Article

Automatic Power Quality Disturbance Diagnosis Based on Residual Denoising Convolutional Auto-Encoder

Jie Liu ¹ , Qiu Tang ^{1,*}, Wei Qiu ^{1,2,*}, Jun Ma ¹ , Yuhong Qin ¹ and Biao Sun ¹

¹ College of Electrical and Information Engineering, Hunan University, Changsha 410082, China; liu_jie@hnu.edu.cn (J.L.); 19090277@hnu.edu.cn (J.M.); holyqin@hnu.edu.cn (Y.Q.); sunbiao1993@hnu.edu.cn (B.S.)

² Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37996, USA

* Correspondence: tangqiu@hnu.edu.cn (Q.T.); qiuwei@hnu.edu.cn (W.Q.); Tel.: +86-130-7736-8836 (Q.T.)

Abstract: With the increasing integration of non-linear electronic loads, the diagnosis and classification of power quality are becoming crucial for power grid signal management. This paper presents a novel diagnosis strategy based on unsupervised learning, namely residual denoising convolutional auto-encoder (RDCA), which extracts features from the complex power quality disturbances (PQDs) automatically. Firstly, the time–frequency analysis is applied to isolate frequency domain information. Then, the RDCA with a weight residual structure is utilized to extract the useful features in the contaminated PQD data, where the performance is improved using the residual structure. A single-layer convolutional neural network (SCNN) with an added batch normalization layer is proposed to classify the features. Furthermore, combining with RDCA and SCNN, we further propose a classification framework to classify complex PQDs. To provide a reasonable interpretation of the RDCA, visual analysis is employed to gain insight into the model, leading to a better understanding of the features from different layers. The simulation and experimental tests are conducted to verify the practicability and robustness of the RDCA.

Keywords: automatic extracted features; power quality disturbances; residual denoising convolutional auto-encoder; single-layer convolutional neural network

Citation: Liu, J.; Tang, Q.; Qiu, W.; Ma, J.; Qin, Y.; Sun, B. Automatic Power Quality Disturbance Diagnosis Based on Residual Denoising Convolutional Auto-Encoder. *Appl. Sci.* **2021**, *11*, 7637. <https://doi.org/10.3390/app11167637>

Academic Editor: Juan-Carlos Cano

Received: 12 July 2021

Accepted: 17 August 2021

Published: 20 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The primary purpose of the power system is to provide industry and individuals with a stable and clean electricity signal [1]. However, the power grid suffers from power quality disturbances (PQDs) due to the increasing involvement of non-linear inductive and capacitive loads in the distribution network. These disturbances that pollute the original clean AC sinusoidal waveform have a negative impact on other consumers and even the stability of the power grid. Moreover, different kinds of PQDs, such as flicker, harmonics, swells, transients, etc., result in not only economic losses, but also an unreasonable waste of energy [2,3]. Unfortunately, the combination of different types of disturbances increases the difficulties of accurate detection. Therefore, to improve the accuracy of power quality diagnosis, it is necessary to take automatic and effective measures to cope with more complex situations.

Taking into account the requirements for accurate diagnosis of power quality, a large number of studies have been proposed. In general, the PQD diagnosis and classification steps can be approximately divided into three steps, i.e., time–frequency analysis, feature extraction, and classification [4]. Various time–frequency analysis algorithms, such as short-time Fourier transform (STFT), transform (GT), wavelet transform (WT), Hilbert–Huang transform (HHT), the Stockwell transform (ST), and some optimized methods have been proposed for the digital signal processing [5]. However, the fixed parameters setting will reduce its applicability, especially for complex PQ disturbances.

In the feature extraction stage, different statistical features are then extracted and fed to the classifier, such as peak and harmonic amplitude [6]. Different kinds of literature have adopted different quantities of statistical features. For example, the identification and classification method under multiple disturbances was presented in [7], by combining the adaptive linear networks and feedforward neural networks together with two statistical features. It was reported in [8] that multiple PQDs were analyzed along with six manually selected features and nearly 94.19% accuracy was achieved. Furthermore, 62 features were designed for classification in [9], and the model has better performance when only partial features were used. However, the design of these manually selected features lacks reference and may lead to information loss owing to human factors. Hereby, automatic feature extraction is becoming increasingly popular.

In the classification stage, various classifiers for PQDs, including support vector machine (SVM), random forests (RF) decision tree (DT), etc., have been also developed [10]. Combined with time–frequency analysis and feature extraction, these classifiers can identify the manually selected features and have a good classification performance. For instance, a practical arcing fault identification method proposed in [11] combines the GT and ANN, but the time–frequency domain window width is fixed, which restricts the analysis ability for signals with rich frequency domain components. Researchers in [12,13] utilized WT and SVM to conduct power quality events classification. Additionally, low-pass scaling and high-pass scaling parameters were used to accommodate signal resolution [12], while the performance of WT degraded in noisy conditions. Then, ST with DT in [14] was developed to evaluate and detect the power quality, nevertheless, the high computational demand makes it difficult to meet the real-time requirement. To balance the trade-off between speed and accuracy, a modified ST with RF and double resolution ST (DRST) with directed acyclic graph SVM in [15,16] were proposed respectively, where the key frequency point selection techniques were used to reduce the computational burden. However, there is still the limitation for the classifiers to deal with ever-changing automatic feature.

The conventional models have been effectively applied to the field of PQD diagnosis since one problem further prevented the development of these methods, namely the feature extraction process [17]. So far, all the features of power quality have been manually extracted from original signal or time–frequency information. It is found that these manually selected features are various for different researchers. Not only that, the process of manually selected feature extraction is time-consuming, complex, and may not be able to completely distinguish disturbance signals, especially for composite signals [18]. Recently, the deep neural network, an automatic feature extraction method, has attracted attention for image analysis and speech recognition [19]. For instance, the convolutional neural network (CNN) has been successfully used in feature learning for big data [20,21].

Unlike most previous models, which mainly focused on classification problems, a novel method is introduced for the purpose of automatic feature extraction. Our contributions are stated as follows:

1. The time–frequency information of PQDs is initially separated by DRST. To refine the feature extraction capability, a noise reduction, and weight residual structure referred as residual denoising convolutional auto-encoder (RDCA) is proposed to automatically extract features from time–frequency information.
2. In terms of classification, the single-layer CNN (SCNN) classifier is used to fulfill the classification task. The batch normalization (BN) layer and dynamic learning rate are integrated to speed up convergence.
3. We also propose a classification framework based on RDCA and SCNN. Especially, we perform two visual verification methods to investigate how RDCA could benefit from the diagnosis of PQDs. Features extracted from different layers of RDCA are visualized and analyzed to verify the meaning of the extracted features.
4. To validate the framework, the simulation and experimental results on both single and combined PQDs data sets are conducted under different noise conditions.

