CARAVEL

Publish open access at Caravel Press

Global Decarbonisation

Journal homepage: globaldecarbonisation.com

Technical article

A Deep Learning Approach to Power Quality Monitoring for Energy Efficiency and Carbon Reduction

Kevser İrem Danaci [1]* and Sıtkı Akkaya [1]

[1] *Department of Electrical Electronics Eng., Sivas University of Science and Technology, Sivas, Turkey.*

ABSTRACT

A classification system using deep learning for detecting power quality disturbances (PQDs), which are commonly experienced in electric power systems, has been proposed in this study. Because reliable and efficient classification methods are becoming increasingly critical as new forms of energy are integrated into the electric grid and as smart grid technology advances, accurate detection of PQDs will significantly contribute to the stability, efficiency, and robustness of current electric power systems. The SEED power quality disturbance dataset has been used for analysis; it contains 17 different types of disturbances, including pure sinusoidal signals. Both the raw 1D time domain signals and the 2D time frequency representations have been used to train the models (the time-frequency representations were created with the Short Term Fourier Transform, or STFT). The models used were convolutional neural network (CNN) architectures adapted for both 1D and 2D input data; both types of architectures were compared using accuracy, precision, and recall as measures of performance. The experimental results of the study show that for test accuracy, the 1D CNN model performed better than the 2D CNN model (i.e., 95% versus 82%, respectively). This research differs from most previous work because it simultaneously examines a common set of time domain and time-frequency feature representations while keeping all other test conditions constant. Results suggest that temporal representations of raw PQ signals may provide a classification advantage over time-frequency representations when used for PQD classification. Such benefits associated with using temporal representations will aid in the design of effective and accurate PQ monitoring systems, which, together with improved efficiencies from energy conservation, can help contribute towards global carbon emission reduction objectives by both reducing and eliminating wasted energy based on accurate and timely detection of power quality problems using intelligent monitoring techniques in modern smart grids.

ARTICLE INFO

Article History:
Received: 16 February 2026
Revised: 09 March 2026
Accepted: 31 March 2026

Keywords:
Power Quality Disturbance Detection
Deep Learning
Carbon Emissions
AI for Energy Efficiency
Signal Processing in Power Systems
Smart Grid
Neural Networks

Article Citation:
Danaci, K. İrem, & Akkaya, S. (2026). A Deep Learning Approach to Power Quality Monitoring for Energy Efficiency and Carbon Reduction. Global Decarbonisation, 2(1), 92–119. https://doi.org/10.65582/gd.2026.007

1. INTRODUCTION

As worldwide energy needs increase, carbon dioxide emissions are on the rise. According to International Energy Agency (IEA) statistics, energy-related CO_2 emissions globally increased 1.1% in 2023 and 0.8% in 2024 (International Energy Agency (IEA) 2024), with total global CO_2 emissions reaching 37.8 gigatonnes

* Corresponding author. Email address: kiremdanaci@sivas.edu.tr (Kevser İrem Danaci)

(Gt). A major cause of this continual growth is associated with the electricity generation and distribution industries (Ding et al., 2025; Jiang and Mao 2025). Smart grids are the newest types of secure and sustainable systems for supporting the growth of electricity generation, transmitting electricity, and distributing electricity while decreasing CO_2 emissions and increasing efficiency. As smart grids combine renewable energy generation with traditional electricity generating resources, wind and solar energy combined are substantial contributors to reducing CO_2 emissions (Wang, and Chen, 2019; Sekar et al., 2022).

Smart grids significantly enhance energy efficiency by enabling real-time monitoring of the quality and efficiency of transmitted electrical energy. The main factors affecting energy efficiency in this context are power quality disturbances (PQDs) such as voltage sags, swells, harmonics, flickers, and transient currents, which can cause energy losses, equipment damage, and system inefficiencies. PQDs not only harm consumer devices but also degrade overall system performance. It is also expected that as more energy systems are integrated, the PQDs will become worse throughout all stages of energy generation, conversion, transmission, and end-use (Wang and Chen 2019). Therefore, early detection and accurate classification of such PQDs are of great importance (Khetarpal and Tripathi 2020).

Moreover, under the concept of grid decarbonisation, PQD still plays an indirect yet significant role in the overall carbon footprint of transmitting and generating electricity. High technical loss due to low-quality electricity decreases the efficiency of devices accountable for consuming electricity, and an earlier depreciation of electricity assets will contribute negatively to carbon emissions associated with electricity generation and transmission systems. Continually and accurately monitoring PQD can help improve the efficiency of our electric grid as a whole, especially with more renewable resources being integrated into today's electricity generation and transmission systems. PQD will be critical to ensuring a long-term future as a low-carbon and sustainable one.

Deep learning (DL) models have shown significant success in solving many energy-related challenges in recent times. Convolutional neural networks (CNNs) are commonly used for classifying signals and recognising patterns, with many successful outcomes (Lecun, Bengio and Hinton, 2015). DL models provide the advantages of automatically extracting and learning features from raw input data, allowing the identification and classification of PQDs (Garcia et al., 2020; Albalooshi and Qader, 2025) more accurately and at an earlier stage than what would typically be possible. This enables the development of fully automated means for monitoring all types of disturbances, thus indirectly enhancing energy efficiency and promoting the creation of smarter energy management systems. Accurate and timely identification of PQDs leads to both economic advantages and reductions in energy used and greenhouse gases emitted by using less electricity.

Detection and classification of PQD is crucial to ensuring the reliability and efficiency of power systems sufficient to meet customers (Albalooshi & Qader 2025). The literature also shows that deep learning methods are becoming increasingly meaningful when addressing PQD detection and classification. Different architectures such as CNN, recurrent neural networks (RNN), long short term memory (LSTM) and Autoencoders have all been used to detect and classify this problem. A brief review of current research on the classification of PQD can be seen below.

Ma et al. (Ma et al. 2017) presented a model based on a Stacked Autoencoder (SAE), that can automatically extract high level features, and achieved promising performance in PQD classification. A classification of seven different PQD types with SAE models integrated with particle swarm optimization (PSO) assisted classification schemes has obtained high accuracy and robustness against noise. Khetarpal et al. (Khetarpal et al. 2023) provided a comparison between various types of deep convolutional autoencoders (DCAEs). As part of their study, they also provided a method for using the results of a DCAE combined with Gabor features (Gabor filters) as an input for classifying PQDs into 15 different types. They demonstrated that the DCAE plus Gabor-filter features had good classification accuracy and could withstand the effects of noise quite well.

Using an integrated framework to combine the traditional stages of signal processing, feature extraction, and classification into one architecture has provided improvements in PQD classification through deep CNNs with supervised training. Improved PQD classification with deep CNNs has been shown to produce better accuracy and operational reliability of multienergy systems (Singh, 2009). According to Topaloglu (Topaloglu, 2023), the new deep CNN architecture was enhanced by including an attention model. A second, unique CNN model uses synthetic image datasets for 9 different classes of PQD. Each input image is processed; 256×256 pixels are used for each input image over five hidden layers, producing spatial and structural features of the image, and the model utilises an attention network to help evaluate key locations within the input images to produce higher weighted outputs. The findings suggest that the combination of an attention-based CNN architecture produces improved advantages such as better accuracy, robustness to noise, and faster responses to PQDs than other types of detection systems. Wang et al. (Wang and Chen 2019) developed a closed-loop method based on a 1-dimensional deep CNN. 16 different single and compound PQD types were synthetically generated. The results revealed that the proposed CNN outperformed other deep learning methods (ResNet50, LSTM, GRU, SAE) and traditional methods in terms of both noise immunity and accuracy. Wang et al. (Wang, Xu and Che 2019) developed a method based on compression sensing (CS) and DCNN. The 1D-DCNN model performs on compressed data, thereby avoiding the necessity for preprocessing or making use of manual feature extraction techniques. A total of fifteen different PQD types were simulated to produce classification accuracies that exceeded those possible with conventional techniques such as SVM. Albalooshi et al. (Albalooshi and Qader, 2025) developed a 1D CNN with an integrated channel-attention mechanism. Due to the limited number of labelled PQD datasets in the literature, the aforementioned study employed an open-source dataset generator to create two synthetic datasets, each comprising 50,000 signals: the first containing Gaussian noise (SNR levels of 20–50 dB) and the second not. All datasets contain 10 defined PQD types (normal, sag, swell, harmonics, etc.) with 5000 records per type, following IEEE standards for the classification of PQ disturbance signals, including IEEE-1159. The proposed CNN model is compared to a baseline CNN as well as ResNet50, BiLSTM, and GRU using a confusion matrix and classification accuracy for both datasets (noisy and noise-free). Results demonstrate that this model is more robust to noise than other competing systems and performs better than existing methods.

Time frequency domain methods have also been widely explored. Yigit et al. (Yigit et al. 2021) proposed a GRU (Gated Recurrent Unit) supported CNN architecture that uses time frequency data obtained via the Short-Time Fourier Transform (STFT). By incorporating GRU layers instead of fully connected layers, the proposed model achieves high accuracy with fewer parameters and demonstrates high performance even with noisy signals. In this study, seven different single and compound PQD types were generated in a simulation environment. Channa et al. (Channa and Li 2023) proposed that features extracted via STFT and fed into 2D CNN models have enabled the successful classification of both single and composite PQDs, as demonstrated in simulations on the IEEE 13 node system. Baig et al. (Baig et al. 2024) converted 1D signals into time-frequency images using Continuous Wavelet Transform (CWT), which were processed using a transfer learning approach with four different pre-trained deep convolutional neural networks (ResNet-50, VGG-16, AlexNet, and SqueezeNet). The output probabilities of each model were combined using a method to create an ensemble classifier, thus improving the generalization ability of individual models. Perez-Anaya et al. (Perez-Anaya et al., 2024) proposed a CWT and CNN-based method. Both synthetic and real data were used in the study. Despite the increased noise level, the proposed method showed high classification success.

Hybrid deep learning models have been created to provide a comprehensive understanding of PQD signals through the analysis of spatial and temporal features. Sekar et al. (Sekar et al., 2022) addressed 16 different scenarios of PQDs, including 10 single disturbances and 6 multiple disturbances, generating a total of 128,000 data sets. The combination of 1D CNNs (to capture spatial characteristics of PQDs) with LSTM (to interpret the time sequence) makes up the structure of their hybrid model. Cai et al. (Cai, Zhang, and Jiang 2023) also created an innovative parallel deep learning network known as CNN-GRU-P. The model simultaneously utilizes a CNN with a squeeze and excitation block (SE) to extract short-term features and a GRU network with an attention mechanism to extract long-term features. The results demonstrate that the model provides high accuracy for both single and compound disturbances and is robust to noise. Bai et al. (Bai et al. 2025) proposed S-transformation (FST) and an improved CNN–LSTM hybrid model. The model's hyperparameters were optimized using the Sparrow Search Algorithm (SSA). The method provided high accuracy for both single-type and mixed disturbances. Chiam et al. (Chiam, Lim, and Law 2023) developed a hybrid model based on WT, attention mechanism, and LSTM. They used 16 different synthetic PQD classes. Signals are separated into frequency bands using a four-level multiresolution decomposition, and then important features are extracted using spatial or temporal attention mechanisms. The resulting features are processed with LSTM layers. The model was found to be more robust to noise compared to deep CNNs and other benchmark models. Garcia et al. (Garcia et al. 2020) compared the performance of LSTM, CNN, and CNN-LSTM. A MATLAB/Simulink model of a microgrid was used to simulate voltage and current signals with six PQD types, such as sag, swell, harmonics, transient, notch, and interruption. The LSTM model is designed for time-series data, while the CNN model processes signals as images via STFT to capture spectral features. After that, they combined a CNN for feature extraction with an LSTM for temporal modelling. The study demonstrates that the hybrid model's ability to combine CNN's feature extraction with LSTM's temporal modelling makes it superior, especially for challenging transients.

In recent years, transformer-based and attention-driven architectures have gained significant interest in time-series analysis due to their capability to model long-range temporal dependencies more effectively than conventional convolutional and recurrent networks. Unlike CNNs, which primarily capture local patterns, attention mechanisms enable adaptive weighting of informative time segments, making them suitable for complex and non-stationary signals. Zerveas et al. (Zerveas et al. 2021) introduced a transformer-based framework for multivariate time-series representation learning and demonstrated that self-attention can effectively capture temporal dependencies across multiple variables. Wen et al. (Wen et al., 2023) provided a comprehensive survey of transformer models in time-series applications, highlighting their growing adoption and advantages in handling long-term dependencies. They also identified several efficient attention mechanisms developed by Zhou et al. (Zhou et al., 2021) that significantly reduce the computational complexity of traditional transformers when used with long sequences. These efficient attention networks are key advancements for enabling the successful application of transformer-based models to large amounts of temporal data.

More recently, attention-based deep learning models have been utilised for time-series classification projects, allowing for gains in their performance on datasets containing overlapping patterns or in the case of nonstationarity. As regards the field of power systems, previous studies have shown that transformer/attention enhanced models produce improved disturbance recognition accuracy by focusing on vital temporal characteristics; however, there is usually an associated increase in difficulty with computing requirements. These studies indicate that there is potential for transformer/attention-based models, while lightweight CNN-based models also offer advantages to users concerned with optimising monitoring operations from both an energy-efficient and real-time perspective (Wang and Chen 2019; Lim et al. 2021; Wen et al. 2023).

