Article

Power Quality Disturbance Recognition Using Empirical Wavelet Transform and Feature Selection

Sihan Chen, Ziche Li, Guobing Pan * and Fang Xu *

College of Mechanical Engineering, Zhejiang University of Technology, Hangzhou 310023, China; 1111902002@zjut.edu.cn (S.C.); lzc20000826@163.com (Z.L.)
* Correspondence: gbpan@zjut.edu.cn (G.P.); fangx@zjut.edu.cn (F.X.); Tel.: +86-571-8795-3851 (G.P. & F.X.)

Abstract: With the growth of nonlinear electrical equipment, power quality disturbances (PQDs) often appear in electrical systems. To solve this, a practical heuristic methodology for PQD detection and classification based on empirical wavelet transform has been proposed. By using a multiresolution analysis tool, empirical wavelet transform, the voltage waveform signal is decomposed into several sub-signals, and some potential features are extracted in the statistical method. To reduce the feature vector dimensions, the ReliefF algorithm is used for feature selection and optimized for dimensionality reduction, which reduces the complexity of system calculation while ensuring accuracy. Finally, a classifier based on support vector machines (SVM) was built, and with the ranked feature vectors’ input, the PQD can be recognized. The experimental results verify that the classification results achieved high accuracy, which confirms the properties and robustness of the proposed approach in noisy environments.

Keywords: power quality; empirical wavelet transform; feature selection; pattern recognition; disturbance detection

1. Introduction

Power quality disturbances (PQDs) are generated with the growth of nonlinear loads, such as solid-state switching equipment, electronically switched devices, industrial rectifiers, and inverters. Warped voltage waveforms adversely affect electronic devices, such as electrical system failures, disk crashes, and microcontroller failures [1]. Therefore, it is essential to evaluate the power quality of the electric system by recognizing detailed disturbance events. By recognizing power quality events, an efficient strategy can be carried out to stabilize power grids. The patterns and the reasons for PQDs are multiple. For example, a short trouble may result in a voltage fluctuation, which creates sag or other events [2]. The use of DC–AC inverters may result in harmonics. High-power cost equipment may lead to voltage flickers. Transient and spike events may be caused by electric discharge or sudden switching-off with overrides.

In this paper, a new algorithm is proposed for PQD detection and classification on an experimental microgrid platform. Considering the deficiencies in the reviewed literature, this proposed method is based on the EWT for decomposition, permutation entropy, and the ReliefF algorithm to obtain the relative classification features. The inspiration for the proposed approach comes from [3], in which the EWT applies three practical criteria to segment the noisy and nonstationary signal spectrum. In particular, the proposed method is based on five steps: analytical PQD signal generation, signal segmentation, feature extraction, feature selection, and pattern classification. By using a multiresolution analysis tool, an enhanced empirical wavelet transform, the voltage waveform signal is decomposed into several sub-signals. Then, some potential features are extracted from decomposed signals in the statistical method. To reduce feature vector dimensions, permutation entropy and ReliefF algorithm are used for feature selection. Finally, the optimal features are selected as the input of the SVM model to prove PQ recognition. Several experiments are
performed on both simulation data and a real test platform to verify the effectiveness of the proposed method.

The main contributions of this article are as follows:

1. Compared with EMD, the EWT is more suitable and robust to a segmented PQD signal, especially in the selection of decomposed parameter optimization.
2. Based on experiment results, the permutation entropy and ReliefF algorithm are vital for feature selection, which help to filter the significant components, eliminating redundant features.
3. The proposed algorithm observes the characteristics of the PQD signal from a multiresolution perspective and improves the robustness and the precision of the classifier model.

2. Related Work

Different studies have been promoted for detecting power quality events. Most PQD-detecting solutions are implemented with the following processes: feature extraction, feature dimension optimization, and pattern recognition. Feature extraction is essential for PQD classification. It is noted that features can be extracted directly from the original signal or the transformed formalization. Fourier transform (FT) is a famous method, but only suitable for stationary signal processing. In other words, FT cannot resolve local information with fluctuations or transient components. Short-time Fourier transform (STFT) utilizes a sliding window on the original signal. Then, the small slice signal can be treated as many small stationary segments for FT analysis. However, time–frequency resolutions are settled along with the determined window size [4]. In [5], the authors combine wavelet transforms (WTs) to characterize PQD events. Additionally, it was certified that WT performance is better in less-noisy environments but sensitive to noise interference when detecting PQDs [6]. In a practical scene, the noises are always coupled with PQD signals, which would limit the application of a WT based algorithm. Hence, it is urgent to search for an effective multiresolution decomposition method with noisy-tolerance robustness.

To overcome the problems of the algorithms referred to above, S-transform (ST) is imported as an optimal time–frequency analysis tool by utilizing an adaptive Gaussian window. In [7,8], Mahela et al. and He et al. presented a method for PQD classification by adopting ST, respectively. In [9], the ST technique was combined with long short-term memory network structures for automatic PQD classification. Furthermore, an S-transform connected with a fuzzy system for PQD recognition was presented in [10]. Moreover, wavelet packet transform (WPT) is also popular in PQD detection due to its excellent performance [11,12]. These referred algorithms have achieved in PQD events recognition. Nevertheless, it is known that STFT, ST and WPD all utilized a fixed transform base for signal decomposition. They may be effective only in some specific scenario due to its fixed decomposition. Therefore, recently, the adaptive decomposition algorithm has been widely used in signal processing. For instance, empirical mode decomposition (EMD) shows excellent local adaptive performance in PQD research. In [13], Manjula et al. employed EMD to recognize voltage sag disturbances. The core of EMD is that it decomposes a nonstationary signal into monocomponent and symmetric signals, which are called intrinsic mode functions (IMFs). In [14], the authors investigated ensemble empirical mode decomposition performance for PQD signal processing. Although EMD has achieved many successes in PQD detection, the restriction cannot be ignored for practical application. To be specific, some IMFs cannot reveal effective details in frequency domains. Additionally, most decomposed IMFs are not analytical in mathematical theory.