The rest of the paper is structured as follows. In Section 2, the proposed RDCA and SCNN are introduced. The PQD diagnostic framework is presented in Section 3. Section 4 explores the interpretability of extracted features by RDCA. Thereafter, Section 5 elucidates the effectiveness of the proposed method through a large number of experiments. Finally, Section 6 concludes this paper.

2. Automatic PQD Diagnosis Method

2.1. Time–Frequency Analysis of Disturbance Signals

Single and combined PQDs contain abundant time domain and frequency domain information. Hereby, different manually selected features are applied to extract as complete feature information as possible from the time–frequency analysis since the classification accuracy of PQDs is highly affected by the performance of time–frequency analysis [22].

The advantage of DRST is that it has high energy concentration at both low and high frequencies. In our work, we use the DRST, which has a double resolution, as a time–frequency decomposition algorithm for PQDs [16]. The DRST of input PQD signal $x(t)$ can be defined as

$$\text{DRST}(\mu, f) = \int_{-\infty}^{\infty} x(t) \frac{\sqrt{\beta_{1,2}|f|}}{\sqrt{2\pi}} e^{-\frac{(\mu-t)^2\beta_{1,2}|f|}{2}} e^{-j2\pi ft} dt \quad (1)$$

where $\beta_{1,2}$ denotes the adjustable parameters, f denotes the signal frequency, μ is the time shift factor. Essentially, the time–frequency spectrum of PQDs is divided into two parts, where $f \leq 1.5f_0$ and $f > 1.5f_0$ represent the low-frequency part and high-frequency part, respectively. It is worth mentioning that to avoid the spectrum leakage and provide more accurately signal analysis result, the separation value is set to 1.5. The f_0 is the fundamental frequency. Besides, to improve the frequency resolution and noise reduction capability, the values of parameter $\beta_{1,2}$ are set to 121 and 10, respectively, according to [16]. To verify the effectiveness, we compared the disturbance signal analysis result with different parameter $\beta_{1,2}$. The disturbance signal is the harmonic disturbance with the harmonic level is the third and the fifth harmonic in Figure 1a–c show the signal analysis result with different parameters of $\beta_{1,2}$.

Figure 1. Comparison result on harmonic disturbance with different parameter $\beta_{1,2}$. (a) input signal, (b) $\beta_1 = 150, \beta_2 = 40$, (c) $\beta_1 = 121, \beta_2 = 10$.

From Figure 1, the parameters $\beta_{1,2}$ directly affect the time–frequency result. For example, when parameters $\beta_{1,2}$ are set to 121 and 10, respectively, the third harmonic and the fifth harmonic can be identified accurately. When parameters $\beta_{1,2}$ are set to 150 and 40, respectively, the time–frequency resolution is lower. Besides, there is also the spectrum leakage problem. Thus, to improve the frequency resolution and noise reduction capability, the $\beta_{1,2}$ are set to 121 and 10, respectively.

To further verify the detection performance, we compared the transients disturbance analysis result with different parameters of $\beta_{1,2}$. Figure 2a shows the transients input signal. Figure 2b,c show the signal analysis result with different parameters of $\beta_{1,2}$.

Figure 2. Comparison result on transients disturbance with different parameter $\beta_{1,2}$. (a) Input signal, (b) $\beta_1 = 150, \beta_2 = 40$, (c) $\beta_1 = 121, \beta_2 = 10$.

From Figure 2, when parameters $\beta_{1,2}$ are set to 121 and 10, respectively, the transients disturbance can be identified accurately, indicating a suitable frequency resolution. When parameters $\beta_{1,2}$ are set to 150 and 40, respectively, the frequency resolution is lower. Besides, there is also the spectrum leakage problem. Thus, to improve the time–frequency resolution, the $\beta_{1,2}$ are set to 121 and 10, respectively.

Different from the manually selected features perspective, the output of DRST, which is an amplitude time–frequency matrix, can be regarded as an image perspective instead of a conventional single amplitude value feature. At this time, the amplitude and time of the time–frequency can be considered as pixel value and location of the image, respectively. In this work, the sampling frequency is set to 2.56 kHz where the fundamental frequency $f_0 = 50$ Hz and the length of the sampling period is 10 cycles of the fundamental component. Then the size of the amplitude time–frequency matrix is 256×512 (height $\times$ width). It is worth mentioning that the 256 and 512 are the length of the frequency axis and the time axis, respectively. For the 50 Hz fundamental frequency, there will be 512 points in 10 periods. Meanwhile, the length of the frequency axis of the time–frequency matrix is half of the time axis since the results of the FFT is symmetrical. Thus, the size of the amplitude time–frequency matrix is 256×512 . The equations of discrete DRST are listed as follows.

Consider the input signal $x(t)$ sampled at a fixed sampling frequency f_s corresponding to the time interval of $T_s = 1/f_s$. The discrete Fourier transform (DFT) of sampled signal is

$$X(k) = \frac{1}{N} \sum_{l=0}^{N-1} x(l) e^{-j2\pi lk/N} \quad (2)$$

where $x(l)$ represents the discrete-time sampling sequence of $x(t)$ and $k = 0, 1, \dots, N-1$.

By using the Equation (2) to the sampled signal, we get an N -point vector $\mathbf{X}_N$ of frequency samples as

$$\mathbf{X}_N = [X_0 \quad X_1 \quad \dots \quad X_{N-1}] \quad (3)$$

Define a 2-D matrix $P_{N \times M}$ as

$$P_{N \times M} = \begin{bmatrix} X_1 & X_2 & \dots & X_M \\ X_2 & X_3 & \dots & X_{M+1} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N-1} & X_0 & \dots & X_{M-2} \\ X_0 & X_1 & \dots & X_{M-1} \end{bmatrix} \quad (4)$$

where $M = 2N$ is owing to the symmetry property of DRST.

Another 2-D matrix $G_{N \times M}$ is defined as

$$G_{N \times M} = \begin{bmatrix} G(0,1) & G(0,2) & \dots & G(0,M) \\ G(1,1) & G(1,2) & \dots & G(1,M) \\ \vdots & \vdots & \ddots & \vdots \\ G(N-1,1) & G(N-1,2) & \dots & G(N-1,M) \end{bmatrix} \quad (5)$$

where $G(n, m) = e^{-(2\pi^2 n^2 / m \lambda_{1,2})}$ is the FFT for the redefined Gaussian window, $n = 0, 1, \dots, N-1$ and $m = 0, 1, \dots, M$.

