International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:2, No:10, 2008

Artificial Intelligence Techniques Applications for Power Disturbances Classification

K.Manimala, Dr.K.Selvi and R.Ahila

***Abstract*—**Artificial Intelligence (AI) methods are increasingly being used for problem solving. This paper concerns using AI-type learning machines for power quality problem, which is a problem of general interest to power system to provide quality power to all appliances. Electrical power of good quality is essential for proper operation of electronic equipments such as computers and PLCs. Malfunction of such equipment may lead to loss of production or disruption of critical services resulting in huge financial and other losses. It is therefore necessary that critical loads be supplied with electricity of acceptable quality. Recognition of the presence of any disturbance and classifying any existing disturbance into a particular type is the first step in combating the problem. In this work two classes of AI methods for Power quality data mining are studied: Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs). We show that SVMs are superior to ANNs in two critical respects: SVMs train and run an order of magnitude faster; and SVMs give higher classification accuracy.

***Keywords*—**back propagation network, power quality, probabilistic neural network, radial basis function support vector machine

I. Introduction

THE Quality of Electrical Power delivered by utility companies has been a matter of growing concern in recent times. Modern equipment like Computers, Electronic Sensors, Programmable Logic Controllers and other sensitive industrial equipment are at risk of malfunction, instability or reduction of their life-spans in the presence of power quality deviations. In order to improve the quality of power, electric utilities continuously monitor power delivered at customer sites. Thus the raw transient power signal disturbance data could be analyzed using data mining techniques to provide knowledge about the captured waveforms[1-3]. The data mining technique proposed in this paper for knowledge discovery in power quality data consists of two stages. Feature extraction using wavelet transform and classification using data mining techniques like SVM and ANN.

An artificial neural network (ANN) is an abstract computational model of the human brain. Similar to the brain ANN is composed of artificial neurons, regarded as the processing units, and the massive interconnection among them. It has the unique ability to learn from examples and to generalize, i.e., to produce reasonable outputs for new inputs not encountered during a learning process. The distinct features of ANN are as following: learning from examples, generalization ability, non-linearity of processing units, adaptability, massive parallel interconnection among processing units and fault tolerance. ANNs had attracted a great deal of attention because of their inherent pattern recognition capabilities and their ability to handle noisy data. Support Vector machine (SVM) is a two layer neural network employing hidden layer of radial units and one output neuron. The procedure of creating this network and learning its parameters is organized in the way in which we deal only with kernel functions instead of direct processing of hidden unit signals.

This paper will summarize and compare these two networks: ANN and SVM. The comparison will be done with respect to the complexity of the structure as well as the accuracy of results for the solution of Power disturbances classification problem. Special emphasis will be given to the generalization ability of the learned structures acquired in different learning processes. Test results taken from three different NN's training methods and three SVM methods indicate the higher ability of the SVM for five types of disturbances considered in this paper for classification.

K.Manimala is with Dr.Sivanthi Aditanar College of Engineering, Tiruchendur,India (phone: 9942172538; e-mail: s_monimala@yahoo.com).

Dr.K.Selvi, is with Thiagarajar College of Engineering, Madurai (e-mail: kseee@tce.edu).

R.Ahila is with Dr.Sivanthi Aditanar College of Engineering, Tiruchendur,India (e-mail: r.ahila@yahoo.co.in).

II. Feature Extraction

A. Wavelet Analysis

Wavelet analysis is a technique for carving up function or data into multiple components corresponding to different frequency bands. This allows one to study each component separately. Wavelet analysis is a form of time-frequency technique as it evaluates signal simultaneously in the time and frequency domains [4]. It uses wavelets, "small waves," which are functions with limited energy and zero average,

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

The functions are typically normalized, $\|\psi\|=1$ and centered in the neighborhood of $t = 0$. It plays the same role as the sine and cosine functions in the Fourier analysis. In wavelet transform, a specific wavelet is first selected as the basis function commonly referred to as the mother wavelet. Dilated (stretched) and translated (shifted in time) versions of the mother wavelet are then generated [2]. Dilation is denoted

by the scale parameter a while translation is adjusted through b

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

where a is a positive real number and b is a real number. The wavelet transform of a signal $f(t)$ at a scale a and time translation b is the dot product of the signal $f(t)$ and the particular version of the mother wavelet, $\psi_{a,b}(t)$. It is computed by circular convolution of the signal with the wavelet function

$$w\{f(a,b)\} = \langle f, \psi_{a,b}\rangle = \int_{-\infty}^{+\infty} f(t)\cdot\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \quad (3)$$

