Electric Power Systems Research 252 (2026) 112352

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Visually inspired power quality disturbances recognition via Gramian Angular Difference Field, swin transformer and temporal–frequency–symmetry attention

Jiajian Lin [a], Jalal Tavalaei [a,*], Mehran Motamed Ektesabi [b], Hadi Nabipour Afrouzi [c]

[a] Faculty of Engineering, Computing and Science, Swinburne University of Technology, Sarawak Campus, Kuching 93350, Malaysia
[b] School of Science, Computing and Engineering Technologies, Swinburne University of Technology, Hawthorn, Victoria, Australia
[c] Faculty of Computing, Engineering and The Built Environment, Birmingham City University, Birmingham B5 5JU, United Kingdom

ARTICLE INFO

Keywords:
Power quality disturbance
Deep learning
Gram matrix
Swin transformer
Attention mechanism

ABSTRACT

Accurate and real-time identification of power quality disturbances (PQDs) remains a pressing challenge in modern power systems, especially with the increased penetration of renewable energy sources and the resulting complexity of electrical networks. This study proposed a novel hybrid framework for PQD recognition, integrating Gramian Angular Difference Field (GADF) image encoding, the Swin Transformer for hierarchical local feature extraction, and a Temporal-Frequency-Symmetry Enhanced Global Attention Mechanism (TFSGAM) for capturing global and domain-specific features. The one-dimensional PQD signals are first converted into two-dimensional images using GADF, effectively preserving temporal dependencies. The Swin Transformer exploits local contextual information, while TFSGAM further enhances feature representation by incorporating temporal position encoding, frequency-domain awareness, and symmetry-based spatial attention. Experimental results on synthetic and real-world datasets demonstrated that the proposed framework achieved classification accuracy exceeding 98 % under most noise conditions, while maintaining strong robustness across 25 PQD types and ensuring real-time applicability with an average inference time of 169 ms/sample. Comparative studies with state-of-the-art methods and extensive ablation analyses confirmed that this approach exhibits strong robustness in noise scenarios with SNR = 20/30/40 dB.

1. Introduction

With the rapid advancement of renewable energy technologies, wind and solar power generation have emerged as the primary focus in the ongoing transformation of modern power systems [1]. The increasing integration of renewable energy sources into the grid inevitably introduces a higher proportion of nonlinear and fluctuating loads, posing significant challenges to maintaining power quality [2]. In microgrid systems, the high penetration of inverter-based resources and frequent islanding transitions exacerbate power quality challenges. Disturbances such as harmonics, voltage fluctuations, and transients degrade power quality indices and trigger instability through complex inverter interactions and control loop coupling. Recent advances in coordinated predictive secondary control for DC microgrids have highlighted the need to mitigate such disturbances to maintain voltage regulation and system stability under high-order dynamic conditions [3]. Moreover, integrating inverter-based distributed resources and the frequent transitions between grid-connected and island modes in microgrids amplify PQDs, leading to instability risks through control loop interactions and communication vulnerabilities. Beyond conventional control strategies, emerging paradigms such as blockchain-enabled consensus mechanisms have been introduced to enhance the resilience of renewable energy systems. For example, the recently proposed Proof of Task protocol has demonstrated the ability to achieve secure real-time regulation in renewable energy power systems, ensuring trustworthy data exchange and system stability under cyber threats [4]. These findings underline that accurate PQD recognition is crucial to ensuring reliable operation and resilience of microgrids with high renewable penetration. Currently, approaches for PQD recognition can be broadly classified into traditional signal processing methods and deep learning-based techniques utilizing neural networks, each with distinct advantages and limitations.

Traditionally, PQD recognition has relied on feature engineering

* Corresponding author.
E-mail addresses: jlin@swinburne.edu.my (J. Lin), jtavalaei@swinburne.edu.my (J. Tavalaei).

https://doi.org/10.1016/j.epsr.2025.112352
Received 20 August 2025; Received in revised form 2 October 2025; Accepted 6 October 2025
Available online 15 October 2025

techniques grounded in digital signal processing. In such approaches, characteristic features are manually extracted from raw electrical signals using methods such as the Short-Time Fourier Transform (STFT), Wavelet Transform, or Hilbert-Huang Transform and are subsequently classified using conventional machine learning classifiers like support vector machines or decision trees [5–7]. However, these traditional approaches are often heavily dependent on expert signal processing and feature selection knowledge. The performance of such methods is sensitive to the choice of signal decomposition technique and may result in redundant or insufficient feature representations. Consequently, they are increasingly inadequate in the era of artificial intelligence and large language models, where automated and intelligent feature extraction is essential for achieving high recognition, accuracy and adaptability across complex scenarios.

In addition to traditional strategies, recent research has explored a variety of advanced material designs to enhance both environmental remediation and energy storage performance. For example, the controllable preparation of coal gangue-based SAPO-5 molecular sieves has provided a low-cost solution for removing heavy metal ions from wastewater, demonstrating the potential of porous frameworks in pollution control [8]. In energy storage, localized highly concentrated electrolytes with ionic liquid solvents were shown to form rigid and Li^{+}-conductive interphases, effectively suppressing dendrite growth and enabling stable cycling of lithium metal batteries across wide temperature ranges [9]. Similarly, oxygen-vacancy engineering in C-WO_3/BiOBr heterojunctions has significantly improved visible-light photocatalytic degradation of benzene by enhancing charge separation and reactive radical generation [10]. These studies highlighted the pivotal role of structural control and interfacial engineering in advancing high-performance functional materials. Inspired by these advances, this study focused on the intelligent classification of complex power quality disturbances by combining GADF-based signal encoding with Swin Transformer and a temporal-frequency-symmetry enhanced global attention mechanism, thereby extending the concept of optimization and integration into modern power systems.

Deep learning architecture can autonomously learn salient features from training data, capturing complex non-linear dependencies without manual feature engineering. In PQD recognition, the CNN-BiLSTM model first employed convolutional layers to extract local spatial–temporal features from raw signals, then used a bidirectional LSTM to model long-term dependencies and global temporal correlations [11]. The CNN-Transformer model similarly used convolutional layers to capture local spatial–temporal patterns, followed by a Transformer module with self-attention to model global dependencies and long-range relationships [12]. Although effective, directly employing raw sampled signals may introduce redundant information, hinder the discrimination of subtle disturbance patterns, and exacerbate overfitting risks, particularly in limited or noisy datasets, compromising model generalizability and robustness in practical scenarios. When models are applied to PQD signals directly, part of the extracted features may correspond to noise components inherent in the raw signals, which provide little value for classification. Moreover, the hierarchical feature extraction process can repeatedly capture similar local patterns across multiple layers, increasing model complexity without yielding additional informative content. In the case of combined disturbances, overlapping feature characteristics among constituent signals may further generate duplicated or correlated representations in the feature space. Such redundancy reduces the efficiency of feature utilization. Despite the stable performance observed under noiseless and high-SNR scenarios, it amplified the models' sensitivity to low signal-to-noise conditions, as evidenced by the notable accuracy degradation at 20 dB.

With advances in computer vision, converting one-dimensional PQD signals into two-dimensional representations has shown superior classification accuracy over traditional 1D methods. Techniques such as the adaptive superlet transform yielded high-resolution time–frequency images for CNN-based classification [13], while the Gramian Angular Field encoded nonlinear spatial–temporal features for grouped residual CNNs with attention, complemented by LSTM-based temporal modelling and feature fusion [14]. These approaches captured both spatial and temporal characteristics of complex disturbances. Still, single-model solutions may lack robustness, generalization, and multi-scale contextual awareness, underscoring the need for integrated or multi-modal strategies for reliable real-world PQD recognition.

To address the challenges, the principal contributions of this article are summarized as follows.

1) Proposed a Temporal-Frequency-Symmetry Enhanced Global Attention Mechanism specifically tailored for PQD image representations. The proposed attention module can capture multi-dimensional correlations and domain-specific structural features beyond conventional attention mechanisms by integrating temporal positional encoding, frequency-aware enhancement, and symmetry-aware convolution.
2) This study proposed a hybrid framework that transforms one-dimensional PQD into Gramian Angular Difference Field images, extracts local and global features using a Swin Transformer and a TFSGAM, and fuses these features for robust and discriminative PQD classification.
3) Rigorous experiments on synthetic and real-world datasets demonstrate that the proposed framework consistently achieves high classification accuracy, exceeding 98 % across 25 distinct power quality disturbance categories and a wide range of noise environments. Moreover, the model exhibits rapid inference speeds, averaging 169 milliseconds per sample, which satisfies real-time operational requirements.

2. Feature extraction

In this section, we described the key methodologies for signal-to-image conversion and advanced feature representation that drove the effectiveness of the proposed model.

2.1. Gramian angular difference field

The Gramian Angular Field (GAF) encodes one-dimensional time series into two-dimensional images by mapping disturbance signals from Cartesian to polar coordinates and constructing a Gramian matrix via trigonometric operations, thereby preserving temporal correlations and enhancing feature extraction for image-based classification tasks [15]. In this study, the GADF encoding approach was employed, defined mathematically as follows:

$$G_{gadf} = \begin{bmatrix} \sin(\phi_1 - \phi_1) & \cdots & \sin(\phi_1 - \phi_n) \\ \sin(\phi_2 - \phi_1) & \cdots & \sin(\phi_2 - \phi_n) \\ \vdots & \sin(\phi_i - \phi_j) & \vdots \\ \sin(\phi_n - \phi_1) & \cdots & \sin(\phi_n - \phi_n) \end{bmatrix} \tag{1}$$

Where, ϕ_i denotes the angular value corresponding to the i^{th} element

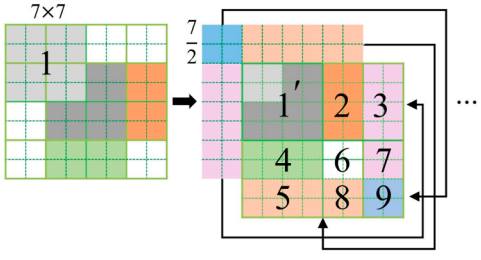

Fig. 1. Shifted windows mechanism in the Swin Transformer.

