International Journal on Electrical Engineering and Informatics - Volume 13, Number 4, December 2021

Classification of Power Quality Disturbances using Wavelet Packet Information Entropy Feature Vectors and Probabilistic Neural Network

Laxmipriya Samal, Hemanta Kumar Palo, Badrinarayan Sahu

Department of Electronics and Communication Engineering,
Siksha 'O' Anusandhan Deemed to be University,
Bhubaneswar, Odisha, India
laxmipriyasamal@soa.ac.in, hemantapalo@soa.ac.in, badrinarayansahu@soa.ac.in

Abstract: This paper illustrates the automatic recognition of power quality disturbances (PQDs) using Wavelet Packet Tuned Probabilistic Neural Network (WP-PNN) structure. Eight statistical parameters are extracted from the WP by decomposing a signal into the fourth level. The PNN is simulated with these statistical parameters for performance appraisal. The Information Gain (IG) feature selection algorithm has been applied to rank the feature subsets based on high IG entropy. 16 significant WP-IG features are extracted from 128 features computed from the WP statistical coefficients. The objective is to discard redundant data for better accuracy and lower computational complexity. A two-stage experimental analysis has been carried out to validate the WP-IGPNN structure. Initially, the PQ data set is procured by regularly varying the parameters of mathematical models and choosing the best features using IG to experience the highest accuracy. Finally, the coefficient values which are not chosen in the earlier case have been used to generate a new data set. The classification accuracy has been observed using the new data set with the same chosen feature set as used in the first stage. The noisy signals are also investigated to simulate the classifier to validate the proposed WP-IGPNN structure.

Keywords: Classification algorithms; Feature extraction; Information entropy; Power quality; Wavelet packets

1. Introduction

In recent years Power Quality (PQ) has become an important issue for both services and customers. The increasing use of equipment conscious of power system disturbances and their related economic aspects has forced the distribution services to adopt new strategies to audit PQ in electrical grids. Poor PQ causes overheating of lines, inaccurate metering, and reduced efficiency of appliances.

For timely management and improved PQ, it is crucial to locate the sources and causes of the disturbances. It requires the extraction of discriminative and reliable features from these signals for efficient characterization and classification. The community has focused on several appropriate signal processing tools to detect, localize, and model these signals. The multi-resolution analysis (MRA) of Wavelet Transform (WT) has been widely explored to assess the PQ supplied to the consumers [1-5]. Online detection of these PQDs has been attempted using WT in [2]. Although the technique is quite popular to classify PQ disturbances, its performance degrades in the presence of noise [5-10]. For improved PQ monitoring, many researchers have proposed several wavelet de-noising schemes such as the change point approach [3], spatial correlation [4], short-time correlation transforms, etc. [11]

The decomposition of the PQD signal at different levels has been attempted to train many types of NNs during the last two decades [10-19]. The WT-based NNs have been effective in automatically identifying the waveform of the disturbances [12]. The WT-statistical parameters computed using MRA have improved the PNN classification performance with faster response [13-14]. The WT-PNN structure has been shown to outperform the conventional NNs such RBFN or MLP in classifying PQ events as reported in [15-17]. The decomposition of the PQD signal up to the 4^{th} level has been carried out to extract 10-features for developing effective PNN models [16]. Up to the 13^{th} level, WT decomposition has been performed to study the energy

Received: August 22^{th}, 2021. Accepted: November 24^{th}, 2021
DOI: 10.15676/ijeei.2021.13.4.6

distribution along with the time duration of a PQ signal [17]. The authors have extracted 14 features to successfully classify 7-types of PQ events. Similarly, the combination of the WT decomposition up to the 11th level and the self-organizing learning array to enhance the PQD accuracy has been attempted in [18]. From these pieces of literature, it can be inferred that the PNN remains a near-optimal classifier. It is faster to train and insensitive to outliers. It does not require extensive retraining with the addition or removal of training samples. The network requires a single parameter adjustment and single processing during training. It is capable of tolerating erroneous samples and can be trained with sparse input data. These properties of the PNN make it a potential candidate to classify the PQDs in this work.

The authors have further observed that the WP has been popular in the diagnosis of aperiodic signals such as emotional speech [19], and also PQDs [5-6],[20-21]. This work investigates the MRA of WT by decomposing the signal up to the 4th level to extract appropriate statistical parameters from each level. Further, to improve the recognition performance an optimal feature vector has been derived from these statistical parameters by applying the Information Gain (IG). The IG ratio signifies the amount of information gained from a feature while observing another feature. It is similar to the Kullback-Leibler divergence or relative entropy [22-23]. It indicates how the probability distribution of a feature vector differs from that of another feature vector [24]. The objective is to rank the WT-statistical features based on their relevance by discarding redundant data for effective outcomes.

The rest of the article is formalized as follows. The Wavelet Packets (WP) algorithm is provided in section 2. The algorithms of statistical feature extraction and the feature ranking algorithm using IG are also covered in this section. Section 3 explains the PNN classification scheme. Section 4 analyses and discusses the simulated results, whereas section 5 concludes the work.

2. Research Methodology

The detailed descriptions of the applied research methodology have been elaborated in the following subsections.

