Article

An Ensemble Deep CNN Approach for Power Quality Disturbance Classification: A Technological Route Towards Smart Cities Using Image-Based Transfer

Mirza Ateeq Ahmed Baig [1], Naeem Iqbal Ratyal [1], Adil Amin [1], Umar Jamil [2,*], Sheroze Liaquat [3], Haris M. Khalid [4,5,*] and Muhammad Fahad Zia [6,*]

1   Department of Electrical Engineering, Mirpur University of Science & Technology (MUST), Mirpur 10250, Pakistan; ateeq.ee@must.edu.pk (M.A.A.B.); naeemratyal@must.edu.pk (N.I.R.); adil.pe@must.edu.pk (A.A.)
2   Department of Electrical and Computer Engineering, University of Texas at San Antonio (UTSA), San Antonio, TX 78249, USA
3   Eaton Research Labs, Eaton Corporation, Golden, CO 80401, USA; sherozeliaquat@eaton.com
4   College of Engineering and Information Technology, University of Dubai, Dubai 14143, United Arab Emirates
5   Department of Electrical and Electronic Engineering Science, University of Johannesburg, Johannesburg 2006, South Africa
6   Department of Electrical and Computer Engineering, American University in Dubai, Dubai 28282, United Arab Emirates
*   Correspondence: umar.jamil@my.utsa.edu (U.J.); mkhalid@ud.ac.ae (H.M.K.); mzia@aud.edu (M.F.Z.)

Abstract: The abundance of powered semiconductor devices has increased with the introduction of renewable energy sources into the grid, causing power quality disturbances (PQDs). This represents a huge challenge for grid reliability and smart city infrastructures. Accurate detection and classification are important for grid reliability and consumers' appliances in a smart city environment. Conventionally, power quality monitoring relies on trivial machine learning classifiers or signal processing methods. However, recent advancements have introduced Deep Convolution Neural Networks (DCNNs) as promising methods for the detection and classification of PQDs. These techniques have the potential to demonstrate high classification accuracy, making them a more appropriate choice for real-time operations in a smart city framework. This paper presents a voting ensemble approach to classify sixteen PQDs, using the DCNN architecture through transfer learning. In this process, continuous wavelet transform (CWT) is employed to convert one-dimensional (1-D) PQD signals into time–frequency images. Four pre-trained DCNN architectures, i.e., Residual Network-50 (ResNet-50), Visual Geometry Group-16 (VGG-16), AlexNet and SqeezeNet are trained and implemented in MATLAB, using images of four datasets, i.e., without noise, 20 dB noise, 30 dB noise and random noise. Additionally, we also tested the performance of ResNet-50 with a squeeze-and-excitation (SE) mechanism. It was observed that ResNet-50 with the SE mechanism has a better classification accuracy; however, it causes computational overheads. The classification performance is enhanced by using the voting ensemble model. The results indicate that the proposed scheme improved the accuracy (99.98%), precision (99.97%), recall (99.80%) and F1-score (99.85%). As an outcome of this work, it is demonstrated that ResNet-50 with the SE mechanism is a viable choice as a single classification model, while an ensemble approach further increases the generalized performance for PQD classification.

Keywords: ensemble learning; deep convolution neural networks; deep learning; machine learning; power quality disturbances; revolutionary technologies; smart cities; transfer learning

1. Introduction

1.1. Motivation

The widespread integration of renewable energy sources along with the extensive introduction of non-linear loads and semiconductor devices in the power system causes the degradation of power quality in terms of distortions in voltage, current, and frequency. Power quality disturbances (PQDs) lead to system protection malfunction and system instabilities and limit the lifetime of electrical equipment [1]. As a result, these disruptions significantly impact the reliability of the power grid. In smart city environments, where interconnected devices and systems rely on stable power, poor power quality can result in decreased productivity, compromised safety, and increased manufacturing costs [2]. In the context of a smart grid, as shown in Figure 1, power quality issues across the generation, transmission, and distribution systems are primarily caused by (1) intermittent renewable energy sources, (2) the presence of power electronics, 3) system faults, (4) variations in load, (5) switching operations, (6) imbalances in load distribution, and (7) non-linear loads. The non-linear loads include (1) variable-speed drives, (2) computers, (3) fluorescent lighting, and (4) other electronics. By addressing these issues, smart cities can ensure reliable energy delivery, enhance operational efficiency, and improve the overall quality of life for residents. Hence, compliance with power quality standards is integral to monitoring setups, which require accurate detection and classification of various kinds of PQDs, including sag, swell, flicker, interruption, harmonics, and transients. This critical problem motivates research scholars to design and investigate a framework for the correct classification of PQDs. In the context of smart cities, the integration of advanced PQD classification methods into smart grid communication systems is essential for ensuring grid stability. Smart grids require efficient and reliable data exchange between key components such as distributed energy resources, smart meters, and control systems. This enables quicker decision-making and response to grid anomalies, ultimately improving the resilience and efficiency of smart city infrastructures. Similarly, interoperability between diverse systems and devices is crucial for ensuring the smooth operation of smart grids. These devices must be able to communicate and function together seamlessly. A customary communication standard being followed in this work ensures compatibility across different platforms and enables smooth integration into existing infrastructures. Furthermore, the security of data traffic within smart grids is also a critical concern, especially given the increasing volume of sensitive information exchanged between components. By implementing robust encryption, access control, blockchain for data integrity, secure communication protocols, and regular security audits, this study ensures that both the PQD data and classification results remain secure throughout the entire classification process. These security measures support the proposed method and make it resilient to cyber threats in smart grid applications.
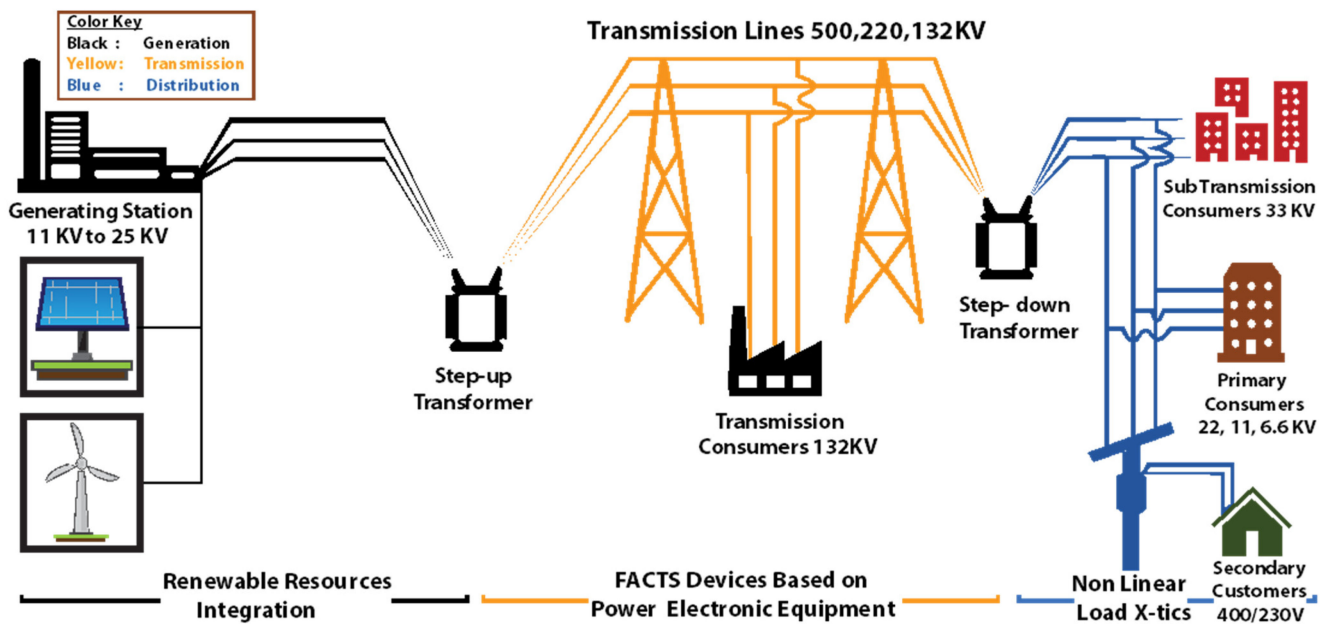
Figure 1. Power quality disturbance sources in smart grid.

1.2. Related Work

Researchers have shown great interest in developing more accurate methods for detecting and classifying PQDs. According to the literature, the classification of PQDs can be divided into two different types based on the techniques employed: (i) signal processing with intelligent classifier-based methods and (ii) deep learning techniques. Various signal transformation techniques including S-transform (ST) [3], Fourier transform (FT) [4], wavelet transform (WT) [4,5], Hilbert–Huang transform (HHT) [6], and Decomposition methods [7] are used for processing stationary and non-stationary PQD signals for features extraction. ST is a widely used signal processing technique for feature extraction because of its suitability in dealing with noisy signals; however, it involves a lot of computational effort. On the other hand, FT is not a good choice for non-stationary PQDs due to the set window resolution. In a noisy environment, WT does not perform well in the extraction of distinguished PQD features. HT emerged as a promising tool for the time–frequency-based classification of PQDs with increased computational complexity. The increasing density of PQDs due to the occurrence of multiple PQDs limits the efficacy of signal processing techniques to meet the classification requirements [8]. Additionally, these methods face the challenges of balancing resolution, requirements of feature engineering, and classifier design issues. The literature reported various studies on classifiers, such as the Artificial Neural Network (ANN) [9,10], Decision Tree (DT) [11,12], Bayesian Network (BN) [13], Fuzzy Logic (FL) [14], Intelligent Expert System (IES) [15], and support vector machine (SVM) [16,17], used to establish a correlation between extracted features and unique targets for the classification of PQDs. The choice of a classifier for PQD classification largely depends on its ability to adapt to various disturbances and maintain computational stability.

Classical classifiers have shown satisfactory performance in the recognition of PQDs. However, due to shallowness, deep learning methods are replacing them to meet evolving classification needs. Deep learning models such as long short-term memory (LSTM) networks [18,19] and CNNs [20,21] have shown dominance over classical methods due to their high-level feature learning and enhanced mapping capabilities. Compared to LSTM, CNNs perform better in capturing spatial and temporal features, making them suitable choices for tasks that require instantaneous feature extraction [22]. Recently, extensive research has been carried out to explore the potential of CNNs to monitor and classify PQDs. Studies using CNN models for PQD classification can be divided into two categories: (i) PQD classification based on signal data and (ii) PQD classification based

on image data. The authors of [23] conducted a comparative analysis of different CNN models for identifying PQDs with synthetic and recorded signals. The results showed that the classifier achieved higher performance with synthetic data and lower accuracy with real data. Jidong Wang et al. [24] proposed a bagging LSTM approach classification of fifteen different types of PQ signals. The results showed that the proposed approach requires less training time without compromising accuracy. The authors of [25] developed a four-stage disturbance detection system for six types of voltage disturbance signals. The first two stages are designated as data processing stages. The third stage deals with the training of CNN models, and an experimental circuit for analysis of online disturbance is presented in the fourth stage. The experimental results established that the proposed system is architecturally efficient. In another work [26], a squeeze-and-excitation network (SENet) is presented for self-learning of each channel feature. The authors claim to have achieved an average accuracy of 98.95% in noisy environments. On the other hand, there are limited studies in the literature that deal with PQD classification using their images. In the study by Santhosh Manikonda et al. [27], a transfer learning approach is used for classifying PQDs through image classification. VGG-16 is used as a pre-trained model to classify five PQDs through scalograms. The results confirmed the efficacy of the proposed strategy. In another work [28], the authors developed a deep transfer learning framework for PQD classification. A small dataset containing voltage waveform images was used to train the model and classify four different types of PQDs. With the limited training dataset, the model achieved high accuracy for PQD classification. A signal transformation from 1-D to two-dimensional (2-D) time–frequency images is performed in [29]. A CNN-based deep learning model is employed to classify sixteen different types of PQDs. Deep learning models are extremely important for generalization and problem learning. Additionally, the generalization ability of a single classification model can be enhanced by combining multiple classification models via an ensemble approach. The idea of this approach is to enhance the accuracy and stability of the overall model, thereby achieving better performance. Some studies [30–32] have employed an ensemble approach to classify various PQDs. However, ensembles of shallow learners are trained using PQD signal data instead of deep learning models trained on PQD image data. The stated work reveals that using an ensemble approach with deep learning models is a promising way to improve the accuracy, robustness, and generalization ability of the PQD classification model. This strategy has the potential to simplify the training process, reducing complexity and time while maintaining a high classification performance.

This paper proposes a voting ensemble model that utilizes the potential of pertained DCNNs by using time–frequency images of various types of PQDs as training samples. The proposed approach decreases the complexity of the training course and achieves high classification performance.

