FAST SVMS FOR POWER QUALITY DATA MINING

K.Manimala[1], Dr.K.Selvi[2], R.Ahila[3]
[1,3] Assistant Professor, CSE, Dr.Sivanthi Aditanar College of Engineering, Tiruchendur, Tamilnadu, India
[2] Assistant Professor, EEE, Thiagarajar College of Engineering, Madurai, Tamilnadu ,India
Email: [1] s_monimala @yahoo.com

Abstract

Recognition of the presence of any disturbance and classifying any existing disturbance into a particular type is the first step in combating the power quality problem. Support Vector Machines (SVMs) have gained wide acceptance because of the high generalization ability for a wide range of classification applications. Although SVMs have shown potential and promising performance in power disturbances classification, they have been limited by speed particularly when the training data set is large. The hyper plane constructed by SVM is dependent on only a portion of the training samples called support vectors that lie close to the decision boundary (hyper plane). Thus, removing any training samples that are not relevant to support vectors might have no effect on building the proper decision function. We propose the use of clustering techniques such as K-mean to find initial clusters that are further altered to identify non-relevant samples in deciding the decision boundary for SVM. This will help to reduce the number of training samples for SVM without degrading the classification result and classification time can be significantly reduced.

Key words: S transform, Support Vector Machine, K-means Clustering, Power Quality

I. INTRODUCTION

Most industries and commercial facilities are affected by power quality problems [1], [2]. Specifically industries requiring ultra-high availability of service (e.g., internet-data and telecommunication switching centers) and precision manufacturing systems (e.g., semiconductor, chemicals and medicinal industries) are very sensitive to power quality problems. The term "power quality" covers several types of problems of electricity supply and power system disturbances such as voltage sags, swells, interruptions, flicker, transient, capacitive switching, harmonics, etc. Prompt and accurate diagnosis of problems will help maintenance personnel to respond to the alarms efficiently. The diagnosis will ensure quality of power and will reduce the risk of interruptions by reducing the time to diagnose and rectify failures. Power quality monitors gather vast amounts of data; however, this raw data directly provides little information on power quality. Intelligent signal processing techniques are applied to extract useful information from the raw data. Once the information is extracted, problems concerning power quality can be identified and precisely categorized by employing detection and classification methods [1]. In particular, when the disturbance type has been classified accurately, the power quality engineers can define the major effects of the disturbance at the load and analyze the source of the disturbances so that an appropriate solution can be formulated.

For the purpose of classification, artificial intelligence based techniques like probabilistic neural networks, and SVM are being used widely [3-6]. ANNs had attracted a great deal of attention because of their inherent pattern recognition capabilities and their ability to handle noisy data. However, ANNs suffer from a number of weaknesses which include difficulty in obtaining a stable solution, the danger of over fitting, falling in local minima and their inherent need of a large numbers of training cycles. The support vector machine (SVM) approach is considered a good candidate because of its high generalization performance without the need to add *a priori* knowledge, even when the dimension of the input space is very high [7].

A serious problem of Support Vector Machine (SVM) is its low classifying speed. The speed depends on the number of support vectors [8]. Power disturbances classification is an online problem and so the computing time should be drastically reduced. Hence there is a need to improve the performance of SVM for power disturbances classification. This paper proposes a new method called the Clustering SVM to reduce the number of support vectors and hence the classification time. The idea behind this method is to reduce the size of the dataset when taking the best effort to keep the class information unchanged from the original dataset. To reach this purpose, we use a clustering method to assign the data of each class to a given number of groups, and then create a new dataset consisting of only the central vectors of each group. The new dataset size is smaller than the original one. Moreover, the size of the new dataset can be controlled arbitrarily by setting the number of groups in the clustering technique.

Fig. 1. illustrates the principle block diagram of the proposed classifier. The features that distinguish the distorted signals are extracted using S transform in the feature extraction block [9,10]. The disturbances are classified by applying data mining technique using Support Vector Machine (SVM) on the relevant features

and the reduced training samples obtained from data selection phase using kmeans clustering.

