A Self-Organizing Learning Array System for Power Quality Classification Based on Wavelet Transform

Haibo He, *Student Member, IEEE,* and Janusz A. Starzyk, *Senior Member, IEEE*

***Abstract*—This paper proposed a novel approach for the Power Quality (PQ) disturbances classification based on the wavelet transform and self organizing learning array (SOLAR) system. Wavelet transform is utilized to extract feature vectors for various PQ disturbances based on the multiresolution analysis (MRA). These feature vectors then are applied to a SOLAR system for training and testing. SOLAR has three advantageous over a typical neural network: data driven learning, local interconnections and entropy based self-organization. Several typical PQ disturbances are taken into consideration in this paper. Comparison research between the proposed method, the support vector machine (SVM) method and existing literature reports show that the proposed method can provide accurate classification results. By the hypothesis test of the averages, it is shown that there is no statistically significant difference in performance of the proposed method for PQ classification when different wavelets are chosen. This means one can choose the wavelet with short wavelet filter length to achieve good classification results as well as small computational cost. Gaussian white noise is considered and the Monte Carlo method is used to simulate the performance of the proposed method in different noise conditions.**

***Index Terms*—Noise, power quality (PQ), self-organizing learning array (SOLAR), support vector machine (SVM), wavelet transform.**

I. Introduction

POWER QUALITY (PQ) is becoming prevalent and of critical importance for power industry recently. The fast expansion in use of power electronics devices led to a wide diffusion of nonlinear, time-variant loads in the power distribution network, which cause massive serious power quality problems. At the same time, the wide use of accurate electronic devices require extremely high quality power supplies. According to the data provided by Electrical Power Research Institute (EPRI), the US economy is losing between $104 billion and $164 billion a year to outages, and another $15 billion to $24 billion to PQ phenomena [1]. Therefore, the research of power quality issues has captured exponentially increasing attention in the power engineering community in the past decade.

This paper is focused on the PQ disturbances classification problem. Artificial intelligence (AI) and machine learning is one of the most powerful tools to deal with this issue. Ibrahim *et al.* provided an excellent survey of the advanced AI techniques for PQ application [2]. The most interesting AI tools for PQ problems include Expert Systems (ES) [3], Fuzzy Logic (FL) [4]–[7], Artificial Neural Networks (ANNs) [5], [8]–[10] and Genetic Algorithms (GA') [11].

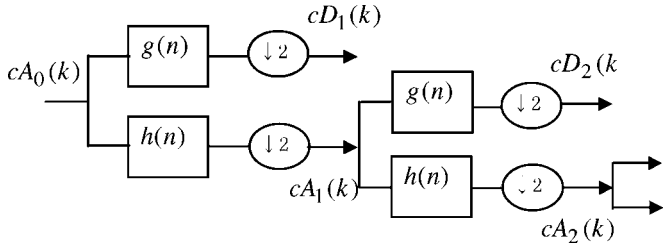

Fig. 1. Two-stage wavelet analysis filter bank.

Recent advances in wavelet transforms provide another powerful tool for PQ classification. Unlike the Short-Time Fourier Transform (STFT) with a fixed window function, the wavelet transform involves a varied time-frequency window, which yields nice performance in PQ classification. A two-stage system that employs the wavelet transform and the adaptive neuro-fuzzy networks for power quality identification is proposed in [5]. Gaouda *et al.* proposed an effective wavelet multiresolution signal decomposition method for analyzing the power quality transient events based on the standard deviation [12] and root mean square value [13]. Several typical power quality disturbances are correctly localized and classified in these papers. Wavelet based online disturbance detection for power quality applications are discussed in papers [14] and [15]. It is shown that these methods have the advantages of speed and precision discrimination in the type of transient event over conventional approaches. Huang *et al.* proposed an arithmetic coding approach based on wavelet packet transform to compress the power quality disturbance data in paper [16]. Since noise is omnipresent in a real electrical power distribution network, Yang and Liao presented a de-noising scheme for enhancing wavelet-based power quality monitoring system [17]. In this scheme, Gaussian white noise is considered and a threshold to eliminate the noise influence is determined adaptively according to the background noise.

Although lots of research achievements have been reported, the objective of classifying different kinds of power quality disturbances is still both difficult and challenging. Motivated by the recent research in the area of intelligent system and wavelet analysis, this paper aims to propose an effective classification method for power quality disturbances based on the wavelet transform and the self organizing learning array (SOLAR) system. Although wavelet transforms show nice performance in PQ analysis, direct use of it involves the problem of too large

Manuscript received May 24, 2004; revised December 16, 2004. This work was supported in part by School of Electrical Engineering and Computer Science, Ohio University, Athens, OH. Paper no. TPWRD-00314-2004.

The authors are with the School of Electrical Engineering and Computer Science, Ohio University, Athens, OH 45701, USA (e-mail: haibohe@bobcat.ent.ohiou.edu; starzyk@bobcat.ent.ohiou.edu).

Digital Object Identifier 10.1109/TPWRD.2005.852392

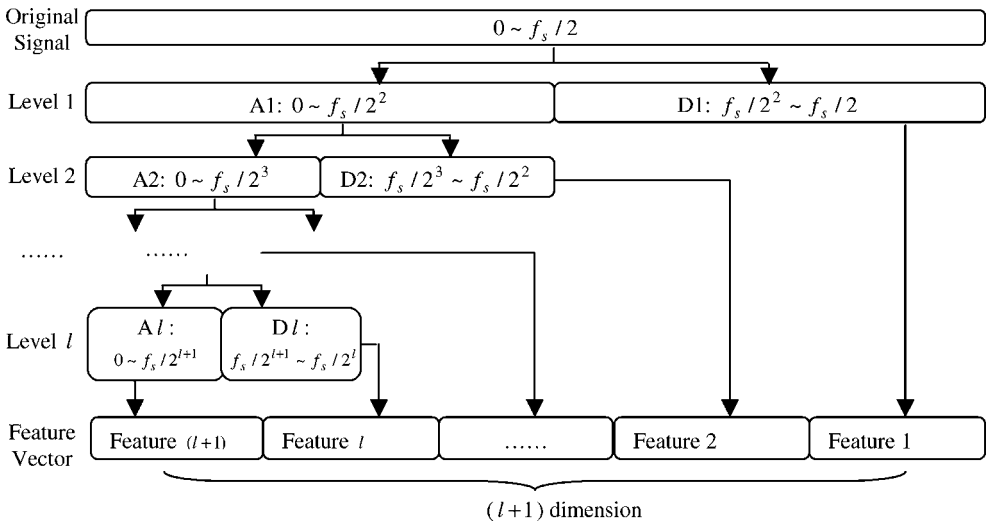

Fig. 2. MRA analysis and feature construction.

data set size [4]. In the proposed method, after using multiresolution analysis on the original sampled power waveform, the energy of the detail and approximation information at each decomposition level is calculated to construct the feature vector for future training and testing. In this way, the feature size of each sampled waveform is dramatically reduced and is only decided by the number of decomposition levels. SOLAR is a sparsely connected, multi layer and information theory based learning system. Based on the entropy estimation, neuron parameters and connections that correspond to minimum entropy are adaptively set for each neuron. By choosing connections for each neuron, the system sets up its wiring and completes its self-organization. It is shown in this paper that the combination of the wavelet transform and the SOLAR system can achieve a good PQ classification performance.

