Power Quality Disturbance Classification Using the Inductive Inference Approach

T. K. Abdel-Galil, *Member, IEEE*, M. Kamel, *Senior Member, IEEE*, A. M. Youssef, E. F. El-Saadany, *Member, IEEE*, and M. M. A. Salama, *Fellow, IEEE*

***Abstract*—This paper presents a novel approach for the classification of power quality disturbances. The approach is based on inductive learning by using decision trees. The wavelet transform is utilized to produce representative feature vectors that can accurately capture the unique and salient characteristics of each disturbance. In the training phase, a decision tree is developed for the power quality disturbances. The decision tree is obtained based on the features produced by the wavelet analysis through inductive inference. During testing, the signal is recognized using the rules extracted from the decision tree. The classification accuracy of the decision tree is not only comparable with the classification accuracy of artificial Neural networks, but also accounts for the explanation of the disturbance classification via the produced if… then rules.**

***Index Terms*—Decision tree, disturbance classification, monitoring techniques, power quality, wavelet transforms.**

I. Introduction

POWER QUALITY monitors handle a large amount of raw data, which renders analysis of power quality records prohibitively expensive due to time and expertise. Therefore, the automated classification of power quality disturbances has become a significant issue especially in the deregulated era. Recently, many automated systems for the classification of power quality have been designed. The proposed systems are based mainly on the artificial neural networks (ANNs) [1], [2], template matching [3], dynamic time warping [4], hidden Markov models [5], and rule-based systems, which rely on the knowledge of an expert in the field [6], [7]. ANNs have attracted a great deal of attention because of their inherent pattern recognition capabilities and their ability to handle noisy data. Template matching techniques rely on comparing the unknown signal with a set of signatures, and selecting the class of the signal according to the similarities of that signal and the stored signatures. The time-alignment problem, as well as the change of frequency, amplitude, and duration of the same disturbance, adds to the difficulties of the template matching task. Dynamic time warping and hidden Markov models are introduced to overcome such difficulties. All the previously mentioned techniques are data driven techniques, since they rely on data for training and storing suitable signatures for the power quality disturbances. These data driven techniques, although providing a reasonable classification errors, lack the capability to explain the obtained results.

On the contrary, rule-based systems for power quality disturbance classification depend on the development of a set of rules that describe the knowledge that is used by the human expert. The human expert provides the reasoning behind the classification decision. Then, a computer program is designed to incorporate the expert reasoning in the form of a set of rules. In [6], the authors have used this concept to develop a rule-based system to identify ten different power quality disturbances. However, as the dimensionality of such problem increases, the inference mechanism may become complicated, and additional rules may be required to obtain the required classification accuracy. As the authors have pointed out [6], the construction of such systems is both costly and time consuming. Moreover, the obtained rules rely mainly on features which are based on the statistical nature of the values of the wavelet coefficients. These statistics are function in both the wavelet that is used and the parameters of the system under consideration, which means that the rules must be adjusted every time the system configuration or parameters change. Also, obtaining the characteristics of the features for similar and combined events is a tedious task.

In this paper, a data mining approach to generate a set of inference rules by adopting a power quality disturbance database is proposed. Since the system infer the classification rules directly from the data, this strategy combines the advantages of a rule-based system and minimizes the cost and time associated with building such a system. This rule-based system can be easily modified to accommodate any future change in the system parameters. In this paper, the C4.5 machine learning algorithm [8], [9]is used to generate the production rules in the form of if…then rules.

The rest of this paper is organized as follows. The proposed system is detailed in Section II.The wavelet transform multiresolution analysis (MRA) for feature extraction is presented in Section III. A detailed overview of inductive inference along with a description of the C4.5 algorithm is given in Section IV. The testing results are examined in Section V, and Section VI concludes the paper.

II. System Architecture

The power quality automated classification system which uses a decision tree is divided into the following stages: segmentation, feature extraction, data reduction, and a decision tree as shown in Fig. 1.

Manuscript received March 25, 2003.

T. K. Abdel-Galil, E. F. El-Saadany, and M. M. A. Salama are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: tkabdelg@hivolt.uwaterloo.ca).

M. Kamel is with the Department of System Design, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

A. M. Youssef is with the Department of Combinatorics and Optimization, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

Digital Object Identifier 10.1109/TPWRD.2003.822533

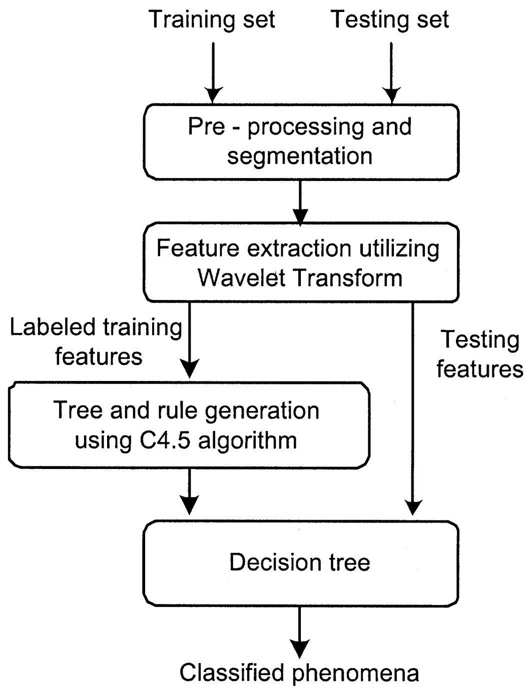

Fig. 1. System architecture block diagram.