Considering modal aliasing and a lack of mathematical applications, Gilles [15] developed empirical wavelet transform (EWT) to decompose a signal based on wavelet transform and spectrum dividing. The core idea of EWT is utilizing the local maximum criterion to divide the spectra into several segmentations. Then, a series of wavelet filters can be adopted to adaptively reconstruct different sub-components from the target signal. In [16,17], EWT is used to analyze voltage and current signals for PQD recognition. As concluded from the experimental results, it was noted that the EWT produces a better result
than DWT and WPT in terms of the percentage of errors. Additionally, the modes revealed by the EWT benefit in terms of feature extraction.

Thereafter, the features can be extracted from the processed signals, which reveal significant information for PQDs. Common statistical features include root-mean-square values (RMS), variances, standard deviations, means, ranges, maximum values, minimum values, skewness, kurtosis, and entropies [18]. Then, the extracted features are fed into a classifier model to recognize different PQD events. Decision tree (DT), support vector machines (SVM), and artificial neural networks (ANNs) are widely used for PQD classification [19]. In [20], Naderian et al. proposed a novel algorithm to detect PQ events by combining Discrete Gabor Transform with Type-2 Fuzzy Kernel-based Support Vector Machine. Liu et al. utilized ensemble empirical mode decomposition and Ranked Wavelet SVM to classify complex PQD events [21]. In [22], authors adopted WT, hyperbolic S-transform, and DT to detect PQD events on modified Nordic 32-bus test systems. Experimental results revealed that the proposed algorithm showed better classification accuracy. In [23], an ST combined with an ANN classifier in conjunction with a Kalman filter based on maximum likelihood was developed to recognize PQD events. The ANN exhibits superior performance but struggles in small sample learning problems. The DT requires less training samples but may produce a complex model with less generalization, which would lead to a overfitting result. Compared with these classification methods, the SVM shows excellent performance for the following reasons: nonlinear classification boundary, comparatively multi-dimensional space, dependent data.

Through the literature review, numerous works have achieved outstanding results, but several limitations still exist. For instance, most of the research has discussed the importance of signal processing and feature extraction. Feature-extraction algorithms are usually based on experiential knowledge or are directly handcrafted. Unfortunately, few studies developed feature selection algorithms for PQD recognition. It is apparent that not all features are valid for classifier models. Redundant features carrying useless information may decrease PQD classification accuracy and increase the computational expense.

3. Theoretical Background

3.1. Basic Theory of Empirical Wavelet Transform

EWT’s main principle is decomposing a signal \( f(t) \) into several intrinsic mode functions (IMFs) based on wavelet transform and spectrum dividing. According to the local maximum characteristics in frequency spectrums, a series of wavelet filters are adaptively adopted to extract different IMFs. The extracted IMFs are defined as \( c_k \). Additionally, a residual error, defined as \( r_n \), is reserved after decomposition. Specifically, the frequency spectrum of the target signal is adaptively segmented based on local maximum values by conducting empirical wavelet filters. Then, the separation of AM–FM components in the Fourier domain is concluded. Among them, the instantaneous frequency and the instantaneous amplitude are procured from the AM–FM module by Hilbert transformation. Based on the definition, EWT has a clear wavelet theoretical framework, which avoids modal aliasing in EMD. Moreover, the decomposed IMFs are explicable for further processing.

\[
f(t) = \sum_{k=1}^{N} c_k + r_n
\]

(1)

To adaptively separate the Fourier spectrum, the signal is divided into \( N \)-continuous components, and the local maximum values in the spectrogram are detected and arranged in descending order. Then, we can use the midpoint of two consecutive maxima as the boundaries for segments. As shown in Figure 1, the spectrum range is limited to between 0 and \( \pi \), and the boundary of each adjacent segment is defined as \( \omega_n \). Meanwhile, we can
define a transition zone with a width of $T_n = 2\tau_n$. Each segment is expressed as between $\omega_{n-\frac{1}{2}}$ and $\omega_n$ as Equation (2).

\[
\Lambda_n = [\omega_{n-\frac{1}{2}}, \omega_n], \quad n = 1, 2, \ldots, N
\]

\[
\bigcup_{n=1}^{N} \Lambda_n = [0, \pi]
\]

The Meyer wavelet is selected by Gilles [15] based on the construction criterion. The empirical scale function and the empirical wavelet function are defined as Equations (3) and (4), respectively.

\[
\hat{\phi}_n(\omega) = \begin{cases} 
1 & , \quad |\omega| \leq (1-\gamma)\omega_n \\
\cos\left(\frac{\pi}{\pi\omega_n}\left(|\omega| - (1-\gamma)\omega_n\right)\right), & \frac{1}{2}(1-\gamma)\omega_n \leq |\omega| \leq (1+\gamma)\omega_n \\
0 & , \quad \text{others}
\end{cases}
\]

(3)

\[
\hat{\psi}_n(\omega) = \begin{cases} 
1 & , \quad |\omega| \leq (1-\gamma)\omega_n \\
\cos\left(\frac{\pi}{\pi\omega_n}\left(|\omega| - (1-\gamma)\omega_n\right)\right), & \frac{1}{2}(1-\gamma)\omega_n \leq |\omega| \leq (1+\gamma)\omega_n \\
0 & , \quad \text{others}
\end{cases}
\]

(4)

The parameter $\tau_n$ is chosen to guarantee the transform that provides a tight frame. The function $\beta(x)$ is the auxiliary function of the Meyer wavelet, which is defined as follows:

\[
\beta(x) = x^4 \left(35 - 84x + 70x^2 - 20x^3\right)
\]

(5)

\[
\tau_n = \gamma \omega_n
\]

(6)

\[
\gamma < \min_n \left(\frac{\omega_{n+1} - \omega_n}{\omega_{n+1} + \omega_n}\right)
\]

(7)

After deriving the empirical wavelets and the scaling function, the approximated coefficients are defined as the inner product of the signal and the scaling function:

\[
W^0_f(0, t) = \langle f, \phi_1 \rangle = \int f(\tau)\hat{\phi}_1(\tau - t) d\tau
\]

(8)

The detail coefficients are defined as the inner product of the signal and the empirical wavelets:

\[
W^k_f(n, t) = \langle f, \psi_n \rangle = \int f(\tau)\hat{\psi}_n(\tau - t) d\tau
\]

(9)

Then, the empirical modes $f_k(t)$ decomposed from the signal are given by:

\[
f_0(t) = W^0_f(0, t) \ast \phi_1(t)
\]

\[
f_k(t) = W^k_f(n, t) \ast \psi_k(t)
\]