A. Wavelet Packet Transform

The WP can be considered as the generalization of WT which provides a richer range of possibilities in analyzing a signal. It decomposes a signal $x(t)$ into high-frequency $h(k)$ and low frequency, $l(k)$ components. These are known as the approximation (low pass filter) and detailed coefficients (high-pass filter) respectively. While the $l(k)$ coefficients are estimated using the scaling function, the $h(k)$ coefficients are computed using both the scaling and wavelet functions. The scaling and wavelet functions can be represented as

$$\varphi(k) = \sqrt{2}\sum_n l(n)\,\psi(2k-n) \quad (1)$$

$$\psi(k) = \sqrt{2}\sum_n h(n)\,\psi(2k-n) \quad (2)$$

The variable n denotes the number of samples and is an integer.

The steps of WP analysis are briefed below.

- Perform level-1 decomposition of the signal to compute the approximation and detailed coefficients. The choice of wavelet family, type and order of the mother wavelet, decomposition level, etc. are generally task-dependent. Literature survey shows, the fourth-order Daubechies wavelet (dB4) provides the best performance in processing PQ events [14], hence chosen in this work.

- Decompose both the approximation and detail in level-2 to form a binary tree-like structure as given in Figure 1.

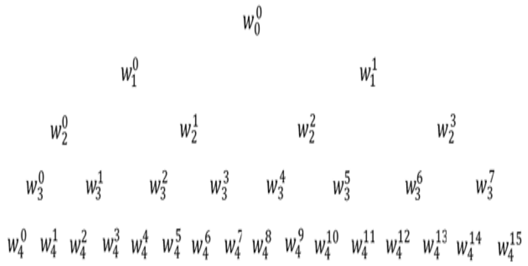


Figure 1. Tree Structure of the Wavelet Packet Decomposed up to the 4th Level

- Apply this procedure recursively to both approximation and detailed up to the k^{th} level based on the frequency resolution. For every level $i \in k$, there exists a 2^i number of nodes after the WP decomposition. In Fig.1, w_i^q indicates the q^{th} node corresponding to the i^{th} level.

There are more than 2^{k-1} alternate ways of encoding a signal in k^{th} level WP. The Shannon entropy-based coding scheme is quite popular in choosing the most effective signal decomposition [20], hence considered in this work. The WP decomposition up to level-2 is shown in Figure 2.

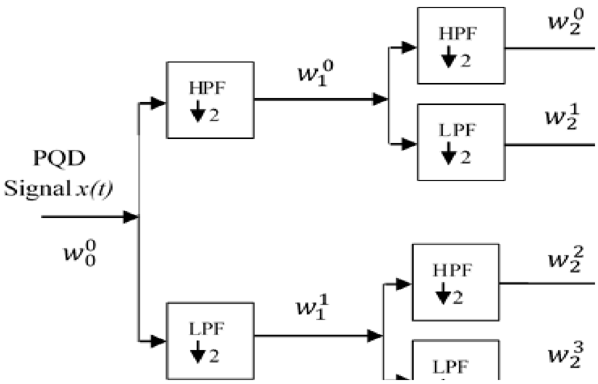


Figure 2. The Wavelet Packet Decomposition up to Level-2

B. Feature Extraction

Eight statistical parameters are estimated from the WP coefficients D_{ji} of a signal using the following equations.

$$Energy(E_j) = \sum_{i=1}^{N} |D_{ji}|, j = 1, 2, \dots, J \quad (3)$$

$$Mean\ (\mu_j) = \frac{1}{N}\sum_{j=1}^{N} D_{ji}, j = 1, 2, \dots, J \quad (4)$$

$$Standard\ deviation\ (\sigma_j) = \left(\frac{1}{N-1}\sum_{j=1}^{N}(D_{ji} - \mu_j)^2\right)^{1/2}, j = 1, 2, \dots, J \quad (5)$$

$$Skewness\ (SK_j) = \frac{E(D_{ji}-\mu_j)^3}{\sigma_j^4}, j = 1, 2, \dots, J \quad (6)$$

$$Kurtosis\ (KT_j) = \frac{E(D_{ji}-\mu_j)^4}{\sigma_j^4}, j = 1, 2, \dots, J \quad (7)$$

$$Entropy\ (EP_j) = -\sum_{j=1}^{N} D_{ji}^2 \log(D_{ji}^2), j = 1, 2, \dots, J \quad (8)$$

$$Crest\ factor\ (CF_j) = \frac{Rms\ value(D_{ji})}{Peak\ value(D_{ji})}, j = 1, 2, \dots, J \quad (9)$$

$$From\ factor\ (FF_j) = \frac{Peak\ value(D_{ji})}{RMs\ value(D_{ji})}, j = 1, 2, \dots, J \quad (10)$$

where $j = 1, 2, \dots, J$ denotes the number of nodes at the 4^{th}-level decomposition, N is the number of coefficients in a decomposition level, and $E(D_{ji} - \mu_j)^k$ is the expected value or the k^{th} moment about the mean.

C. Feature Ranking using Information Gain

There are several feature selection techniques to judge the goodness of a feature in a training vector. Among these, the Ranking methods are simple and considerable success has been reported in classifying different patterns [22-24]. In this method, the features are ranked based on their relevance. This way, it removes redundant data and retains only the selected features. It is based on the principle that while a feature may not depend on the input feature vector, it must depend on class labels. Thus, the features that do not influence the class labels are generally discarded.

