viable solution for intelligent power quality monitoring in practical smart grid applications.

Keywords: Power quality disturbances, Deep Learning, Dual-Pathway Architecture, Cross-Attention Mechanism, Time-Frequency Fusion, Smart Grid Monitoring.

Received on 08 September 2025, accepted on 19 December 2025, published on DD MM YYYY

Copyright © 2026 Fei Yu *et al.*, licensed to EAI. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/ew.12734

1. Introduction

The increasing integration of distributed energy resources (DERs) and power electronics converters has introduced new challenges to power system operation and control [1-2]. Power quality disturbances (PQDs) represent a major concern, as they can lead to renewable generation tripping, sensitive equipment malfunctions, and substantial economic losses [3-4]. PQDs manifest as deviations in voltage, current, or frequency from standard operating conditions, encompassing both single-type events (e.g., harmonics, voltage sags/swells) and composite disturbances involving multiple simultaneous phenomena [5-6]. Conventional

*Corresponding author. Email: d24180802@nue.edu.cn

detection techniques face limitations when addressing the diverse PQDs patterns and high noise levels characteristic of converter-dominated grids. Developing accurate and noise-resilient PQDs classification methods is therefore critical for enhancing grid operational security and power supply reliability.

Deep learning has become the predominant approach for PQDs recognition, demonstrating superior performance over traditional signal processing methods. Recent studies have explored various architectures including Convolutional Neural Networks (CNNs) [7-8], attention mechanisms [9-10], and hybrid frameworks [11-12]. Wavelet-based deep learning models achieved high accuracy but introduced substantial preprocessing overhead limiting real-time deployment [7]. Graph Neural Networks (GNNs) combined with multi-scale feature extraction demonstrated competitive performance but required extensive computational resources [8]. Attention mechanisms have been incorporated to enhance feature discrimination, with self-attention CNNs reaching notable accuracy on multi-class disturbances, yet computational complexity remains a concern for edge deployment [9-10]. Several approaches employed Long Short-Term Memory (LSTM) and Transformer architectures to capture temporal dependencies [11-12], while others utilized residual networks with time-frequency image transformations [13]. However, these methods predominantly process time and frequency domain information sequentially or independently, lacking effective cross-domain fusion mechanisms. Furthermore, the trade-off between classification accuracy and computational efficiency remains unresolved—sophisticated preprocessing or deep architectures achieve high accuracy but are unsuitable for resource-constrained environments [13-14]. This motivates the need for lightweight architectures that efficiently integrate complementary time-frequency features while maintaining real-time inference capability and noise robustness.

To address composite PQDs classification challenges in converter-dominated grids, this paper proposes a lightweight time-frequency fusion network with cross-attention mechanism. The key innovations include:

(i) Dual-branch feature extraction—time-domain CNN-BiLSTM captures temporal dynamics while FFT-based frequency-domain CNN-BiLSTM extracts spectral characteristics, avoiding computationally intensive preprocessing.

(ii) Cross-attention fusion mechanism that establishes adaptive correlation between time and frequency features, enhancing discrimination of complex composite disturbances.

(iii) Lightweight architecture achieving 99.56% accuracy with only 0.97M parameters and 0.251ms inference latency, enabling real-time edge deployment.

The method demonstrates robust performance across 24 disturbance classes under mixed noise conditions effectively balancing accuracy, efficiency, and noise resilience.

2. Proposed method

2.1. Overall architecture

This paper proposes a dual-branch PQDs recognition model based on FFT-CNN-BiLSTM with cross-attention fusion mechanism. The overall framework is illustrated in Fig.1. The raw signal directly enters the time-domain branch to preserve transient features (voltage sags/swells, impulses, oscillatory transients), while FFT is applied to extract frequency-domain representations characterizing steady-state attributes (harmonics, inter-harmonics). Both branches employ CNN-based convolution-pooling structures for hierarchical feature extraction, followed by BiLSTM networks to capture long-range temporal dependencies. In the fusion stage, a cross-attention mechanism achieves adaptive alignment between time-frequency features: the frequency-domain output serves as *Query* while the time-domain output serves as *Key-Value*, enabling frequency positions to retrieve and aggregate the most relevant temporal segments. This mechanism highlights discriminative components while suppressing noise, yielding representations with both frequency discriminability and temporal consistency. Finally, fused features are compressed via adaptive average pooling and classified through fully connected layers with Softmax.

2.2. Fast fourier transform

Fast Fourier Transform (FFT) is employed to convert time-domain signals into frequency-domain representations, enabling characterization of spectral components and energy distribution [15]. The formal definition of FFT is given in Equation (1):

$$X(k) = \sum_{n=0}^{N-1} x(n)\ e^{-i2\pi kn/N} \tag{1}$$

Unlike computationally intensive alternatives such as Short-Time Fourier Transform (STFT) or Wavelet Transform (WT), FFT reduces computational complexity from $O(N^2)$ to $O(N\ log\ N)$ while maintaining consistent frequency resolution across the entire spectrum. The proposed architecture leverages the complementary nature of frequency-domain features (FFT) and time-domain features (CNN-BiLSTM): the CNN-BiLSTM module captures temporal dynamics and local patterns, while FFT provides critical spectral information that is difficult to obtain through time-domain analysis alone. This dual-domain fusion strategy is particularly advantageous for PQDs recognition—transient events exhibit precise localization characteristics in the time domain, whereas steady-state disturbances such as harmonic distortion are more prominent in the frequency domain. Furthermore, FFT integrates seamlessly into deep learning frameworks, supporting end-to-end optimization with cross-attention mechanisms for enhanced feature fusion.

