AIMS Energy, 13(5): 1052–1075.
DOI: 10.3934/energy.2025039
Received: 19 June 2025
Revised: 05 August 2025
Accepted: 28 August 2025
Published: 11 September 2025

https://www.aimspress.com/journal/energy

Research article

Detection and classification of power quality disturbances: Vision transformers vs. CNN

Muhammad Hassan Anwar, Mirza Muhammad Ali Baig, Abdurrahman Javid Shaikh* and Abdul Ghani Abro

Department of Electrical Engineering, NED University of Engineering and Technology, Karachi-75270, Sindh, Pakistan

* **Correspondence:** Email: arjs@neduet.edu.pk; Tel: +923333065581.

Abstract: The increasing integration of renewable energy sources and widespread use of nonlinear power-electronic devices have amplified the occurrence of power quality disturbances (PQDs), which can disrupt sensitive equipment and jeopardize the safe and efficient operation of smart grids. Existing approaches for PQD classification, including traditional signal processing methods and deep learning models, often face limitations in accurately handling the nonlinear and non-stationary nature of PQDs. This study proposes a novel two-step approach that combines the smoothed pseudo Wigner-Ville distribution (SPWVD) with a vision transformer (ViT) model for effective PQD detection and classification. In the proposed method, synthetic PQD signals were generated using MATLAB in accordance with IEEE 1159 standards, and 1-D signals were transformed into 2-D time-frequency images using SPWVD to enhance feature representation. These images were then classified using a ViT model, leveraging the self-attention mechanism to capture global relationships in the data. Experimental results demonstrated that the proposed ViT-SPWVD approach achieved a high classification accuracy of 98.94%, indicating its capability and promise for accurate PQD detection. This work represents the first application of a vision transformer for PQD analysis, offering a new direction for transformer-based models in power system monitoring.

Keywords: convolutional neural networks; deep learning; power delivery; power quality disturbances; smoothed pseudo Wigner-Ville distribution; vision transformer

1. Introduction

Rising energy demand is compelling a shift toward renewable energy. In response, the extensive adoption of clean energy, pervasive usage of nonlinear loads, and integration of distributed generation networks in power grids introduce numerous challenges, particularly, in maintaining power quality [1]. The international standards encompass a diverse range of PQDs, including *voltage sag*, *voltage swell*, *harmonics*, *interruptions*, *flickers,* and many other waveform anomalies, highlighting the importance of firm initiatives to address them [2–5].

PQDs have profound consequences, impacting both power grids and consumers [6,7]. Voltage-related PQDs such as *swell*, *sag,* and *harmonics* are common causes of overheating and equipment failure. Transients in the voltage level can lead to insulation failure of connected equipment. Harmonics are also known for adversely affecting the performance of protective systems and measuring instruments [8,9]. Therefore, the development of an intelligent technique with the skills to automatically detect and classify PQDs is crucial.

1.1. Background

The detection and classification of PQDs conventionally follow a two-step procedure involving feature extraction and classification. Numerous researchers extensively employed a combination of signal-processing techniques and intelligent classifiers to detect and classify PQDs. Signal-processing techniques for features extraction such as the S-transform (ST), Fourier transform (FT) [10], short-time Fourier transform (STFT) [11,12], wavelet transform (WT) [13–15], Wigner-Ville distribution (WVD) [16], and empirical mode decomposition (EMD) [17] have been widely adopted.

The FT is straightforward to implement but lacks time resolution, making it unsuitable for analyzing localized disturbances. The STFT addresses this with a sliding window, but the fixed window size imposes a trade-off between time and frequency resolution—leading to low accuracy for fast-changing or short-lived events. The WT provides multi-resolution capabilities and improved localization, but its performance is highly dependent on the choice of mother wavelet and is often sensitive to noise. The S-transform (ST) offers better noise robustness but at higher computational cost, making it less suitable for real-time systems. EMD, though adaptive, suffers from mode-mixing, where overlapping signal modes lead to unreliable spectral separation [18].

The Wigner-Ville distribution (WVD) offers excellent resolution but is affected by severe cross-term interference. To mitigate this, we adopt the smoothed pseudo Wigner-Ville distribution (SPWVD) in our study. The SPWVD applies separate smoothing windows in both the time and frequency domains, significantly reducing cross-terms while preserving the high-resolution characteristics of the WVD. Among various time-frequency methods, the SPWVD has been shown to outperform the STFT and even advanced transforms like the Gabor-Wigner in analyzing harmonics and transient disturbances in power signals. Thus, it is particularly well-suited for analyzing nonlinear, non-stationary signals such as PQDs.

Recent comparative studies in biomedical and electrical systems have validated the effectiveness of the SPWVD over other time-frequency techniques. In PQD applications, the SPWVD has been shown to outperform the WT and ST in terms of localization and classification accuracy—especially when used with advanced classifiers like CNNs and transformers [19]. Table 1 shows a comprehensive comparison of the SPWVD and other methods.

Table 1. Comparison of the SPWVD with others.

Feature	Time-Frequency Resolution	Cross-Term Suppression	Noise Robustness
STFT [20]	Moderate	Good	Low
WT [21,22]	High	Good	Moderate
S-Transform [23]	Moderate	Good	Moderate
WVD [20]	Very High	Poor	Low
SPWVD [24]	Very High	Excellent	High

Several intelligent classifiers, such as support vector machines (SVMs), decision trees (DTs), k-nearest neighbors (KNNs), and artificial neural networks (ANNs) have been employed for the classification of PQDs. The classification of PQDs by utilizing a combined approach of the WT and SVM (WSVM) is explored in [25]. This WSVM technique involves a pre-processing approach and can handle PQDs in the presence of noise. This method aims to detect and classify 100 cases of nine distinct PQDs by extracting six dimensional features from the wavelet transform and SVM. The experimental evaluation shows a maximum classification accuracy of 98.85% for 50 dB noisy PQDs. In [26], a pre-processing method of the fast discrete S-transform (FDST) algorithm and a simple decision tree classifier-based approach has been implemented for PQD classification. The method's performance has been evaluated on 13 PQDs at 40 dB noise and achieved accuracy of 98.85%. In [27], PQDs using an S-transform, ANN classifier, and rule-based decision tree algorithm evaluated 13 PQDs and yielded 99.9% accuracy. Khokhar S et al. in [28] applied a discrete wavelet transform (DWT) and artificial bee colony optimized probabilistic neural network (ABC-PNN) for the classification of PQDs. This method resulted in accuracy of 99.8% on 16 noiseless PQDs and 98.6% on 40 dB noise PQD cases.

Nevertheless, in the case of SVMs, with an increase in the input size, the training time increases, resulting in high computational complexities which persist even after applying the kernel trick corrections [29]. Another drawback of SVMs is their sensitivity to noisy data, making them ineffective when noise is high [30] and having higher cumulative errors [18]. In contrast to SVMs, ANNs do not have cumulative error issues in PQD classification [18]. Therefore, deep learning methods have now become preferred choices. Moreover, deep neural networks (DNNs) can simultaneously deal with the nonlinear and non-stationary behavior of real-world signals [18].

Convolutional neural networks (CNNs) and long short-term memory (LSTM) networks are the deep learning methods that have ruled over the decade, specially CNNs [31]. Deep learning approaches successfully detect and classify PQDs with or without pre-processing techniques [18,32,33]. In [32], a deep convolutional neural network on a 1-D approach was used without any pre-processing stage. The method classified 16 PQDs with an accuracy of 99.96%. In [18], a Wigner-Ville distribution (WVD) and CNN-based two-step approach was employed. This method was used to investigate nine PQDs and achieved an accuracy of 99.67%. Sindi H et al. [33] used the combination of 1-D and 2-D CNN architectures, which resulted in 99.97% accuracy on 13 PQDs. In [34], a comparison of the CNN, LSTM, and a hybrid approach of using CNN-LSTM for PQD problems led to the classification accuracy of 79.14% for LSTM, 84.58% for CNN, and 84.76% for CNN-LSTM networks.