Thus, by utilizing the convolution theorem and frequency shifting property of Fourier transform, the N -point discrete DRST can be described in matrix form as

$$\text{DRST}_{N \times M} = \text{IDFT}\{P_{N \times M} \circ G_{N \times M}\} \quad (6)$$

where $\circ$ represents the Hadamard product of the matrix and IDFT representation of the inverse DFT (IDFT). In applications, the FFT and inverse FFT (IFFT), instead of DFT and IDFT, respectively, are used to improve the computational efficiency.

The $\text{DRST}(n, m)$ is, in general, a complex-valued matrix. Therefore, it can be expressed as

$$\text{DRST}(n, m) = |\text{DRST}(n, m)| e^{j\phi(n, m)} \quad (7)$$

where $|\text{DRST}(n, m)|$ is the magnitude of $\text{DRST}(n, m)$ and $\phi(n, m) = \angle \text{DRST}(n, m)$ is the phase of $\text{DRST}(n, m)$.

Clearly, the amplitude associated with DRST matrix can be written as

$$A(n, m) = 2|\text{DRST}(n, m)| \quad (8)$$

Next, the amplitude time–frequency matrix of DRST is fed to RDCA.

2.2. Regular Convolutional Auto-Encoder

The auto-encoder (AE) serves as a data dimensionality reduction technique [23]. Traditionally, the framework of AE consists of two components, the encoder and decoder. The encoder is mainly responsible for segmenting the input image and extracting valid features, and the decoder is used to reconstruct input data with feature information [24]. The number of nodes in the hidden layer is much lower than the input layer so that the encoder can provide effective representation for the input data.

Given an input amplitude time–frequency data set $\mathbf{x} = \{x_1, x_2, \dots, x_l\}$ with shape size $m \times n$ for each element, where x_l denotes the l th sample of $\mathbf{x}$. Note that the $\mathbf{x}$ is the two-dimensional matrix set, which represents the time–frequency matrix caused by the DRST. The m and n represent the height and width of the time–frequency matrix, namely, the m and n are 256 and 512, respectively. In the auto-encoder, the data can be viewed as multiple two-dimensional images stacked on top of each other. Each image can be viewed as a feature mapping. The output mapping $\mathbf{z}_i^r$ of the encoder in the r th layer can be expressed as

$$\mathbf{z}_i^r = f(\mathbf{W} * \mathbf{x} + \mathbf{b}) \quad (9)$$

where $\mathbf{W}$ and $\mathbf{b}$ are the $d \times d \times h$ dimensional convolutional kernel (also called convolution filter) weight and offset vector, respectively. The h is the depth of the convolution layer, $i = 0, 1, \dots, h$. The d is the size of the convolution filter. The sign $*$ is the convolution operation, $f(\cdot)$ is the activate function. We choose the rectified linear unit as the activate function to improve the convergence speed [25]. The size of $\mathbf{z}_i^r$ becomes $m \times n \times h$ which the zero is padding into the blank space.

Following the convolutional layer in the encoder, a max-pooling layer, which plays a role of the downsampling operation, is applied to reduce the input dimension [20]. The output feature $\mathbf{z}_{pi}^r$ of the max-pooling layer is defined as $\mathbf{z}_{pi}^r = \max(\mathbf{z}_i(p \times p \times h)_s)$, where $p \times p \times h$ is the size of the pooling layer with the stride length s . It indicates that the maximum value per $p \times p$ size in $\mathbf{z}_i$ is the output of max-pooling. The size of pooled feature vectors $\mathbf{z}_{pi}^r$ is determined with the formula $\lfloor (m-p)/s \rfloor \times \lfloor (n-p)/s \rfloor \times h$, where $\lfloor \cdot \rfloor$ denotes rounding down. The target features $\mathbf{z}_{pi}^r$ extraction by the AE are fed to the classifier.

To reconstruct the signal, the decoder constructs a symmetric AE structure by up-sampling. Similarly, the hidden layer representation $\mathbf{z}_{pi}^r$ is then reversely mapped to the

same input size. This inverse mapping process is called decoder, the reconstructed output $\mathbf{x}'$ of the nonlinearity form is

$$\mathbf{x}' = f(\mathbf{W}' * \mathbf{z}_{pi}^r + \mathbf{b}') \quad (10)$$

where $\mathbf{W}'$ and $\mathbf{b}'$ are the weight and offset vector, respectively. The dimension of $\mathbf{x}'$ is expanded in up-sampling by interpolation so that the $\mathbf{W}'$ and $\mathbf{b}'$ have the same dimension corresponding to the encoder.

The regular AE can be used to compress complex PQDs to obtain representative features. However, it is difficult for AE to maintain effectiveness in shallow layers when the real-time diagnosis of PQDs is taken into account.

2.3. Residual Denoising Convolutional Auto-Encoder

One of the competitive methods to improve the optimization and feature expression ability is residual structure using shortcuts [26]. However, the effect of residual structure will be limited, especially for shallow AE. Therefore, we introduce a weighting strategy to residual structure and design a residual denoising convolutional auto-encoder.

The weight residual structure is shown in Figure 3. Figure 3 shows that the weight residual structure is an identity mapping where a weight parameter λ to feature $\mathbf{z}_i^r$ and $f(\mathbf{z}_i^r)$ are directly added. The red dotted line represents the original residual structure. It is worth mentioning that the feature $f(\mathbf{z}_i^r)$ are more separable than $\mathbf{z}_i^r$ after passing through multiple convolutional layers. It is reasonable to assign a lower weight for $\mathbf{z}_i^r$. Thus we introduce a weight parameter λ to feature $\mathbf{z}_i^r$ as shown in Figure 3. The output of the weight residual structure can be obtained as

$$\mathbf{z}_i^{r+1} = \lambda \mathbf{z}_i^r + f(\mathbf{z}_i^r) \quad (11)$$

where $\mathbf{z}_i^r$ and $\mathbf{z}_i^{r+1}$ are the input and output features of the weight residual structure. Additionally, we add a 1×1 convolution layer and max-pooling for $\mathbf{z}_i^r$ to fit the data dimension.