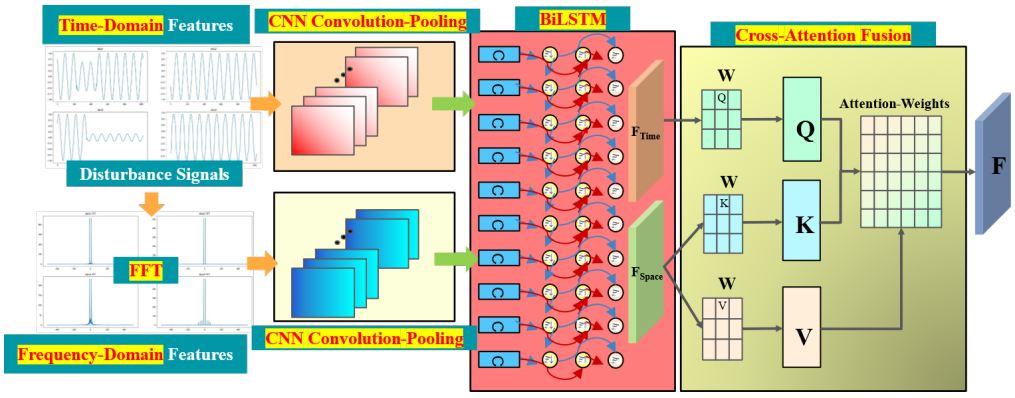


Figure 1. Overall architecture of the model recognition process

2.3. CNN architecture in Time-Frequency dual branches

CNNs originally designed for computer vision tasks, have proven equally effective for one-dimensional time-series feature learning [16]. The CNN module in this work adopts a dual-branch architecture for time-domain and frequency-domain feature extraction respectively, as illustrated in Fig.2. Each branch consists of stacked convolution-pooling blocks: convolutional layers employ multiple kernels sliding over the input with predefined strides, extracting hierarchical local patterns; pooling layers perform spatial downsampling via max or average operations, reducing dimensionality while enhancing robustness against minor input variations. The output of the j-th neuron in the l-th convolutional layer is formulated in Equation (2):

$$a_j^l = f\left(b_j^l + \sum_{i \in M_j^l} a_i^{l-1} \cdot k_{ij}^l\right) \tag{2}$$

This multi-scale feature extraction mechanism progressively learns abstract representations from low-level to high-level, providing rich feature information for subsequent BiLSTM temporal modeling and cross-attention fusion.

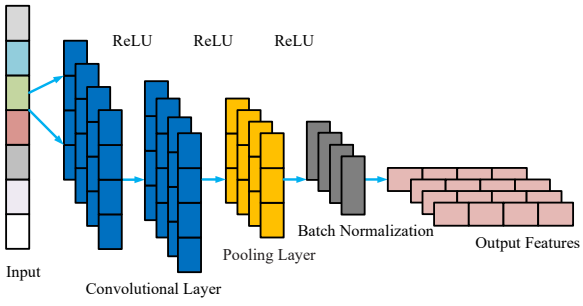


Figure 2. Diagram of CNN feature extraction module structure

2.4. Bidirectional context modeling and feature enhancement via BiLSTM

Bidirectional Long Short-Term Memory (BiLSTM), a variant of recurrent neural networks, is specifically designed for sequential data modeling through its innovative bidirectional architecture: the forward LSTM processes historical context while the backward LSTM captures future context, enabling comprehensive temporal dependency modeling at each time step [17]. The detailed structures of LSTM unit and BiLSTM network are illustrated in Fig.3(a) and Fig.3(b) respectively.

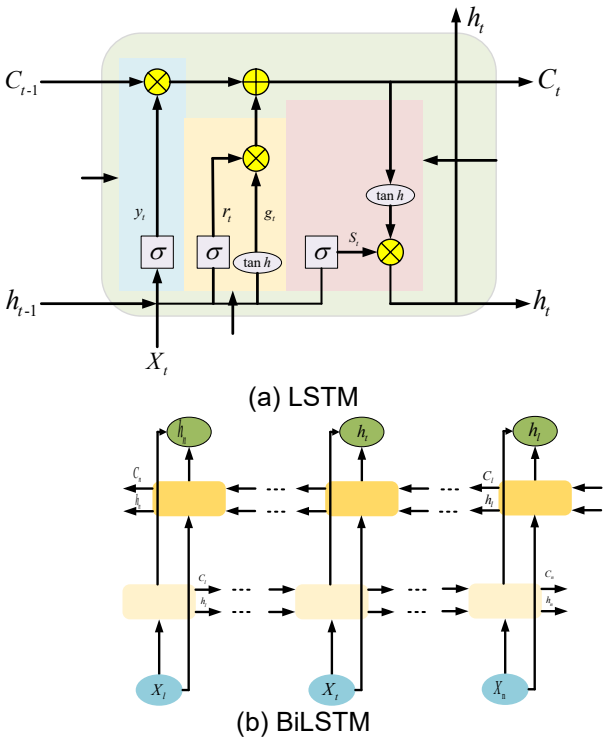


Figure 3. Structure diagram

With the gate computations formulated in Equation (3):

$$\begin{cases} y_t = \sigma(\boldsymbol{W}_f \cdot [h_{t-1}, X_t] + \boldsymbol{b}_f) \\ r_i = \sigma(\boldsymbol{W}_i \cdot [h_{t-1}, X_t] + \boldsymbol{b}_i) \\ s_t = \sigma(\boldsymbol{W}_o \cdot [h_{t-1}, X_t] + \boldsymbol{b}_o) \\ g_t = \tan h(\boldsymbol{W}_g \cdot [h_{t-1}, X_t] + \boldsymbol{b}_g) \\ C_t = y_t C_{t-1} + r_t g_t \\ h_t = s_t \tan h(C_t) \end{cases} \quad (3)$$