1.3. Research Gap

Although machine learning models have demonstrated effectiveness in PQD classification, there are several challenges to consider:

Firstly, the classification of PQDs is primarily treated as a signal processing problem. In this regard, machine learning methods process time series signals as the input and require feature engineering to extract domain-specific information. Exclusion of frequency domain information when processing time domain signals can limit classification accuracy because important features may be missed. Moreover, training a classification model from scratch requires a large dataset and computational resources, resulting in a longer training time. Additionally, tuning hyperparameters is a challenging task, especially to mitigate overfitting.

Secondly, the classification performance of a single model often faces generalization problems across different datasets, which can be improved by integrating several deep learning models.

These challenges can be addressed by transforming the time domain signals into time–frequency images utilizing continuous wavelet transform (CWT). The proposed approach instigates an ensemble technique employing DCNNs with transfer learning. The pre-trained DCNN architecture offers significant advantages in image classification, showing robustness in capturing significant features that are usually ignored in conventional techniques. Moreover, the need for feature engineering is avoided by simultaneously performing feature learning and classification. Hence, converting PQD signals to images allows the model to visually classify disturbances and consider complex characteristics that are not present in time series data. Additionally, the generalization capability of individual learners can be improved by utilizing an ensemble of various DCNNs.

Table 1 presents a summary of the reported work of the most recent studies in the literature. It is evident from Table 1 that there are only a few studies that have used both time and frequency features for PQD classification, i.e., that consider the PQD classification problem as an image processing problem. Moreover, there is a notable gap in leveraging ensemble DCNN architectures to classify PQDs using both time and frequency features for improved accuracy.

Table 1. Summary of recent studies related to PQD classification [1].

| Ref. | Publication Year | Methodology | | No. of PQD Classes | Features | |
		Ensemble Approach	Unified Model Approach		Time Domain	Frequency Domain
[12]	2021	×	DWT, MLP, SVM	9	√	×
[30]	2021	DWT, LR, NB, DT,	×	9	√	×
[33]	2021	×	Hybrid CNN	13	√	√
[24]	2022	Bagging-LSTM	×	15	√	×
[28]	2022	×	CNN	3	√	√
[34]	2023	×	CNN-LSTM	14	√	×
[29]	2023	×	S-transform- CNN	16	√	×
[26]	2023	×	HT-CNN	16	√	×
[35]	2023	×	CWT-CNN	7	√	×

[1] Here, DWT, LR, NB, DT, LSTM, MLP, SVM, CNN, HT, and CWT are acronyms of the discrete wavelet transform, linear regression, naïve Bayes, discrete time, long short-term memory, multi-layer perceptron, support vector machine, convolution neural network, heterogenous, and continuous wavelet transform, respectively.

1.4. Problem Statement

Accurate classification of PQDs is essential for grid reliability and stability. Traditional methods for classifying PQDs are based on time series signals, which ignore information in the frequency domain. The time–frequency information is important for feature enhancement. Moreover, the performance of a single classification model often suffers from generalization issues across dissimilar datasets. Therefore, there is a need for an inclusive approach that transforms time domain signals to the time–frequency domain to improve feature extraction and uses ensemble deep learning methods to achieve accurate PQD classification.

1.5. Contributions

This work contributes to this field as follows:

(1) We present a method for transforming time domain PQD signals to time–frequency domain images based on CWT. This transformation allows deep models to more effectively identify and extract high-level disturbance features.

(2) We propose an ensemble classification framework based on transfer learning with DCNN models to classify PQDs using time–frequency images. The framework includes four pre-trained DCNN models, ResNet-50, VGG-16, AlexNet, and SqueezeNet, which were selected after rigorous experimental evaluation. We evaluate their performance across a spectrum of sixteen different PQD classes.

(3) The proposed ensemble approach uses the voting approach to improve the accuracy and generalization capabilities of individual classifiers. This method aggregates predictions from multiple classifiers using a voting scheme.

The paper is ordered as follows: the proposed approach is given in Section 2. The experiments along with the results are discussed in Section 3, while the study is concluded and future work is proposed in Section 4.

2. Proposed Methodology

An overview of the proposed method for classifying PQDs is shown in Figure 2. It consists of five essential steps: (1) PQD dataset generation; (2) time–frequency transformation; (3) DCNN-based transfer learning; (4) ensemble learning; and (5) evaluation. In the first step, various PQD signals are generated by using the open-source PQD signal generator given in [36]. Once the data are formulated, the signals are then transformed into a time–frequency spectrum of signals using CWT. In the third step, various pre-trained DCNN models including ResNet-50, VGG-16, AlexNet, and SqueezeNet are employed as classification models. After this, a voting-based ensemble model is exercised to achieve the final classification of PQDs. Finally, the model's performance is evaluated by considering various performance evaluation matrices. The details of each step involved in the proposed framework are described in the following sub-sections.
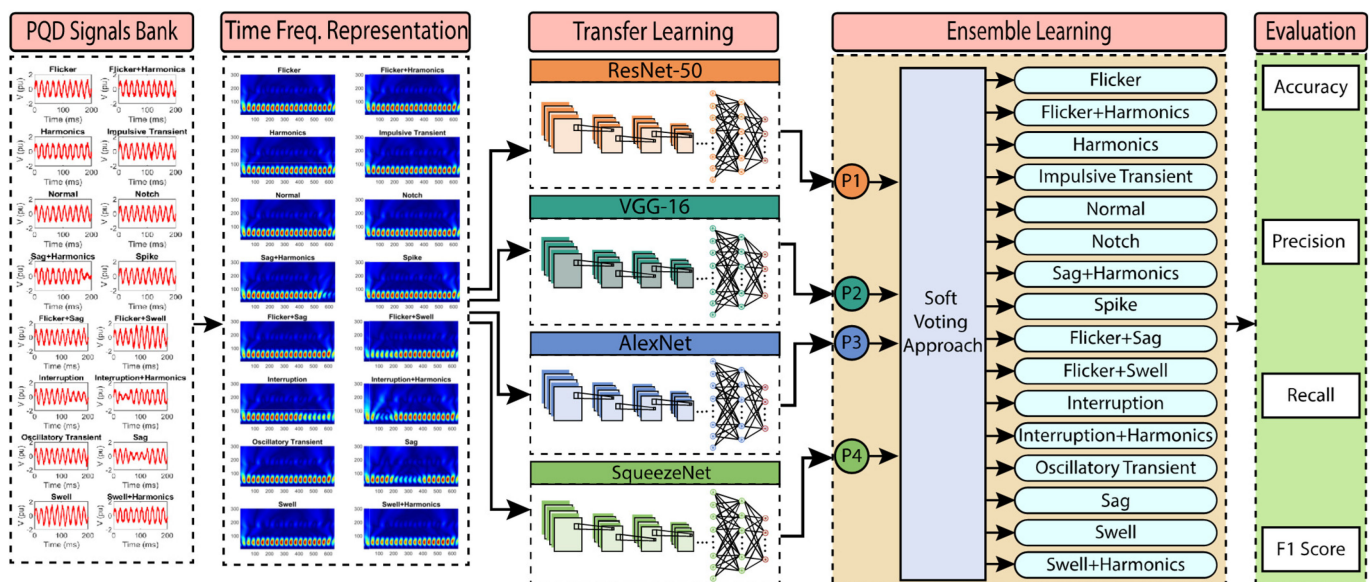
Figure 2. The proposed ensemble classifier based on DCNN models for PQD classification. Here, PQD and ResNet are abbreviations of power quality disturbance and residual neural network, respectively.

2.1. PQDs Dataset Generation

In this work, a dataset of sixteen different kinds of PQDs including single and multiple disturbances is constructed according to the IEEE-1159, EN 50160, and IEC 61000 standards [37]. It has been widely used in previous studies [36,38] to evaluate classifier performance. The parameters and their specifications are given in Table 2 and configured using an open-source PQD dataset generator [36].

Table 2. Parameters and their specifications for PQD dataset generation [36].

Parameters	Specifications		
Number of PQD classes	16		
PQD class, characteristics, equation and parameter range	Flicker (D1)	$[1 + \alpha f \sin(\omega t)] \sin(\omega t)]$	$0.1 \leq \alpha f \leq 0.2, 5 \leq \beta \leq 20\ Hz$
	Flicker + Harmonics (D2)	$[1 + \alpha f \sin(\beta \omega t)] \times [\alpha 1 \sin(\omega t) + \alpha 3 \sin(3\omega t) + \alpha 5 \sin(5\omega t) + \alpha 7 \sin(7\omega t)]$	$0.1 \leq \alpha f \leq 0.2, 5 \leq \beta \leq 20$ $0.05 \leq \alpha 3, \alpha 5, \alpha 7 \leq 0.15, \Sigma(\alpha_i^2) = 1$
	Flicker + Sag (D3)	$[1 + \alpha f \sin(\beta \omega t)][1 - \alpha(u(t - t1) - u(t - t2))] \sin(\omega t)]$	$0.1 \leq \alpha f \leq 0.2, 5 \leq \beta \leq 20$ $0.1 \leq \alpha \leq 0.9, T \leq (t2 - t1) \leq 9T$
	Flicker + Swell (D4)	$[1 + \alpha f \sin(\beta \omega t)][1 + \alpha(u(t - t1) - u(t - t2))] \sin(\omega t)]$	$0.1 \leq \alpha f \leq 0.2, 5 \leq \beta \leq 20$ $0.1 \leq \alpha \leq 0.8, T \leq (t2 - t1) \leq 9T$
	Harmonics (D5)	$\alpha 1 \sin(\omega t) + \alpha 3 \sin(3\omega t) + \alpha 5 \sin(5\omega t) + \alpha 7 \sin(7\omega t)$	$0.05 \leq \alpha 3, \alpha 5, \alpha 7, \leq 0.15, \Sigma(\alpha_i^2) = 1$
	Impulsive transient (D6)	$[1 - \alpha(u(t - t1) - u(t - t2))] \sin(\omega t)]$	$0.1 \leq \alpha \leq 0.414, T/20 \leq (t2 - t1) \leq T/10$
	Interruption (D7)	$[1 - \alpha(u(t - t1) - u(t - t2))] \sin(\omega t)]$	$0.9 \leq \alpha \leq 1, T \leq (t2 - t1) \leq 9T$
	Interruption + Harmonics (D8)	$[1 - \alpha(u(t - t1) - u(t - t2))] \times [\alpha 1 \sin(\omega t) + \alpha 3 \sin(3\omega t) + \alpha 5 \sin(5\omega t) + \alpha 7 \sin(7\omega t)]$	$0.9 \leq \alpha \leq 1, T \leq (t2 - t1) \leq 9T$ $0.05 \leq \alpha 3, \alpha 5, \alpha 7 \leq 0.15, \Sigma(\alpha_i^2) = 1$
	Normal (D9)	$[1 \pm \alpha(u(t - t1) - u(t - t2))] \sin(\omega t)$	$\alpha < 0.04, T \leq (t2 - t1) \leq 9T$
	Notch (D10)	$\sin(\omega t) - sign(\sin(\omega t)) \times \Sigma k[u(t - (t1 - 0.02n)) - u(t - (t2 - 0.02n))]$	$0 \leq t1, t2 \leq 0.5T, 0.1 \leq K \leq 0.4, 0.01T \leq t2 - t1 \leq 0.05T$
	Oscillatory transient (D11)	$\sin(\omega t) + \alpha - (t - t1)/\tau \sin(\omega n(t - t1))(u(t2) - u(t1))$	$0.1 < \alpha \leq 0.8, 0.5T \leq (t2 - t1) \leq 3T,$ $8 \leq \tau \leq 40, 300 \leq 2\pi \omega n \leq 900$
	Sag (D12)	$[1 - \alpha(u(t - t1) - u(t - t2))] \sin(\omega t)$	$0.1 \leq \alpha < 0.9, T \leq (t2 - t1) \leq 9T$
	Sag + Harmonics (D13)	$[1 - \alpha(u(t - t1) - u(t - t2))] \times [\alpha 1 \sin(\omega t) + \alpha 3 \sin(3\omega t) + \alpha 5 \sin(5\omega t) + \alpha 7 \sin(7\omega t)]$	$0.1 \leq \alpha < 0.9, T \leq (t2 - t1) \leq 9T,$ $0.05 \leq \alpha 3, \alpha 5, \alpha 7 \leq 0.15, \Sigma(\alpha_i^2) = 1$
	Spike (D14)	$\sin(\omega t) + sign(\sin(\omega t)) \times \Sigma k[u(t - (t1 - 0.02n)) - u(t - (t2 - 0.02n))]$	$0 \leq t1, t2 \leq 0.5T, 0.1 \leq K \leq 0.4, 0.01T \leq t2 - t1 \leq 0.05T$
	Swell (D15)	$[1 + \alpha(u(t - t1) - u(t - t2))] \sin(\omega t)$	$0.1 \leq \alpha \leq 0.8, T \leq (t2 - t1) \leq 9T$
	Swell + Harmonics (D16)	$[1 + \alpha(u(t - t1) - u(t - t2))] \times [\alpha 1 \sin(\omega t) + \alpha 3 \sin(3\omega t) + \alpha 5 \sin(5\omega t) + \alpha 7 \sin(7\omega t)]$	$0.1 \leq \alpha < 0.8, T \leq (t2 - t1) \leq 9T,$ $0.05 \leq \alpha 3,$ $\alpha 5, \alpha 7 \leq 0.15, \Sigma(\alpha_i^2) = 1$
Samples for each class	500		
Reference frequency	50 Hz		
Sampling frequency	3.2 kHz		
Number of cycles/class sample	10		
Magnitude of the signal	1 p.u.		
Noise levels	20 dB, 30 dB and random noise		

