Multi - Class Power Quality Disturbances Classification by Using Ensemble Empirical Mode Decomposition Based SVM

Turgay Yalcin[1], Okan Ozgonenel[1], and Unal Kurt[2]

[1]Ondokuz Mayis University, Electrical & Electronic Eng. Depart., Kurupelit, Samsun, Turkey
yalcnturgay@yahoo.com.tr, okanoz@omu.edu.tr
Amasya University, Technical & Vocational Higher School, Amasya
unalkurt@msn.com

Abstract

This paper presents performance comparisons of Support Vector Machine (SVM) and different classification method for power quality disturbance classification. The first goal of this study is to investigate EEMD (ensemble empirical mode decomposition) performance and to compare it with classical EMD for feature vector extraction and selection of power quality disturbances. Features are extracted from the power electrical signals by using Hilbert Huang Transform (HHT). This technique is a combination of ensemble empirical mode decomposition (EEMD) and Hilbert transform (HT). The outputs of HT are instantaneous frequency (IF) and instantaneous amplitude (IA). Characteristic features are obtained from first IMFs', IF and IA. The ten features, i.e. mean, standard deviation, singular values, maxima and minima of IF and IA, are then calculated. These features are normalized and the inputs of SVM and other classifiers.

1. Introduction

The power monitoring/diagnostics has become an important role in power systems. Power quality, or more specifically, a power quality disturbance, is generally defined as any change in power (voltage, current, or frequency) that interferes with the normal operation of electrical equipment. Power quality concerns various kinds of electric disturbances such as voltage sag/swell, flicker, impulse, harmonics, DC component of voltage, oscillatory transient analysis, electromagnetic interference (EMI) and power system outage [1].

To analyze these electric power system disturbances, data is often available as a form of sampled time function that is represented by a time series of amplitudes. When dealing with such data, the Fourier transform (FT) based approach is most often used. FT assumes periodicity of a given signal and loses the time axis account; it is not capable of providing time information of signal disturbances. Short time Fourier transform (STFT) provides both time and frequency information, but it suffers severely from the Heisenberg uncertainty principle [2] causing it to undergo a "trade-off" between time resolution and frequency resolution. Wavelet transform [3], which is a popular signal analysis method, offers continuous and discrete wavelet transforms (CWT and DWT) [4] and wavelet packet transform (WPT) [5] for the feature extraction of signals.

On the contrary, many of the former decomposition methods, EMD is intuitive and direct, with the basis functions based on and derived from the data. The assumptions for this method are (a) the signal has at least a pair of extrema; (b) the characteristic time scale is defined by the time between the successive extrema; and (c) if there are no extrema, and only inflection points, then the signal can be differentiated to realize the extrema, whose IMFs can be extracted. Integration may be employed for reconstruction. The time between the successive extrema was used by Huang et al. [6] as it allowed the decomposition of signals that were all positive, all negative, or both. This implied that the data did not have to have a zero mean. This also allowed a finer resolution of the oscillatory modes.
Many classification algorithms have been developed by researchers for classification of power electrical disturbances. Integrated Fourier linear combiner and fuzzy expert system [7] were used for the classification of transient disturbance waveforms in a power system. S-Transform and two dimensional time–time (TT) transform [8] have been implemented for electrical fault identification. Patterns generated by S-Transform and TT transform are unique and hence accuracy of identification is high. An adaptive neural network approach for the estimation of harmonic distortions and power quality in power networks are implemented [9]. A hybrid system to automatically detect, locate and classify disturbances affecting power quality in an electrical power system is presented [10]. Least absolute value (LAV) State Estimation algorithm has been used to measure the flicker voltage magnitude [11]. The Simulated Annealing (SA) optimization algorithm has been used for measuring the voltage flicker magnitude, frequency and the harmonics contents of the voltage signal for power quality analysis [12]. An algorithm to detect the fundamental frequency is proposed. It is based on the chirp-z transform (CZT) spectral analysis and is able to observe all standards in force because of its accuracy and working characteristics [13]. Wavelet multi-resolution decomposition that combines frequency domain with time-domain analysis for power disturbance feature extraction is proposed [14]. A wavelet norm entropy-based effective feature extraction method for power quality disturbance classification problem has been studied by [15]. Multiwavelet- based neural networks with learning vector quantization network are used for power quality disturbances as a powerful classifier [16]. Fuzzy ARTMAP, Back propagation algorithm (BPA) and Radial Basis Function (RBF) network in combination with S Transform, Wavelet transform and Hilbert Transform (HT) for classifying power faults have been used [17].