Based on the original residual structure, we can simply constrain the weight parameter λ to (0, 1). The weight residual structure is equivalent to the original residual structure when $\lambda = 0$. The detailed parameter λ value can be determined experimentally.

To learn useful features, a training criterion that contributes to the improvement of feature performance, namely denoising auto-encoder, is introduced to RDCA [23]. The key step is to clean partially corrupted input or add some noise to the corrupted input during training. For instance, the isotropic Gaussian noise (GN) σ is superimposed to the input $\mathbf{x}$, then the input of RDCA becomes $\mathbf{x} + \sigma$ while the output is still $\mathbf{x}$.

Thereafter, the proposed RDCA generates PQD features by stacking multiple convolutional layers. Finally, the model parameters are selected by optimizing the mean square error between the input $\mathbf{x}$ and reconstructed output $\mathbf{x}'$.

Figure 3. The weight residual structure of RDCA.

2.4. Single-Layer Convolutional Neural Network

After the RDCA is established, an effective classifier is required to classify the encoded features. The general classifier uses the softmax classifier with a fully connected (FC)

layer [27]. However, this classifier is easily troubled by high dimensional data, especially when the output of RDCA is high dimensional.

In this section, we propose a single-layer CNN, which consists of a convolutional layer, a pooling layer, a BN, and a FC layer. The single convolutional layer also enhances flexibility of auto-encoder, without considering the depth of the convolution filter in encoder. Followed by the pooling layer, we use the BN layer to speed up training. The output of the BN layer is set as $\mathbf{z}_c$, and it is reshaped to a sequence to connect the FC layer.

Then the feature vector $\mathbf{z}_c$, an image containing abstract features, is fed to the softmax classifier. We set k as the total number of PQDs classes. The output of softmax classifier is obtained as

$$P(y = j | \mathbf{z}_c, \theta) = \frac{e^{\theta_j^T \mathbf{z}_c}}{\sum_{i=1}^k e^{\theta_i^T \mathbf{z}_c}} \quad (12)$$

where θ and y denote the parameters of softmax classifier and output class, respectively. Normally, the category with the highest probability of PQDs can be specified as the final class. To further reduce the over-fitting under multiple combined disturbances, the dropout is applied to softmax classifier, which can randomly omit certain probability neural nodes [27].

3. Classification Framework

By using the proposed RDCA and SCNN, this section presents a new framework of automatic PQD diagnosis and classification. Its framework is illustrated in Figure 4. The detailed parameter settings are as follows: the size of C1, C2, C3, and C4 are $11 \times 11 \times 8$, $9 \times 9 \times 16$, $9 \times 9 \times 24$, and $7 \times 7 \times 4$, respectively. The P1, P2, and P3 are all set to 2×2 . The stride s are all set to 1×1 . We set the learning algorithm as the Adam. The learning rate is 0.001 and the dropout rate is 0.5. The batch size is 50.

Figure 4. The structure of automatic PQD diagnosis and classification based on RDCA.

The sign C_i and P_i are the sizes of convolutional filter and max-pooling in the i th layer, respectively. The F_i denotes the number of nodes in the FC layer. For example, when the input size $m \times n$ is set as 128×256 , it can be inferred from Figure 4 that the output of the first max-pooling layer is $64 \times 128 \times 8$ according to Section 2.2.

It is worth noting that the original shape of the amplitude matrix $\mathbf{x}$ is 256×512 . However, when there is an output of RDCA model with the shape $64 \times 128 \times 8$, the super-parameters of RDCA can exceed millions. Thus, one compromise to reduce the RDCA model parameters is to down-sampling DRST, i.e., the size of amplitude time–frequency matrix decreases from 256×512 to 128×256 .

Schematically, the classification of PQDs can be divided into three steps.

1. Time–frequency analysis: Obtain the amplitude matrix data set $\mathbf{x}$ based on DRST. Each amplitude matrix is normalized to $[0, 1]$ for better training.
2. Automatic feature extraction: A multilayer RDCA is designed and trained using the amplitude time–frequency matrix. Then obtain the output features $\mathbf{z}_{pi}^r$ of the encoder.
3. Classification: Remove the decoder structure of RDCA, and then feed features $\mathbf{z}_{pi}^r$ to the well-designed SCNN.

To reduce over-fitting, a dynamic learning rate is used in the training process of RDCA and SCNN. Specifically, the learning rate is reduced by half when the training loss is no longer reduced by five consecutive iterations.

4. Feature Visualization Analysis Based on RDCA

4.1. PQD Signal Model and Parameter Setting

In this work, a total of sixteen types of PQDs are considered to verify the performance of RDCA, including different single and combined PQDs as shown in Table 1. It should be noted that every single model is established based on the IEEE Standard [8,28] for the combined PQDs. It is worth mentioning that all harmonic levels are the combination of the third and the fifth harmonic.

For the better separation of the useful information through RDCA, the PQDs are randomly generated with MATLAB. Different signal-to-noise ratios (SNR) of GN are superimposed on the original sequence signal. The input SNR of RDCA is 10 dB higher than the reconstructed signal, e.g., the target reconstructed signal is 30 dB when the input is set 20 dB.

Table 1. The sixteen types of PQDs.

Label	Categories	Label	Categories
L1	Normal	L9	Voltage Spike
L2	Voltage Swell	L10	Swell with Harmonics
L3	Voltage Sag	L11	Flicker with Harmonics
L4	Voltage Interrupt	L12	Interrupt with Harmonics
L5	Harmonics	L13	Sag with Harmonics
L6	Transient	L14	Swell with Transient
L7	Flicker	L15	Sag with Transient
L8	Voltage Notch	L16	Spike with Transient

4.2. Convolution Filter and Extracted Features Visualization

It is particularly crucial to have a reliable interpretation of the extracted features from the black box in RDCA. Moreover, from the perspective of PQDs diagnosis structure optimization, the interpretation can also contribute to design model expectations.

In this work, we sought to establish the connection between PQD diagnosis and extracted features. Concretely, we take two methods to visualize the features in different layers from convolution filter and feature output aspects, namely gradient ascent and de-convolutional network (Deconvnet) [25,29].

The first method is gradient ascent [30], which aims at perturbing the input with noise and observing the change of RDCA prediction in every layer. The basic principle is applying gradient ascent on the input of the ConvNet so as to maximize the activation value of filter. Thus, a three-layer RDCA (RDCA-3) is built. The size of convolution filters are set to C1: $9 \times 9 \times 8$, P1: 2×2 , C2: $9 \times 9 \times 16$, P2: 2×2 , C3: $9 \times 9 \times 24$, P3: 2×2 . The corresponding size of output features per max-pooling layer are $64 \times 128 \times 8$, $32 \times 64 \times 16$, and $16 \times 32 \times 24$, respectively.

