On the choice of wavelet based features in power quality disturbances classification

Marija Markovska, Dimitar Taskovski
Ss. Cyril and Methodius University in Skopje
Faculty of Electrical Engineering and Information Technologies
Skopje, R. Macedonia
marijam, dtaskov@feit.ukim.edu.mk

***Abstract*— In this study we investigate the effectiveness of some wavelet based features used for classification of power quality (PQ) disturbances, and check the difference in their efficiency when they are used in combinations, in order to perform optimal wavelet based feature extraction method. The investigation was made using three well known classification techniques, which are support vector machine (SVM), decision tree (DT) and random forest (RF), for classification in case of 7 and 11 types of PQ disturbances. In both cases it is shown that the effectiveness of a given feature is not general, but it depends on the type of other features it is used with and the kind of the applied classification method.**

***Keywords*— Power quality, disturbances, wavelets, classification, optimal feature extraction**

I. Introduction

The increasing number of polluting loads requires higher power quality (PQ) in the generation, transmission and distribution systems, due to the undesirable effects caused by PQ disturbances. These effects are plainly noticeable at industrial or public facilities, where the PQ disturbances cause malfunction in the equipment and may lead to interruption of the working process. In order to ensure high quality power supply of the power grid and prevent the devices from damage, it is very important to analyze and recognize these PQ disturbances where the main issue is how to extract the feature vectors automatically from large amount of PQ data and classify the PQ events accurately [1].

The researchers have introduced various classification methods for automated classification of PQ disturbances. Some commonly used artificial intelligence (AI) based cassifiers are the rule-based expert systems (ES) [2], artificial neural networks (ANN) [3], fuzzy logic (FL) classification systems [4], support vector machines (SVM) [5], and data mining based classifiers [6]. All these classification methods use features extracted from PQ disturbance signals. The performance and the accuracy of the used classifier depend from the feature vector constructed with the selected features. Hence, with proper selection of feature extraction technique it is possible to achieve higher classification accuracy.

In [7] several digital signal processing techniques are proposed for extraction of features that are typical for certain PQ event, such as the fast working Fourier transform (FT), short-time Fourier transform (STFT), S-transform (ST), Neural Network, Fuzzy logic and, the wavelet transform (WT) [8], [9]. Among all these processing techniques, in the last years the wavelet transform has been most extensively used. Wavelet transform analysis give time and frequency information accurately, by convolving the dilated and translated wavelet with the input signal. This feature makes the wavelet transform approach suitable for analysis of transients that occur in the power system caused by different kinds of PQ disturbances [10].

The feature selection should be performed for the purpose of not only increase the classification accuracy, but also to reduce the response time of the protection system. Therefore, low dimensional feature space is needed. Different feature construction methods for selection of ‘‘better” features among the given feature set were proposed for this purpose. However, all these methods introduce additional mathematics and increase the computational cost of the classification algorithm. In this paper a research using the comparison of ‘‘how useful these wavelet based features are with respect to each other” for classification of 7 and 11 kinds PQ disturbance signals is presented. The classification was performed using three different classifiers, which are support vector machine (SVM), decision tree (DT) and random forest (RF) [11].

This paper is organized as follows: The feature extraction technique and construction of the feature vector are given in Section II. In Section III are theoretically presented the classification methods which are used for the experimental classification purposes. Some of the obtained results for experimental classification of 7 and 11 types of PQ disturbances are presented in Section IV. The conclusion from the presented research and future discussion are given in Section V.

II. Feature vector construction

In the feature extraction process PQ disturbance signals are decomposed using discrete wavelet transform (DWT). The wavelet analysis is in fact a measure of similarity between the mother wavelet and the input signal. Hence, the proper selection of wavelet mother function is one of the main issues for performing a successful DWT application. In this work Daubechies4 wavelet filter is used, as one of widely used mother wavelets in power quality problems [12].

The number of decomposition levels, l, is also of significant importance. Higher number of decompositions levels results in increased computational cost. Additionally,

978-1-5386-3917-7/17/$31.00 ©2017 IEEE

choosing higher l will bring more information in the system i.e. extended feature vector, and in that way will provide classification with higher accuracy. Experimental results have shown that the increasing the number of decomposition levels after l=6 do not significantly affect the accuracy of algorithm [12]. Therefore, in this research we have chosen $l = 6$.

With the detail coefficients obtained at each decomposition level and approximation coefficients at the last level, feature vector is constructed. Several feature extraction techniques have been proposed using different wavelet based features, given in Table I. In [12]-[14] a feature extraction approaches are presented using only energy. In [15] energy, standard deviation and Shannon entropy are used. In [16] all nine features presented in the table are used.