This process results in a dataset with dimensions of 8000 × 1600. To approximate the realistic conditions and enable comparative analysis, random noise in the range of 20–30 dB is added to the generated data. Figure 3 shows an example of the sixteen PQDs with 20 dB noise along with respective class information.
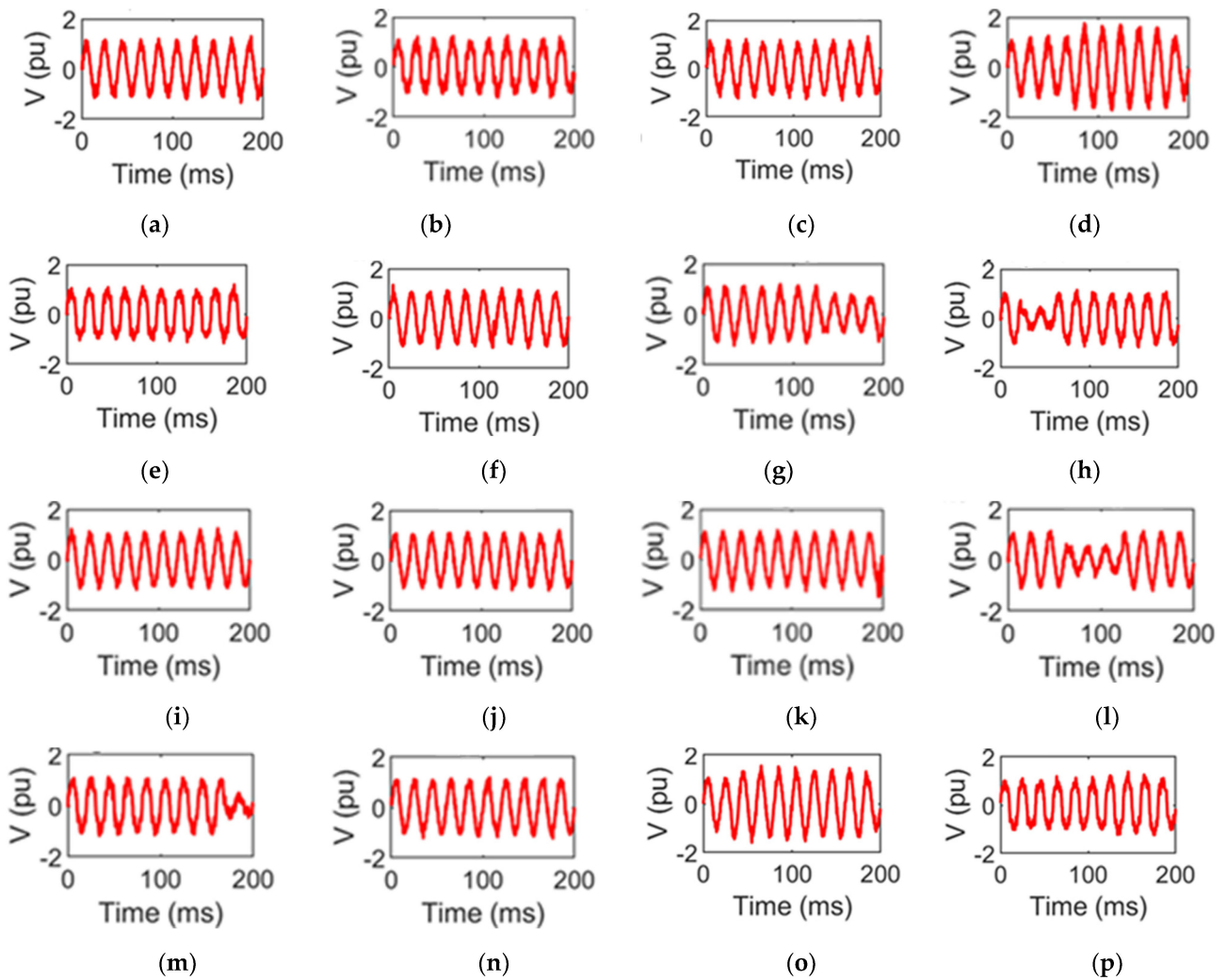
Figure 3. An example of PQDs with 20 dB noise: (**a**) flicker; (**b**) flicker + harmonics; (**c**) flicker + sag; (**d**) flicker + swell; (**e**) harmonics; (**f**) impulsive transient; (**g**) interruption; (**h**) interruption + harmonics; (**i**) normal; (**j**) notch; (**k**) oscillatory transient; (**l**) sag; (**m**) sag + harmonics; (**n**) spike; (**o**) swell; and (**p**) swell + harmonics.

2.2. Time–Frequency Transformation

Numerous signal processing techniques have been developed to obtain a time–frequency demonstration of a time domain signal. Among these techniques, CWT is widely recognized for its effectiveness in generating a time–frequency representation of transient signals. Due to the transient nature of most PQDs, CWT is particularly beneficial for PQD studies.

Let us assume that $\phi(t)$ is the mother wavelet, a set of wavelet basis functions can be constructed as Equation (1) [35].

$$\phi_{a,b} = \frac{1}{\sqrt{a}}\phi\left(\frac{t-b}{a}\right) \tag{1}$$

In this context, *a* represents the scaling factor and *b* denotes the translation time, with ϕ serving as the mother wavelet. The CWT for a given continuous signal $x(t)$ is expressed by Equation (2) [35].

$$CT(a,b)) = \frac{1}{\sqrt{a}}\left[\int_{-\infty}^{\infty} x(t)\phi\left(\frac{t-b}{a}\right)dt\right] \tag{2}$$

After the transformation, the amplitude scale of the wavelet coefficients is calculated by Equation (3) [35].

$$A(a,b) = |C(a,b)| \tag{3}$$

After applying CWT, the time–frequency representations of the signals for Figure 3 are illustrated in Figure 4.
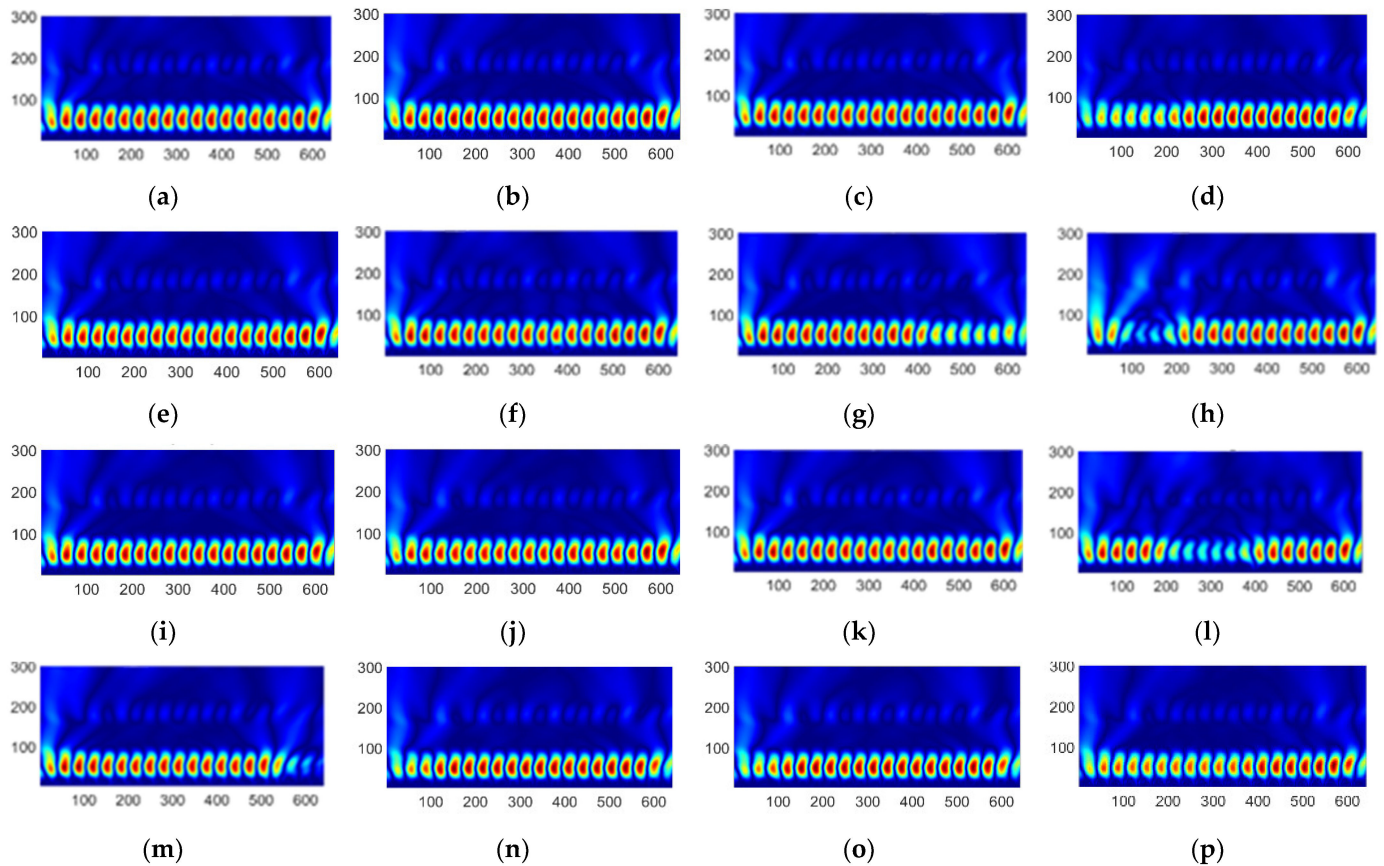
Figure 4. An example of a time–frequency representation of PQDs with 20 dB noise: (**a**) flicker; (**b**) flicker + harmonics; (**c**) flicker + sag; (**d**) flicker + swell; (**e**) harmonics; (**f**) impulsive transient; (**g**) interruption; (**h**) interruption + harmonics; (**i**) normal; (**j**) notch; (**k**) oscillatory transient; (**l**) sag; (**m**) sag + harmonics; (**n**) spike; (**o**) swell; (**p**) swell + harmonics.

2.3. DCNN Models

Currently, DCNNs are applied to a wide variety of applications in image classification tasks. In this work, four types of DCNNs were explored, i.e., ResNet-50, VGG-16, AlexNet and SqueezeNet, for the classification of PQDs. The basic architecture of each network is illustrated in sub-sections.

2.3.1. ResNet-50

ResNet-50 is a residual network deep learning model proposed by [39], shown in Figure 5. It has fifty layers, including a (1) convolutional layer, (2) max pooling layer, (3) deconvolution layer, and (4) Softmax layer. It accepts input images with a size of 227 × 227. ResNet-50 was selected for the PQD classification task due to its superior performance and effective solution to the vanishing gradient problem.
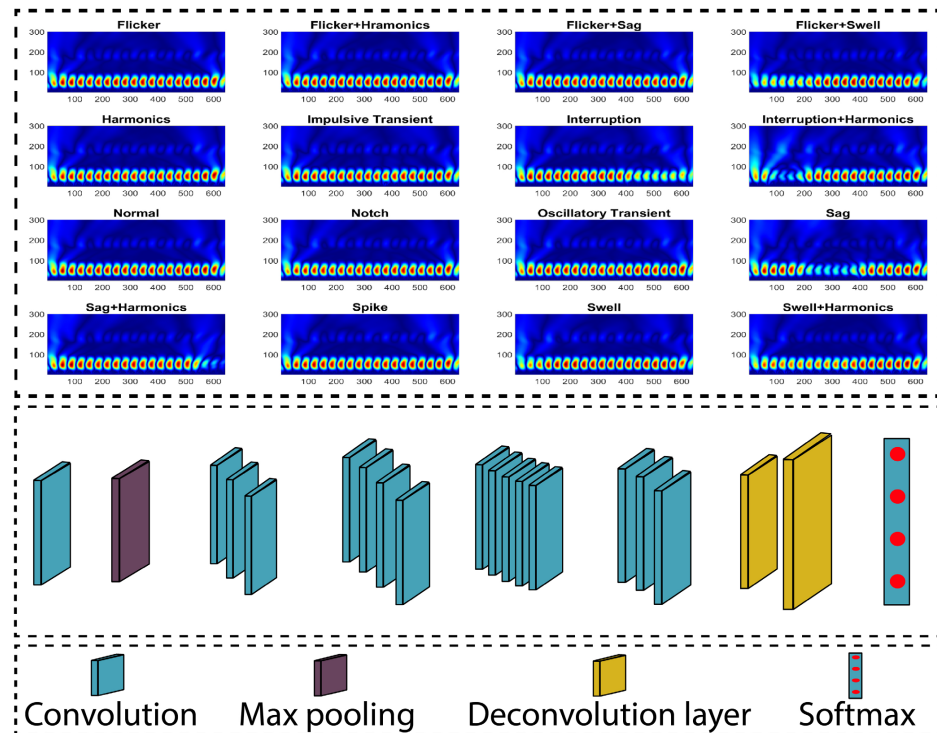
Figure 5. ResNet-50 architecture for PQD classification.