In this work, the potential of a relatively recent method of ensemble empirical mode decomposition (EEMD) based SVM classification for analyzing nonlinear and non-stationary power disturbances is applied.

The classification process of real-time eight power quality disturbances are shown in Fig. 1.

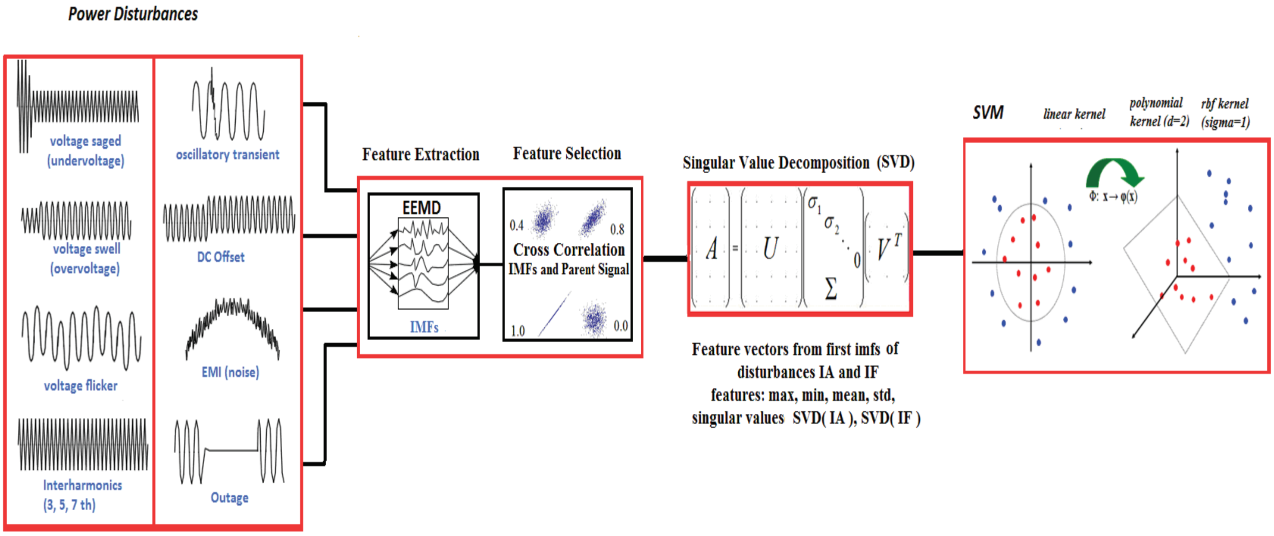

Fig. 1. Basic stages of the classification process

2. EMD, EEMD and Hilbert Huang Transform

2.1. EMD

The EMD algorithm used in this study comprises the following steps:

i. Identify all the extrema (maxima and minima) of the signal, x(t).
ii. Generate the upper and lower envelope by the cubic spline interpolation of the extrema point developed in step (1).
iii. Calculate the mean function of the upper and lower envelope, $m(t)$.
iv. Calculate the difference signal $d(t) = x(t) - m(t)$.
v. If $d(t)$ becomes a zero-mean process, then the iteration stop and $d(t)$ is an IMF1, named $c_1(t)$; otherwise, go to step (1) and replace $x(t)$ with d(t).
vi. Calculate the residue signal $r(t) = x(t) - c_1(t)$.
vii. Repeat the procedure from steps (i) to (vi) to obtain IMF2, named $c_2(t)$. To obtain $c_n(t)$, continue steps (i) – (vi) after n iterations. The process is stopped when the final residual signal $r(t)$ is obtained as a monotonic function.