A contracted version of the mother wavelet would correspond to high frequency and is typically used in temporal analysis of signals, while a dilated version corresponds to low frequency and is used for frequency analysis. With wavelet functions, only information of scale $a < 1$ corresponding to high frequencies is obtained. In order to obtain the low-frequency information necessary for full representation of the original signal $f(t)$, it is necessary to determine the wavelet coefficients for scale $a > 1$. This is achieved by introducing a scaling function $\varphi(t)$ which is an aggregation of the mother wavelets $\psi(t)$ at scales greater than 1.The scaling function can also be scaled and translated as the wavelet function,

$$\phi_{a,b}(t) = \frac{1}{\sqrt{a}}\phi\left(\frac{t-a}{b}\right) \quad (4)$$

With scaling function, the low-frequency approximation of $f(t)$ at a scale a is the dot product of the signal and the particular scaling function [3], and can be computed by circular convolution given by (5).

$$L\{f(a,b)\} = \langle f, \psi_{a,b}\rangle = \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{a}}\phi\left(\frac{t-a}{b}\right)dt \quad (5)$$

Implementation of these two transforms (3) and (5) can be done smoothly in continuous wavelet transform (CWT) or discretely in discrete wavelet transform (DWT).

Successful application of wavelet transform depends heavily on the mother wavelet. The most appropriate one to use is generally the one that resembles the form of the signal. Among the several wavelet functions that were mentioned in the literature, the Daubechies families of wavelets are the most widely used. Among the different dbN (N-order) wavelets, db4 is the most widely adopted wavelet in power quality applications. It has sufficient number of vanishing moments to bring out the transients while maintaining a relatively short support to avoid having too many high-valued coefficients. In our work db4 mother wavelet is chosen.

B. Multiresolution Analysis

One important trait of wavelet transform is that its nonuniform time and frequency spreads across the frequency plane. They vary with scale a but in the opposite manner, with the time spread being directly proportional to a while frequency spread to $1/a$. The resolutions of DWT vary across the planes. At low frequency when the variation is slow, the time resolution is coarse while the frequency resolution is fine. This enables accurate tracking of the frequency while allowing sufficient time for the slow variation to transpire before analysis. On the contrary, in the high-frequency range, it is important to pinpoint when the fast changes occur. The time resolution is therefore small, but the frequency resolution is compromised. This adjustment of the resolutions is inherent in wavelet transform as the wavelet basis is stretched or compressed during the transform [4]. This ability to expand function or signal with different resolutions is termed as multiresolution analysis, which forms the cornerstone of many wavelet applications. In this sense, a recorder-digitized function $a_0(n)$, which is a sampled signal of $f(t)$, is decomposed into its smoothed version $a_1(n)$ (containing low-frequency components), and detailed version (containing higher-frequency components) $d_1(n)$, using filters $h(n)$ and $g(n)$, respectively. This is first-scale decomposition. The next higher scale decomposition is now based on signal $a_1(n)$ and so on Fig. 1.

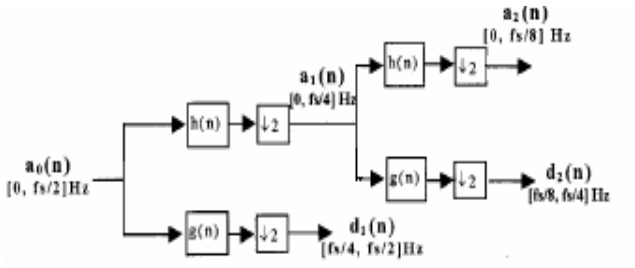

Fig. 1 Multiresolution signal decomposition

C. Classification Of Various Powerquality Events

The Daubauchie "db4" wavelet function was adopted to perform the DWT. The different levels of wavelet coefficient over the scales can be interpreted as uneven distribution of energy across the multiple frequency bands [2,5]. If the selected wavelet and scaling functions form an orthonormal (independent and normalized) set of basis, then the Parseval theorem relates the energy of the signal to the values of the coefficients. This means that the norm or energy of the signal can be separated according to the following multiresolution expansion

$$\int |f(t)|^2 dt = \sum_k |A_{j0}(k)|^2 + \sum_{j\le j0}\sum_k |D_j(k)|^2 \quad (6)$$

These squared wavelet coefficients were shown to be useful features for identifying power quality events. The energy distribution pattern in the wavelet domain can be computed as sums of the squared coefficients as in (6).