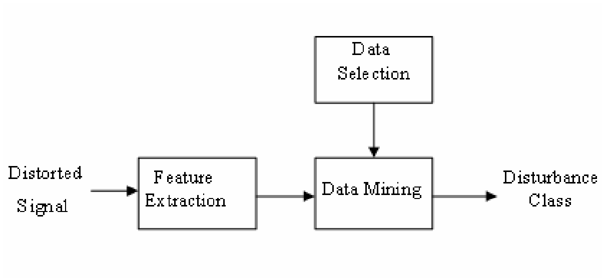

Fig. 1. Diagram of Proposed classifier

II. FEATURE EXTRACTION

A. Discrete S-Transform

Let $p[KT]$,k=0,1,..N-1 denote a discrete time series corresponding to a signal $p(t)$ with a time sampling interval of T. The discrete Fourier transform of the signal can be obtained as follows:

$$P\left[\frac{n}{NT}\right] = \frac{1}{N}\sum_{k=0}^{N-1} p[kT]e^{-(i2\Pi nk/N)} \quad [1]$$

where n=0,1,…N-1 and the inverse discrete Fourier transform is

$$p[kT] = \sum_{n=0}^{N-1} P\left[\frac{n}{NT}\right]e^{i2\Pi nk/N} \quad [2]$$

In the discrete case, the S-Transform is the projection of the vector defined by the time series $p[kT]$, onto a spanning set of vectors [9,10].The spanning vectors are not orthogonal and the elements of the S-Transform are not independent. Each basis vector (of the Fourier transform) is divided into N localized vectors by an element- by- element product with the N shifted Gaussians, such that the sum of these N localized vectors is original basis vector. The S-Transform of a discrete time series $p[kT]$, is given by

$$S\left[\frac{n}{NT}, jT\right] = \sum_{m=o}^{N-1} P\left[\frac{m+n}{NT}\right]G(n,m)e^{i2\Pi mj/N} \quad [3]$$

where $G(n,m) = e^{-(2\Pi^2 m^2/n^2)}$ =Gaussian function and j ,m, n = 0,1,….,N-1.

The following steps are adapted for computing the discrete S-Transform

1) Perform the discrete Fourier transform of the original time series $p[kT]$ with N points and sampling interval T to get $P[m/NT]$ using the FFT routine. This is only done once.
2) Calculate the localizing Gaussian $G(n,m)$ for the required frequency n/NT.
3) Shift the spectrum $P[m/NT]$ to $P[(m+n)/NT]$ for the frequency n/NT
4) Multiply $P[(m+n)/NT]$ by $G[n,m]$ to get B[n/NT, m/NT] (N multiplications)
5) Inverse Fourier transform of B[n/NT,m/NT]m/NT to j to give the row of $S[n/NT,jT]$ corresponding to the frequency n/NT
6) Repeat steps 3, 4 and 5 until all the rows of $S[n/NT,jT]$ corresponding to all discrete frequencies n/NT have been defined

From (3), it is seen that the output from the S-Transform is an N×M matrix called the S-matrix whose rows pertain to frequency and columns thus represents the "local spectrum" for that point in time. Each element of the S-matrix is complex valued. The choice of windowing function is not limited to the Gaussian function; other windowing functions were also implemented successfully. Also, frequency-time contours having the same amplitude spectrum are obtained to detect, and localize power disturbance events. A mesh three-dimensional view of the S-transform output yields frequency-time, amplitude-time, and frequency-amplitude plots.

In all of the plots, the frequency magnitude is normalized with respect to the sampling frequency and is given by f_s. From these results, it is quite obvious that in case of S-transform output, one can detect, localize, and quantify the disturbance completely. The S-transform contours for the disturbances swell and sag are shown in figures 2-5.