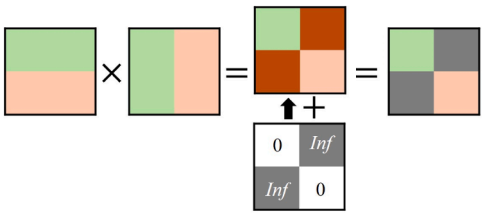

Fig. 2. Masking mechanism for discontinuous windows.

in the sequence. Each element $\sin(\phi_i - \phi_j)$ represents the sine of the angular difference between a normalized time series' i^{th} and j^{th} samples. The diagonal entries are always zero since $\phi_i - \phi_i = 0$, while the off-diagonal entries encode the relative phase relationships between different time points. This construction yielded a skew-symmetric matrix that captured pairwise temporal dependencies in a two-dimensional form, where positive or negative values reflected the direction of angular differences. This encoding method preserved the temporal order of the one-dimensional sequence within a two-dimensional image, with the main diagonal retaining the original signal and off-diagonal elements representing inter-time-step relationships. Each matrix value, computed through convolution-like operations, reflected the intensity of a specific feature. For time-dependent PQD signals, the GAF thus integrated the time vector and signal into a unified feature representation.

2.2. Swin transformer

The Swin Transformer, an enhanced Vision Transformer (ViT), improves fine-grained spatial understanding in image recognition by adopting a hierarchical feature representation akin to VGG [16]. Unlike ViT, which processes fixed-size patches globally, the Swin Transformer partitions feature maps into 7 × 7 non-overlapping windows for local self-attention, reducing computational complexity but initially limiting cross-window interaction, as shown in Fig. 1. To overcome this, a Shifted Windows mechanism shifts windows by half their size after each attention step, enabling overlap and information exchange between previously independent regions, thereby mitigating information loss and enhancing model accuracy.

However, these newly formed windows are composed of patches not naturally adjacent to the original feature map. For example, windows 4 and 5 contain discontinuous patches, which makes direct attention computation invalid. To resolve this, a masking mechanism is introduced [17]. An attention mask is applied to the attention matrix, where interactions between non-contiguous patches are assigned a negative value $Inf = -100$, which prevents attention from being computed across unrelated patches, as depicted in Fig. 2.

2.3. Temporal-frequency-symmetry enhanced global attention mechanism

The Global Attention Mechanism (GAM) jointly models channel–spatial attention and their cross-dimensional interactions, enhancing image recognition performance [18]. Its channel attention submodule applies three-dimensional permutation to preserve spatial–channel dependencies, followed by a two-layer MLP to capture higher-order interactions. In contrast, the spatial attention submodule employs two convolutional layers for effective spatial feature aggregation. For PQD analysis, GAM is further refined through structural modifications, as shown in Fig. 3.

The GADF encodes time series into two-dimensional representations whose pixel values reflect the sequence's temporal order. To preserve this structure, a Temporal Positional Encoding (TPE) layer is introduced before the GAM [19], embedding explicit time-dependent information to enhance long-range dependency modelling and sensitivity to sequence order. Additionally, to address the limitation of conventional channel attention, which focuses solely on spatial statistics and neglects informative frequency characteristics, frequency-aware enhancement is incorporated into the channel attention submodule, as illustrated in Fig. 4. Specifically, a two-dimensional Discrete Cosine Transform (2D-DCT) is applied to each channel F_c^{freq}, allowing the model to extract and emphasize frequency domain characteristics critical to identifying PQD patterns. The transformation is defined as follows:

$$F_c^{freq}(u,v) = \sum_{x=0}^{H-1}\sum_{y=0}^{W-1} F(x,y)\cos\left[\frac{\pi(2x+1)u}{2H}\right]\cos\left[\frac{\pi(2y+1)v}{2W}\right], \quad \forall c \in [1,C] \tag{2}$$

$$F_{concat}^{freq} = concat\left(F_1^{freq},\ldots,F_C^{freq}\right) \in R^{C\times H\times W} \tag{3}$$

$$z_{freq} = \frac{1}{H\cdot W}\sum_{i=1}^{H}\sum_{j=1}^{W} F_{concat}^{freq}(i,j) \tag{4}$$

$$M_c = \sigma\left(W_2\cdot\delta\left(W_1\cdot z_{freq}\right)\right) \in R^{C\times 1\times 1},\ W_1 \in R^{\frac{C}{r}\times C}, W_2 \in R^{C\times\frac{C}{r}} \tag{5}$$

Where $F(x,y)$ denotes the input pixel at position (x,y), u and v respectively represent the horizontal frequency index and vertical frequency index in the output frequency domain, r is the compression ratio, δ is the ReLU activation function, σ is the Sigmoid function, $W_1$and W_2 are two layers of weights used for dimensionality reduction and enhancement and H, W represent the height and width of the feature map, respectively.

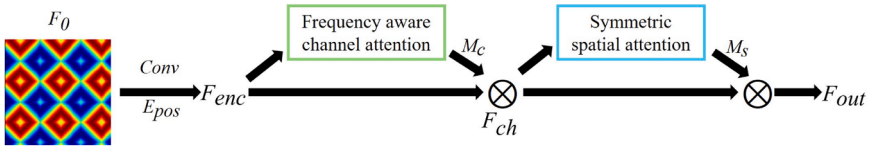

Fig. 3. Overview of the proposed Temporal–Frequency–Symmetry Enhanced Global Attention Mechanism.

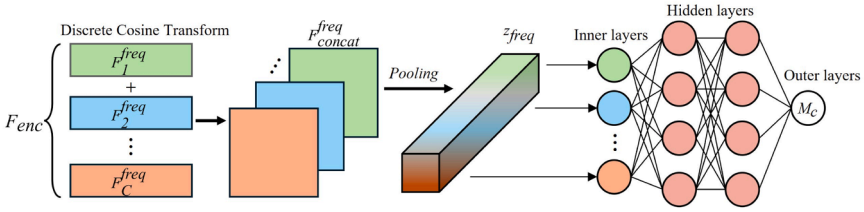

Fig. 4. Frequency-aware channel attention submodule.

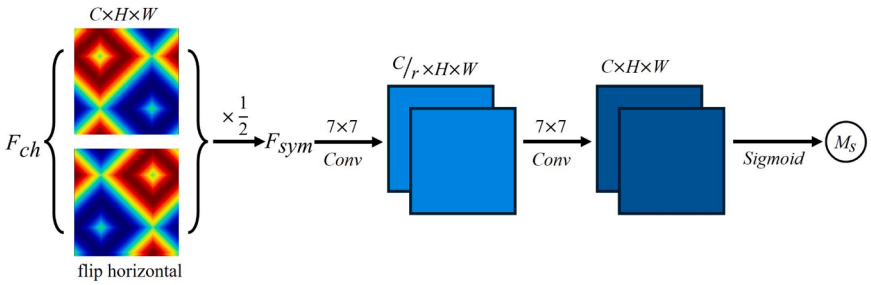

Fig. 5. Symmetry-aware spatial attention submodule.

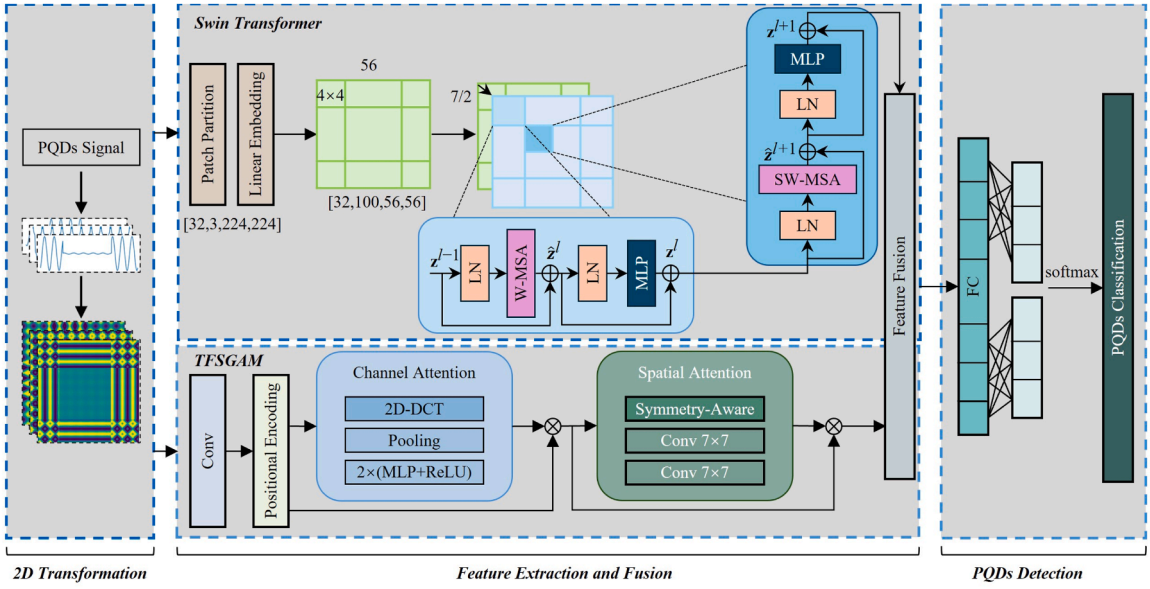

Fig. 6. Framework of the proposed GADF-Swin T-TFSGAM hybrid model.