The rest of this paper is organized as follows. In Section II, a wavelet based feature extraction scheme is proposed. For an l level wavelet multiresolution analysis (MRA), an $(l+1)$ dimensional feature vector is constructed. This not only dramatically reduces the feature size, but also keeps the necessary PQ characteristics for future classification. Section III discusses the self organizing learning array system for classification. Section IV presents simulation results based on the method proposed in this paper. Comparison research between the proposed method, the support vector machine (SVM), and the latest literature reports are discussed in detail in this section. Further research about the relationships of the classification performance with the wavelet decomposition levels is presented in this part. By using the hypothesis test of the average values, we showed that there is no statistically significant difference in performance of the proposed method for PQ classification when different wavelets are chosen. This means we can choose the wavelet with short wavelet filter length to achieve nice classification results as well as small computational cost. In Section V, noise is taken into consideration and the Monte Carlo method is used to show the effectiveness of the proposed method in a noisy environments. Finally, a conclusion is given in Section VI.

II. Wavelet Based Feature Extraction

The mathematics of the wavelet transform were extensively studied and can be referred in papers [18] and [19]. The multiresolution analysis was introduced by Mallat and a detailed study about MRA can be found in [20].

Basically, a wavelet is a function $\psi \in L^2(R)$ with a zero average

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0. \tag{1}$$

The continuous wavelet transform (CWT) of a signal $x(t)$ is then defined as

$$\mathrm{CWT}_\psi x(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} x(t)\psi^* \left(\frac{t-b}{a}\right) dt \tag{2}$$

where $\psi(t)$ is called the mother wavelet, the asterisk denotes complex conjugate, a and b $(a, b \in R)$ are scaling (dilation) and translation parameters, respectively. The scale parameter a will decide the oscillatory frequency and the length of the wavelet, the translation parameter b will decide its shifting position.

In a practical application, we will use the discrete wavelet transform (DWT) instead of the CWT. This is implemented by using discrete values of the scaling parameter a and translation parameter b. To do so, set $a = a_0^m$ and $b = nb_0a_0^m$, then we get

$$\psi_{m,n}(t) = a_0^{-m/2}\psi\left(a_0^{-m}t - nb_0\right) \tag{3}$$

where $m, n \in Z$, and m indicating frequency localization and n indicating time localization. Generally, we can choose $a_0 = 2$ and $b_0 = 1$. This choice will provide a dyadic-orthonormal wavelet transform and provide the basis for multiresolution analysis.

In MRA, any time series $x(t)$ can be completely decomposed in terms of the approximations, provided by scaling functions $\phi_m(t)$ and the details, provided by the wavelets $\psi_m(t)$, where $\phi_{m,n}(t)$ and $\psi_{m,n}(t)$ are defined as the following:

$$\begin{aligned} \phi_{m,n}(t) &= 2^{-m/2}\phi(2^{-m}t - n) \\ \psi_{m,n}(t) &= 2^{-m/2}\psi(2^{-m}t - n). \end{aligned} \tag{4}$$

The scaling function is associated with the low-pass filters with filter coefficients $\{h(n), n \in Z\}$, and the wavelet function is associated with the high-pass filters with filter coefficients $\{g(n), n \in Z\}$. The so-called Two Scale Equations (TSE) give rise to these filters.

$$\phi(t) = \sum_n h(n)\sqrt{2}\phi(2t-n)$$
$$\psi(t) = \sum_n g(n)\sqrt{2}\phi(2t-n). \quad (5)$$

There are some important properties of these filters

- $$\sum_n h(n)^2 = 1 \quad \text{and} \quad \sum_n g(n)^2 = 1 \quad (6)$$
- $$\sum_n h(n) = \sqrt{2} \quad \text{and} \quad \sum_n g(n) = 0. \quad (7)$$
- Filter $g(n)$ is an alternating flip of the filter $h(n)$, which means there is an odd integer N such that
$$g(n) = (-1)^n h(N-n). \quad (8)$$

Daubechie's [18] gives a detailed discussion about the characteristics of these filters and how to construct them. The decomposition procedure starts with passing a signal through these filters. The approximations are the low-frequency components of the time series and the details are the high-frequency components.

Multiresolution analysis leads to a hierarchical and fast scheme. This can be implemented by a set of successive filter banks as shown in Fig. 1, where $h(n)$ and $g(n)$ are the low-pass and high-pass filters as defined in (5)–(8). $\downarrow 2$ means the down sampling with a factor of 2, k is the coefficient index at each decomposition level.

Considering the filter bank implementation in Fig. 1, the relationship of the approximation coefficients and detail coefficients between two adjacent levels are given as

$$cA_j(k) = \sum_n h(2k-n)cA_{j-1}(n) \quad (9)$$
$$cD_j(k) = \sum_n g(2k-n)cA_{j-1}(n) \quad (10)$$

where cA_j and cD_j represent the approximation coefficients and detail coefficients of the signal at level j, respectively.

In this way, the decomposition coefficients of MRA analysis can be expressed as

$$\begin{aligned}[A_0] &\leftrightarrow [cA_1, cD_1] \\ &\leftrightarrow [cA_2, cD_2, cD_1] \\ &\leftrightarrow [cA_3, cD_3, cD_2, cD_1] \\ &\leftrightarrow \cdots \end{aligned} \quad (11)$$

which correspond to the decomposition of signal $x(t)$ as

$$\begin{aligned} x(t) &= A_1(t) + D_1(t) \\ &= A_2(t) + D_2(t) + D_1(t) \\ &= A_3(t) + D_3(t) + D_2(t) + D_1(t) \\ &= \cdots \end{aligned} \quad (12)$$

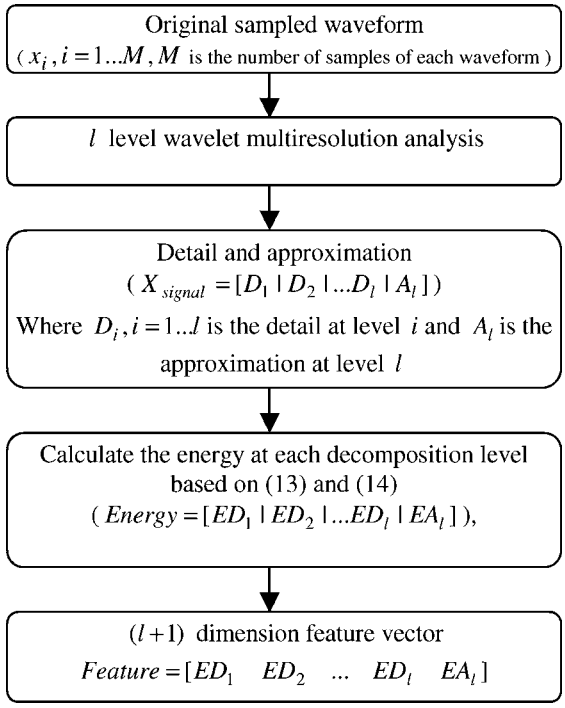

Fig. 3. Wavelet based feature extraction.

where $A_i(t)$ is called the approximation at level i, and $D_i(t)$ is called the detail at level i. Since both the high pass filter and the low pass filter are half band, the MRA decomposition in frequency domain for a signal sampled with the sample frequency f_s can be demonstrated in Fig. 2.