Most existing research on classifying PQD typically uses only one form of input signal (i.e., raw time domain or time-frequency). This lack of diversity can lead to unclear results when comparing different types of signal representations as to which one is more advantageous under comparable test conditions. This study focuses on comparing the effectiveness of raw (one-dimensional) and time-frequency (two-dimensional) representations by analysing the classification performance of both signals using a comprehensive dataset containing multiple PQDs. This research uses both types of representations in a custom-designed 1D and 2D CNN; although the results from prior studies appear promising, the majority of current studies are limited to either simulated data, address only a select few degradation forms, or utilise only one type of representation (time domain or time-frequency). There are a limited number of studies that have comprehensively compared the performances of both the 1D and 2D CNN architectures with the same raw and transformed signal inputs under the same experimental scenario. The importance of bridging this gap is that improved PQD product detections allow for greater energy efficiency through less wasted energy and quicker responses for corrective actions that also contribute to a reduced environmental footprint from operating a power system more sustainably.

Despite many studies performed on PQD classification with the use of deep learning methods, the key contribution of this research lies in establishing a comparison between existing techniques rather than introducing an additional network architecture. In our work, we evaluate the standard time-domain raw signals and the STFT-based time–frequency representations using the same dataset and identical data splits while using similar CNN architectures. This provided an opportunity for a fair and comparable evaluation, which has often not been achieved in prior studies. The results have provided an indication from a practical perspective regarding the performance of raw waveforms vs time–frequency images for the purpose of smart grid monitoring for the SEED PQD dataset (SEED - Power Quality Disturbance Dataset 2023; Khan, Aziz, and Usman 2023). Based on our findings, raw waveforms provide a much higher level of classification accuracy than do time–frequency images, indicating that it is possible to achieve reliable PQD detection with much simpler preprocessing and that this is particularly useful for real-time and energy-efficient monitoring systems.

The contributions of the study are as follows:

- This study presents a direct and systematic comparison of 1D and 2D CNN architectures for PQD classification under the same experimental conditions.
- Both raw time domain signals and STFT-derived time-frequency representations are integrated into the classification framework, providing an in-depth performance comparison between temporal and time-frequency feature sets.

- The SEED Dataset has many real-world scenarios that include 17 different distortions, and it helps to provide a range of real-world scenarios and improve generalisability.
- Unlike most existing approaches that focus solely on simulated data, process a limited number of distortion types, or rely solely on a single data representation, this study fills a critical gap by jointly evaluating multiple distortion types, data representations, and CNN input structures in a unified experimental setup.
- According to the experimental results, the classification accuracy of PQD is significantly enhanced by using raw signals with 1D CNN models compared to the classification accuracy of PQD when using 2D CNN models based on STFT-based spectrograms. This supports the hypothesis that 1D CNN models learn discriminative features in time that can be used to distinguish between different types of PQD more effectively than the temporal and frequency features provided by STFT-based spectrograms.
- Implementing this new way to detect PQD allows utilities to have accurate information so they can act quickly to minimise energy losses and reduce carbon emissions by operating their power systems more sustainably.

The remainder of this paper is organised as follows: Section 2 provides an overview of PQD and DL. Section 3 presents the methods used in the study. Finally, Section 4 covers the obtained results.

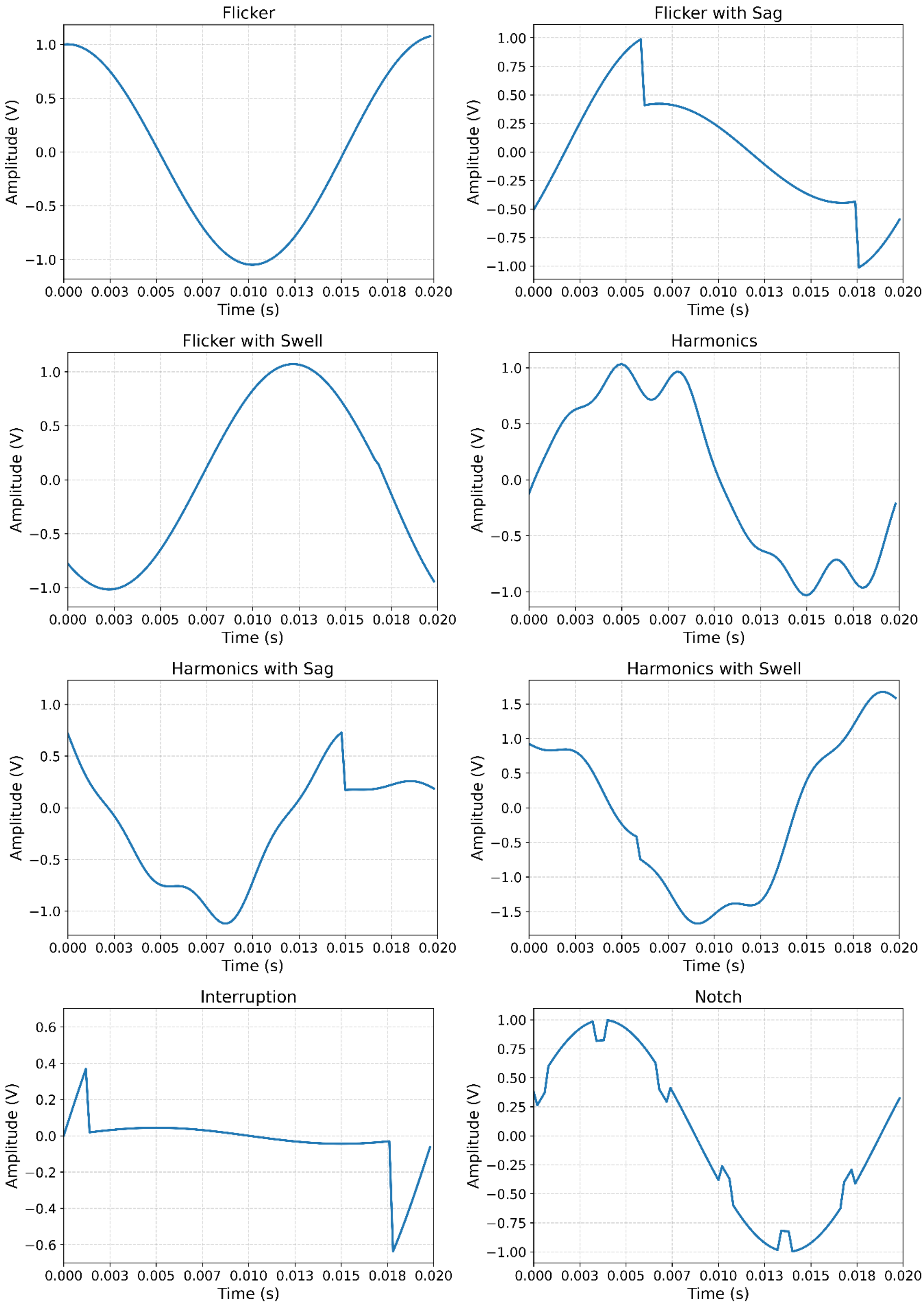
2. MATERIALS AND METHODS

This section provides a theoretical and methodological background for the study. It first introduces the concept of PQDs and their impact on energy systems. Subsequently, conventional detection and classification techniques are discussed, followed by an overview of DL approaches applied in this domain. The section also explains the use of STFT for time-frequency analysis and presents the characteristics of the SEED PQD dataset used in this study.

2.1. Power Quality Disturbances

An ideal voltage or current waveform is a single-frequency sinusoidal wave with constant amplitude and frequency. However, when the voltage or current deviates from this ideal form, a PQD occurs (Bollen 2003). These deviations can arise from various causes within power systems. For instance, short circuit faults, abrupt load switching, capacitor bank operations, the presence of nonlinear loads, system overloading, transformer energization, power electronic converters, arc producing devices, and the integration of distributed generation units may lead to sudden rises, drops, oscillations, or other anomalies in voltage or current (Wang and Chen 2019; Khetarpal and Tripathi 2020), (Martinez et al. 2022), (Mahela, Shaik and Gupta 2015).

The presence of PQDs in the electrical power system negatively affects the stability, efficiency, and dependability of modern electrical power systems. PQDs can cause malfunction or premature failure of electrical equipment, increase operational and maintenance costs, misoperation of protection equipment, overheating, failure of motors, loss of data, and reduction of system performance (Chiam, Lim, and Law 2023). To preserve the reliability and high performance of power systems, it is necessary to correctly detect and classify the PQDs, particularly due to the increased risk to power systems caused by the presence of highly sensitive electronic equipment and the bidirectional flow of power between the generating sources and the end-users that are characteristic of smart grid environments (Wang and Chen, 2019). Common types of PQDs include voltage sags, swells, fluctuations, notching, harmonics, interruptions, and oscillatory transients (Khetarpal and Tripathi 2020; Caicedo et al. 2023). 16 examples of PQD distortions are displayed in Figure 1. Each example is accompanied by a graphical representation of the corresponding waveform. Table 1 lists the PQD events, the mathematical models, the class labels used in this study, and IEEE 1159 (IEEE SA - IEEE 1159-2019, 2019) compatible parameters.


Flicker
Flicker with Sag
Flicker with Swell
Harmonics
Harmonics with Sag
Harmonics with Swell
Interruption
Notch
Amplitude (V)
Time (s)

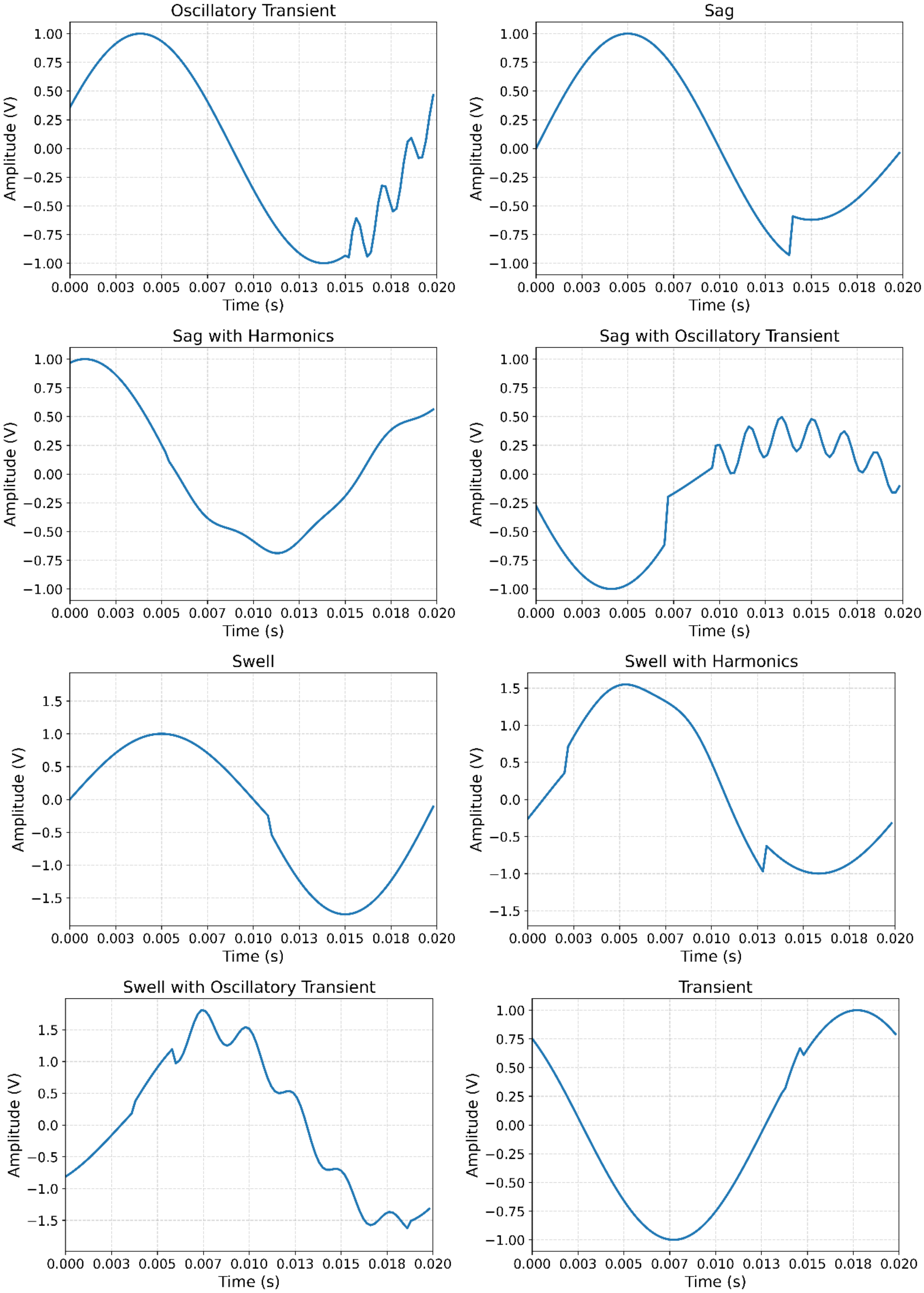


Figure 1. Sample waveforms of 16 types of PQDs used in this study.

Table 1. PQD events and mathematical models (Igual et al. 2018; Wang, Xu and Che 2019).