GADF encodes the angular information of time series data into a two-dimensional image, exhibiting prominent diagonal symmetry and periodic patterns. These characteristics are distinct from the texture structures commonly found in natural images. As a result, conventional spatial attention mechanisms, which are primarily designed for natural image textures, may not capture the inherent structural properties of GADF representations. To address this issue, a symmetry aware enhancement mechanism is incorporated into the spatial attention submodule. This design facilitates the recognition of repetitive and mirrored patterns by explicitly modeling the bilateral dependencies along the horizontal axis, as illustrated in Fig. 5. The corresponding formulation of the mirror-aware convolution operation is given as follows:

$$F_{sym} = \frac{1}{2}(F_{ch} + flip(F_{ch})) \tag{6}$$

$$M_s = \sigma(Conv_{7\times7}(F_{sym})) \in R^{C\times H\times W} \tag{7}$$

Where *flip* represents horizontal flipping of the $H \times W$ dimension.

3. Methodology

A novel PQD classification framework was proposed, combining GADF-based signal-to-image transformation, the Swin Transformer for local contextual feature extraction, and TFSGAM for global spatial representation across temporal, frequency, and symmetrical dimensions. Features from both branches are fused and processed via adaptive average pooling, as illustrated in Fig. 6.

3.1. Power quality disturbances and datasets

A synthetic PQD dataset was generated in Python following IEEE Std. 1159 [20], comprising 10 single-disturbance and 15 multi-disturbance types as listed in Table 1, with 1000 samples each for a total of 25, 000 waveforms. Each sample contains 10 cycles at 50 Hz, sampled at 5120 Hz with 1024 data points over 0.2 s, and the dataset is split into training 70 %, validation 20 %, and testing 10 %. To improve robustness, noise at 20, 30, and 40 dB SNRs was added during waveform generation.

3.2. Conversion of GADF for power quality disturbances

To facilitate the image-based classification of PQDs, one-dimensional time-series signals were transformed into two-dimensional images using GADF. Before the transformation, each series was normalized to the range $[-1, 1]$ and encoded into angular values using the transformation $\varphi_i = arccos(x_t)$. Subsequently, the GADF matrix was visualized as a two-dimensional colour image using the *jet* colour map, effectively highlighting subtle differences in temporal dynamics. The resolution of the GADF matrix was set to 500, and the resulting image was resized to 224 × 224 pixels using the DPI scaling factor provided by the *matplotlib* library, which was also done by other researchers [21]. The workflow is illustrated in Fig. 7.

3.3. Feature extraction of visual models

The Swin Transformer, employed as the visual mechanism to extract

Table 1
Classification accuracy of 25 PQDs under noise levels of noiseless, 20, 30, and 40 dB.

Types	PQDs name	Noiseless (%)	20 dB (%)	30 dB (%)	40 dB (%)
C1	Normal	100.00	97.00	100.00	100.00
C2	Swell	100.00	100.00	100.00	100.00
C3	Sag	99.00	100.00	97.00	100.00
C4	Harmonics	100.00	97.00	100.00	100.00
C5	Flicker	99.00	100.00	99.00	100.00
C6	Interruption	100.00	96.00	100.00	99.00
C7	Transient impulsive	100.00	95.00	100.00	99.00
C8	Transient oscillatory	100.00	100.00	100.00	100.00
C9	Notching	100.00	98.99	100.00	98.98
C10	Spike	96.00	96.00	97.00	99.00
C11	Harmonics + Swell	100.00	94.00	100.00	100.00
C12	Harmonics + Sag	97.00	98.99	97.00	100.00
C13	Harmonics + Interruption	100.00	100.00	100.00	98.99
C14	Harmonics + Flicker	100.00	100.00	100.00	100.00
C15	Harmonics + Transient impulsive	99.00	98.00	99.00	100.00
C16	Harmonics + Transient oscillatory	100.00	96.00	100.00	100.00
C17	Flicker + Swell	97.93	78.00	98.93	99.00
C18	Flicker + Sag	98.00	93.00	98.00	98.00
C19	Flicker + Transient oscillatory	96.00	98.00	96.00	99.00
C20	Flicker + Transient impulsive	99.00	99.00	99.00	100.00
C21	Transient oscillatory + Swell	97.00	97.00	97.00	98.00
C22	Transient oscillatory + Sag	98.00	96.00	98.00	99.00
C23	Harmonics + Transients oscillatory + Swell	99.00	88.00	95.00	98.00
C24	Harmonics + Transients oscillatory + Sag	97.98	95.96	97.98	98.96
C25	Harmonics + Transients oscillatory + Flicker	92.93	94.00	92.93	99.00
	Overall average accuracy is 98.13 %	**98.57**	**96.20**	**98.41**	**99.33**

local spatial features from GADF images, partitions each [3224,224] RGB input into 4 × 4 non-overlapping patches, yielding 56×56 tokens projected into a 100-dimensional embedding space. Configured with a single encoder layer and four self-attention heads, it computes local attention within non-overlapping 7 × 7 windows using a shifted-window mechanism, enabling efficient hierarchical feature representation.

For each window, multi-head self-attention is performed independently. Given a sequence of embedded tokens $X \in R^{N\times d}$, where $N = 49$ and $d = 100$, the attention is computed as:

$$Attention(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}} + B\right)V \tag{8}$$

Where $Q = XW^Q$, $K = XW^K$, $V = XW^V$ are query, key, and value projections, d_k is the dimension of each attention head, $B = \frac{7}{2}$ is the relative positional bias specific to each window.

To allow cross-window communication, the shifted window mechanism shifts the feature map by an offset of $B = \frac{7}{2}$. This enables attention to span across previously non-overlapping windows while preserving computational efficiency. After self-attention, each token is passed through a MLP with a hidden layer dimension of $4d$. The MLP operation is defined as:

$$MLP(x) = \sigma(xW_1 + b_1)W_2 + b_2 \tag{9}$$

Where $W_1 \in R^{d\times 4d}$ and $W_2 \in R^{4d\times d}$ are weight matrix, σ is the GELU activation function, b_1 and b_2 are the offset constants of the first layer and the second layer. The resulting spatial features retain a resolution of 56 × 56, and the output is formatted as $[32, 56, 56, 100]$, where 32 represents the batch size.

3.4. Feature extraction of deep learning models

A convolutional neural network (CNN) augmented with the TFSGAM module was employed to extract complementary global spatial features from PQD images. The architecture consists of two 3 × 3 convolutional blocks (padding = 1) with ReLU and 2 × 2 max-pooling, increasing feature channels from 3 to 50 and 100, yielding a [32,100,56,56] feature map aligned with the Swin Transformer output for fusion. A Temporal Positional Encoding layer preceded channel attention to embed time-dependent positional information. In the channel attention

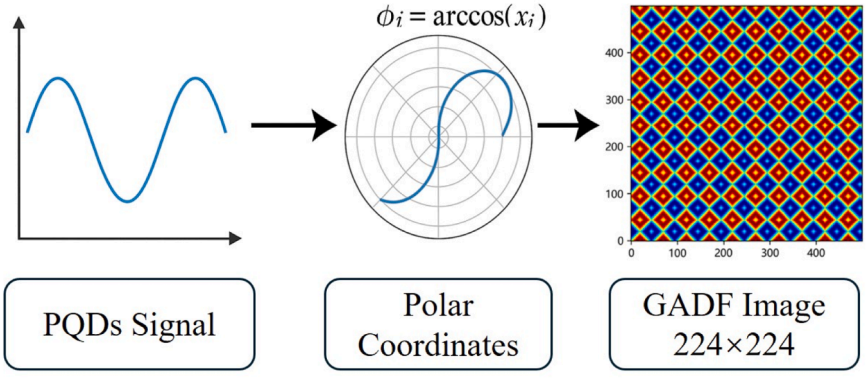

Fig. 7. Workflow of converting PQD signals into GADF images.