Unlike CNNs that focus on local spatial features, BiLSTM effectively addresses gradient vanishing and exploding problems, significantly enhancing long-range sequence modeling capability. In the proposed model, BiLSTM is placed after both time-domain and frequency-domain CNN branches to perform cross-temporal context modeling and long-range dependency capture. Through simultaneous forward and backward sequence channels, BiLSTM enables local segments to be reorganized and enhanced within the global temporal context, thereby more accurately characterizing disturbance evolution trajectories and duration patterns. The enriched sequential representations provide high-quality inputs for subsequent Cross-Attention fusion, ensuring semantic coherence during time-frequency alignment and weighting.

2.5. Time-Frequency complementary fusion via Cross-Attention

This section establishes a Cross-Attention fusion mechanism to integrate multi-source features extracted from time-domain and frequency-domain branches. As illustrated in Fig.4, unlike conventional feature concatenation, cross-attention uses frequency-domain features as *Query (Q)* and time-domain features as *Key-Value (K-V)*, enabling directional information flow for enhanced feature extraction [18].

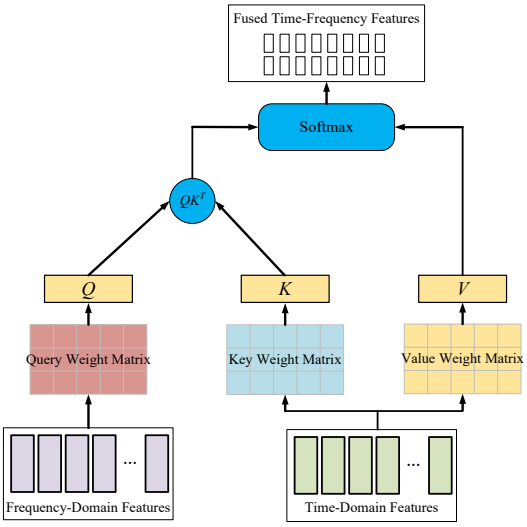


Figure 4. Structure of Cross-Attention fusion mechanism

The attention computation is formulated in Equation (4):

$$selfAttention(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = sotmax(\frac{\boldsymbol{Q} \times \boldsymbol{K}^T}{\sqrt{d_k}}) \cdot \boldsymbol{V} \quad (4)$$

Attention weights are obtained through linear projections, matrix multiplication, and Softmax normalization, adaptively adjusting the relative contributions of both domains to highlight discriminative components while suppressing irrelevant noise. Specifically, frequency-domain positions can adaptively "retrieve" the most relevant temporal segments through the attention matrix, which characterizes correlations between spectral components and different time slices. The fused result preserves steady-state spectral information (harmonics, inter-harmonics) while explicitly aligning and enhancing transient semantics such as disturbance onset/offset and transition slopes. The fused time-frequency sequence is then compressed into a global discriminative vector via adaptive average pooling and fed into fully connected layers for classification. Compared to simple concatenation, Cross-Attention dynamically allocates feature contributions in a "frequency-guided, time-supported" manner.

3. Simulation experiments

3.1. Dataset construction

Based on the IEEE-1159 standard for power quality disturbance classification and parameter specifications, 24 types of PQDs are synthesized using Python. The dataset comprises 9 single disturbances: Normal (C1), Voltage Swell (C2), Voltage Sag (C3), Harmonics (C4), Flicker (C5), Interruption (C6), Impulsive Transient (C7), Oscillatory Transient (C8), and Notch (C9); and 15 composite disturbances: Harmonics+Swell (C10), Harmonics+Sag (C11), Harmonics+Interruption (C12), Harmonics+Flicker (C13), Harmonics+Impulsive (C14), Harmonics+Oscillatory (C15), Flicker+Swell (C16), Flicker+Sag (C17), Flicker+Oscillatory (C18), Flicker+Impulsive (C19), Swell+Oscillatory (C20), Sag+Oscillatory (C21), Harmonics+Oscillatory+Swell (C22), Harmonics+Oscillatory+Sag (C23), and Harmonics+Oscillatory+Flicker (C24). Representative waveforms of typical PQDs are illustrated in Fig.5.

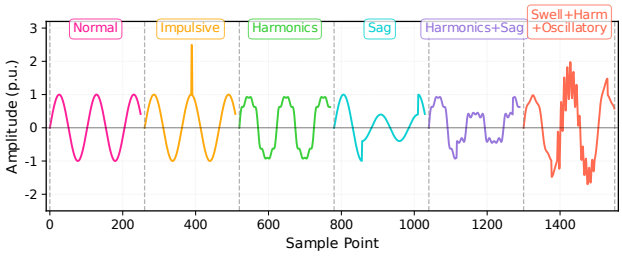


Figure 5. Representative waveforms of typical PQDs

The fundamental frequency is set to 50 Hz with a sampling frequency of 5,120 Hz, resulting in a sample length of 1,024 points covering 10 fundamental cycles. For training and validation, 350 clean samples per class are generated and mixed 1:1 with 30 dB SNR noisy versions, with an additional 5% augmentation at 50 dB SNR. This mixed dataset is then split into training and validation sets at an 80:20 ratio using stratified sampling. For testing, independent test sets are constructed separately at four SNR levels (no noise, 30 dB, 40 dB, 50 dB), with 100 samples per class at each level. All test sets are generated from the same baseline signals to ensure paired consistency across noise conditions.