(10)
Then, the signal reconstruction is calculated as follows, where:

\[ f(t) = W_f^e(0, t) \ast \phi_1(t) + \sum_{n=1}^{N} W_f^e(n, t) \ast \psi_n(t) \]

\[ = F^{-1} \left[ W_f^e(0, \omega) \hat{\phi}_1 + \sum_{n=1}^{N} W_f^e(n, \omega) \hat{\psi}_n(t) \right] \tag{11} \]

3.2. Permutation Entropy

As concluded from information theory, entropy is an effective method to calculate the uncertainty degree of stochastic systems. For an observation system, when the randomness increases, the entropy value increases as well. According to entropy’s definition, it can be adopted to evaluate chaotic degrees for PQD signals. Moreover, entropy is also able to qualify for feature extraction in signal processing. Specifically, permutation entropy is an effective tool in feature statistics [24]. By calculating the sequence signal in time domains, the transient features can be extracted. In another word, the permutation entropy can assess the sequence complicacy after balancing the neighboring values. According to a mathematical model, the permutation entropy is computed by the time sequence probability density function, which is similar to Shannon entropy. For a time data series \( x(t) \) with a limited length of \( N \), where \( t \) is increased from 1 to \( N \), a fragment can be structured by an \( m \)-order dimension. Then, the obtained fragment is ranked in an increasing sequence that is annotated as \( \pi \). After, we can set the time delay as \( \alpha \), and the \( m \)-order fragment \( X \) is defined as follows:

\[ X^m_i = \{ x(i + \alpha(j_1 - 1)), x(i + \alpha(j_2 - 1)), \ldots, x(i + \alpha(j_m - 1)) \} \tag{12} \]

From experimental trials, the time delay parameter \( \alpha \) is selected as 1, and the value \( m \) is selected from three to seven. Hence, the number of \( m \)-order permutations \( \pi \) is the factorial of \( m \). The relative occurrence frequency of each permutation \( \pi \) is presented as follows:

\[ p(\pi) = \frac{H\{X^m_i \text{ has type } \pi, i | 1, 2, \ldots, N - m + 1\}}{N - m + 1} \tag{13} \]

where \( H \) represents the count number. After, the permutation entropy is estimated by the probability density function for the relative frequency:

\[ PE(m) = -\sum_{i=1}^{m} p(\pi_i) \log(p(\pi_i)) \tag{14} \]

3.3. ReliefF-Based Feature Selection Method

The ReliefF algorithm [25], which is a part of the Relief series, is a supervised approach for selecting features. The gap between cases from a similar classification and those from a different classification is closer for a distinguishing feature. As a result, by analyzing the relevant weight between features and classifications, the Relief algorithm can be used as a filter to select the best characteristics. The ReliefF algorithm was created to address multi-class problems because the original Relief algorithm was limited to binary classification. It is more robust and noise tolerant.

Each instance in a training dataset, \( D = d_1, d_2, \ldots, d_m \), has \( p \) features, which are defined as the parameter \( i \). The \( i \) value ranges from 1 to \( m \). The ReliefF algorithm selects an instance \( d_i \) from the dataset when utilizing binary classification. The next step is to check for two neighbors at a reasonable distance. One of the neighbors belongs to the same class as the nearest hit \( H \), while the other belongs to a different class as the nearest miss \( M \). For each feature \( t \), the weight coefficient \( W_t \) is calculated as:

\[ W_t = W_t - \text{diff}(t, d_i, H) / r + \text{diff}(t, d_i, M) / r \tag{15} \]
For numerical features, \( \text{diff}(t, R_i, R_j) \) is defined as:

\[
\text{diff}(t, R_i, R_j) = \left| \frac{R_{i,t} - R_{j,t}}{\max_i - \min_i} \right|
\]

(16)

where \( \min_i \) and \( \max_i \) are the minimum and maximum values of feature \( t \) in dataset \( S \), respectively. The class label for a multiclass problem is \( C = C_1, C_2, \ldots, C_i \). In the same class, the Relief\(^F\) algorithm looks for the \( k \) closest neighbors. Additionally, it searches for the \( k \) closest neighbors in each class.

4. Proposed Method

Figure 2 depicts the suggested methodology’s implementation process, which is broken down into five steps: PQD signal collection, signal pre-processing, feature extraction on IMFs, feature-optimizing selection, and PQD event recognition. Then, the performance of the algorithm is discussed to verify its effectiveness. Although the proposed methodology is effective in PQD recognition, it still has limitations in some application scenarios when dealing with the wrong training tags. In other words, our classifier model is obtained under supervised learning, which requires that the training samples are well-tagged. When facing wrongly tagged samples, our model may not be applicable. It is noted that this situation is universal for all supervised learning models. However, we can import a semi-supervised learning strategy, such as cooperative training, to solve this problem.

![PQD detection and classification Process](image)

**Figure 2.** Main flowchart of the proposed method.

4.1. PQD Signal Simulation

Following the IEEE-1159 standard [26], a collection of signals is generated artificially in this work using mathematical models. Ten different PQD signals are selected, and the detailed models are shown in Figure 3. Specifically, the waveform is collected at 2000 data points per second at a frequency of 10 kHz. The amplitude of the fundamental sine signal, annotated as \( A \), is normalized as 1, which randomly varies within \( \pm 10\% \). Moreover, to simulate real scenes, the artificial signal is coupled with Gaussian white noise. The different signal-to-noise ratios (SNRs) are set at 30 dB.

Among Table 1, there are many parameters for PQD mathematical models [27]. The explanations are listed as follows. Symbol \( \omega \) represents the fundamental angular frequency in the power grid. Symbol \( a \) is used to adjust the wave’s amplitude. For instance, in the sag disturbance event, we set \( a \) ranging from 0.1 to 0.9. Then, the voltage amplitude would drop from 10% to 90%. Symbols \( l_1 \) and \( l_2 \) represent the start and end times for the occurring disturbances. Symbol \( T \) represents a cycle time for power grids, which is 20 milliseconds under 50 Hz. Symbol \( f_n \) represents the oscillatory frequency for the oscillatory transient event. Symbols \( \alpha \) and \( \beta \) are used to tune the flickering shape. Symbol \( K \) is used to tune the amplitude for the notch events.
Table 1. PQD mathematical model [27].