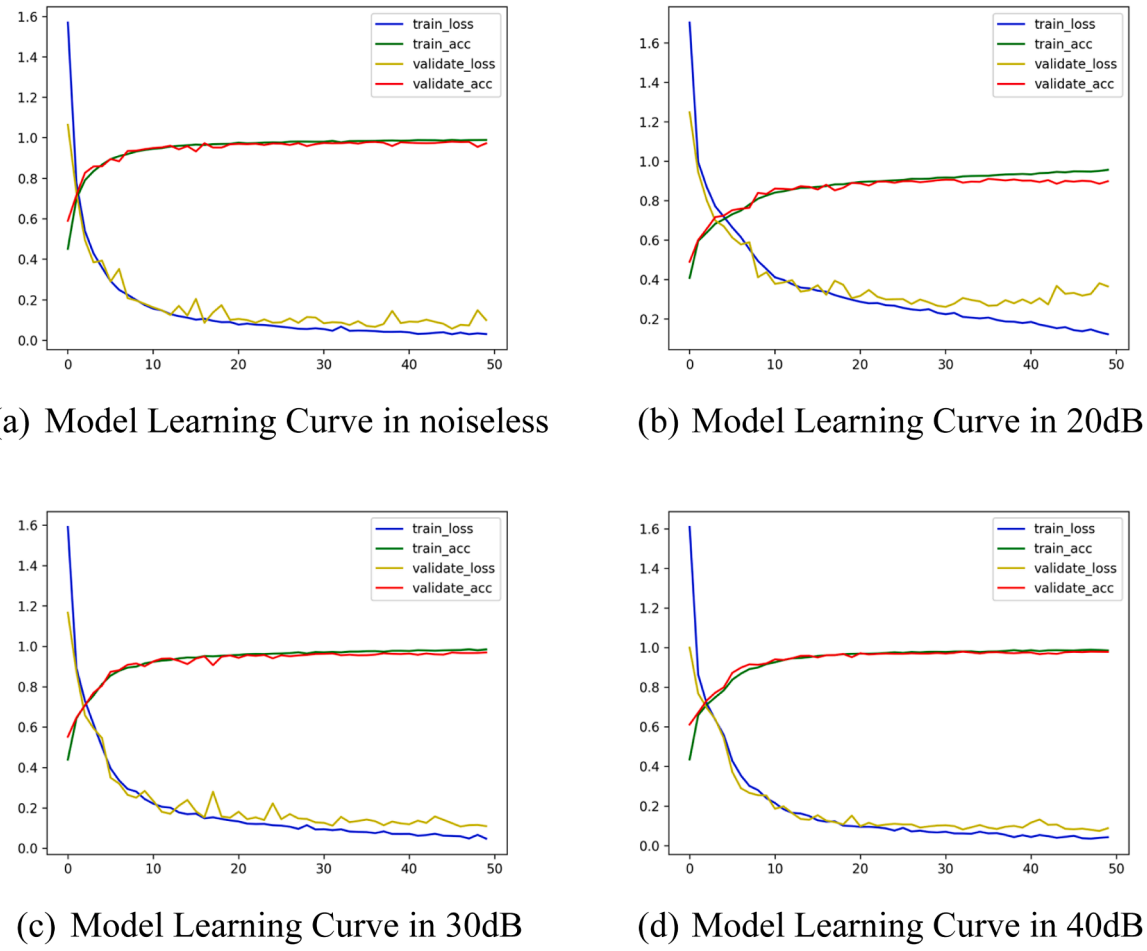

Fig. 8. Training and validation curves under noise levels of noiseless, 20, 30, and 40 dB.

submodule, a 2D-DCT was applied per channel, and the resulting frequency coefficients passed through a two-layer MLP (reduction ratio $r = 4$) to capture higher-order, frequency-sensitive dependencies. In the spatial attention submodule, the channel-refined features were processed by two 7×7 convolutions with batch normalisation, ReLU, and Sigmoid activation to produce an attention mask, enhancing spatially significant regions.

3.5. Model training process

After the model was constructed, we used an AMD Ryzen 9 5900HS CPU running at 3.30 GHz, 16 GB of RAM, and a 12 GB NVIDIA GeForce RTX3060 GPU to train the model. Then, the fusion feature from the two branches was trained with a batch size set to 32, a learning rate initialized at 0.001, and the Adam optimizer employed for gradient updates. The training performance of the proposed model under different noise conditions, specifically at noiseless, 20 dB, 30 dB, and 40 dB SNRs, is illustrated in Fig. 8.

After 50 training epochs, the proposed framework achieved validation losses of 0.1067, 0.3345, 0.1350, and 0.0725 with corresponding accuracies of 0.9743, 0.9046, 0.9623, and 0.9835 under noise levels of noiseless, 20 dB, 30 dB, and 40 dB, respectively, demonstrating its robustness across varying noise environments. Performance degradation at 20 dB was accompanied by a marked divergence between training and validation losses after the 19th epoch, indicating overfitting, wherein the model captured training-specific patterns more precisely than it generalized to unseen data. Despite this, validation accuracy continued to improve, suggesting preserved classification capability. Regularization techniques were introduced to enhance generalization and mitigate overfitting under moderate noise.

4. Results and analysis

This chapter comprehensively evaluates the proposed framework, including experimental results, comparative analyses, and ablation studies, to demonstrate its effectiveness, robustness, and superiority over existing methods.

4.1. Robustness to noise

Table 1 presents the classification accuracies of the proposed framework across 25 distinct types of PQDs under four noise conditions: noiseless, 20 dB, 30 dB, and 40 dB. The results demonstrate the model's robustness and high discriminative capability across single and composite disturbance categories. At a noise-free condition, the average classification accuracy reaches 98.13 %, with 11 out of 25 disturbance types achieving 100 % accuracy. This confirms the model's strong capacity to distinguish various PQD patterns in clean signal environments.

Under the 20 dB noise condition, the average classification accuracy decreased to 96.20 %, which is the lowest performance among all tested noise levels, as shown in Fig. 9. This decline indicates moderate noise may interfere more significantly with feature extraction and generalization than low and high noise levels. Furthermore, the two composite PQD types with the lowest classification accuracies under this condition are C17 and C23, with 78.00 % and 88.00 %, respectively. Notably, both of these disturbance types involve the swell component. This

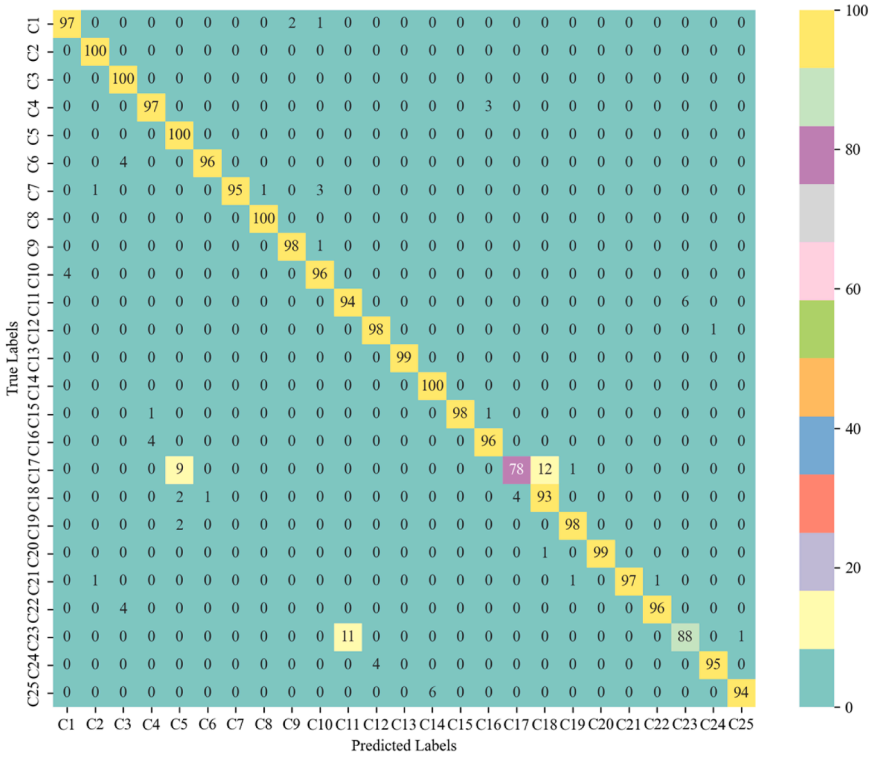

Fig. 9. Confusion matrix of classification results under 20 dB noise.

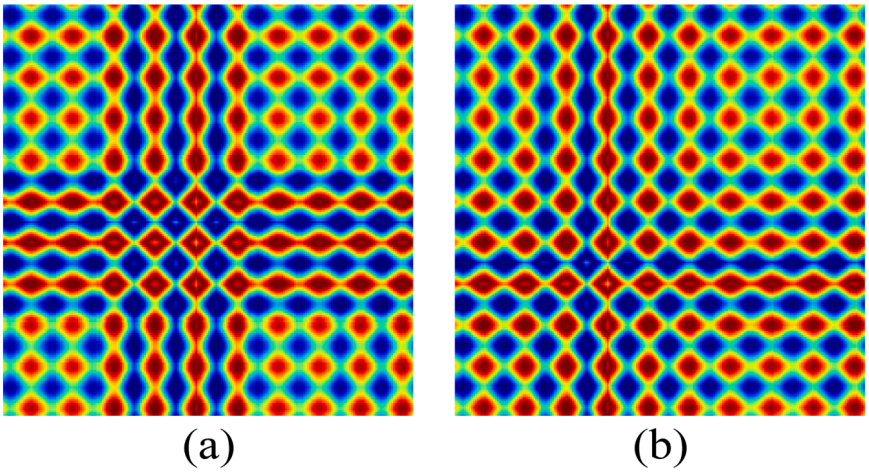

Fig. 10. GADF-transformed two-dimensional feature maps under 20 dB noise: (a) C17 (Flicker + Swell) and (b) C18 (Flicker + Sag).

observation suggests that voltage swell, particularly when combined with other dynamic disturbances, may reduce feature separability under moderate noise. The relatively smooth and gradual swell might be more easily masked by background noise at this level, thereby hindering the model's ability to effectively capture distinguishing characteristics. This highlights the need for more refined feature enhancement or denoising strategies when dealing with compound PQDs involving swell components under intermediate noise conditions.

A Temporal Positional Encoding layer preceded channel attention to embed time-dependent positional information. In the channel attention submodule, a 2D-DCT was applied per channel, and the resulting frequency coefficients passed through a two-layer MLP (reduction ratio $r = 4$) to capture higher-order, frequency-sensitive dependencies. In the spatial attention submodule, the channel-refined features were processed by two 7×7 convolutions with batch normalization, ReLU, and Sigmoid activation to produce an attention mask, enhanced spatially significant regions.