The IG is an entropy-based feature evaluation method that signifies the amount of information provided by a feature item for the text category. It calculates the importance of lexical items for classifying a pattern. In this work, the IG is applied to rank the extracted WP statistical features based on their importance. It emphasizes the training vector attributes which minimize the entropy and maximize the IG ratio. Based on these attributes, the difference between the average entropy before and after splitting the training vector is computed as given below

$$IG(x, y) = Entropy(x) - Entropy(y) \quad (11)$$

Or

$$IG = Entropy(before) - \sum_{l=1}^{L} Entropy(i, after) \quad (12)$$

Or

$$IG = H(P) - \sum_{l=1}^{L} \frac{P_i}{P} H(P_l) \quad (13)$$

where L denotes the number of splitting subsets, $H(P)$ and $H(P_l)$ are the entropy of the given and the i^{th} subset using the partitioning P of the training vector. Thus, the IG is a filtering approach that aims at ranking feature subsets based on high IG entropy in descending order. It measures the degree of disorder of a system based on the entropy. In this, the higher-ranked features can prominently reflect the disturbance classes than low-ranked features. Thus, retaining only the high-ranked features can classify the PQDs better.

3. The Probabilistic Neural Network Architecture

The paper aims to classify power quality disturbances (PQDs) using Wavelet Packet Tuned Probabilistic Neural Network (WP-PNN). It investigates the Information Gain (IG) feature selection algorithm to rank the WP-statistical feature subsets based on high IG entropy. The algorithm assumes that the features depend on class labels, however, all the input vectors may not contain relevant information. Thus, the features that do not influence the class labels are generally discarded and the remained vector remains more discriminative. The low-dimensional feature vector will consume less memory, provide faster system response, and reduce the computational complexity of the classifier in real-time. Further, the chosen PNN classifier uses a Parzen window that does not demand any apriori assumptions or knowledge of any underlying distribution of a feature vector. It is a feed-forward NN that approximates the probability distribution function (PDF) of a class using non-parametric functions. It approaches Bayes' optimization while estimating the class probability by allocating the highest posterior probability to the new feature input. The reason is, the structure requires a single pass training and can approximate any PDF using the sum of multivariate Gaussian functions. The combination of lower computational complexity, ease of training, flexibility, and faster response makes the classifier superior to other state-of-the-art NNs. The absence of any local minima issues guarantees the classifier to convergence to optimal solutions.

The PNN classifier is simulated using the WP-IG ranked feature vectors. The structure has the input, pattern, summation, and output layers as shown in Figure 3.

The input PQ pattern $'s'$ of z-dimension is fed to the pattern layer of the PNN, whose neuron vector s_{ih} provides the output as

$$\vartheta_{i,h}(s) = \frac{1}{(2\pi)^{\frac{z}{2}}\rho^{z}} exp\left[-\frac{(s-s_{i,h})^{T}(s-s_{i,h})}{2\rho^{2}}\right] \tag{14}$$

where ρ denotes the smoothing parameter and $i = 1, 2, \ldots, 8$ is the number of PQD states. The summation layer computes the maximum likelihood of the chosen PQ pattern. It summarizes and estimates the averages of all the output neurons corresponding to the chosen PQ state and classifies the pattern $'s'$ as belonging to that state.

$$p_i(x) = \frac{1}{(2\pi)^{\frac{z}{2}}\rho^{z}} \frac{1}{L_i} \sum_{h=1}^{L_i} exp\left[-\frac{(s-s_{i,h})^{T}(s-s_{i,h})}{2\rho^{2}}\right] \tag{15}$$

where $p_i(x)$ and L_i denote the class conditional probability and the total amount of features in the actual PQ state O_i. The decision layer obeys the Bayes' rule and classifies the required PQ pattern on receiving the summation layer output by assuming an equal *apriori* probability of all the eight-PQ states. Assuming the loss incurred in predicting a wrong decision to be equal for all the PQD states, the estimated state $\widehat{O}(s)$ is expressed as