In the equations from the table, $i = 1,2,3...l$ represents the level of decomposition, C_{ij} stands for detail coefficients at each multiresolution level and approximation coefficients from the last level, while N is the number of these coefficients.

TABLE I. EXPRESSIONS OF THE DWT BEASED FEATURES

Labels	Features based on DWT	
F1	Mean	$\mu_i = \frac{1}{N}\sum_{j=1}^{N} C_{ij}$
F2	Standard deviation	$\sigma_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(C_{ij} - \mu_i\right)^2}$
F3	Skewness	$SK_i = \sqrt{\frac{1}{6N}\sum_{j=1}^{N}\left(\frac{C_{ij} - \mu_i}{\sigma_i}\right)^3}$
F4	Kurtosis	$KRT_i = \sqrt{\frac{N}{24}}\left(\frac{1}{N}\sum_{j=1}^{N}\left(\frac{C_{ij} - \mu_i}{\sigma_i}\right)^4 - 3\right)$
F5	RMS	$rms_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N} C_{ij}^2}$
F6	Energy	$E_i = \sum_{j=1}^{N}\left\|C_{ij}\right\|^2$
F7	Shannon Entropy	$SE_i = -\sum_{j=1}^{N} C_{ij}^2 \log(C_{ij}^2)$
F8	Log-energy entropy	$LOE_i = \sum_{j=1}^{N} \log(C_{ij}^2)$
F9	Norm entropy	$NE_i = \sum_{j=1}^{N} (C_{ij})^P \ ,1 \le P$

The above features reflect the disturbance characteristics of different types of PQ disturbances. When a disturbance occurs, the values of some features will have significant difference between the different kinds of PQ disturbances.

In this work we investigate the influence of different wavelet based features and their combinations with the commonly used features [12]-[16] on the accuracy of the classification methods. From the large number of combinations in Section IV are given the most optimal results.

III. CLASSIFICATION METHODS

For the aim of experimental choice of optimal wavelet based features for PQ classification, three classification methods were used. A theoretical overview for these methods is given in addition.

A. Support vector machine (SVM)

SVM is a supervised machine learning technique which is based on structural minimization method and is frequently used in application where regression and classification analysis are needed. The main goal of the SVM is to build a model using a set of training samples, where each sample is denoted as belonging into one of the possible classes.

For M–dimension of training samples $x_i (i = 1,2,3,...N)$, N is the sample number present in one of the two classes $c_i = \{-1,1\}$. Using the trained SVM, a prediction model can be created which can separate new samples of the classes using a hyperplane given as:

$$f(x) = \omega \cdot x - b = 0 \tag{1}$$

where $\omega \cdot x$ represents the dot product, ω is the normal vector to the hyperplane. The parameter $b / \|\omega\|$ defines the offset of the given hyperplane from the origin along ω. Hence, the main task of classification process is to choose a hyperplane which can best separate the two classes. With the M–dimensional vector the above equation can be written as:

$$f(x) = \sum_{j=1}^{M} \omega_j \cdot x_j - b = 0 \tag{2}$$

When the two classes are not linearly separable a soft margin SVM technique is used [17]. This technique can choose a hyperplane which can still best separate the data.

B. Decision tree (DT)

DT is widely used in classification which is based on the choice of a feature that maximizes and fixes data division [18]. These features are split into several branches recursively, until the termination and classification is reached. The mathematical illustration of the DT algorithm is constructed using the following definitions:

$$\bar{X} = \{X_1, X_2, ..., X_m\}^T \tag{3}$$

$$X_i = \{x_1, x_2, ..., x_{ij}, ..., x_{in}\} \tag{4}$$

$$S = \{S_1, S_2, ..., S_i, ..., S_m\} \tag{5}$$

where m is the available observation number, n is the number of independent variables, S is the m-dimensional vector of the variable forecasted from $\bar{X}$, X_i is the i-th component of the n-dimension autonomous variables, $x_1, x_2, ..., x_{ij}, ..., x_{in}$ are autonomous variables of the pattern vector X_i, and T is the transpose notation vector. The main objective of DT is to forecast S based on the observations of $\bar{X}$.