PQD event	Label (Class)	Equation	Parameters
Pure Sinusoidal	0	$v(t) = A sin(\omega t - \emptyset)$	$T \leq t_2 - t_1 \leq 9T$
Sag	1	$v(t) = A sin(\omega t - \emptyset) - \propto A sin(\omega t - \emptyset)[u(t - t_1) - u(t - t_2)]$	$0.1 \leq \propto \leq 0.9,$

PQD event	Label (Class)	Equation	Parameters
Swell	2	$v(t) = Asin(\omega t - \emptyset) + \beta Asin(\omega t - \emptyset)[u(t - t_1) - u(t - t_2)]$	$0 \leq t_2 \leq t_1 \leq 9T$ $0.1 \leq \beta \leq 0.8,$ $0 \leq t_2 \leq t_1 \leq 9T$
Interruption	3	$v(t) = Asin(\omega t - \emptyset) - \rho Asin(\omega t - \emptyset)[u(t - t_1) - u(t - t_2)]$	$0.9 \leq \rho \leq 1.0,$ $0 \leq t_2 \leq t_1 \leq 9T$
Transient	4	$v(t) = Asin(\omega t - \emptyset) - \propto Asin(\omega t - \emptyset)[u(t - t_1) - u(t - t_2)]$	$0 \leq \propto \leq 0.414,$ $0 \leq t_2 \leq t_1 \leq 9T$
Oscillatory transient	5	$v(t) = Asin(\omega t - \emptyset) + \propto sin(\omega t - \emptyset) + sin\,(\omega_n t) e^{-\frac{(t-t_1)}{T}}[u(t - t_1) - u(t - t_2)]$	$0.1 \leq \propto \leq 0.8,$ $0.5T \leq (t_2 - t_1) \leq 3T$
Harmonics	6	$v(t) = A[sin(\omega t - \emptyset) + \sum_{n\in\{3,5,7\}} \propto_n sin(n\omega t - \partial_n)]$	$\propto_3, \propto_5, \propto_7 \in [0.05, 0.15],$ $\propto_i \leq 1,$ $\partial_n \in [-\pi, \pi]$
Harmonics with Sag	7	$v(t) = A\,(1 - \propto [u(t - t_1) - u(t - t_2)])[sin(\omega t - \emptyset) + \sum_{n\in\{3,5,7\}} \propto_n sin(n\omega t - \partial_n)]$	$0.1 \leq \propto \leq 0.9,$ $0 \leq t_2 \leq t_1 \leq 9T$
Harmonics with Swell	8	$v(t) = A\,(1 + \beta[u(t - t_1) - u(t - t_2)])[sin(\omega t - \emptyset) + \sum_{n\in\{3,5,7\}} \propto_n sin(n\omega t - \partial_n)]$	$0.1 \leq \propto \leq 0.8,$ $0 \leq t_2 \leq t_1 \leq 9T$
Flicker	9	$v(t) = A\left[1 + \gamma sin\left(\omega_f t\right)\right] sin(\omega t - \emptyset)$	$\gamma \in [0.1, 0.2],$ $\omega_f = 2\pi f_f$, $f_f \in [40, 100]$
Flicker with Sag	10	$v(t) = A\left[1 + \beta sin\left(\omega_f t\right)\right] sin(\omega t - \emptyset)[1 - \propto (u(t - t_1) - u(t - t_2))]$	$\beta \in [0.1, 0.2],$ $\propto \in [0.1, 0.9],$ $0 \leq t_2 \leq t_1 \leq 9T,$ $f_f \in [5, 20]$
Flicker with Swell	11	$v(t) = A\left[1 + \beta sin\left(\omega_f t\right)\right] sin(\omega t - \emptyset)[1 + \propto (u(t - t_1) - u(t - t_2))]$	$\beta \in [0.1, 0.2],$ $\propto \in [0.1, 0.8],$ $0 \leq t_2 \leq t_1 \leq 9T,$ $f_f \in [5, 20]$
Sag with Oscillatory Transient	12	$v(t) = Asin(\omega t - \emptyset)\left[1 - \propto \left(u(t - t_1) - u(t - t_2)\right)\right] + \beta\, sin(\omega_n(t - t_1) - \partial)\, e^{-\frac{(t-t_1)}{T}}[u(t - t_1) - u(t - t_2)]$	$\beta \in [0.1, 0.8],$ $0 \leq t_2 \leq t_1 \leq 9T,$ $T \in [8, 40],$ $f_n \in [300, 900]$
Swell with Oscillatory Transient	13	$v(t) = A\,[1 + \beta(u(t - t_1) - u(t - t_2))]\, sin(\omega t - \emptyset) + \beta\, sin(\omega_n(t - t_1) - \partial)\, e^{-\frac{(t-t_1)}{T}}[u(t - t_1) - u(t - t_2)]$	$\beta \in [0.1, 0.8],$ $0 \leq t_2 \leq t_1 \leq 9T,$ $T \in [8, 40],$ $f_n \in [300, 900]$
Sag with Harmonics	14	$v(t) = A\,(1 - \propto [u(t - t_1) - u(t - t_2)])[sin(\omega t - \emptyset) + \sum_{n\in\{3,5,7\}} \propto_n sin(n\omega t - \partial_n)]$	$0.1 \leq \propto \leq 0.9,$ $0 \leq t_2 \leq t_1 \leq 9T$
Swell with Harmonics	15	$v(t) = A\,(1 + \beta[u(t - t_1) - u(t - t_2)])[sin(\omega t - \emptyset) + \sum_{n\in\{3,5,7\}} \propto_n sin(n\omega t - \partial_n)]$	$0.1 \leq \propto \leq 0.8,$ $0 \leq t_2 \leq t_1 \leq 9T$
Notch	16	$v(t) = A\, sin(\omega t - \emptyset) - k\, sign(sin\,(\omega t - \emptyset)) \sum_{n=0}^{cN-1}[u\left(t - (t_c + n\,s)\right) - u\left(t - (t_d + n\,s)\right)]$	$k \in [0.1, 0.4], c \in [1, 2, 4, 6],$ $s = \frac{T}{c}, 0 \leq t_c,\ t_d \leq s,$ $0.01T \leq (t_d - t_c) \leq 0.05T$

With the increasing integration of battery energy storage systems and electric vehicle charging stations in microgrids, the use of power electronic devices such as solid-state switches and electronic inverters has significantly expanded. This development has led to increased harmonic distortion within the system. In particular, series components such as transformers and cables are prone to overheating due to current harmonics. These harmonics adversely affect the performance of electronic loads. Furthermore, events such as the sudden activation of wind turbines or large loads, motor startup, and transformer energization can cause voltage sags. Similarly, sudden increases in the output power of wind turbine generators or photovoltaic generators may result in voltage fluctuations (Bollen 2003; Wang and Chen 2019). As observed, there are various types of PQDs, each caused by different underlying events. Some PQD types may occur at the same time; it is important to identify each PQD type correctly and to understand the reasons for each to avoid having imbalances in a system, which can ultimately result in a reduction in energy efficiency and an increase in carbon emissions.

Manual feature extraction and analysis using manually created features have long been the traditional means of discovering PQDs; however, these approaches are time-consuming and may be prone to error due to human involvement. The advancement of technology and the increasing volume of data generated over time have made it increasingly difficult to use traditional methods to identify, classify, and analyse all forms of disturbance (Wang and Chen 2019). DL techniques have emerged as a possible substitute for traditional approaches to identifying and classifying disturbances by automating feature extraction and increasing classification accuracy in the classification and pattern-recognition phases of PQD detection studies. The literature indicates that the ability to discern highly complex and nonlinearly correlated data using DL has increased. For PQD detection, the potential benefits of this approach could be both greater accuracy and shorter time required for completion.

2.2. Deep Learning and Short-Time Fourier Transform

DL allows automatic feature extraction directly from unprocessed signal data, eliminating the need for manual extraction and conventional approaches. Consequently, DL not only improves the accuracy of classified events in PQD investigations but also saves considerable time in processing them. Additionally, a DL model has the potential to learn more difficult-to-understand features than conventional techniques, leading to better accuracy in the classification of signals (Sekar et al. 2022; Channa and Li 2023).

Standard analyses of PQDs employ analytical techniques based on the time-frequency domain, primarily Discrete Fourier Transformations (DFTs), STFTs, S-Transforms (STs), and Wavelet Transformations (WTs) (Khetarpal and Tripathi 2020). Typically, each of these techniques is incapable of producing adequate feature representations on its own and is computationally heavy (Wang and Chen, 2019). Of these four standard time-frequency analytical techniques for the detection of PQDs, the DFTs and STFTs are by far the most commonly found in the literature (Wang and Chen 2019; Sekar et al. 2022; Channa and Li 2023), (Mahela, Shaik and Gupta 2015).

The Fourier Transform is a widely used method for analyzing stationary harmonics in voltage and current waveforms. However, due to its lack of time frequency resolution, it becomes insufficient for analyzing nonstationary events. In such cases, the STFT, which divides the signal into short-time segments to provide simultaneous time and frequency information, offers a more suitable approach (Khetarpal and Tripathi, 2020). FT and STFT are expressed by Eqs. 1 and 2.

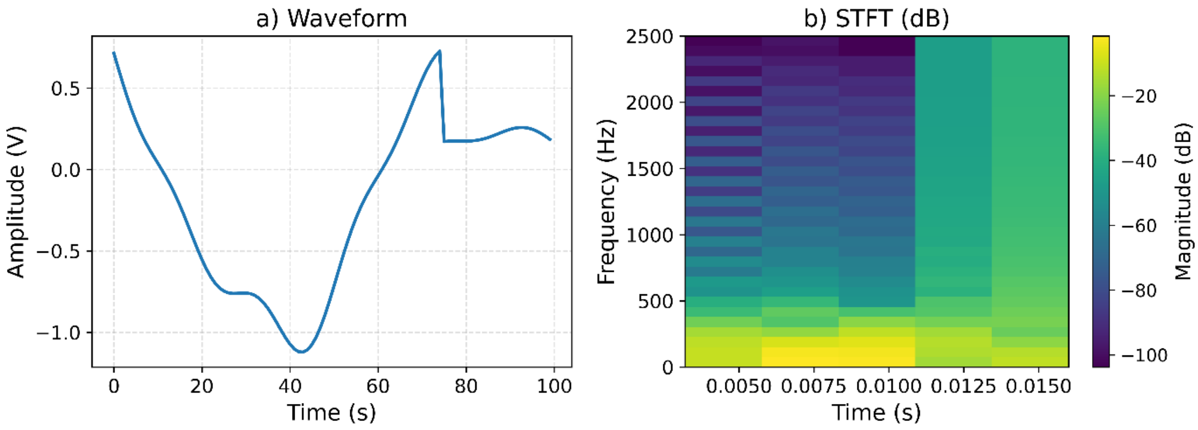
$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi f t \pi} dt \tag{1}$$

$$X(t,f) = \sum_{-\infty}^{\infty} x[n] w[n-t] e^{-j2\pi f n} \tag{2}$$

where w is the window function, f is the frequency.

In this study, the STFT method was combined with the CNN architecture, a DL model known for its success in classification, recognition, and detection tasks on images and signals. Additionally, the CNN architecture was applied directly to raw signal data. The aim of this work is to enable complex feature extraction through the CNN architecture and to evaluate the results for both 1D and two-dimensional (2D) data types. The mathematical summary of the method is given in Eq.3. Figure 2 represents raw signal and spectrogram pairs of a pure sinusoidal signal.

$$Feature\ Extraction = \{CNN_{1D}(x),\ Raw\ signal\ input\ \ CNN_{2D}(STFT(x)),\ Time\ frequency\ input \tag{3}$$



(a) **(b)**

Figure 2. **(a)** Raw signal, **(b)** Spectrogram.

3. PROPOSED METHODOLOGY

Traditional PQD detection methods rely on signal processing techniques such as the FT or WT combined with manual feature extraction and rule-based classifiers (Khetarpal and Tripathi 2020). Therefore, the traditional method of PQD detection is limited in its applicability and performance when detecting complex, noisy, and overlapping disturbances. On the other hand, DL will fundamentally change how we do feature extraction and classification by automating the feature extraction process and producing more accurate classification results, even in the presence of noise and multiple types of disturbances (Pan et al. 2024). In this research study, the STFT was used as the feature extraction method to minimise the time required for classifying PQDs and reduce the number of feature extraction parameters (Channa and Li 2023). Additionally, training on raw signal data was performed to allow for comparison on the level of performance achieved. The overall flowchart of this research study is illustrated in Figure 3.