$$\widehat{O}(s) = \arg\max\{p_i(s)\}, i = 1, 2, \ldots ,8 \tag{16}$$

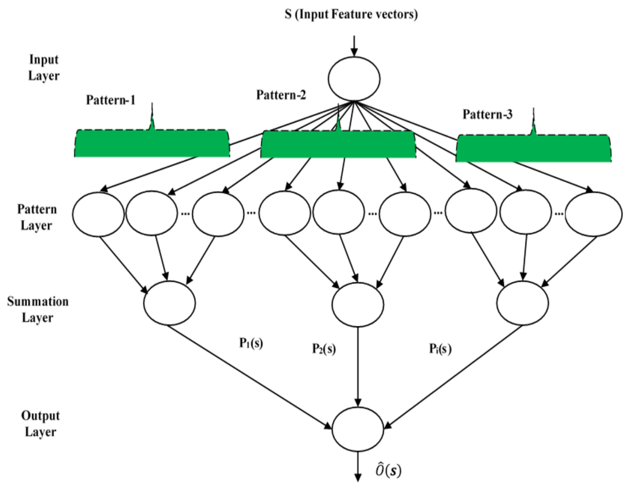


Figure 3. A Generalized PNN Architecture

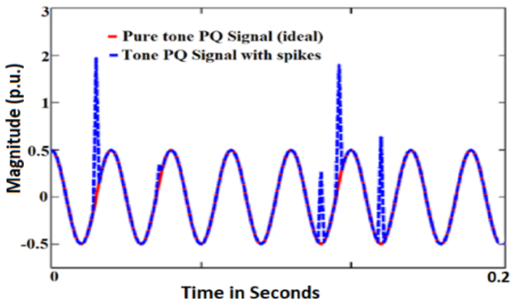
4. Results and Discussion

A. Preparation of PQD Data

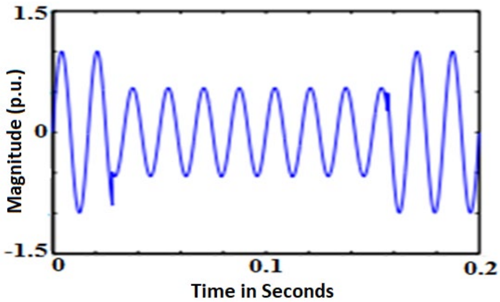
This work synthetically generates the PQ disturbance signals using mathematical models in MATLAB for the intended analysis as presented in Table 1.

Table 1. Different PQD Signal Models

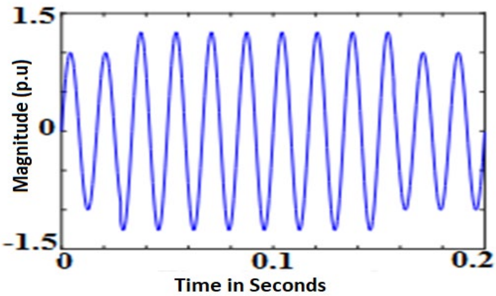
PQ disturbances	Equation	Parameters
Pure Sine signal	$s(t) = A[1 \pm \beta(u(t-t_1) - u(t - t_2))]\sin(wt)$	$T \leq t_2 - t_1 \leq 9T$ $\beta \leq 0.1$
Swell Signal	$s(t) = A[1 + \beta(u(t-t_1) - u(t - t_2))]\sin(wt)$	$0.1 \leq a \leq 0.8$, $T \leq t_2 - t_1 \leq 9T$
Sag Signal	$s(t) = A[1 - \beta(u(t-t_1) - u(t - t_2))]\sin(wt)$	$0.1 \leq \beta \leq 0.9$ $T \leq t_2 - t_1 \leq 9T$
Transient Signal	$s(t) = A[1 - \beta(u(t-t_1) - u(t - t_2))]\sin(wt)$	$\frac{T}{20} \leq t_2 - t_1 \leq \frac{T}{10}$ $0 \leq \beta \leq 0.414$
Harmonics	$s(t) = A[\beta_1 \sin(wt) + \beta_3 \sin(3wt) + \beta_5 \sin(5wt) + \beta_7 \sin(7wt)]$	$0.05 \leq \beta_1, \beta_3, \beta_5, \beta_7 \leq 0.15$ $\sum \beta_i^2 = 1$
Flicker	$s(t) = A\left[1 + \beta_f sin(\alpha_f wt)\right]\sin(wt)$	$0.1 \leq \beta_f \leq 0.2$ $5 \leq \alpha_f \leq 20Hz$
Momentary Interruptions	$s(t) = A\left[1 - \beta\left(u(t-t_1) - u(t - t_2)\right)\right]\sin(wt)$	$0.9 \leq \beta \leq 1$, $T \leq t_2 - t_1 \leq 9T$
Notch	$s(t) = \sin(wt) - sign(\sin(wt)) \times \left\{\sum_{n=0}^{9} K \times [u(t - (t_1 - 0.02n)) - u(t - (t_2 - 0.02n))]\right\}$	$0 < t_1, t_2 \leq 0.5T$ $0.01T \leq t_2 - t_1 \leq 0.05T$ $0.01 \leq K \leq 0.04$
Spike	$s(t) = \sin(wt) + sign(\sin(wt)) \times \left\{\sum_{n=0}^{9} K \times [u(t - (t_1 - 0.02n)) - (u(t - (t_2 - 0.02n))]\right\}$	$0 < t_1, t_2 \leq 0.5T$ $0.01T \leq t_2 - t_1 \leq 0.05T$ $0.01 \leq K \leq 0.04$


Pure tone PQ Signal (ideal)
Tone PQ Signal with spikes
Magnitude (p.u.)
Time in Seconds
3
2
1
0.5
0
-0.5
0
0.2

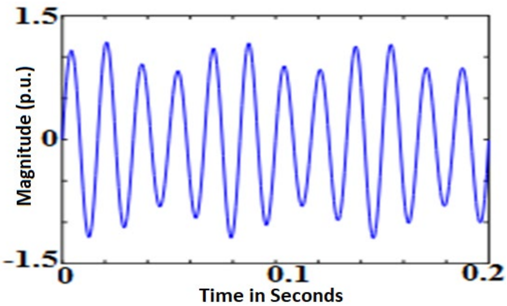
(a).


Magnitude (p.u.)
Time in Seconds
1.5
0
-1.5
0
0.1
0.2

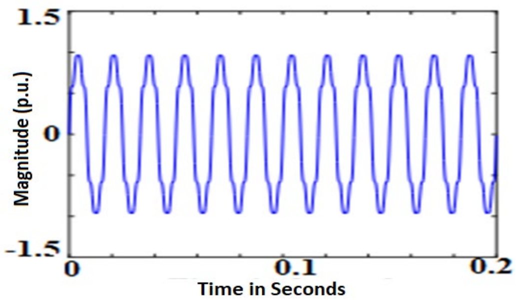
(b).