A. Disturbance Segmentation

The disturbance segments must be extracted from the digitized voltage signal where the power quality disturbance occurs before the classification process is run. This can be accomplished by implementing a tracking and detecting algorithm to achieve a fast triggering of the classification schemes. For power quality disturbances that are associated with sudden changes, a modified Kalman filter can be employed to achieve this goal [10]; for power quality disturbances, which are characterized with relatively slow changes such as flickers, the Teager energy operator [11] can be utilized. In this paper, the decision tree is applied in the classification stage, and it is assumed that the clustering of the data which contains the disturbance has already been achieved.

B. Feature Extraction

The wavelet transform MRA is used as a tool to extract the most salient features that represent the power quality phenomenon [12], [14]. Before the application of this transform, a preprocessing of the signals is required to normalize them, because they are collected from various voltage levels in the distribution system. The size of the generated tree can be greatly reduced with the proper choice of a small set of features. In this paper, the MRA is employed as a clustering technique to generate a small set of features which are accurate, efficient, and reliable.

C. Decision Tree

The C4.5 algorithm [9] generates the decision tree by using the algorithm described in Section IV. The algorithm allows the decision tree to be simplified by pruning techniques which reduce the size of the tree according to a user-defined level. The C4.5 algorithm can produce either decision tree or rules of the form *if…then…*; both representation schemes are much more readable and understandable than other black-box machine learning approaches such as ANNs.

III. Wavelet Transform MRA

In this section, the discrete wavelet transform (DWT) is introduced as the basic tool for the decision tree feature extraction. DWT is the discrete counterpart of the continuous wavelet transform (CWT). The CWT of a continuous time signal x_t is defined as

$$CWT_{\psi}x(a,b) = \int_{-\infty}^{\infty} x(t)\psi^{*}_{a,b}(t)dt, \quad a, b \epsilon R, \ a \neq 0 \quad (1)$$

where

$$\psi^{\star}_{a,b}(t) = \frac{1}{\sqrt{a}}, \quad \psi^{\star}\left(\frac{t-b}{a}\right). \quad (2)$$

The function $\psi(t)$ is the mother wavelet, and the asterisk denotes a complex conjugate. a, b are the scaling and translating parameters, respectively. In practical applications, the DWT (3) of the sampled signal x_k is used to replace the CWT of x_t

$$\mathrm{DWT}_{\psi}(m,n) = \sum_k x_k \psi^{\star}_{m,n}(k) \quad (3)$$

where

$$\psi^{\star}_{m,n}(k) = \frac{1}{\sqrt{a_0^m}} \psi^{\star}\left(\frac{k - nb_0 a_0^m}{a_0^m}\right). \quad (4)$$

Both the scaling factor $a = a_0^m$ and the shifting factor $b = nb_0 a_0^m$ are functions of an integer parameter m, where m and n are scaling and sampling numbers, respectively, and $m = 0, 1, 2, \ldots$. By selecting $a_0 = 2$ and $b_0 = 1$, a representation of any signal x_k at various resolution levels can be developed by using the MRA. It is implemented by a set of successive filter banks [15] with the mother wavelet as a low-pass filter $h(n)$ and its dual as a high pass filter $g(n)$. Consider a general stage of the filter bank; then the relation between the input sequence $c_{j-1}(n)$ to this stage and its output, which is the detail coefficients $d_j(n)$ and approximated coefficients $c_j(n)$ is given as [15], [16]

$$c_j(n) = \sum_k h(k-2n)c_{j-1}(k) \quad (5)$$

and

$$d_j(n) = \sum_k g(k-2n)c_{j-1}(k) \quad (6)$$

where c_j represents the coefficients of the approximate signal at level j, and $d_j(k)$ represents the detailed coefficients of the signal at level j. The same procedure can be applied recursively. Thus, the signal is mapped by the following set of coefficients:

$$c_{\mathrm{signal}} = [c_0|d_0|d_1|\cdots d_{J-1}|]. \quad (7)$$

The energy of the signal W_{signal} is partitioned in terms of the wavelet coefficients as [13]

$$W_{\mathrm{signal}} = W_{c_0} + \sum_{j=0}^{J-1} W_{d_j} \quad (8)$$

and

$$W_{\text{signal}} = \sum_k |c_0(k)|^2 + \sum_{j=0}^{J-1} \sum_k |d_j(k)|^2 \tag{9}$$

where W_{c_0} is the energy of the approximate level and W_{d_j} is the disturbance energy at the detailed level j. The Haar wavelet is utilized to decompose the signal into its resolution levels. The energy at the different resolution levels represents one feature vector. The sampling frequency used with the ten power cycles permit decomposing the signal into 11 levels of resolution with the fundamental frequency in the eighth level. After the calculation of the energy vector, the energy vectors for the training and testing sets are normalized by dividing on the eighth coefficient of the normal case energy vector. This level contains most of the energy of the normal signal, and the normalized training vectors are utilized to deduce the tree. It is worth mentioning that the proposed set of features is not affected by the normal frequency variation, since the split of the energy among the different resolution levels are not affected by those variations around the nominal frequency.