In order to reduce the feature dimension, we will not directly use the detail $(D_i(t))$ and approximate $(A_i(t))$ information for future training and testing. Instead, we propose to use the energy at each decomposition level as a new input variable for SOLAR classification. The energy at each decomposition level is calculated using the following equations:

$$\mathrm{ED}_i = \sum_{j=1}^{N} |D_{ij}|^2, i = 1 \ldots l \quad (13)$$
$$\mathrm{EA}_l = \sum_{j=1}^{N} |A_{lj}|^2 \quad (14)$$

where $i = 1 \ldots l$ is the wavelet decomposition level from level 1 to level l. N is the number of the coefficients of detail or approximate at each decomposition level. ED_i is the energy of the detail at decomposition level i and EA_l is the energy of the approximate at decomposition level l.

In this way, for a l level wavelet decomposition, we construct a $(l+1)$ dimensional feature vector for future analysis. Fig. 3 shows data flow in the proposed wavelet feature extraction.

III. SOLAR for Classification

A self-organizing learning array (SOLAR) system was proposed in [21], [22]. As a parallel learning structure, SOLAR has several advantages over a typical neural network:

1) Data driven learning. Each neuron performs its specific arithmetic or logic function concurrently.

2) Local interconnections. Every neuron of the system is prewired to its local neighbors and the final wiring configuration is decided during the learning stage. Neurons exchange information during learning and in operation, extracting features from the processed signals.
3) Entropy based self-organization. Each neuron selects a specific arithmetic (linear or nonlinear) or logic function and applies it to its input data to reach the maximal information index in its learning space based on data entropy.

The SOLAR implemented in this paper employs a feed forward (FF) structure with all neurons arranged in multiple layers. SOLAR neurons are event-driven processors responding to their selected data and control inputs. With control input high, a neuron only reacts to data from a selected part of the entire input space, which forms a local input space for the neuron. In this way, a control input plays the role of an inhibitory connection in biological neuron, preventing a controlled neuron from firing. Each neuron performs a simple operation like adding, subtracting, shifting, or a simple approximation of logarithm and exponent functions. In SOLAR, neurons self-organize by adapting to information included in their input data. The information index (INI) is maximized in each neuron based on probabilities of data from various classes and subspaces in neuron's input space. The INI is locally defined as (15), shown at the bottom of the page where $P_{\rm sc} = n_{\rm sc}/n_t$ is the probability of a class c satisfying threshold, $P_{\rm sic} = n_{\rm sic}/n_t$ is the probability of a class c not satisfying threshold. $P_s = n_s/n_t$ is the probability of an input satisfying threshold, $P_{\rm si} = 1 - P_s$, and $P_c = n_c/n_t$ is the class probability. Here $n_{\rm sc}$ is the number of input data that belong to class c and satisfy threshold while $n_{\rm sic}$ are those that do not satisfy threshold. n_t is the total number of input data, n_s is the number of input data that satisfy threshold, and n_c is the number of input data that belong to class c. The information index defined in (15) is normalized to a (0, 1) interval. The physical meaning of the information index is that of a normalized measure of completeness for the problem solution. It is an opposite of entropy that measures disorder, and it is directed toward a specific problem solution. When INI$= 0$, there was no reduction in data entropy, while INI$= 1$ indicates that data entropy in neurons' input space was reduced to 0. In this way, the value of information index measures the quality of a neuron's subspace separation. The INI is also related to the definition of information deficiency (IND) introduced in (16), which quantifies the amount of information left in the subspaces

$$\delta_s = \frac{\Delta \mathrm{E}_s}{\mathrm{E}_{\max}} = \frac{\sum_{\rm sc} P_{\rm sc}\log(P_{\rm sc}) - P_s\log(P_s)}{\sum_c P_c \log(P_c)}. \tag{16}$$

Information deficiency helps to self-organize the learning process. A subspace with zero information deficiency does not require any learning, that is to say, data is already well classified in the subspace.

Different combinations of data inputs and transformation operations result in different information index values. Each neuron learns data distribution in its input space and adjusts its connections and function to maximize the local space information index. The optimized configuration information is stored inside each neuron in order to provide the best separation of various classes in the neuron's input training data. At the first layer of neurons, it is assumed that the input information deficiency is one. As subsequent neurons extract information from the input data, there is less and less independent information left in the data. The learning array grows by adding more neurons until the information deficiency in the subsequent neurons falls below a set threshold value.

As a result of training, neurons internally save the correct recognition probabilities of all the classes. If an input data point falls in a voting neuron's input subspace, it is going to vote for this data using its estimated probabilities for that class. The voting scheme combines all the information and classifiers the input signal using a weight function designed after the Maximum Ratio Combination (MRC) technique as presented in [23]

$$B_c = 1 - \frac{1}{1 + \sqrt{\sum_{i=1}^{n}\left(\frac{1}{1/P_{cc_i} - 1 + \varepsilon}\right)^2}} \tag{17}$$

where $P_{cc_i}(i = 1 \ldots n)$ is each vote's correct classification for class c [it is $P_{\rm sc}$ or $P_{\rm sic}$ as defined in (15)], n is the number of voting neurons, and ε is a small number preventing division by zero. This weight function provides a statistically robust fusion of individual neurons vote. The classifier chooses a class c with the maximum weight B_c. A detailed discussion about the structure and operation of the SOLAR system can be found in paper [21] and [22].