Magnitude (p.u)
Time in Seconds
1.5
0
-1.5
0
0.1
0.2

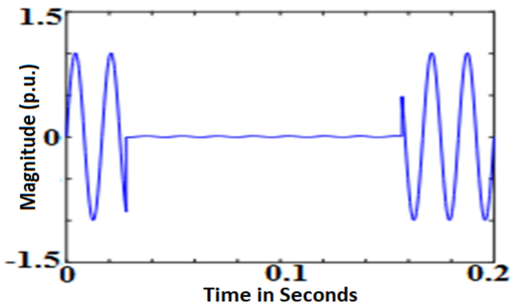
(c).


Magnitude (p.u.)
Time in Seconds
1.5
0
-1.5
0
0.1
0.2

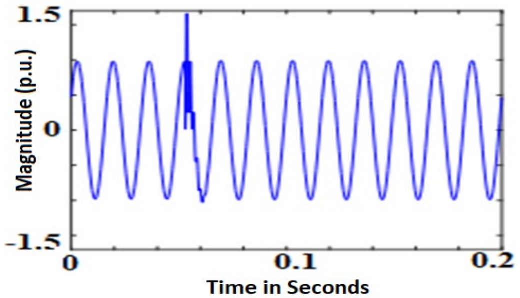
(d).



(e).



(f)



(g)

Figure 4. The Waveform of Power quality Disturbances (a) Pure Tone and Spike (b) Voltage Sag (c) Voltage Swell (d) Flicker (e) Harmonics (f) Momentary Interruption (g) Transient

The waveforms of the PQ disturbance signals are shown in Figure 4. The coefficient β denotes the swell or sag level. The disturbance duration that exists in the pure sine signal is represented using the unit step function $u(t)$. To generate the data set, both the parameters β and $u(t)$ have been suitably varied on several points of the signal wave (by varying $t_1$and t_2) and disturbance duration $(t_2 - t_1)$. In this way, signals of different magnitude are obtained. The combinations of 2nd, 3rd, 5th, and 7th order harmonics are used to generate the harmonic signals. The amplitude β_i and the frequency α_f are varied accordingly to generate the flicker signals. The amplitude of the momentary interruption having parameter β is varied to generate the signal. Similarly, other signals are generated by varying suitable parameters. A sampling frequency of 3.2 kHz with a data length of 14 cycles corresponding to each disturbance class has been chosen for this purpose. A total of 100 samples for each PQ event has been used to develop the database.

The WP is applied to each disturbance signal by decomposing it up to the 4th level. The Shannon entropy-based WP present in the MATLAB toolbox has been used for this purpose. There are a total of 16-different nodes corresponding to different sub-band frequencies (each 100 Hz) of WP and 8-statistical parameters are extracted from each sub-band.

B. Comparison of Recognition Accuracy

The WP-IG features are extracted by ranking the statistical parameters based on their relevance. The extraction of the proposed optimal feature matrix has been formalized as follows.

- Decompose each signal using 4th level WP which generates 16 nodes.
- Concatenate the WP features of all the $M = 100$ signals in a PQ event.
- Extract 8-statistical features corresponding to each WP node of a PQ event which provides a feature vector of size 8×16 as in equation (17).

$$Feature = \begin{bmatrix} E_1\, E_2 \dots E_{16} \\ \mu_1\, \mu_j \dots \mu_{16} \\ \sigma_1 \sigma_2 \dots \sigma_{16} \\ SK_1 SK_2 \dots SK_{16} \\ KT_1 KT_2 \dots KT_{16} \\ EP_1 EP_2 \dots EP_{16} \\ CF_1\, CF_2 \dots\, CF_{16} \\ FF_1 FF_2 \dots\, FF_{16} \end{bmatrix} \tag{17}$$

The features are normalized to prevent the influence of high-magnitude features over the lower ones.

- Apply the IG algorithm to each column of the feature vector of a PQ event so that it selects the highest-ranked feature of that node. This way, 16 most relevant features based on their rank are extracted from each PQ event and displayed in Table 2.

Table 2. Ranking of WP-statistical Features of a PQ Event using IG

Selected Feature	Node	Ranking	Ranking Weight
Kurtosis	0	R_1	0.504
Mean	1	R_2	0.428
Energy	2	R_3	0.213
Mean	3	R_4	0.194
Skewness	4	R_5	0.190
Energy	5	R_6	0.186
Form factor	6	R_7	0.180
Entropy	7	R_8	0.177
Kurtosis	8	R_9	0.169
Mean	9	R_{10}	0.150
Peak factor	10	R_{11}	0.133
Skewness	11	R_{12}	0.129
Skewness	12	R_{13}	0.125
Energy	13	R_{14}	0.093
Entropy	14	R_{15}	0.076
Peak factor	15	R_{16}	0.072

The aim of using IG is to maximize the 10-fold average recognition accuracy using the PNN structure. For this, the WP-IG features of all the 8-PQ events are concatenated one below the other. It provides a single 10-fold feature vector that contains the relevant features of all the chosen classes. The selection of smoothing parameter ρ in this structure is purely task-dependent and there doesn't exist any theoretical basis for this. Several values of ρ such as 0.25, 0.5, 0.75, 1.1, 1.25, and 1.5 have been trailed for a comparison of PQ accuracy. A value of $\rho = 1.1$ has provided the highest average accuracy and is hence retained for this work. To achieve the desired network convergence, the classifier has been simulated with 20, 30, 40, 50, and 100 epochs. It has been observed that the computation time and the network complexity increase with an increase in the number of epochs with a meager improvement in classification accuracy. Thus, the number of epochs has been fixed at 30 as a trade-off between the accuracy and the complexity.