(a) (b)

Figure 1. (a) Actual face image and (b) shuffled face image.

LSTM and CNNs have shortcomings, too. First, CNNs are unable to capture long-distance dependency among the input elements [35]. Second, common CNN architectures consist of convolutional, pooling, and fully connected layers. The pooling layer helps CNNs to achieve translational invariance, but it may also lead to wrong prediction [36]. The problem is exemplified in Figure 1, highlighting the shuffling in an image of a face. CNNs may classify Figure 1(b) as a face because of the loss of spatial information. To overcome these drawbacks, the Google Brain team introduced another competitive deep learning method, the transformer [37], which works on the global relationship among input elements with positional encoding.

The transformer has resulted in astonishing results in natural language processing (NLP). Moreover, the vision transformer (ViT), developed for image recognition in 2020, outperformed the state-of-the-art CNNs while requiring substantially fewer computations [38]. ViTs exhibit higher robustness toward occlusion, perturbation, and domain shifts than CNNs. Furthermore, their shape bias is nearly as high as a human's ability to recognize shapes.

All of the above-mentioned intriguing properties and outstanding performance of the ViT model have urged researchers to apply this approach in the power system field. Hence, it is reasonable to believe that a new ViT-based intelligent technique would overcome the drawbacks of previously applied methods, and detect and classify PQDs accurately. To the best of our knowledge, this is the first time a ViT model is being employed for PQD problems. The rest of the research work is divided into proposed methodology, classification methods, results, and discussions.

1.2. Methodology

The two-dimensional (2-D) method is superior at providing distinct and diverse features helping with classifying disturbances that do not show abrupt waveform discontinuities, such as *sag*, *flicker*, and *swell*. Therefore, the 2-D approach exemplified in Figure 2 has been adopted in this work. A 1-D raw voltage data sequence is provided to the SPWVD to transform it into a 2-D (time-frequency) image. In the following sub-section, fundamentals of the SPWVD will be discussed.

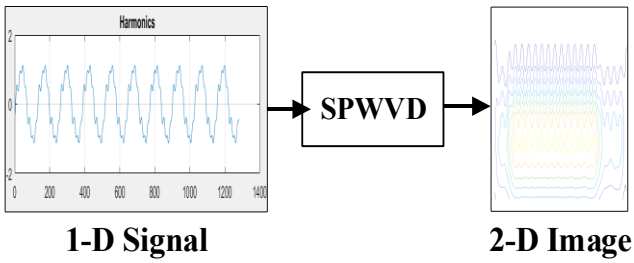

Figure 2. Transformation of 1-D raw data to a 2-D image.

2. Methodology

2.1. The smoothed pseudo Wigner-Ville distribution (SPWVD)

The Wigner-Ville distribution [24] is a tool to generate high-resolution time-frequency information out of a 1-D signal. The WVD faces a cross-term interference drawback which is overcome by introducing a window. Therefore, the smoothed pseudo Wigner-Ville distribution (SPWVD), a variant of the WVD, is implemented in this work. The SPWVD utilizes an independent window to smooth the transformation result in the time and frequency domains. In this study, a Kaiser window has been used as both the time and frequency smoothing windows. The SPWVD and Kaiser [39,40] expressions are given here:

$$SPWVD_s\,(t,f) \;=\; \int_{-\infty}^{\infty} k \int_{-\infty}^{\infty} (l \times m) dt' \cdot e^{-j2\pi f\tau}\, d\tau \tag{1}$$

$$\hat{s} \;=\; \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{s(\tau)}{t-\tau}\, d\tau \tag{2}$$

$$w(n) \;=\; \frac{I_o\left(\beta\sqrt{1-\left(\frac{n-0.5N}{0.5N}\right)}\right)}{I_o(\beta)} \tag{3}$$

where,

$$k \;=\; h\left(\frac{\tau}{2}\right)\, h^*\left(\frac{\tau}{2}\right)$$

$$l \;=\; g(t-t')\, s\left(t'+\frac{\tau}{2}\right)$$

$$m \;=\; s^*\left(t'-\frac{\tau}{2}\right)$$

β: the shape factor, which controls side lobe suppression and main lobe width (β has been set at 8)
N: total number of samples in the window
n: sample index ranging from 0 to N - 1
I_o: modified Bessel function of zeroth order
h(t): window reducing crossterms in the time domain
g(t): window reducing crossterms in the frequency domain
s(t): investigated signal processed with the help of Hilbert transformation (2) [39]

Once the 2-D images are generated for every PQD, the ViT model is trained for PQD classification. The following well-established steps were adopted to implement the SPWVD in the time-frequency domain [19]:

i) In the time domain, the autocorrelation function (ACF) of the signal is computed.
ii) A Fourier transform is applied to the ACF to achieve frequency-domain representation.
iii) Typically, a Fourier-transformed ACF is multiplied by the smoothing window function in the time domain.
iv) The inverse Fourier transform of the smoothed result is calculated to achieve the SPWVD in the time-frequency domain.

2.2. Vision transformer (ViT)

The architecture of the ViT is illustrated in Figure 3 and explained below:

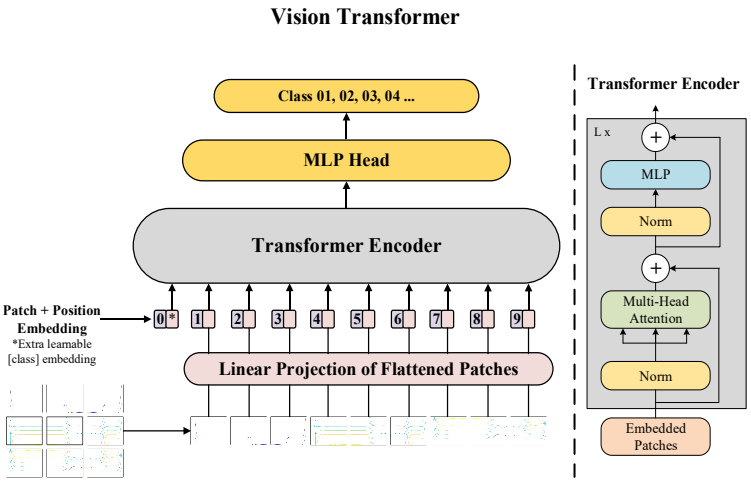

Figure 3. Vision transformer architecture.

2.2.1. Number of patches

The input image is broken into patches to leverage the self-attention mechanism through positional information. The input image of height H, width W, and channel C is cut into N number of 2-D patches of height P, width P, and channel C as expressed in Eq (4).

$$N = \frac{HW}{P^2} \tag{4}$$

2.2.2. Patch and positional embedding

Each patch $\boldsymbol{x}$ is flattened, unrolled into a 1-D vector with length CP^2, and mapped to D dimensions with the trainable linear projection embedding matrix $\boldsymbol{E}$ (Figure 3). With m number of flattened patches, an extra learnable embedding (a special *class* token) is then fed to the $m + 1$ vector sequence by adding positional encoding to retain the original position of each patch, as presented

in Eq (5). A *class* token ($z_0^0 = x_{class}$) is randomly initialized at the encoder's input state with the patch embedding [38]. It will be used at the final classification stage as z_L^0 in the multilayer perceptron (MLP) to represent image y in (6).

$$z^0 = \left[x_{class};\ x_p^1\boldsymbol{E};\ x_p^2\boldsymbol{E};\ \ldots;\ x_p^N\boldsymbol{E}\right] + \boldsymbol{E}_{pos} \tag{5}$$

where,

$$\boldsymbol{E} \in \mathbb{R}^{(P^2.C)\times D}, \boldsymbol{E}_{pos} \in \mathbb{R}^{(N+1)\times D}$$

$$y = LN(z_L^0) \tag{6}$$

2.2.3. Vision transformer encoder

The vision transformer encoder [38], depicted in Figure 4, contains blocks of multi-head self-attention (MHSA) and multilayer perceptron (MLP) along with a normalization layer (LN). After each block, a residual connection [41] is implemented. The output of combined embedding with a dimension of $((N+1)\times P^2\times C)$ is fed to the encoder. First, inputs are passed to the LN and then fed to the MHSA block. *Query*, *Key,* and *Value* matrices are obtained from the converted input matrix with a dimension of $((N+1)\times(P^2C\times 3))$ using a linear layer. Each *QKV* matrix of dimension $((N+1)\times P^2C)$ is reshaped into a new dimension of $\left((N+1)\times 3\times P^2C\right)$. The final shape of the *QKV* matrix with a dimension of $\left(h\times(N+1)\times\left(\frac{P^2C}{h}\right)\right)$ is achieved by utilizing *h* heads for multilayer self-attention.