TABLE II. MATHEMATICAL MODELS FOR THE USED PQ DISTURBANCES

Disturbance	Model	Parameters
Normal	$x(t)=\sin(\omega t)$	$f=50Hz, \omega=2\pi f$
Sag	$x(t)=[1-\alpha(u(t-t_1)-u(t-t_2))]\sin(\omega t)$	$0.1\le\alpha\le0.9, T\le t_2-t_1\le 9T$
Swell	$x(t)=[1+\alpha(u(t-t_1)-u(t-t_2))]\sin(\omega t)$	$0.1\le\alpha\le0.8, T\le t_2-t_1\le 9T$
Outage / Interruption	$x(t)=[1-\alpha(u(t-t_1)-u(t-t_2))]\sin(\omega t)$	$0.9\le\alpha\le1, T\le t_2-t_1\le 9T$
Flicker	$x(t)=[1+\alpha\sin(2\pi\beta t)]\sin(\omega t)$	$0.1\le\alpha\le0.2, 5Hz\le\beta\le 20Hz$
Oscillatory transient	$x(t)=\sin(\omega t)+\alpha\exp(-(t-t_1)\tau)(u(t-t_1)-u(t-t_2))\sin(2\pi f_n t)$	$0.1\le\alpha\le0.8, 8ms\le\tau\le 40ms$ $0.5T_n\le t_2-t_1\le 3T_{n,}$ $300Hz\le f_n\le 900Hz$
Harmonics	$x(t)=\alpha_1\sin(\omega t)+\alpha_3\sin(3\omega t)+\alpha_5\sin(5\omega t)+\alpha_7\sin(7\omega t)$	$0.05\le\alpha_3,\alpha_5,\alpha_7\le 0.15, \sum\alpha_i^2=1$
Notch	$x(t)=\sin(\omega t)-\text{sgn}(\sin(\omega t))\left\{\sum_{k=0}^{9}k[u(t-(t_1+0.2n))-u(t-(t_2+0.2n))]\right\}$	$0.1\le k\le 0.4, 0\le t_1,t_2\le 0.5T$ $0.01T\le t_2-t_1\le 0.05T$
Spike	$x(t)=\sin(\omega t)+\text{sgn}(\sin(\omega t))\left\{\sum_{k=0}^{9}k[u(t-(t_1+0.2n))-u(t-(t_2+0.2n))]\right\}$	$0.1\le k\le 0.4, 0\le t_1,t_2\le 0.5T$ $0.01T\le t_2-t_1\le 0.05T$

The operation of the DT starts from the node called “rootnode”. The next step is division of the records according to their features. The assigned node is called “intermediatenode”. Similarly the next step is assignment of a class label to the nodes. The class label assigned node is called “leafnode”. Once the classifier is designed, testing of a new record is a simple approach.

C. Random forest (RF)

RF represents a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The RF classifier is defined as:

$$\left\{f(x,\delta_k), k=1,\ldots\right\} \tag{6}$$

where $f(x,\delta_k)$ is a meta classifier and it is a tree construct classifier that can be formed by several algorithms, x is the input vector, while δ_k are k random vectors and each of them determines the growth of a single decision tree. The RF summarizes the classification results of different decision trees in order to achieve the optimal classification results.

RF is an effective data mining based classifier, especially when dealing with massive amounts of data. It has more advantages, compared to the other type of classifiers, such as less parameters, less influence of the over-fitting problem and good anti-noise performance [19]. It can satisfy the demands of situations when real-time performance is highly needed.

IV. EXPERIMENTAL CLASSIFICATION RESULTS

The generation of the PQ signals was performed according to the mathematical definitions presented in Table II, using MATLAB 8.5. The experimental classification was made in case of 7 and 11 different classes of PQ disturbances. The class labels of the used classes are given in Table III. In every PQ disturbance signal ten cycles are included, with a sampling frequency of 3.2 KHz (64 samples/cycle), so that every signal consists of 640 samples. The fundamental frequency is 50 Hz.

TABLE III. 11 TYPES OF PQ DISTURBANCES

PQ Disturbance	Class Labels
Normal	D1
Swell	D2
Sag	D3
Harmonic	D4
Outage	D5
Sag + Harmonic	D6
Swell + Harmonic	D7
Flicker	D8
Oscillatory transient	D9
Notch	D10
Spike	D11

In the case of 7 PQ disturbances, the learning of the classifiers was done using training and testing sets consisted of 7000 generated PQ signals, while in case of 11 disturbances each set was consisted of 11000 PQ signals. In both cases, every training and testing class has size of 1000 samples.

Some of the classification accuracies obtained using different combinations of the wavelet based features and SVM, DT and RF classifiers for 7 classes of PQ disturbances are given in Table IV, Table V and Table VI, respectively.