2.3.2. VGG-16

VGG16 [40] is a sixteen-layer deep model. It presents thirteen convolutional layers combined with three fully connected layers. The convolutional layers are further split into five parts, each with numerous convolutional layers and a max pooling layer. It accepts images with a size of 224×224, and its design is displayed in Figure 6.
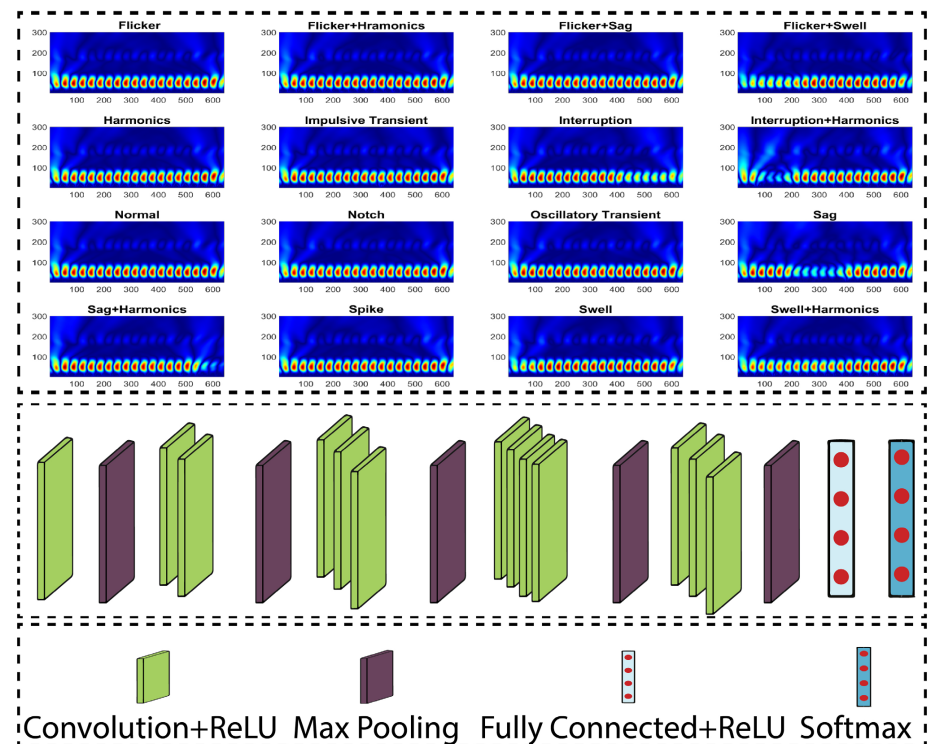
Figure 6. VGG-16 architecture for PQD classification.

2.3.3. AlexNet

According to [41], AlexNet comprises five convolutional layers, three pooling layers, two fully connected layers, and one Softmax layer. The network accepts input images with a size of 227 × 227. The basic design of AlexNet is given in Figure 7.
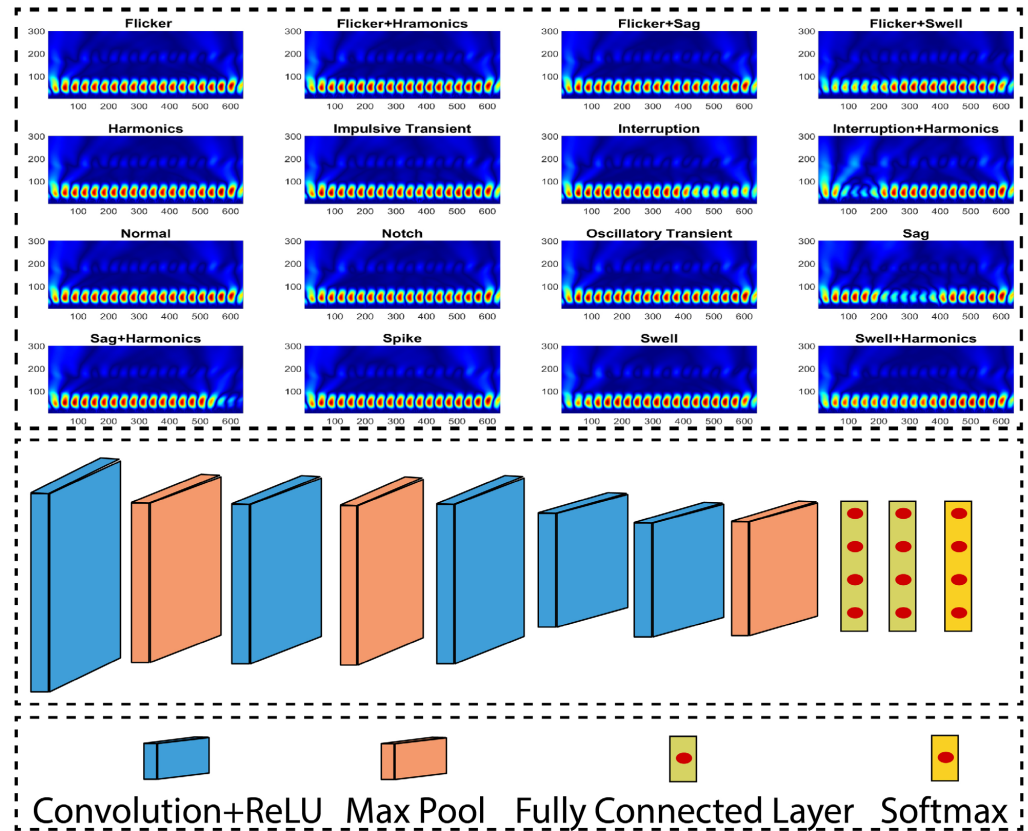
Figure 7. AlexNet model for PQD classification.

2.3.4. SqueezeNet

The SqueezeNet model was developed by [42] and consists of eighteen layers. The model has a convolutional layer at the start, followed by fire modules, and ends with an additional convolutional layer. The model supports various activation functions including ReLU, Tanh, and Sigmoid. The architecture of SqueezeNet used for PQD classification is illustrated in Figure 8.
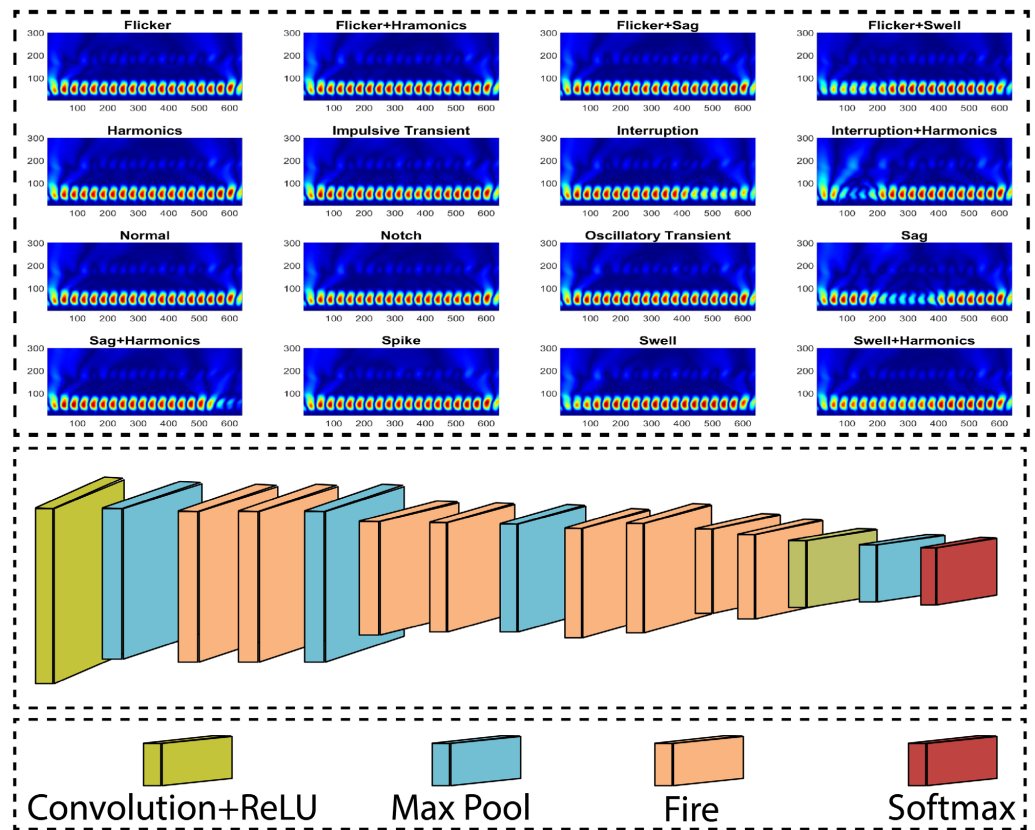
Figure 8. SqueezeNet architecture for PQDs classification. Here, ReLU is an acronym for activation function.

2.4. Soft Voting Ensemble Approach

The soft voting approach considers the confidence levels of each classifier's predictions. It converts each classifier's predictions into probabilities for the respective classes and then aggregates these probabilities to determine the final classification result. Using a voting ensemble approach often results in better performance than a single model. In this study, DCNN models were ensembled and the final decision was determined using a soft voting technique, as given by Equation (4).

$$\varepsilon_j = \arg max_m \sum_{i=1}^{n} \alpha_i Y_{j,m}^{i} \tag{4}$$

where n denotes the total number of classifiers, α_i represents the weight of the ith classifier, and $Y_{j,k}^{i}$ is the prediction probability of the ith classifier for the jth sample of the mth class.

2.5. Performance Evaluation Metrices

Accuracy assessment is very important in evaluating the classification performance of machine learning algorithms. The performance of the proposed ensemble classification model was valued using specific performance metrics selected from [43].

- Accuracy (A): This is the ratio of the model's true predictions to the overall prediction. Mathematically, it can be formulated as Equation (5).

$$Accuracy\ (A) = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{5}$$

where T_P denotes true positive, representing the accurately predicted positive cases. T_N represents true negatives, indicating the accurately predicted negative cases. F_P corresponds to false positives, showing negative events wrongly predicted as posi-

tive; F_N represents false negatives, representing the positive occurrences incorrectly predicted as negative.

- Precision (P): this denotes the ratio of accurately predicted positive occurrences out of the total number of predicted positive occurrences and is expressed as Equation (6).

$$Precision = \frac{T_P}{T_P + F_P} \tag{6}$$

- Recall (R): this refers to the proportion of accurately predicted positive instances among all instances in the class and can be stated as Equation (7).

$$Recall = \frac{T_P}{T_P + F_N} \tag{7}$$

- F1-score: this denotes a weighted mean of the precision and recall, formulated as Equation (8).

$$F1 - score = \frac{(2 \times P \times R)}{(P + R)} \tag{8}$$

3. Experimental Results and Discussion

This part discusses the experimental setup developed to execute the proposed methodology, followed by the obtained results and discussion. Finally, the proposed model's performance is compared with the literature.

3.1. Experimental Setup

In this sub-section, the PQD dataset of different noise levels is presented. In addition, the parameter settings of the DCNN models are also given in this part. The performance of the presented model is assessed on the synthetic database comprising sixteen different PQD signals, including the single and composite disturbances derived from [36] given in Section 2. The generated dataset consists of a total of 8000 signals, including 500 signals for each class. It is instrumental in adding noise to the PQD signals to create a more realistic dataset. In light of this, various noise levels such as 20 dB noise, 30 dB noise, and random noise are added to the original PQD signals of each class. Afterward, the 1-D signals are converted into time–frequency 2-D images by employing CWT. Thus, for each noise level, an image dataset of 500×16 is prepared, out of which 7200 images are used for training and validation and the remaining 800 images are reserved for testing the proposed framework. Various experiments with different training options have been performed to achieve the optimal performance of each deep model. The best settings of the parameters for each deep model are given in Table 3.