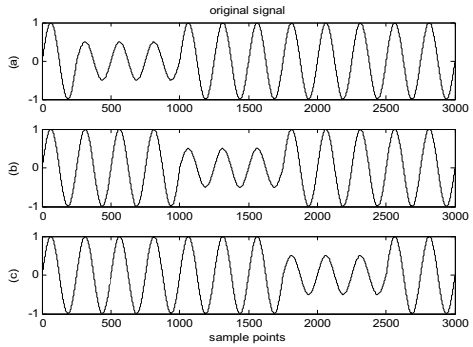

Fig. 2 Sag at different instants

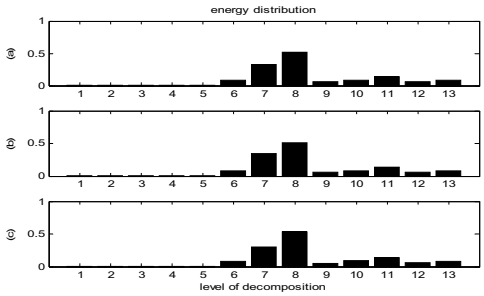

Fig. 3 Energy distribution diagram for sag at different instants

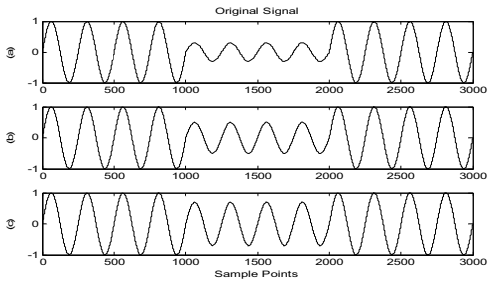

Fig. 4 Sag with different amplitudes

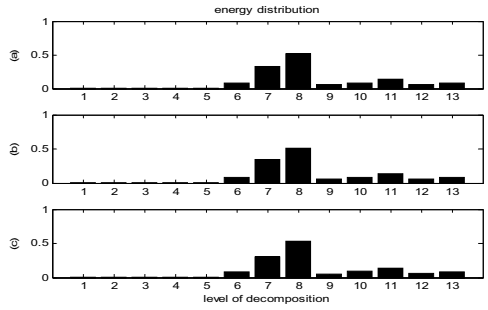

Fig. 5 Energy distribution diagram for sag with different amplitudes

The properties of energy disturbance features are

• The energy distribution remains unaffected by the time of disturbance occurrence.

• The outline of energy distribution remains the same despite variations in the amplitude of the same disturbance type.

• The low-level energy distribution will show obvious variations when the distorted signal contains high-frequency elements. On the contrary, the high- level energy distribution will show obvious variations when the distorted signal contains low-frequency elements.

Figures 2-5 shows that the energy distribution pattern remains the same for the event sag despite occurring at different instants and with different amplitudes.

D. Duration of Transients

In general, when a transient disturbance occurs, the stable power signal will generate a discontinuous state at the start and end points of the disturbance duration. Employing the DWT technique to analyze the distorted signal through one level decomposition of the MRA will cause the wavelet coefficients at the start and end points of the disturbance to generate severe variation [2]. Therefore, we can easily obtain the start time and end time of the disturbance from the variations in absolute wavelet coefficients and calculate the disturbance duration. Fig 6 and 7 shows the plot of level 1, level 2 and level 3 DWT coefficients for the disturbance swell and pure sine wave. The coefficients show variation for the disturbance swell from which the disturbance duration can be determined but there is no variation for pure sine wave because the signal is smooth.

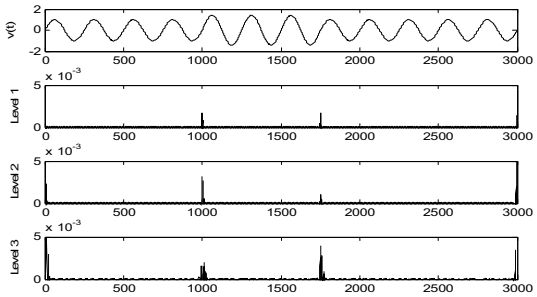

Fig. 6 Plot of DWT coefficients for swell

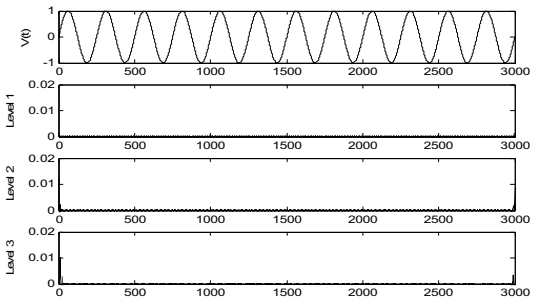

Fig. 7 Plot of DWT coefficients for pure sine

III. Disturbance Classification Using ANN

A. Back Propagation Network

A two layer feedforward neural network is used for learning the feature vectors. From experience, 27 neurons in the hidden

layer gave the best results. The tan-sigmoid function is used as the transfer function in the hidden-layer neurons. The output layer is comprised of one neuron to identify the disturbance class. The transfer function used in the output layer is purelin because the output should indicate the classes from 1 to 5. Levinberg-Marquardt training function (TrainLM) is used to speed up the training process.