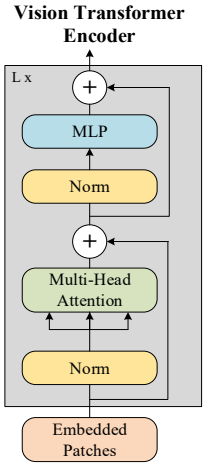

Figure 4. Vision transformer encoder.

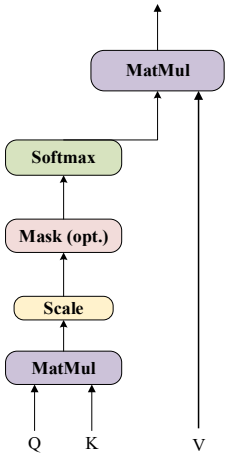

Figure 5. Self-attention mechanism.

After achieving the *QKV* matrix, the MHSA block performs attention concatenated by Eqs (7) and (8) [37]. The output from the MHSA block is passed through a skip connection, and then the input goes through a normalization layer before feeding to the MLP block. The scaled-dot product attention, Figure 5, has an input of queries and keys with a dimension of d_k and values with a dimension of d_v. The dot product is computed on queries with all keys and scaled by a factor of $1/\sqrt{d_k}$. Then, to have weights on the values, a *softmax* function is applied. Therefore, queries, keys, and values are packed

together into matrices *Q, K,* and *V,* respectively. The output matrix is computed as:

$$Attention\ (Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)\cdot V \tag{7}$$

In MHSA, the attention function is performed in parallel by h times linearly projecting the queries, keys, and values with learned and different linear projections to $d_q, d_k,$ and d_v dimensions, respectively, as shown in Eq (7) and Figure 6.

$$MultiHead(Q,K,V) = Concat(head_1,\dots,head_h)W^O \tag{8}$$

The projections are parameter matrices:

$$W_i^Q \in \mathbb{R}^{d_{model}\times d_k},$$

$$W_i^K \in \mathbb{R}^{d_{model}\times d_k},$$

$$W_i^V \in \mathbb{R}^{d_{model}\times d_v}.$$

In the MLP block, two layers with Gaussian error linear unit (GELU) nonlinearity are implemented as defined in Eqs (9–12) [38].

$$z_l' = MSA\big(LN(z_{l-1})\big) + z_{l-1} \tag{9}$$

$$z_l = MLP\big(LN(z_l')\big) + z_l' \tag{10}$$

$$GELU(x) = x\,P(X \le x) = x\,\Phi(x) \tag{11}$$

$$GELU(x) \approx 0.5x\left(1+\tanh\left(\sqrt{\frac{2}{\pi}}(x+0.044715x^3)\right)\right) \tag{12}$$

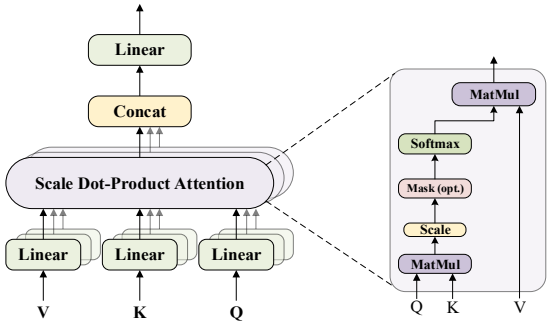

Figure 6. Multi-head self-attention.

The output of the MLP block in the encoder is then passed to a skip connection. This process repeats for multiple transformer layers of a deep ViT structure. To reduce training duration, normalization of neurons' activities is applied. For this purpose, the layer normalization (LN) technique is used in the encoder around each sublayer.

2.3. Convolutional neural networks (CNNs)

CNNs are primarily utilized for pattern recognition tasks within an image and address the problem of computational complexity which ANNs show while computing image data. Typically, CNN frameworks contain fundamental components, involving an input layer, convolutional layer, pooling layer, fully connected layer, and output layer [18]. CNN architectures are well-known and may be found in multiple books and research articles.

The low-level local input features of images are first transferred to high-level global features by employing a mathematical 2-D convolutional operation. The convolutional layer also contributes in reduction of the model's complexity by utilizing three hyperparameters, involving depth, stride, and setting zero-padding. Further, to achieve the desired outcome, an elementwise activation function rectified linear unit (ReLu) is employed, outputted by the preceding layer [18]. This is a crucial and completely connected layer in CNNs to process global information extracted from the feature map.

3. Proposed classification method

3.1. Numerical models of PQDs

Table 2. Numerical model of the simulated PQDs.

Label	PQDs	Numerical Model	Parameters
Sag	Sag	$y(t) = \left[1 - \alpha\left(u(t-t_1) - u(t-t_2)\right)\right] \times \sin(\omega t - \varphi)$	$0.1 \le \alpha \le 0.9$, $T \le t_2 - t_1 \le (N-1)T$, $-\pi \le \varphi \le \pi$
Swell	Swell	$y(t) = \left[1 + \beta\left(u(t-t_1) - u(t-t_2)\right)\right] \times \sin(\omega t - \varphi)$	$0.1 \le \beta \le 0.8$, $T \le t_2 - t_1 \le (N-1)T$, $-\pi \le \varphi \le \pi$
Inter	Interruption	$y(t) = \left[1 - \rho\left(u(t-t_1) - u(t-t_2)\right)\right] \times \sin(\omega t - \varphi)$	$0.9 \le \rho \le 1.0$, $T \le t_2 - t_1 \le (N-1)T$, $-\pi \le \varphi \le \pi$
Spike	Spike	$y(t) = [\sin(\omega t - \varphi) - \psi\left(e^{-750(t-t_a)} - e^{-344(t-t_a)}\right) \times (u(t-t_a) - u(t-t_b))]$	$0.222 \le \psi \le 1.11$, $-\pi \le \varphi \le \pi$, $t_b = t_a + 1ms$