Table 3. Training options for DCNN models.

DCNN Model	Optimizer	Training Parameters							
		Hyperparameter with Search Space			Optimized Value				
		Learning Rate	Batch Size	Epoch	Learning Rate	Batch Size	Epoch	Number of Layers	Input Image Size (Pixel)
ResNet-50	SGD							177	224 × 224
VGG-16	SGD							41	224 × 224
AlexNet	SGD	[0.01, 0.001, 0.00015]	[16, 32, 48]	[10, 20, 30]	0.0001	32	30	25	227 × 227
SqueezeNet	SGD							68	227 × 227
ResNet-50 with attention mechanism	SGD							177	224 × 224

In deep learning, hyperparameter tuning plays a critical role in optimizing the performance of models. The minimum batch size, max epochs, and learning rate were considered for hyperparameter optimization for the different models in this study, as shown in Table 2.

A random search algorithm was performed, starting with a default learning rate of 0.001, and fine-tuned through multiple iterations, eventually selecting 0.0001 to ensure stable convergence without overshooting during training, while the mini batch size varied between 16 and 48 with steps of 16. Similarly, the model was tested with 10 epochs to observe the performance and convergence trends. After analyzing the validation performance, 30 epochs were settled on to avoid overfitting, as further training did not significantly improve accuracy. By using pre-trained CNN architectures and fine-tuning only the final layers, the robustness of these models was utilized to avoid overfitting the dataset. By following this comprehensive hyperparameter tuning and avoiding an overfitting strategy, optimal performance for PQD classification was ensured, exhibiting a balance between performance and generalization. The proposed technique is implemented in MATLAB using a PC model Intel Core i9-9820X CPU (3.3.0 GHz), DDR4 32 GB RAM, and NVIDIA GeForce RTX 2080 8G GPU.

3.2. Training and Evaluation of DCNNs

The training and evaluation of each model for different noise levels are presented in the following sub-sections.

3.2.1. ResNet-50 Classification Results

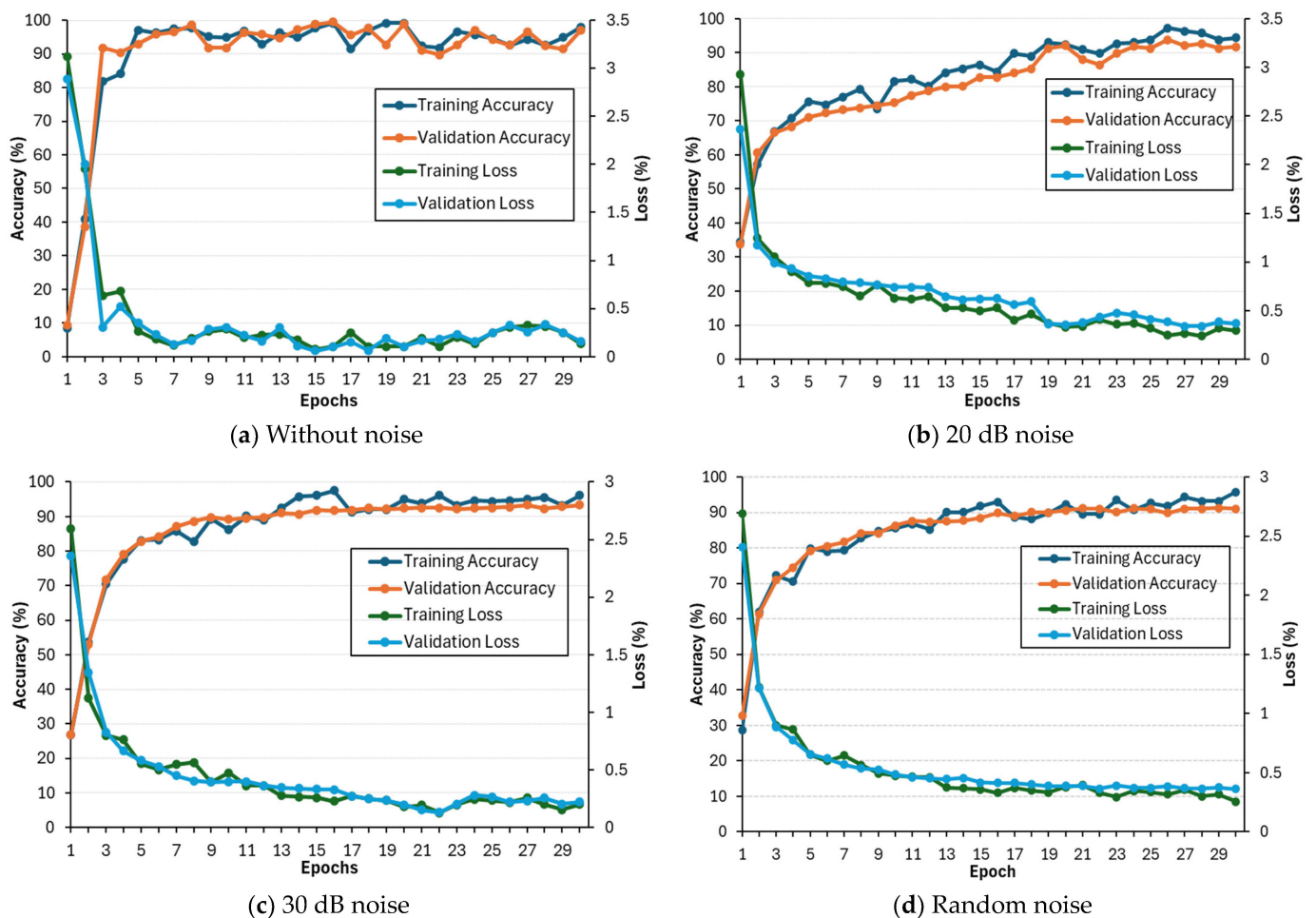The training performance of the ResNet-50 model for 30 epochs, with and without noise, is shown in Figure 9.

Figure 9. ResNet-50 training performance for noisy and noiseless datasets.

It is observed that under no noise conditions, the ResNet-50 achieved a maximum training and validation accuracy of 98.01% and 97.14%, respectively, with 0.133% training and 0.158% validation losses. Its performance degraded under noisy conditions, and for 20 dB noise, the training and validation accuracies decreased to 94.38% and 91.79%, respectively. The confusion matrices for noisy and noiseless datasets are illustrated in Figure 10, which demonstrates that the classification accuracy is relatively high for noiseless testing datasets as compared to other noisy datasets. The diagonal entries of the confusion matrix represent the true classification of testing instances, while the off-diagonal entries depict the misclassification of these samples. Additionally, accuracy, precision, recall, and F1-score were computed using Equations (4)–(7). The results indicate that ResNet-50 achieves superior performance on noiseless datasets, with an average value of accuracy (99.75%), precision (98.04%), recall (98%), and F1-score (98%), whereas with noisy test datasets, its performance decreases, resulting in an average value of accuracy (99.25%), precision (94.55%), recall (94%), and F1-score (94.13%) with 20 dB noise as shown in Table 4.
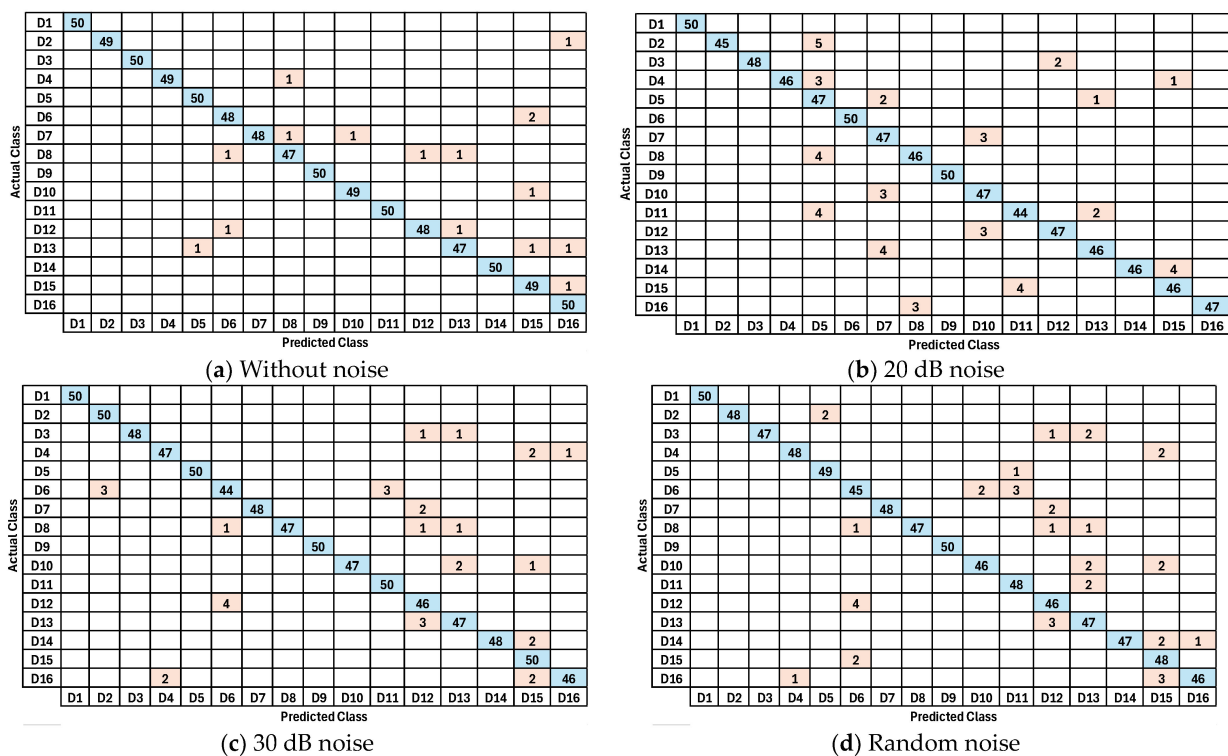
(a) Without noise

Actual \ Predicted	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16
D1	50															
D2		49														1
D3			50													
D4				49				1								
D5					50											
D6						48								2		
D7							48	1		1						
D8							1	47				1	1			
D9									50							
D10										49				1		
D11											50					
D12							1					48	1			
D13						1							47		1	1
D14														50		
D15															49	1
D16																50

(b) 20 dB noise

Actual \ Predicted	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16
D1	50															
D2		45					5									
D3			48										2			
D4				46	3										1	
D5					47		2					1				
D6						50										
D7							47			3						
D8				4				46								
D9									50							
D10							3			47						
D11				4							44		2			
D12							3					47				
D13				4									46			
D14														46	4	
D15											4				46	
D16									3							47

(c) 30 dB noise

Actual \ Predicted	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16
D1	50															
D2		50														
D3			48										1	1		
D4				47										2	1	
D5					50											
D6		3				44				3						
D7							48				2					
D8							1	47				1	1			
D9									50							
D10										47		2		1		
D11											50					
D12						4						46				
D13												3	47			
D14														48	2	
D15															50	
D16				2											2	46

(d) Random noise

Actual \ Predicted	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16
D1	50															
D2		48			2											
D3			47									1	2			
D4				48											2	
D5					49						1					
D6						45				2	3					
D7							48					2				
D8							1	47				1	1			
D9									50							
D10										46			2		2	
D11											48		2			
D12						4						46				
D13												3	47			
D14														47	2	1
D15						2									48	
D16				1											3	46

Figure 10. ResNet-50 confusion matrices for noisy and noiseless testing datasets.

Table 4. Performance evaluation of DCNNs and voting ensemble technique for noisy and noiseless test dataset.