The network is trained with a backpropagation algorithm. The error measure is given as

$$E = \sum_{j=1}^{Q} \sum_{k=1}^{M} \left(d_{kj} - x_{kj} \right)^2 \quad (7)$$

where Q is the number of training samples, M the number of output neurons, and d_{kj} is the target value. The weights are updated as

$$\Delta w_{k+1} = -\eta \nabla wE + \alpha \Delta w_k \quad (8)$$

where ∇wE the gradient of E with respect to w, and α is is the momentum constant and η is the learning rate.

The gradient-descent algorithm was implemented in batch mode. The performance of a gradient-descent algorithm is very dependent on the learning rate. If the learning rate were too large, the training would oscillate back and forth. If the learning rate were too small, it would take a long time to reach convergence [11]. To overcome this problem an adaptive learning rate that attempts to keep the step size as large as possible without causing oscillation is used. The learning rate is made responsive to the complexity of the local error surface. The learning rate used in this work is 0.1.

B. Probabilistic Neural Network

A PNN is a variant of a radial basis function neural network. It consists of a radial basis layer and a competitive layer[11]. The transfer function in the hidden layer can be written as

$$h(x) = \phi(\|x - x_c\|) \quad (9)$$

where φ is the radial basis function, x is the input of the neuron, x_c is the center of the neuron, and $\|x - x_c\|$ is usually taken to be the Euclidean distance. In this particular case, a Gaussian function was used as the radial basis function. It is written as

$$h(x) = e^{-(\|x - x_c\|^2 / 2\sigma^2)} \quad (10)$$

where σ is the spread of the Gaussian. If the spread of the Gaussian function is too small, The Gaussian is sharply peaked and the neurons are not able to cover the input space well. This will result in poor generalization. However, if the spread is too large, there would be large overlaps in the input space and all neurons will give large values for all inputs. The network was trained for spread value 0.01.

The input layer has m units to which the m dimensional input vector is applied. The first hidden layer has one pattern unit for each exemplar. The second hidden layer contains one summation unit for each class. The output layer is the decision layer which contains one neuron to specify the class. The PNN [1,6,7] can function as a classifier and has the advantage of being a fast learning process as it requires only a single-pass network training stage without any iteration for adjusting weights. Further, it can itself adapt to architectural changes. As the structure of PNN is simple and learning efficiency is very fast it is suitable for signal classification problems. Hence PNN is considered as the best neural network for power disturbances classification. The input layer contains 14 neurons (13 energy features from DWT, time of duration), the first hidden layer contains 300 neurons (300 training exemplars) and the second hidden layer contains 5 neurons (5 classes) and the final output layer contains one neuron.

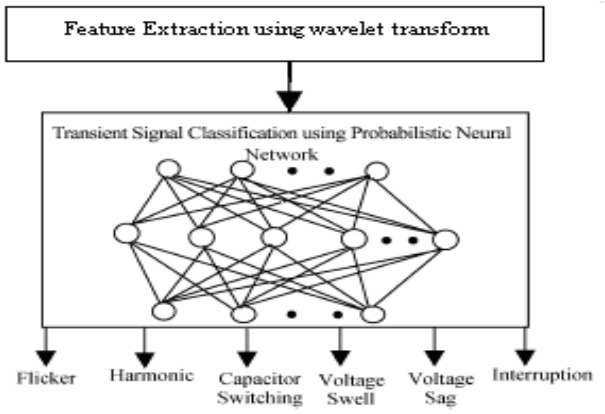

Fig. 8 PNN Classifier

C. RBF Networks

RBF networks can be used as universal function approximations. This consists of a network with a single hidden layer and a structure similar to back propagation networks [8,9]. Each hidden layer unit has a centroid c_i and smoothing factor σ_i. These neurons compute the distance between the input x_i and the centroid c_i rather than the vector product of the weights and inputs. The outputs are nonlinear, radial symmetric functions of the distance. Thus the output is the strongest when x_i is the closest to the value c_i. RBF networks apply real mapping functions f_m which have the general form

$$f_m(x) = \sum_{i=1}^{M} w_i K[(x_i - c_i)/\sigma_i] \quad (11)$$

The function K is a radial symmetric kernel function computed by M kernel units. The Gaussian exponential function is commonly used in RBF networks

$$f(x) = \beta \exp\left(- \sum_i [(x_i - c_i)/\sigma_i]^2 \right) \quad (12)$$

The centroid c_i and constants β and σ_i have to be chosen accordingly to the training data set. General Gaussian activation functions are superior to sigmoid functions in estimating a broad class of functions. The network has only one hidden layer and the fact that the hidden nodes receive input directly from the input layer without having to calculate the weighted sums, makes it much faster to train than a back propagation network of comparable size.