IV. Inductive Inference Algorithm

A decision tree is a tree data structure consisting of a root node, decision nodes, and leaf nodes. A deterministic decision tree classifies power quality disturbances by sorting them from the root to some leaf node, which provides the identification of a power quality disturbance. A decision node specifies a test over one of the features which is called the attribute (feature)that is selected at the node, and each branch descending from that node corresponds to one of the possible values of the selected feature. A leaf identifies a class value (sag, swell,…,etc.). Decision tree algorithms use a set of training power quality cases. Each training case is specifically assigned values for the normalized energy at different resolution levels for different power quality disturbances which are considered as the features of interest. Associated with each training case is a label that represents the name of a class (sag, swell,…, etc.). Classes are denoted by $[S_1, S_2, \ldots, S_N]$. Fig. 2 illustrates a simple decision tree for power quality disturbance classification; this tree can be viewed as a partitioning of the feature space $[v_{rms}, duration]$. Each partition, represented by a leaf, contains the objects that are similar in relevant respects (sag, swell, overvoltage, and undervoltage), and thus are expected to belong to the same class.

A. Decision Tree Learning Algorithm

A divide and conquer strategy is used to construct the decision tree in the C4.5 algorithm [17]. Each node in a tree is associated with a set of cases which may belong to different classes $[S_1, S_2, \ldots, S_N]$. At the beginning, the algorithm begins with the root node and the whole training cases. Then, the following divide and conquer algorithm is executed to exploit the locally best choice; no backtracking is permitted.

Let T be a set of cases associated at the node of interest. First, the frequency $freq(S_i, T)$ of the cases of T whose class S_i for $i = [1, 2, \ldots, N]$ is calculated. If all or most of the cases in T belong to a same class S_m, then the node is a leaf with the associated class S_m, or else if T contains cases belonging to

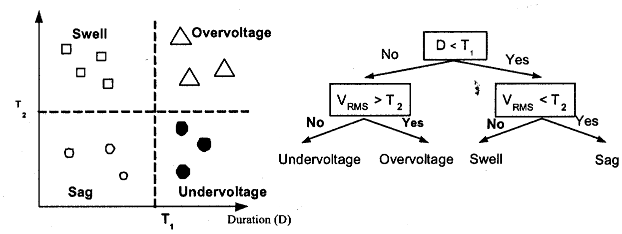

Fig. 2. Simple decision tree using continuous features.

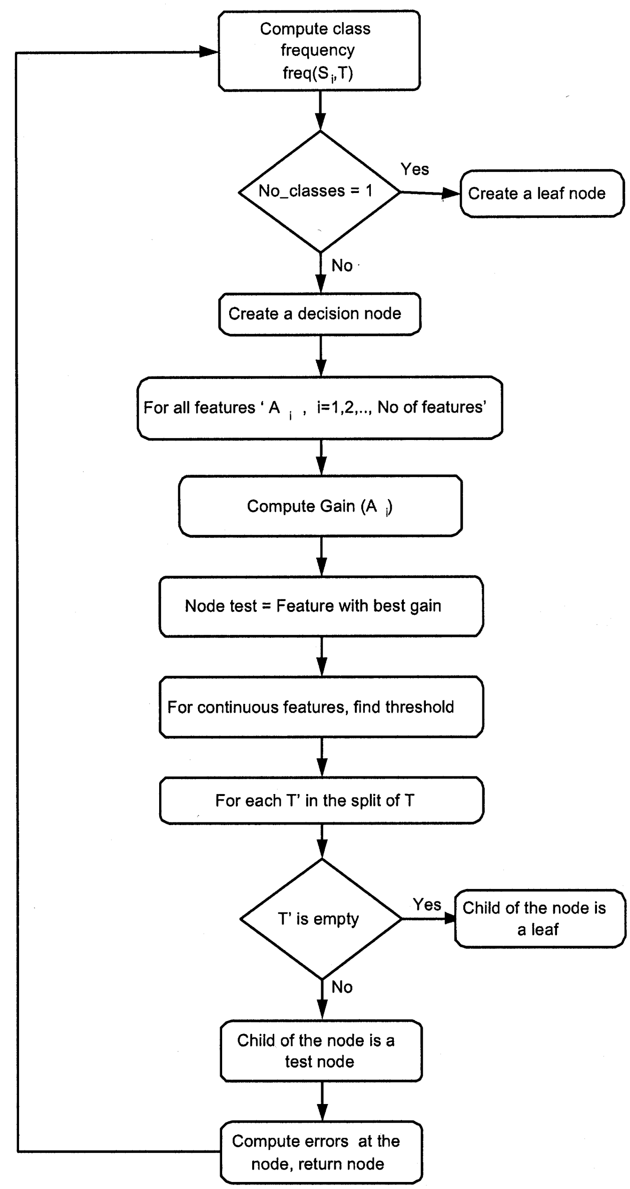

Fig. 3. Decision tree learning algorithm.

two or more classes, then the information gain for each feature is calculated. The calculation of the information gain for discrete, as well as continuous features, is discussed in Section IV-B The feature with the highest information gain is chosen as the test feature at the node. In general, a decision node has k children $[T_1, T_2, \ldots, T_k]$, with $k = 2$ in the case of continuous feature and $k = D$ for the discrete feature with D possible values.