Continued on next page

Label	PQDs	Numerical Model	Parameters
Osc	Oscillatory Transient	$y(t) = [\sin(\omega t - \varphi) + \beta e^{-\left(\frac{t-t_I}{\tau}\right)} \sin(\omega_n(t-t_I) - \upsilon) \times \left(u(t-t_{II}) - u(t-t_I)\right)]$	$0.1 \le \beta \le 0.8$, $-\pi \le \varphi \le \pi$, $\omega_n = 2\pi f_n$, $300\ Hz \le f_n \le 900\ Hz$, $8ms \le \tau \le 40ms$, $-\pi \le \upsilon \le \pi$, $0.5T \le t_{II} - t_I \le \frac{N}{3.33}$,
Har	Harmonics	$y(t) = \alpha_1 \sin(\omega t - \varphi) + \alpha_3 \sin(3\omega t - \vartheta_3) + \alpha_5 sin5(5\omega t - \vartheta_5)$	$0.05 \le \alpha_n \le 0.15$, $-\pi \le \varphi \le \pi$, $-\pi \le \vartheta_3, \vartheta_5 \le \pi$
Har with Sag	Harmonics with Sag	$y(t) = \left[1 - \alpha\left(u(t-t_1) - u(t-t_2)\right)\right] \times (\alpha_1 \sin(\omega t - \varphi) + \alpha_3 \sin(3\omega t - \vartheta_3) + \alpha_5 sin5(5\omega t - \vartheta_5))$	$0.1 \le \alpha \le 0.9$, $T \le t_2 - t_1 \le (N-1)T$, $0.05 \le \alpha_n \le 0.15$, $where\ n = 3, 5, and\ 7$, $-\pi \le \varphi \le \pi$, $-\pi \le \vartheta_3, \vartheta_5 \le \pi$
Har with Swell	Harmonics with Swell	$y(t) = \left[1 + \beta\left(u(t-t_1) - u(t-t_2)\right)\right] \times (\alpha_1 \sin(\omega t - \varphi) + \alpha_3 sin(3\omega t - \vartheta_3) + \alpha_5 sin5(5\omega t - \vartheta_5))$	$0.1 \le \beta \le 0.8$, $T \le t_2 - t_1 \le (N-1)T$, $0.05 \le \alpha_3, \alpha_5 \le 0.15$, $-\pi \le \varphi \le \pi$, $-\pi \le \vartheta_3, \vartheta_5 \le \pi$
Flicker	Flicker	$y(t) = \left[1 + \alpha_f \sin(\beta\omega t)\right] \sin(\omega t - \varphi)$	$0.05 \le \alpha_f \le 0.1$, $8 \le \beta \le 25\ Hz$, $-\pi \le \varphi \le \pi$
Flicker with Sag	Flicker with Sag	$y(t) = \left[1 + \alpha_f \sin(\beta\omega t) - \alpha\left(u(t-t_1) - u(t-t_2)\right)\right] \times \sin(\omega t - \varphi)$	$0.1 \le \alpha \le 0.9$, $T \le t_2 - t_1 \le (N-1)T$, $0.05 \le \alpha_f \le 0.1$, $8 \le \beta \le 25\ Hz$, $-\pi \le \varphi \le \pi$
Flicker with Swell	Flicker with Swell [45–47]	$y(t) = \left[1 + \alpha_f \sin(\beta\omega t) + \beta\left(u(t-t_1) - u(t-t_2)\right)\right] \times \sin(\omega t - \varphi)$	$0.1 \le \beta \le 0.8$, $T \le t_2 - t_1 \le (N-1)T$, $0.05 \le \alpha_f \le 0.1$, $8 \le \beta \le 25\ Hz$, $-\pi \le \varphi \le \pi$

A comprehensive set of 11 PQDs have been synthesized with the help of a numerical model of PQDs listed in Table 2 by utilizing MATLAB, in compliance with IEEE 1159 standards [2]. *Sag, swell, interruption, spike, oscillatory transient, harmonics, harmonics with sag, harmonics with swell, flicker, flicker with sag,* and *flicker with swell* PQDs were selected for this research work. The selection of these PQDs for the classification task is a preferred and widely adopted choice among researchers. These PQDs are constructed with an amplitude of 1 p.u. and cover seven single and four combined PQDs. The fundamental frequency is set at 50 Hz, while PQDs are sampled at the most common sampling rate of 6.4 kHz as used by several researchers [18,42–44]. Each PQD contains 10 cycles, resulting in 128 sampling points in each cycle. Figure 2 visually presents the 1-D signal and its corresponding converted 2-D image obtained using the SPWVD method. Each image is of dimension 781×625 pixels, which serves as the input.

3.2. Vision transformer model of the proposed method

In this work, the ViT model is trained on 2-D RGB images of PQDs from scratch. Self-attention is implemented on patches, so it makes the ViT capable of using high-resolution images. To reduce computational cost, each image is resized from its original dimensions of 781×625 to a size of 224×224, while retaining 3 channels. The square of patch size is inversely proportional to the transformer's sequence length, therefore a patch size value of 32×32 is chosen in this work, resulting in a total of 49 patches (Eq (4)). To ensure consistency in information encoded by the patches, the latent vector *D* with a constant length of 128 is employed across all layers of the model.

After flattening the patches, the trainable linear projection of Eq (5) is utilized to map them to the *D* vector. To ensure the preservation of spatial information, the position embedding is integrated into the patch embedding. In addition, an extra learnable embedding ($z_0^0 = x_{class}$) is introduced alongside the patches to enrich image representation, denoted as *y,* illustrated in Eq (5). This fusion makes the ViT model capable of effectively classifying PQDs based on extracted features.

The input matrix is processed in the transformer encoder through MHSA and MLP blocks. With 8 heads, MHSA performs multi-dimensional attention and capture diverse information of the input data. The captured information is forwarded to the MLP block, which utilizes nonlinear transformation to acquire complex patterns and relationships within the data. To ensure consistent flow of information, each block is accompanied by layer normalization (LN) and a residual connection. This process is iterated for 8 layers for the encoder to facilitate comprehensive and hierarchal handling of the input matrix.

The final output obtained from the ViT encoder is subsequently fed to the MLP head, which plays a key role in classifying PQDs. To address the overfitting issue in network, a residual dropout of 0.1 is employed throughout the process [48]. The value of the learning rate is set at a constant value of 0.0001 and an Adam optimizer is used.

3.3. Convolutional neural network model

A CNN model is also developed in this work to make a comparison with the proposed method. The model contains key components, including an input layer, three convolutional layers, three max-pooling layers, a fully connected layer, and an output layer. The model utilizes 32, 64, and 64 convolutional layers, respectively, each with a fixed kernel size of 3×3. To ensure comprehensive

feature extraction, a zero-padding approach is implemented using the 'same' setting to cover all input elements. Each convolutional layer is employed with the rectified linear unit (ReLU) activation function to capture the nonlinear property of the input. Pooling layers are introduced to reduce the parameter count and prevent overfitting in the network. A fully connected layer implements the softmax function for the classification task and the resulting output is then fed to the output layer.

In this work, tests have been conducted on two datasets as listed in Table 2. The first set contained eight PQDs including six single and two combined PQDs, while the second dataset is comprised of 11 PQDs with seven single and four combined PQDs, as explained previously. A two-step method combining the smoothed pseudo Wigner-Ville distribution with a vision transformer is proposed to detect and classify PQDs in this work. First, 1-D raw data is converted into a 2-D image file and then the ViT model classifies the PQDs. The performance of the proposed method is evaluated on 11 types of synthetic PQDs based on international standards IEEE 1159 [2], IEC 61000 [3], and EN 50160 [4].

3.4. Deployment architecture for real-time PQs classification

The conceptual architecture for real-time integration into the power system can be deployed via cloud-edge coordination. A phasor measurement unit or intelligent electronic devices can be used to capture the power quality signal and then, with the help of FPGA or an embedded system, the SPWVD will transform the signal into a 2-D image. The 2-D image data will be sent on a cloud-based system where the ViT will perform real-time classification of the PQDs. This setup allows long-term storage of the classification results for further analysis eventually leading to enhanced reliability of power systems.

4. Results and discussion

The results of the convolutional neural network and vision transformer models are illustrated with utmost clarity by utilizing confusion matrices and a comprehensive comparison table.

4.1. CNN model results

4.1.1. Confusion matrix of eight PQDs' test data

In the first phase, a comparison of the CNN-SPWVD and CNN-WVD methods has been made. For better comparative study, the eight PQDs listed in Table 3 are the same as in [18]. Table 4 illustrates that each PQD was evaluated with 5000 images. The CNN-SPWVD method demonstrated outstanding performance in detecting four PQDs: swell, oscillatory transient, harmonics, and flicker with 100% accuracy. However, the remaining four PQDs, namely *sag*, *interruption*, *harmonics with sag,* and *harmonics with swell,* could not achieved 100% accuracy individually. Nevertheless, they were identified with accuracy surpassing 99%, which is highly satisfactory. Notably, *sag* and *interruption* are the only two PQDs that exhibited great confusion with each other in this dataset. The comprehensive analysis in Table 5 revealed that the CNN-SPWVD demonstrated remarkable overall test accuracy of 99.86%, outperforming the CNN-WVD [18]. The SPWVD is a superior version of the WVD as it effectively overcomes the problem of cross-term interference, which in many cases occurs in the WVD yielding distorted results.