RBF networks can be used as universal function approximations. This consists of a network with a single hidden layer and a structure similar to back propagation networks. Each hidden layer unit has a centroid c_i and smoothing factor σ_i. These neurons compute the distance between the input x_i and the centroid c_i rather than the vector product of the weights and inputs. The outputs are nonlinear, radial symmetric functions of the distance. Thus the output is the strongest when x_i is the closest to the value c_i. The network has only one hidden layer and the fact that the hidden nodes receive input directly from the input layer without having to calculate the weighted sums, makes it much faster to train than a back propagation network of comparable size. Smoothing factor is chosen as 0.03.Radial basis function with exact fit is used as the training network.

IV. Disturbances Classification Using SVM

In recent years, a new approach to construct and train neural networks (NNs) was developed, which is free of many disadvantages. The new networks are called SVMs. Support vector machines (SVMs) [8] were originally designed for binary classification. How to effectively extend it for multiclass classification is still an ongoing research issue. Currently there are two types of approaches for multiclass SVM. One is by constructing and combining several binary classifiers while the other is by directly considering all data in one optimization formulation. In general, it is computationally more expensive to solve a multi-class problem than a binary problem with the same number of data. This work is devoted to the second approach, i.e. it solves a multi-class problem by decomposing it to several binary problems in a hierarchical way. The three methods considered in this paper are "one-against-all" and "one-against-one" and "dendogram based SVM".

A. One against all method

The earliest used implementation for SVM multiclass classification is probably the one-against-all method (for example, [8],[9]). It constructs k SVM models where k is the number of classes. The ith SVM is trained with all of the examples in the ith class with positive labels, and all other examples with negative labels. Thus given l training data, $(x_1,y_1),\ldots.(x_l,y_l)$, where $x_i \varepsilon R_n$, $i=1,\ldots.l$ and $y_k \varepsilon \{1,\ldots,k\}$ is the class of x_i, the ith SVM solves the following problem:

$$\min_{w^i,b^i,\xi^i} \frac{1}{2}(w^i)^T w^i + C\sum_{j=1}^{l} \xi_j^i (w^i)^T$$
$$(w^i)^T \phi(x_j) + b^i \geq 1 - \xi_j^i, \quad \text{if } y_j = i$$
$$(w^i)^T \phi(x_j) + b^i \leq -1 + \xi_j^i, \text{ if } y_j \neq i$$
$$\xi_j^i \geq 0, j = 1,....,l \qquad (13)$$

where the training data x_i are mapped to a higher dimensional space by the function φ and C is the penalty parameter.

$$\text{Class of } x \equiv \arg\max_{i=1,..k}((w^i)^T \phi(x) + b^i) \qquad (14)$$

B. One against one method

This method constructs $k(k\text{-}1)/2$ classifiers where each one is trained on data from the ith and jth classes, we solve the following binary classification problem[10]:

$$\min_{w^{ij}b^{ij}\xi^{ij}} \frac{1}{2}(w^{ij})^T w^{ij} + C\sum_i \xi_t^{ij} (w^{ij})^T$$
$$(w^{ij})^T \phi(x_t) + b^{ij} \geq 1 - \xi_t^{ij}, \quad \text{if } y_t = i$$
$$(w^{ij})^T \phi(x_t) + b^{ij} \leq -1 + \xi_t^{ij}, \text{ if } y_t = j$$
$$\xi_t^{ij} \geq 0 \qquad (15)$$

There are different methods for doing the future testing after all $k(k\text{-}1)/2$ classifiers are constructed. Since we have considered 5 classes of disturbances, the total number of SVMs is 10.

C. Dendogram based SVM (DSVM)

The proposed DSVM takes advantage of both the efficient computation of the ascendant hierarchical clustering of classes and the high classification accuracy of SVM for binary classification. Although DSVM needs (N-1) SVMs for N class problem in the training phase, for the testing phase DSVM requires an optimal set of SVMs selected in a descendant way from the root of the taxonomy through the selected class among the "leaf" nodes. The DSVM method consists of two major steps:(1) computing a clustering of the known classes and (2) associating a SVM at each node of the taxonomy obtained by (1).