The noticeable performance degradation at 20 dB can be attributed to combined feature masking and generalization difficulty. As shown in Fig. 8(b), the divergence between training and validation losses after the 19th epoch indicates a tendency toward overfitting under moderate noise conditions. Moreover, Table 1 reveals that disturbance types involving voltage swell (C17 and C23) suffered the largest accuracy decline, dropping to 78 % and 88 %, respectively. This can be explained by the relatively smooth and gradual characteristics of swell, which are more easily masked by moderate background noise, especially when compounded with other disturbances such as flicker or transients. In contrast, at noiseless the model learns strong anti-noise patterns, while

Table 2
Comparison with existing methods.

Method	PQD No.s	Performance			
		Accuracy (%)	Recall (%)	Precision (%)	F1-score (%)
Proposed model	25	98.13	99.14	98.26	98.33
URPM-CWT+MCFFN [22]	28	97.65	97.65	97.72	98.00
SWT and AlexNet [23]	15	96.05	97.06	96.31	–
Improved ResNet and attention [24]	20	–	96.37	96.40	96.38
EfficientNet [25]	8	84.72	84.33	90.00	86.00

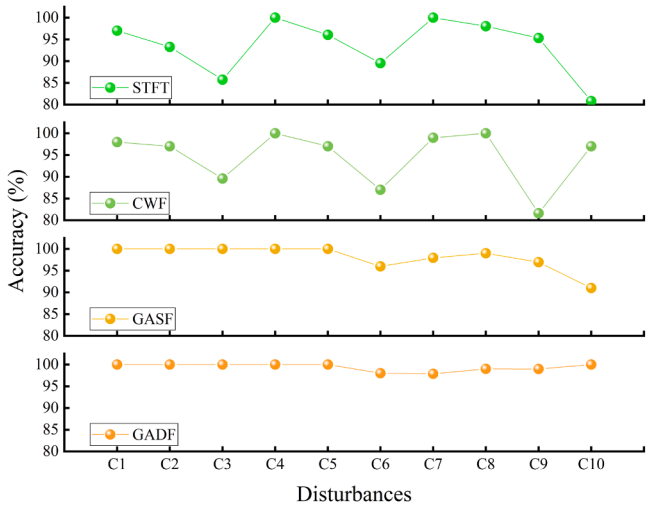

Fig. 11. Comparison of signal-to-image transformations (GADF, GASF, CWT, STFT) for 10 single PQDs under noiseless.

at 30–40 dB the original disturbance features remain clearly distinguishable. The intermediate 20 dB condition thus represents the most challenging regime.

The performance degradation at 20 dB SNR can be more clearly understood by examining specific misclassification cases. As shown in Table 1, disturbance type C17 (Flicker + Swell) achieved the lowest recognition accuracy, dropping to 78 %. Within the validation set, 12 samples of C17 were incorrectly classified as C18 (Flicker + Sag). Fig. 10 presents the GADF-transformed two-dimensional feature maps of C17 and C18, respectively. Both images exhibit strong diagonal and symmetric patterns, with periodic oscillations distributed across the feature space. The subtle differences lie mainly in the oscillatory bands' localized wave intensity and orientation. However, under 20 dB noise, these differences are significantly blurred, resulting in highly similar spatial structures.

This observation supports our earlier explanation that moderate noise levels, unlike extreme or weak noise, produce a particularly challenging condition in which smooth or gradually varying features are masked without completely dominating the signal. The confusion between C17 and C18 demonstrates how overlapping temporal–frequency signatures reduce separability in the learned feature space.

Table 4
Ablation study of different modules under varying noise levels.

Method	Accuracy (%)				Parameters (M)	Latency (ms/sample)
	Noiseless	20 dB	30 dB	40 dB		
GADF-Swin T-TFSGAM	98.57	96.20	98.41	99.33	0.45	167
GADF-Swin T	95.54	93.92	95.48	96.35	0.09	158
GADF-TFSGAM	95.67	92.84	94.38	95.42	0.36	160
GADF-GAM	91.47	89.63	90.92	92.06	0.35	155

4.2. Comparison with existing methods

Table 2 compared the proposed model with other methods regarding accuracy, recall, precision, F1-score, and the number of PQD categories. The proposed model achieves the best overall performance, with 98.13 % accuracy, 99.14 % recall, 98.26 % precision, and a 98.33 % F1-score. Notably, this performance is achieved on a dataset encompassing 25 distinct types of PQDs, reflecting a broader and more complex classification task than other methods.

URPM-CWT+MCFFN achieved 97.65 % accuracy and recall on 28 PQD types but remained slightly inferior to the proposed model. SWT+AlexNet attained 96.05 % accuracy on 15 types, with lower recall and precision and no reported F1-score. Improved ResNet with attention yielded 96.40 % precision and 96.38 % F1-score on 20 types but lacked accuracy data, limiting comparability. EfficientNet performed worst, with 84.72 % accuracy and an F1-score of 86.00 on eight types, underscoring its limited suitability for complex multi-class PQD tasks.

4.3. Ablation experiment

Fig. 11 compares four signal feature extraction methods, namely GADF, Gramian Angular Summation Field (GASF), Continuous Wavelet Transform (CWF), and STFT, for classifying ten single PQDs under noiseless with identical model settings. GADF achieved the best performance with an average accuracy of 99.38 %, reaching 100 % in six categories and only minor declines in C6 to 98.00 % and C7 to 97.88 %. GASF ranked second at 98.09 % but dropped to 91.00 % in C10. In comparison, CWF and STFT recorded lower average accuracies of 94.61 % and 93.57 %, with significant decreases in C9 to 81.62 % and C10 to 80.80 %, indicating limitations in capturing certain temporally localized features.

As evidenced by the results in Table 3, visual transformation methods such as GADF and GASF consistently outperform traditional time–frequency analysis techniques like CWT and STFT in single PQD classification under noiseless noise. Specifically, GADF achieved an average accuracy of 99.38 %, whereas CWF and STFT were limited to 94.61 % and 93.57 %, respectively, with pronounced performance degradation in categories such as C9 and C10. This advantage arose because visual transformations embedded temporal dependencies and phase relationships into structured two-dimensional images, producing

Table 3
Classification accuracy of single PQDs under noiseless using different signal feature extraction methods.

Clean (%)	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	Avg
GADF	100.00	100.00	100.00	100.00	100.00	98.00	97.88	99.00	98.99	100.00	**99.38**
GASF	100.00	100.00	100.00	100.00	100.00	96.00	97.96	99.00	96.97	91.00	**98.09**
CWF	97.98	97.00	89.59	100.00	97.00	87.00	98.98	100.00	81.62	97.00	**94.61**
STFT	97.00	93.27	85.73	100.00	96.04	89.53	100.00	98.04	95.31	80.80	**93.57**

Table 5
Paired T-test results comparing GADF-Swin T-TFSGAM with ablated models.

Methods	Average difference	T value	P value	Standard error	95 % confidence interval of difference
GADF-Swin T	2.8050 %	15.921	5.39×10^{-4}	0.176	[2.245, 3.365]
GADF-TFSGAM	3.5500 %	13.592	8.61×10^{-4}	0.261	[2.719, 4.381]
GADF- GAM	7.1075 %	36.236	4.62×10^{-5}	0.196	[6.484, 7.731]

distinctive diagonal and symmetric patterns that aligned with the intrinsic characteristics of PQDs. Such structured representations facilitated more effective feature extraction by advanced computer vision models, enabling them to capture multi-scale spatial correlations and subtle variations even under noisy conditions. In contrast, traditional decomposition approaches suffered from resolution trade-offs and noise sensitivity, often leading to redundant or insufficient feature representations. These findings provided a conceptual explanation for the superiority of visual transformations, highlighting their potential as a robust foundation for intelligent PQD recognition.

For the ablation experiments, the Swin Transformer branch was configured with a single encoder layer and four self-attention heads, which was consistent with the base model. The TFSGAM branch was implemented with a two-layer MLP in the channel attention submodule, using a reduction ratio $r = 4$. These fixed configurations ensure that differences among GADF-Swin T, GADF-TFSGAM, and the combined model are attributable to the presence or absence of the respective modules, rather than changes in hyperparameters.

Table 4 presents an ablation study of the Swin Transformer, TFSGAM, and the baseline GAM under noiseless, 20, 30, and 40 dB. The complete model achieved the highest accuracies of 98.57, 96.20, 98.41, and 99.33 %, confirming strong noise robustness. Removing TFSGAM reduced accuracy, showing that while the Swin Transformer captures local features, global attention is essential for modelling long-range and frequency-sensitive dependencies. Using only TFSGAM slightly lowered accuracy compared to the full model but outperformed the Swin-only variant in low-noise conditions, highlighting its role in enhancing global feature representation. The baseline with GADF and the original GAM performed worst, with accuracies between 91.47 % and 92.06 %, indicating the limited capacity of the standard attention mechanism without the proposed enhancements.

To further assess the efficiency–accuracy trade-off, Table 4 reports the classification accuracy under different noise levels, the number of trainable parameters, and the average inference latency of each variant. The results show that the complete GADF-Swin T-TFSGAM model requires 0.45 M parameters and achieves a latency of 167 ms/sample, which is slightly higher than the lighter variants such as GADF-Swin T (0.09 M, 158 ms/sample) and GADF-TFSGAM (0.36 M, 160 ms/sample). However, this modest increase in complexity is accompanied by significant accuracy gains across all noise levels, particularly under challenging 20 dB conditions. These findings confirm that the proposed architecture balances accuracy and efficiency, making it suitable for real-time PQD recognition tasks.