The effect of the filter on a certain disturbance image with the size of $128 \times 256 \times 1$ is illustrated in Figure 5. Here we set the L13 as input of the RDCA-3. Two pictures of each convolutional layer are picked by setting the threshold value that has the maximum loss value. The orange line at the top of the image represents the amplitude of the fundamental component. The second layer has some preliminary and uncertain shapes. It shows that the texture is clearer in the third layer, and it can be regarded as multiple complex combinations of the former layers. In the CNN model, the lower the convolutional layers' number is, the closer the signal is to the original signal, which means no useful information has been learned. The more the convolutional layers' number is, the higher the feature order is and more abstract, and many typical local features can be extracted. Therefore, the features in Layer3 are some typical local features in L13, indicating that information has been learned.

The details of the L13 disturbance image become increasingly abstract and high level when the layer of RDCA goes deeper. Particularly, the harmonic component becomes blurred while the characteristics of the fundamental wave are retained in the third layer. Figure 5 manifests that the convolution operation has the benefit to extract edge information and amplitude information of PQDs through a combination of features.

To further explore the feature information, we focus on the output of the feature maps in the encoder. Taking the single PQDs L1, L4, L5, L6, and L9, and combined PQDs L13, L14, and L15 as examples, the output features of the third layer in RDCA-3 are illustrated in Figure 6. The size of feature maps in the third layer is 16×32 .

According to the comparison of automatically extracted features in Figure 6, it can be seen that different PQD signals have unique characteristics. For instance, it clearly shows that there is a distinct recessed area and a small block in class L4 and L6, respectively, these corresponding high-level features indicate that the signal is voltage interrupt and transient disturbances. Furthermore, the alternating amplitude of L7 indicates that this feature is closest to the characteristic of the flicker signal. Both L5 and L13 demonstrate that the harmonics have more than one line in the feature visualization. Thus, it is revealed that the RDCA can accurately determine the location of the PQDs and categories.

Figure 5. The output of the convolution filter when L13 is the input.

Figure 6. The feature visualization of the third layer in RDCA.

Nevertheless, some of the features are hard to be recognized visually. For example, it is difficult to intuitively judge whether L13 contains a sag or swell signal in Figure 6. Besides, owing to the amplitude of spike being small, it is not obvious for finding the difference of L9 by our intuitive feeling. To investigate the meaning of the unintuitive features, we utilize the Deconvnet approach proposed by Zeiler [29]. The Deconvnet gives an insight into the intermediate feature layers which maps the extracted features back into the input pixel space.

The strategy of Deconvnet is comprised of deconvolution and unpooling operations. Typically, the Deconvnet can be obtained by stacking deconvolution and unpooling operations with the weight parameters that have been trained in the RDCA. Following this technology, an exemplary plot of four extracted features and the corresponding Deconvnet analysis results have been presented in Figure 7. The symbol L11- i and L11- $d-i$ represent the feature of RDCA-3 in the i th layer of L11 and the feature of Deconvnet in i th layer of L11, respectively.

Figure 7. Deconvnet analysis on the features of RDCA-3.

It can be observed from Figure 7 that each feature corresponds to a part of the original PQD's input image. For example, the part of harmonic in L11-2 is arduous to identify, however, the flicker and harmonic features become clear after the Deconvnet as shown in L11-d-2. Furthermore, those visually illegible abstract features especially for the high-level features in the third layer, such as the L6-3 and the L11-3, both have a clear and explicable meaning. Even though the features are sometimes illegible, it is to be mentioned that each feature corresponds to a portion of the PQD signal.

Overall, the meaning of automatically extracted features has a tremendous difference from the traditional manually selected features. The visualization results as illustrated make automatic feature extraction possible through their differences. However, it is still unknown whether these abstract features can be subdivided. Therefore, a suitable visualization analysis of feature classification is necessary.

4.3. Visual Analysis of Feature Classification Ability

To verify the separability of the features z_{pi} , the stochastic neighbor embedding based on Student-t distribution (t-SNE) is utilized to analyze the high dimensional feature data [31]. Particularly, the t-SNE converts affinities of data points into the Student-t distribution probabilities. Thus, it is suitable for the visualization of high-dimensional PQDs feature images.

For a clearer view of the distribution of features, six types of PQDs are selected and fed into the t-SNE including L1, L5, L10, L13, L14, and L15. Three hundred feature samples are selected per type. Two parameters of the t-SNE, i.e., perplexity and iterations, are set to 40 and 3000, respectively. Figure 8 shows the results of t-SNE on feature data z_{pi}^3 , generating by the third convolutional layer in the RDCA-3.

Figure 8. Visualization of PQD features z_{pi}^3 based on t-SNE and RDCA-3.

Schematically, it indicates from Figure 8 that the distance and location of features for different types of PQDs are distinct. The purpose is to explain the change of disturbance features through visual analysis. Three hundred feature samples are selected per type. For instance, the classes L5 and L14 are distributed on both sides of the graph, respectively. Conversely, some features overlap each other, such as the L14 and the L15, which indicates that they have similar characteristics. The L1 and L5 are gathered together indicating that the corresponding original signals can be reconstructed with fewer features. It can be revealed from the comparison that the features extracted from the RDCA-3 are separable, which provides a guarantee for PQDs classification.

5. Experimental Analysis and Verification

In this section, many experiments are conducted to verify the RDCA and the diagnosis and classification results of PQDs. Concretely, 1600 samples per type are generated. The fold cross-validation is utilized to average output results, where 1040 samples (65%) for each type of disturbance are used for training and 560 samples (35%) each for validation and testing, respectively. To reduce the attenuation rate of feature size and allow multi-layer structure stacking in encoder process, all the size and stride of max-pooling are set to $P_i = 2 \times 2$ and $s = 2$. All the strides of the convolution filter are also set to 1 in RDCA. Grid search is used to select the rest of the model parameters. The training process is implemented on a laptop with the GTX 1060 GPU card. In the present study, the 20 epochs are set for RDCA training based on the Keras framework [32].

5.1. Sensitivity of RDCA Model Parameters

The structure of the model determines the classification performance of the extracted features. In this light, we conducted many experiments to test the performance of diverse layers, sizes of convolution filters, and convergence speeds as is shown in Figure 9. The 20 dB GN is added to the PQDs data set in this test. Moreover, the depth of each layer in the encoder is set to 8, 16, 24, and 32. In addition, we set the equal size of convolution filter per layer in each experiment.