The ranking of the WP-statistical features has indeed improved the recognition accuracy. The IG applies to the statistical feature vectors to rank the features based on their relevance. The features with the highest ranking by minimizing the entropy are retained for simulating the PNN classifier. The improvement in PQ accuracy has been experienced as the applied technique discards those features with no influence on the class labels although are present in the input dataset. It can be concluded that a large-feature dimension with redundant data requires a large PNN structure which increases the computational complexity. Further, the structure may be oversensitive to the input samples resulting in poor generalization. Thus, the optimal feature vector has provided an improved accuracy.

The recognition accuracy by simulating the PNN using the investigated method is shown in Table 3. This table also compares the recognition accuracy of the PQ events using only the entropy or energy features obtained from all the sub-bands (16-nodes). There is very little variation in recognition accuracy, using either the energy or entropy features as revealed from our results. However, consideration of both these features increases the feature dimension and yields better accuracy due to more available information. The overall recognition accuracy is further improved by involving all the statistical parameters for a similar reason.

Table 3. Comparison of PNN Recognition Accuracy using Different Instances

Instance	Entropy (16)	Energy (16)	Entropy (16) + Energy (16)	All WP-statistical parameters (8×16)	WP-IG ranked Features (16)
Pure Signal	97.95%	97.95%	98.14%	99.1%	99.52%
Transitional Signals	93.41%	93.59%	95.83%	94.43%	96.49%
Transitional Signals have the same magnitude but a different point on the wave	93.33%	93.47%	95.80%	93.72%	97.07%
Transitional Signals have different magnitude but the same point on a wave	94.95%	94.51%	96.18%	94.88%	98.12%

The 10-fold average recognition accuracy of the pure signal data set is presented in the first instance in Table 3. Here, the training and the testing data set involve the features extracted corresponding to all the PQD signals having different durations, magnitude, and points on the wave. In the second instance, the training and testing data involve the features extracted from the original and transitional signals respectively. However, the duration, magnitude, and point on the wave of the transitional signals used for testing are different than the training signals. Table 4 provides the individual PQ event recognition accuracy, using these transitional signals. The performance of the discussed algorithm has been further validated using case 3 and case 4 having different conditions of the transitional signals of the second instance. In the third instance, the point on wave of the transitional signal is varied while the magnitude β is fixed. The fourth instance is considered the opposite. From Table 3 and Table 4, it can be concluded that the PQ recognition accuracy has been improved using the proposed WP-IGPNN algorithm.

Table 4. Comparison of PNN Recognition Accuracy using the Transitional Signals

Disturbance Category	Entropy (16)	Energy (16)	Entropy (16) + Energy (16)	All WP-statistical parameters (8 × 16)	WP-IG ranked Features (16)
Swell	96.6%	96.0%	97.7%	97.7%	98.9%
Sag	95.3%	93.7%	95.3%	94.3%	96.5%
Transient	94.9%	95.6%	97.1%	95.8%	98.3%
Harmonic	95.1%	94.9%	96.6%	96.1%	97.8%
Flicker	92.7%	92.3%	94.4%	93.7%	95.6%
Momentary Interruption	93.4%	93.1%	94.6%	94.5%	95.8%
Notch	91.3%	91.5%	93.5%	92.2%	94.7%
Spike	90.2%	91.3%	93.0%	91.2%	94.2%
Average	93.68%	93.55%	95.27%	94.43%	96.81%

A comparison of the classification time has been studied using the WP-statistical and WP-Statistical-IG feature vectors respectively. A 7th Generation Intel Core i3 processor with 4 GB DDR4 RAM, 1TB HDD, laptop, and Windows 8 operating system under MATLAB R2016a, win64 platform has been used for this purpose. The involvement of the IG feature ranking algorithm arguably increases the complexity of forming the feature vector. However, it reduces the classification time. The time for classification has been 0.6712s and 0.0845s respectively, with WP and WP-IG feature sets when 1-fold test data (10 for testing and 100 for training) have been used. The PNN structure is inherently parallel and requires a single parameter adjustment. The addition or deletion of features does not require further retraining and the classifier can converge to an optimal solution without overfitting. These factors make the training instantaneous, easy, and faster. Further, it does not update the weights unlike RBFN and the possibility to set all the weights to one makes the PNN suitable for reduced feature vectors. Finally, the application of Kernel Fisher discriminant statistical analysis in PNN reduces the misclassification errors. The WP-IG vector is low-dimensional (16) as compared to the WP feature vector (8×16), hence the latter has shown a faster PNN response.

C. Comparison of Recognition Accuracy using Noisy Data

In general, practical data is always associated with noise in any power distribution system. Thus, the authors intend to evaluate the proposed method under noisy conditions. The Gaussian white noise has been widely popular in analyzing PQ disturbance signals. The features are extracted from the noisy data having SNR (signal to noise ratio) of 25, 35, and 40-dB noise levels and tested with 20, 25, 30, and 40-dB. A comparison of the noisy PQ signals to classify the PQD events has been made in Table 5.