4.4. Discussion

A paired T-test analysis was conducted to further quantify the trade-off between accuracy improvements and architectural modifications by comparing the full GADF-Swin T-TFSGAM model with its ablated counterparts across four noise levels. Each comparison was based on four paired observations, giving a degree of freedom of 3. The statistical results are summarized in Table 5. Compared with GADF-Swin T, the proposed model achieved an average improvement of 2.8050 %, with a T value of 15.921, a P value of 5.39×10^{-4}, a standard error of 0.176, and a 95 % confidence interval of [2.245, 3.365], clearly demonstrating a highly significant difference. Against GADF-TFSGAM, the improvement was 3.5500 % ($T = 13.592$, $P = 8.61\times10^{-4}$, standard error = 0.261, 95 % CI = [2.719, 4.381]), again confirming the superiority of the integrated framework. The largest margin was observed against GADF-GAM, where the improvement reached 7.1075 % ($T = 36.236$, $P = 4.62\times10^{-5}$, standard error = 0.196, 95 % CI = [6.484, 7.731]). The confidence intervals excluded zero in all cases, providing strong statistical evidence that the observed gains are not incidental.

These results indicate that while Swin Transformer and TFSGAM contribute independently to improved classification, their integration yields the most substantial benefits. Notably, the comparison between GADF-TFSGAM and GADF-GAM shows a 3.5575 % improvement, highlighting that TFSGAM is the most critical module in enhancing the recognition of power quality disturbances. This is attributable to TFSGAM's explicit modeling of temporal continuity, frequency-domain sensitivity, and symmetrical characteristics inherent in PQD signals, which are not fully exploited by conventional attention mechanisms. The superior performance of the full model thus arises from the complementary strengths of Swin Transformer's local feature extraction and TFSGAM's domain-specific global representation, resulting in robust and interpretable classification across diverse noise conditions.

The effectiveness of TFSGAM can be explained by aligning its design with the physical characteristics of PQDs. Let the input feature map be denoted as $F_0 \in R^{C\times H\times W}$, obtained from the GADF encoding of voltage waveforms. The GADF representation preserves the original signal's phase relations and periodic structures through its skew-symmetric property, which is essential for distinguishing events such as harmonics, flicker, and transients. To ensure that disturbance duration and onset information are retained, a temporal positional encoding is added:

$$\begin{cases} F_T = F_0 + P_t \\ P_t(i,2k) = \sin\left(\dfrac{i}{\tau^{2k/d}}\right) \\ P_t(i,2k+1) = \cos\left(\dfrac{i}{\tau^{2k/d}}\right) \end{cases} \tag{10}$$

where i is the temporal index, d is the feature dimension, and τ is the wavelength. This ensures that the model captures temporal continuity, vital for differentiating PQDs with similar shapes but different onsets or durations. PQDs are also characterized by distinctive frequency components. TFSGAM employs a two-dimensional discrete cosine transform:

$$C(u,v) = \alpha(u)\alpha(v)\sum_{i=0}^{H-1}\sum_{j=0}^{W-1} F_T(i,j)\cos\left(\frac{\pi(2i+1)u}{2H}\right)\cos\left(\frac{\pi(2j+1)v}{2W}\right) \tag{11}$$

where $(0) = 1/\sqrt{H}, \alpha(u>0) = \sqrt{2/H}$. The DCT coefficients represent the spectral energy distribution, which is then aggregated into low, mid, and high frequency bands. The band energies are passed through a multilayer perception to generate frequency weights w_c. Physically, these weights correspond to emphasizing flicker envelopes and voltage swells (low frequency), oscillatory transients (mid frequency), and inverter-induced harmonics (high frequency). GADF-encoded PQD images inherently exhibit diagonal or mirror-like symmetry due to the periodicity of electrical waveforms. To exploit this property, a mirror operator is defined as:

$$M(F_T)(i,j) = F_T(H-1-i, W-1-j) \tag{12}$$

A composite input is constructed as $Z = [F_T, M(F_T), F_T - M(F_T)]$. Then apply 7×7 convolution and sigmoid function to obtain spatial mask A, highlighting the regions where symmetry has been disrupted. These

Table 6
Real-world validation on the IEEE PES dataset.

Types	Identification accuracy (%)	Average accuracy (%)
C4	96.00	
C11	94.83	
C12	94.36	
C13	100.00	96.74
C16	92.00	
C23	100.00	
C24	100.00	

regions physically correspond to non-ideal behaviors, such as residual DC offset, phase imbalance, or composite PQD. The overall TFSGAM output can be expressed as:

$$F_{\text{out}}(i,j,c) = A(i,j)\cdot(w_c\cdot(F_0(i,j,c) + P_t(i))) \tag{13}$$

Where $F_0(i,j,c)$ represents the raw GADF-derived features that capture phase relations, $P_t(i)$ preserves temporal order and disturbance duration, w_c emphasizes frequency bands associated with PQD spectral signatures, and $A(i,j)$ highlights symmetry-breaking regions related to waveform distortion or composite disturbances.

This formulation indicates that TFSGAM is underpinned by empirical effectiveness and theoretical consistency. The incorporation of temporal encoding, frequency weighting, and symmetry masking is directly aligned with the intrinsic physical characteristics of PQDs, thereby enhancing the model's robustness and interpretability in disturbance classification.

5. Experiment validation

This chapter comprehensively validated the proposed framework through real-world measurement data and simulated scenarios, demonstrating its effectiveness, generalization capability, and practical feasibility for intelligent power quality disturbance recognition.

5.1. Real-world data validation

To validate the proposed approach on real-world data, the IEEE Power and Energy Society (PES) power quality disturbance dataset, comprising seven disturbance types each represented by 1536 data points sampled at 256 points per cycle [26], was employed. Following data cleaning and normalization using the dplyr and caret packages in R, the model achieved an overall identification accuracy of 96.74 %, shown in Table 6, confirming its reliability and practical applicability for real-world power quality monitoring.

5.2. Simulation data validation

To further verify the generalization performance of the proposed GADF-Swin T-TFSGAM under experimental signals, this study used MATLAB/Simulink to simulate and model power quality disturbances [27].

Fig. 12 depicts five simulation models specifically developed to represent distinct power quality disturbances (PQDs). Model 1 is a line fault simulation encompassing various fault scenarios such as single line-to-ground, double line-to-ground, line-to-line, three-phase, and multi-stage faults. Model 2 emulates capacitor bank energization, capturing the voltage oscillatory transients associated with capacitor switching events typically employed for power factor correction. Model 3 represents a three-phase nonlinear load scenario in which a six-pulse three-phase rectifier was used to simulate voltage sags and the introduction of harmonic components. Model 4, the electric arc furnace model, is designed to reproduce the flicker disturbances commonly induced by arc furnace operation. Model 5, known as the lightning impulse model, simulates impulsive transient events from lightning strikes near transmission lines.

The experimental evaluation was conducted in a controlled environment to ensure reproducibility and comparability of the reported inference latency. As summarized in Table 7, the hardware platform consisted of a Windows 10 (64-bit) operating system equipped with an

Table 7
Hardware and software environment used in model training and inference.

Hard/Soft ware	Version or setting value
OS	Win10 64bit
CPU	AMD Ryzen 9 5900HS
GPU	RTX 3060
RAM	DDR4 16GB×1
Python	3.9.7
Torch	2.3.0
Batch size	32
Learning rate	0.001

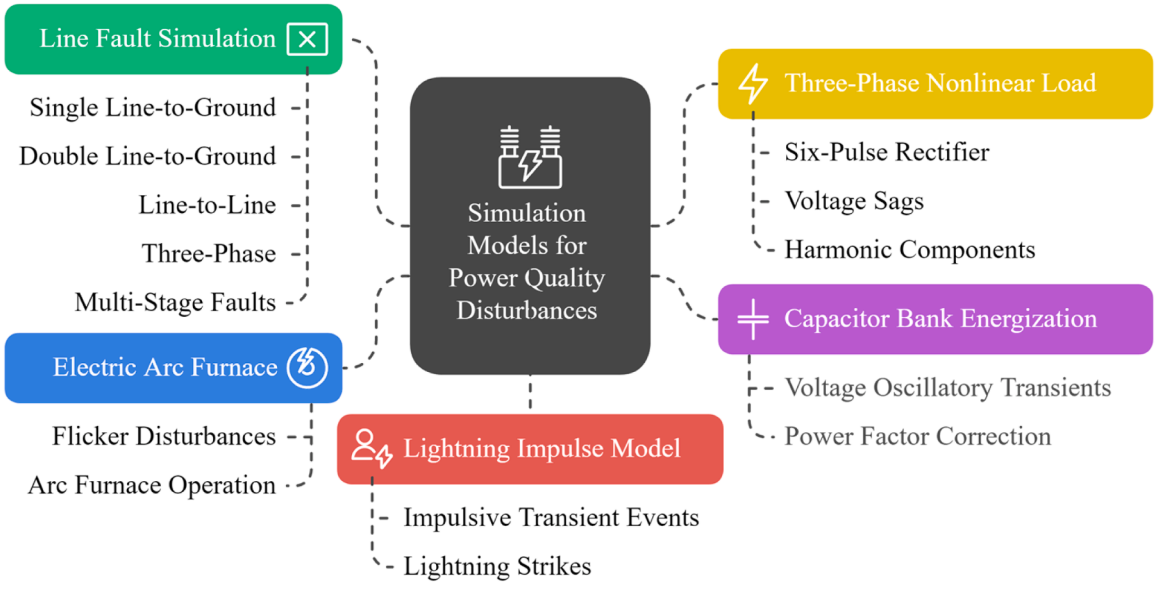

Fig. 12. Simulation models for generating PQD scenarios in MATLAB/Simulink.

Table 8
Simulation validation results.