TABLE IV. ON THE CHOICE OF FEATURE EXTRACTION METHOD USING SVM CLASSIFIER FOR 7 CLASSES

	F6	**F6 + F8**	**F2+F6+ F7**	**F2+F6+ F7+F8**	**ALL**
	7 features	**14 features**	**21 features**	**28 features**	**63 features**
D1	100	100	100	100	100
D2	100	100	100	100	100
D3	83,7	99,3	96,7	99,3	99,2
D4	100	100	100	100	100
D5	90,4	99,1	98,1	99,4	99,4
D6	99,4	99,9	99,8	100	99,9
D7	100	100	100	100	100
TA%	**96,21**	**99,76**	**99,23**	**99,81**	**99,79**

TABLE V. ON THE CHOICE OF FEATURE EXTRACTION METHOD USING DT CLASSIFIER FOR 7 CLASSES

	F6	F6 + F8	F2+F6+ F7	F2+F6+ F7+F8	ALL
	7 features	14 features	21 features	28 features	63 features
D1	100	100	100	100	100
D2	98,6	98,7	99,3	99,3	99,3
D3	83,7	95,7	91,6	95,6	94,4
D4	100	100	99,9	100	100
D5	90,3	98,8	96,2	97,1	97
D6	92,4	98,3	97,8	99,1	98,7
D7	98,9	98,9	99,9	99,9	99,9
TA%	**94,84**	**98,63**	**97,81**	**98,71**	**98,47**

TABLE VI. ON THE CHOICE OF FEATURE EXTRACTION METHOD USING RF CLASSIFIER FOR 7 CLASSES

	F6	F6 + F8	F2+F6+ F7	F2+F6+ F7+F8	ALL
	7 features	14 features	21 features	28 features	63 features
D1	100	100	100	100	100
D2	99,8	100	100	100	100
D3	85,5	97,3	91,2	96,5	95,8
D4	100	100	100	100	100
D5	93,1	99,6	99,3	99,7	99,8
D6	96,5	99,6	99,3	99,7	99,5
D7	99,8	99,9	100	100	100
TA%	**96,39**	**99,49**	**98,54**	**99,41**	**99,30**

For better comparison between the effectiveness of different feature combinations on the classification accuracy, the same results are graphically shown in Fig. 1.

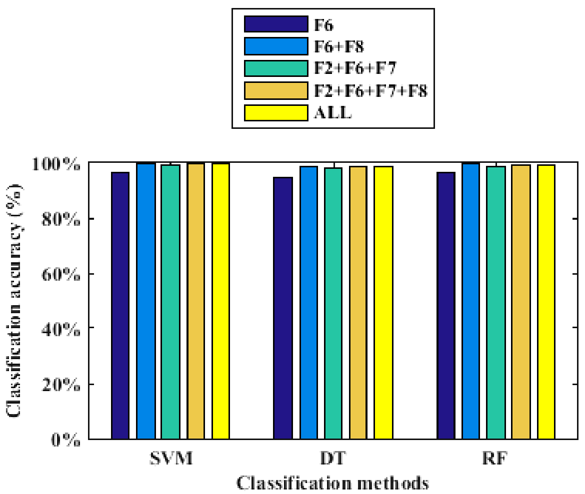

Fig. 1 Graphical representation of classification accuracies for 7 PQ classes

From the presented results it is evident that the accuracies of the feature vectors constructed using four wavelet based features, i.e. energy, standard deviation, Shannon entropy and log-energy entropy and the accuracies of the feature vectors constructed using two wavelet based features i. e. energy and log-energy entropy are almost identical with the accuracies obtained when all nine features are used. Moreover, in most of the cases they are even slightly higher.

Same experimental classifications were made for 11 PQ disturbance classes. Some of the obtained results, using SVM, DT and RF classification methods and different wavelet based features, are given in Table VII, Table VIII and Table IX, respectively.