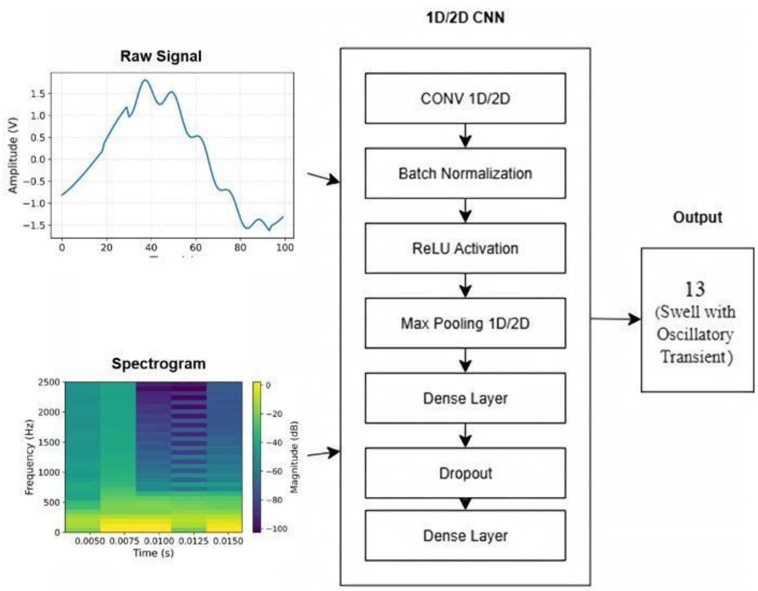
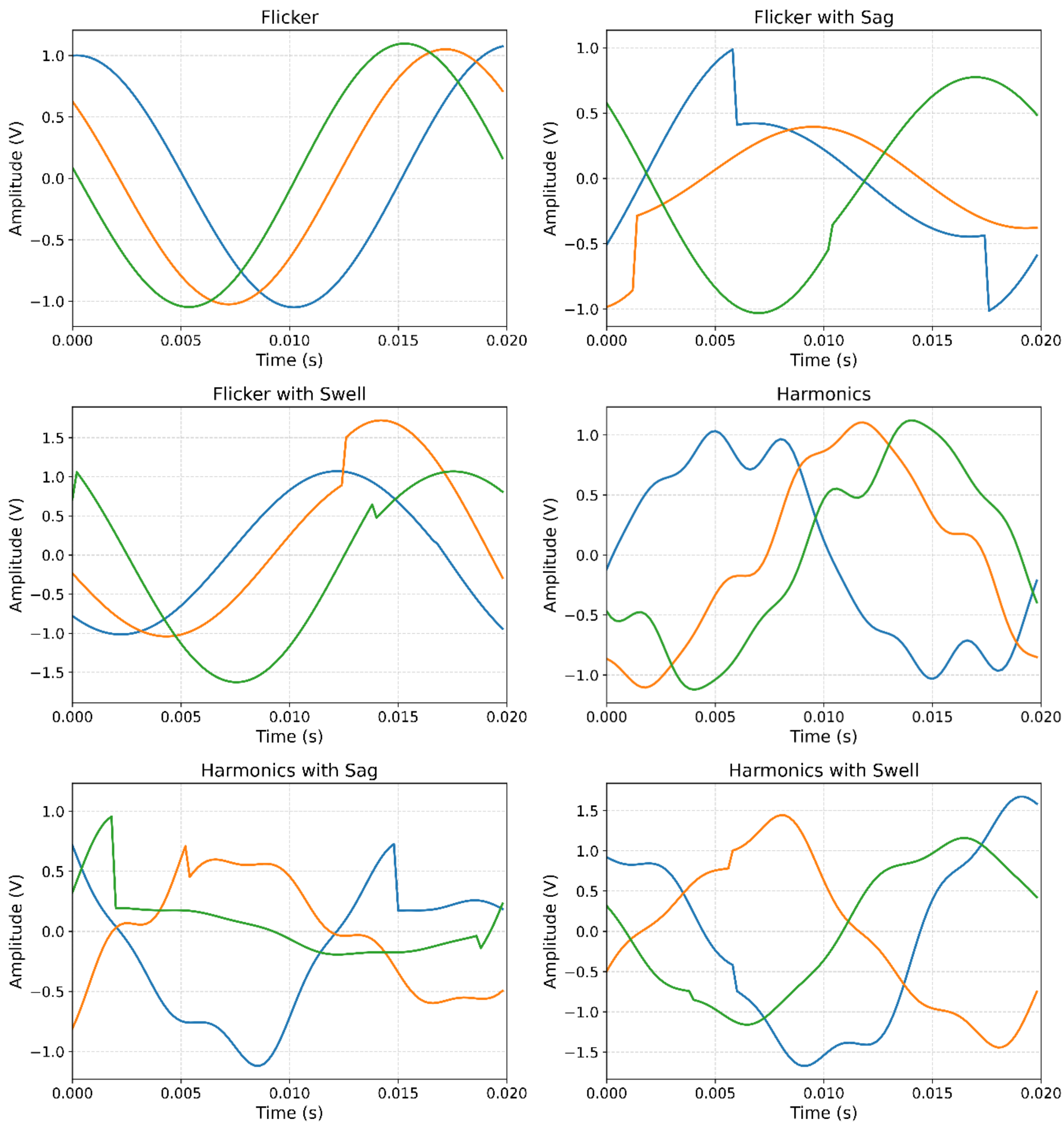
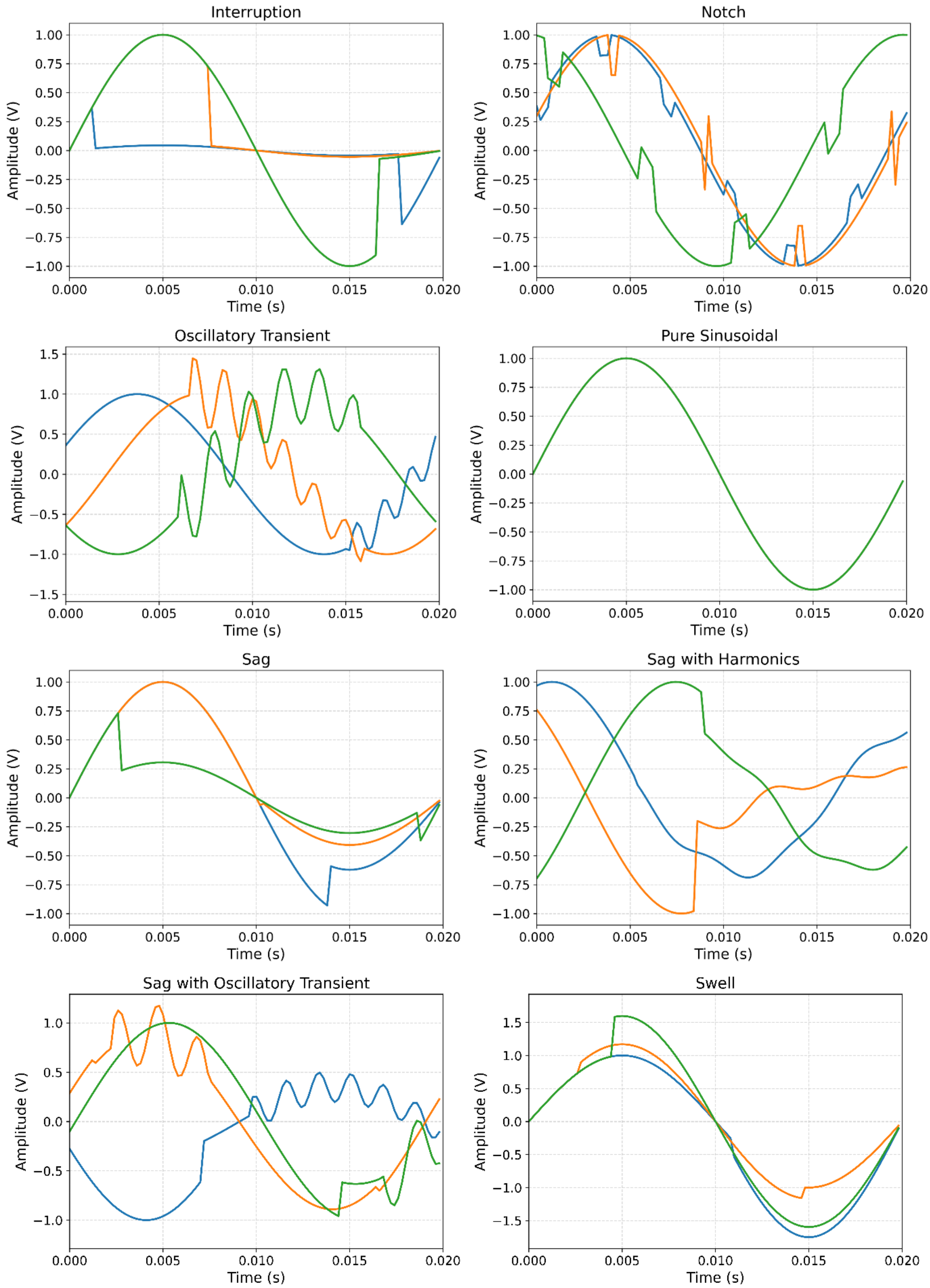


Figure 3. Flowchart of the study.

3.1. Data Preprocessing and Feature Extraction with STFT

The SEED Power Quality Disturbance Dataset is the primary data source used in this research and can be found on Kaggle. (SEED - Power Quality Disturbance Dataset 2023). The dataset consists of artificially created power disturbance signals that were specifically designed for the assessment and training of machine learning models. The dataset contains both single and multiple disturbances; the presence of single and multiple disturbances allows for a more accurate representation of real-world scenarios. There are 17 classes of PQDs in the dataset, including voltage sags, swells, interruptions, and transients, among others. The dataset consists of 8 multi-PQD signals (harmonic sag, harmonic swell, flicker sag, flicker swell, sag oscillatory transient, swell oscillatory transient, harmonic sag, and harmonic swell), and there are 1,000 labelled signals for each class; therefore, this dataset can be used for supervised learning and provide a reliable ground truth for this task. The signals in the dataset are represented as time series voltage waveforms, and they are uniformly sampled at 5 kHz. The data set is made up of 17000 signals, each with an average length of 100 data points. Figure 4 shows a graphical representation of the input signal distribution.




Interruption
Notch
Oscillatory Transient
Pure Sinusoidal
Sag
Sag with Harmonics
Sag with Oscillatory Transient
Swell
Amplitude (V)
Time (s)
0.000
0.005
0.010
0.015
0.020

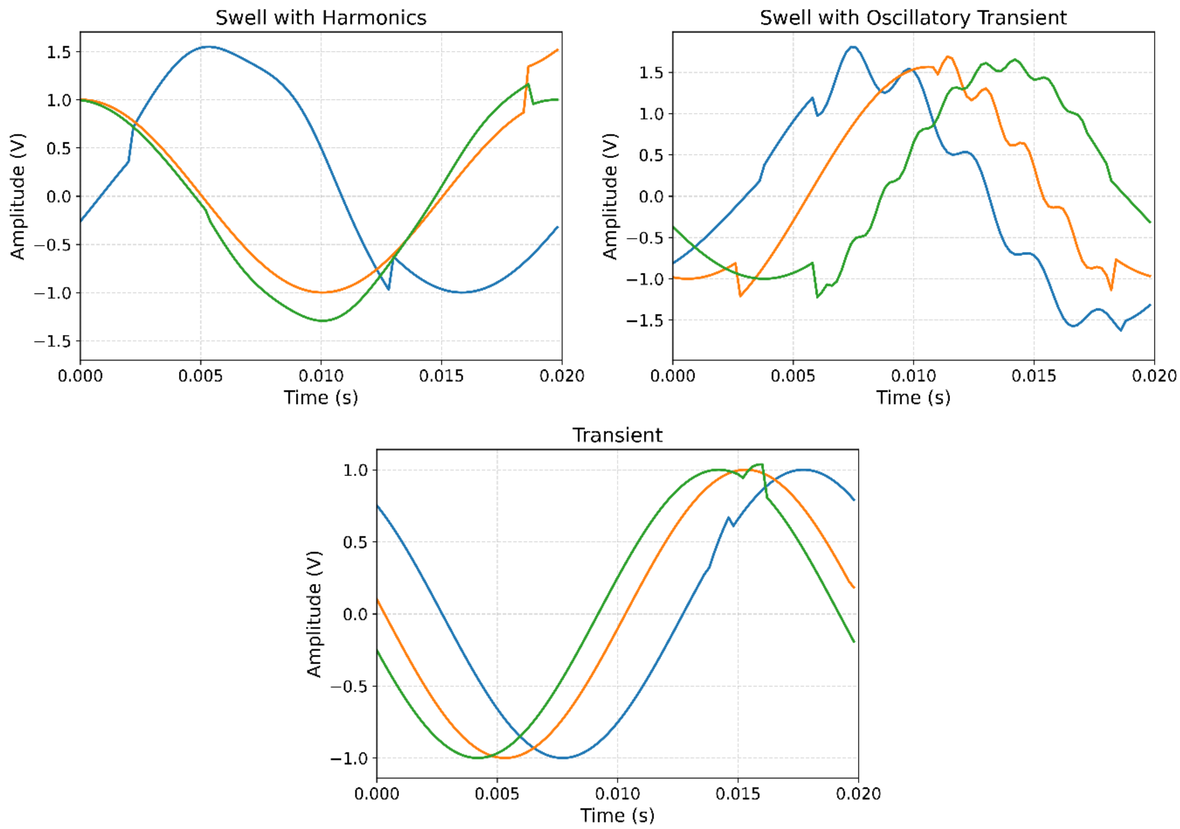


Figure 4. The signal classes (3 samples for each).

The SEED dataset consists entirely of artificially created signals; thus, the dataset contains well-defined and easily identifiable PQD patterns, making it ideal for benchmark and comparative performance evaluation. Conversely, artificially created datasets do not reproduce all of the conditions experienced when a power system is operating in the field, such as, for example, measurement noise, background harmonic content, sensor error, or a range of sampling frequency variations that might be encountered in the normal operation of a real-world power system. Although many researchers have used synthetic datasets to develop algorithmic techniques and to perform architectural comparisons between algorithms, the limitations of synthetic datasets to represent field conditions have been well known. Therefore, the results of this study can be interpreted primarily as an architectural comparison of 1D time-domain signals with 2D time-frequency representations rather than as a direct comparison of how well the systems would perform in a real-world scenario. Other studies have identified similar constraints of using synthetic PQD datasets, so it is essential to validate these PQD datasets against real-world measurements (Khetarpal and Tripathi 2020; Caicedo et al. 2023).

In this study, each 20 ms voltage waveform is transformed into a time-frequency representation using the STFT to generate spectrogram images. As a result, both the raw time domain signals (1D) and their corresponding STFT-based spectrograms (2D) are used to train CNN models.

The STFT converts raw time series signals into a joint time–frequency representation, enabling the analysis of non-stationary signal components. In this process, a sliding window is applied across the signal, and the Fourier spectrum is computed at each step, resulting in a spectrogram, a 2D representation where the horizontal axis corresponds to time, the vertical axis to frequency, and the pixel intensity to signal magnitude, as shown in Figure 5.