Disturbance term	Accuracy (%)	Average accuracy (%)	Test time per sample (ms)
C1	98.34		
C2	98.12		
C3	94.52		
C4	99.79		
C5	99.84		
C6	98.53		
C7	98.61		
C8	99.37		
C9	97.52		
C10	97.48		
C11	99.75	98.37	169
C12	96.88		
C13	99.74		
C14	99.86		
C16	98.64		
C17	95.83		
C19	99.70		
C21	98.45		
C22	98.69		
C23	96.58		
C24	98.62		
C25	99.28		

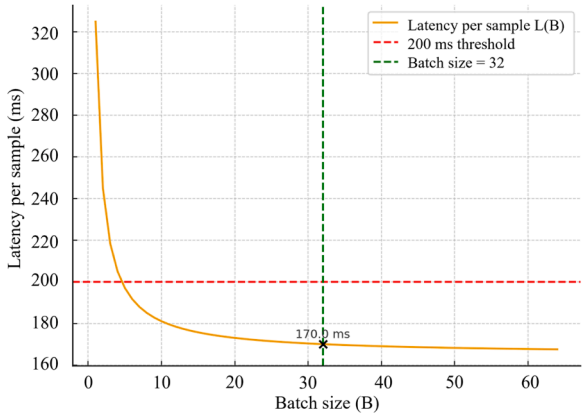

Fig. 13. Latency per sample as a function of batch size.

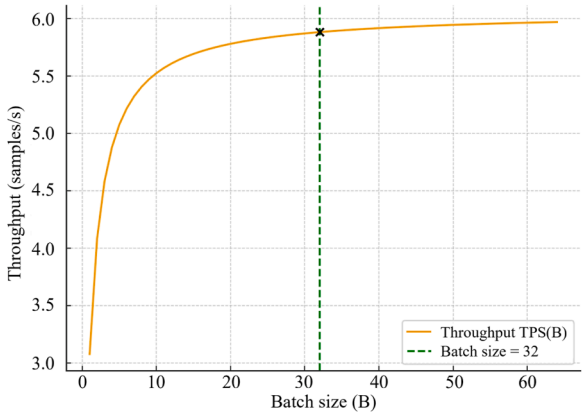

Fig. 14. Throughput performance as a function of batch size.

AMD Ryzen 9 5900HS CPU, an NVIDIA RTX 3060 GPU, and 16 GB of DDR4 RAM. The software stack included Python 3.9.7 and PyTorch 2.3.0. During inference, a batch size of 32 was adopted to measure the per-sample latency, while the learning rate during model training was set to 0.001.

One thousand samples were generated for each simulated PQD type, with results in Table 8 showing an average classification accuracy of 98.37 %, indicating strong reliability and generalization in simulated environments. The average inference time per sample was 169 ms, well below the 200-millisecond real-time threshold [28]. This confirms that the proposed framework delivers high accuracy and low latency and is suitable for practical power quality monitoring and disturbance diagnosis.

5.3. Scalability and real-world applicability

The framework achieved an average per-sample inference latency of 169 ms when evaluated with batch size $B = 32$. To analyze scalability under high sensor density, we model the batch processing time as $T(B) = T_0 + kB$, where T_0 is the fixed overhead and k is the compute cost per sample. The average latency per sample within the batch is $L(B) = \frac{T(B)}{B} = \frac{T_0}{B} + k$, and the effective samples per second is $TPS(B) = \frac{B}{T(B)} = \frac{1}{k+\frac{T_0}{B}} \times 1000$. Calibration is performed to match the experimental observation at $B = 32$, Choosing $k = 165ms$ and $T_0 = 160ms$ yields:

$$\begin{cases} L(32) = \dfrac{160}{32} + 165 = 169ms; \\ T(32) = 160 + 165 \times 32 = 5440ms; \\ TPS(32) = \dfrac{32}{5440} \times 1000 \approx 5.9 samples \Big/ s. \end{cases} \tag{14}$$

As illustrated in Fig. 13 and Fig. 14, the proposed framework exhibits a clear trade-off between latency and throughput when varying the batch size. The latency curve rapidly declines as the batch size increases, approaching an asymptotic value determined by the intrinsic per-sample computational cost. At batch size $B = 32$, the average latency is approximately 170 ms, remaining below the 200 ms threshold required for real-time PQD recognition. In contrast, the throughput curve steadily increases with larger batch sizes, reaching about 5.9 samples/s at $B = 32$. These results indicate that the framework can simultaneously achieve real-time latency compliance and efficient utilization of computational resources.

For PQD monitoring with a reporting period $\Delta t = 200ms$, real-time operation at each measurement point requires $L(B) \le \Delta t$, which is satisfied at $B = 32$ because $170ms < \Delta t$. To assess system capacity, the maximum number of measurements points that one GPU can handle is approximate as:

$$N_{max}(B) = \lfloor \mathrm{TPS}(B) \cdot \Delta t \rfloor \tag{15}$$

Substituting the calibrated values yields $N_{max}(32) = \lfloor 5.9 \times 0.2 \rfloor = \lfloor 1.18 \rfloor = 1$, meaning that, under the current implementation, a single GPU reliably supports one high-frequency measurement stream in strict real time.

This baseline can be extended in two ways. First, with multiple GPUs working in parallel, the total serviceable sensor counts increases approximately linearly as $N_{\max}^{(\mathrm{total})}(32) \approx G \cdot N_{\max}(32)$. For instance, $G = 8$ GPUs would support at least eight real-time data streams. Second, reducing either T_0 or k directly enhances scalability. For example, halving the overhead to $T_0 = 80\,ms$ and reducing compute to $k = 120\,ms$ lowers the latency to $L(32) = \frac{80}{32} + 120 \approx 122.5ms$, while boosting throughput to $TPS(32) = \frac{32}{80+120\times32} \times 1000 \approx 8.2 samples/s$, which increases the single-GPU capacity to $N_{max}(32) = \lfloor 8.2 \times 0.2 \rfloor = 1$, but allows greater margin for robustness and smoother scaling to multi-GPU deployments.

Although the current implementation supports one strict real-time stream per GPU at $B = 32$ with a latency of approximately 170 ms, the analytical results demonstrate clear scalability pathways. By leveraging multi-GPU parallelism and reducing fixed overhead and per-sample computation, the framework can maintain less than 200 ms latency while extending to large numbers of concurrent measurement points,

Table 9
Performance evaluation on the inverter-based photovoltaic substation dataset.

Disturbance term	Count	Accuracy (%)	Recall (%)	Precision (%)	F1 (%)	Average test time (ms)
C1	122	98.12	97.95	98.32	98.13	169
C2	98	97.92	97.54	98.21	97.87	165
C3	21	94.31	94.05	94.67	94.36	171
C4	136	99.61	99.48	99.70	99.59	160
C5	174	98.86	98.54	99.01	98.77	168
C6	92	98.35	98.02	98.61	98.31	164
C7	48	98.45	98.23	98.58	98.40	170
C8	56	99.13	98.95	99.21	99.08	163
C9	34	97.35	97.10	97.63	97.36	166
C10	41	97.29	97.01	97.50	97.25	161
C11	39	99.53	99.42	99.62	99.52	167
C12	52	97.58	97.26	97.89	97.57	162
C13	28	99.64	99.50	99.72	99.61	171
C14	45	99.43	99.21	99.58	99.40	164
C15	37	99.03	98.88	99.20	99.04	160
C16	63	98.43	98.10	98.61	98.35	172
C17	26	95.32	95.05	95.67	95.36	165
C18	44	98.34	98.15	98.50	98.32	168
C19	19	99.53	99.39	99.61	99.50	161
C20	23	99.10	98.91	99.23	99.07	170
C21	31	98.27	98.02	98.47	98.24	162
C22	27	98.45	98.15	98.62	98.38	166
C23	24	96.36	96.05	96.71	96.38	163
C24	17	98.39	98.12	98.58	98.35	169
C25	12	99.07	98.89	99.23	99.06	160

ensuring its applicability in high-density grid environments.

5.4. Validation on renewable-dominated microgrids

To further validate the applicability of the proposed model in renewable energy microgrids under evolving conditions, we selected a dataset from a solar photovoltaic system-based microgrid. Photovoltaic systems are inverter-based generators composed of PV panels that generate direct current electricity and inverters that continuously convert DC into alternating current. The inverter enables the PV system to be connected to AC electrical installations but can also be a source of power quality issues.

The dataset used in this study is a publicly available European microgrid dataset that includes voltage, current, power, energy, and weather data collected from a low-voltage substation and surrounding households with high rooftop PV penetration [29]. We selected the three-phase voltage measurements from the Alverston Close substation, comprising 10,990 hourly samples recorded over 480 days. Following our methodology, each three-phase voltage window was transformed using the GADF method to generate two-dimensional feature images. These were then fed into the proposed GADF-Swin T-TFSGAM classifier to identify 25 categories of PQDs, covering both single events and composite events typically associated with inverter-dominated feeders.

The validation results are summarized in Table 9. Overall, the proposed model demonstrates strong performance across all 25 PQD categories, with accuracies consistently above 94 %, and most classes exceeding 98 %. The recall, precision, and F1-scores also remain high, with minimal class-to-class variation, while the average test time per sample remains within 160–172 ms, meeting real-time application requirements. Importantly, frequent disturbances such as harmonics (C5), voltage imbalance (C4), and voltage sag (C1) were effectively recognized, reflecting the dominant impact of inverter-based PV systems. Less frequent but more complex composite disturbances also achieved satisfactory recognition rates, albeit slightly lower than single-event categories. These findings demonstrate the generalizability of the proposed framework: it performs robustly not only on synthetic and benchmark datasets but also under realistic microgrid conditions with high renewable penetration.