The first step of DSVM method consists of calculating N gravity centers for the N known classes. Then AHC clustering is applied over these N centers. Dendogram is constructed through the AHC method to classify PQ disturbances. The basic thought is as follows: firstly the PQ disturbance set needing to be classified is divided into two subsets according to the similarity of the chosen feature vectors, and then the two subsets are divided into two subsets separately again according to the same principle. The division will continue until the classification task is finished. The multi-class SVM classification tree of PQ disturbances is shown in Figure 9. It can be seen that there are 4 SVMs in the multi-class SVM application tree, and each SVM chooses different feature vector to implement binary classification

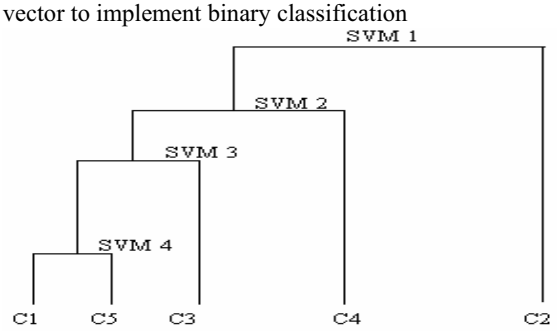

Fig. 9 Classification way by DSVM

C1-sag, C2-swell, C3-Interruption,
C4-Harmonics , C5- Flicker

Figure 9 shows an example of a taxonomy done by AHC [16] algorithm over the *N* classes. In the second step, each SVM is associated to a node and trained with the elements of the two subsets of this node. For example, in Figure10 which illustrates clustering of 5 classes SVM1 is trained by considering elements of {C2*}* as positives and elements of {C1*,C5,C3,C4}* as negatives; SVM2 is trained by considering elements of *{C4}* as positives and elements of *{C1,C3,C5}* as negatives. SVM4 is trained by considering elements of *{C1}* as positives and elements of *{C5}* as negatives.

In final, we will train (*N* - 1) SVM for *N*-class problem. The advantage of training in DSVM is to a priori separate the classes in a hierarchical way. That facilitates the class separation for the SVM. In fact, SVM1 found easily boundary separation between *{C2}* and *{C1,C3,C4,C5}*. The level of difficulty for boundary separation increases from the root through the leaves. The idea of DSVM is that it is preferable to solve many small problems in a hierarchical way than to solve a complex great problem. For classifying a pattern query, DSVM presents it to the "root" SVM which provides an output for right or left on the taxonomy. The procedure is repeated for each selected node in the "way" of the classification (Figure 3) until arriving to a leaf which finally represents the associated class for our pattern query.

V. Application and Results

A. Signal Modeling

Signal modeling by parametric equations for classifiers tests was advantageous in some aspects. It was possible to change testing and training signal parameters in a wide range and in a controlled manner. Signals simulated that way were very close to reality. On the other hand, different signals belonging to the same class gave the opportunity to estimate the generalization ability of classifiers based on Neural Networks. Signals belonging to six main groups of disturbances [9], were simulated. The classes and respective parametric equations for simulation of signals are summarized in Table1. As a special group, the nondisturbed signals were chosen (pure sinusoid).The ranges of signals parameter variation are shown in Table 2.The variation range corresponds to values measured in real power systems. The parameters α_{ss} and α_{sw} correspond to the depth of sag and swell respectively. The step function 1(t) is used to determine sag and swell duration. Flicker is characterized by its frequency β_{fw} and α_f amplitude.

In order to obtain representative signals for the most common power quality disturbances to serve the purpose of training, as well as the testing of the PNN classifier, power quality disturbance signals are simulated using Matlab. Six categories of disturbances are simulated, namely, undisturbed sinusoid, sudden swell, sudden sag, interruption, harmonics, and voltage flicker. The disturbances are based on ten cycles of voltage waveform. These waveforms are generated at a sampling rate of 256 samples / cycle for a total of 3001 points.

TABLE I PARAMETRIC EQUATIONS FOR SIMULATION OF DISTURBED SIGNALS

Event	Equation
Pure Sinusoid	$v(t) = \sin(wt)$
Sag	$v(t) = 1 - \alpha_{ss}((1(t-t_1) - 1(t-t_2)))\sin(wt)$
Swell	$v(t) = 1 + \alpha_{sw}((1(t-t_1) - 1(t-t_2)))\sin(wt)$
Interruption	$v(t) = 1 - ((1(t-t_1) - 1(t-t_2)))\sin(wt)$
Harmonics	$v(t) = \begin{pmatrix} \alpha_{k1}\sin(wt) + \alpha_{k3}\sin(3wt) + \\ \alpha_{k5}\sin(5wt) + \alpha_{k7}\sin(7wt) \end{pmatrix}$
Flicker	$v(t) = (1 + \alpha_f \sin(\beta_f wt))\sin(wt)$