3.2. Experimental setup

The PQDs dataset construction procedure and the software/hardware specifications for model training are consolidated in Table 1 to facilitate experimental reproducibility and comparative analysis.

Table 1. Overall environment

Component	Specification
Operating System	Windows11(64)24H2
Python	3.11
PyTorch	2.7.1
CUDA	12.6
CPU	Ultra7 265K
GPU	NVIDIA GeForce RTX 5070Ti/16G
RAM	DDR5 32GB

3.3. Model training and ablation study

The proposed model employs an end-to-end deep learning framework with the Adam optimizer at a learning rate of 3×10^{-4}, leveraging its adaptive gradient properties for efficient parameter updates. To ensure reproducibility, a fixed random seed (seed=100) is applied, and deterministic computation is enabled in the GPU environment.

Training adopts mini-batch gradient descent with a batch size of 64. Data loading utilizes multi-process parallelization (num_workers=2), where the training set undergoes random shuffling to enhance generalization, while the validation set maintains a fixed order for consistent evaluation. Each epoch comprises four sequential steps: forward propagation, loss computation, backpropagation, and parameter optimization.

Model outputs are converted into probability distributions via the Softmax function, and classification errors are quantified using cross-entropy loss, which exhibits favorable mathematical properties for optimization in multi-class tasks, as expressed in Equation (5):

$$L_{oss} = -\frac{1}{BS}\sum_{d=1}^{BS}\sum_{f=1}^{N_{cls}} y_{df}\log(\hat{y}_{df}) \tag{5}$$

To mitigate overfitting, an early stopping mechanism is implemented: training halts when validation loss improvement falls below 0.001 over 10 consecutive epochs, and the model weights corresponding to the highest validation accuracy are automatically restored. The maximum training duration is set to 100 epochs, with actual termination determined dynamically by early stopping. Throughout training, metrics including training accuracy (Acc), loss, validation accuracy (Val-acc), and validation loss (Val-loss) are monitored in real-time to facilitate performance analysis.

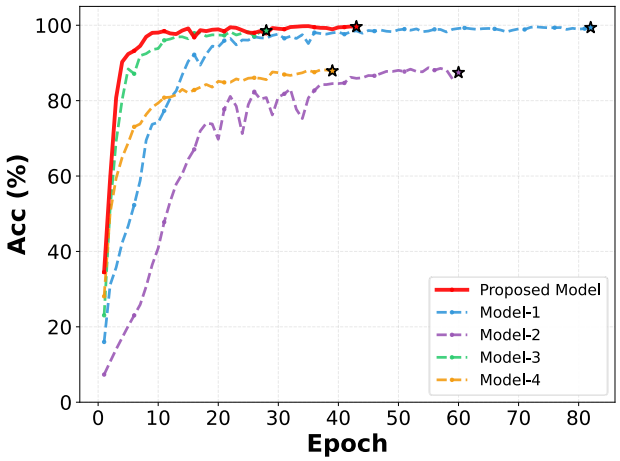
To validate the contribution of each component in the FFT-CNN+BiLSTM-Cross-Attention architecture, four ablated variants are constructed using a controlled variable approach:

- Model-1: Time-domain branch only (removes FFT branch) — evaluates frequency-domain contribution.
- Model-2: Frequency-domain branch only (removes time-domain branch) — assesses time-domain contribution.
- Model-3: Dual-branch with direct concatenation (removes Cross-Attention) — quantifies the benefit of adaptive feature fusion.
- Model-4: Dual-branch CNN without BiLSTM — examines the impact of temporal dependency modeling on complex disturbance evolution.

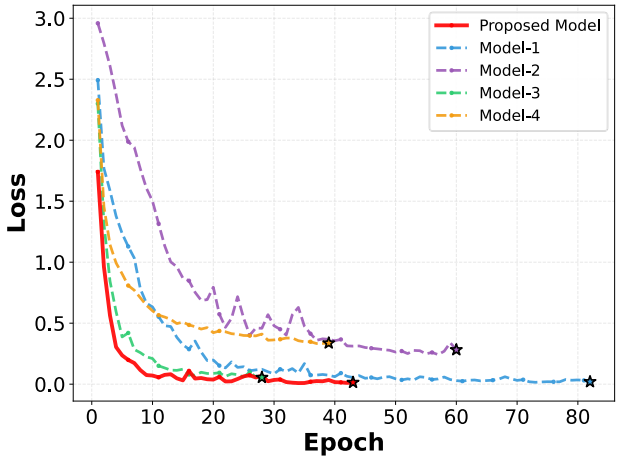
As illustrated in Fig.6, the complete model achieves 99.56% validation accuracy with early stopping at epoch 43. Model-1 exhibits a marginal accuracy drop of 0.06%, yet requires 82 epochs—nearly double the convergence time. This indicates that while frequency-domain features are non-essential for accuracy, they significantly accelerate training convergence. Conversely, Model-2 suffers a substantial accuracy degradation to 89.59%, revealing the dominant role of time-domain features in PQDs recognition. This asymmetry stems from the fact that transient characteristics and amplitude variations are more pronounced in the time domain, whereas frequency-domain information primarily captures periodic disturbances and harmonic components.