At the end of the procedure, we have a residue $r(t)$ and a collection of n IMF, named from $c_1(t)$ to $c_n(t)$. Now, the original signal can be represented as:

$$x(t) = \sum_{i=1}^{n} c_i(t) + r(t) \qquad (1)$$

Often, $r(t)$ is regadred as $c_{n+1}(t)$ [6].

2.2 EEMD

The steps for the EEMD algorithm are as follows:

i. Add a white noise series $n(t)$ to the targeted signal, named $x_1(t)$ in the following description, and $x_2(t)$=$x_1(t)$+$n(t)$.
ii. Decompose the data $x_2(t)$ by EMD algorithm, as described in *Section 2.1.1.*
iii. Repeat Steps (i) and (ii) until the trial numbers, each time with different added white noise series of the same power at each time. The new IMF combination $C_{ij}(t)$ is achieved, where i is the iteration number and j is the IMF scale.
iv. Estimate the mean (ensemble) of the final IMF of the decompositions as the desired output:

$$EEMD\_c_j(t) = \sum_{i=1}^{ni} c_{ij}(t) \qquad (2)$$

where ni denotes the trial numbers, according to Wu [18].

2.3. Comparison of EMD and EEMD decomposition for feature extraction and selection

The typical EMD and EEMD decomposition and extracted IMF are shown in Fig. 2. The low level IMF contained high frequency components; while the high level IMF contained low frequency components.

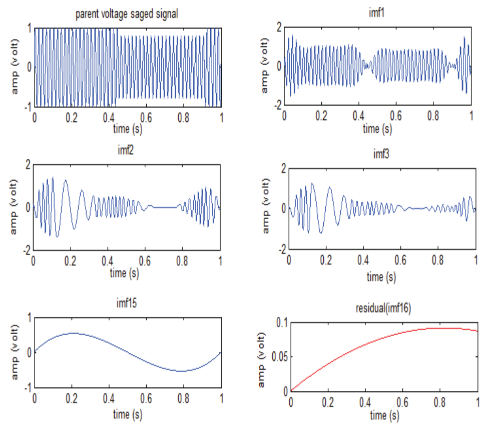

Fig. 2a. Feature vectors (IMFs) for a voltage saged signal with EMD

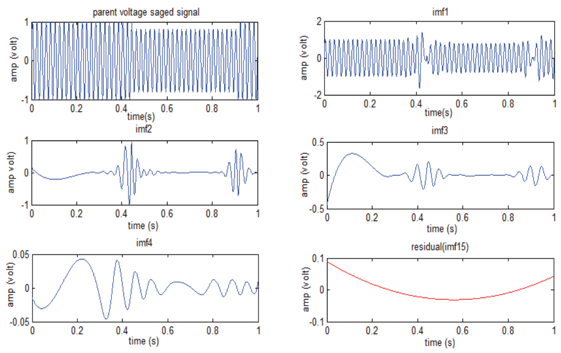

Fig.2b. Feature vectors (IMFs) for a voltage saged signal with EEMD

2.3.1. Pearson product-moment correlation coefficient (PMCC):

The Pearson product-moment correlation coefficient (R) is a measure of the correlation (linear dependence) between two variables X and Y, giving a value between +1 and −1 inclusive.

An equivalent expression gives the correlation coefficient as the mean of the products of the standard scores. Based on a sample of paired data (X_i, Y_i), the sample Pearson correlation coefficient is defined in Eq. (3)

$$R = \frac{1}{n-1}\sum_{i=1}^{n}[(\frac{X_i - \bar{X}}{\sigma_X})(\frac{Y_i - \bar{Y}}{\sigma_Y})] \quad (3)$$

where $\frac{(X_i - \bar{X})}{\sigma_x}$, $\bar{X}$ and σ_X are the standard score, sample mean, and sample standard deviation of X. Table 1 shows the R values for EMD and EEMD.