Table 3. The datasets.

PQDs	Total Images/PQD (Train + Valid)		Training		Validation		Test (where applicable)
	1st set	2nd set	1st set	2nd set	1st set	2nd set	
Sag	20,000	155,000	Randomly	Randomly	Randomly	Randomly	5000
Swell	20,000	155,000	selected 92%	selected	selected 8%	selected 5%	5000
Interruption	20,000	151,667	of total	95% of total	of total	of total	5000
Spike	-	155,000	images	images	images	images	5000
Oscillatory Transient	20,000	155,000					5000
Harmonics	20,000	155,000					5000
Harmonics with Sag	20,000	155,000					5000
Harmonics with Swell	20,000	155,000					5000
Flicker	20,000	155,000					5000
Flicker with Sag	-	155,000					5000
Flicker with Swell	-	155,000					5000
Total Images	160,000	1,701,667	147,200	1,616,584	12,800	85,083	55,000

Table 4. CNN model confusion matrix on eight PQDs' test data.

Sag	4970		30					
Swell		5000						
Inter	22		4978					
Osc				5000				
Har					5000			
Har with Sag	1					4997	2	
Har with Swell					1	1	4998	
Flicker								5000
	Sag	Swell	Inter	Osc	Har	Har with Sag	Har with Swell	Flicker

Table 5. Comparison of this work with others.

Model	No. of PQDs	Table/Ref.	Test Acc.	Training Acc.	No. of Samples			
					Total	Training	Validation	Test
CNN-SPWVD	8	IV	99.86%	99.94%	200,000	147,200	12,800	40,000
CNN-SPWVD	11	V	99.26%	98.89%	1,741,667	1,616,584	85,083	40,000
ViT-SPWVD	11	VI	98.94%	98.73%	1,741,667	1,616,584	85,083	40,000
EMD + Balanced Neural Tree	8	[49]	97.90%	-	-	-	-	-
Hybrid ST + DT	11	[26]	94.36%	-	-	-	-	-
ST + NN + DT	13	[27]	99.90%	-	-	-	-	-
ADALINE + FNN	12	[50]	90.58	-	2400	1200	-	1200

Continued on next page

Model	No. of PQDs	Table/Ref.	Test Acc.	Training Acc.	No. of Samples			
					Total	Training	Validation	Test
CNN-WVD	9	[18]	99.67%	-	400	240	60	100
Deep CNN-1-D	16	[32]	99.96%	-	860,800	768,000	76,800	16,000
CNN + PSR	10	[42]	99.80%	100%	400	320	80	-
IPCA-1-D CNN	12	[51]	99.92%	-	2400	200	-	200
STBOACNN	14	[52]	99.20%	-	7000	3500	-	3500
Regulated 2-D-DCNN	14	[53]	99.97%	99.25%	11,000	9000	1000	1000
CNN-GRU-P	12	[54]	99.60%	-	36,000	28,800	3600	3600

Table 6 presents the confusion matrix of the CNN model on 11 PQDs, which displays remarkable performance in detecting six PQDs—*swell, harmonics, harmonics with sag, harmonics with swell, flicker,* and *flicker with swell*—with 100% accuracy. Further, the accuracy for the remaining PQDs, namely, *sag, interruption,* and *flicker with sag,* were classified with accuracy beyond 99.5%, except *spike* and *oscillatory transient,* which were classified with accuracies up to 94.4% and 98.14%, respectively.

Evidently from Table 6, the CNN could not detect *spike* and *oscillatory transient* PQDs accurately. The CNN model detected 280 occurrences of *spike* incorrectly and labeling them as *oscillatory transient*. Conversely, 93 *oscillatory transient* instances were mislabeled as *spike*. The CNN model achieved overall test accuracy of 99.26%.

Table 6. CNN model confusion matrix on 11 PQDs' test data.

Sag	4979		21								
Swell		5000									
Inter	2		4998								
Spike				4720	280						
Osc				93	4907						
Har						5000					
Har with Sag							5000				
Har with Swell								5000			
Flicker									5000		
Flicker with Sag	1		2							4988	9
Flicker with Swell											5000
	Sag	Swell	Inter	Spike	Osc	Har	Har with Sag	Har with Swell	Flicker	Flicker with Sag	Flicker with Swell

4.2. ViT model results

A further study was undertaken to assess the ViT model with the CNN model in classifying PQDs. The accuracies of the ViT model are presented in Table 7 and Figure 7. As the number of classes were incrementally raised, it became evident that the ViT required more data and a more complex model, as reflected in the increasing hyperparameters. Consequently, a second dataset containing a substantial amount of PQD images was employed.

In the initial phase, we evaluated a lightweight ViT model on a dataset containing 20,000 images per class for four PQDs—*swell*, *flicker*, *harmonics with swell*, and *flicker with sag*. Using a configuration with a hidden size of 16, MLP dimension of 16, and 12 encoder layers, the model achieved a high-test accuracy of 99.73%. However, when a fifth PQD class was added and the MLP dimension increased to 32 (to accommodate the added complexity), accuracy dropped to 92.81%, indicating that even a modest increase in class diversity could significantly affect model performance. When this same architecture was extended to the full set of 11 PQDs, accuracy dropped further to 85.61%, confirming that the lightweight model (hidden size = 16) was underpowered for more complex classification tasks, as shown in Table 7. To explore further, we also tested a much larger ViT model with hidden size = 768 and MLP dimension = 3072 using a significantly larger dataset (160,000 images per PQD). This configuration improved accuracy to 90.94%, but still underperformed compared to the final optimized configuration (hidden size = 128), indicating that model depth alone was not sufficient.

The work in [38] established a fact that to outperform a CNN, the ViT needs to be trained on an extensive amount of data such as the JFT-300M [55] dataset. Therefore, a second set of 11 PQDs, as listed in Table 3, was employed on the base variant of the ViT [38] with hyperparameters presented in Table 7. With the bigger size of the hyperparameters, a slower training rate, and a much higher volume of training data, the accuracy of 90.94% was achieved, revealing this as an underfit model. It is evident that the ViT with lower hyperparameters as compared to the base variant will be the best fit for this. So multiple combinations were tried and finally reached to correct combination of hyperparameters, listed in Table 8, that achieved significant improvement in accuracy. The iteration-wise accuracies for both hyperparameter sets are given in Table 8 and Figure 8.

Table 7. Initial test accuracies at different hyperparameters.

PQDs	Accuracy	No. of Heads	Hidden Layer	MLP Dim	Encoder Layer	No. of Images per PQD
4	99.73%	12	16	16	12	20,000
5	92.81%	12	16	32	12	20,000
11	85.61%	12	16	32	12	20,000
11	90.94%	12	768	3072	12	160,000

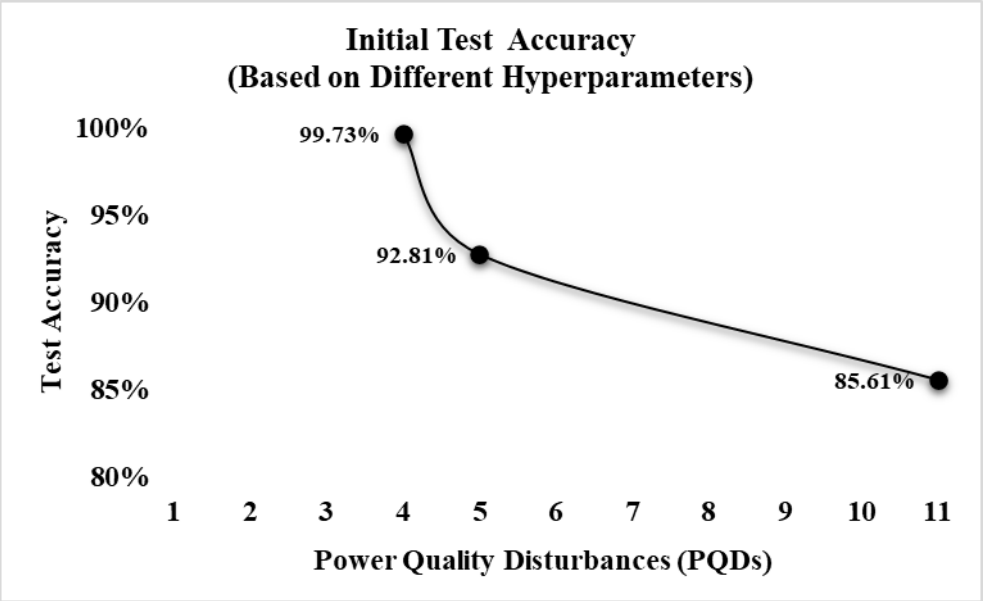

Figure 7. Initial test accuracies at different hyperparameters.