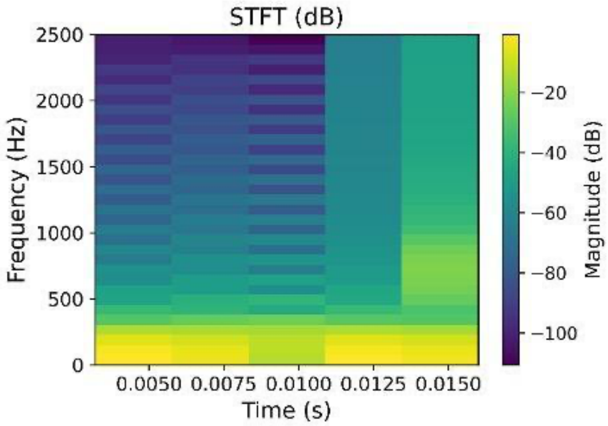
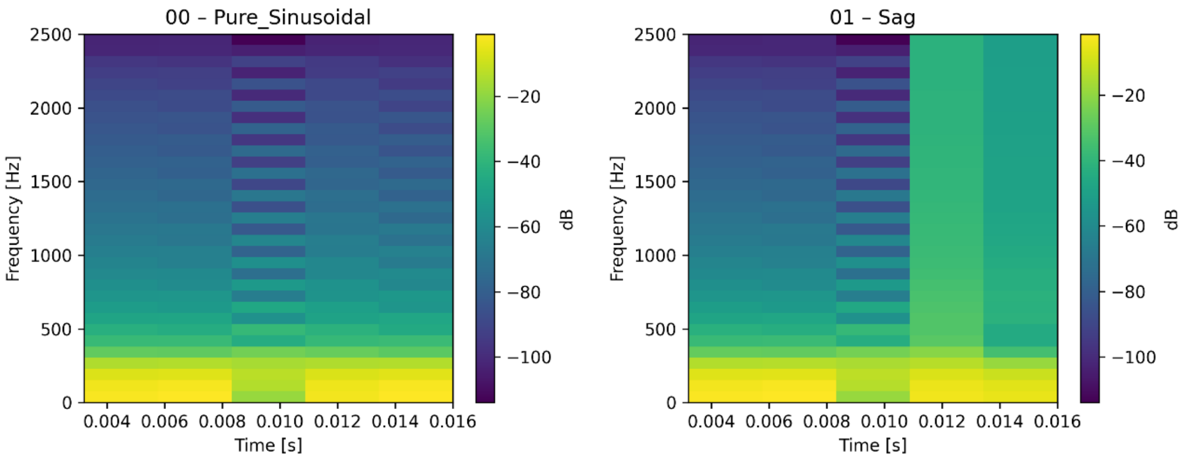


Figure 5. STFT spectrogram.

In this paper, STFT is used to create a spectrogram from each signal. The time frequency representations of the signals are used to train a 2D CNN model. STFT transformation was performed with a sampling frequency of 5 kHz. Since the signal length is 100 samples (20 ms), a short window is used to capture the temporal changes with 50% overlap. The window size is chosen as 32 samples. According to this window size, it produces a spectrogram with 156 Hz per bin and several time steps at 5 kHz sampling. The resulting spectrograms are stored as 64×64 pixel grayscale images. Some examples of STFT spectrograms of signal classes are given in Figure 6. STFT successfully captures the unique characteristic patterns associated with different types of disturbances. For instance, transients and notches produce wide frequency content because of abrupt changes in the waveform. Discrete frequency components of harmonics are located on the integer multiples of the fundamental frequency (50 Hz), with pure sinusoidal signals having 100 percent of their total energy at 50 Hz and minimum amounts at other frequencies. The sag still shows the 50 Hz band throughout, but the energy (yellow intensity) might fluctuate due to the amplitude change; since sag is an amplitude reduction, no new frequencies are introduced (the spectrogram remains mostly a single band). The notch introduces high-frequency components: the sudden drop causes a spread of energy across a broad range of frequencies (seen as the greenish bands up to high frequencies around the moment of the notch at ~10 ms).



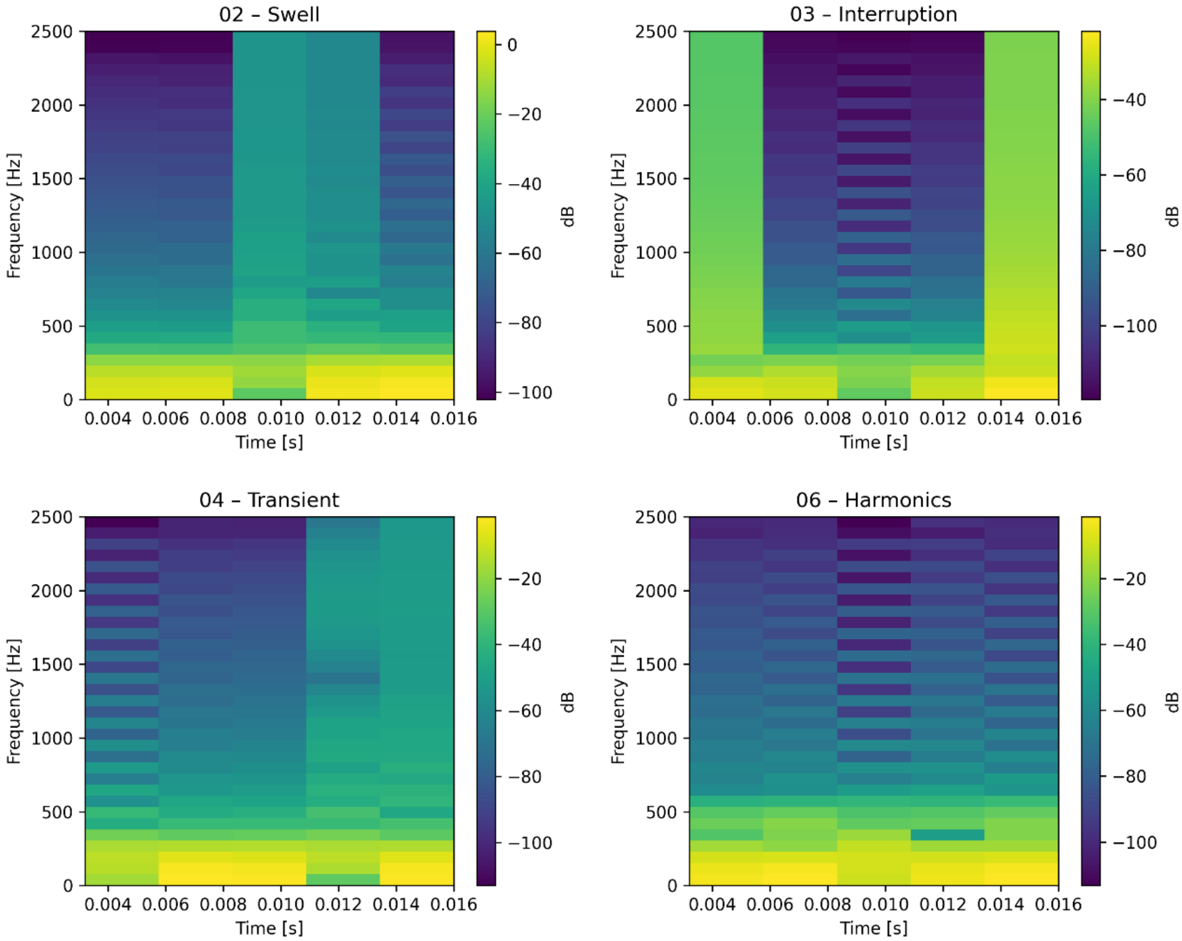


Figure 6. STFT transforms of the signal classes (6 classes were selected to show the differences).

3.2. CNN Model Architecture

Considering the complex nature of PQD signals, a DL model with strong generalization capability is essential (Wang and Chen 2019). In this study, a deep CNN approach is proposed for the classification of PQDs. Both 1D and 2D CNN architectures were implemented to evaluate model performance for raw signal inputs and time–frequency spectrogram inputs, respectively, using a comparable architectural design. The base architecture consists of multiple convolutional layers for feature extraction, pooling layers for dimensionality reduction, and dense layers for classification. Each convolutional block has an associated pooling layer to emphasise discriminating features and minimise noise.

- 1D CNN-Based Model: The model is defined as an input layer (100, 1) with three convolutional blocks of increasing filter sizes (32, 64, and 128). Every convolution block has two consecutive convolutional layers and a max pool layer, with the addition of batch normalisation and ReLU activation functions to reduce noise while providing feature extraction from the dataset. The convolution of an input signal x with a convolutional kernel k of size M can be seen in Eq. 4. The addition of batch normalisation and L2 regularisation are applied to mitigate overfitting and improve the model's generalisation performance. Max pooling was used to reduce the dimensions of the feature maps and highlight important features, as max pooling has shown significantly better performance than average pooling for the classification of PQD data (Wang and Chen, 2019). The feature maps generated by the max pooling layer were then sent to the global average pooling layer (which eliminates many parameters). In the model trained on data through the last pooling layers, the next layer consists of the fully connected/FC layers as well as a Softmax output layer. In the output layer, multi-class PQD classification was performed using softmax activation. The softmax function calculates the probability that each input signal belongs to one of 17 different PQD classes. A combination of batch normalisation and dropout is applied to further increase model depth (care was taken not to degrade performance).

$$y(i) = \sum_{m=0}^{M-1} x(i+m)k(m) \tag{4}$$

where x(i) represents the input signal at position i, k(m) is the convolution kernel coefficient at index m, and y(i) is the output feature map after convolution.

● 2D CNN-Based Model: The 2D CNN model processes greyscale spectrogram images of size 64×64 pixels generated via STFT. The architecture follows a conventional Conv2D–MaxPooling2D design for feature extraction and dimensionality reduction. After feature maps are flattened, classification is performed through fully connected layers with dropout regularisation. The remaining structure of the 2D CNN is analogous to the 1D CNN model. For a 2D CNN applied to spectrogram images, the convolutional process is given as in Eq. 5.

$$y(i) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x(i+m, j+n)k(m,n) \tag{5}$$

Where x(i, j) denotes the input image intensity at pixel (i, j), k(m, n) is the 2D convolution filter of size MxN, y(i, j) represents the convolved feature map.

The layer-by-layer summary of this architecture is presented in Table 2. The 2D CNN model architecture is given in Figure 7.

Table 2. Summary of the model.

Layer (type)	Output shape for 1D CNN	Output shape for 2D CNN	Filter	Activation	Regularization
Input Layer	(None, 100, 1)	(None, 64, 64, 1)	-	-	-
Conv1D/2D + BN + ReLU	(None, 100, 32)	(None, 64, 64, 32)	32	ReLU	L2
Conv1D/2D + BN + ReLU	(None, 100, 32)	(None, 64, 64, 32)	32	ReLU	L2
MaxPooling	(None, 50, 32)	(None, 32, 32, 32)	-	-	-
Conv1D/2D + BN + ReLU	(None, 50, 64)	(None, 32, 32, 64)	64	ReLU	L2
Conv1D/2D + BN + ReLU	(None, 50, 64)	(None, 32, 32, 64)	64	ReLU	L2
MaxPooling	(None, 25, 64)	(None, 16, 16, 64)	-	-	-
Conv1D/2D + BN + ReLU	(None, 25, 128)	(None, 16, 16, 128)	128	ReLU	L2
Conv1D/2D + BN + ReLU	(None, 25, 128)	(None, 16, 16, 128)	128	ReLU	L2
MaxPooling	(None, 12, 128)	(None, 8, 8, 128)	-	-	-
GAP	(None, 128)	(None, 128)	-	-	-
Dense	(None, 256)	(None, 256)	256	ReLU	L2
Dropout	(None, 256)	(None, 256)	-	-	0.3
Dense (Output)	(None, 17)	(None, 17)	17 classes	Softmax	-

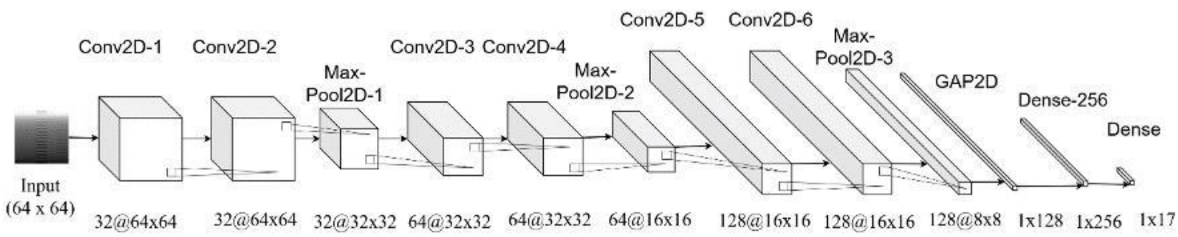


Figure 7. 2D CNN model architecture.

The main advantages of the proposed architecture are:

- Flexible use for both 1D signals and 2D spectrograms,
- Strong generalization capabilities with L2 regularization, batch normalization, and dropout,
- Prevents overfitting by reducing the number of parameters with the GAP layer,
- Provides sufficient depth for PQD classification with three consecutive convolution blocks.

3.3. TRAINING AND TESTING

The dataset was divided into training, validation, and test subsets. For testing, 100 samples from each class were reserved, resulting in a total of 1,700 test instances. The remaining 15,300 samples were split into a training set of 12,240 signals (80%) and a validation set of 3,060 signals (20%). The separation of training and validation from the testing set ensured that the test images used during the evaluation were never seen by the model during training or validation and ultimately provided for an unbiased assessment of the predictive performance of the CNN model. The model was created using the Adam optimiser with a learning rate of 1×10^{-5}, a dropout of 0.3, and a batch size of 128 during the training phase. The Adam optimizer functions as an adaptive learning rate, which improves convergence stability. The parameters described above were tuned using hyperparameter tuning techniques to help minimize overfitting of the model and, at the same time, to maximize the computational efficiency of the model. The model used sparse categorical cross-entropy as the loss function and used classification accuracy, precision, recall, and the F1 score as the evaluation metrics.