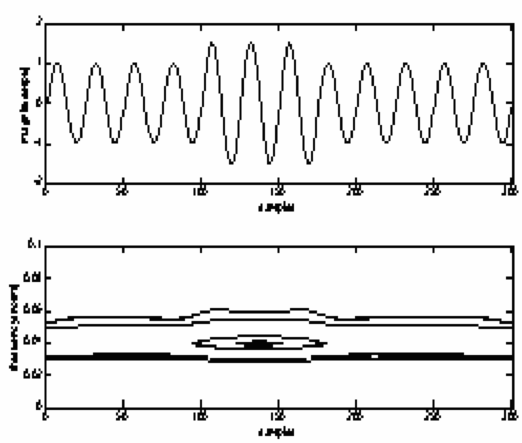

Fig. 2. voltage swell and S-transform contours

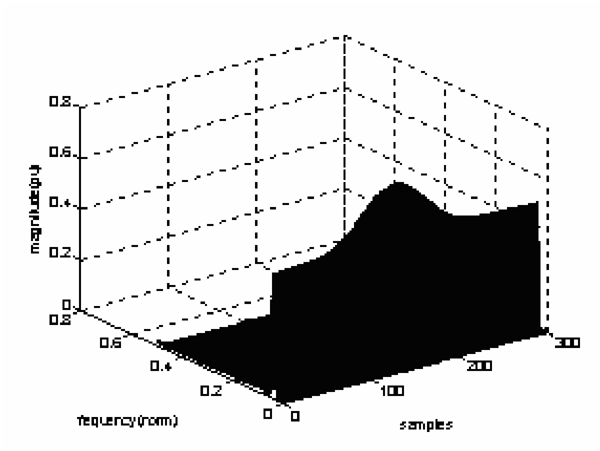

Fig. 3. 3-D S-transform plot of voltage swell

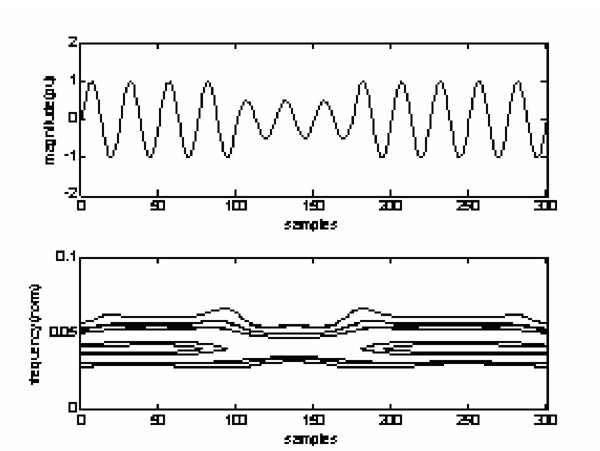

Fig. 4. Voltage sag and S-transform contours

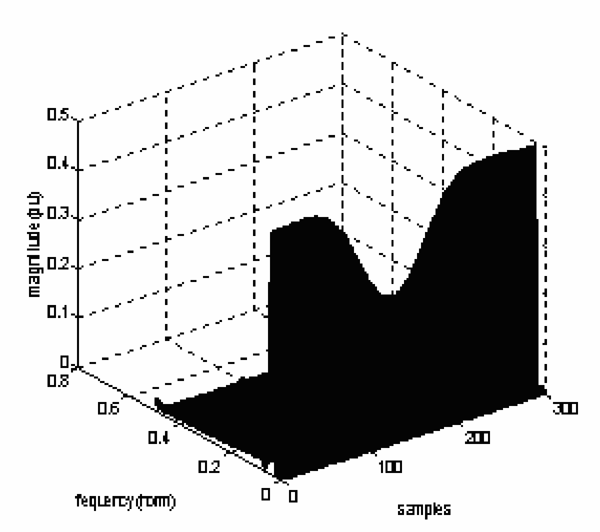

Fig. 5. 3-D S-transform plot of voltage sag

B. Feature Extraction Using S-Transform

The S-transform performs multiresolution analysis on a time varying signal as its window width varies inversely with frequency [9,10]. This gives high time resolution at high frequency and high frequency resolution at low frequency. Since power quality disturbances make the power signal a nonstationary one, the S-Transform can be applied effectively. In this paper, the signals are simulated using MATLAB. The signals are sampled at 25 points per cycle. Five types of power quality disturbances are simulated and the features of all the types of disturbances are extracted from the S-matrix. Further, from the S-matrix important information in terms of magnitude, phase and frequency can be extracted. These are shown in Figs.6,7. The time-frequency contours of the S-transform output shows a decrease or increase in magnitude for voltage sag and swell, and interruption , which provide a better visual classification strategy in comparison to the wavelet transform (similar to time versus rms or peak value of voltage). The S-transform output at different frequency resolutions will be required for classification of high-frequency transients, impulses, notches, etc., since it yields some more parameters for discriminating various types of transient disturbances. It is observed that the standard deviation of second contour is an important parameter to distinguish between transients, impulses, and notches. Feature extraction is done by applying standard statistical techniques onto the S-matrix. Many features such as amplitude, slope (or gradient) of amplitude, time of occurrence, mean, standard deviation and energy of the transformed signal are widely_used for proper classification [10,11]. Since, the aim is to obtain a satisfactory classification accuracy, features based on standard deviation (S.D.) and energy of the transformed signal are extracted as follows.

Feature 1: Standard deviation (S.D.) of the data set comprising of the elements corresponding to maximum magnitude of each column of the S- matrix.

Feature 2: Energy of the data set comprising of the elements corresponding to maximum magnitude of each column of the S- matrix.

Feature 3: Standard deviation (S.D.) of the data set values corresponding to maximum value of each row of the S-matrix.

Feature 4: Standard deviation (S.D.) of the phase contour.

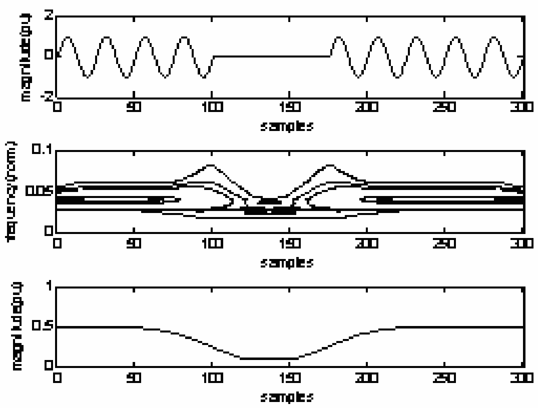

Fig. 6. Voltage interruption

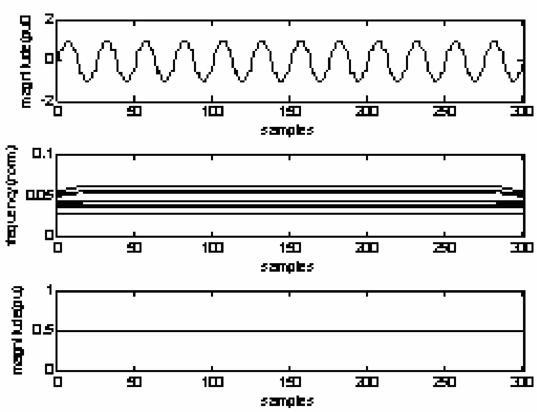

Fig. 7. Voltage harmonics

S-Transform can detect the disturbance correctly in both pure as well as noisy environments. Also S-transform requires only 4 features compared to wavelet transform which requires 13 features for classification.

III. DATA REDUCTION PHASE

A. K Means Clustering

Clustering of data is a method by which large sets of data are grouped into clusters of smaller sets of similar data. A clustering algorithm attempts to find natural groups of components (or data) based on some similarities. To determine cluster membership, most algorithms evaluate the distance between a point and the cluster centroids. The output from a clustering algorithm is basically a statistical description of the cluster centroids with the number of components in each cluster. The centroid of a cluster is a point whose parameter values are the mean of the parameter values of all the points in the clusters. The k-means algorithm used in this work is one of the most non-hierarchical methods used for data clustering.