IV. Simulation and Analysis

A. Data Generation

The simulation data was generated in MATLAB based on the model in paper [24]. Seven classes (C1–C7) of different PQ disturbances, named undisturbed sinusoid (normal), swell, sag, harmonics, outage, sag with harmonic and swell with harmonic, were considered. Table I gives the signal generation models and their control parameters. Two hundred cases of each class with different parameters were generated for training and another 200 cases were generated for testing. Both the training and testing signals are sampled at 256 points/cycle (the same as in [24] for results comparison) and the normal frequency is 50 Hz. Ten power frequency cycles which contain the disturbance are used for a total of 2560 points.

$$\begin{aligned}\mathrm{INI} &= 1 - \frac{\Delta \mathrm{E}_s}{\mathrm{E}_{\max}} \\ &= 1 - \frac{\left[\sum_{\rm sc} P_{\rm sc}\log(P_{\rm sc}) - P_s\log(P_s)\right] + \left[\sum_{\rm sic} P_{\rm sic}\log(P_{\rm sic}) - P_{\rm si}\log(P_{\rm si})\right]}{\sum_c P_c \log(P_c)}\end{aligned} \tag{15}$$

TABLE I
POWER QUALITY DISTURBANCE MODEL

PQ disturbances	Class Symbol	Model	Parameters
normal	C1	$x(t)=\sin(\omega t)$	
Swell	C2	$x(t)=A(1+\alpha(u(t-t_1)-u(t-t_2)))\sin(\omega t)$ $t_1<t_2, u(t)=\begin{cases}1, t\geq 0\\ 0, t<0\end{cases}$	$0.1\leq\alpha\leq 0.8$ $T\leq t_2-t_1\leq 9T$
Sag	C3	$x(t)=A(1-\alpha(u(t-t_1)-u(t-t_2)))\sin(\omega t)$	$0.1\leq\alpha\leq 0.9; T\leq t_2-t_1\leq 9T$
Harmonic	C4	$x(t)=A(\alpha_1\sin(\omega t)+\alpha_3\sin(3\omega t)+$ $\alpha_5\sin(5\omega t)+\alpha_7\sin(7\omega t))$	$0.05\leq\alpha_3\leq 0.15, 0.05\leq\alpha_5\leq 0.15$ $0.05\leq\alpha_7\leq 0.15, \sum\alpha_i^2=1$
Outage	C5	$x(t)=A(1-\alpha(u(t-t_1)-u(t-t_2)))\sin(\omega t)$	$0.9\leq\alpha\leq 1; T\leq t_2-t_1\leq 9T$
Sag with Harmonic	C6	$x(t)=A(1-\alpha(u(t-t_1)-u(t-t_2)))$ $(\alpha_1\sin(\omega t)+\alpha_3\sin(3\omega t)+\alpha_5\sin(5\omega t))$	$0.1\leq\alpha\leq 0.9, T\leq t_2-t_1\leq 9T$ $0.05\leq\alpha_3\leq 0.15, 0.05\leq\alpha_5\leq 0.15; \sum\alpha_i^2=1$
Swell with Harmonic	C7	$x(t)=A(1+\alpha(u(t-t_1)-u(t-t_2)))$ $(\alpha_1\sin(\omega t)+\alpha_3\sin(3\omega t)+\alpha_5\sin(5\omega t))$	$0.1\leq\alpha\leq 0.9, T\leq t_2-t_1\leq 9T$ $0.05\leq\alpha_3\leq 0.15, 0.05\leq\alpha_5\leq 0.15; \sum\alpha_i^2=1$

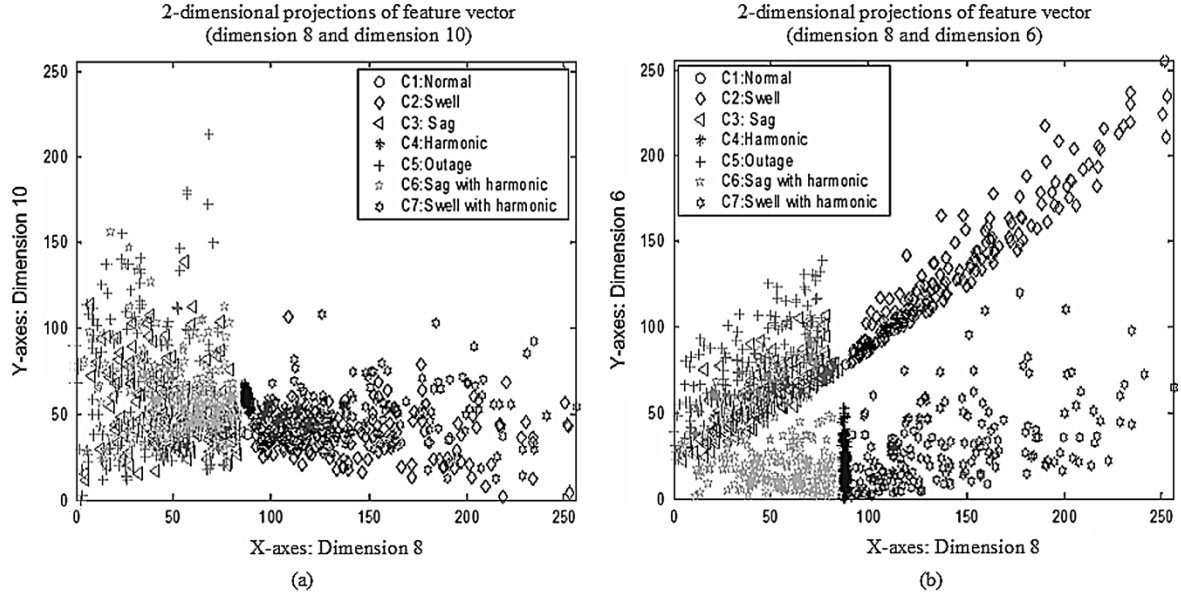

Fig. 4. Two-dimensional projections of the feature vector. (a) Dimension 8 and dimension 10. (b) Dimension 8 and dimension 6.

B. Simulation Results

Daubechie's 4 (Db4) wavelets with ten levels of decomposition were used for analysis ($l = 10$). Based on the feature extraction shown in Fig. 3, 11-dimensional feature sets for training and testing data were constructed. The dimensions here describe different features resulting from the wavelet transform, that is to say, the total size of the training data or testing data set is 1400×11, where 1400 comes from 200 cases per class multiplied by 7 classes and 11 is the dimension of the feature size of each case. All data sets were scaled to the range of (1–255) before being applied to SOLAR for training and testing. Figs. 4(a) and 4(b) shows two two-dimensional projections of the training set. As we can see from Fig. 4, in some dimensions, such as dimension 8 (x axes) and dimension 10 (y axes) shown in Fig. 4(a), the data sets are mixed up, while in some other dimensions, such as dimension 8 (x axes) and dimension 6 (y axes) shown in Fig. 4(b), the data are better separated. Since SOLAR will dynamically choose its functionality and connection structure based in its learning stage, it will easily handle this kind of data classification problem.

In order to evaluate the performance of the proposed method, we compared the classification result with the recent results reported in literature [24] and the classification using the Support Vector Machine (SVM) method. Simulation of the SVM for classification is based on the modification of the Ohio State University (OSU) SVM Classifier Matlab Toolbox [25]. Two types of SVM were implemented in our simulation: C-Support Vector Classification and ν-Support Vector Classification. For each type of SVM, the following four kernel functions $K(x_i, x_j) = \phi(x_i)^T\phi(x_j)$ are taken into account, where x_i and x_j are the feature vectors in the input space (i and j denotes the index of the features) and ϕ is the mapping function

Linear:

$$K(x_i, x_j) = x_i^T x_j \tag{18}$$

Polynomial:

$$K(x_i, x_j) = \left(\gamma x_i^T x_j + r\right)^d, \gamma > 0 \tag{19}$$

Radial Basis Function (RBF):

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right), \gamma > 0 \tag{20}$$

Sigmoid:

$$K(x_i, x_j) = \tanh\left(\gamma x_i^T x_j + r\right) \tag{21}$$

where γ, r and d are kernel parameters. The detailed discussion about these types of SVMs and the kernel functions can be referred in [26]–[28].