<table>
<thead>
<tr>
<th>Label</th>
<th>PQD Type</th>
<th>Mathematical Model</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>Pure Sine</td>
<td>$y(t) = A \sin(\omega t)$</td>
<td>$A = 1$; $\omega = 2\pi \times 50$</td>
</tr>
<tr>
<td>C2</td>
<td>Sag</td>
<td>$y(t) = A[1 - a(u(t - t_1) - u(t - t_2))] \sin(\omega t)$</td>
<td>$0.1 \leq a \leq 0.9$; $T \leq t_2 - t_1 \leq 97$</td>
</tr>
<tr>
<td>C3</td>
<td>Swell</td>
<td>$y(t) = A[1 + a(u(t - t_1) - u(t - t_2))] \sin(\omega t)$</td>
<td>$0.1 \leq a \leq 0.8$; $T \leq t_2 - t_1 \leq 97$</td>
</tr>
<tr>
<td>C4</td>
<td>Interrupt</td>
<td>$y(t) = A[1 - a(u(t - t_1) - u(t - t_2))] \sin(\omega t)$</td>
<td>$0.9 \leq a \leq 1$; $T \leq t_2 - t_1 \leq 97$</td>
</tr>
<tr>
<td>C5</td>
<td>Harmonics</td>
<td>$y(t) = A[a_1 \sin(\omega t) + a_3 \sin(3\omega t) + a_5 \sin(5\omega t) + a_7 \sin(7\omega t)]$</td>
<td>$0.05 \leq a_3, a_5, a_7 \leq 0.15$; $\sum a_i^2 = 1$</td>
</tr>
<tr>
<td>C6</td>
<td>Sag with harmonics</td>
<td>$y(t) = A[1 - a(u(t - t_1) - u(t - t_2))][a_1 \sin(\omega t) + a_3 \sin(3\omega t) + a_5 \sin(5\omega t)]$</td>
<td>$0.1 \leq a \leq 0.9$; $T \leq t_2 - t_1 \leq 97$</td>
</tr>
<tr>
<td>C7</td>
<td>Interrupt with harmonics</td>
<td>$y(t) = A[1 - a(u(t - t_1) - u(t - t_2))][a_1 \sin(\omega t) + a_3 \sin(3\omega t) + a_5 \sin(5\omega t)]$</td>
<td>$0.05 \leq a_3, a_5, a_7 \leq 0.15$; $\sum a_i^2 = 1$</td>
</tr>
<tr>
<td>C8</td>
<td>Flicker</td>
<td>$y(t) = A[1 + a_5 \sin(5\omega t)] \sin(\omega t)$</td>
<td>$0.1 \leq a \leq 0.5$; $5 \leq \beta \leq 20$ Hz</td>
</tr>
<tr>
<td>C9</td>
<td>Oscillatory transient</td>
<td>$y(t) = A \left[\sin(\omega t) + a e^{-t(t_1-t_2)/\tau} \sin(\omega_n(t_1-t_2))(u(t_1) - u(t_2))\right]$</td>
<td>$0 \leq a \leq 0.5$; $500 \leq f_s \leq 900$ Hz</td>
</tr>
<tr>
<td>C10</td>
<td>Periodic notch</td>
<td>$y(t) = \sin(\omega t) - \text{sign}(\sin(\omega t))$</td>
<td>$0.01 \leq \omega \leq 0.057$; $0 \leq K \leq 0.4$</td>
</tr>
</tbody>
</table>

Figure 3. The waveform of 10 representative PQD signals.

4.2. EWT Signal Segmentation

As can be observed from the mathematical models, the PQD signal is coupled with distinct components. Hence, it is essential to decompose the signal for feature extraction. In this work, the EWT is imported for signal processing. As mentioned in Section 2, the EWT can separate the objective signal into AM–FM components, which guarantees that the PQD changes can be traced more accurately. Moreover, the EWT can reserve useful components effectively and remove noisy components. The EWT parameters are optimized for PQD signals by experimental trials, which will be discussed in the following sections.
In this work, the decomposition number selected is six, which means the target signal is decomposed into six segmentations in spectral domains.

4.3. Feature Extraction and Selection

After EWT decomposition, nine statistical features are extracted from the sub-components. The detailed features are listed in Table 2, which is inspired by [28].

<table>
<thead>
<tr>
<th>No.</th>
<th>Statistical Feature</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RMS</td>
<td>[ \sqrt{\frac{1}{K} \sum_{k=1}^{K} [x(k)]^2} ]</td>
</tr>
<tr>
<td>2</td>
<td>Average</td>
<td>[ \frac{\sum_{k=1}^{K}</td>
</tr>
<tr>
<td>3</td>
<td>Standard Deviation</td>
<td>[ \sqrt{\frac{\sum_{k=1}^{K} [x(k) - \bar{x}]^2}{K(K-1)}} ]</td>
</tr>
<tr>
<td>4</td>
<td>Variance</td>
<td>[ \frac{\sum_{k=1}^{K} x(k)}{K} ]</td>
</tr>
<tr>
<td>5</td>
<td>Range</td>
<td>[ \max {x(k)} - \min {x(k)} ]</td>
</tr>
<tr>
<td>6</td>
<td>Kurtosis</td>
<td>[ \sum_{k=1}^{K} \left( \frac{[x(k) - \bar{x}]}{\sigma} \right)^4 / K ]</td>
</tr>
<tr>
<td>7</td>
<td>Skewness</td>
<td>[ \sum_{k=1}^{K} \left( \frac{[x(k) - \bar{x}]}{\sigma} \right)^3 / K ]</td>
</tr>
<tr>
<td>8</td>
<td>Average Deviation</td>
<td>[ \sum_{k=1}^{K} [x(k) - \bar{x}] / K ]</td>
</tr>
<tr>
<td>9</td>
<td>Permutation Entropy</td>
<td>[ -\sum_{i=1}^{m!} p(\pi_i) \log(p(\pi_i)) ]</td>
</tr>
</tbody>
</table>

Table 2. Statistical feature models.