Table 8. Epoch-wise accuracy based on two different hyperparameters.

No. of PQDs	No. of Epoch	Training Acc.	Validation Acc.	No. of Heads	Hidden Layer	MLP Dim	Encoder Layer
11	1	50.39%	85.55%	12	768	3072	12
(160,000 images per PQD)	2	86.53%	88.69%				
	3	87.89%	88.90%				
	4	89.07%	89.93%				
	5	89.81%	90.51%				
	6	90.36%	90.70%				
	7	90.85%	90.94%				
	1	78.14%	95.14%	8	128	512	8
	2	95.89%	96.69%				
	3	96.99%	97.05%				
	4	97.73%	97.94%				
	5	98.07%	98.30%				
	6	98.31%	98.33%				
	7	98.57%	98.81%				

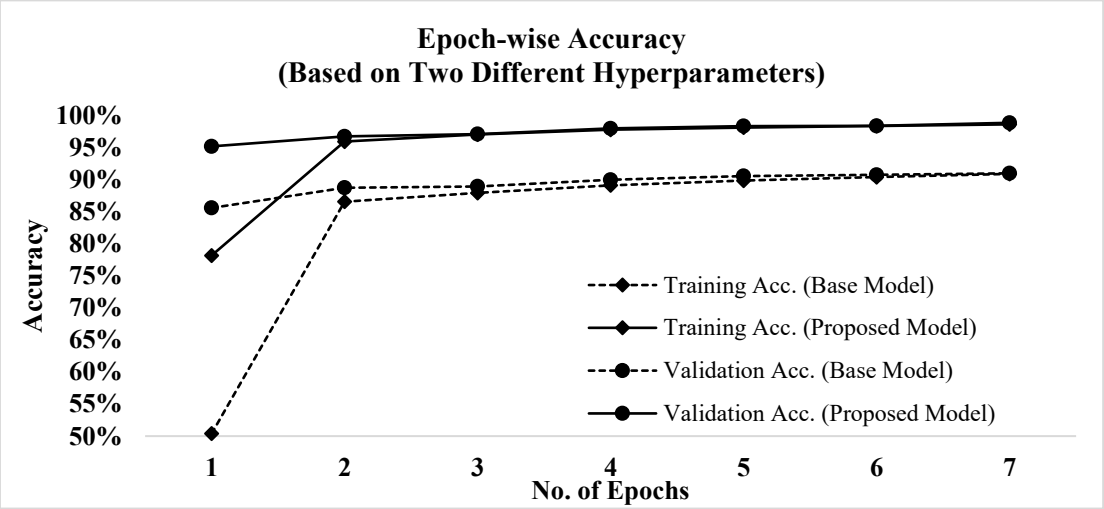

Figure 8. Epoch-wise accuracy based on two different hyperparameters.

4.2.1. Confusion matrix of 11 PQDs' test data

Table 9. Hyperparameters for ViT model training.

PQDs	Accuracy	No. of Heads	Hidden Layer	MLP Dim	Encoder Layer	No. of Images per PQD
11	98.94%	8	128	512	8	160,000

Table 10. ViT model confusion matrix of test data.

Sag	4954		44				1			1	
Swell		5000									
Inter	26	1	4971				1			1	
Spike				4760	240						
Osc				254	4746						
Har						5000					
Har with Sag							4997	1		2	
Har with Swell							2	4998			
Flicker									5000		
Flicker with Sag	1		1				3			4992	3
Flicker with Swell										1	4999
	Sag	Swell	Inter	Spike	Osc	Har	Har with Sag	Har with Swell	Flicker	Flicker with Sag	Flicker with Swell

The performance of the ViT model, based on the hyperparameters listed in Table 9, was evaluated by putting it through the same set of 11 PQDs leading to the generation of a confusion matrix, presented in Table 10. The ViT model showed remarkable performance in detecting three PQDs, that is, *swell, harmonics,* and *flicker,* which were detected with flawless 100% accuracy. The accuracy for the remaining eight PQDs could not attain perfection but remained at a satisfactory level, with the exception of *spike* and *oscillatory transient*. Notably, the ViT model could also not succeed in accurately labeling *spike* and *oscillatory transient* events and confused them. The overall test accuracy attained by the ViT model was 98.94%.

4.3. Comparison of the proposed method with other works

The comparative analysis of the CNN-SPWVD was made with various existing techniques, involving the CNN as the main classifier, presented in Table 5. The CNN-SPWVD exhibited overall higher test accuracy against a few approaches, namely, ViT-SPWVD, EMD–balanced neural tree, hybrid ST-DT, CNN-WVD, STBOACNN, and CNN-GRU-P. It displayed slightly lower accuracy than other methods including ST-NN-DT, deep CNN-1-D, IPCA 1-D CNN, and regulated 2-D-DCNN. This disparity is due to the distinctive nature of the compared methods and the size of the data per PQD class. In this study, the size is almost 2 million samples for 11 PQD classes, which is the greatest amongst all.

Moreover, traditional methods that depend on manually crafted feature extraction processes are primarily tailored for processing 1-D data with constrained dataset sizes. Their drawback lies in their susceptibility to accurately detect PQDs with diverse characteristics. In contrast, the model in this study employs automatic feature extraction on diverse 2-D data, as opposed to the limited scope of 1-D data. This innovative approach demonstrates remarkable proficiency in detecting a broader spectrum of PQDs when compared to alternative methods, listed in Table 5, emphasizing its superior capabilities.

The performance of deep learning models is based on different factors such as the architecture of the model, complexities of the data, optimal values of the hyperparameters, and quality and quantity of the data for the training model. The vision transformer model implementation in the PQD domain has demonstrated its immense potential for various image classification tasks. Additionally, the CNN has been the state-of-the-art technique for image classification tasks and is still ruling. The CNN requires low computational resources and less data, and works on locally acquire features. It is a well-known fact that the ViT performs outstandingly on larger datasets while the CNN performs remarkably when the dataset is smaller. This happens because the CNN has strong inductive bias [56] and captures local patterns and spatial hierarchies in images effectively. The ViT possesses weak image specific inductive bias as it works on globally acquired features. In this study, rigorous experiments have shown that the ViT requires a lengthier duration for training and evaluating PQDs as compared to the CNN. In this work, the ViT achieved the results in a time span which was 28-times more than the CNN for the same data size. Therefore, the findings of this work empirically establish the superiority of convolutional neural networks (CNNs) over vision transformer (ViT) methods for the classification of PQDs.

Moreover, while working on smaller datasets, locally capturing features is crucial for better performance while globally learned patterns may be difficult to find from scratch. Therefore, inductive bias of the CNN is beneficial when the dataset is smaller, but when the dataset increases, a direct learning approach is useful and utilizes weak inductive bias. It was observed that when the ViT-*Large*

model was pre-trained on the ImageNet dataset, a small dataset in the context of a ViT, it underperformed as compared to the ViT-Base model [38]. The same observation was also made while carrying out the experiments in this work. Therefore, the hyperparameters used in this work are of lower value as compare to the ViT-Base model because, with a smaller dataset, the model does not generalize well and tends to overfit. Tables 7 and 9 clearly present this fact that the model achieved 90.94% accuracy when it was trained on the ViT-Base model and showed significant improvement by achieving 98.94% with the hyperparameters listed in Table 9. Epoch-wise training and validation accuracy are also presented in Table 8 and Figure 8. The hyperparameters used in this work were achieved after trying multiple combinations and reached a final hyperparameter set by reducing the hidden layer and MLP dimension to one-sixth while the number of heads and encoder layer were reduced to two-thirds. The model achieved remarkable test accuracy of 98.94%.