Tables II–IV gives the simulation result for this seven-class PQ disturbance problem based on the proposed method (SOLAR based on wavelet feature extraction), the method reported in paper [24] (inductive inference approach), and SVM classification, respectively. As the results in Tables II–IV, a 7×7 confusion matrix C is constructed to show the classification performance for each method. The diagonal elements represent the correctly classified PQ types. The off-diagonal elements represent the misclassification. As we can see here, the method proposed in this paper can effectively classify different kinds of PQ disturbances. One thing that should be noted here is that, it is recognized that an optimum selection of the SVM type, kernel function and parameter setting may have better results than those listed in Table IV. However, it is not easy to choose the optimum SVM kernel functions and parameters in advance. The proposed method in this paper can automatically provide statistically stable results. Further discussion about the selection of SVM kernel functions and parameters can be found in papers [26]–[28].

In order to get the detailed performance evaluation, two more questions should be investigated:

1) What is the relationship between the classification performance and the choice of the decomposition levels?
2) How to choose a suitable wavelet for analysis in the proposed method?

The following parts (C) and (D) will discuss these two issues.

C. Discussion About the Classification Performance and Wavelet Decomposition Levels

As discussed in Section II, l level wavelet decomposition will decrease the $(l+1)$ dimensional feature vector for future analysis. Obviously, more levels of decomposition will increase the computational cost. So, how should a reasonable number of decomposition levels be chosen?

In this simulation, we choose Db4 wavelet and scan the decomposition levels from 1 to 10. For each type of PQ class (C1 to C7 as in Table I), 200 cases were generated with different parameters for training and another 200 cases were generated for testing. Sample frequency was 256 points/cycle. Fig. 5 gives the simulation results.

TABLE II
CLASSIFICATION RESULT FOR THE PROPOSED METHOD IN THIS PAPER (SOLAR BASED ON WAVELET FEATURE EXTRACTION)

	C1	C2	C3	C4	C5	C6	C7
C1	200	0	0	0	0	0	0
C2	0	200	0	0	0	0	0
C3	1	0	174	0	24	1	0
C4	0	0	0	200	0	0	0
C5	15	0	16	0	161	8	0
C6	0	0	2	1	2	194	1
C7	0	0	0	0	0	0	200
Overall accuracy				***94.93%***			

TABLE III
CLASSIFICATION RESULTS AS REPORTED IN PAPER [24]: INDUCTIVE INFERENCE APROACH

	C1	C2	C3	C4	C5	C6	C7
C1	200	0	0	0	0	0	0
C2	0	194	0	0	0	0	6
C3	0	0	153	0	11	36	0
C4	0	0	0	200	0	0	0
C5	0	0	1	0	180	19	0
C6	0	0	42	0	15	143	0
C7	0	4	0	0	0	0	196
Overall accuracy				***90.4%***			

As we can see from Fig. 5, when the wavelet decomposition levels are relatively small, such as $l \leq 3$, the overall classification accuracy is about 60%. When the decomposition level is greater than 6, we can achieve about 90% classification accuracy. Further investigation shows that when $l \geq 6$, the increase of the classification probability is small. Fig. 5 provides a reference for trade off between the decomposition levels and classification performance.

D. Choice of Suitable Wavelet

Another import issue related to the proposed method is the choice of a suitable wavelet. Obviously, the longer the wavelet filter length, the higher the computational cost. In this part, we investigate the influence of different kinds of wavelets to the classification accuracy of the proposed method.

Conjecture 1: No single wavelet transform has a statistically significant advantage over other wavelets in performance of the proposed method for PQ classification.

To verify conjecture 1, four commonly used wavelets, named Haar wavelet, Daubechie's wavelet, Symlets and Coiflets wavelet are taken into account. Table V shows their corresponding characteristics. Here the orthogonal means that the inner products of wavelet basis functions are zero. The compact support and support width is used to measure the domain of wavelet function with nonzero values on it. Compact support means these nonzero values are only for a finite duration and support width means how long this nonzero duration is. These parameters will influence the frequency characteristics of the wavelet transform. A wavelet with small support width is fast to compute, but the narrowness in time domain implies a large width in frequency domain. Conversely, wavelets with large compact support are smoother and have finer frequency

TABLE IV
CLASSIFICATION RESULTS FOR SVM METHOD

	C-SVM								ν-SVM							
		C1	C2	C3	C4	C5	C6	C7		C1	C2	C3	C4	C5	C6	C7
Linear kernel function	C1	200	0	0	0	0	0	0	C1	200	0	0	0	0	0	0
	C2	32	168	0	0	0	0	0	C2	43	157	0	0	0	0	0
	C3	36	0	132	0	31	1	0	C3	55	0	113	0	31	1	0
	C4	0	0	0	200	0	0	0	C4	0	0	0	177	0	10	13
	C5	0	0	11	0	189	0	0	C5	3	0	13	0	184	0	0
	C6	0	0	1	0	10	189	0	C6	0	0	1	0	15	184	0
	C7	0	0	0	0	0	0	200	C7	0	0	0	0	0	0	200
	Overall accuracy	**91.29%**							Overall accuracy	**86.79%**						
Polynomial Kernel function		C1	C2	C3	C4	C5	C6	C7		C1	C2	C3	C4	C5	C6	C7
	C1	200	0	0	0	0	0	0	C1	200	0	0	0	0	0	0
	C2	31	169	0	0	0	0	0	C2	44	156	0	0	0	0	0
	C3	36	0	140	0	23	1	0	C3	54	0	117	0	28	1	0
	C4	0	0	0	196	0	4	0	C4	0	0	0	155	0	45	0
	C5	0	0	21	0	179	0	0	C5	3	0	18	0	179	0	0
	C6	0	0	2	0	7	191	0	C6	0	0	1	0	14	185	0
	C7	0	0	0	0	0	0	200	C7	0	0	0	0	0	0	200
	Overall accuracy	**91.07%**							Overall accuracy	**86.14%**						
RBF kernel function		C1	C2	C3	C4	C5	C6	C7		C1	C2	C3	C4	C5	C6	C7
	C1	200	0	0	0	0	0	0	C1	200	0	0	0	0	0	0
	C2	12	188	0	0	0	0	0	C2	38	162	0	0	0	0	0
	C3	17	0	158	0	23	2	0	C3	52	0	121	0	26	1	0
	C4	0	0	0	200	0	0	0	C4	0	0	0	181	0	0	19
	C5	0	0	7	0	193	0	0	C5	0	0	9	0	191	0	0
	C6	0	0	3	0	11	186	0	C6	0	0	1	0	24	175	0
	C7	0	0	0	0	0	0	200	C7	0	0	0	0	0	0	200
	Overall accuracy	**94.64%**							Overall accuracy	**87.86%**						
Sigmoid kernel function		C1	C2	C3	C4	C5	C6	C7		C1	C2	C3	C4	C5	C6	C7
	C1	200	0	0	0	0	0	0	C1	200	0	0	0	0	0	0
	C2	0	200	0	0	0	0	0	C2	39	161	0	0	0	0	0
	C3	0	0	171	0	28	1	0	C3	55	0	112	0	32	1	0
	C4	0	0	0	200	0	0	0	C4	0	0	0	179	0	0	21
	C5	0	0	33	0	164	3	0	C5	3	0	10	0	187	0	0
	C6	0	0	1	0	6	193	0	C6	0	0	2	0	15	183	0
	C7	0	0	0	0	0	0	200	C7	0	0	0	0	0	0	200
	Overall accuracy	**94.86%**							Overall accuracy	**87.29%**						