In the study of the proposed CNN model during its test phase, 1,700 images were tested on the proposed CNN model, and the CNN produced a loss value of 0.1062 and an accuracy of 97%. The CNN-produced data were processed during the entirety of each step in approximately 25 ms (average), and considering the average batch size (~121 images), the classification of a single image required an overall minimum of 0.206 ms. The image classifications represent 20-ms signal segments, and each classification was completed by the CNN model within an extremely high standard of accuracy during each classification, thus proving to be very fast and providing reliable detection capabilities. The resulting detection performance provides many advantages, including the ability to shorten processing times, lower energy use, and minimize carbon emission levels. For time-sensitive applications such as PQD classification, rapid detection is critical for the reliability and stability of the power grid. All experiments conducted were performed on the Google Colab environment using an NVIDIA Tesla T4 (16 GB GDDR6 memory) GPU. Due to its high computational power and low energy consumption, this hardware allowed for the rapid training and testing of several models. Model calculations were performed via equations 6 (Noureddine, Rouvoy, and Seinturier 2013) and 7 (Rabczuk, Łosiński and Cenian 2020) as referenced in the appropriate articles below.

$$E(kWh) = \frac{P(W)\ x\ t(s)}{3.600.000} \tag{6}$$

Energy values were calculated using average power supplied to the GPU (P-GPU), which consumed an average of 70 W throughout the testing process (PNY Technologies Inc. 2018). Since the total time spent performing testing (t) was approximately 0.35 s, it was possible to calculate energy consumption during this period.

$$CO_2\ Emission\ (kg) = E(kWh) x\ EF_N \tag{7}$$

The selected emission factor (EF), measured in grams of CO_2 equivalent per kWh (CO_2eq/kWh), was determined using the EF value for Turkey, which was derived from the local hourly grid carbon intensity (405 g CO_2/kWh). Real-time and historical data related to CO_2 emissions from electricity use were retrieved using the Electricity Maps platform (Electricity Maps ApS 2025).

Calculations show that the entire test process consumed only approximately 6.8×10^{-6} kWh of energy and produced a negligible carbon emission of 2.7×10^{-6}kg $CO_{2.2}$.

The estimates of carbon emissions presented in this research have been calculated only for the carbon footprint during the inference phase of a continuous PQD monitoring system, which is the primary operational phase of PQD monitoring systems deployed within a smart grid or edge devices, and therefore excludes the carbon footprints for the model training, data preprocessing, and manufacturing of the hardware needed to run the inference. Although these decisions were intentional (the inference phase is ongoing for long time periods, whereas the training phase is done offline, sporadically), the emission values presented here are meant to provide a basis for comparison only and will not provide a measure of total lifecycle greenhouse gas emissions.

3.4. EVALUATION

Accuracy alone is not sufficient to evaluate the performance of classifiers in supervised learning, especially for imbalanced datasets; therefore, it is necessary to evaluate classifiers using other metrics, such as Precision, Recall, and F1 Score.

- Precision

Precision is defined as the ratio of successfully predicted positive instances to all predicted positive instances. This value can be determined by using Eq. 8.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

where TP denotes true positives, and FP denotes false positives. A high-precision score indicates a low false-positive rate, meaning the model rarely misclassifies negative instances as positive.

- Recall

Recall, also referred to as sensitivity, the true positive rate, indicates what percentage of the model's outputs were actually positive. This can be modelled mathematically per Eq. 9 (where the FN value indicates false negatives).

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

Thus, a high value of recall indicates the model is able to capture most actual positive cases.

- F1 Score

The harmonic mean of both precision and recall is described by the F1 Score; it provides a balance of both measures. The F1 Score is especially useful when there are imbalanced class distributions. The F1 Score can be calculated using Eq. 10.

$$Precision = \frac{Precision * Recall}{Precision + \ Recall} \tag{10}$$

If the F1 Score is large, it indicates a good balance between minimising false positives and false negatives (referring to high precision and high recall).

4. EXPERIMENTAL RESULTS

The results from the experiments using the raw 1D signal and the 2D spectrogram via the STFT to train the model were evaluated on their performance in terms of classification accuracy and classification loss. All models were trained for two durations of time (60 epochs and 120 epochs) to examine the benefits of additional epochs on both training accuracies and training losses as well as any potential overfitting. The results from the two data types (i.e., 1D raw signal and 2D spectrogram) can be found in Table 3, while the respective training curves for the two data types can be found in Figures 8 and 9.

Table 3. Train and validation accuracy/loss values.

Full Training	**Raw Signal Data**		**Spectrogram Data**	
	Train	**Validation**	**Train**	**Validation**
60 epochs	accuracy: 0.7819 loss: 0.7215	val_accuracy: 0.8327 val_loss: 0.6349	accuracy: 0.7130 loss: 0.9471	val_accuracy: 0.7536 val_loss: 0.8754
120 epochs	accuracy: 0.9734 loss: 0.0985	val_accuracy: 0.9699 val_loss: 0.0988	accuracy: 0.8112 loss: 0.6484	val_accuracy: 0.8141 val_loss: 0.6379

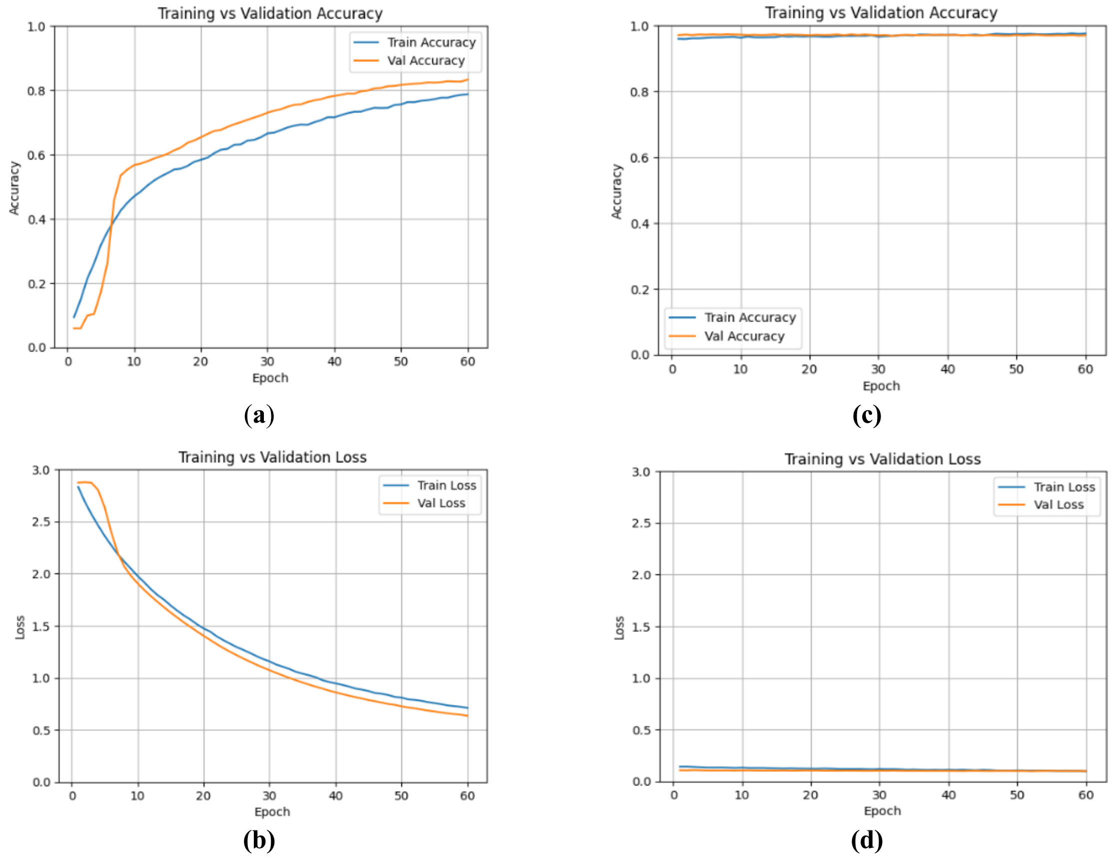


Figure 8. Training and validation accuracy and loss graphs for full signal data training (**(a)** and **(b)** for 60 epochs, **(c)** and **(d)** for 120 epochs.)

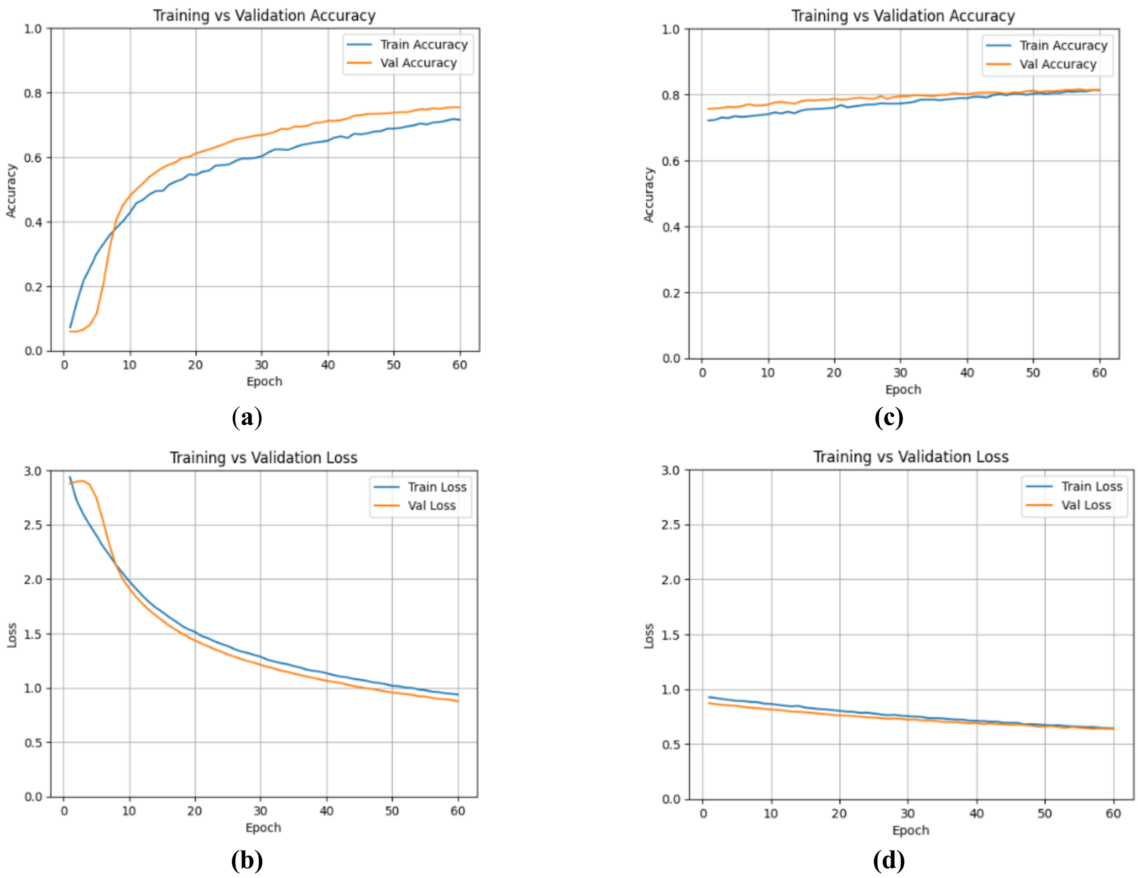


Figure 9. Training and validation accuracy and loss graphs for full spectrogram data training (**(a)** and **(b)** for 60 epochs, **(c)** and **(d)** for 120 epochs.)

After completing 60 training epochs on the 1D raw signal data, the model achieved a validation accuracy of 83.27%. After completing 120 training epochs, the model achieved an increased validation accuracy of 96.99% along with a corresponding decrease in validation loss from 0.6349 to 0.0988 over the same time period. The significant increases in validation accuracy and the large decrease in validation loss suggest that longer training is beneficial to the model's ability to differentiate between classes and that the model has learned robustly discriminative temporal features without any significant overfitting to the training set, as evidenced by the close correspondence between the performance achieved on the training set and performance achieved on the validation set. For 2D spectrogram data, the validation accuracy after 60 epochs was about 75%, and after 120 epochs it was about 81%. Additionally, the corresponding validation loss decreased from 0.8754 to 0.6379. Longer training provided better performance for the spectrogram model; however, the overall accuracy was much lower than that of the 1D model. This indicates that in this dataset, raw signals' preserved temporal information provides a more discriminative cue to classify PQD than the time-frequency patterns of the spectrograms. The results of the comparison between both methods demonstrate a clear superiority of the 1D CNN model compared to the 2D model for all training durations.

It is essential to take noise from reality into consideration, as well as aspects of sampling, when designing practical PQD systems to monitor PQD. Noise was not added or changed in this particular work because one of the main goals of this work is to do a fair comparison and controlled comparison of 1D and 2D CNNs using the same conditions. Adding extra noise models or different sampling rates would create bias in how one representation or another is compared. There are many methods available to improve noise resistance, such as data augmentation (adding random data), using attention mechanisms, or hybrid architectures, and they were purposely not included in this study to provide a more straightforward method of construction, as well as for comparison purposes (Chiam, Lim and Law 2023; Albalooshi and Qader 2025). Future research will expand on this research to examine the robustness of these results using signals that have been damaged by noise and have been sampled asynchronously.

To assess how well each model generalizes, both were evaluated against an independent subset of the data. Average precision, recall, F1 scores, and accurate results for the evaluation are summarized in Tables 4 and 5, respectively. The 1D CNN trained on raw signals produced an overall test accuracy of 95.06%, compared to 82.88% for the 2D CNN trained on spectrograms, resulting in a substantial difference between the two. In addition, 1D model precision, recall, and F1 scores were approximately 11–12% better than the 2D model across all evaluation metrics, indicating consistently superior classification performance by the 1D model.

Table 4. Test results for raw signal data.