To be specific, since the PQD signal is decomposed into six IMFs, the referred nine statistical features are imported for each IMF. In other words, fifty-four features are obtained to construct feature vectors. As referred to in Equation (17), the constructed feature vectors are defined as \( F(i) = [F_1, F_2, \ldots, F_9] \), where \( i \) is the \( i \)-th order of the data sample. \( F_1, F_2, \ldots, F_9 \) are the RMS feature vectors, standard deviation annotated as \( \sigma \), variance annotated as \( \text{var} \), range, skewness annotated as \( \text{SK} \), kurtosis annotated as \( \text{KT} \), mean annotated as \( \mu \), average deviation annotated as \( \overline{\text{var}} \), and permutation entropy annotated as \( \text{PE} \), respectively.

\[
F_1 = [\text{RMS}_1 \ \text{RMS}_2 \ \text{RMS}_3 \ \text{RMS}_4 \ \text{RMS}_5 \ \text{RMS}_6]
\]
\[
F_2 = [\mu_1 \ \mu_2 \ \mu_3 \ \mu_4 \ \mu_5 \ \mu_6]
\]
\[
F_3 = [\sigma_1 \ \sigma_2 \ \sigma_3 \ \sigma_4 \ \sigma_5 \ \sigma_6]
\]
\[
F_4 = [\text{var}_1 \ \text{var}_2 \ \text{var}_3 \ \text{var}_4 \ \text{var}_5 \ \text{var}_6]
\]
\[
F_5 = [\text{range}_1 \ \text{range}_2 \ \text{range}_3 \ \text{range}_4 \ \text{range}_5 \ \text{range}_6]
\]
\[
F_6 = [\text{KT}_1 \ \text{KT}_2 \ \text{KT}_3 \ \text{KT}_4 \ \text{KT}_5 \ \text{KT}_6]
\]
\[
F_7 = [\text{SK}_1 \ \text{SK}_2 \ \text{SK}_3 \ \text{SK}_4 \ \text{SK}_5 \ \text{SK}_6]
\]
\[
F_8 = [\bar{\sigma}_1 \ \bar{\sigma}_2 \ \bar{\sigma}_3 \ \bar{\sigma}_4 \ \bar{\sigma}_5 \ \bar{\sigma}_6]
\]
\[
F_9 = [\text{PE}_1 \ \text{PE}_2 \ \text{PE}_3 \ \text{PE}_4 \ \text{PE}_5 \ \text{PE}_6]
\]
\[
F(i) = [F_1 \ F_2 \ F_3 \ F_4 \ F_5 \ F_6 \ F_7 \ F_8 \ F_9]
\]

One more point should be noticed in that these statistical features are calculated in different units. Therefore, the extracted features should be normalized, ranging from 0 and 1, for classifier models. As shown in Equation (18), the min–max normalization method is used in this work. \( Z_i \) is defined as the normalized data, \( F_i \) is the corresponding feature of the \( i \)-th training sample, \( F_{\text{max}} \) is the maximum value of the features, and \( F_{\text{min}} \) is the minimum value.

\[
Z_i = \frac{F_i - F_{\text{min}}}{F_{\text{max}} - F_{\text{min}}}
\]
also be regarded as a dimensionality reduction. By importing the ReliefF algorithm, the vector dimensions are shortened, which benefits computational expenses.

4.4. Multiclass Pattern Recognition

In this study, the multiclass SVM is imported to recognize PQD events. Considering that the basic SVM is only able to achieve binary classification, it is essential to expand a multiclass strategy for PQD recognition. Specifically, the “one-versus-rest” (OVR) technique is widely used for multiclass categorization [27]. When facing an N-class problem, N numbers of hyperplanes are essential for the SVM classifier model. When dealing with the k-th pattern recognition, the multiclass SVM model regards the samples in the k-th class as the positive samples for training. Obviously, the rest of the samples are negative. Then, the nearest distance searching strategy is utilized to judge where the new sample belongs.

Moreover, the SVM parameters are vital for the recognition results. For this research, the radial basis function (RBF) is selected as the SVM kernel function. The RBF kernel has smooth characteristics, which are applicable for training data without prior knowledge. For the RBF, the penalty parameter, annotated as C, is used to tune the tolerance for classification error. When the penalty value is set high, it means the SVM model has a small error tolerance to the training results. In this circumstance, it may make the classifier overfitted, which means the generalization ability will be poor. Oppositely, when the penalty value is small, it means the SVM model allows a relatively large error in the classification results. In this circumstance, the SVM model may exhibit bad precision. Based on the definition, the grid search strategy is performed to optimize the penalty parameter C. After experimental trials, the penalty parameter C is selected as 25.6.

5. Experimental Results and Discussion

5.1. Experimental Setup

A microgrid platform is developed in a laboratory environment to test the suggested method as shown in Figure 4. The microgrid platform adopts a double-layer structure approach with a flexible topology, which is composed of loads, distributed power sources, and energy storage devices. Specifically, typical loads in the microgrid platform include electric motors, transmission mechanisms, induction motors, electric vehicle charging loads, and a 100 kW programmable PLC load. Additionally, distributed power sources include photovoltaic cells, photovoltaic simulators, diesel generator simulators, and wind generators, and energy storage devices consisting of lead batteries. The microgrid platform is able to work at 400 V at 50 Hz. The microgrid topology used in the simulation verification adopts the standard low-voltage microgrid system proposed by the European Union Microgrid Project (ENK5-CT-2002-00610). A variety of lines and load types can be configured in the system, with two sub-micro grid layers coupled with bidirectional converters, allowing the bidirectional flow of power transmission. More specifically, the microgrid platform can use the line impedance simulator to recreate different line faults under the logic program control.

5.2. EWT Decomposition

A simulated signal representing the sag with harmonics is shown to verify the efficiency for EWT. The PQD signal waveform is presented in Figure 5. Observed from the time-series signal, the voltage sag occurs from 40 milliseconds and 110 milliseconds. Meanwhile, the signal is coupled with third, fifth, and seventh harmonics. Figures 6–9 show the decomposition results by EWT in time domains, where the band number N is selected as 4, 5, 6, and 8. The spectrum with segmentation lines is also shown. When N = 4 in Figure 6a, the No. 1, No. 2, and No. 3 sub-components are separated effectively in each IMF. However, the No. 4 subcomponent reveals that two different harmonics are mixed together. It is also more obvious in Figure 6b that the No. 4 spectral segmentation contains the fifth and seventh harmonics. Based on this result, it was found that the decomposition results are less decomposed. This means that the decomposition
number N should increase. When N = 5 in Figure 7, although the No. 5 component still contains Gaussian noise, the decomposition results show that all the main components are separated effectively. Furthermore, there is less noise influence with the No. 2, No. 3, No. 4, and No. 5 components. Figure 8a shows that components are separated cleanly when N = 6. It is further confirmed in Figure 8b that the fundamental, third, fifth, and seventh harmonics are reserved in No. 2, No. 3, No. 4, and No. 5 spectral segmentations. Figure 8 proves that the segmentation bands are well fitted. The significant information is extracted by the EWT.