Figure 9. The classification accuracies and loss curve, where the RDCA- i denotes the RDCA has i layers convolution and max-pooling. (a) Accuracy with different layers and filter sizes when $\lambda = 0$, (b) accuracy with different λ , (c) verification loss of SCNN with and without BN layer.

First of all, it is clear from Figure 9a that the classification accuracy increases significantly when the convolution filter size is lower than 5×5 , and the performance of size 7×7 to 13×13 maintain at the same level for a certain layer. Secondly, when the layers are lower than 3, it has been witnessed that the precision is the corresponding increase with the rise of the layer number. The RDCA-3 has the best performance compared with the RDCA-4, indicating that the RDCA-4 may be overfitting due to lots of parameters. Thus a three layers RDCA with the 11×11 convolution filter in the first layer and 9×9 filter in the next two layers are selected.

Based on the RDCA-3, we further verify the validity of the weight residual structure as shown in Figure 9b. It shows that the RDCA performs better than the regular AE ($\lambda = 0$). We set $\lambda = 0.4$ because the RDCA-3 has the highest accuracy. Followed by the RDCA-3, we verify the convergence speed of SCNN. Figure 9c shows that the SCNN with the BN achieves stability when iterations is 15. The loss is also smaller, indicating that the SCNN has learned more information.

5.2. Experiments on the RDCA with Noise

To verify the noise immunity performance of the model, we trained both the three-layer RDCA (RDCA-3) and the three layer regular convolutional auto-encoder (CAE-3) without a denoising strategy. It is to be noted that the reconstructed signal in the decoder is exactly the same as the input for the CAE-3, where the size of the filter is optimally set to 11×11 for the first convolutional layer and 9×9 for the rest layers. Further, the depth of the RDCA and CAE in different layers are sequentially 8, 16, and 24, respectively. The input signal SNR range from 10 dB to 30 dB with a sliding stride of 5 dB. The classification result with different layers and noise is depicted in Figure 10, where the error bar denotes the standard deviation.

It is clear from Figure 10 that the classification accuracy increases as the noise level decreases for both RDCA and CAE. The result is in line with expectations, in which a lower noise level means a cleaner signal. Specifically, when the SNR value is below 20 dB, the advantages of the RDCA-3 are more obvious than the CAE-3. Further, the performance of the RDCA-3 is close to the CAE-3 when the SNR value is higher than 20 dB. Given the difference in reconstruction goals, the classification result indicates that the RDCA has stronger PQD feature extraction capabilities.

Figure 10. The classification accuracies with varying noise.

5.3. Comparison of RDCA, CAE, and CNN

Considering the proposed RDCA integrate the advantages of CAE and CNN and these three methods can all extract the input feature automatically, to verify the validity

of the RDCA, we compare the proposed RDCA with some advanced feature extraction methods, including auto-encoder (AE), CAE [17], and DCNN [33]. The detailed structure of different methods is as follows.

For traditional AE, the input matrix x is directly flattened, i.e., the input layer has 32,768 neuron nodes. In this test, a single-layer (AE-1) and a three-layer AE (AE-3) model are used for comparison. We set the structure of 32,768 to 2000 for the AE-1. Therefore, it means that there are 2000 nodes in the second layer of the encoder. Similarly, the structure of AE-3 is set to 32,768 to 5000 and then 1000 to 512.

For the CAE [17], an optimal single-layer CAE and three-layer CAE with L2 regularization are used to compare. The softmax classifier is used to classify both for AE and CAE.

For the DCNN, a five-layer AlexNet with fewer parameters is used to classify time-frequency information directly [33]. The depth of each layer is set to 8, 16, 32, 64, and 128.

In particular, the PQDs' signal is superimposed with the GN and the SNR is 20 dB. The quantitative performance and the test time are obtained, as listed in Table 2. The test time calculates the part of feature extraction and classification for all test sets (8960 samples), ignoring the time-frequency transform.

It is shown in Table 2 that the proposed RDCA outperforms the AE, CAE, and DCNN. The performance of the RDCA-3 with the softmax classifier is lower than that with the SCNN. Particularly, the accuracy of RDCA-1 is better than DCNN, which indicates that the operation of refactoring contributes to the extraction of effective PQDs features. The performance of AE is not satisfactory due to the lack of convolution operations. On the other hand, the three-layer network always has a better classification result compared to the single-layer models. For the test time analysis, the AE and CAE take less time because the structures of models are simpler with fewer parameters. It can be witnessed that the total time of the RDCA models in Table 2 is similar to the CNN and CAE.

Table 2. Comparison of different feature extraction methods.

Model	Structure	Parameters	Accuracy (%)	Test Time (s)
AE-1	32768-2000	131 106 k	94.02	2.12
AE-3	32768-5000-1000-512	338 749 k	96.32	2.39
CAE-1	L2 regularization	31 k	97.85	5.54
CAE-3 [17]	L2 regularization	136 k	98.07	6.35
DCNN [33]	AlexNet: five layers	3807 k	97.81	8.48
RDCA-1	Classifier: SCNN	31 k	98.5	5.24
RDCA-3	Classifier: Softmax	137 k	98.06	7.19
RDCA-3	Classifier: SCNN	137 k	99.34	6.31

5.4. Comparison with Other PQD Diagnosis Methods

In order to further verify the validity of the proposed method, we compare the RDCA with some recent research results are compared, as listed in Table 3.

Table 3. PQD classification performance comparison.

Diagnosis Methods	Num. of Features	Num. of PQDs	SNR (dB)	Accuracy (%)
FDST and FDT [34]	5	13	30	97.95
ST, TT and PFS [9]	62	11	20	95.2
WT and PNN [30]	9	16	20	93.6
DRST, DAG-SVM [16]	9	9	20	97.77
Hybrid ST and DT [35]	5	11	20	94.36
VMD and DT [36]	4	14	30	96.73
ADLINE and FFNN [7]	2	12	20	90.58
DWT and FFT [37]	10	12	20	99.04
RDCA and SCNN	Automatic	16	20	99.34

It is observed that different methods have different types and quantities of manually selected features. For instance, the method proposed in [9] has 62 kinds of features, while the proposal in [36] has 4 features. However, the accuracy of manually selected feature extraction methods is still lower than the automatic feature extraction. Moreover, sixteen types of disturbances are reported in [30], it is obviously difficult to distinguish all the PQDs by nine features. Particularly, only 30 dB noise conditions of the PQDs are reported in [34,36]. It can be seen from Table 3 that the proposed method performs with the highest classification accuracy under the condition of 20 dB noise level.