Table 5. Comparison of PNN Recognition Accuracy (%) with Noisy Signals

Disturbance Category	Noise				
	20 dB	25 dB	30 dB	35 dB	40 dB
Swell	78.8200	88.1404	97.5806	100.0000	100.0000
Sag	67.1225	85.2089	96.0130	96.0428	100.0000
Transient	94.5912	98.6397	93.5633	96.9607	99.9643
Harmonic	89.0472	85.0000	83.3519	90.0338	98.8580
Flicker	98.5700	98.3438	100.0000	100.0000	100.0000
Momentary Interruption	87.3584	88.6307	99.7300	92.9407	89.7884
Notch	99.82	99.3551	89.2526	87.6721	90.2241
Spike	99.82	99.1799	90.5413	88.2285	92.2566
Average	89.39	92.81	93.75	93.98	96.38

Table 6. Comparison of the Proposed Technique with that of the Existing Techniques

	Feature	Feature Selection	Classifier	Recognition Accuracy	Number of Events
Khokhar et al. 2017 [14]	WT	ABC	PNN	98.62	8
Khokhar et al. 2018 [16]	WT	Norm Entropy	PNN	96.42%	8
Zhang et. Al 2012 [26]	WP	Nil	SVM	97.7%	8
	WP		QNN	96%	10
Panigrahi et al. 2009 [20]	WP	GA-FKNN	SVM	96.25%	8
Hu et al. 2008 [25]	WP	Energy Entropy	Weighted SVM	98.4%	5
Manimala et. Al 2012 [27]	WP	GA	SVM PNN	98.33% 91.25%	8
Proposed Method	WP	Statistical-IG	PNN	99.52%	9

Further, the authors also consider generating a transitional data set to validate the proposed algorithm. To procure the transitional data the parameter values of the mathematical models that are not considered while generating the first data set have been used.

A comparison of the earlier work attempted by different researchers in this field with the proposed technique has been provided in Table 6.

5. Conclusions

This work illustrates the applicability of the MRA-WP to extract appropriate statistical parameters from a few synthetically generated PQ disturbance signals. The IG has been applied to rank the features based on their relevance and remove the redundant data. The proposed WP-IG method has provided a low-dimension feature vector containing an optimal feature set. The PNN classifier has been employed to classify the PQDs automatically using the derived feature sets. It has shown better results for low-dimensional WP-IG-ranked feature vectors with lower classification time and higher recognition accuracy. The optimal feature set has been validated for its efficacy in describing the PQDs at different instances of time. It has also been verified in presence of both parametric, and noisy PQ data.

6. References

[1]. S. P. Singh and A. H. Bhat, "Adaptive LMS theory-based energy minimized dynamic voltage restorer for mitigating various power quality problems in a distribution system," *International Journal on Electrical Engineering and Informatics*, vol. 13, no. 2, Jun. 2021.

[2]. K. Thirumala, M. S. Prasad, T. Jain, and A. C. Umarikar, "Tunable-Q wavelet transform and dual multiclass SVM for online automatic detection of power quality disturbances," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3018-3028, Nov. 2016.

[3]. U. D. Dwivedi, and S. N. Singh, "Denoising Techniques with change-point approach for wavelet-based power-quality monitoring," *IEEE Trans Power Deliv*. vol. 24, no. 3, pp. 1719-1727, Jun. 2009.

[4]. P. A. Karafotis, and P. S. Georgilakis, "Power quality monitoring and evaluation in power systems under non-stationary conditions using wavelet packet transform," High Voltage, vol. 4, no. 3, pp. 186-196, Sep. 2019.

[5]. S. Choe, and J. Yoo, "Wavelet packet transform modulus-based feature detection of stochastic power quality disturbance signals," *Appl. Sci*. vol. 11, no. 6, 2825, Jan. 2021.

[6]. G. Chandrashekar, and F. Sahin, "A survey on feature selection methods," *Comput. Electr. Eng*. vol. 1, no. 40, pp. 16-28, Jan. 2014.