TABLE I
DISTURBANCE SIGNAL MODELING

Event	Symbol	Controlling Parameters	Equation
Pure Sinusoidal	S_1	N/A	$v(t) = \sin(\omega t)$
Sudden Swell	S_2	$0.1 \leq \alpha \leq 0.8, T \leq t_2 - t_1 \leq 9T$	$v(t) = A(1 + \alpha(u(t_2) - u(t_1))) \sin(\omega t)$
Sudden Sag	S_3	$0.1 \leq \alpha \leq 0.9, T \leq t_2 - t_1 \leq 9T$	$v(t) = A(1 - \alpha(u(t_2) - u(t_1))) \sin(\omega t)$
Harmonics	S_4	$0.05 \leq \alpha_3 \leq 0.15, 0.05 \leq \alpha_5 \leq 0.15,$ $0.05 \leq \alpha_7 \leq 0.15, \Sigma\alpha_i^2 = 1$	$v(t) = A(\alpha_1 \sin(\omega t) + \alpha_3 \sin(3\omega t) +$ $\alpha_5 \sin(5\omega t) + \alpha_7 \sin(7\omega t))$
Outage	S_5	$0.9 \leq \alpha \leq 1, T \leq t_2 - t_1 \leq 9T$	$v(t) = A(1 - \alpha(u(t_2) - u(t_1)))$
Sag with Harmonics	S_6	$0.1 \leq \alpha \leq 0.9, T \leq t_2 - t_1 \leq 9T$ $0.05 \leq \alpha_3 \leq 0.15, 0.05 \leq \alpha_5 \leq 0.15$	$v(t) = A(1 - \alpha(u(t_2) - u(t_1)))$ $(\sin(\omega t) + \alpha_3 \sin(3\omega t) + \alpha_5 \sin(5\omega t))...$
Swell with Harmonics	S_7	$0.1 \leq \alpha \leq 0.8, T \leq t_2 - t_1 \leq 9T$ $0.05 \leq \alpha_3 \leq 0.15, 0.05 \leq \alpha_5 \leq 0.15$	$v(t) = A(1 + \alpha(u(t_2) - u(t_1)))$ $(\sin(\omega t) + \alpha_3 \sin(3\omega t) + \alpha_5 \sin(5\omega t))...$

For $i = [1, k]$, if T_i is empty, the child node is directly set to be a leaf with the associated class set to be the most frequent class at the parent node with zero classification error. If T_i is not empty, the previous operations are recursively applied. Finally the classification error of the last step is calculated as the sum of the errors of the child nodes. If the result is greater than the error of classifying all the cases in T, then the node is set to be a leaf and all the subtrees are removed.

B. Feature Selection by Information Gain Measurement

C4.5 uses the concept of information gain to build the decision tree. The information gain can be described as the effective decrease in entropy, resulting from making a choice as to which attribute to use and at what level. Consider a feature A for a set of cases T. The information gain of this attribute is calculated as follows:

Discrete Feature: Let N be the number of classes and $p(T, j)$ be the proportion of cases in T that belong to the jth class. The residual uncertainty about the class to which a case in T belongs can be expressed by the entropy function as

$$\text{info}(T) = -\sum_{j=1}^{N} p(T, j) * \log_2(p(T, j)) \tag{10}$$

where

$$p(T, j) = \frac{\text{freq}(S_j, T)}{|T|} \tag{11}$$

and the corresponding information gained by a test feature A with k outcomes as

$$\text{Gain}(T, A) = \text{Info}(T) - \sum_{i=1}^{k} \frac{|T_i|}{|T|} * \text{info}(T_i). \tag{12}$$

The information gained by a test is strongly affected by the number of outcomes k, and is maximal when there is one case in each subset T_i, which means that the information gain is biased to the attributes with many values over those with few values. To overcome this difficulty, a measure which is called split information is calculated as

$$\text{Split}(T, A) = -\sum_{i=1}^{k} \frac{|T_i|}{|T|} * \log_2\left(\frac{|T_i|}{|T|}\right). \tag{13}$$

This split information tends to increase with the number of outcomes of a test. Then, the gain ratio criterion of a feature A, defined as the ratio of its information gain to its split information, is used instead of the information gain measure. The gain ratio of every possible test is determined, and among those with at least an average gain, the split with maximum gain ratio is selected.