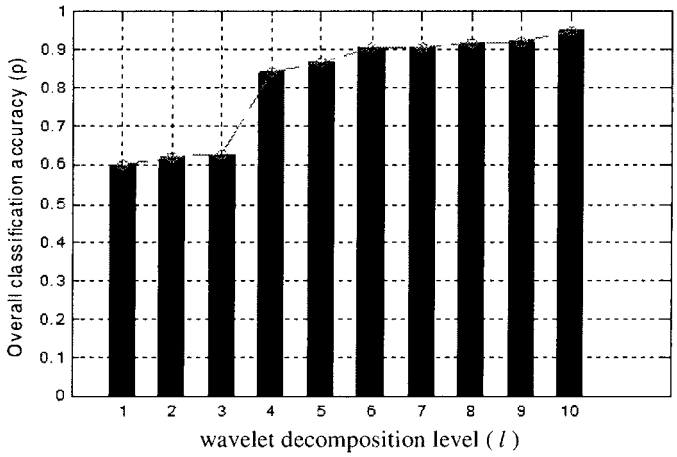

Fig. 5. Relationship between the wavelet decomposition levels and overall classification accuracy.

resolution. Filter length is the number of filter coefficients as discussed in Section II. This parameter will affect the computational cost of the wavelet transform and the longer the wavelet filters length, the larger the computational cost. Symmetry is another desired property for wavelets, however, symmetry prevents orthogonality. The only filters with properties of both orthogonality and symmetry are Haar filters (with 2 filter coefficients). A detailed discussion about these wavelet characteristics can be found in paper [18] and [19].

TABLE V
WAVELET CHARACTERISTICS

Wavelet Name	Orthogonal	Compact support	Support Width	Filters length	Symmetry
Haar	Yes	Yes	1	2	Yes
Daubechies	Yes	Yes	2N-1	2N	Far from
Coiflets	Yes	Yes	6N-1	6N	Near from
Symlets	Yes	Yes	2N-1	2N	Near from

In order to evaluate the performance of different kind of wavelets, we fix the decomposition levels at 6. The reason we choose six levels of decomposition is based on the results in Fig. 5, in which we can see that six levels decomposition can provide nice classification accuracy (about 90%) as well as relatively small computational cost compared to higher levels of decomposition, such as nine levels or ten levels. The tradeoff between the decomposition levels and classification performance is also discussed in part C of this section. Seven classes of different PQ disturbances as defined in Section IV part A were considered here. For each choice of wavelet, we took ten simulation runs. In each run, 200 cases of each class (seven classes total) of PQ disturbances based on the model in Table I were generated for training and another 200 cases of each class were generated for testing. The averaged accuracy over the ten runs for each choice of wavelet were presented in Table VI.

To test if there is any significant difference among different wavelet families, or among different wavelets within one

TABLE VI
AVERAGED CLASSIFICATION RESULTS FOR DIFFERENT WAVELETS

Wavelet name	Haar	Db2	Db3	Db4	Db5	Db6	Db7	Db8	Db9	Db10	Coif1
Average accuracy	0.8700	0.8557	0.8746	0.8908	0.8909	0.8786	0.8784	0.8776	0.8996	0.8882	0.8903
Wavelet name	Coif2	Coif3	Coif4	Coif5	Sym2	Sym3	Sym4	Sym5	Sym6	Sym7	Sym8
Average accuracy	0.8921	0.8819	0.8694	0.8883	0.8646	0.8856	0.8692	0.8644	0.8755	0.8894	0.8650

wavelet family, we use hypothesis testing of the average values [29]. Here we show the case of the testing among different wavelet families.

Table VII gives the mean and standard deviation results of each family of wavelet as presented in Table VI. The mean μ and the standard deviation σ of the population are calculated using the following equations:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} p_i \tag{22}$$

$$\sigma = \sqrt{\frac{n\sum_{i=1}^{n} p_i^2 - \left(\sum_{i=1}^{n} p_i\right)^2}{n(n-1)}} \tag{23}$$

where p_i is the classification accuracy of different wavelets in each wavelet family in Table VI and n is the number of wavelets in each family.

We now formulate the hypothesis:

Null hypothesis:

$$H_0 : \mu_1 = \mu_2 \tag{24}$$

Alternative hypothesis:

$$H_1 : \mu_1 \neq \mu_2 \tag{25}$$

The test statistic is calculated as follows:

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{26}$$

For a two-tailed test, we will reject H_0 if $|Z| > 1.96$. (1.96 is for a two-tailed test where the results are significant at a level of 0.05). Table VIII shows the hypothesis testing result.

From the analysis results in Table VIII we can see, we will accept the null hypothesis H_0, which means there is no difference in the mean values. This shows that there is no statistically significant difference in the classification performance when different wavelet families are chosen. The same test can be performed for the case of different wavelets within one wavelet family and we can get the same results. Fig. 6 shows the analysis results for different wavelets within one wavelet family, namely Daubechies wavelet and their corresponding wavelet filter length.

So far, we have shown that there is no statistically significant difference in performance of the proposed method for PQ classification when different wavelets are chosen. This means the proposed method has good robustness characteristics and has no strict requirements for the choice of a particular wavelet. However, different wavelets have different filter lengths. The longer the wavelet filters length, the larger the computational cost. Based on the results in this part, we can choose the wavelet with short wavelet filter length, such as Haar wavelet, Db2 wavelet to achieve both good classification results as well as small computational cost.