**Figure 4.** Experimental platform for electronic system simulation.

**Figure 5.** Original sag with harmonics signal.
Figure 6. PQD signal decomposition results by EWT when $K = 6$. (a) Results in the time domain; (b) Segmented spectrum.

Figure 7. PQD signal decomposition results by EWT when $K = 5$. (a) Results in the time domain; (b) Segmented spectrum.

Figure 8. PQD signal decomposition results by EWT when $K = 6$. (a) Results in the time domain; (b) Segmented spectrum.
which corresponds to the disturbance times. Then, we can import the entropy calculation problems. (Figure 10. Decomposition results in time domain by EMD. ) Segmented spectrum. Importantly, No. 6, No. 7, and No. 8 components reveal some valid information. For instance, the No. 8 component exhibits two pulses in 40 milliseconds and 110 milliseconds, which corresponds to the disturbance times. Then, we can import the entropy calculation to evaluate the chaos degree for each IMF. The permutation entropy of each IMF is 0.1246, 0.1434, 0.1980, 0.2468, 0.2874, 0.3876, 0.3868, and 0.9723, respectively. Due to the entropy analysis, the No. 8 component is chaotic and should be removed for further processing.

Moreover, as mentioned in [29], EMD is adopted to present the effectiveness of the EWT. Figure 10 presents the decomposition result by EMD. As can be observed from Figure 10, the decomposition level is nine in self-adaption. Unfortunately, most decomposed components are chaotic, and it is difficult to extract useful features. From the comparison, it was found that EMD has some limitations in PQD recognition. Specifically, the number of intrinsic mode functions (IMFs) changes according to the waveform of PQ events, which proves that the EMD does not behave well in dealing with noisy and nonstationary signals. Besides this, the decomposed IMF is not analytical compared with the IMF from EWT. According to the above analysis, the EWT is adaptable and robust for the PQD signal process, which exhibits excellent performance in mode mixing and decomposed-number-selection problems.

Figure 9. PQD signal decomposition results by EWT when K = 8. (a) Results in the time domain; (b) Segmented spectrum.

Figure 10. Decomposition results in time domain by EMD.
Additionally, two other PQD events, swell with harmonics and periodic notch, are imported to verify the effectiveness of EWT in further investigations. The results are exhibited in Figures 11 and 12. In these two events, the decomposition number N is selected as eight. This is to guarantee that the PQD signal can be decomposed sufficiently. As can be observed from Figure 11 for swell with harmonics, it was found that the disturbance components are separated adaptively by EWT. The results are similar to the sag with harmonics. The difference is only in the No. 2 component. In Figure 9a, the fundamental wave at 50 Hz is reserved, while the amplitude is dropped during 40 milliseconds to 110 milliseconds. Inversely, in Figure 11a, the fundamental wave also remained, but the amplitude rose during the same time. In Figure 12, the periodic notch event is also decomposed by EWT. As can be seen from Figure 12a, it is found that the periodic notch signal consists of numerous periodic impulse components in the time domain, which is obviously detected in the frequency domain. After EWT decomposition, the key information of these PQD signals can be extracted in further processes.

As in [15], “if for a given class of signals it should be possible to guess the best number of modes, this is not the case in general where no a priori information are available”. In such cases, it is essential to estimate the appropriate number of modes. In the referred article, the author suggested offering a deeper analysis but using a simple method to estimate the number of mode N. To be specific, Gills suggested that “the most important maxima in the magnitude of the Fourier transform of the input signal (corresponding to the center of each desired Fourier segments) are significantly larger than the other existing maxima”. In this framework, the above idea is equivalent to keep all maxima that are greater than the amount of the difference between the bigger maximum and the smaller maximum. We can formulate this as “keeping all maxima larger than the threshold”. Moreover, we also have the prior information of the PQD events from the mathematical model. Hence, we can select several different numbers of EWT for trials.

However, the EWT has a problem with the local maxima method. In other words, the EWT has issues with spectrum division. Therefore, we can firstly adopt the EWT to process the PQD signals. The decomposition results are presented in Figures 5–12. As can be observed from these figures, it was found that the EWT has a flexible ability to divide the spectrum into segmentations for each component. The time and spectrum domains show that the main components are well separated. The corresponding boundary obtained will fall in the largest support of the first mode due to the local maxima algorithm. Therefore, we chose the “poly” global trend removal to estimate the global trend by a polynomial interpolation with 6 degrees for the first mode. Then, we can see the second mode of the EWT is obtained as the fundamental signal wave of the power grid. Besides that, it was also found that the amplitude of the high-order harmonics is larger than the background noise, which also guarantees the effectiveness of the local maxima method.
Figure 11. Swell with harmonics decomposition results by EWT when K = 8. (a) Results in the time domain; (b) Results in the frequency domain.

Figure 12. Periodic notch decomposition results by EWT when K = 8. (a) Results in the time domain; (b) Results in the frequency domain.
5.3. ReliefF Feature Selection

As an optimal feature selection approach, the ReliefF algorithm was adopted to achieve dimension reduction for the original feature vectors. The embedded dimension parameter of the permutation entropy was set to six in this case. The permutation entropy’s time delay parameter was set as one. The permutation entropy hard barrier was set to 0.6.

Two feature selection methods, random feature selection and ReliefF algorithm, were used for comparison. Figure 13a,b present the precision and recall results through random feature selection and ReliefF algorithm, respectively. All the accuracy trends rise with the feature number increasing at the early stage. This is because the first selected features would help to train the classifier model directly. In Figure 13a, the random feature selection algorithm obtains the highest accuracy when selecting 22 top-ranked features. At this time, the average precision and average recall ratio reached 97.3 percent and 96.5 percent, respectively. As a comparison, in Figure 13b, the ReliefF feature selection algorithm obtains the highest accuracy when selecting 17 top-ranked features. Meanwhile, the average precision and average recall ratio achieved by ReliefF reached 98.2 percent and 97.9 percent, respectively. The reason for the rising accuracy is easy to understand. When the feature number is small, the classifier model lacks training. Hence, increasing features would help to construct the classification hyperplane for SVM models.