Continuous Feature: If A is a continuous feature, the task is converted to finding a threshold γ such that $A \leq \gamma$ results in two outcomes, true and false. To find the threshold γ that maximizes the splitting criterion, the cases in T are sorted according to their values of feature A in order to give ordered distinct values $[v_1, v_2, \ldots, v_m]$. Consider for $i \in [1 : m - 1]$, every pair of adjacent values suggests a threshold $\gamma = (v_i + v_{i+1})/2$ and the splitting

$$[T_1^{\gamma} = v_j | v_j \leq \gamma, T_2^{\gamma} = v_j | v_j \geq \gamma]. \tag{14}$$

For each value γ, the information gain $gain_{\gamma}$ is computed by considering the above splitting. The value γ' for which $gain_{\gamma'}$ is maximum is set to be the local threshold, and the information gain for attribute A is defined as $gain_{\gamma'}$.

V. APPLICATION AND RESULTS

A. Signal Modeling

To generate the tree, 1400 cases are used, and another 1400 examples are used for testing purposes. These cases are generated by using algebraic equations (Table I) and EMTDC/PSCAD3.08 simulation software.

The advantage of using algebraic equations is that the flexibility it introduces covering a wide range of variations, which is a required feature in building the tree. However, the use of EMTDC/PSCAD3.08 simulation software helps to generate cases which simulate real-time operation and mimic real power system disturbances. The algorithm shows consistency in the results with both algebraic equations and EMTDC/PSCAD3.08 simulation. Seven categories of disturbances are simulated, namely, undisturbed sinusoid, sudden swell, sudden sag, harmonics, outage, sag with harmonics, and swell with harmonics. The wavelet transform MRA is performed on the preprocessed signal. Both the test and training signals are sampled at 256 samples/cycle. Ten power frequency cycles which contain the disturbance is used for a total of 2560 points. This sampling rate can detect up to 7.6 kHz for power frequency equal to 60

TABLE II
RULES GENERATED FROM THE DECISION TREE

Rule	Description
1	$C_4 \le 0.0002 \wedge C_8 > 0.9285 \rightarrow S_1$
2	$C_5 \le 0.0099 \wedge C_6 > 0.0258 \wedge C_8 > 1.0669 \wedge C_{10} \le 0.0104 \rightarrow S_2$
3	$C_5 \le 0.0029 \wedge C_8 > 1.0669 \rightarrow S_2$
4	$C_1 > 0.0004 \wedge C_6 > 0.0258 \wedge C_8 > 0.9285 \wedge C_{10} \le 0.0104 \rightarrow S_2$
5	$C_4 > 0.0002 \wedge C_8 > 0.9285 \wedge C_8 \le 1.0669 \rightarrow S_4$
6	$C_5 > 0.0029 \wedge C_{10} > 0.0104 \rightarrow S_7$
7	$C_1 \le 0.0004 \wedge C_5 > 0.0099 \wedge C_8 > 1.0669 \rightarrow S_7$
8	$C_6 \le 0.0258 \wedge C_8 > 1.0669 \rightarrow S_7$
9	$C_5 > 0.0016 \wedge C_8 \le 0.9285 \wedge C_9 > 0.0083 \wedge C_{11} > 0.0007 \rightarrow S_5$
10	$C_5 > 0.0016 \wedge C_8 \le 0.9285 \wedge C_{11} > 0.002 \rightarrow S_5$
11	$C_4 > 0.0073 \wedge C_8 \le 0.9285 \wedge C_{11} > 0.0007 \rightarrow S_5$
12	$C_3 > 0.0037 \wedge C_6 > 0.0291 \wedge C_8 \le 0.9285 \rightarrow S_5$
13	$C_5 > 0.0016 \wedge C_8 \le 0.9285 \wedge C_9 > 0.0133 \wedge C_{11} \le 0.0005 \rightarrow S_5$
14	$C_5 > 0.0016 \wedge C_8 \le 0.9285 \wedge C_{10} > 0.0071 \rightarrow S_5$
15	$C_5 > 0.0016 \wedge C_6 > 0.0269 \wedge C_{10} \le 0.0039 \rightarrow S_5$
16	$C_3 > 0.0037 \wedge C_7 \le 0.2365 \rightarrow S_5$
17	$C_4 > 0.0033 \wedge C_5 \le 0.0014 \rightarrow S_5$
18	$C_1 \le 0.0001 \wedge C_4 > 0.0008 \wedge C_5 > 0.0016 \wedge C_8 \le 0.9285 \wedge C_{11} \le 0.0007 \rightarrow S_6$
19	$C_5 > 0.0016 \wedge C_7 > 0.4519 \wedge C_8 \le 0.9285 \rightarrow S_6$
20	$C_3 \le 0.0037 \wedge C_4 > 0.0015 \wedge C_5 > 0.0016 \wedge C_6 \le 0.0269 \wedge C_7 > 0.4026 \wedge C_8 \le 0.9285 \wedge C_9 \le 0.0133 \rightarrow S_6$
21	$C_3 \le 0.0037 \wedge C_4 \le 0.0051 \wedge C_5 > 0.0016 \wedge C_8 \le 0.9285 \wedge C_{10} > 0.0048 \wedge C_{10} \le 0.0071 \wedge C_{11} \le 0.0007 \rightarrow S_6$
22	$C_3 \le 0.0037 \wedge C_5 > 0.0016 \wedge C_6 > 0.014 \wedge C_8 \le 0.9285 \wedge C_9 > 0.0096 \wedge C_9 \le 0.0133 \wedge C_{10} > 0.0018 \wedge C_{10} \le 0.0071 \wedge C_{11} \le 0.0007 \rightarrow S_6$
23	$C_4 > 0.0004 \wedge C_4 \le 0.0033 \wedge C_5 > 0.0013 \wedge C_5 \le 0.0016 \wedge C_7 > 0.2635 \wedge C_8 \le 0.9285 \wedge C_{10} > 0.0032 \rightarrow S_6$
24	$C_5 > 0.0016 \wedge C_8 \le 0.9285 \wedge C_9 \le 0.0026 \wedge C_{11} \le 0.0007 \rightarrow S_6$
25	$C_2 \le 0 \wedge C_4 > 0.0004 \wedge C_5 > 0.0013 \rightarrow S_6$
26	$C_1 \le 0.0037 \wedge C_3 > 0.0037 \wedge C_5 > 0.0016 \wedge C_6 \le 0.0291 \wedge C_7 > 0.2365 \wedge C_8 \le 0.9285 \wedge C_{11} \le 0.0007 \rightarrow S_6$
27	$C_4 \le 0.0033 \wedge C_5 \le 0.0013 \wedge C_6 > 0.0154 \wedge C_8 \le 0.9285 \rightarrow S_3$
28	$C_4 \le 0.0004 \wedge C_5 \le 0.0016 \wedge C_8 \le 0.9285 \rightarrow S_3$
29	$C_1 > 0.0001 \wedge C_3 \le 0.0037 \wedge C_4 \le 0.0015 \wedge C_7 \le 0.4519 \wedge C_9 > 0.0026 \wedge C_9 \le 0.0096 \wedge C_{10} \le 0.0048 \wedge C_{11} \le 0.0007 \rightarrow S_3$
30	$C_1 > 0.0001 \wedge C_6 \le 0.0269 \wedge C_9 \le 0.0133 \wedge C_{10} \le 0.0018 \wedge C_{11} \le 0.0007 \rightarrow S_3$
31	$C_4 \le 0.0073 \wedge C_5 \le 0.0111 \wedge C_8 > 0.4032 \wedge C_9 \le 0.0083 \wedge C_{10} \le 0.0032 \wedge C_{11} > 0.0007 \wedge C_{11} \le 0.002 \rightarrow S_3$
32	$C_1 > 0.0001 \wedge C_3 \le 0.0037 \wedge C_6 \le 0.0269 \wedge C_7 \le 0.4026 \wedge C_9 > 0.0026 \wedge C_9 \le 0.0096 \wedge C_{10} \le 0.0048 \wedge C_{11} \le 0.0007 \rightarrow S_3$