The ablation experiments on fusion mechanism and sequence modeling further validate the network design rationale. Model-3 exhibits only a 0.36% validation accuracy drop but converges in merely 28 epochs, suggesting that direct feature concatenation preserves dual-branch information yet lacks adaptive weighting capability, resulting in insufficient exploitation of discriminative features. More critically, Model-4 (without BiLSTM) suffers a substantial performance degradation to 90.24%, approaching the frequency-only baseline. Analysis of training curves reveals pronounced oscillations in Model-4's validation loss,

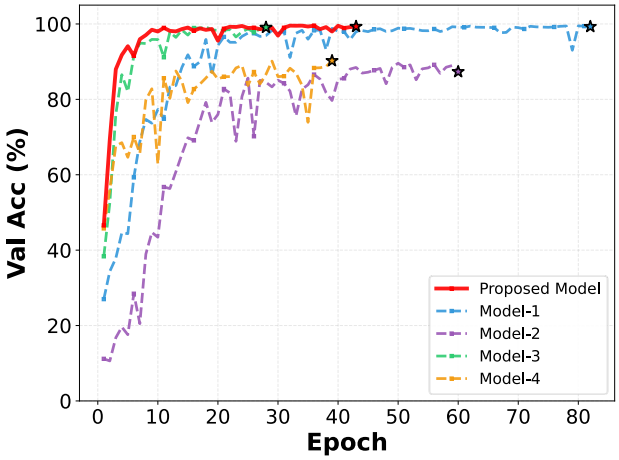
indicating poor stability. This confirms that BiLSTM is indispensable for capturing long-term dependencies in disturbance signals. Collectively, the four ablation studies demonstrate that the dual-branch architecture, Cross-Attention fusion, and BiLSTM sequence modeling synergistically establish the performance foundation of the proposed model.



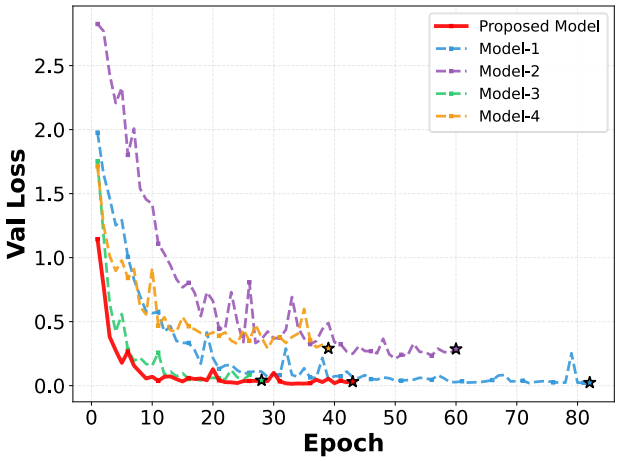
(a) Training accuracy



(b) Training loss



(c) Validation accuracy



(d) Validation loss

Figure 6. Training process of each classification model

3.4. Noise robustness evaluation

To systematically evaluate model robustness under realistic noise conditions, this section employs four independent test sets constructed in Section 3.1: noise-free, 50 dB, 40 dB, and 30 dB SNR levels. Equal-sized sampling is performed across all PQDs classes at each SNR level to ensure fair evaluation and mitigate class imbalance effects on performance metrics.

The comparative study includes the four ablation variants (Model-1 to Model-4) presented in Section 3.3, alongside the 1D-ConvNeXt-Tiny classification model from [19] as a strong baseline to benchmark against modern convolutional architectures on time-series signals. All models are evaluated using their best-performing weights on the validation set. To ensure fairness and reproducibility, all experiments adopt identical data partitioning, preprocessing pipelines, and hyperparameter configurations.

Four widely-used classification metrics are employed for performance evaluation: Accuracy, Precision, Recall, and F1-Score. Accuracy quantifies the overall correctness of predictions across all samples. Precision measures the proportion of true positives among samples classified as positive, emphasizing prediction purity. Recall captures the proportion of actual positive samples correctly identified, reflecting the model's coverage capability. F1-Score, as the harmonic mean of Precision and Recall, balances both metrics and is particularly effective in scenarios with class imbalance or asymmetric costs between false positives and false negatives. These complementary metrics collectively provide a comprehensive assessment of classifier performance and robustness, as expressed in Equation (6):

$$\begin{cases} A_{accuracy} = \dfrac{N_C}{N_C + N_I} \\ P_{precision} = \dfrac{T_p}{T_p + F_p} \\ R_{recall} = \dfrac{T_P}{T_P + F_N} \\ F_{1-score} = \dfrac{1}{1/R_{recall} + 1/P_{precision}} \end{cases} \tag{6}$$

Experimental results are summarized in Table 2. Under moderate-to-high SNR conditions (≥30 dB), the proposed model demonstrates superior classification performance. On the noise-free test set, it achieves 99.20% Accuracy, 99.23% Precision, 99.20% Recall, and 99.20% F1-Score, outperforming the 1D-ConvNeXt-Tiny baseline by approximately 1 percentage point. As noise intensity increases to 30 dB SNR, the proposed model maintains 98.94% accuracy compared to the baseline's 97.38%, widening the performance gap to 1.56 percentage points. This validates the superiority of the dual-branch fusion strategy in noisy environments.