Class	Precision	Recall	F1 Score	Support
0	0.9804	1.0000	0.9901	100
1	0.9798	0.9700	0.9749	100
2	0.9703	0.9800	0.9751	100
3	0.9709	1.0000	0.9852	100
4	0.8909	0.9800	0.9333	100
5	0.9091	0.9000	0.9045	100
6	0.9709	1.0000	0.9852	100
7	0.9505	0.9600	0.9552	100
8	0.9798	0.9700	0.9749	100
9	0.9612	0.9900	0.9754	100
10	0.9143	0.9600	0.9366	100
11	0.8991	0.9800	0.9378	100
12	0.9889	0.8900	0.9368	100
13	1.0000	0.8800	0.9362	100
14	0.9053	0.8600	0.8821	100
15	0.9239	0.8500	0.8854	100
16	0.9802	0.9900	0.9851	100
Accuracy			0.9506	1700
Loss			0.1571	1700
Macro avg	0.9515	0.9506	0.9502	1700

Table 5. Test results for spectrogram data.

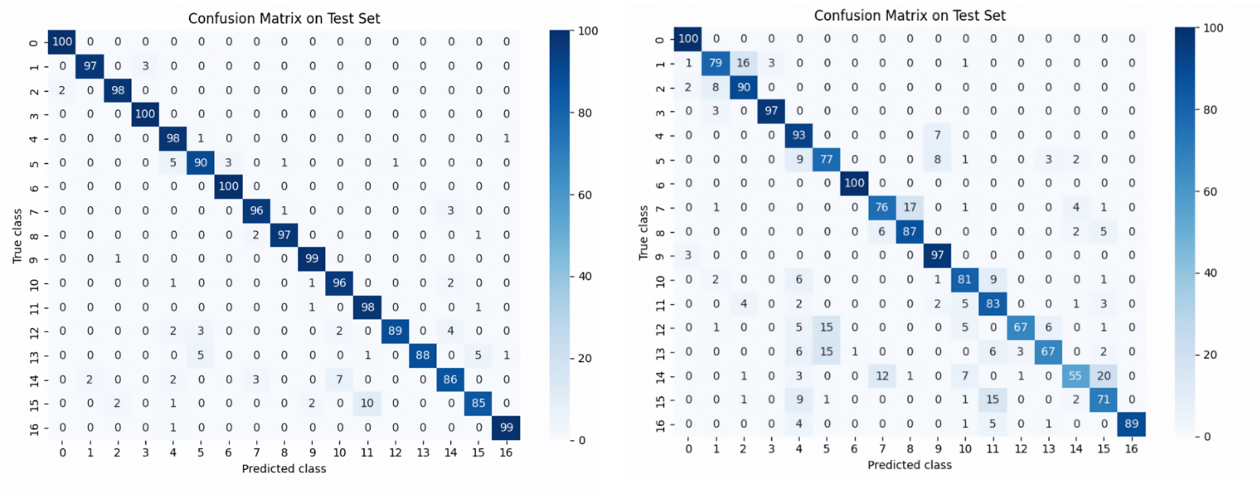
Class	Precision	Recall	F1 Score	Support
0	0.9434	1.0000	0.9709	100
1	0.8404	0.7900	0.8144	100
2	0.8036	0.9000	0.8491	100
3	0.9700	0.9700	0.9700	100
4	0.6788	0.9300	0.7848	100
5	0.7130	0.7700	0.7404	100
6	0.9901	1.0000	0.9950	100
7	0.8085	0.7600	0.7835	100
8	0.8286	0.8700	0.8488	100
9	0.8435	0.9700	0.9023	100
10	0.7864	0.8100	0.7980	100
11	0.7034	0.8300	0.7615	100
12	0.9437	0.6700	0.7836	100
13	0.8701	0.6700	0.7571	100
14	0.8333	0.5500	0.6627	100
15	0.6827	0.7100	0.6961	100
16	1.0000	0.8900	0.9418	100
Accuracy			0.8288	1700
Loss			0.6179	1700
Macro avg	0.8376	0.8288	0.8271	1700

Table 4 contains the classification report of the raw signal data, showing that the overall classification accuracy was 95.06% on average, the macro-average precision was 95.15%, the recall was 95.06%, and the F1-score was 95.02 %. These metrics indicate that the model can recognize meaningful patterns from the raw signals, thus showing strong generalisation abilities. Classes 0 (Pure Sinusoidal), 3 (Interruption), 6 (Harmonics), and 9 (Harmonics with Swell), in particular, showed almost perfect classification performance (F1-scores of 0.985 or greater). Moreover, the signal characteristics of these classes were distinct from one another, leading to an ability to identify them accurately. In some classes, precision was high, while recall was comparatively low. This indicates that the model was quite accurate in predicting examples for the relevant class, but missed some real examples. Conversely, in Class 12 (Sag with Oscillatory Transient), recall was high, but precision was low, resulting in a high number of false positives when predicting the class. The results show that the model can classify raw signals with high accuracy and performs strongly for most of the PQD classes, with room for improvement only for some mixed PQD classes. The test loss value of 0.1571 demonstrates that the model achieves a very low level of not only classification accuracy but also prediction error. This result confirms that the model does not exhibit overfitting during the learning process and exhibits high generalization ability on the test set. Therefore, the low loss value obtained supports the model's reliable performance in accurately distinguishing PQD classes.

The results presented in Table 5 demonstrate the classification performance of the CNN model on spectrogram data. This model achieved an overall accuracy of 82.88 percent and an average F1 score of 0.8271. These results show that the model can accurately separate most PQD classes. In particular, the F1 scores for classes 0 (pure sinusoidal), 3 (interruption), 6 (harmonic), 9 (flicker), and 16 (notch) are all above 90%; therefore, these classes have very high recognition success rates. However, there is a significant drop in performance on other complex types of PQD; for example, class 14 (sag with harmonic), class 15 (swell with harmonic), and class 5 (oscillatory transient). The loss model value for this model is 0.6179, meaning that certain classes have reduced performance due to misclassifications; thus, the overall model performance has been restricted by this factor. The data indicate that the spectrogram-based model has excellent performance for some classes but will continue to need additional improvements, especially when dealing with mixed PQD types.

Tables 4 and 5 combined show that the CNN model learning directly from the raw signal outperformed the spectrogram-based model. In terms of how each model performed by class (ie, classes 1-15), the CNN model achieved balanced performance across all classes, while the spectrogram-based model experienced large performance losses for mixed PQD type classes (eg, classes 14 and 15). This comparison shows that learning from the raw signal provides a better representation for PQD classification than does learning from spectrograms and therefore provides more reliable results.

The practical impact of the comparison of raw signal-based models to spectrogram models is thus an important consideration when exploring sustainability issues. Using raw signal data rather than generating time–frequency images as a preprocessing stage reduces the amount of time and computation required, thus supporting reduced energy consumption and faster decision-making when deploying PQD detection models in continuous monitoring systems (e.g., edge-based systems).



(a) **(b)**

Figure 10. Confusion matrices for **(a)** 1D and **(b)** 2D models.

In Figures 10a and 10b, we present the confusion matrices generated from the trained models with both raw signal data (10a) and spectrogram data (10b). The classification performance for the raw signal-based model was quite good; classification accuracy was greater than 97% on average for 5 of the 17 (0, 1, 2, 3, and 6) PQD categories within the raw data model classification, whereas compound disturbances such as 7, 8, 10, and 11 were each greater than 85%. Substantial confusion between compound events did occur, as there were a few instances of confusion between categories that had a similar spectral signature; for example, the confusion between 14 (Sag with Harmonics) and 15 (Swell with Harmonics). Overall, as a result of strong diagonal dominance in the classification results of the confusion matrix generated with the raw waveform data input, we conclude that the raw waveforms preserve both temporal and spectral characteristics well enough to allow for robust discrimination of the various PQD classes despite no explicit feature transformation preceding the classification.

On the other hand, the spectrogram-based approach shows good results in classification rates when considering certain types of compound PQDs. However, the classification rates for these types of compound PQDs will decrease. For example, three single disturbances – (0) pure sinusoidal disturbance, (1) sag, and (2) swell – are classified very well. As the spectrograms provide an ambiguous representation of short-duration events, such as (4) transient and (5) oscillatory transient events, and overlapping events, this is likely due in part to the trade-off between time and frequency resolutions experienced in the STFT process, where important aspects of rapid-appearing disturbances may become blurred by this time and frequency trade-off. Nonetheless, the spectrogram-based method still provides a unique frequency-domain view into how disturbances operate and can be acceptable for use in applications that require the use of spectral interpretations.

In Table 6, we will find a summary of a comparison between previously proposed PQD classification methods and other representative PQD classification methods found in the literature. The comparisons indicate the differences in terms of input representation, number of disturbance classes, and classification accuracy. The previous studies either focus on a few number of PQD signal types or use only one signal representation; however, in this study, we investigate both raw and Time Frequency inputs with a unified experimental framework.

Table 6. Comparative summary of PQD classification studies

Ref.	Study / Approach	Model Type	Input / Representation	Dataset Type	#Classes	Reported Performance
(Wang and Chen, 2019)	Closed-loop 1D-DCNN	1D CNN (DCNN)	Raw 1D signal	Synthetic (IEEE-1159) + real field data	16	Accuracy varies with

Ref.	Study / Approach	Model Type	Input / Representation	Dataset Type	#Classes	Reported Performance
						SNR; no single overall value reported)
(Sekar et al., 2022)	Hybrid CNN–LSTM	1D CNN + LSTM	Raw 1D signals	Synthetic (IEEE 13-node hybrid system, MATLAB/Simulink)	16	Accuracy: 99.90% (clean / low noise), 99.31% (20 dB SNR)
(Albalooshi and Qader, 2025)	Multi-scale CNN with channel attention	1D CNN + Attention mechanism	Raw 1D signal	Synthetic (noisy & noise-free, 20–50 dB SNR)	10	High accuracy reported (varies with SNR; no single overall value reported)
(Garcia et al., 2020)	CNN vs LSTM vs CNN–LSTM	CNN / LSTM / CNN–LSTM	Raw time-domain voltage & current signals	MATLAB/Simulink (synthetic) + real field measurements	6	Accuracy: 84.76% (CNN–LSTM, best), 84.58% (CNN), 79.14% (LSTM)
(Ma et al., 2017)	SAE-based PQD classification	Stacked Autoencoder (SAE) (+ PSO)	Learned features from signals	Synthetic	7	Accuracy: 99.75% (noiseless), 99.61% (30 dB), 98.53% (20 dB)
(Khetarpal *et al.*, 2023)	DCAE + Gabor features	Deep Autoencoder (DAE) + Gabor + Softmax	PQD images with Gabor features	Synthetic (noisy & noise-free)	15	Accuracy: >97% (20 dB SNR)
(Topaloglu, 2023)	Attention-supported CNN	CNN + Attention model	Time-domain → image-like tensors	Synthetic (MATLAB/Simulink)	9	Accuracy: 99.92%
(Wang, Xu and Che, 2019)	CS + 1D-DCNN	Compressive sensing + 1D DCNN	Compressed measurements	Synthetic	15	Accuracy: ≈91% (measured data); higher on simulation data
(Yiğit *et al.*, 2021)	CNN + GRU on STFT	CNN + GRU	STFT-based time–frequency matrices	Synthetic	7	Accuracy: 98.44%
(Channa and Li, 2023)	STFT features + 2D CNN	2D CNN	STFT time–frequency	Synthetic + IEEE-13 node system	10	Accuracy: 99.99% (noiseless), 97.50% (30 dB), 96.25% (20 dB)
(Baig *et al.*, 2024)	Transfer learning + ensemble	TL-based DCNNs + Voting Ensemble	CWT time–frequency images	Synthetic (noise-free & noisy datasets)	16	Accuracy: 99.98%
(Perez-Anaya *et al.*, 2024)	CWT-based PQD diagnosis	CNN	CWT time–frequency images (128×128)	Synthetic + real measurements	6	Accuracy: 99.48%
(Cai, Zhang and Jiang, 2023)	Parallel CNN–GRU with attention	CNN–GRU-P (SE + Attention)	Raw time-domain signals	Synthetic (MATLAB/Simulink)	12	Accuracy: 99.6%

Ref.	Study / Approach	Model Type	Input / Representation	Dataset Type	#Classes	Reported Performance
(Bai *et al.*, 2025)	FST + improved CNN–LSTM	FST + CNN–LSTM (SSA tuned)	Time–frequency transform	Synthetic + IEEE-33 node simulation	5	Accuracy: >97%
(Chiam, Lim and Law, 2023)	WT + attention + LSTM	Wavelet + Attention + LSTM	Multi-resolution wavelet coefficients (TFA / SFA)	Synthetic	16	Accuracy: 91.78% (20 dB)
This study	1D CNN (raw)	1D CNN	Raw time-domain (100×1)	SEED (Kaggle)	17	Accuracy: 95.06%
This study	2D CNN (STFT)	2D CNN	STFT spectrogram (64×64×1)	SEED (Kaggle)	17	Accuracy: 82.88%

Table 6 summarizes findings in existing studies on the classification of disturbances due to PQ with reference predominantly to the synthetically generated datasets; this is of predominantly mathematical origin based upon the IEEE-1159 standards for creating synthetic datasets, being capable of controlled analysis and having large-scale training capability. Synthetic datasets may not capture the true variability and uncertainty of real power grid environments. Additionally, the majority of studies report having used a small number of disturbance class types, with the majority of studies reporting between 6 and 16 classes. As such, a majority of existing research will have limited generalization capacity when the studies are extended to complex or real-world environments. Many studies utilize advanced hybrid architectures, including but not limited to multi-stage feature extraction, wavelets, attention mechanisms, recurrent layers, and many other advanced techniques, primarily to increase robustness to noise. However, the proposed research utilizes a publicly available dataset containing 17 PQD classes to assess both the delivered raw time-domain signal and STFT-based time-frequency representation under the same experiment conditions. According to what was found, it appears that using simple representations and building your own model from scratch makes for good classification performance. Such a conclusion should serve as motivation for others who want to develop practical/efficient methods for PQD monitoring.