Hz. However, in order to avoid aliasing, this sampling rate is used for the accurate detection of frequencies around 4.0 KHz.

B. Results

The application of the training signals generates a tree which can be represented by 32 rules presented in Table II. It can be seen that some of these rules can be extracted by using the knowledge of the power quality information and how energy is distributed at different resolution levels [12]. However, for combined events; example, sag contaminated with harmonics and swell contaminated with harmonics, it is very difficult to identify rules to classify these events. The program generates just one rule to identify the normal class. Since the coefficients are normalized so that the energy at the eighth resolution level for the normal case is equal to unity, the program identifies that in order to attain normal case, a c_8 should exceed a threshold of 0.92. However, this condition is not enough since there exist many phenomena which have the same characteristic. The search reveals that the normal case has a leakage energy which is less than 0.0002 at the fourth resolution level. This single rule is used to classify the two hundred examples correctly. Another interesting issue is that the program generates only one rule to identify the harmonic phenomena. Since, in simulating this phenomena, it is assumed that the THD is within 5% and the harmonic energy is distributed along different resolution levels. The generated rule identifies the harmonic such that c_8 is greater than 0.92 and less than 1.06. The 5% in the THD is mapped as a variation around unity at the 8th resolution level; at the same time, the rule states that c_8 should be greater than 0.0002 to classify the harmonic correctly. The two rules can be easily identified by a power quality expert, but it is a difficult task to infer the other rules for the remaining classes. This difficulty arises from the fact that the boundaries that separate sag from outage, or from sag contaminated with harmonic, are not crisp enough. This occurs because the value of attributes are very close to each other at the boundaries. The algorithm infers too many rules to classify and distinguish between these classes as shown in Table II. The detailed use of these rules for identifying and classifying the training examples is explained in Table III. The same argument holds also in the cases of swell and swell contaminated with harmonics, but there is less confusion since there are only two classes (not three) which may have similar attributes at the boundaries. In addition, this