Ablation analysis reveals the contribution of each component: Model-1 (time-domain only) maintains 98.78%–99.11% accuracy under high SNR, comparable to the full model, while Model-2 (frequency-domain only) degrades to 85.60%–87.80%. This contrast confirms the dominance of time-domain features in disturbance characterization, though frequency-domain information provides non-negligible complementary benefits. Model-3 (concatenation fusion) exhibits 0.5–1 percentage point accuracy loss, whereas Model-4 (no BiLSTM) drops sharply to 89%–92%, demonstrating the necessity of attention-based fusion and temporal modeling for robust feature representation.

Table 2. Classification performance comparison under different SNR conditions

SNR	Model	Evaluation Metrics			
		Accuracy	Precision	Recall	F1-Score
0dB	PM	0.9920	0.9923	0.9920	0.9920
	1	0.9911	0.9914	0.9911	0.9912
	2	0.8780	0.8865	0.8781	0.8739
	3	0.9848	0.9856	0.9848	0.9847
	4	0.9185	0.9284	0.9184	0.9186
	BL	0.9821	0.9828	0.9821	0.9821
50dB	PM	0.9924	0.9927	0.9924	0.9924
	1	0.9911	0.9914	0.9911	0.9912
	2	0.8767	0.8859	0.8769	0.8728
	3	0.9852	0.9859	0.9852	0.9851
	4	0.9134	0.9246	0.9134	0.9137
	BL	0.9825	0.9831	0.9825	0.9825
40dB	PM	0.9911	0.9914	0.9911	0.9911
	1	0.9899	0.9903	0.9899	0.9899
	2	0.8780	0.8867	0.8781	0.8739
	3	0.9844	0.9852	0.9844	0.9843
	4	0.9046	0.9189	0.9045	.9051
	BL	0.9829	0.9835	0.9829	0.9829
30dB	PM	0.9894	0.9899	0.9894	0.9895
	1	0.9878	0.9882	0.9878	0.9878
	2	0.8560	0.8685	0.8562	0.8518
	3	0.9848	0.9855	0.9848	0.9848
	4	0.8974	0.9121	0.8973	0.8971
	BL	0.9738	0.9745	0.9737	0.9737

Note: PM - Proposed Model; BL - 1D-ConvNeXt-Tiny; M-1 to M-4 - Ablation variants.

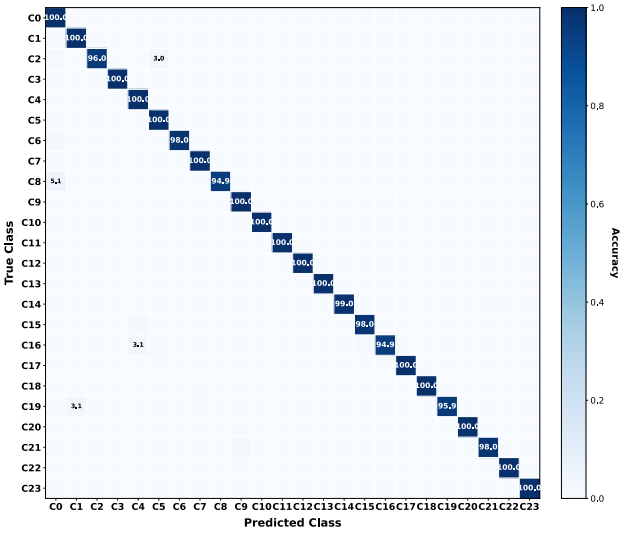


Figure 7. Detailed classification results of the proposal with a 30dB noise environment

To further reveal classification patterns under noisy conditions, Fig.7 presents the confusion matrix of PM at 30 dB SNR. This noise level is representative as it introduces moderate interference while maintaining high accuracy, enabling detailed performance analysis. The confusion matrix, as a multi-class diagnostic tool, decomposes overall accuracy into class-specific performance profiles: employing row normalization, diagonal elements directly represent per-class Recall, facilitating rapid identification of weak classes; off-diagonal elements reveal misclassification pathways, such as 4.04% of C2 samples being confused with similar waveform classes. This visualization transcends single-metric limitations, enabling quantitative assessment of confusion mechanisms between feature-similar disturbances

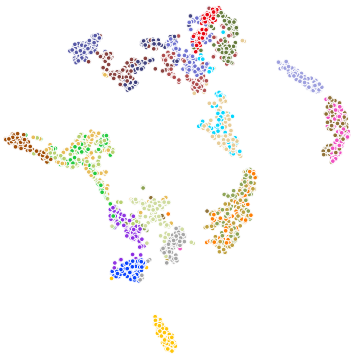
and the model's discriminative capability for composite events, thereby providing precise targets for feature refinement and decision boundary optimization.

The confusion matrix reveals that PM achieves 98.94% overall accuracy on the 24-class PQDs recognition task at 30 dB SNR, with 18 classes attaining ≥98% accuracy. This demonstrates the dual-branch architecture's capability to capture discriminative features across diverse disturbance types, particularly for harmonics (distinct frequency signatures) and interruptions (abrupt time-domain transitions). The high recognition rates for composite disturbances (C9-C23) further validate the Cross-Attention mechanism's effectiveness in fusing complementary time-frequency information. However, minor misclassifications occur in specific cases: C8 (notch) exhibits relatively lower accuracy due to its subtle localized waveform distortion, easily confused with normal signals under noise; C16 and C19 (flicker-based composites) show confusion with their constituent single disturbances (C2, C7), indicating challenges in distinguishing composites from dominant components when feature imbalance exists. Overall, the 1.06% misclassification rate concentrates among feature-similar class pairs, providing interpretable insights for targeted model refinement.