The K-means algorithm [12] is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early grouping is done. At this point we need to recalculate k new centroids as barycentre of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated; as a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more. Finally, this algorithm aims at minimizing an objective function, in this case a squared error function. The objective function:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{j} - c_j \right\|^2 \qquad [4]$$

where $\left\| x_i^{j} - c_j \right\|^2$ is a chosen distance measure between a data point and the cluster centre c_j, is an indicator of the distance of the n data points from their respective cluster centers.

The general algorithm is composed of the following steps:

(1) Place k points into the space represented by the objects that are being clustered. These points represent initial group centroids.

(2) Assign each object to the group that has the closest centroid.

(3) When all objects have been assigned, recalculate the positions of the K centroids.

(4) Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

B. Clustering SVM

This paper proposes a new method, called the Clustering SVM, to reduce the number of support vectors. The idea behind this method is to reduce the size of the dataset when taking the best effort to keep the class information unchanged from the original dataset. To reach this purpose, we use kmeans clustering method to assign the data of each class to a given number of groups, and then create a new dataset consisting of only the central vectors of each group [8]. The new dataset size is smaller than the original one. As described above, the initial centroid values are chosen randomly from the training data set. Table1 lists the final chosen parameters for each class. It is a small portion of the initial dataset (14% only).The proposed algorithm reduces the training data size from 600 samples to 84 samples.

Table 1. **Samples selected using Clustering**

Class	Initial Samples	Number Of clusters
Sag	150	24
Swell	100	12
Interruption	150	24
Harmonics	100	12
Flicker	100	12

IV. SUPPORT VECTOR MACHINE (SVM)

SVM based classifier is built to minimize the structural misclassification risk, whereas conventional classification techniques often apply minimization of the empirical risk. Therefore, SVM is claimed to lead enhanced generalization properties. Further, application of SVM results in the global solution for a classification problem. Thirdly, SVM based classification is attractive, because its efficiency does not directly depend on the dimension of classified entities. Support vector machines (SVMs) [13] were originally designed for binary classification. How to effectively extend it for multiclass classification is still an ongoing research issue. This work 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 [13]. 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,….l and $y_k \varepsilon \{1,\ldots,k\}$ is the class of x_i, the ith SVM solves the following problem:

$$\min_{w^i,b^i,\xi^i} \frac{1}{2}(w^i)^T w^i + C\sum_{j=1}^{l}\xi_j^i (w^i)^T \tag{5}$$
$$(w^i)^T\phi(x_j)+b^i \geq 1-\xi_j^i, \quad \text{if } y_j = i$$
$$(w^i)^T\phi(x_j)+b^i \leq -1+\xi_j^i, \text{ if } y_j \neq i, \xi_j^i \geq 0, j=1,\ldots l,$$

where the training data x_i are mapped to a higher dimensional space by the function ϕ and C is the penalty parameter. We say x is in the class which has the largest value of the decision function

$$\text{Class of } x \equiv \arg\max_{i=1,..k}((w^i)^T\phi(x)+b^i) \tag{6}$$

B. One against one method

This method [13] constructs $k(k$-1)/2 classifiers where each one is trained on data from the ith and jth classes, we solve the following binary classification problem:

$$\min_{w^{ij}b^{ij}\xi^{ij}} \frac{1}{2}(w^{ij})^T w^{ij} + C\sum_i \xi_t^{ij}(w^{ij})^T \tag{7}$$
$$(w^{ij})^T\phi(x_t)+b^{ij} \geq 1-\xi_t^{ij}, \quad \text{if } y_t = i$$
$$(w^{ij})^T\phi(x_t)+b^{ij} \leq -1+\xi_t^{ij}, \quad \text{if } y_t = j$$
$$\xi_t^{ij} \geq 0$$