TABLE III
RULE EVALUATION ON TRAINING DATA

Rule	Size	Used	Wrong	Class
1	2	200	0 (0.0%)	S_1
2	4	184	0 (0.0%)	S_2
3	2	9	0 (0.0%)	S_2
4	4	6	0 (0.0%)	S_2
5	3	200	0 (0.0%)	S_4
6	2	184	1 (0.5%)	S_7
7	3	3	0 (0.0%)	S_7
8	2	14	0 (0.0%)	S_7
9	4	89	1 (1.1%)	S_5
10	3	14	0 (0.0%)	S_5
11	3	16	0 (0.0%)	S_5
12	3	16	0 (0.0%)	S_5
13	4	28	0 (0.0%)	S_5
14	3	16	0 (0.0%)	S_5
15	3	9	2 (22.2%)	S_5
16	2	3	0 (0.0%)	S_5
17	2	7	0 (0.0%)	S_5
18	5	64	0 (0.0%)	S_6
19	3	21	0 (0.0%)	S_6
20	7	11	0 (0.0%)	S_6
21	7	27	3 (11.1%)	S_6
22	9	13	0 (0.0%)	S_6
23	7	12	0 (0.0%)	S_6
24	4	7	0 (0.0%)	S_6
25	3	9	1 (11.1%)	S_6
26	7	9	1 (11.1%)	S_6
27	4	91	0 (0.0%)	S_3
28	3	20	3 (15.0%)	S_3
29	8	30	1 (3.3%)	S_3
30	5	15	0 (0.0%)	S_3
31	7	13	1 (7.7%)	S_3
32	8	30	11 (36.7%)	S_3

TABLE IV
CLASSIFICATION RESULTS FOR TRAINING SIGNALS

	S_1	S_2	S_3	S_4	S_5	S_6	S_7
S_1	200	0	0	0	0	0	0
S_2	0	199	0	0	0	0	1
S_3	0	0	183	0	2	15	0
S_4	0	0	0	200	0	0	0
S_5	0	0	0	0	195	5	0
S_6	0	0	16	0	1	183	0
S_7	0	0	0	0	0	0	200
Accuracy	97.1%						

TABLE IV
CLASSIFICATION RESULTS FOR TESTING SIGNALS

	S_1	S_2	S_3	S_4	S_5	S_6	S_7
S_1	200	0	0	0	0	0	0
S_2	0	194	0	0	0	0	6
S_3	0	0	153	0	11	36	0
S_4	0	0	0	200	0	0	0
S_5	0	0	1	0	180	19	0
S_6	0	0	42	0	15	143	0
S_7	0	4	0	0	0	0	196
Accuracy	90.4%						

fact is clarified with the number of the deduced rules for each class and the use of these rules.

Table IV indicates the classification accuracy by using the training data with the generated decision tree. It is evident that the tree can correctly classify all the normal and harmonic cases. The tree misclassifies only one swell case, and the error increases when classifying the three similar phenomena, which are sag, outage, and sag with harmonics.

Table V shows the classification accuracy when the decision tree is used with the testing data. As it is expected, the classification accuracy drops to 91% for the unseen examples, and again, the maximum error occurs when classifying between sag, outage, and sag contaminated with harmonics.

VI. CONCLUSIONS

The identification and classification of power quality disturbances are crucial tasks for the precise monitoring of distribution systems. A new classification methodology that is based on machine inductive learning implemented using the C4.5 algorithm is proposed in this paper. The size of the generated tree is substantially decreased by the use of MRA. The proposed algorithm outperforms other classification techniques, since it can infer the classification tree from the training data. The C4.5 algorithm shows great potential for the future development of fully automated monitoring systems with on-line classification capabilities. In addition, the algorithm overcomes the need of a power quality expert to generate a rule-based power quality classifier, since it discovers its rule and infers it directly from the data.

REFERENCES

[1] A. K. Ghosh and D. L. Lubkeman, "The classification of power system disturbance waveforms using a neural network approach," *IEEE Trans. Power Delivery*, vol. 10, pp. 109–115, Jan. 1995.

[2] S. Santoso, E. J. Powers, W. M. Grady, and A. C. Parsons, "Power quality disturbance waveform recognition using wavelet-based neural classifier. I.Theoretical foundation," *IEEE Trans. Power Delivery*, vol. 15, pp. 222–228, Jan. 2000.

[3] T. K. Abdel-Galil, E. F. El-Saadany, A. M. Youssef, and M. M. A. Salama, "On-line disturbance recognition utilizing vector quantization based fast match," in *Proc. IEEE Power Eng. Soc. Summer Meeting*, Chicago, IL, July 21–25, 2002.

[4] A. M. Youssef, T. K. Abdel-Galil, E. F. El-Saadany, and M. M. A. Salama, "Disturbance classification utilizing dynamic time warping classifier," *IEEE Trans. Power Delivery*, to be published.

[5] J. Chung, E. J. Powers, W. M. Grady, and S. C. Bhatt, "Power disturbance classifier using a rule-based method and wavelet packet-based hidden markov model," *IEEE Trans. Power Delivery*, vol. 17, pp. 233–241, Jan. 2002.

[6] S. Santoso, J. Lamoree, W. M. Grady, E. J. Powers, and S. C. Bhatt, "A scalable PQ event identification system," *IEEE Trans. Power Delivery*, vol. 15, pp. 738–743, Apr. 2000.

[7] M. Kezunovic and Y. Liao, "A novel software implementation concept for power quality study," *IEEE Trans. Power Delivery*, vol. 17, pp. 544–549, Apr. 2002.