Model	Without Noise				20 dB Noise				30 dB Noise				Random Noise			
	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ResNet-50	99.75	98.04	98	98	99.25	94.55	94	94.13	99.5	96.18	96	96.02	99.38	95.29	95	95.07
VGG-16	99.48	96.17	95.88	95.90	99.13	93.50	93	93.07	99.39	95.36	95.13	95.16	99.25	94.26	94.	94.05
AlexNet	99.38	95.21	95	95.01	98.91	91.73	91.25	91.23	99.08	93.23	92.63	92.71	98.94	92.05	91.49	91.60
SqueezeNet	98.75	90.75	90	90.01	98.59	89.31	88.75	88.75	98.66	90.10	89.25	89.30	98.59	89.21	88.75	88.75
ResNet-50 with SE mechanism	99.86	98.46	98	98.23	99.35	94.66	94	94.33	99.5	96.22	96	96.11	99.68	95.51	95	95.25
Voting Ensemble	**99.98**	**99.97**	**99.80**	**99.85**	**99.73**	**98.23**	**97.23**	**97.78**	**99.90**	**99.83**	**99.65**	**99.80**	**99.88**	**98.68**	**98.10**	**98.05**

3.2.2. VGG-16 Classification Results

The training performance of VGG-16 for 30 epochs, with and without noise, is shown in Figure 11. Without the noise dataset, VGG-16 achieved maximum training and validation accuracies of 96.20% and 93.50%, respectively, with 0.140% training and 0.190% validation losses. For 20 dB noise, the training and validation accuracies decreased to 93.50% and 92.30%, respectively. The confusion matrices for the noise and noiseless datasets are shown in Figure 12, which demonstrate that 99.48% is the highest testing accuracy achieved for the noiseless dataset followed by 99.39% for 30 dB noise, 99.25% for random noise and 99.13% for 25 dB noise. Furthermore, for the VGG-16 model, the precision, recall, and F1-score are also calculated. The results indicate that VGG-16 achieves superior performance on the noiseless test dataset, with (1) an average value of accuracy (99.48%), (2) precision (96.17%), (3) recall (95.88%), and (4) F1-score (95.90%), whereas the lowest performance is recorded with the 20 dB noise test dataset with the following average values: (1) accuracy (99.13%), (2) precision (93.50%), (3) recall (93%), and (4) F1-score (93.07%), as shown in Table 4.

(**a**) Without noise

(**b**) 20 dB noise

(**c**) 30 dB noise

(**d**) Random noise

Figure 11. VGG-16 training performance for noisy and noiseless datasets.

(**a**) Without noise

(**b**) 20 dB noise

(**c**) 30 dB noise

(**d**) Random noise

Figure 12. VGG-16 confusion matrices for noisy and noiseless testing datasets.

3.2.3. AlexNet Classification Results

The training performance of AlexNet for 30 epochs, with and without noisy datasets, is shown in Figure 13.

(**a**) Without noise

(**b**) 20 dB noise
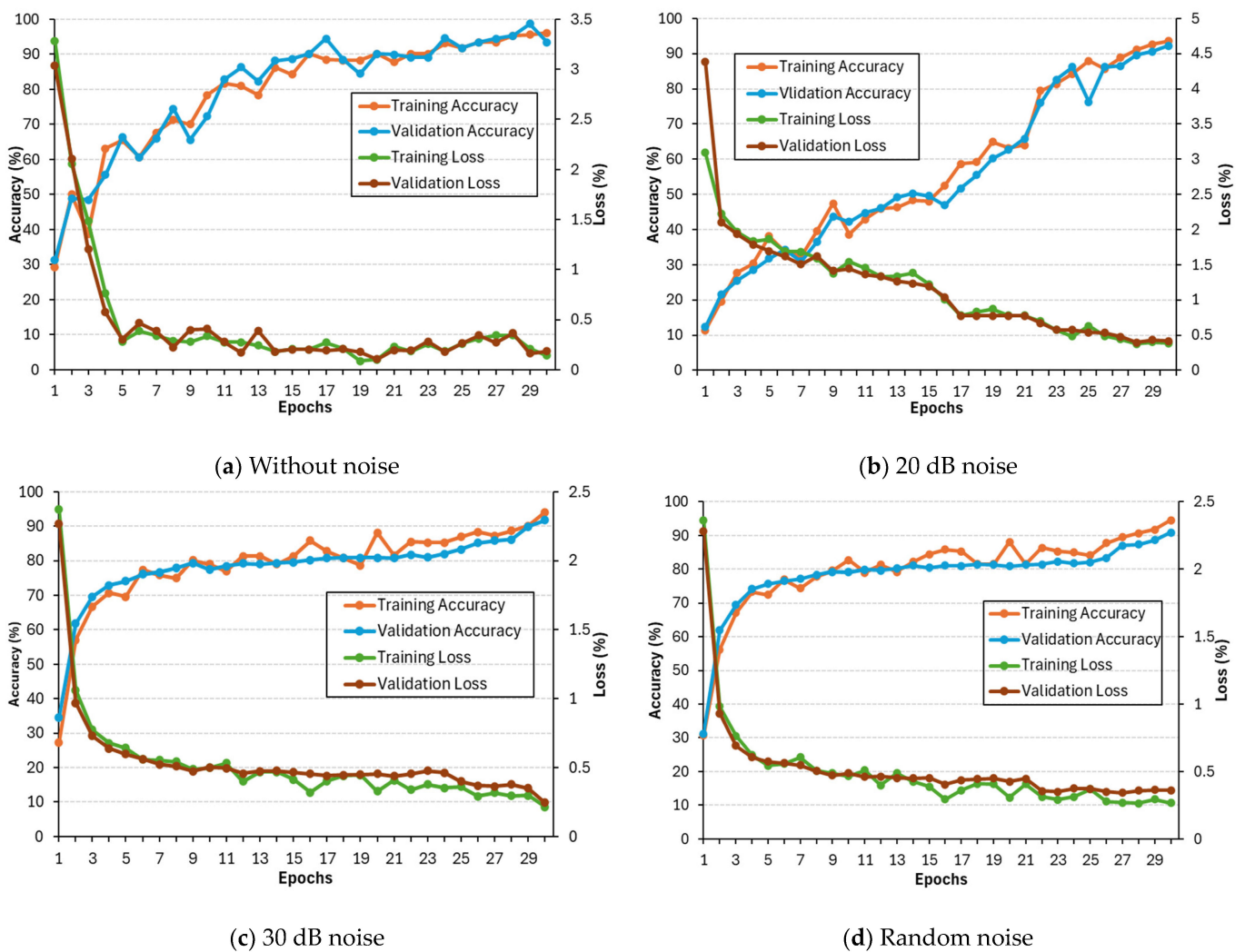
(**c**) 30 dB noise

(**d**) Random noise

Figure 13. AlexNet training performance for noisy and noiseless datasets.

The performance trend of AlexNet is in line with the ResNet-50 and VGG-16 models. It obtained maximum training and validation accuracies of 93.92% and 92.19%, respectively, with 0.205% training and 0.269%, validation losses. The training and validation accuracies decreased to 92.50% and 90.938%, respectively, for 20 dB noise. The confusion matrices for noisy and noiseless datasets are illustrated in Figure 14. The test accuracy is relatively high for noiseless datasets as compared to other noisy datasets. Moreover, the performance matrices including accuracy, precision, recall, and F1-score are given in Table 4. The results specify that AlexNet attains the highest performance for the noiseless test dataset, giving an average value of accuracy (99.38%), precision (95.21%), recall (95%), and F1-score (95.01%). In the case of noisy test datasets, it produces the lowest average results of accuracy (99.91%), precision (91.73%), recall (91.25%), and F1-score (91.23%) with 20 dB noise, as shown in Table 4.
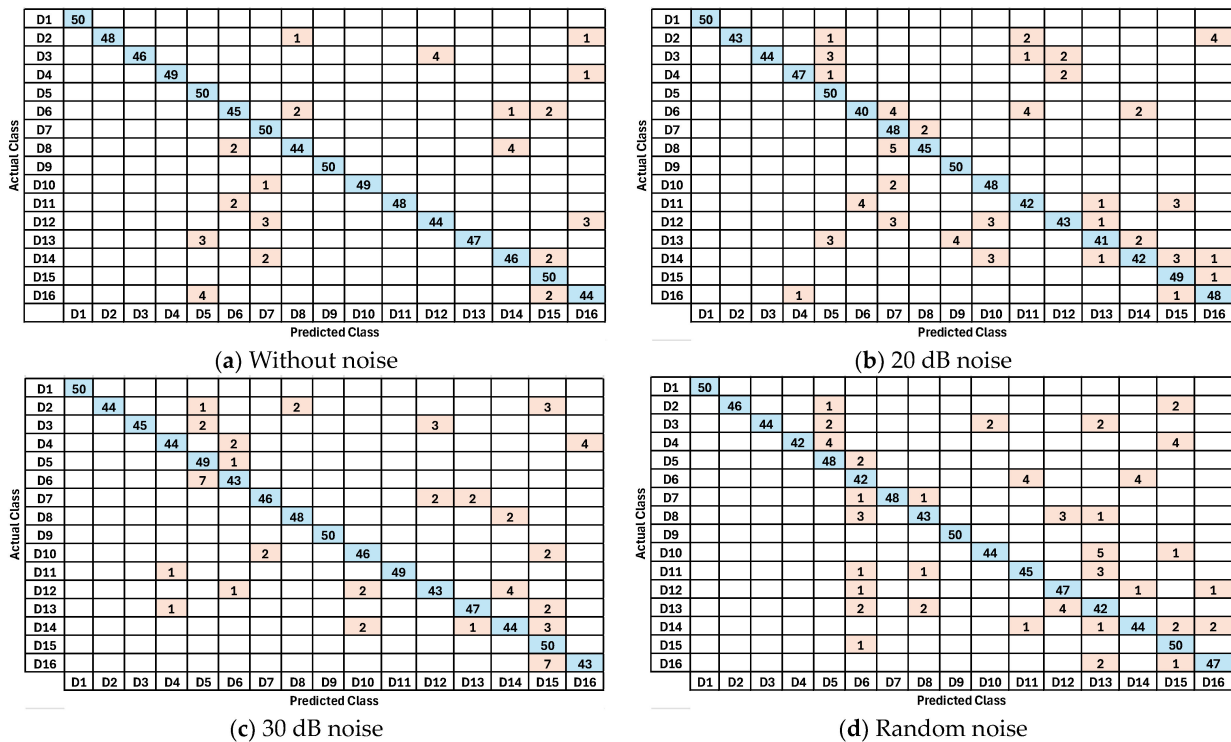
(a) Without noise

Actual Class	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16
D1	50															
D2		48								1						1
D3			46									4				
D4				49											1	
D5					50											
D6						45		2						1	2	
D7							50									
D8						2		44						4		
D9									50							
D10								1		49						
D11						2					48					
D12							3				44					3
D13					3							47				
D14							2							46	2	
D15															50	
D16						4									2	44

(b) 20 dB noise

Actual Class	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16
D1	50															
D2		43				1							2			4
D3			44			3						1	2			
D4				47		1								2		
D5					50											
D6						40	4						4			2
D7							48	2								
D8							5	45								
D9									50							
D10								2		48						
D11						4		3			42		1			
D12						3		3				43	1			
D13						3		4					41	2		
D14								3				1	42	3	1	
D15															49	1
D16						1									1	48

(c) 30 dB noise

Actual Class	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16
D1	50															
D2		44		1			2							3		
D3			45		2						3					
D4				44		2										4
D5					49	1										
D6						7	43									
D7							46						2	2		
D8								48						2		
D9									50							
D10								2		46				2		
D11				1							49					
D12						1				2		43		4		
D13				1									47		2	
D14										2		1		44	3	
D15															50	
D16															7	43

(d) Random noise

Actual Class	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16
D1	50															
D2		46				1									2	
D3			44			2					2			2		
D4				42		4										4
D5					48		2									
D6						42				4				4		
D7						1	48	1								
D8						3		43					3	1		
D9									50							
D10										44			5		1	
D11						1		1			45		3			
D12						1						47		1		1
D13						2		2				4	42			
D14										1		1		44	2	2
D15															50	
D16													2		1	47

Figure 14. AlexNet confusion matrices for noisy and noiseless testing datasets.

3.2.4. Ensemble Model Results

The classification accuracy of the deep model can be improved using the ensemble technique. For this purpose, an ensemble of the above-stated deep learning models has been used, utilizing the weights from the Softmax function for voting. The mean values of the performance matrices for every type of dataset obtained for all deep models and the proposed approach are given in Table 4.

The performance matrices, including accuracy, precision, recall and F1-score, for the developed model are given in Table 4. The voting ensemble approach showed a better outcome compared to the single DCNN models. Although the evaluation metrics for the noiseless dataset using individual deep models are closely related to the results of the ensemble approach, a notable difference between the proposed approach and DCNN models becomes evident when dealing with a noisy dataset. It improved the results for each case and produced the best results for noiseless cases, giving the following values: accuracy (99.98%), precision (99.97%), recall (99.80%), and F1-score (99.85%), while the lowest performance is recorded for 20 dB noise, with accuracy (99.73%), precision (98.23%), recall (97.23%), and F1-score (97.78%), as compared to other cases.