TABLE VII. ON THE CHOICE OF FEATURE EXTRACTION METHOD USING SVM CLASSIFIER FOR 11 CLASSES

	F6	F6+F8	F2+F6+ F7	F2+F6+ F7+F8	ALL
	7 features	14 features	21 features	28 features	63 features
D1	100	100	100	100	100
D2	94,4	100	97,2	100	100
D3	83,7	99,3	96,7	99,3	99,2
D4	100	100	100	100	100
D5	90,4	99,1	98,1	99,4	99,4
D6	99,4	99,5	99,8	99,9	99,9
D7	100	100	100	100	100
D8	98,6	100	100	100	100
D9	77,6	86	84,8	87,9	89,9
D10	89,9	91,5	91,6	93,6	96
D11	92,2	91,6	92,5	91,7	94,6
TA%	**93,29**	**97,00**	**96,43**	**97,44**	**98,09**

TABLE VIII. ON THE CHOICE OF FEATURE EXTRACTION METHOD USING DT CLASSIFIER FOR 11 CLASSES

	F6	F6 + F8	F2+F6+ F7	F2+F6+ F7+F8	ALL
	7 features	14 features	21 features	28 features	63 features
D1	100	100	100	100	100
D2	98,1	99,1	98,8	99,6	99,6
D3	85,6	95,7	91,6	95,6	94,5
D4	100	100	100	100	100
D5	85,4	98,8	96,6	97,2	97,2
D6	94,1	98,9	98	99,5	99
D7	99	98,8	99,6	99,7	99,7
D8	99,7	100	99,8	99,8	99,8
D9	86,7	89,5	88,6	88,9	88,4
D10	94,2	95,2	94,2	94,5	94,8
D11	93,4	95,4	92,6	95,1	94,5
TA%	**94,20**	**97,40**	**96,35**	**97,26**	**97,05**

TABLE IX. ON THE CHOICE OF FEATURE EXTRACTION METHOD USING RF CLASSIFIER FOR 11 CLASSES

	F6	F6 + F8	F2+F6+ F7	F2+F6+ F7+F8	ALL
	7 features	14 features	21 features	28 features	63 features
D1	100	100	100	100	100
D2	98,8	100	99,4	100	100
D3	85,5	97	91,5	96,6	95,7
D4	100	100	100	100	100
D5	92,7	99,6	99,5	99,7	99,9
D6	96,1	99,5	99,2	99,7	99,7
D7	99,8	100	100	100	100
D8	99,7	100	99,9	100	100
D9	91,1	93,5	92,4	93,7	94,3
D10	96,2	96,1	96,2	96,3	96,5
D11	96	95,3	96	95,5	96
TA%	95,99	98,27	97,65	98,32	98,37

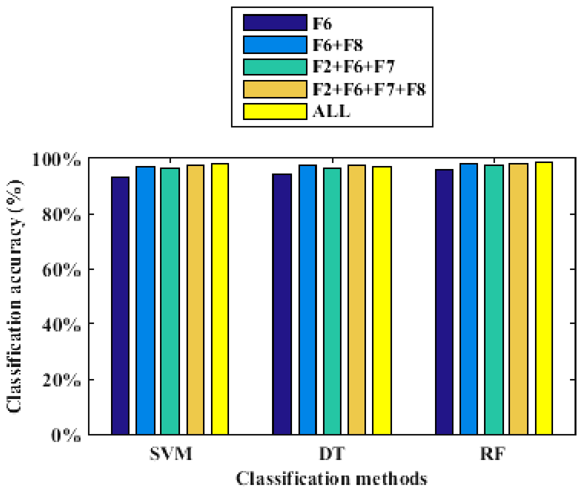

Fig. 2 Graphical representation of classification accuracies for 11 PQ classes

For comparative purposes of the experimental classification, the results given in Table VII, Table VIII and Table IX are graphically presented in Fig. 2. From the figure it is noticeable that the results obtained for classification of 11 PQ disturbances classes are very similar to those obtained for 7 classes. Namely, the accuracies that were achieved with feature vectors constructed using four (i.e. energy, standard deviation, Shannon entropy and log-energy entropy) and two (i.e. energy and log-energy entropy) wavelet based features are very close to those achieved using all of the features given in Table I.

On the other hand, the effectiveness of different wavelet based features and their combinations affect the computational cost of the classification algorithm. For instance, using nine features at each decomposition level results in 63 features in the feature extraction process which, compared with 14 features, obtained when using only two wavelet based features, will have a significantly increased calculation time. This time–accuracy tradeoff is essential for applications where real-time PQ disturbances classification is needed.

V. CONCLUSION

In this study we investigated the influence of various feature vectors constructed using different wavelet based features and their combinations on the accuracy of the classification of PQ disturbances. The effectiveness of the features was experimentally tested using three different classification methods, SVM, DT and RF, in case of 7 and 11 types of PQ disturbances. With the obtained classification results was shown that in most of the situations when using two or four features, instead of all nine, there is no significant impact on the accuracy, while the calculation efficiency of the classification will be significantly increased.