6. Conclusion

This study proposed a hybrid framework integrating GADF for time-series encoding, Swin Transformer for local spatial feature extraction, and TFSGAM for comprehensive global feature learning to accurately classify transient and steady-state PQDs. Experiments on synthetic and real-world datasets demonstrated robustness, generalizability, and efficiency, achieving over 98 % average accuracy in most settings, rapid inference suitable for real-time use, and superior performance over conventional and state-of-the-art methods. The model adapted to complex disturbance patterns, validated successfully on the IEEE PES database and simulated environments, and offered a practical solution for reliable, real-time power quality monitoring.

While the present study demonstrated the robustness and generalizability of the proposed framework on inverter-based PV substation data, future research should extend validation to additional renewable energy scenarios. In particular, incorporating wind power and hybrid microgrid datasets would allow evaluation under broader intermittency patterns and diverse inverter control dynamics. Moreover, exploring real-time online deployment is essential to address the operational requirements of modern smart grids, where disturbance detection must be accurate and computationally efficient for large-scale, high-density sensor networks. These directions will further consolidate the framework's applicability to evolving renewable-dominated distribution systems.

CRediT authorship contribution statement

Jiajian Lin: Writing – review & editing, Writing – original draft, Methodology. **Jalal Tavalaei:** Writing – review & editing, Supervision. **Mehran Motamed Ektesabi:** Supervision. **Hadi Nabipour Afrouzi:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] T.A. Rajaperumal, C.C. Columbus, Transforming the electrical grid: the role of AI in advancing smart, sustainable, and secure energy systems, Energy Inform. 8 (2025) 51, https://doi.org/10.1186/s42162-024-00461-w.

[2] F. Malik, M. Khan, T. Rahman, M. Ehtisham, M. Faheem, Z. Haider, M. Lehtonen, A comprehensive review on voltage stability in wind-integrated power systems, Energies. (Basel) 17 (2024) 644, https://doi.org/10.3390/en17030644.

[3] Y. Yu, G.-P. Liu, Y. Huang, J.M. Guerrero, Coordinated predictive secondary control for DC microgrids based on high-order fully actuated system approaches, IEEe Trans. Smart. Grid. 15 (2024) 19–33, https://doi.org/10.1109/TSG.2023.3266733.

[4] Y. Yu, G.-P. Liu, Y. Huang, C.Y. Chung, Y.-Z. Li, A blockchain consensus mechanism for real-time regulation of renewable energy power systems, Nat. Commun. 15 (2024) 10620, https://doi.org/10.1038/s41467-024-54626-y.

[5] M. Satyanrayana, V. Veeramsetty, D. Rajababu, The analysis of short duration power quality disturbances using short time fourier transform, in: 2025 IEEE 1st International Conference on Smart and Sustainable Developments in Electrical Engineering (SSDEE), IEEE, 2025, pp. 1–6, https://doi.org/10.1109/SSDEE64538.2025.10967837.

[6] H. Bai, R. Yao, W. Zhang, Z. Zhong, H. Zou, Power quality disturbance classification strategy based on fast S-transform and an improved CNN-LSTM hybrid model, Processes 13 (2025) 743, https://doi.org/10.3390/pr13030743.

[7] B. Rathore, A.K. Raghav, R. Soni, Power quality event detection and classification using wavelet-alienation based scheme," E-Prime - Advances in Electrical Engineering, Electron. Energy 13 (2025) 101040, https://doi.org/10.1016/j.prime.2025.101040.

[8] L. Kang, B. Xu, P. Li, K. Wang, J. Chen, H. Du, Q. Liu, L. Zhang, X. Lian, Controllable preparation of low-cost coal gangue-based SAPO-5 molecular sieve and its adsorption performance for heavy metal ions, Nanomaterials 15 (2025) 366, https://doi.org/10.3390/nano15050366.

[9] N. Cao, H. Du, J. Lu, Z. Li, Q. Qiang, H. Lu, Designing ionic liquid electrolytes for a rigid and Li+-conductive solid electrolyte interface in high performance lithium metal batteries, Chem. Phys. Lett. 866 (2025) 141959, https://doi.org/10.1016/j.cplett.2025.141959.

[10] H. Zhang, Z. Li, Y. Liu, X. Du, Y. Gao, W. Xie, X. Zheng, H. Du, Oxygen vacancies-modulated C-WO3/BiOBr heterojunction for highly efficient benzene degradation, Vacuum. 234 (2025) 114117, https://doi.org/10.1016/j.vacuum.2025.114117.

[11] J. Lin, J. Tavalaei, L.Y. Yeo, Y. Zhou, H.N. Afrouzi, M.M. Ektesabi, Hybrid CNN-BiLSTM model for power quality disturbance classification, in: 2024 IEEE International Conference on Advanced Power Engineering and Energy (APEE), IEEE, 2024, pp. 112–116, https://doi.org/10.1109/APEE60256.2024.10790890.

[12] G. Wang, H. Zhang, M. Gao, W. Ding, Y. Qian, Identification and classification of power quality disturbances using CNN-transformer, J. Electric. Eng. Technol. 20 (2025) 2993–3007, https://doi.org/10.1007/s42835-025-02213-6.

[13] S. Mukherjee, S. Chatterjee, R. Mandal, Deep learning aided power quality disturbance detection with improved time–frequency resolution employing adaptive superlet transform, Electric. Eng. 107 (2025) 8101–8113, https://doi.org/10.1007/s00202-025-02961-8.

[14] W. Liu, X. Ye, W. Yan, Power quality disturbance classification based on dual-parallel 1D2D fusion of improved ResNet and attention mechanism, Measurement. (Lond) 252 (2025), https://doi.org/10.1016/j.measurement.2025.117358.

[15] C.-L. Yang, Z.-X. Chen, C.-Y. Yang, Sensor classification using convolutional neural network by encoding multivariate time series as two-dimensional colored images, Sensors 20 (2019) 168, https://doi.org/10.3390/s20010168.

[16] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, others, An image is worth 16×16 words: Transformers for image recognition at scale, ArXiv Preprint ArXiv: 2010.11929 (2020).

[17] B. Cheng, I. Misra, A.G. Schwing, A. Kirillov, R. Girdhar, Masked-attention mask transformer for universal image segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 1290–1299.

[18] Y. Liu, Z. Shao, N. Hoffmann, Global attention mechanism: retain information to enhance channel-spatial interactions, ArXiv Preprint ArXiv:2112.05561 (2021).

[19] Z. Wang, S. Zhou, J. Chen, Z. Zhang, B. Hu, Y. Feng, C. Chen, C. Wang, Dynamic graph transformer with correlated spatial-temporal positional encoding, in: Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining, ACM, New York, NY, USA, 2025, pp. 60–69, https://doi.org/10.1145/3701551.3703489.

[20] I.P. Quality, IEEE recommended practice for monitoring electric power quality, in: IEEE Recommended Practice for Monitoring Electric Power Quality, 1995.

[21] L. Huang, F. Wang, Y. Zhang, Q. Xu, Fine-grained ship classification by combining CNN and swin transformer, Remote Sens. (Basel) 14 (2022) 3087, https://doi.org/10.3390/rs14133087.

[22] J. Jiang, H. Wu, C. Zhong, Y. Cai, H. Song, A novel methodology for microgrid power quality disturbance classification using URPM-CWT and Multi-Channel feature Fusion, IEEe Access. 12 (2024) 35597–35611, https://doi.org/10.1109/ACCESS.2024.3350170.

[23] Y.S.U. Vishwanath, S. Esakkirajan, B. Keerthiveena, R.B. Pachori, A generalized classification framework for power quality disturbances based on synchrosqueezed wavelet transform and convolutional neural networks, IEEe Trans. Instrum. Meas. 72 (2023) 1–13, https://doi.org/10.1109/TIM.2023.3308235.

[24] W. Liu, X. Ye, W. Yan, Power quality disturbance classification based on dual-parallel 1D2D fusion of improved ResNet and attention mechanism, Measurement 252 (2025) 117358, https://doi.org/10.1016/j.measurement.2025.117358.

[25] M.D. Fitri Mat Zabidi, S. Shahbudin, S.I. Sulaiman, F.Y. Abdul Rahman, H. Saad, Power quality disturbances classification analysis using EfficientNet architecture, in: 2024 IEEE 15th Control and System Graduate Research Colloquium (ICSGRC), IEEE, 2024, pp. 64–69, https://doi.org/10.1109/ICSGRC62081.2024.10691145.

[26] Z. Duan, Z. Peng, J. Song, S. Lu, An intelligent complex power quality disturbance recognition method based on two dimension encoding conversion and machine vison, Electr. Power Syst. Res. 232 (2024) 110413, https://doi.org/10.1016/j.epsr.2024.110413.

[27] R.H.G. Tan, V.K. Ramachandaramurthy, A comprehensive modeling and simulation of power quality disturbances using MATLAB/SIMULINK. Power Quality Issues in Distributed Generation, InTech, 2015, https://doi.org/10.5772/61209.

[28] J. Ma, Q. Tang, M. He, L. Peretto, Z. Teng, Complex PQD classification using time-frequency analysis and multiscale parallel attention residual network, IEEE Trans. Ind. Electron. 71 (2024) 9658–9667, https://doi.org/10.1109/TIE.2023.3323692.

[29] UK Power Networks, Photovoltaic (PV) Solar Panel Energy Generation Data, European Data Portal, 2025. http://data.europa.eu/88u/dataset/photovoltaic-pv-solar-panel-energy-generation-data.