![Figure 13](image1.png)

**Figure 13.** Recognition accuracy under different selected features. (a) Training result by random selection; (b) Training result by ReliefF selection.

Nevertheless, after the best-feature dimension, both classifier models drop in accuracies due to the redundant features involved. After decreasing, the precision and the recall values rise a little again. The reason for this is that we selected the RBF as the SVM kernel. The slack variable and the penalty parameters import soft intervals, which contribute to building a decision boundary under a tolerant error. The soft interval boundary improves the robustness and generalization of SVM models.

Then, the classifier accuracy is tested under different ratios of training sets to further investigate the effectiveness and robustness of the feature selection procedure. The training set’s percentages were increased by 10 percent, 20 percent, 30 percent, 40 percent, 50 percent, and 60 percent, respectively. For each training set percentage, 1000 trials were performed to reduce random effects. Figure 14a,b presents the average accuracy bar graph for different percentages of verification and testing datasets. In the figures, the error standard deviations are also indicated. To be specific, 22 top-ranked features were chosen for training and testing by random feature selection, which is shown in Figure 14a. Meanwhile, 17 top-ranked features were chosen for training and testing by the ReliefF algorithm, as shown in Figure 14b.
When the training percentage increased to 60%, the training and testing accuracies were used as follows: precision for correctly recognizing the PQD event, defined as 5.4. Performance with Real Experiments

Based on this comparison, it was found that the accuracy of the ReliefF algorithm was always better than random selection. For instance, the verification accuracy error by ReliefF is 0.83%, 0.72%, 0.70%, 0.65%, 0.59%, and 0.48% when the training percentage is 10%, 20%, 30%, 40%, 50%, and 60%, respectively. However, for random feature selection in Figure 14b, the error between the verification and test accuracy was not reduced with an increasing training percentage. The average accuracy difference between verification and test was 4.1%, 2.7%, 2.3%, 1.8%, 1.7%, and 1.9%. Comparing the average accuracy difference between verification and test, it was found that the accuracy of the ReliefF algorithm was always better than random selection, even on the same training percentage. As referred to above, for each training set percentage, 1000 trials were performed to reduce random effects. Therefore, each training result varies, and the error results are marked in a black line on the bar. For a specific training percentage, the error bar shows that the accuracy error by ReliefF is much smaller than random selection. For instance, the verification accuracy error by ReliefF is 0.83%, 0.72%, 0.70%, 0.65%, 0.59%, and 0.48% when the training percentage is 10%, 20%, 30%, 40%, 50%, and 60%, respectively. However, the verification accuracy error by random selection is 0.93%, 0.90%, 0.91%, 0.92%, 0.90%, and 0.94% when the training percentage is 10%, 20%, 30%, 40%, 50%, and 60%. Obviously, the error does not decrease, which also shows that the randomly selected classifier is less robust than ReliefF. Based on this comparison, it was verified the constructed model by random selection is always underfitting. In other words, the randomly selected features are not optimal for the classifier model. Therefore, it is essential to perform the feature selection procedure before constructing the classifier model.

5.4. Performance with Real Experiments

To present the results of the proposed algorithm, the following indicators [30] were used as follows: precision for correctly recognizing the PQD event, defined as \( P_{\text{event}} \); recall for correctly recognizing the PQD event, \( R_{\text{event}} \); precision for incorrectly recognizing the

![Graph](image-url)

**Figure 14.** Accuracy comparison with different training percentages. (a) Accuracy with feature selection; (b) Accuracy without feature selection.

Overall, it was observed that both training and testing accuracies with the optimal selection perform better than the randomly selected results. In Figure 14a,b, the training accuracy was greater than the testing accuracy when the training percentage was only 10%. This indicates that the classifier is underfitting. Then, the accuracy for test sets rises with the increasing percentage for training sets. It is understandable that the classifier is well trained, with increasing training samples for both optimal selection and random selection. When the training percentage increased to 60%, the training and testing accuracies were closed by the ReliefF selection algorithm in Figure 14a. Moreover, the accuracy error between verification and test results decreased gradually for ReliefF selection. To be specific, when the training percentage was 10%, 20%, 30%, 40%, 50%, and 60%, the average accuracy difference between verification and test was 1.8%, 1.4%, 1.6%, 0.8%, 0.6%, and 0.3%, respectively. However, for random feature selection in Figure 14b, the error between the verification and test accuracy was not reduced with an increasing training percentage. The average accuracy difference between verification and test was 4.1%, 2.7%, 2.3%, 1.8%, 1.7%, and 1.9%. Comparing the average accuracy difference between verification and test, it was found that the accuracy of the ReliefF algorithm was always better than random selection, even on the same training percentage. As referred to above, for each training set percentage, 1000 trials were performed to reduce random effects. Therefore, each training result varies, and the error results are marked in a black line on the bar. For a specific training percentage, the error bar shows that the accuracy error by ReliefF is much smaller than random selection. For instance, the verification accuracy error by ReliefF is 0.83%, 0.72%, 0.70%, 0.65%, 0.59%, and 0.48% when the training percentage is 10%, 20%, 30%, 40%, 50%, and 60%, respectively. However, the verification accuracy error by random selection is 0.93%, 0.90%, 0.91%, 0.92%, 0.90%, and 0.94% when the training percentage is 10%, 20%, 30%, 40%, 50%, and 60%. Obviously, the error does not decrease, which also shows that the randomly selected classifier is less robust than ReliefF. Based on this comparison, it was verified the constructed model by random selection is always underfitting. In other words, the randomly selected features are not optimal for the classifier model. Therefore, it is essential to perform the feature selection procedure before constructing the classifier model.