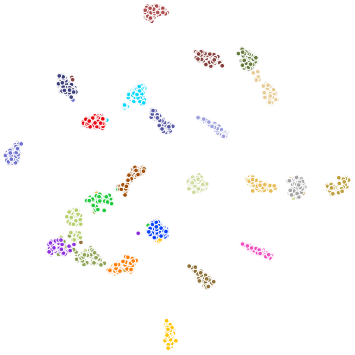
3.5. Time-Frequency feature visualization

Beyond classification performance, the distribution of feature vectors in the embedding space serves as a critical indicator of feature effectiveness [20]. Model representations evolve progressively across network layers, intuitively revealing each module's contribution to discriminative information. However, PQDs signals originate as one-dimensional time-series data and are subsequently mapped to high-dimensional spaces through time-frequency convolution and temporal modeling, rendering direct observation of their distribution patterns impractical. To address this challenge, t-SNE (t-distributed Stochastic Neighbor Embedding), a manifold-based dimensionality reduction algorithm, is employed to project high-dimensional features into two-dimensional embedding space, enabling visual comparison of discriminative structures across network layers, as illustrated in Fig.8.

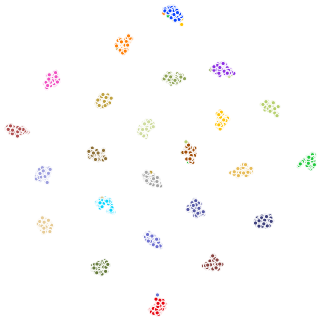
(a) Raw fignal



(b) Time-Frequency convolutional features



(c) BiLSTM temporal features



(d) Cross-Attention fusion

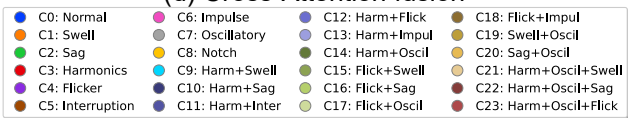


Figure 8. t-SNE visualization

Fig.8 illustrates the progressive enhancement of feature discriminability across network layers through t-SNE visualization under 30dB noise. The raw signal space (Figure 8a) exhibits highly chaotic distributions with extensive overlap, particularly among complex disturbances (C15-C23), indicating insufficient discriminative capacity. Time-frequency convolution (Figure 8b) partially alleviates this chaos as simple disturbances (C0-C4) begin clustering, though substantial overlap persists. BiLSTM temporal modeling (Figure 8c) markedly improves structural organization with enhanced cluster cohesion, reflecting bidirectional temporal regularization effects. The cross-

attention fusion stage (Figure 8d) achieves qualitative transformation: previously overlapping classes separate into 24 well-defined regions with compact, near-Gaussian distributions. Even triple-composite disturbances (C21-C23) maintain distinguishable boundaries, validating the dual-pathway architecture's effectiveness. The substantially increased ratio of inter-cluster distance to intra-cluster density provides superior linear separability. This evolutionary trajectory—local feature extraction → temporal dependency modeling → cross-domain fusion—demonstrates progressive discriminative enhancement, with cross-attention serving as the critical component achieving discriminability breakthrough through dynamic complementary information weighting.

3.6. Model computational complexity evaluation

To assess the engineering feasibility of the proposed model, a unified computational complexity evaluation is conducted on all models investigated in Sections 3.3 and 3.4, with training time and inference latency recorded for each architecture. The statistical results presented in Table 3 provide quantitative insights into the trade-offs between resource consumption and inference efficiency across different model variants, offering practical guidance for subsequent deployment and application scenarios.

Table 3. Time-consuming comparisons of different models

Model	Total Epochs	Avg. Training Time per Epoch (s)	Avg. Inference Time per Sample (μs)
Proposed Model	43	8.22	251
model-1	82	7.02	31.3
model-2	60	5.7	19.1
model-3	28	7.07	48
model-4	28	7.07	15.7
1D-ConvNeXt-Tiny	41	13.61	153.9

From a training efficiency perspective, the proposed model achieves convergence after 43 epochs, substantially faster than Model-1 (82 epochs) and Model-2 (60 epochs), demonstrating that the dual-pathway architecture with cross-attention effectively accelerates feature learning. Although Model-3 and Model-4 converge within 28 epochs, their inferior accuracy indicates that moderate architectural complexity enables both performance preservation and rapid convergence. Regarding per-epoch training time, the proposed model requires 8.22s, slightly higher than ablation variants yet considerably lower than 1D-ConvNeXt-Tiny (13.61s), attributed to the superior computational efficiency of parallel dual-branch processing compared to deep convolutional networks. In terms of inference performance, the proposed model exhibits an average latency of 251μs per sample, which, despite being marginally higher than structurally simplified Model-2 and Model-4, remains acceptable given its substantial accuracy advantage. Furthermore, the 251μs latency fully satisfies real-time requirements for PQDs monitoring applications. Collectively, considering convergence speed, classification accuracy, and inference latency, the proposed model achieves a favorable trade-off between performance and computational efficiency.