Overall, there is no doubt that one-dimensional CNNs still outperformed two-dimensional CNNs on every evaluation metric evaluated, but in terms of picking out all of the major classes of PQD, both methods performed reasonably well to good. However, raw signal-based CNNs produced superior accuracy, precision, and F1 scores compared to STFT-based spectrogram networks, and this advantage was particularly significant for detecting compound disturbances for both networks used for all three metrics above. Therefore, temporal features present in raw waveforms are likely to contain more discriminatory information than those contained within STFT-based spectrogram representations when trained under identical experimental conditions.

5. CONCLUSION

This research evaluated the use of deep learning to classify PQDs through the analysis of both raw time-domain signal data and time-frequency data in the form of spectrograms. The results indicate that 1D CNNs have better performance than 2D CNNs trained on spectrogram data, demonstrating how well raw time domain input can detect PQDs. In addition, because of their ability to operate with minimal preprocessing requirements, 1D CNNs are especially suited for use as part of real-time monitoring systems within resource-constrained smart grids. These results provide support for the notion that raw signal features can assist in creating faster, more effective, and more economical monitoring solutions that support energy savings and reductions in greenhouse gas emissions in contemporary electrical networks. Further, when comparing performance using the same dataset and model complexity, 1D CNNs trained using only raw input data outperformed 2D CNNs trained using time-frequency data. This demonstrates the significant impact that the representation type plays in PQD classification research. The performance evaluations showed that the proposed model had a high degree of accuracy with minimal computational resources, thereby enabling the model to have significantly low energy consumption and carbon emissions during testing. This level of efficiency illustrates that DL models have the potential to assist sustainable energy management by limiting unnecessary power loss and environmental harm. The proposed framework will allow for efficient monitoring of power quality disturbance (PQD) using fewer computational resources; therefore, it fits with decarbonisation-based smart grid strategies seeking to reduce energy losses and unnecessary resource consumption.The increased accuracy achieved by using a 1D CNN model can be utilized to support and improve the operation of smart grids. By being able to provide more accurate detection of PQD, the number of false alarms or missed occurrences will be reduced. Consequently, this allows operators to take quicker corrective action when an event occurs, thus improving the overall stability of the power system. Even though the research does not give a numerical value for the total amount of energy that would have been saved at a system level from this work, the presence of accurate and low-cost overhead monitoring of PQDs enables utility companies to operate their grid in a more efficient manner while also reducing system technical losses. As a result, the contribution to energy efficiency and reducing carbon emissions from improved operational performance should be viewed as indirect contributions rather than a direct measure or calculation of emissions that were avoided.

6. LIMITATIONS AND FUTURE WORK

While the proposed method shows good results for classification, there are some limitations. First, this study used a synthetic PQD dataset; therefore, it has limited disturbance patterns (disturbances were controlled and defined and have specific criteria) to fully capture the variability, noise, and operating uncertainties of real power grids. As such, the results for the overall classification should be interpreted only as a comparison of how well various representations of the input signals match the output classes used as ground truth and not how well they would perform in the field if they were deployed in a full-scale power system. Also, the analysis is done on short, fixed-length segments of the input signal, thereby not including longer-duration disturbances or the evolving PQD events that typically occur in practical power grids.

A limitation of this study is that it did not include any verification or validation of the method using actual PQD signals measured in real time from the power system. Power system measurements may also contain noise from sensors and background harmonics, as well as from variation due to the operation of equipment, which are not captured in our synthetic datasets. Therefore, future work will be aimed at verifying the new approach using actual measurements from the power grid and determining its robustness under actual conditions of operation.

In the future, research should concentrate on optimizing deep learning models to identify PQD events across multiple grid environments in real time, as well as reducing overfitting by using more real-world datasets. Another avenue for enhancing accuracy and robustness will be through hybrid architectures that combine the processing of raw signals and advanced methods of extracting features. Additionally, implementing these models on edge devices could provide low-latency detection of PQD events and help make power systems more resilient, efficient, and environmentally sustainable. Another area for future work is to test and validate proposed models using actual grid measurements from distribution networks and microgrids. Investigating the use of proposed frameworks for real-time monitoring systems and edge-based devices will also be an area of exploration. Such developments have the potential to enable early detection of PQD events with low computation overheads, which will ultimately lead to increased energy efficiency, reduced energy loss, and contribute toward decarbonizing modern power systems.

AUTHOR CONTRIBUTIONS

Kevser İrem Danaci: Conceptualization, Methodology, Software, Data Curation, Formal Analysis, and Writing Original Draft. **Sıtkı Akkaya**: Supervision, Validation, Formal Analysis, and Writing, Review & Editing.

COMPETING INTERESTS

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 ACCESSIBILITY

The dataset used in this study is publicly available through the Kaggle platform and corresponds to the SEED Power Quality Disturbance dataset (Khan, Aziz, and Usman, 2023). The dataset can be accessed via its official Kaggle repository and is cited accordingly in the reference list.

ETHICAL APPROVAL

This study utilizes a publicly available synthetic dataset (SEED Power Quality Disturbance Dataset) ; therefore, ethics committee approval and informed consent were not required.

REFERENCES

Albalooshi, F. A. and Qader, M. R., 2025. Deep Learning Algorithm for Automatic Classification of Power Quality Disturbances, Applied Sciences (Switzerland). Multidisciplinary Digital Publishing Institute, 15(3), p. 1442. DOI: doi.org/10.3390/APP15031442.

Bai, H. et al., 2025. Power Quality Disturbance Classification Strategy Based on Fast S-Transform and an Improved CNN-LSTM Hybrid Model, Processes, 13(3), pp. 1–20. DOI: doi.org/10.3390/pr13030743.

Baig, M. A. A. et al., 2024. An Ensemble Deep CNN Approach for Power Quality Disturbance Classification: A Technological Route Towards Smart Cities Using Image-Based Transfer, Future Internet, 16(12), pp. 1–24. DOI: doi.org/10.3390/fi16120436.

Bollen, M. H. J., 2003. What is power quality?, Electric Power Systems Research, 66(1), pp. 5–14. DOI: doi.org/10.1016/S0378-7796(03)00067-1.

Cai, J., Zhang, K. and Jiang, H., 2023. Power Quality Disturbance Classification Based on Parallel Fusion of CNN and GRU, Energies, 16(10). DOI: doi.org/10.3390/en16104029.

Caicedo, J. E. et al., 2023, A systematic review of real-time detection and classification of power quality disturbances, Protection and Control of Modern Power Systems. Springer Nature Singapore, 8(1). DOI: doi.org/10.1186/s41601-023-00277-y.

Channa, I. A. and Li, D., 2023. Detection and classification of power quality disturbances using STFT and deep neural Network, ACM International Conference Proceeding Series, pp. 461–466. DOI: doi.org/10.1145/3638584.3638596.

Chiam, D. H., Lim, K. H. and Law, K. H., 2023. LSTM power quality disturbance classification with wavelets and attention mechanism, Electrical Engineering. Springer Berlin Heidelberg, 105(1), pp. 259–266. DOI: doi.org/10.1007/s00202-022-01667-5.

Ding, Y. et al., 2025. Carbon management for modern power system: An overview, Smart Power & Energy Security. The Authors, 1(1), pp. 12–24. DOI: doi.org/10.1016/j.spes.2024.06.001.

Electricity Maps ApS, 2025. Electricity Maps. Available at: https://app.electricitymaps.com/map/72h/hourly (Accessed: 21 August 2025).

Garcia, C. I. et al., 2020. A Comparison of Power Quality Disturbance Detection and Classification Methods Using CNN, LSTM and CNN-LSTM, Applied Sciences. Multidisciplinary Digital Publishing Institute, 10(19), p. 6755. DOI: doi.org/10.3390/APP10196755.

IEEE SA - IEEE 1159-2019, 2019. Available at: https://standards.ieee.org/ieee/1159/6124/ (Accessed: 15 August 2025).

Igual, R. et al., 2018. Integral mathematical model of power quality disturbances, Proceedings of International Conference on Harmonics and Quality of Power, ICHQP, 2018-May, pp. 1–6. DOI: doi.org/10.1109/ICHQP.2018.8378902.

International Energy Agency (IEA), 2024. Global Energy and CO_2 Status Report 2024. Available at: https://www.iea.org/reports/global-energy-and-co2-status-report-2024.

Jiang, Y. and Mao, Z., 2025. A novel carbon emission monitoring method for power generation enterprises based on hybrid transformer model, Scientific Reports, 15(1), pp. 1–16. DOI: doi.org/10.1038/s41598-024-82188-y.

Khan, M. U., Aziz, S. and Usman, A., 2023. XPQRS: Expert power quality recognition system for sensitive load applications, Measurement: Journal of the International Measurement Confederation. Elsevier B.V., 216. DOI: doi.org/10.1016/J.MEASUREMENT.2023.112889.

Khetarpal, P. et al., 2023. Power Quality Disturbances Detection and Classification Based on Deep Convolution Auto-Encoder Networks, IEEE Access. Institute of Electrical and Electronics Engineers Inc., 11, pp. 46026–46038. DOI: doi.org/10.1109/ACCESS.2023.3274732.

Khetarpal, P. and Tripathi, M. M., 2020. A critical and comprehensive review on power quality disturbance detection and classification, Sustainable Computing: Informatics and Systems. Elsevier Inc., 28. DOI: doi.org/10.1016/j.suscom.2020.100417.

Lecun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning, Nature. Nature Publishing Group, 521(7553), pp. 436–444. DOI: doi.org/10.1038/NATURE14539.

Lim, B. et al. 2021. Temporal Fusion Transformers for interpretable multi-horizon time series forecasting, International Journal of Forecasting. Elsevier, 37(4), pp. 1748–1764. DOI: doi.org/10.1016/J.IJFORECAST.2021.03.012.

Ma, J. et al., 2017. Classification of Power Quality Disturbances via Deep Learning, IETE Technical Review, 34(4), pp. 408–415. DOI: doi.org/10.1080/02564602.2016.1196620.

Mahela, O. P., Shaik, A. G. and Gupta, N., 2015. A critical review of detection and classification of power quality events, Renewable and Sustainable Energy Reviews. Elsevier, 41, pp. 495–505. DOI: doi.org/10.1016/j.rser.2014.08.070.

Martinez, R. et al. 2022. Techniques to Locate the Origin of Power Quality Disturbances in a Power System: A Review, Sustainability. Multidisciplinary Digital Publishing Institute, 14(12), p. 7428. DOI: doi.org/10.3390/SU14127428.

Noureddine, A., Rouvoy, R. and Seinturier, L., 2013. A review of energy measurement approaches, Operating Systems Review (ACM), 47(3), pp. 42–49. DOI: doi.org/10.1145/2553070.2553077.

Pan, S. et al., 2024. Classification of Power Quality Disturbances Using Resnet with Channel Attention Mechanism, arXiv preprint.

Perez-Anaya, E. et al., 2024. Methodology for the Detection and Classification of Power Quality Disturbances Using CWT and CNN, Energies, 17(4). DOI: doi.org/10.3390/en17040852.

PNY Technologies Inc., 2018. NVIDIA T4. Available at: https://www.pny.com/nvidia-tesla-t4 (Accessed: 26 August 2025).

Rabczuk, G., Łosiński, J. and Cenian, A., 2020. Description and analysis of methods for calculation of CO_2 emission.

SEED - Power Quality Disturbance Dataset, 2023. Available at: https://www.kaggle.com/datasets/sumairaziz/seed-power-quality-disturbance-dataset/data (Accessed: 8 August 2025).

Sekar, K. et al., 2022. An Improved Power Quality Disturbance Detection Using Deep Learning Approach, Mathematical Problems in Engineering, 2022(1). DOI: doi.org/10.1155/2022/7020979.

Singh, G. K., 2009. Power system harmonics research: a survey, Euro. Trans. Electr. Power, 19, pp. 151–172.

Topaloglu, I., 2023. Deep Learning Based a New Approach for Power Quality Disturbances Classification in Power Transmission System, Journal of Electrical Engineering and Technology. Springer Nature Singapore, 18(1), pp. 77–88. DOI: doi.org/10.1007/s42835-022-01177-1.

Wang, J., Xu, Z. and Che, Y., 2019. Power Quality Disturbance Classification Based on Compressed Sensing and Deep Convolution Neural Networks, IEEE Access. IEEE, 7(c), pp. 78336–78346. DOI: doi.org/10.1109/ACCESS.2019.2922367.

Wang, S. and Chen, H., 2019. A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network, Applied Energy. Elsevier, 235, pp. 1126–1140. DOI: doi.org/10.1016/j.apenergy.2018.09.160.

Wen, Q. et al., 2023. Transformers in Time Series: A Survey, IJCAI International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence, 2023-August, pp. 6778–6786. DOI: doi.org/10.24963/IJCAI.2023/759.

Yiğit, E. et al., 2021. Automatic Detection of Power Quality Disturbance Using Convolutional Neural Network Structure with Gated Recurrent Unit, Mobile Information Systems, 2021(Ml). DOI: doi.org/10.1155/2021/7917500.

Zerveas, G. et al., 2021. A Transformer-based Framework for Multivariate Time Series Representation Learning, Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, pp. 2114–2124. DOI: doi.org/10.1145/3447548.3467401.

Zhou, H. et al., 2021. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting, Proceedings of the AAAI Conference on Artificial Intelligence. Association for the Advancement of Artificial Intelligence, 35(12), pp. 11106–11115. DOI: doi.org/10.1609/AAAI.V35I12.17325.