TABLE II PARAMETERS VARIATION IN SIMULATED SIGNALS

Event	Parameters Variation
Pure Sinusoid	*Amplitude* : 1 *Frequency* : 50 *Hz*
Sag	*Duration* : $(t_2 - t_1) = (0 - 9)T$ *Amplitude* : $\alpha_{ss} = 0.3 - 0.8$
Swell	*Duration* : $(t_2 - t_1) = (0 - 8)T$ *Amplitude* : $\alpha_{ss} = 0.3 - 0.7$
Interruption	____
Harmonics	*Order* : 3,5,7 *Amplitude* : $0 - 0.9$
Flicker	*Frequency* : $(5 - 10)Hz$ $\alpha_f = 0.1 - 0.2$

The described experiments were implemented using the MATLAB7 environment, on a Pentium IV, 2.88 GHz, with 256 Mb of memory. The parameters used to vary the classes of events are the depth, the angle, the starting time and the duration of the events, which are defined below.

The depth of the event is defined as the change in the amplitude of a signal. The angle represents the phase shift at which the signal is captured. The starting time is the time at which the event starts. The duration is the time period of the event. In this study the above described parameters were varied according to the IEEE recommended practice in [12].

The angle of the signal is varied from 0% to 100% of the entire period (which is a realistic assumption since the captured waveforms in a practical monitoring system could have a phase shift that may vary from 0 to 2Π). In addition, the starting time of the sag is varied from 0% to 80% of the total length of the signal. Moreover, the duration of the sag is varied from 5% to 100% of the total length of the waveform.

For the interruption and the swell events, the four parameters (depth, angle, starting time and duration) were also varied as described above. The depth of the swell is increased from 10% to 90% of the magnitude of the pure sine waveform. For the harmonics events, 2nd, 3rd, 5th, 7th, 9th, and 11^{th} harmonics are used to randomly contaminate the ideal waveforms. During the generation of such events, the total harmonic distortion (THD) of the waveform was kept greater than 5%, as suggested in [12].

In the case of the flicker events, the amplitude of the simulated signals was changed periodically to introduce the effect of a flicker. To achieve this, the magnitude of the target waveform was varied as a function of another sine wave. This results an oscillation in the amplitude of the target waveform, which varied randomly from 50% to 70% of the fundamental frequency. In order to obtain representative signals for the most common power quality disturbances to serve the purpose of training, as well as the testing of the PNN classifier, power quality disturbance signals are simulated using Matlab. Five categories of disturbances are simulated namely, sudden swell, sudden sag, interruption, harmonics, and voltage flicker. Undisturbed pure sine wave is considered as a special case. The disturbances are based on twelve cycles of voltage waveform. These waveforms are generated at a sampling rate of 256 samples / cycle for a total of 3001 points.

B. Result of the Proposed Classification Method

Based on the feature extraction by the Wavelet-Transform method, we will perform a 13-level decomposition of each discrete distorted signal to obtain the detailed version coefficients. The features extracted from wavelet transform were applied to the ANN Classifiers and SVMs for recognizing and classifying the distorted signals. The 10-Fold Cross Validation Evaluation Results of the ANN and SVM based Classifiers for the five data sets is shown in Table 3&5 respectively.

Cross-validation, is the practice of partitioning a sample of data into subsets such that the analysis is initially performed on a single subset, while the other subset(s) are retained for subsequent use in confirming and validating the initial analysis.The test result shows that the SVM classifier attains better recognition rates when compared with the ANN classifier.

TABLE III 10-FOLD CROSS VALIDATION EVALUATION RESULT OF THE ANN CLASSIFIER

Class	BPN	RBF	PNN
Sag	90	80	90
Swell	70	80	100
Interruption	100	80	80
Harmonics	100	100	100
Flicker	100	100	100
Overall	**92**	**88**	**94**

TABLE IV PERFORMANCE OF ANN

Classifier	Training Time (sec)	Testing Time (sec)	Accuracy (%)
BPN	120	5	92
RBF	5	1	88
PNN	0.02	0.005	94

TABLE V 10-FOLD CROSS VALIDATION EVALUATION RESULT OF THE SVM CLASSIFIER

Class	One Against All	One Against One	DSVM
Sag	90	90	90
Swell	100	100	100
Interruption	70	90	90
Harmonics	100	100	100
Flicker	100	100	100
Overall	**92**	**96**	**96**

TABLE VI PERFORMANCE OF SVM

Classifier	Training Time(sec)	Testing Time(sec)	Accuracy (%)
One Against One	0.7813	0.6563	96
One Against All	0.7969	0.2031	92
Dendogram SVM	0.6094	0.1875	96