5. Conclusions

This study introduces a novel ViT-SPWVD framework for classifying power quality disturbances (PQDs), combining the high-resolution, time-frequency analysis capabilities of the smoothed pseudo Wigner-Ville distribution (SPWVD) with the global feature learning power of vision transformers (ViTs). Synthetic PQD signals were generated in compliance with the IEEE 1159 standard and transformed into 2-D time-frequency images for classification.

To the best of our knowledge, this is the first application of a ViT in PQD image classification. The proposed model achieved a test accuracy of 98.94%, demonstrating strong performance and competitive results compared to a CNN baseline (99.26%). The ViT's ability to model long-range dependencies and global features presents a promising direction for intelligent monitoring in power systems.

This work establishes a foundation for further exploration of transformer-based architectures in the PQD domain. Although the current study used synthetic, noiseless data and focused on base model comparisons, a concrete cloud-edge deployment framework has been proposed to illustrate practical integration. Future extensions will focus on robustness under low signal-to-noise ratio (SNR) and noise conditions, testing on public and field-acquired datasets, and exploring temporal transformers (e.g., informer, autoformer) for raw signal classification. Additionally, optimized ViT variants such as the DeiT and Swin transformers will be considered to enhance computational efficiency and edge-device compatibility.

Overall, this research opens new opportunities for deploying deep transformer models in smart-grid environments and lays the groundwork for advancing real-time, data-driven power quality assessment.

Use of AI tools declaration

The authors declare that no generative AI tools were used in the generation of data, analysis, or interpretation of results in this study. However, AI-assisted tools (such as ChatGPT and Grammarly) were used only for language refinement and grammar improvement during manuscript preparation. All intellectual content, results, and conclusions are the authors' own responsibility.

Acknowledgments

No funding was received for this study.

Conflict of interests

The authors declare no conflict of interests.

Author contributions

Muhammad Hassan Anwar: Conceptualization, Methodology, Software, Validation, Investigation, Writing — Original Draft, Visualization, Writing — Review & Editing;

Mirza Muhammad Ali Baig: Conceptualization, Methodology, Investigation, Resources, Writing — Original Draft, Supervision, Project Administration, Writing — Review & Editing;

Abdurrahman Javid Shaikh: Software, Visualization, Resources, Writing — Original Draft, Writing — Review & Editing, Supervision, Project Administration;

Abdul Ghani Abro: Data Curation, Writing — Original Draft, Writing — Review & Editing, Project Administration and Supervision.

References

1. Khalid MR, Alam MS, Sarwar A, et al. (2019) A comprehensive review on electric vehicles charging infrastructures and their impacts on power-quality of the utility grid. *eTransp* 1: 100006. https://doi.org/10.1016/j.etran.2019.100006
2. IEEE (2019) IEEE recommended practice for monitoring electric power quality. IEEE Std 1159-2019 (Revision of IEEE Std 1159-2009), 1–98. https://doi.org/10.1109/IEEESTD.2019.8796486
3. IEC (2015) Electromagnetic compatibility (EMC)-Part 4-30: Testing and measurement techniques-power quality measurement methods. IEC 61000-4-30:2003. Available from: https://webstore.iec.ch/en/publication/21844.
4. CENELEC (2022) Voltage characteristics of electricity supplied by public electricity networks. EN 50160:2022. Available from: https://standards.cencenelec.eu/dyn/www/f?p=CENELEC:110:::::FSP_PROJECT,FSP_ORG_ID:71003,1258595&cs=177F89A233554A3CA651BC5AAA21C3EB3.
5. Zhang Y, Zhang L, Liu B, et al. (2024) Voltage sag sensitive load type identification based on power quality monitoring data. *Int J Electr Power Energy Syst* 158: 109936. https://doi.org/10.1016/j.ijepes.2024.109936
6. Liao J, Liu Y, Guo C, et al. (2024) Power quality of DC microgrid: Index classification, definition, correlation analysis and cases study. *Int J Electr Power Energy Syst* 156: 109782. https://doi.org/10.1016/j.ijepes.2024.109782
7. Rezapour H, Fathnia F, Fiuzy M, et al. (2024) Enhancing power quality and loss optimization in distorted distribution networks utilizing capacitors and active power filters: A simultaneous approach. *Int J Electr Power Energy Syst* 155: 109590. https://doi.org/10.1016/j.ijepes.2023.109590

8. Mishra M (2019) Power quality disturbance detection and classification using signal processing and soft computing techniques: A comprehensive review. *Int Trans Electr Energy Syst* 29: e12008. https://doi.org/10.1002/2050-7038.12008
9. Zhong T, Zhang S, Cai G, et al. (2019) Power quality disturbance recognition based on multiresolution S-transform and decision tree. *IEEE Access* 7: 88380–88392. https://doi.org/10.1109/ACCESS.2019.2924918
10. Heydt GT, Fjeld PS, Liu CC, et al. (1999) Applications of the windowed FFT to electric power quality assessment. *IEEE Trans Power Delivery* 14: 1411–1416. https://doi.org/10.1109/61.796235
11. Gu YH, Bollen MHJ (2000) Time-frequency and time-scale domain analysis of voltage disturbances. *IEEE Trans Power Delivery* 15: 1279–1284. https://doi.org/10.1109/61.891515
12. Gaouda AM, Salama MMA, Sultan MR, et al. (1999) Power quality detection and classification using wavelet-multiresolution signal decomposition. *IEEE Trans Power Delivery* 14: 1469–1476. https://doi.org/10.1109/61.796242
13. Gao W, Ning J (2011) Wavelet-based disturbance analysis for power system wide-area monitoring. *IEEE Trans Smart Grid* 2: 121–130. https://doi.org/10.1109/TSG.2011.2106521
14. Ray PK, Kishor N, Mohanty, SR (2012) Islanding and power quality disturbance detection in grid-connected hybrid power system using wavelet and S-transform. *IEEE Trans Smart Grid* 3: 1082–1094. https://doi.org/10.1109/TSG.2012.2197642
15. Wang H, Wang P, Liu T (2017) Power quality disturbance classification using the S-transform and probabilistic neural network. *Energies* 10: 107. https://doi.org/10.3390/en10010107
16. Climente-Alarcon V, Antonino-Daviu JA, Haavisto A, et al. (2015) Diagnosis of induction motors under varying speed operation by principal slot harmonic tracking. *IEEE Trans Ind Appl* 51: 3591–3599. https://doi.org/10.1109/TIA.2015.2413963
17. Huang NE, Shen Z, Long SR, et al. (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc: Math Phys Eng Sci* 454: 903–995. Available from: http://www.jstor.org/stable/53161.
18. Cai K, Cao W, Aarniovuori L, et al. (2019) Classification of power quality disturbances using Wigner-Ville distribution and deep convolutional neural networks. *IEEE Access* 7: 119099–119109. https://doi.org/10.1109/ACCESS.2019.2937193
19. Wu X, Liu T (2019) Spectral decomposition of seismic data with reassigned smoothed pseudo Wigner-Ville distribution. *J Appl Geophys* 68: 386–393. https://doi.org/10.1016/j.jappgeo.2009.03.004
20. Cohen L (1995) Time-frequency analysis, Prentice Hall signal processing series, Englewood Cliffs, New Jersey: Prentice Hall.
21. Mallat S (1999) A wavelet tour of signal processing, 2nd Eds., San Diego, CA: Academic Press. https://doi.org/10.1016/B978-0-12-466606-1.X5000-4
22. Uyar M, Yildirim S, Gencoglu MT (2008) An effective wavelet-based feature extraction method for classification of power quality disturbance signals. *Electr Power Syst Res* 78: 1747–1755. https://doi.org/10.1016/j.epsr.2008.03.002
23. Stockwell RG, Mansinha L, PLowe R (2002) Localization of the complex spectrum: The S transform. *IEEE Trans Signal Process* 44: 998–1001. https://doi.org/10.1109/78.492555