Table 1. Relation with first IMFs and current power signal

Cross-Correlation	EMD	EEMD
IMF- healthy state	0.9762	1.
IMF- voltage sag	0.8819	0.9081
IMF- voltage swell	0.9300	0.9685
IMF- voltage flicker	0.9700	0.9950
IMF- harmonics (3, 5, 7)	0.7855	0.7861
IMF- transient	0.5685	0.6104
IMF- dc component	0.9244	0.9357
IMF- EMI	0.5334	0.6429
IMF- outage	0.9313	0.9856

Table 1 clearly shows that correlation coefficient increases in EEMD method and a better decomposition is obtained.

2.4. Hilbert-Huang Transformation Method (HHT)

HHT represents the original signal $X(t)$ into the time frequency domain by combining the empirical mode decomposition with the Hilbert transform. EMD can decompose signal into a finite number of n IMFs C_j , $j=1,2,3...,n$ which extract the energy associated with various intrinsic time scales and residual r_n , this step is called *sifting process* described as in Eq. (1).

The IMFs as a class of functions that satisfy two definitions:
In the whole data set, the number of extrema and the number of zero-crossings must be either equal or differ at most by one.
At any point, the mean value of the envelop defined by the local maxima and minima is zero.
The Hilbert transform is then applied to each IMF component C_j:

$$v_j(t) = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{c_j(\tau)}{t-\tau}d\tau \quad (4)$$

Where $c_j(t)$ and $v_j(t)$ are real part and imaginary part of an analytic signal $z_j(t)$:

$$z_j(t) = c_j(t) + jv_j(t) \quad (5)$$
$$z_j(t) = A_j(t)\exp(jw_j(t))$$

With *amplitude* and *phase* defined by the expressions:

$$A_j(t) = \sqrt{c_j(t)^2 + v_j(t)^2}$$
$$\theta_j(t) = \arctan(\frac{v_j(t)}{c_j(t)}) \quad (6)$$

Therefore, the *instantaneous frequency* $w_j(t)$ was given by:

$$w_j(t) = \frac{d\theta_j(t)}{dt} \quad (7)$$

Thus, the original data can be expressed in the following form:

$$x(t) = \sum_{j=1}^{n} A_j(t)\exp\{i\int w_j(t)\,dt\} \quad (8)$$

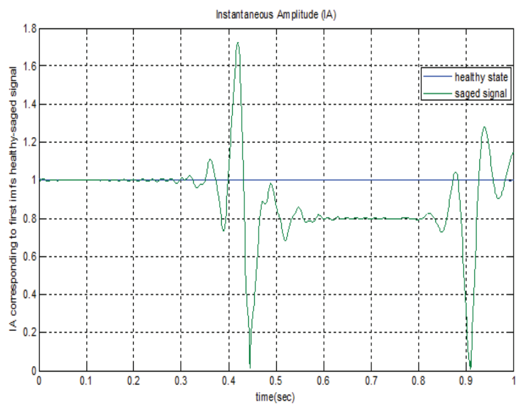

Fig. 3. IA corresponding to first IMF for healthy signal and saged signal

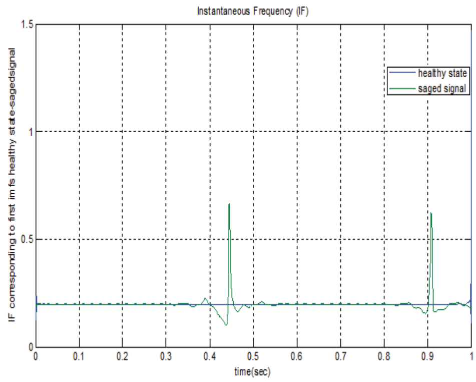

Fig. 4. IF corresponding to first IMF healthy signal and saged signal

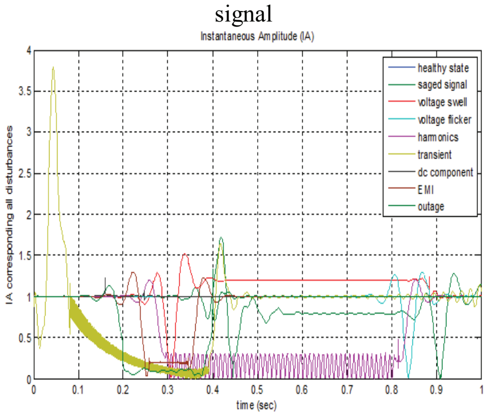

Fig. 5. IA corresponding to first IMF for all disturbances

The outputs of Hilbert Transform are instantaneous frequency (IF) and instantaneous amplitude (IA). Candidate characteristic features are obtained from first IMFs', IF and IA such kind of disturbances as shown Fig. 3, 4, and 5.

3. Singular Value Decomposition

Singular values give more information for identification of the system. We use this ability to use for future studies *Hierarchical classifying* on multiple feature groups.
For any real ***mxn*** matrix A, then there exist orthogonal matrices

$$U = [u_1, u_2, \ldots, u_m] \in R^{mxm}\ , \qquad (9)$$
$$V = [v_1, v_2, \ldots, v_m] \in R^{nxn}\ ,$$

such that;

$$A = U \Sigma V^T \qquad (10)$$

Where

$$\Sigma = diag(\sigma_1, \sigma_2, \ldots, \sigma_{\min(m,n)})) \in R^{mxn} \qquad (11)$$
$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(m,n)} \geq 0$$

The σ_i is the *i-th* singular value of A in non-increasing order and the vectors u_i and v_i are the *i-th* left singular vector and the *i-th* right singular vector of A for $i = min(m,n)$, respectively [19]. The singular values of matrix A are unique, the singular vectors corresponding to distinct singular values are uniquely determined up to the sign.

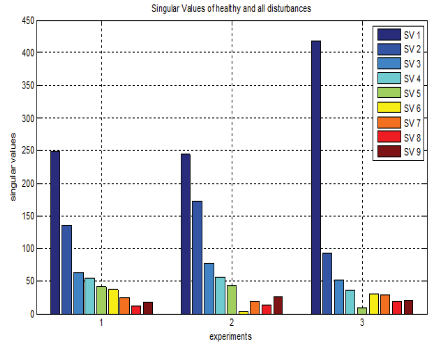

Fig. 6. Singular values for healthy state and all quality disturbances (IA)

4. Support Vector Machines (SVMs)

SVMs are a class of learning machines that aim at finding optimal hyperplanes among different classes of input data or training data in a high dimensional feature space F, and new test data can be classified using the separating hyperplanes. The optimal hyperplane, found during a *training* phase, makes the smallest number of training errors. Fig. 7 illustrates an optimal hyperplane for two classes of training data.

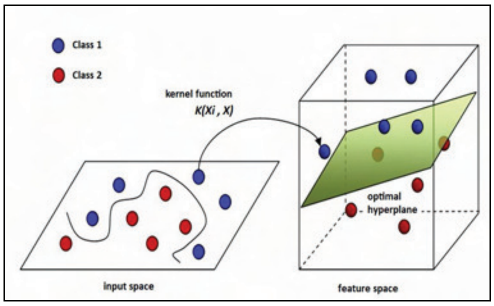

Fig. 7. Optimal Hyperplane

Let $\{x_i, y_i\}$, $i = 1,2,\ldots,L$ be L training data vectors x_i with class label y_i. Given an input vector x, an SVM constructs a classifier of the form:

$$f(x) = sign(\sum_{i=1}^{L} a_i y_i K(x_i, x) + b) \qquad (12)$$

where $\{a_i\}$ are non-negative Lagrange multipliers each of which corresponds to a training data, b is a bias constant, and $K(.\ ,\ .)$ is a kernel satisfying the conditions of *Mercer's theorem* [20]. Frequently used kernel functions are the polynomial kernel $K(x_i, x_j) = (x_i . x_j + 1)^d$ and Gaussian Radial Basis Function (RBF), $K(x_i, x_j) = e^{-\frac{|x_i - x_j|^2}{2\sigma^2}}$.