Our future research will be focused on constructing an optimal feature vector suitable for real-time classification of different types of PQ disturbances with high classification accuracy.

REFERENCES

[1] IEEE Recommended Practice for Monitoring Electric Power Quality, IEEE Std. 1159-1995, 1995.

[2] S. Emmanouil, M.H.J. Bollen, I.Y.H. Gu: “Expert system for classification and analysis of power system events”. IEEE Trans. Power Del., vol. 17, no. 2, pp. 423–428, 2002.

[3] I. Monedero, C. León, J. Ropero, A.García, J.M. Elena, J.C. Montaño: “Classification of electricaldisturbances in real time using neural networks”. IEEE Trans. Power Del., vol. 18, no. 2, pp. 406–1296, 2003.

[4] P.K. Dash, K.S. Mishra, M.M.A. Salama: “Classification of power system disturbances using a fuzzy expert system and a Fourier linear combiner”. IEEE Trans. Power Del., vol. 15, no. 2, pp. 472–477, 2000.

[5] P. Janik, T. Lobos: “Automated classification of power-quality disturbances using SVM and RBF networks.” IEEE Trans. Power Del., vol. 21, no. 3, pp. 1663–1669, 2006.

[6] Dehghani, H., Vahidi, B., Naghizadeh, R., and Hosseinian, S.. Power quality disturbance classification using a statistical and wavelet-based hidden markov model with dempster-shafer algorithm. International Journal of Electrical Power & Energy Systems,47,368–377, 2013.

[7] M.H.J. Bollen, I.Y.H. Gu, Signal processing of power quality disturbances, Wiley –IEEE Press, 2006

[8] Gaing ,Z.L..Wavelet-based neural network for power disturbance recognition and classification. Power Delivery, IEEE Transactionson, 19(4), 1560–1568, 2004.

[9] J.G., Decanini, M.S., Tonelli-Neto, F.C., Malange and ,C.R., Minussi. Detection and classification of voltage disturbances using a fuzzy-artmap-wavelet network. Electric Power Systems Research, 81(12), 2057–2065, 2011.

[10] G. Rata, M. Rata, C. Filote, “Theoretical and experimental aspects concerning Fourier and wavelet analysis for deforming consumers in power network”, Elektronika ir Elektrotechnika (Electronics and Electrical Engineering), no. 1, pp. 62–66, 2010.

[11] L., Breiman. Random Forests. Mach. Learn. 45, 5–32, 2001.

[12] A.Milchevski, D.Taskovski, Classification of power quality disturbances using wavelet transform and SVM decision tree, 11th IEEE Electrical Power Quality and Utilization Conference, 17–19 Oct. 2011, Lisbon.

[13] T. K. A. Galil, M. Kamel, A. M. Youssef, E. F. E. Saadany, and M. M. A. Salama, “Power quality disturbance classification using the inductive inference approach,” IEEE Trans. Power Deliv. vol. 19 No. 10, pp. 1812-1818, 2004.

[14] M. Uyara, S. Yildirima, and M. T. Gencoglub, “An effective wavelet-based feature extraction method for classification of power quality disturbance signals,” J. Elec. Power Syst. Res., vol. 78, pp. 1747-1755, 2008.

[15] S. Upadhyaya, S. Mohanty, Localization and Classification of Power Quality Disturbances using Maximal Overlap Discrete Wavelet

Transform and Data Mining based Classifiers. Elsevier,IFAC-PapersOnLine 49-1, 437–442, 2016.

[16] H., Eri¸sti, Ö., Yıldırım, B., Eri¸sti, Y., Demir. Optimal feature selection for classification of the power quality events using wavelet transform and least squares support vector machines. Int. J. Electr. Power Energy Syst. 49, 95–103, 2013.

[17] B. E. Boser, I. Guyon, V. Vapnik, “A training algorithm for optimal margin classifiers”, in Proc. of the Fifth Annual Workshop on Computational Learning Theory, ACM Press, 1992, pp. 144–152.

[18] P.K. Ray, S.R. Mohanty, N. Kishor, and J.P.S. Catalao, “Optimal feature and decision tree-based classification of power quality disturbances in distributed generation systems,” IEEE Trans. on Sustainable Energy, vol. 5, no. 1, Jan. 2014, pp. 200-208.

[19] T., Li, B., Ni, X., Wu, Q., Gao, Q., Li, D., Sun. On random hyper-class random forest for visual classification. Neurocomputing, 172, 281–289, 2016.