24. Szmajda M, Górecki K, Mroczka J (2010) Gabor transform, SPWVD, Gabor-Wigner transform and Wavelet transform—Tools for power quality monitoring. *Metrol Meas Syst* XVII: 383–396. Available from: http://www.metrology.pg.gda.pl/full/2010/M&MS_2010_383.pdf.
25. Moravej Z, Abdoos AA, Pazoki M (2009) Detection and classification of power quality disturbances using wavelet transform and support vector machines. *Electr Power Compon Syst* 38: 182–196. https://doi.org/10.1080/15325000903273387
26. Biswal M, Dash PK (2013) Detection and characterization of multiple power quality disturbances with a fast S-transform and decision tree based classifier. *Digital Signal Process* 23: 1071–1083. https://doi.org/10.1016/j.dsp.2013.02.012
27. Kumar R, Singh B, Shahani DT, et al. (2015) Recognition of power-quality disturbances using S-transform-based ANN classifier and rule-based decision tree. *IEEE Trans Ind Appl* 51: 1249–1258. https://doi.org/10.1109/TIA.2014.2356639
28. Khokhar S, Zin AAM, Memon AP, et al. (2017) A new optimal feature selection algorithm for classification of power quality disturbances using discrete wavelet transform and probabilistic neural network. *Measurement* 95: 246–259. https://doi.org/10.1016/j.measurement.2016.10.013
29. Cervantes J, Lamont FG, López-Chau A, et al. (2015) Data selection based on decision tree for SVM classification on large data sets. *Appl Soft Comput* 37: 787–798. https://doi.org/10.1016/j.asoc.2015.08.048
30. Ackerman A, Sen PK, Oertli C (2013) Designing safe and reliable grounding in AC substations with poor soil resistivity: An interpretation of IEEE Std. 80. *IEEE Trans Ind Appl* 49: 1883–1889. https://doi.org/10.1109/TIA.2013.2256092
31. Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. *Commun ACM* 60: 84–90. https://doi.org/10.1145/3065386
32. Wang S, Chen H (2019) A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network. *Appl Energy* 235: 1126–1140. https://doi.org/10.1016/j.apenergy.2018.09.160
33. Sindi H, Nour M, Rawa M, et al. (2021) A novel hybrid deep learning approach including combination of 1D power signals and 2D signal images for power quality disturbance classification. *Expert Syst Appl* 174: 114785. https://doi.org/10.1016/j.eswa.2021.114785
34. Garcia CI, Grasso F, Luchetta A, et al. (2020) A comparison of power quality disturbance detection and classification methods using CNN, LSTM and CNN-LSTM. *Appl Sci* 10: 6755. https://doi.org/10.3390/app10196755
35. Li C, Huang X, Song R, et al. (2022) EEG-based seizure prediction via transformer guided CNN. *Measurement* 203: 111948. https://doi.org/10.1016/j.measurement.2022.111948
36. Han Z, Liu H, Liu Z, et al. (2019) 3D2SeqViews: Aggregating sequential views for 3D global feature learning by CNN with hierarchical attention aggregation. *IEEE Trans Image Process* 28: 3986–3999. https://doi.org/10.1109/TIP.2019.2904460
37. Vaswani A, Shazeer N, Parmar N, et al. (2017) Attention is all you need. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 6000–6010. Available from: https://dl.acm.org/doi/10.5555/3295222.3295349#core-history.
38. Dosovitskiy A, Beyer L, Kolesnikov A, et al. (2021) An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations*. Available from: https://openreview.net/forum?id=YicbFdNTTy.

39. Feldman M (2001) Hilbert transforms, In: S. Braun (Ed.) Encyclopedia of Vibration, Oxford: Elsevier, 642–648. https://doi.org/10.1006/rwvb.2001.0057
40. Lin YP, Vaidyanathan PP (1998) A Kaiser window approach for the design of prototype filters of cosine modulated filterbanks. *IEEE Signal Process Letters* 5: 132–134. https://doi.org/10.1109/97.681427
41. Wang Q, Li B, Xiao T, et al. (2019) Learning deep transformer models for machine translation. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics,* 1810–1822. Available from: https://aclanthology.org/P19-1176/.
42. Cai K, Hu T, Cao W, et al. (2019) Classifying power quality disturbances based on phase space reconstruction and a convolutional neural network. *Appl Sci* 9: 3681. https://doi.org/10.3390/app9183681
43. Wang J, Zhang D, Zhou Y (2022) Ensemble deep learning for automated classification of power quality disturbances signals. *Electr Power Syst Res* 213: 108695. https://doi.org/10.1016/j.epsr.2022.108695
44. Liu Y, Jin T, Mohamed MA, et al. (2021) A novel three-step classification approach based on time-dependent spectral features for complex power quality disturbances. *IEEE Trans Instrum Meas* 70: 1–14. https://doi.org/10.1109/TIM.2021.3050187
45. Igual R, Medrano C, Arcega FJ, et al. (2018) Integral mathematical model of power quality disturbances. *2018 18th International Conference on Harmonics and Quality of Power (ICHQP)*, 1–6. https://doi.org/10.1109/ICHQP.2018.8378902
46. Hooshmand R, Enshaee A (2010) Detection and classification of single and combined power quality disturbances using fuzzy systems oriented by particle swarm optimization algorithm. *Electr Power Syst Res* 80: 1552–1561. https://doi.org/10.1016/j.epsr.2010.07.001
47. Wang J, Xu Z, Che Y (2019) Power quality disturbance classification based on compressed sensing and deep convolution neural networks. *IEEE Access* 7: 78336–78346. https://doi.org/10.1109/ACCESS.2019.2922367
48. Srivastava N, Hinton G, Krizhevsky A, et al. (2014) Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 15: 1929–1958. Available from: https://jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf.
49. Biswal B, Biswal M, Mishra S, et al. (2013) Automatic classification of power quality events using balanced neural tree. *IEEE Trans Ind Electron* 61: 521–530. https://doi.org/10.1109/TIE.2013.2248335
50. Valtierra-Rodriguez M, Romero-Troncoso RJ, Osornio-Rios RA, et al. (2014) Detection and classification of single and combined power quality disturbances using neural networks. *IEEE Trans Ind Electron* 5: 2473–2482. https://doi.org/10.1109/TIE.2013.2272276
51. Shen Y, Abubakar M, Liu H, et al. (2019) Power quality disturbance monitoring and classification based on improved PCA and convolution neural network for wind-grid distribution systems. *Energies* 12: 1280. https://doi.org/10.3390/en12071280
52. Eristi B, Eristi H (2022) A new deep learning method for the classification of power quality disturbances in hybrid power system. *Electr Eng* 104: 3753–3768. https://doi.org/10.1007/s00202-022-01581-w
53. Chen CI, Berutu SS, Chen YC, et al. (2022) Regulated two-dimensional deep convolutional neural network-based power quality classifier for microgrid. *Energies* 15: 2532. https://doi.org/10.3390/en15072532

54. Cai J, Zhang K, Jiang H (2023) Power quality disturbance classification based on parallel fusion of CNN and GRU. *Energies* 16: 4029. https://doi.org/10.3390/en16104029
55. Sun C, Shrivastava A, Singh S, et al. (2017) Revisiting unreasonable effectiveness of data in deep learning era. *2017 IEEE International Conference on Computer Vision (ICCV)*, 843–852. https://doi.org/10.1109/ICCV.2017.97
56. Utgoff PE (1986) Machine learning of inductive bias, New York: Springer. https://doi.org/10.1007/978-1-4613-2283-2