The problem of finding the optimal hyperplane is specified by the following quadratic programming problem:

Minimizing:

$$W(a) = -\sum_{i=1}^{L} a_i + \frac{1}{2}\sum_{i=1}^{L}\sum_{j=1}^{L} a_i a_j y_i y_j K(x_i, x_j)$$

(13)

Subject to:

$$\sum_{i=1}^{L} a_i y_i = 0 \quad (14)$$

$$0 \le a_i \le C, i = 1,2,..., L$$

The above quadratic programming problem can be solved with traditional optimization techniques. The vectors for which $a_i > 0$ after optimization are called *support vectors.*

A comparative study is carried out in the next section comparing the performances of the OAO (one against one) and OAA (one against all) decomposition methods in training the proposed multiclass SVMs. The results of the application of the proposed algorithm to the data set are presented at Table 2.

4. Experimental Results and Discussions

To evaluate the performance of the proposed power quality classification algorithm, a total number of 115 PQ disturbances data sets segments. The PQ signals at each class are divided into the train 90 and test sets 25. In this part of study, the polynomial kernel is used as the kernel function and the parameter *d* is chosen 2. The exact number of the train and test segments for each class is shown in Table 2. The most accurate multi classifiers (OAA) for each PQ dataset are indicated by Red font.

Table 2. Comparisons of the performances of the one-against-all (OAA) and the one-against-one (OAO) SVM classifiers in terms of the average correct classification results obtained for each class

Total number of train segments	**Classification (SVM) One Against All (OAA) – One Against One (OAO)**								
90/25	**HS**	**VSG**	**VSW**	**VF**	**VH**	**VTR**	**VDC**	**VEM**	**VO**
HS	**3 - 2**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
VSG	**1**	**2 - 2**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
VSW	**0**	**0**	**2 - 2**	**1**	**0**	**0**	**0**	**0**	**0**
VF	**0**	**0**	**0**	**3 - 2**	**0**	**0**	**0**	**0**	**0**
VH	**0**	**0**	**0**	**0**	**3 - 3**	**0**	**0**	**0**	**0**
VTR	**0**	**0**	**0**	**0**	**0**	**4 - 3**	**0**	**1**	**0**
VDC	**0**	**0**	**0**	**0**	**0**	**0**	**3 - 3**	**0**	**0**
VEM	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**3 - 2**	**0**
VO	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**2 - 2**

HS: healthy state, **VSG:** voltage saged signal, **VSW:** voltage swell, **VF:** voltage flicker, **VH:** Voltage Harmonics

As seen from the Table 2, the OAA method (Error = 0%, Precision = 100%) is superior to the OAO method (Error = 16%, Precision = 84 %). Therefore, the OAA decomposition method is chosen to the multiclass SVM in this study.

According to the results of test data, the effectiveness of the proposed SVM algorithm is suitable by increasing RBF kernel parameter's standard deviation (sigma 0.1 to 1). RBF polynomial kernel is worse than RBF kernel because of CPU time. Decision tree technique is superior than the others. It is another contribution of this research work.

5. Conclusion

In this paper, EEMD-HHT based Support Vector Machine (SVM) and different classification method for power quality disturbance classification algorithm was presented. EEMD is superior to EMD on IMFs decomposition. It is able to solve the problem of feature extraction selection method for PQ disturbances classification. By using EEMD method, a better decomposition can be obtained.