The performance of the ResNet-50 model was also evaluated with an SE attention mechanism for PQD classification. The SE mechanism improves feature learning by amplifying important features while suppressing less useful ones. The training results of ResNet-50 with and without the attention mechanism are shown in Figure 15. The model with SE produced comparatively better results than the standard ResNet-50. Both the training and validation accuracy are recorded as 98.71% and 98.40%, respectively, with 0.121% training and 0.144% validation losses. In noisy conditions (20 dB noise), the attention-enhanced model also performed better, giving 95.89% and 93.11% training and validation accuracy, respectively, as compared to the standard ResNet-50. Additionally, accuracy, precision, recall, and F1-score are calculated according to Equations (4)–(7) in Table 4, which demonstrate that ResNet-50 with the attention mechanism consistently delivers better results compared to the baseline model without any attention. However, despite improved performance, the integration of the SE mechanism introduces computational overheads. The attention module requires additional computational resources, leading to increased model complexity and longer training and inference times, as explained in Section 3.3. While the accuracy and feature learning is enhanced, the added computation may impact real-time applications, especially in systems like smart grids where rapid decision-making is crucial. Therefore, while ResNet-50 with the attention mechanism shows improved results, its computational overheads must be carefully considered, particularly in resource-constrained environments.

(**a**) Without noise

(**b**) 20 dB noise

(**c**) 30 dB noise

(**d**) Random noise
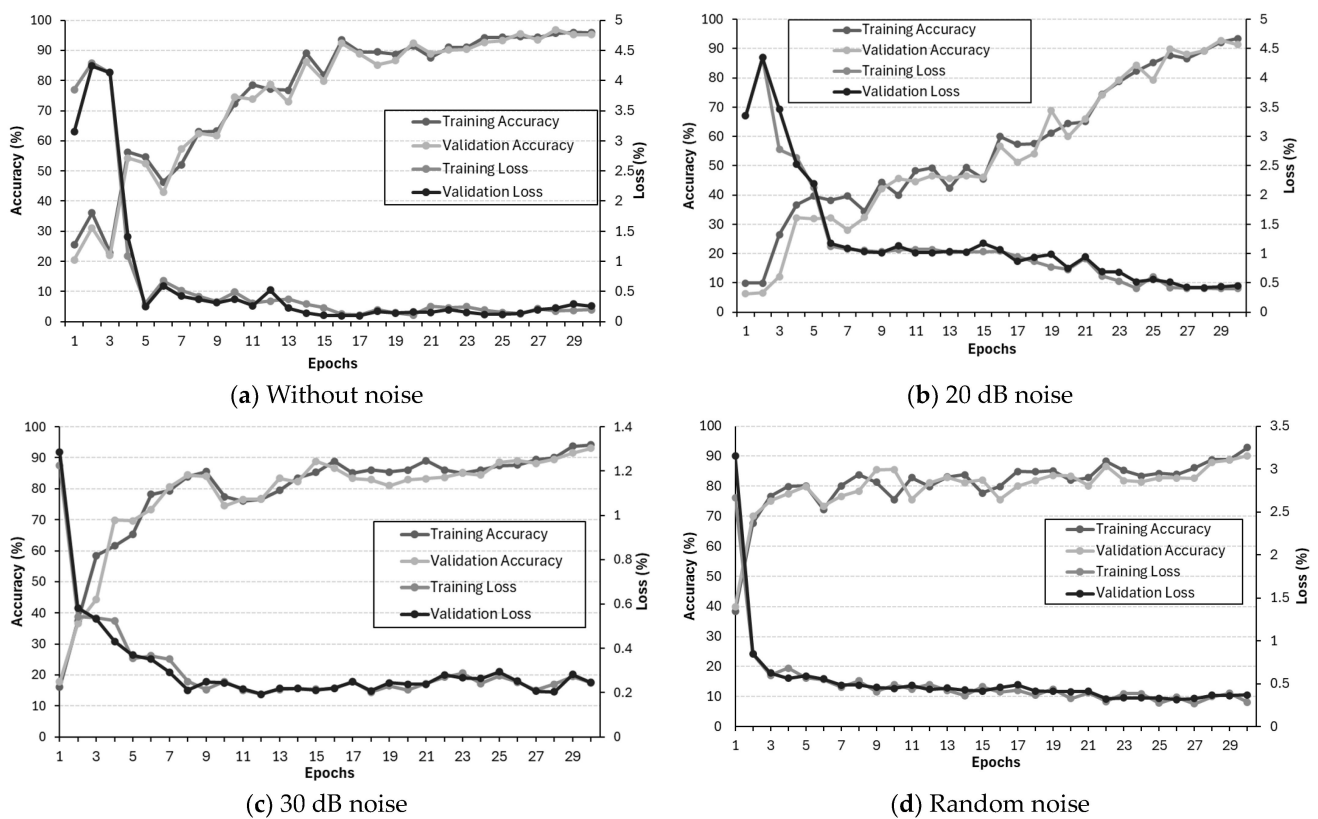
Figure 15. ResNet-50 with SE mechanism's training performance for noisy and noiseless datasets.

3.3. Comparative Analysis with Literature

In this sub-section, a comparison of the proposed approach with the literature is presented in Table 5 to highlight the standing of the developed model. For this purpose, references utilizing 1-D signal data and 2-D image data, both with and without noisy data, are considered. A thorough evaluation of the proposed model is presented in terms of accuracy, precision, recall, and F1-score. The reported work lacks detailed performance evaluation metrics, which may result in improper classification outcomes due to class imbalance. In [24,26,30,38,44], the authors use 1-D signals for the classification of PQDs

using machine learning models, whereas in [28,29,33], the authors consider a 2-D image dataset for PQD classification using transfer learning. The highest classification accuracy of 99.71% is achieved in [44] for noiseless 1-D signal fast discrete S-transform with a memetic firefly algorithm-based light gradient boost machine. On the other end, the authors in [28] achieved a maximum accuracy of 99.80% based on a PQD image dataset using a transfer learning approach. In comparison with the literature, our voting ensemble model has outperformed the reported methods in all of the performance evaluation matrices. Thus, the proposed method can improve the accuracy of pre-trained models for classifying PQD signals using their images. To address the computational cost of our approach and its feasibility for real-time applications in smart grids, it is important to note that CNNs are deep architectures and can be computationally expensive. However, in this method, transfer learning approaches are employed, which significantly reduce training time and computational efforts by using pre-trained CNN architectures. This allows us to focus on fine-tuning the models for PQD classification rather than training from scratch, making the process more efficient and feasible for real-time applications. PQDs can be specified both in frequency and magnitude by different standards such as IEEE 1159, EN 50160, and IEC 61000, [45]. IEEE Std. 1159 focuses on power quality monitoring and classification of disturbances like voltage disturbances, including sags, swells, interruptions, flicker, harmonics, transients, and electrical noise, whereas EN 50160 defines voltage quality, including frequency, voltage variations, flicker, harmonics, and voltage dips and swells, as well as defining thresholds for acceptable voltage variations. IEC 61000 regulates electromagnetic compatibility, including voltage dips, surges, harmonics, flicker, transients, and high-frequency disturbances. The authors in [46,47] were able to achieve 100% (three classes) and 99.78% (nine classes, Stockwell Transform and Decision Tree) classification accuracy, respectively, using the IEC 61000 standard; however, other PQDs are not incorporated in the work. A PQD classification accuracy of 99.26% is achieved for eight classes for the EN 50160 standard using a probabilistic neural network [48]. In another work [25], eight different types of PQDs based on the EN 50160 standard are classified using a CNN model, giving a 97.94% classification accuracy. The proposed approach utilized sixteen different PQDs based on all these standards and produced a 99.98% classification accuracy. Thus, the proposed approach is substantially adaptable for various power quality monitoring environments with a high classification accuracy.

In smart grid applications, real-time performance is critical to ensure timely detection and classification of PQDs. Although the ensemble deep CNN approach shows strong accuracy and robustness in classifying PQDs, it is important to assess the time performance. The training and test times (single sample and batch of fifty samples) for each architecture were recorded on a PC model Intel Core i9-9820X CPU (3.3.0 GHz), DDR4 32 GB RAM and NVIDIA Ge-Force RTX 2080 8 G GPU and are presented in Table 6.

Table 5. Performance comparison of the proposed approach with the literature.

Method	Without Noise				20 dB Noise				30 dB Noise				Random Noise			
	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1-D Signals																
FDST+MFA_LGBM [44]	99.71	-	-	-	96.85	-	-	-	98.45	-	-	-	-	-	-	-
SE with (LR+NB+J48 DT) [30]	91	91.5	91	91.10	-	-	-	-	-	-	-	-	89.33	89.60	89.3	89.3
DR with (KNN, SVM, NB, RF) [38]	-	-	-	-	99.72	-	-	-	99.48	-	-	-	99.65	-	-	-
Bagging-LSTM [24]	-	-	-	-	98.67	-	-	-	99.20	-	-	-	-	-	-	-
HT+DAOM [26]	99.44	99.24	99.15	99.19	-	-	-	-	98.95	98.58	98.05	98.31	-	-	-	-
2-D Images																
Pre-trained deep Networks [28]	99.80	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ST+CNN [29]	99.12	-	-	-	98.57	-	-	-	98.14	-	-	-	83.45	-	-	-
1-D+2-D CNN [33]	99.71	-	99.53	99.80	-	-	-	-	-	-	-	-	-	-	-	-
Proposed Approach	**99.98**	**99.97**	**99.80**	**99.85**	**99.73**	**98.23**	**97.23**	**97.78**	**99.90**	**99.83**	**99.65**	**99.80**	**99.88**	**98.68**	**98.10**	**98.05**

Table 6. Comparison of computational times of deep CNNs models.

Model	Training Time	Test Time (Batch of Fifty Samples)	Test Time (Single Sample)
ResNet-50	298 min 48 s	2.98 s	59.8 ms
VGG-16	293 min. 50 s	2.77 s	55.6 ms
AlexNet	270 min 59 s	2.51 s	50.4 ms
SqueezeNet	330 min 25 s	4.02 s	80.6 ms
ResNet-50 with SA mechanism	283 min 41 s	2.69 s	53.9 ms
Voting Ensemble	-	2.75 s	55.1 ms

ResNet-50, VGG-16, and ResNet-50 with SA mechanisms show higher accuracy compared to other deep models; however, relatively high training times are required, as these are deep and computationally intensive models compared to AlexNet and SqueezeNet. Moreover, a relatively long time is also taken to process the testing images. The training time is not provided for the ensemble model as it relies on the performance of the base models. Due to the complexity involved in combining multiple models, the voting ensemble with the highest classification accuracy requires 2.75 s to process a batch of fifty test images. For a single test sample, the test time is computed as 55.1 ms. The stated inference time illustrates that the proposed system can detect and classify disturbances rapidly, with each model's testing performance for a single test sample making them suitable for real-time grid monitoring. The ability to process each sample in milliseconds provides sufficient time for automated control systems to initiate protective or corrective actions within the necessary dynamics of a power grid. Furthermore, by leveraging edge computing devices equipped with GPUs or TPUs, the low-latency classification can be implemented locally, minimizing communication delays and further enhancing real-time responsiveness. Therefore, the proposed model has real-time application importance. In addition to this, several strategies can be incorporated to meet the rigorous time constraint, including modern hardware such as GPU/TPU and utilizing a sliding window technique, and model pruning and quantization can further enhance the inference speed without sacrificing classification accuracy. The deep CNN-based ensemble approach used in this study for PQD classification can be effectively integrated into a power grid through a three-layered architecture consisting of data acquisition and preprocessing, disturbance detection, and a communication interface. In the first layer, smart sensors and Phasor Measurement Units (PMUs) are deployed across the grid to continuously monitor electrical parameters by capturing voltage and current signals at various locations. These signals are then preprocessed and converted from 1-D time domain signals into time–frequency images using CWT. Depending on the grid's setup and communication capabilities, this preprocessing can be performed either on local edge devices near the measurement points or at centralized data hubs. In the next layer, the preprocessed time–frequency images are fed into the deep CNN-based ensemble model for disturbance classification. This can be implemented on edge computing devices equipped with GPUs/TPUs, ensuring low-latency processing. The trained model can be deployed either on these edge devices or at centralized processing units to perform real-time disturbance classification. Once disturbances are detected and classified, the results are communicated to the central control center or grid operators via a secure communication network. These results can then trigger automated control actions such as switching operations, load balancing, or system protection reconfiguration to mitigate the impact of disturbances. Additionally, the system can be integrated with existing Supervisory Control and Data Acquisition (SCADA) systems to enhance comprehensive monitoring and automated decision-making. By adopting this layered approach, the proposed deep CNN ensemble system can be seamlessly incorporated into the existing power grid infrastructure, enabling enhanced monitoring, rapid disturbance detection, and effective response strategies, thereby improving grid reliability and stability in a smart city context.