There are different methods for doing the future testing after all $k(k$-1)/2 classifiers are constructed. After some tests, the following voting strategy is used to determine the class. If sign $((w^{ij})^T\phi(x)+b^{ij}$ says that x is in the ith class, then the vote for the ith class is added by one. Otherwise, the jth is increased by one. Then we predict x is in the class with the largest vote. The voting approach described above is also called the “Max Wins” strategy [13]. In case those two classes have identical votes, thought it may not be a good strategy, now we simply select the one with the smaller index. Since we have considered 5 classes of disturbances, the total numbers of SVMs are 10.

C. Dendogram based SVM (DSVM)

The DSVM method takes advantage of both the efficient computation of the ascendant hierarchical clustering of classes and the high classification accuracy of SVM for binary classification. 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[10]. The division will continue until the classification task is finished as in fig8.

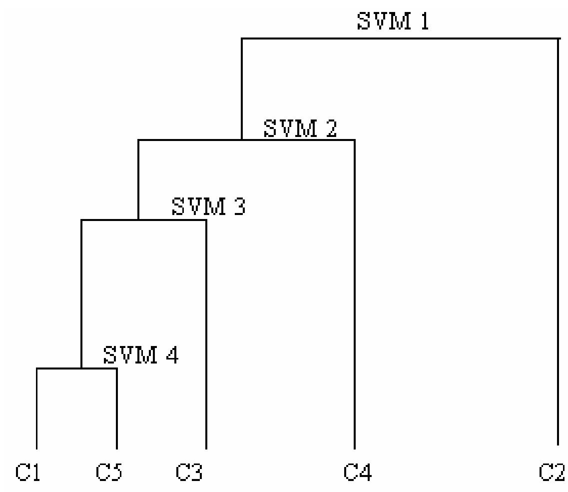

Fig. 8. Classification way by DSVM

V. APPLICATION AND RESULTS

Substantial computer simulations are conducted to optimize data selection phase using clustering. Using the output set of samples obtained from the clustering phase, we construct a new training set and the resulting set is presented to the Support Vector Machine described above. Table4 summarizes the obtained performance results. Because the goal of this work is to enhance the learning capabilities of the support vector machine for classification of power quality disturbances, the proposed method is compared to a classic support vector machine implementation that use the full set of samples.

The 10-fold cross validation evaluation result and performance of the SVM and Clustering SVM classifiers for the five data sets are shown in Table 2-5. 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 with clustering technique attains better recognition rates with less number of training data and classification time when compared with the conventional SVM classifier.

Table 2. Result of SVM Classifier

Class	One Against All	One Against One	DSVM
Sag	80	90	90
Swell	100	100	100
Interruption	80	90	90
Harmonics	100	100	100
Flicker	100	100	100
Overall	**92**	**96**	**96**

Table 3. Result of Clustering SVM Classifier

Class	One Against All	One Against One	DSVM
Sag	100	100	100
Swell	100	100	100
Interruption	80	90	90
Harmonics	100	100	100
Flicker	100	100	100
Overall	**96**	**98**	**98**

Table 4. Performance of SVM Classifier

Classifier	Training Time (sec)	Testing Time (sec)	Accuracy (%)
One Against All	0.5469	0.17	92
One Against One	0.6563	0.4563	96
Dendogram SVM	0.5	0.13	96

Table 5. Performance of Clustering SVM

Classifier	Training Time (sec)	Testing Time (sec)	Accuracy (%)
One Against All	0.1563	0.1875	96
One Against One	0.2031	0.4250	98
Dendogram SVM	0.1250	0.1719	98

VI. CONCLUSION AND FUTURE WORK

Since the high computation intensity and the long training cycles are the main obstacle to speed up SVM, we propose a new learning schema to reduce the amount of used samples using a k-means clustering algorithm. The proposed method can reduce the quantity of training data without losing its property, thus requiring less memory space and computing time for proper classification of disturbance types. The experimental results showed that the proposed method has the ability of recognizing and classifying different power disturbance types efficiently, and it has the potential to enhance the performance of the power transient recorder with real time processing capability. The proposed classifier has provided significant improvement in classifying the PQ events, as compared to the other conventional classifiers.