Singular values feature gives more information for identification of the system. It is used for *Hierarchical classifying* on multiple feature groups. Real time multi-class classification of power quality disturbances is the future work.

6. References

[1] M. D. L. Angrisani, P. Daponte and A. Testa, "A measurement method based on the wavelet transform for power quality analysis,"IEEE Transaction on power delivery, vol.13, no. 4,pp. 990-998,October 1988.

[2] Patrick. J. Loughlin, James. W. Pitton and Les E. Atlas, "Proper time-frequency energy distributions and the Heisenberg uncertainty principle". IEEE Signal Processing, pp. 151-154, Oct.1992

[3] I. Daubechies, Ten Lectures on Wavelets, SIAM, Philadelphia, PA, 1992.

[4] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation", IEEE Pattern Analysis and Machine Intelligence 11, pp. 674-693, 1989.

[5] R. R. Coifman, M.V. Wickerhauser, "Entropy-based algorithms for best basis selection", IEEE Information Theory 38, pp. 713-718, 1992.

[6] Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.;Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear andnon-stationary time series analysis. *Proc. Roy. Soc. Lond.* 1998, *454*, 903–995.

[7] Dash, P.K., R.K. Jena, M.M.A. Salama, 1999. Power quality monitoring using an integrated Fourier linear combiner and fuzzy expert system. Electrical Power and Energy Systems, 21, 497–506.

[8] Suja, S., Jovitha Jerome, 2010. Pattern recognition of power signal disturbances using S Transform and TT Transform. Electrical Power and Energy Systems, 32, 37-53.

[9] Dash, P.K., S.K. Panda, A.C. Liew, B. Mishra, R.K. Jena, 1998. A new approach to monitoring electric power quality. Electric Power Systems Research, 46, 11–20.

[10] Mário Oleskovicz, et al., 2009. Power quality analysis applying a hybrid methodology with wavelet transforms and neural networks. Electrical Power and Energy Systems, 31, 206–212.

[11] Soliman, S.A., M.E. El-Hawary, 2000. Measurement of power systems voltage and flicker levels for power quality analysis: a static

LAV state estimation based algorithm. Electrical Power and Energy Systems, 22, 447–450.
[12] Soliman, S.A., A.H. Mantaway, M.E. El-Hawary, 2004. Simulated annealing optimization algorithm for power systems quality analysis. Electrical Power and Energy Systems, 26, 31–36.
[13] Aiello, M., A. Cataliotti, S. Nuccio, 2005. A Chirp-Z transform-based synchronizer for power system measurements IEEE Trans. Instrum. Meas, 54, 1025–1032.
[14] Zheng, G., D. M'x' Shi, Liu,1. Yao, Miao, Z.M, 2002. Power quality disturbance classification based on rule-based and wavelet-multi-resolution decomposition. Proceedings of 2002 International Conference on Machine Learning and Cybernetics, 4, 2137-2141.
[15] Murat Uyar. Selcuk Yildirim, Muhsin Tunay Gencoglu, 2008. An effective wavelet-based feature extraction method for classification of power quality disturbance signals. Electric Power Systems Research, 78, 1747–1755.
[16] Suriya Kaewarsa. et al., 2008. Recognition of power quality events by using multiwavelet based neural networks. Electrical Power and Energy Systems, 30, 254–260.
[17] Jayasree, T., D. Devaraj, R. Sukanesh, 2010. Power quality disturbance classification using Hilbert transform and RBF networks. Neurocomputing, 73, 1451-1456.
[18].Wu, Z.; Huang, N.E. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv. Adapt. Data. Anal.* 2009, 1, 1–41.
[19] G. H. Golub and C. F. V. Loan. Matrix Computations. The Johns Hopkins University Press, Baltimore, Maryland, 1996.
[20] V. N. Vapnik. Statistical Learning theory. Wiley, New York, 1998.
[21] Hsu C-W, Lin C-J. A comparison of methods for multiclass support vector machines. IEEE Trans Neural Networks 2002; 13(2):415—25.