TABLE VII
PERFORMANCE OF WAVELETS

wavelet	p	wavelet	p	wavelet	p
Haar(Db1)	0.8700	Coif1	0.8903	Sym2	0.8646
Db2	0.8557	Coif2	0.8921	Sym3	0.8856
Db3	0.8746	Coif3	0.8819	Sym4	0.8692
Db4	0.8908	Coif4	0.8694	Sym5	0.8644
Db5	0.8909	Coif5	0.8883	Sym6	0.8755
Db6	0.8786			Sym7	0.8894
Db7	0.8784			Sym8	0.8650
Db8	0.8776				
Db9	0.8996				
Db10	0.8882				
Mean(μ)	0.8804		0.8844		0.8734
Standard deviation(σ)	0.0125		0.0092		0.0105

TABLE VIII
WAVELET FAMILY HYPOTHESIS TEST

Wavelet name	Wavelet name	$\lvert Z\rvert$	Accept or reject H_0
Daubechies	Coiflets	0.7011	Accept
Daubechies	Symlet	1.2497	Accept
Coiflets	Symlet	1.9243	Accept

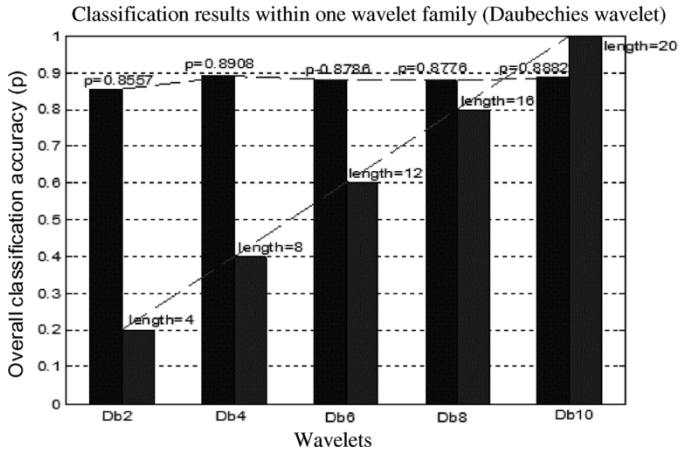

Fig. 6. Classification results and their corresponding wavelet filter length for different wavelets within one family (Daubechies wavelet).

V. NOISE ANALYSIS FOR THE PROPOSED METHOD

Since noise is omnipresent in an electrical power distribution network, we analyze whether the proposed method is still effective in a noisy environment.

Gaussian white noise is widely considered in the research of power quality issues [17], [30], [31]. To test the proposed method performance in different noise environments, different levels of noises with the signal to noise ratio (SNR) values

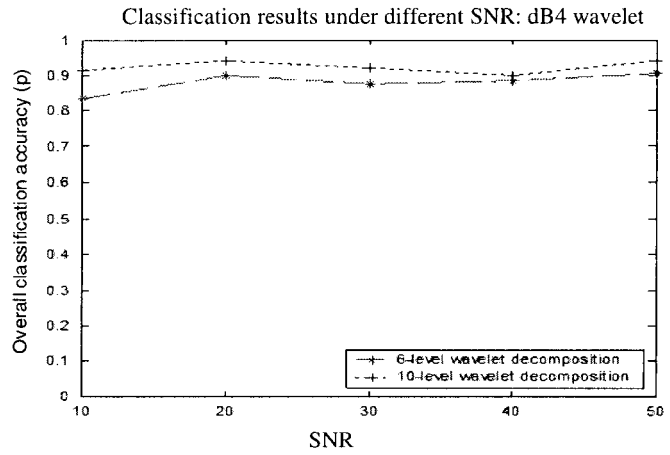

Fig. 7. Classification results under different SNR condition.

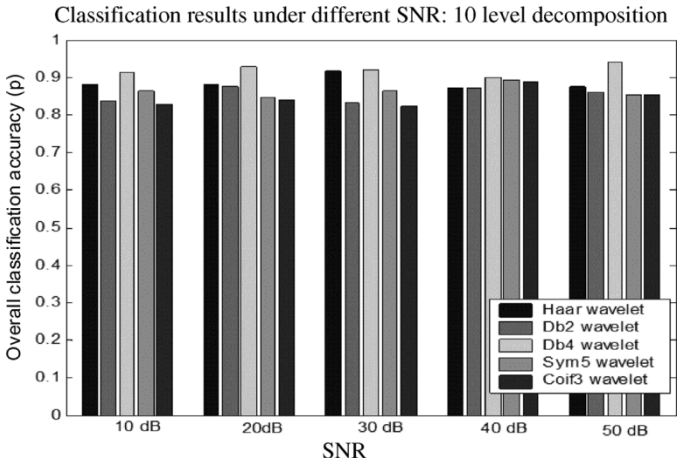

Fig. 8. Different wavelet classification results under different SNR conditions.

ranging from 10 to 50 dB are considered here. The value of SNR is defined as following

$$\text{SNR} = 10\log(P_s/P_n)\,dB \tag{27}$$

where P_s is the power (variance) of the signal and P_n is that of the noise. The Monte Carlo method is used to generate the simulation data set with different parameters as show in Table I. For each class of PQ disturbance (C1 to C7 in Table I), 200 cases were generated for training and another 200 cases were generated for testing. We choose the Db4 wavelet for analysis with six and ten levels of decomposition, respectively. Fig. 7 shows the simulation results. As we can see, even in very low SNR condition, the proposed method can still achieve high overall classification accuracy. The ten levels of decomposition classification result is slightly better than that of 6 levels decomposition, but the improvement is not very big. This is consistent with our previous result as shown in Fig. 5.

Fig. 8 shows the overall classification performance for different wavelets in a noisy environment. We choose ten levels of wavelet decomposition here. As we can see from Fig. 8, the proposed method has a robust anti-noise performance and we can still achieve high classification accuracy in a noisy environment. In addition, although Db4 wavelets show slightly better classification results among the chosen wavelets, there is no statistically significant difference in performance of the proposed method within different wavelets. This is also consistent with our previous results in Section IV part D.

VI. Conclusion

A novel PQ classification approach based on the wavelet transform and self organizing learning array system is proposed in this paper. Wavelet transform is utilized to construct the feature vector based on the multiresolution analysis method. These feature vectors are then applied to a SOLAR system for training and testing. Several typical PQ disturbances are taken into consideration in this paper. Comparison between the proposed method, the support vector machine (SVM) and existing literature reports show that the proposed method can effectively classify different kinds of PQ disturbances.

We have shown that there is no statistically significant difference in performance of the proposed method when different wavelets are chosen. This means that the simplest wavelet to implement the proposed method will do as good a job as any other wavelets, at least for the PQ disturbance classification problem. In addition, simulation results under different noise levels show that the proposed method works well in noisy environment. The ideas of the combining wavelet transform with self organizing learning array system could potentially be used in other problem domains, such as financial data analysis, automatic target recognition, etc.

Acknowledgment

The authors would like to thank the reviewers for their helpful comments and suggestions.

References

[1] M. Samotyj, "The Cost of power disturbance to industrial and digital economy companies," Consortium for electrical infrastructure to support a digital society, an initiative by EPRI and the Electrical Innovation Institute, Jun. 2001. Primen.

[2] W. R. A. Ibrahim and M. M. Morcos, "Artificial intelligence and advanced mathematical tools for power quality applications: A survey," *IEEE Trans. Power Del.*, vol. 17, no. 2, pp. 668–673, Apr. 2002.