More work must be performed to find an optimal way to determine the number of used clusters and selected samples of each class. This work use only heuristics and trays to determine these parameters. 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] Mohammad S. Azam, Fang Tu, Krishna R. Pattipati*, Fellow, IEEE*, and Rajaiah Karanam,, July 2004, "A Dependency Model-Based Approach for Identifying and Evaluating Power Quality Problems", IEEE Transactions On Power Delivery, Vol. 19, No. 3.

[2] Peter G. V. Axelberg, Irene Yu-Hua Gu*, Senior Member,* July 2007*,*"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.

[3] Gaing Zwe-Lee, October 2004, "Wavelet-Based Neural Network for Power Disturbance Recognition and Classification", IEEE transactions on Power delivery, vol. 19, pp1560-1567.

[4] Borras Dolores, Castilla M, Moreno Narciso and J.C.Montano, January 2001, "Wavelet and Neural Structure: A New Tool for Diagnostic of Power System Disturbances", IEEE Transactions on Power Delivery, Volume 37, pp. 184-190.

[5] Chuang Cheng, Lu Yen, Tsong-Liang, Huang Ying, 2005, "Recognition of Multiple PQ Disturbances Using Wavelet-based Neural Networks—art :Implementation and Applications" IEEE/PES Transmission and Distribution Conference .

[6] Gargoom Ameen, Ertugrul Nesimi, Soong Wen., October 2007, "Investigation of Effective Automatic Recognition Systems of Power-Quality Events", IEEE Trans. On Power Delivery, Vol. 22, Pp.2319-2326.

[7] Ganyun LV, Xiaodong WANG, Haoran ZHANG, Changjiang ZHANG, "PQ Disturbances identification Based on SVMs Classifier".

[8] Quanganh tran, Qian-li zhang, Xing li, November 2003, "Reduce the number of support vectors by using clustering techniques", Proceedings of the Second International Conference on Machine Learning and Cybernetics, Wan, 2-5.

[9] R. G. Stockwell, L. Mansinha, and R. P.Lowe, Apr. 1996 ,"Localization of the complex spectrum: The S-transform,"*IEEE Trans. Signal Processing*, vol.144, pp. 998–1001.

[10] M. V. Chilukuri*, Member, IEEE,* and P. K. Dash*, Senior Member, IEEE,* January 2004 *,"* Multiresolution S- transform-Based Fuzzy Recognition System for Power Quality Events", IEEE Transactions On Power Delivery, Vol. 19, No. 1.pp.323-330.

[11] S. Mishra*, Senior Member, IEEE*, C. N. Bhende, and B. K. Panigrahi*, Senior Member, IEEE,* January 2008, *" D*etection and classification of Power Quality 1isturbances Using S-Transform and Probabilistic Neural Network", IEEE Transactions On Power Delivery, Vol. 23, No. 1, pp.280-287.

[12] Faraoun K.M. and Boukelif .A, 2005 , "Neural networksLearning Improvement using the K-Means Clustering Algorithm to Detect Network Intrusions", International Journal of Computational Intelligence, Vol. 3 No. 2.

[13] Chih-Wei Hsu and Chih-Jen Lin, March 2002 ,"A Comparison of Methods for Multiclass Support Vector Machines", IEEE Transactions On Neural Networks, Vol. 13, NO. 2, pp.415-425.

[14] Khalid Benabdeslem and Youn`es Bennani, 2006, "Dendogram- based SVM for Multi-Class Classification", Journal of Computing and Information Technology - CIT 14, 4, 283–289.

Manimala.K

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.