VI. CONCLUSION AND FUTURE WORK

Artificial Neural Networks (ANN) and Support Vector Machines (SVM) are the two most popular machine learning techniques now used in classification. The numerical experiments performed for both: ANN and SVM networks have confirmed that both solutions are very well suited for classification. ANN relies mostly on heuristics whereas SVM is mathematically well founded. Out of the three ANN techniques considered in this paper, the PNN performs well in terms of accuracy and time of computation. But the major drawback of PNN is that all its training vectors must be stored requiring large amount of memory. In the case of SVM, the dendogram based SVM achieves 96% accuracy and outperform the other two methods in terms of computation time. The general perception for this power disturbances classification problem is that SVM may outperform ANN in terms of classification accuracy. The observed differences in performance are in most cases negligible. However the main difference is in the complexity of the neural networks. Because the distorted signals in this study were generated by simulation, employing real distorted signals measured by the digital recorder to improve the proposed method for more number of disturbances is one of our future works.

REFERENCES

[1] Mishra S, C. Bhende N, Panigrahi B. K., "Detection and Classification of Power Quality Disturbances Using S-Transform and Probabilistic Neural Network", IEEE Transactions On Power Delivery, January 2008, Vol. 23,pp.280-287.
[2] Gaing Zwe-Lee, "Wavelet-Based Neural Network for Power Disturbance Recognition and Classification", IEEE transactions on Power delivery,October 2004,vol. 19, pp1560-1567.
[3] Janik Przemyslaw and Lobos Tadeusz, "Automated Classification of Power-Quality Disturbances Using SVM and RBF Networks", IEEE Transactions On Power Delivery, July 2006,Vol. 21, Pp.1663-1669.
[4] S. Chen and H. Y. Zhu, "Wavelet Transform for Processing Power Quality Disturbances", EURASIP Journal on Advances in Signal Processing Volume 2007, Article ID 47695, 20 pages.
[5] Borras Dolores, Castilla M, Moreno Narciso and Montano J.C., "Wavelet and Neural Structure: A New Tool for Diagnostic of Power System Disturbances", IEEE Transactions on Power Delivery, January 2001,Volume 37, pp. 184-190.
[6] Specht, D. F., "Probabilistic neural networks", International Neural Network Society, Neural Networks, Vol3, 1990, pp. 109-1 18.
[7] Z.Chen,Senior Member , IEEE, and P.Urwin "Power Quality Detection and Classification Using Digital Filters" IEEE Porto Power Tech Conference 2001.
[8] Peter G. V. Axelberg, Irene Yu-Hua Gu, *Senior Member,*,"Support Vector Machine for Classification of Voltage Disturbances", *IEEE*, and Math H. J. Bollen, *Fellow, IEEE,* IEEE Transactions On Power Delivery, Vol. 22, No. 3, July 2007
[9] Chih-Wei Hsu and Chih-Jen Lin, A Comparison of Methods for Multiclass Support Vector Machines, IEEE Transactions On Neural Networks, Vol. 13, NO. 2, March 2002 pp.415-425.
[10] Khalid Benabdeslem and Youn`es Bennani, "Dendogram- based SVM for Multi-Class Classification", Journal of Computing and Information Technology - CIT 14, 2006, 283–289.
[11] I.W.C.Lee, Member,IEEE, and P.K. Dash, Senior Member, IEEE "S-Transform-Based Intelligent System for Classification of Power Quality isturbance Signals", IEEE Trans on Industrial Electronics,Vol 50,No.4, Aug 2003.
[12] IEEE Recommended Practice for Monitoring Electric Power Quality, IEEE/Std. 1159-1995.

Manimala .K earned a bachelor of engineering in Electrical and Electronics from Government College of Engineering, Tirunelveli, India. She earned a Master's degree in Computer science and Engineering from Manonmaniam sundaranar University, India in 2004. She has 12 years teaching experience. She is currently an Assistant Professor at the computer Science and Engineering Department of Dr.Sivanthi Aditanar College of Engineering, Tiruchendur, India. She is pursuing PhD in the area of Data Mining.

Dr.K.Selvi obtained B.E(EEE) with Honours,M.E (Power System) with Distin, from Madurai Kamaraj University in the year 1989 and 1995 respectively.She obtained Ph.D in Electricity Deregulation in June 2005 from Madurai Kamaraj University. She is currently working as Assistant Professor in Department of Electrical Engineering, in Thiagarajar college of Engineering, Madurai.,Tamilnadu, India. She has obtained Young Scientist Fellowship from Dept. of Science and Technology. Her research interests are Electricity deregulation and AI techniques.

Ahila.R received the B.E. degree in Electrical and Electronics engineering in 1995 and the M.E. degree in Computer Science and Engineering in 2004 from Manonmanium Sundaranar University,Tirunelveli, India.She is currently an assistant professor at the computer Science and Engineering Department of Dr.SivanthiAditanar College of Engineering, Tiruchendur,,India.