4. Conclusion

This paper presents a dual-pathway deep learning architecture integrating CNN, BiLSTM, and cross-attention mechanisms for power quality disturbance classification. The proposed model constructs parallel time-domain and frequency-domain branches to extract complementary features, with a Cross-Attention fusion module enabling enhanced discriminability for complex composite disturbances.

Experimental validation on 24 PQDs classes demonstrates superior performance: 99.73% validation accuracy and 98.94% accuracy under 30dB noise conditions. Ablation studies confirm that the dual-pathway structure improves accuracy by 6.51 percentage points over single-branch variants, while cross-attention contributes an additional 2.08 percentage points. Feature visualization through t-SNE reveals progressive evolution from chaotic raw signal distributions to well-separated cluster structures, validating the discriminative enhancement mechanism.

The model achieves convergence within 43 epochs with inference latency of 251μs per sample, satisfying real-time monitoring requirements. This balanced integration of accuracy, noise robustness, and computational efficiency establishes the proposed architecture as a viable solution for intelligent power quality monitoring in smart grid applications.

Future research directions include: (1) extending the framework to identify emerging disturbance types from high-penetration renewable energy sources through open-set recognition and incremental learning approaches, and (2) investigating model compression techniques such as structured pruning, quantization, and knowledge distillation to enable deployment on resource-constrained edge devices while maintaining classification accuracy.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (Grant No. 2023YFC3107100).

References

[1] Aleem SHEA, Zobaa AF, Aziz MMA. Optimal C-type passive filter based on minimization of the voltage harmonic distortion for nonlinear loads. IEEE Trans Ind Electron. 2022;69(2):1583-93.

[2] Mishra S, Ray PK, Mallick RK. Power quality event classification under noisy conditions using EMD-based de-noising techniques. IEEE Trans Instrum Meas. 2023;72:1-11.

[3] Kumar C, Liserre M. A LSTM-based deep learning method for predicting grid voltage sag. IEEE Trans Ind Appl. 2023;59(1):954-63.

[4] Tan RHG, Ramachandaramurthy VK, Mansor M, Ekanayake JB. A comprehensive review on power quality disturbances and machine learning applications. Renew Sustain Energy Rev. 2024;183:113515.

[5] Shen L, Wang F, Cao L, Zhao J, Qiao W. Novel fusion-based deep learning network for multi-label power quality disturbances detection. Electr Power Syst Res. 2022;208:107881.

[6] Wang J, Xu Z, Che Y. Power quality disturbance classification via combining advanced S-transform and CNN with feature ranking. Measurement. 2023;203:111975.

[7] Shen L, Wang F, Cao L, Zhao J, Qiao W. Novel fusion-based deep learning network for multi-label power quality disturbances detection. Electr Power Syst Res. 2022;208:107881.

[8] Chen Y, Zhao D, Luo L, Wang Y. A light gradient boosting machine with Shapley additive explanation for power quality disturbances classification. Electr Power Syst Res. 2022;211:108163.

[9] Wang J, Xu Z, Che Y. Power quality disturbance classification via combining advanced S-transform and CNN with feature ranking. Measurement. 2023;203:111975.

[10] Yao Y, Wei H, Chen J, Zhou G, Liu Y. Power quality disturbance classification based on multiresolution convolutional neural networks. IEEE Trans Ind Electron. 2023;70(8):8421-30.

[11] Sha H, Mei F, Zhang C, Pan Y, Zheng J. Identification method for voltage sags based on K-means-singular value decomposition and least squares support vector machine. Energies. 2022;15(8):2831.

[12] Kumar C, Liserre M. A LSTM-based deep learning method for predicting grid voltage sag. IEEE Trans Ind Appl. 2023;59(1):954-63.

[13] Ahila R, Thangavel S, Jeyabharath R. A hybrid particle swarm optimization technique for power quality disturbance classification using wavelet transform and extreme learning machine. Electr Eng. 2022;104(5):3101-16.

[14] Tan RHG, Ramachandaramurthy VK, Mansor M, Ekanayake JB. A comprehensive review on power quality disturbances and machine learning applications. Renew Sustain Energy Rev. 2024;183:113515.

[15] Rodrigues Junior WL, Borges FAS, Rabelo RdAL, Fernandes RAS. A methodology for detection and classification of power quality disturbances using a real-time operating system in a computational embedded platform. Int J Electr Power Energy Syst. 2023;147:108860.

[16] Kiranyaz S, Avci O, Abdeljaber O, Ince T, Gabbouj M, Inman DJ. 1D convolutional neural networks and applications: A survey. Mechanical Systems and Signal Processing. 2021;151:107398.

[17] Jiang Y, Xie B, Wang J, Xia Y. Hybrid LSTM-based deep learning framework for fault diagnosis in power systems using PMU data. IEEE Trans Power Syst. 2023;38(5):4523-35.

[18] Li X, Wang W, Hu X, Yang J. Selective kernel networks with attention mechanism for power quality disturbance recognition. IEEE Trans Ind Inform. 2023;19(4):5aborr-54.

[19] Baig MAA, Ratyal NI, Amin A. A multi-modal deep learning framework for power quality disturbance classification: an integration of 1D time-series signals and 2D scalograms. Comput Electr Eng. 2025;128:110716.

[20] Luo Y, Wong Y, Kankanhalli M, et al. G-Softmax: improving intraclass compactness and interclass separability of features. IEEE Trans Neural Netw Learn Syst. 2020;31(2):685-699.