4. Conclusions

This work presents a voting ensemble approach based on pre-trained deep learning models for PQD classification using time–frequency images. In this process, CWT is used to convert 1-D PQD signals into 2-D time–frequency images. Four DCNNs including ResNet-50, VGG-16, AlexNet and SqueezeNet are employed to classify sixteen PQD classes for four datasets, i.e., without noise, 20 dB noise, 30 dB noise, and random noise. In the case of the individual model, ResNet-50 produced the best results for accuracy (99.75%), precision (98.04%), recall (98%), and F1-score (98%). To further enhance the performance of the classification model, a voting ensemble approach is presented. The proposed approach outperformed the individual deep learning models for accuracy (99.98%), precision (99.97%), recall (99.80%), and F1-score (99.85%) with the noiseless dataset. The results confirm the suitability of the proposed approach using image classification applications for PQDs in the context of smart cities, where reliable energy quality is crucial for the operation of various connected devices and infrastructure. In the future, it is intended to explore new deep learning models with real-world datasets to demonstrate their effectiveness in a real-time smart city environment, contributing to enhanced energy management and resilience.

5. Disclaimer

The views and conclusions contained herein should not be interpreted as necessarily representing the official policies, endorsements, or views of Sheroze Liaquat's current employer, Eaton Corporation.

References

1. Chakravorti, T.; Nayak, N.; Bisoi, R.; Dash, P.; Tripathy, L. A new robust kernel ridge regression classifier for islanding and power quality disturbances in a multi distributed generation based microgrid. *Renew. Energy Focus* **2019**, *28*, 78–99. [CrossRef]
2. Conrado, B.R.; de Souza, W.A.; Liberado, E.V.; Paredes, H.K.; Brandao, D.I.; Moreira, A.C. Towards technical and economic feasibility of power quality compensators. *Electr. Power Syst. Res.* **2023**, *216*, 109020. [CrossRef]
3. Li, J.; Liu, H.; Wang, D.; Bi, T. Classification of power quality disturbance based on S-transform and convolution neural network. *Front. Energy Res.* **2021**, *9*, 708131. [CrossRef]
4. Priyadarshini, M.; Bajaj, M.; Prokop, L.; Berhanu, M. Perception of power quality disturbances using Fourier, Short-Time Fourier, continuous and discrete wavelet transforms. *Sci. Rep.* **2024**, *14*, 3443. [CrossRef]
5. Chen, S.; Li, Z.; Pan, G.; Xu, F. Power quality disturbance recognition using empirical wavelet transform and feature selection. *Electronics* **2022**, *11*, 174. [CrossRef]
6. Rodriguez, M.A.; Sotomonte, J.F.; Cifuentes, J.; Bueno-López, M. Classification of power quality disturbances using hilbert huang transform and a multilayer perceptron neural network model. In Proceedings of the 2019 International Conference on Smart Energy Systems and Technologies (SEST), Porto, Portugal, 9–11 September 2019; IEEE: New York, NY, USA, 2019; pp. 1–6.
7. Dutt, P.V.B.P.; Balaga, H. Detection and Classification of Power Quality Disturbances Using Variational Mode Decomposition and Deep Learning Networks. In Proceedings of the International Conference on Flexible Electronics for Electric Vehicles, Jaipur, India, 28–29 July 2022; Springer: Singapore, 2022; pp. 1–12.
8. Eldar, Y.C.; Hero, A.O., III; Deng, L.; Fessler, J.; Kovacevic, J.; Poor, H.V.; Young, S. Challenges and open problems in signal processing: Panel discussion summary from ICASSP 2017 [panel and forum]. *IEEE Signal Process. Mag.* **2017**, *34*, 8–23. [CrossRef]
9. Ijaz, M.; Shafiullah, M.; Abido, M. Classification of power quality disturbances using Wavelet Transform and Optimized ANN. In Proceedings of the 2015 18th International Conference on Intelligent System Application to Power Systems (ISAP), Porto, Portugal, 11–16 September 2015; IEEE: New York, NY, USA, 2015; pp. 1–6.

10. Borges, F.A.; Fernandes, R.A.; Silva, I.N.; Silva, C.B. Feature extraction and power quality disturbances classification using smart meters signals. *IEEE Trans. Ind. Inform.* **2015**, *12*, 824–833. [CrossRef]

11. Mahela, O.P.; Shaik, A.G.; Khan, B.; Mahla, R.; Alhelou, H.H. Recognition of complex power quality disturbances using S-transform based ruled decision tree. *IEEE Access* **2020**, *8*, 173530–173547. [CrossRef]

12. Puliyadi Kubendran, A.K.; Loganathan, A.K. Detection and classification of complex power quality disturbances using S-transform amplitude matrix–based decision tree for different noise levels. *Int. Trans. Electr. Energy Syst.* **2017**, *27*, e2286. [CrossRef]

13. Luo, Y.; Li, K.; Li, Y.; Cai, D.; Zhao, C.; Meng, Q. Three-layer Bayesian network for classification of complex power quality disturbances. *IEEE Trans. Ind. Inform.* **2017**, *14*, 3997–4006. [CrossRef]

14. Saleem, A.; Khosiljonovich, K.I.; Qizi, K.M.M.; Sokhib, K.; Ugli, S.M.S.; Obidovich, S.S. Estimation of power quality in distribution system using fuzzy logic theory. *Indones. J. Electr. Eng. Comput. Sci.* **2023**, *323*, 1236–1245.

15. Moreira, A.C.; Paredes, H.K.; de Souza, W.A.; Marafao, F.P.; Da Silva, L.C. Intelligent expert system for power quality improvement under distorted and unbalanced conditions in three-phase AC microgrids. *IEEE Trans. Smart Grid* **2017**, *9*, 6951–6960. [CrossRef]

16. Thirumala, K.; Pal, S.; Jain, T.; Umarikar, A.C. A classification method for multiple power quality disturbances using EWT based adaptive filtering and multiclass SVM. *Neurocomputing* **2019**, *334*, 265–274. [CrossRef]

17. Naderian, S.; Salemnia, A. An implementation of type-2 fuzzy kernel based support vector machine algorithm for power quality events classification. *Int. Trans. Electr. Energy Syst.* **2017**, *27*, e2303. [CrossRef]

18. Dekhandji, F.Z.; Recioui, A.; Ladada, A.; Moulay Brahim, T.S. Detection and Classification of Power Quality Disturbances Using LSTM. *Eng. Proc.* **2023**, *29*, 2. [CrossRef]

19. Chiam, D.H.; Lim, K.H.; Law, K.H. LSTM power quality disturbance classification with wavelets and attention mechanism. *Electr. Eng.* **2023**, *105*, 259–266. [CrossRef]

20. Wang, S.; Chen, H. A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network. *Appl. Energy* **2019**, *235*, 1126–1140. [CrossRef]

21. Perez-Anaya, E.; Jaen-Cuellar, A.Y.; Elvira-Ortiz, D.A.; Romero-Troncoso, R.d.J.; Saucedo-Dorantes, J.J. Methodology for the Detection and Classification of Power Quality Disturbances Using CWT and CNN. *Energies* **2024**, *17*, 852. [CrossRef]

22. Ercolano, G.; Rossi, S. Combining CNN and LSTM for activity of daily living recognition with a 3D matrix skeleton representation. *Intell. Serv. Robot.* **2021**, *14*, 175–185. [CrossRef]

23. Mohan, N.; Soman, K.; Vinayakumar, R. Deep power: Deep learning architectures for power quality disturbances classification. In Proceedings of the 2017 International Conference on Technological Advancements in Power and Energy (TAP Energy), Kollam, India, 21–23 December 2017; IEEE: New York, NY, USA, 2017; pp. 1–6.

24. Wang, J.; Zhang, D.; Zhou, Y. Ensemble deep learning for automated classification of power quality disturbances signals. *Electr. Power Syst. Res.* **2022**, *213*, 108695. [CrossRef]

25. El-Rashidy, M.A.; Abd-elhamed, S.A.; El-Fishawy, N.A.; Shouman, M.A. Efficient online detection system of power disturbances based on deep-learning approach. *Alex. Eng. J.* **2023**, *70*, 377–394. [CrossRef]

26. Liu, Y.; Jin, T.; Mohamed, M.A. A novel dual-attention optimization model for points classification of power quality disturbances. *Appl. Energy* **2023**, *339*, 121011. [CrossRef]

27. Manikonda, S.K.; Santhosh, J.; Sreekala, S.P.K.; Gangwani, S.; Gaonkar, D.N. Power quality event classification using transfer learning on images. In Proceedings of the 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Tamilnadu, India, 11–13 April 2019; IEEE: New York, NY, USA, 2019; pp. 1–5.

28. Todeschini, G.; Kheta, K.; Giannetti, C. An image-based deep transfer learning approach to classify power quality disturbances. *Electr. Power Syst. Res.* **2022**, *213*, 108795. [CrossRef]

29. Fu, L.; Deng, X.; Chai, H.; Ma, Z.; Xu, F.; Zhu, T. PQEventCog: Classification of power quality disturbances based on optimized S-transform and CNNs with noisy labeled datasets. *Electr. Power Syst. Res.* **2023**, *220*, 109369. [CrossRef]

30. Radhakrishnan, P.; Ramaiyan, K.; Vinayagam, A.; Veerasamy, V. A stacking ensemble classification model for detection and classification of power quality disturbances in PV integrated power network. *Measurement* **2021**, *175*, 109025. [CrossRef]

31. Dhalaria, M.; Gandotra, E.; Saha, S. Comparative analysis of ensemble methods for classification of android malicious applications. In Proceedings of the Advances in Computing and Data Sciences: Third International Conference, ICACDS 2019, Ghaziabad, India, 12–13 April 2019; Springer: Singapore, 2019; pp. 370–380.

32. Kiruthiga, B.; Narmatha Banu, R.; Devaraj, D. Detection and classification of power quality disturbances or events by adaptive NFS classifier. *Soft Comput.* **2020**, *24*, 10351–10362. [CrossRef]

33. Sindi, H.; Nour, M.; Rawa, M.; Öztürk, Ş.; Polat, K. A novel hybrid deep learning approach including combination of 1D power signals and 2D signal images for power quality disturbance classification. *Expert Syst. Appl.* **2021**, *174*, 114785. [CrossRef]

34. Zhang, L.; Jiang, C.; Chai, Z.; He, Y. Adversarial attack and training for deep neural network based power quality disturbance classification. *Eng. Appl. Artif. Intell.* **2024**, *127*, 107245. [CrossRef]

35. Salles, R.S.; Ribeiro, P.F. The use of deep learning and 2-D wavelet scalograms for power quality disturbances classification. *Electr. Power Syst. Res.* **2023**, *214*, 108834. [CrossRef]

36. Machlev, R.; Chachkes, A.; Belikov, J.; Beck, Y.; Levron, Y. Open source dataset generator for power quality disturbances with deep-learning reference classifiers. *Electr. Power Syst. Res.* **2021**, *195*, 107152. [CrossRef]

37. *1159-1995*; IEEE Recommended Practice for Monitoring Electric Power Quality. IEEE: New York, NY, USA, 1995.

38. Singh, G.; Pal, Y.; Dahiya, A.K. Classification of power quality disturbances using linear discriminant analysis. *Appl. Soft Comput.* **2023**, *138*, 110181. [CrossRef]

39. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; IEEE: New York, NY, USA, 2016; pp. 770–778.

40. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

41. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 84–90. [CrossRef]

42. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. *arXiv* **2016**, arXiv:1602.07360.

43. Çığşar, B.; Ünal, D. Comparison of data mining classification algorithms determining the default risk. *Sci. Program.* **2019**, *2019*, 8706505. [CrossRef]

44. Panigrahi, R.R.; Mishra, M.; Nayak, J.; Shanmuganathan, V.; Naik, B.; Jung, Y.-A. A power quality detection and classification algorithm based on FDST and hyper-parameter tuned light-GBM using memetic firefly algorithm. *Measurement* **2022**, *187*, 110260. [CrossRef]

45. Sepasi, S.; Talichet, C.; Pramanik, A.S. Power quality in microgrids: A critical review of fundamentals, standards, and case studies. *IEEE Access* **2023**, *11*, 108493–108531. [CrossRef]

46. Balouji, E.; Salor, O. Classification of power quality events using deep learning on event images. In Proceedings of the 2017 3rd International Conference on Pattern Recognition and Image Analysis (IPRIA), Shahrekord, Iran, 19–20 April 2017; IEEE: New York, NY, USA, 2017; pp. 216–221.

47. Minh Khoa, N.; Van Dai, L. Detection and classification of power quality disturbances in power system using modified-combination between the stockwell transform and decision tree methods. *Energies* **2020**, *13*, 3623. [CrossRef]

48. Wang, H.; Wang, P.; Liu, T. Power quality disturbance classification using the S-transform and probabilistic neural network. *Energies* **2017**, *10*, 107. [CrossRef]