5.5. Experimental Verification of Proposed Method

It is worthwhile to verify the performance of RDCA under the experimental data set, which is closer to a real power grid environment. We designed a hardware device to capture the PQD signals from the signal source, as shown in Figure 11. This platform consists of three parts: signal generating unit, measurement unit, and monitoring unit. A high-precision three-phase power quality source Fluke 6100A is utilized for signal generating, which can synthesize sinusoidal and non-sinusoidal power with harmonics, fluctuating harmonics, flicker, interrupt, sags, and swells. Then the hardware circuit with the 16-bit analog to digital converter (ADC) ADS8556 is used to sample real-time PQD signal and digital signal processor (DSP) TMS320C6748 with a clock frequency of 375 MHz is used to process the sampled signal via the DRST. After that, the features of real-time data can be extracted by the RDCA as soon as the DSP sends the result data of DRST to the host PC through a USB serial interface.

Figure 11. Hardware experiment platform and structure diagram for real-time PQD signals diagnosis.

Due to the functional limitation of Fluke 6100A, we generate eight types of PQDs including L1, L2, L3, L4, L5, L7, L10, and L13. Table 4 presents the performance of experimental data generated by the hardware platform. It is noted that a total of 50 samples are generated per type of disturbance. The test time consists of DRST and RDCA, which the DRST and RDCA are calculated in DSP and the host PC under all the samples, respectively.

Table 4. The performance of the experimental data.

Class	Accuracy (%)	Test Time (ms)	Class	Accuracy (%)	Test Time (ms)
L1	100	217.13	L5	98	232.28
L2	94	216.09	L7	92	224.31
L3	96	222.83	L10	94	238.76
L4	96	211.14	L13	96	234.49

As listed in Table 4, it is observed that the performance is a bit lower than the simulation because the experimental signal is more complicated than the simulation model. However, the average accuracy still reaches 95.75%. On the other hand, the test time shows that the combined disturbances are time-consuming compared to a single disturbance. In general, it is caused by the fact that the combined PQDs have more points of key frequency in the DRST analysis. The highest detection time namely, L10, is 4.76 ms for each sample, once the training of the RDCA model is completed. The results also reveal that real-time power quality diagnosis can be realized effectively in a high-frequency computing unit.

6. Conclusions

In this paper, new RDCA and SCNN methodologies have been proposed with automatic extracted features and classification for both single and combined PQD signals. The automatically extracted features, which are based on the primary novel idea of treating the time–frequency matrix as the whole angle, solve the uncertainty of the issue of manually selected features. For the purpose of PQD diagnosis and classification, the parameter sensitivity experiments show that the weight residual structure of the RDCA is helpful to improve the accuracy. Furthermore, the convergence experiment reveals that the BN layer of the SCNN contributes to accelerating convergence and makes the structure of the RDCA more flexible. The visualization results show that the feature maps are able to obtain higher abstractness along with the increment of layers. A variety of comparative experiments manifest that the proposed method has superior precision as well as anti-noise capability than the existing approaches. The designed hardware test platform shows the real-time PQD diagnosis using the proposed method. The implementation performance demonstrates the practicability and robustness of the proposed method. To further improve this methodology, future work will focus on optimization of high-level feature extraction to simplify the model structure.

Author Contributions: Conceptualization, J.L. and W.Q.; methodology, J.L.; software, W.Q.; validation, J.M., Y.Q. and B.S.; formal analysis, J.M.; investigation, J.L.; resources, Q.T.; data curation, Q.T.; writing—original draft preparation, J.L.; writing—review and editing, Q.T.; visualization, W.Q.; supervision, J.L.; project administration, Q.T.; funding acquisition, Q.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China OF FUNDER grant number 52077067.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chang, G.; Lin, Y.; Liu, Y.; Sun, G.; Yu, J. A Hybrid Approach for Time-Varying Harmonic and Interharmonic Detection Using Synchrosqueezing Wavelet Transform. *Appl. Sci.* **2021**, *11*, 752. [\[CrossRef\]](#)
2. Garcia, C.; Grasso, F.; Luchetta, A.; Piccirilli, M.; Paolucci, L.; Talluri, G. A Comparison of Power Quality Disturbance Detection and Classification Methods Using CNN, LSTM and CNN-LSTM. *Appl. Sci.* **2020**, *10*, 6755. [\[CrossRef\]](#)
3. Fu, L.; Zhu, T.; Pan, G.; Chen, S.; Zhong, Q.; Wei, Y. Power Quality Disturbance Recognition Using VMD-Based Feature Extraction and Heuristic Feature Selection. *Appl. Sci.* **2019**, *9*, 4901. [\[CrossRef\]](#)
4. Tang, Q.; Qiu, W.; Zhou, Y. Classification of complex power quality disturbances using optimized s-transform and kernel svm. *IEEE Trans. Ind. Electron.* **2020**, *67*, 9715–9723. [\[CrossRef\]](#)
5. Igual, R.; Medrano, C. Research challenges in real-time classification of power quality disturbances applicable to microgrids: A systematic review. *Renew. Sustain. Energy Rev.* **2020**, *132*, 110050. [\[CrossRef\]](#)
6. Sahani, M.; Dash, P.K. Automatic power quality events recognition based on hilbert huang transform and weighted bidirectional extreme learning machine. *IEEE Trans. Ind. Inform.* **2018**, *14*, 3849–3858. [\[CrossRef\]](#)
7. Valtierra-Rodriguez, M.; de Jesus, Romero-Troncoso, R.; Osornio-Rios, R.A.; Garcia-Perez, A. Detection and Classification of Single and Combined Power Quality Disturbances Using Neural Networks. *IEEE Trans. Ind. Electron.* **2014**, *61*, 2743–2482. [\[CrossRef\]](#)
8. Manikandan, M.S.; Samantaray, S.R.; Kamwa, I. Detection and Classification of Power Quality Disturbances Using Sparse Signal Decomposition on Hybrid Dictionaries. *IEEE Trans. Instrum. Meas.* **2015**, *64*, 27–38. [\[CrossRef\]](#)
9. Lee, C.Y.; Shen, Y.X. Optimal Feature Selection for Power-Quality Disturbances Classification. *IEEE Trans. Power Del.* **2011**, *26*, 2342–2351. [\[CrossRef\]](#)