[8] T. Michael, *Machine Learning*. New York: McGraw-Hill, 1997.

[9] J. R. Quinlan, *C4.5: Programs for Machine Learning*. New York: Morgan Kaufmann, 1992.

[10] T. K. Abdel-Galil, E. F. El-Saadany, and M. M. A. Salama, "Power quality assessment in deregulated power systems," in *Proc. IEEE Power Eng. Soc. Winter Meeting*, New York, Jan. 2002, pp. 952–958.

[11] ——, "Energy operator for on-line tracking of voltage flicker levels," in *Proc. IEEE Power Eng. Soc. Winter Meeting*, New York, Jan. 2002, pp. 1153–1157.

[12] A. M. Gaouda, M. M. A. Salama, M. R. Sultan, and A. Y. Chikhani, "Power quality detection and classification using wavelet multi-resolution signal decomposition," *IEEE Trans. Power Delivery*, vol. 14, pp. 1469–1476, Oct. 1999.

[13] A. M. Gaouda, E. F. El-Saadany, M. M. A. Salama, V. K. Sood, and A. Y. Chikhani, "Monitoring HVDC systems using wavelet multi-resolution analysis," *IEEE Trans. Power Delivery*, vol. 16, pp. 662–670, Nov. 2001.

[14] A. M. Gaouda, S. H. Kanoun, and M. M. A. Salama, "On-line disturbance classification using nearest neighbor rule," *J. Elect. Power Syst. Res.*, vol. 57, pp. 1–8, Jan. 2001.

[15] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, pp. 674–693, July 1989.

[16] D. C. Robertson, O. I. Camps, J. S. Mayer, and W. B. Gish, "Wavelets and electromagnetic power system transient," *IEEE Trans. Power Delivery*, vol. 11, pp. 1050–1058, Apr. 1996.

[17] E. B. Hunt, J. Marin, and P. J. Stone, *Experiments in Induction*. New York: Academic, 1966.

T. K. Abdel-Galil (M'04) received the B.Sc. and M.Sc. degrees in electrical engineering from Ain Shams University, Cairo, Egypt, in 1992 and 1998, respectively, and the Ph.D. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 2003.

Currently, he is a Postdoctoral Fellow at the University of Waterloo. His interests include power distribution systems, power quality analysis, application of digital signal processing, and machine learning algorithms in power systems.

M. Kamel (SM'96) received the B.Sc. (Hons.) degree in electrical engineering from the University of Alexandria, Egypt, in 1970, the M.Sc. degree in computation from McMaster University, Hamilton, ON, Canada, in 1974, and the Ph.D. degree in computer science from the University of Toronto, ON, Canada, in 1981.

Since 1985, he has been with the Department of Systems Design Engineering, University of Waterloo. He is currently a Canada Research Chair Professor in Cooperative Intelligent Systems and Director of the Pattern Analysis and Machine Intelligence Laboratory. In 1989/1990, he was a visiting faculty at the University of Toronto and Purdue University, Lafayette, IN. In 1996/1997, he was a visiting professor at the ACE NASA Center, University of New Mexico, Albuquerque, and the University of Porto, Portugal. He is a member of the board of directors and co-founder of Virtek Vision International Inc. His research interests include computational intelligence, pattern recognition, and distributed and multiagent systems. He has authored and co-authored over 200 papers in journals and conference proceedings, two patents, and numerous technical and industrial project reports. Over 45 graduate students have completed their degrees under his supervision.

Dr. Kamel is the Editor-in-Chief of the *International Journal of Robotics and Automation*, Associate Editor of the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS–PART A: SYSTEMS AND HUMANS, *Intelligent Automation and Soft Computing*, *Pattern Recognition Letters*, and the *International Journal of Image and Graphics*.

A. M. Youssef received the B.Sc. and M.Sc. degrees from Cairo University, Cairo, Egypt, in 1990 and 1993, respectively, and the Ph.D. degree from Queens University, Kingston, ON, Canada, in 1997.

He worked for Nortel Networks from 1997 until mid-1999, and then joined the Combinatorics and Optimization Department (Center for Applied Cryptographic Research) at the University of Waterloo, ON, Canada, until the end of 2001. He is currently an Assistant Professor with the Department of Electronics and Communications Engineering, Cairo University.

E. F. El-Saadany (M'98) received the B.Sc. and M.Sc. degrees from Ain Shams University, Cairo, Egypt, and the Ph.D. degree from the University of Waterloo, Waterloo, ON, Canada, all in electrical engineering, in 1986, 1990, and 1998, respectively.

He is an Assistant Professor in the Department of Electrical and Computer Engineering, University of Waterloo. His interests include the operation and control of distribution systems, power electronics application in power systems, power quality analysis, and mechatronics.

M. M. A. Salama (F'02) received the B.Sc. and M.Sc. degrees in electrical engineering from Cairo University, Giza, Egypt, in 1971 and 1973, respectively, and the Ph.D. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 1977.

Currently, he is a Professor in the Electrical and Computer Engineering Department, University of Waterloo. His interests include the operation and control of electric distribution systems, power-quality analysis, and insulation systems. He has consulted widely with government agencies and the electrical authority.