[3] S. Santoso, J. Lamoree, W. M. Grady, E. J. Powers, and S. C. Bhatt, "A scalable PQ event identification system," *IEEE Trans. Power Del.*, vol. 15, no. 4, pp. 738–743, Apr. 2000.

[4] J. Huang, M. Negnevitsky, and D. T. Nguyen, "A neural-fuzzy classifier for recognition of power quality disturbances," *IEEE Trans. Power Del.*, vol. 17, no. 4, pp. 609–616, Apr. 2002.

[5] A. Elmitwally, S. Farghal, M. Kandil, S. Abdelkader, and M. Elkateb, "Proposed wavelet-neurofuzzy combined system for power quality violations detection and diagnosis," *Proc. Inst. Electr. Eng. Gener. Transm. Distrib.*, vol. 148, pp. 15–20, Jan. 2001.

[6] S. A. Farghal, M. S. Kandil, and A. Elmitwally, "Quantifying electric power quality via fuzzy modeling and analytic hierarchy processing," *Proc. Inst. Electr. Eng. Gener. Transm. Distrib.*, vol. 149, pp. 44–49, Jan. 2002.

[7] A. Elmitwally, S. Abdelkader, and M. Elkateb, "Universal power quality manager with a new control scheme," *Proc. Inst. Electr. Eng. Gener. Transm. Distrib.*, vol. 147, pp. 183–189, May 2000.

[8] S. Santoso, E. J. Powers, W. M. Grady, and A. C. Parsons, "Power quality disturbance waveform recognition using wavelet-based neural classifier-part I: Theoretical foundation," *IEEE Trans. Power Del.*, vol. 15, no. 1, pp. 222–228, Jan. 2000.

[9] S. Santoso, E. J. Powers, W. M. Grady, and A. C. Parsons, "Power quality disturbance waveform recognition using wavelet-based neural classifier-part II: Application," *IEEE Trans. Power Del.*, vol. 15, no. 1, pp. 229–235, Jan. 2000.
[10] J. V. Wijayakulasooriya, G. A. Putrus, and P. D. Minns, "Electric power quality disturbance classification using self-adapting artificial neural networks," *Proc. Inst. Electr. Eng. Gener. Transm. Distrib.*, vol. 149, no. 1, pp. 98–101, Jan. 2002.
[11] C. K. Kung, M. J. Devaney, C. M. Huang, and C. M. Kung, "Fuzzy-based adaptive digital power metering using a genetic algorithm," *IEEE Trans. Instrum. Meas.*, vol. 47, no. 1, pp. 183–188, Jan. 1998.
[12] A. M. Gaouda, M. M. A. Salam, M. R. Sultan, and A. Y. Chikhani, "Power quality detection and classification using wavelet-multiresolution signal decomposition," *IEEE Trans. Power Del.*, vol. 14, no. 10, pp. 1469–1476, Oct. 1999.
[13] A. M. Gaouda, S. H. Kanoun, M. M. A. Salama, and A. Y. Chikhani, "Wavelet-based signal processing for disturbance classification and measurement," *Proc. Inst. Electr. Eng. Gener. Transm. Distrib.*, vol. 149, pp. 310–318, May 2002.
[14] M. Karimi, H. Mokhtari, and M. R. Iravani, "Wavelet based on-line disturbance detection for power quality application," *IEEE Trans. Power Del.*, vol. 15, no. 10, pp. 1212–1220, Oct. 2000.
[15] H. Mokhtari, M. K. Ghartemani, and M. R. Iravani, "Experimental performance evaluation of a wavelet-based on-line voltage detection method for power quality application," *IEEE Trans. Power Del.*, vol. 17, no. 1, pp. 161–172, Jan. 2002.
[16] S. J. Huang and M. J. Jou, "Application of arithmetic coding for electric power disturbance data compression with wavelet packet enhancement," *IEEE Trans. Power Syst.*, vol. 19, no. 4, pp. 1334–1341, Aug. 2004.
[17] H. T. Yang and C.-C. Liao, "A de-noising scheme for enhancing wavelet-based power quality monitoring system," *IEEE Trans. Power Del.*, vol. 16, no. 7, pp. 353–360, Jul. 2001.
[18] I. Daubechies, "Ten lectures on wavelets," . Philadelphia, PA: Society for Industrial and Applied Mathematics, 1992.
[19] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms: A Primer*. Englewood Cliffs, NJ: Prentice-Hall, 1998.
[20] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
[21] J. Starzyk, Z. Zhu, and T. H. Liu, "Self-organizing learning array," *IEEE Trans. Neural Netw.*, vol. 16, no. 2, pp. 355–363, Mar. 2005.
[22] J. Starzyk and T. H. Liu, "Design of a self-organizing learning array system," in *Proc. IEEE Int. Symp. Circuits Systems*, Bangkok, Thailand, May 2003, pp. 801–805.
[23] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.
[24] T. K. A. Galil, M. Kamel, A. M. Youssef, E. F. E. Saadany, and M. M. A. Salama, "Power quality disturbance classification using the inductive inference approach," *IEEE Trans. Power Del.*, vol. 19, no. 10, pp. 1812–1818, Oct. 2004.
[25] Ohio State University SVM Classifier Matlab Toolbox [Online]. Available: http://www.ece.osu.edu/-maj/osu_svm/
[26] B. Scholkopf, A. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Computation*, vol. 12, pp. 1207–1245, May 2000.
[27] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.
[28] C. Cortes and V. Vapnik, "Support-vector network," *Mach. Learning*, vol. 20, pp. 273–297, Sep. 1995.
[29] I. Miller and J. E. Fruend, *Probability and Statistics for Engineers*. Englewood Cliffs, NJ: Prentice-Hall, 1965.
[30] H. Zhang, P. liu, and O. P. Malik, "Detection and classification of power quality disturbances in noisy conditions," *Proc. Inst. Electr. Eng. Gener. Transm. Distrib.*, vol. 150, pp. 567–572, Sep. 2003.
[31] P. O'Shea, "A high-resolution spectral analysis algorithm for power-system disturbance monitoring," *IEEE Trans. Power Syst.*, vol. 17, no. 8, pp. 676–680, Aug. 2002.

Haibo He (S'04) received the B.S. and M.S. degrees in electrical engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 1999 and 2002 respectively. He is currently pursuing the Ph.D. degree in electrical engineering at Ohio University, Athens.

His research interests include machine learning, power quality and powerline communication, low power VLSI, and reconfigurable design for wireless communication.

Janusz A. Starzyk (SM'83) received the M.S. degree in applied mathematics and the Ph.D. degree in electrical engineering from Warsaw University of Technology, Warsaw, Poland, in 1971 and 1976, respectively.

He is currently a Professor of the School of Electrical engineering and Computer Science (EECS), Ohio University, Athens. He has been a consultant to AT&T Bell Laboratories, Sarnoff Research, Sverdrup Technology, Magnolia Broadband, and Magnetek Corporation. His current research is in the areas of self-organizing learning machines, neural networks, rough sets, VLSI design and test of mixed-signal CMOS circuits, and reconfigurable design for wireless communication.