5.4. Performance with Real Experiments

To present the results of the proposed algorithm, the following indicators [30] were used as follows: precision for correctly recognizing the PQD event, defined as \( P_{\text{event}} \); recall for correctly recognizing the PQD event, \( R_{\text{event}} \); precision for incorrectly recognizing the
PQD event, defined as $P_{\text{others}}$, and recall for incorrectly recognizing the PQD event, defined as $R_{\text{others}}$.

\[
P_{\text{event}} = \frac{TP}{TP + FP} \\
R_{\text{event}} = \frac{TP}{TP + FN} \\
P_{\text{others}} = \frac{TN}{TN + FN} \\
R_{\text{others}} = \frac{FP}{FP + TN} \\
\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% 
\]  

(19)

In Equation (20), the parameter TP denotes the example number for true positive, which represents that the predicted PQD event was truly recognized. The parameter FP stands for the example number of false positives, which signifies that another PQD was mistakenly identified as the anticipated PQD event. The number of false-negative instances represents the number of times the estimated PQD event was mistakenly identified as another PQD event. TN stands for true negative occurrences, which indicates that another PQD type was correctly identified as the associated PQD event [31]. The confusion matrix is imported to exhibit the PQD classification results, and the validation datasets are shown in Table 3.

The confusion matrix verifies that the classification results achieved high accuracy. Unfortunately, some samples were assigned to incorrect classes. For instance, one C3 (Swell) case was incorrectly recognized as C1 (pure sine). This was because a swelling event occurred less than 3 T. The amplitude parameter $\alpha$ was also close to 0.1. Additionally, three C4 (interrupt) cases were recognized as C2 (sag). As seen in the PQD mathematical model, the definitions of C4 and C2 are the same. The only difference is the value selection for the parameter $\alpha$.

Table 3. Confusion matrix of the classification results.

<table>
<thead>
<tr>
<th>Assigned Class</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>C8</th>
<th>C9</th>
<th>C10</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>423</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C2</td>
<td>0</td>
<td>435</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C3</td>
<td>1</td>
<td>0</td>
<td>406</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C4</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>411</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>428</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>403</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>415</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>437</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>429</td>
<td>0</td>
</tr>
<tr>
<td>C10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>422</td>
<td></td>
</tr>
</tbody>
</table>

5.5. Performance with Noisy Data

Table 4 illustrates the proposed classifier’s performance at 15 dB, 25 dB, and 35 dB SNRs for 500 test samples. As can be observed from the table, the accuracy results verify that the proposed model exhibits excellent noise tolerance in the dataset. The number of PQDs and the corresponding event types were chosen at random from the training dataset. As can be observed from Table 4, the classification accuracy improved with the SNR’s increase. Clearly, the classifier performs more admirably in a less-noisy environment on a 35 dB SNR. Moreover, the test results reveal that the developed method was less sensitive to noise in the practical environment. For instance, the lowest accuracy was achieved at 93.8% at C7 classes under 15dB SNR. This might be credited to the EWT algorithm’s good performance, which uses empirical wavelet filters to divide signals into independent components in frequency spectrums. The signal’s effective elements with noisy disturbance are then extracted. The feature extraction approach that follows is also beneficial to EWT decomposition.
Table 4. Classification accuracy performance in a noisy environment.

<table>
<thead>
<tr>
<th>Recognized Class</th>
<th>Signal-to-Noise Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15 dB</td>
</tr>
<tr>
<td>C1</td>
<td>95.2%</td>
</tr>
<tr>
<td>C2</td>
<td>100%</td>
</tr>
<tr>
<td>C3</td>
<td>98.2%</td>
</tr>
<tr>
<td>C4</td>
<td>100%</td>
</tr>
<tr>
<td>C5</td>
<td>95.2%</td>
</tr>
<tr>
<td>C6</td>
<td>96.4%</td>
</tr>
<tr>
<td>C7</td>
<td>93.8%</td>
</tr>
<tr>
<td>C8</td>
<td>98.2%</td>
</tr>
<tr>
<td>C9</td>
<td>100%</td>
</tr>
<tr>
<td>C10</td>
<td>97.6%</td>
</tr>
<tr>
<td>Overall</td>
<td>97.41%</td>
</tr>
</tbody>
</table>

6. Conclusions

This paper presents a practical method for PQD event recognition. By using empirical wavelet transform and optimized parameters, the obtained signal is transformed into several intrinsic mode function components. After, the statistical features are extracted from the decomposed subcomponents. Then, the feature selection algorithm, ReliefF, is adopted to reconstruct the optimal feature vectors. Finally, the improved feature vectors are loaded into a multiclass SVM model to identify various PQD event instances. To test the effectiveness of the proposed algorithm, a simulation dataset, as well as a real dataset on the microgrid platform, was used for several experimental investigations. By using the ReliefF algorithm, 17 top-ranked features were valid for reconstruction. The final classification accuracy achieved 97.41%, 98.86%, and 99.54% when the SNR is 15 dB, 25 dB, and 35 dB, respectively. The main conclusions of this work are summarized as follows. Compared with the EMD, the EWT is more adaptive and robust in signal segmentation and denoising. Feature selection techniques, such as the ReliefF algorithm, are effective and essential in filtering the significant components and eliminating redundant features. After selection, both the accuracy and the computing efficiency improved when the prominent features were identified. Based on the performance analysis, the proposed method demonstrates significant adaptability in various noisy situations, whether for PQD detection sensitivity or specificity.

Author Contributions: Each author made significant contributions to the production of this manuscript. S.C. and Z.L. designed the experiment; F.X., G.P. and S.C. performed the experiments; Z.L. and S.C. analyzed the data; S.C. wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Basic Research Program of China (973 Program) (Project Name: Comprehensive Control and Cooperative Optimization of Multi-energy Flow in Industrial Zone Park. No. 2017YFA0700300), and the Zhejiang Province Key Research and Development Program (Project Name: The research and application of multi-core modular power management system based on IoT. Grant No. 2021C01112).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We would like to acknowledge Lei Fu for help with this study’s experiment analysis.

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

PQD  Power Quality Disturbances
SVM  Support Vector Machine
FT  Fourier Transform
STFT  Short-time Fourier transform
WPT  Wavelet Packet Transform
EMD  Empirical Mode Decomposition
IMF  Intrinsic Mode Functions
EWT  Empirical Wavelet Transform
RMS  Root-Mean-Square value
DT  Decision Tree
SVR  Support Vector Machine
ANN  Artificial Neural Network
SNR  Signal-to-Noise Ratio
OVR  One-Versus-Rest
RBF  Radial Basis Function

References