[7]. K. Thirumala, A. Kanjolia, T. Jain, and A. C. Umarikar, “Empirical wavelet transform and dual feed-forward neural network for classification of power quality disturbances,” *Int. J. Power Energy Convers*. vol. 11, no. 1, pp. 1-21, 2020.
[8]. M. Ijaz, M. Shafiullah, and M. A. Abido, “Classification of power quality disturbances using Wavelet Transform and Optimized ANN,” 18th International Conference on Intelligent System Application to Power Systems (ISAP), *IEEE, Porto, Portugal*, Sep. 2015, pp. 1-6, Doi: 10.1109/ISAP .2015.7325522.
[9]. M. Aggarwal, and K. Saini, “Designed orthogonal wavelet-based feature extraction and classification of underlying causes of power quality disturbance using probabilistic neural network,”. *Aust. J. Electr. Electron. Eng*. pp. 1-11, Jul. 2021.
[10]. H. Zang, P. Liu, and O. P. Malik, “Detection and classification of power quality disturbances in noisy conditions,” *IEEE Proc. Gener. Transm. Distrib*. vol. 150, no. 5, pp. 567–572, Sep. 2003.
[11]. S. Santoso, E. J. Powers, W. Grady, and A. Parsons, “Power quality disturbance waveform recognition using Wavelet based neural classifier. I.,” *IEEE Trans. Power Deliv*. Vol. 15, no. 1, pp. 222-228, Jan. 2000.
[12]. S. Santoso, E. J. Powers, W. Grady, and A. Parsons, Power quality disturbance waveform recognition using Wavelet-based neural classifier. II., application,” *IEEE Trans. Power Deliv*., vol.15, no.1, pp. 229-235, Jan. 2000.
[13]. L. Samal, H. K. Palo, B.N. Sahu, and D. Samal, “Characterization of power quality disturbances and their efficient classification,” *Advances in Electrical Control and Signal Systems (AECS) Springer*, Singapore, Jul. 2020, pp. 969-981, Doi:10.1007/978-981-15-5262-5.
[14]. L. Samal, H. K. Palo, and B. N. Sahu, “Comparison of classifiers for power quality disturbances with wavelet statistical analysis,” *International Conference on Computational Intelligence for Smart Power System and Sustainable Energy (CISPSSE)*, Keonjhar, India, 29-31 Jul. 2020, pp. 1-5, Doi: 10.1109/CISPSSE49931.2020.9212300.
[15]. S. Khokhar, A. A. Zin, A. P. Memon, and A. S. Mokhtar, “A new optimal feature selection algorithm for classification of power quality disturbances using discrete wavelet transform and probabilistic neural network,” Measurement, vol. 95, pp. 246-259, Jan. 2017.
[16]. S. Khokhar, A. M. Zina, and A. Mokhtar, “Automatic classification of hybrid power quality disturbances using wavelet norm entropy and neural network, WEAS Trans. Power Syst. vol. 13, pp. 163-173, 2018.
[17]. Z. L. Gaing, “Wavelet-based neural network for power disturbance recognition and classification,” *IEEE Trans. Power Deliv*. vol. 19, no. 4, pp. 1560–1568, Oct. 2004.
[18]. H. He, and J. A. Starzykn, “A self-organizing learning array system for power quality classification based on wavelet transform,” *IEEE Trans. Power Deliv*, vol. 21, no.1, pp. 286-295, Dec. 2005.
[19]. H. K. Palo, and M. N. Mohanty, “Wavelet-based feature combination for recognition of emotions,” *Ain Shams Eng. J*. vol. 9, no. 4, pp. 1799-1806, Dec. 2018.
[20]. B. K. Panigrahi, and V. R. Pandi, “Optimal feature selection for classification of power quality disturbances using Wavelet packet-based fuzzy k-nearest neighbour algorithm,” *IET Gener. Transm*. vol. 3, no. 3, pp. 296- 306, Mar. 2009.
[21]. Rodríguez, et al., “A Decision Tree and S-transform based approach for power quality disturbances classification,” *4th IEEE International Conference on Power Engineering Energy and Electrical Drives*, Istanbul, Turkey, 13-17 May 2013, Doi: 10.1109/PowerEng.2013.6635763.
[22]. N. Huang, et al., “Feature selection of power quality disturbance Signals with an entropy-importance-based random forest,” Entropy, vol. 18, no. 2, Feb. 2016.
[23]. Azhagusundari, and A. S. Thanamani, “Feature selection based on information gain. *Int. J. Innovative Techno. Explor. Eng*., vol. 2, no. 2, pp. 18-21, Jan. 2013.

[24]. T. A. Alhaj, M. M. Siraj, A. Zainal, H. T. Elshoush, and F. Elhaj, Feature selection using information gain for improved structural-based alert correlation," Plos One, vol. 11, no. 11, Nov. 2016.
[25]. G. S. Hu, F.F. Zhu, and Z. Ren, "Power quality disturbance identification using wavelet packet energy entropy and weighted support vector machines," Expert Syst. Appl. vol. 35, Iss. 1-2, pp. 143-149, Jul. 2008.
[26]. M. Zhang, K. Li. and Y. Hu, "Classification of power quality disturbances using wavelet packet energy and multiclass support vector machine," *Int. J. Comput. Math. Electr. Electron. Eng*. Jul. 2012.
[27]. K. Manimala, K. Selvi, and R. Ahila, "Optimization techniques for improving power quality data mining using wavelet packet-based support vector machine," Neurocomputing, vol. 77, no. 1, pp. 36-47, Feb. 2012.

Laxmipriya Samal has received the B.E. and M. Tech. degree in Electronics and Communication Engineering from the BPUT, Odisha and the Siksha 'O' Anusandhan Deemed to be University, Odisha in 2003 and 2009 respectively. She is currently working towards a Ph.D. degree at the Siksha 'O' Anusandhan Deemed to be University, Odisha, India. Her research interests include Power quality disturbance and applied signal processing.

Hemanta Kumar Palo has completed his Master of Engineering from "Birla Institute of Technology", Mesra, Ranchi in 2011 and Ph.D. in 2018 from the Siksha 'O' Anusandhan Deemed to be University. He is having 20 years of experience in the field of Electronics and Communication Engineering. Currently, he is working as an Assoc. Prof. in the Dept. of ECE, SOA University, Bhubaneswar, Odisha, India. He has published around 70 papers in international journals and conferences. His research interests include signal processing, speech processing, and power electronics

Badrinarayan Sahu has completed M-Tech in the year 2004 from UCE Burla, Sambalpur, and Ph.D. from Siksha 'O' Anusandhan University. Currently, he is serving as a Professor in the Department of Electronics and Communication Engineering at Siksha 'O' Anusandhan University, Bhubaneswar, Odisha, India. His areas of interest are signal processing, image processing, and power quality recognition