10. Qiu, W.; Sun, K.; Yao, W.; Wang, W.; Tang, Q.; Liu, Y. Hybrid Data-driven based HVDC Ancillary Control for Multiple Frequency Data Attacks. *IEEE Trans. Ind. Informat.* **2021**. [\[CrossRef\]](#)
11. Kawady, T.A.; Elkalashy, N.I.; Ibrahim, A.E.; Taalab, A.I. Arcing fault identification using combined Gabor Transform-neural network for transmission lines. *Int. J. Electr. Power Energy Syst.* **2014**, *61*, 248–258. [\[CrossRef\]](#)
12. Thirumala, K.; Prasad, M.S.; Jain, T.; Umarikar, A.C. Tunable-Q Wavelet Transform and Dual Multiclass SVM for Online Automatic Detection of Power Quality Disturbances. *IEEE Trans. Smart Grid.* **2018**, *9*, 3018–3028. [\[CrossRef\]](#)
13. Eristi, B.; Yildirim, O.; Eristi, H.; Demir, Y. A new embedded power quality event classification system based on the wavelet transform. *Int. Trans. Electr. Energy Syst.* **2018**, *1*, 1–15. [\[CrossRef\]](#)
14. Ray, P.K.; Mohanty, S.R.; Kishor, N.; Catalao, J.P.S. Optimal feature and decision tree based classification of power quality disturbances in distributed generation systems. *IEEE Trans. Sustain. Energy.* **2014**, *5*, 200–209. [\[CrossRef\]](#)
15. Reddy, M.V.; Sodhi, R. A Modified S-Transform and Random Forests-Based Power Quality Assessment Framework. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 78–89. [\[CrossRef\]](#)
16. Li, J.; Teng, Z.; Tang, Q.; Song, J. Detection and Classification of Power Quality Disturbances Using Double Resolution S-Transform and DAG-SVMs. *IEEE Trans. Instrum. Meas.* **2016**, *65*, 2302–2312. [\[CrossRef\]](#)
17. Chen, K.; Hu, J.; He, J. Detection and Classification of Transmission Line Faults Based on Unsupervised Feature Learning and Convolutional Sparse Autoencoder. *IEEE Trans. Smart Grid* **2018**, *9*, 1748–1758.
18. Sun, K.; Qiu, W.; Yao, W.; You, S.; Yin, H.; Liu, Y. Frequency Injection based HVDC Attack-defense Control via Squeeze-Excitation Double CNN. *IEEE Trans. Power Syst.* **2021**. [\[CrossRef\]](#)
19. Huang, H.; Hu, X.; Zhao, Y.; Makkie, M.; Dong, Q.; Zhao, S.; Guo, L.; Liu, T. Modeling Task fMRI Data Via Deep Convolutional Autoencoder. *IEEE Trans. Med. Imag.* **2018**, *37*, 1551–1561. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Wei, Q.; Tang, Q.; Liu, J.; Yao, W. An Automatic Identification Framework for Complex Power Quality Disturbances Based on Multifusion Convolutional Neural Network. *IEEE Trans. Ind. Inform.* **2020**, *16*, 3233–3241.
21. Qiu, W.; Tang, Q.; Teng, Z.; Liang, C.; Yao, W. Automatic Classification of Power Quality Disturbances Via Weighted Convolutional Neural Network. In Proceedings of the 2019 IEEE Power & Energy Society General Meeting (PESGM), Atlanta, GA, USA, 4–8 August 2019; pp. 1–5.
22. Singh, U.; Singh, S.N. Optimal Feature Selection via NSGA-II for Power Quality Disturbances Classification. *IEEE Trans. Ind. Informat.* **2018**, *14*, 2994–3002. [\[CrossRef\]](#)
23. Vincent, P.; Larochelle, H.; Lajoie, I. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *J. Mach. Learn. Res.* **2010**, *11*, 3371–3408.
24. Masci, J.; Meier, U.; Cireşan, D.; Schmidhuber, J. Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction. *Int. Conf. Artif. Neural Netw.* **2011**, *1*, 52–59.
25. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Proc. Int. Conf. Neural Inf. Process. Syst.* **2012**, *60*, 1097–1105. [\[CrossRef\]](#)
26. 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 (CVPR), Las Vegas, NV, USA, 26–30 June 2016; pp. 770–778.
27. Srivastava, N.; Hinton, G.; Krizhevsky, A. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
28. EM Committee. *IEEE Recommended Practice for Monitoring Electric Power Quality*; IEEE: Washington, DC, USA, 2009; pp. 1159–2009.
29. Zeiler, M.D.; Taylor, G.W.; Fergus, R. Adaptive deconvolutional networks for mid and high level feature learning. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 2018–2015.
30. Khokhar, S.; Zin, A.A.M.; Memon, A.P.; Mokhtar, A.S. A new optimal feature selection algorithm for classification of power quality disturbances using discrete wavelet transform and probabilistic neural network. *Measurement* **2017**, *95*, 246–259. [\[CrossRef\]](#)
31. Van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
32. Keras. Available online: <https://github.com/keras-team/keras> (accessed on 27 June 2021).
33. Liu, H.; Hussain, F.; Shen, Y. Complex power quality disturbances classification via curvelet transform and deep learning. *Electr. Power Syst. Res.* **2018**, *163*, 1–9. [\[CrossRef\]](#)
34. Biswal, M.; Dash, P.K. Measurement and Classification of Simultaneous Power Signal Patterns With an S-Transform Variant and Fuzzy Decision Tree. *IEEE Trans. Ind. Inform.* **2013**, *9*, 1819–1827. [\[CrossRef\]](#)
35. He, S.; Li, K.; Zhang, M. A Real-Time Power Quality Disturbances Classification Using Hybrid Method Based on S-Transform and Dynamics. *IEEE Trans. Instrum. Meas.* **2013**, *62*, 2465–2475. [\[CrossRef\]](#)
36. Achlerkar, P.D.; Samantaray, S.R.; Manikandan, M.S. Variational Mode Decomposition and Decision Tree Based Detection and Classification of Power Quality Disturbances in Grid-Connected Distributed Generation System. *IEEE Trans. Smart Grid* **2018**, *9*, 3122–3132. [\[CrossRef\]](#)
37. Deokar, S.A.; Waghmare, L.M. Integrated DWT-FFT approach for detection and classification of power quality disturbances. *Int. J. Electr. Power Energy Syst.* **2014**, *61*, 594–605. [\[CrossRef\]](#)