

ON THE CHOICE OF OPTIMAL METHODS FOR FEATURE EXTRACTION AND CLASSIFICATION OF VOLTAGE DISTURBANCES

Marija Markovska, Dimitar Taškovski

*Faculty of Electrical Engineering and Information Technologies,
"Ss. Cyril and Methodius" University in Skopje,
Rugjer Bošković bb, P.O. box 574, 1001 Skopje, North Macedonia
marijam@feit.ukim.edu.mk*

Abstract: Power quality (PQ) disturbances classification plays an essential role in ensuring high quality power supply of the power grid. Therefore, most of the researchers have aimed to achieve accurate and fast classification. But yet one of the main issues is how to extract the “right” features from massive amount of voltage and current data. The feature extraction should be performed for the aim of not only increasing the classification accuracy, but in the same time reducing the computational complexity of the classification algorithm. Once the optimal feature combination is found, the next challenge is to determine the most appropriate number of input signals for training the classifier, so as the classifier achieves maximum accuracy for the given features. In this paper we investigate the effectiveness of the wavelet based features in order to find less numerous features combination that attains high classification accuracy. Furthermore, using that combination, we investigate the influence of the different number of input voltage signals for training the corresponding classifier, in order to obtain optimal classification method.

Key words: classification; feature extraction; power quality

ЗА ИЗБОРОТ НА ОПТИМАЛНИ МЕТОДИ ЗА ИЗВЛЕКУВАЊЕ НА КАРАКТЕРИСТИКИТЕ И КЛАСИФИКАЦИЈА НА РАСПРЕДЕЛБАТА НА НАПОНОТ

Апстракт: Класификацијата на нарушувањата на електричната енергија има важна улога во обезбедувањето на квалитетот на енергијата во енергетската мрежа. Од таа причина повеќето истражувачи имале за цел да постигнат точна и брза класификација на нарушувањата во напонот и струјата. Но сè уште едно од главните прашања е како да се извлечат „правите“ карактеристики од големата количина податоци. Одбирањето на карактеристиките треба да се изврши не само со цел да се зголеми точноста на класификацијата, туку истовремено и да се намали времето кое му е потребно на алгоритмот да ги изврши пресметките. Кога оптималната комбинација на карактеристиките е пронајдена, следниот предизвик е да се одреди најсоодветниот број влезни сигнали потребни за учење на класификаторот, така што за дадените карактеристики да постигне максимална точност во класификацијата. Во овој труд ја истражуваме ефективноста на карактеристиките базирани на везелет, со цел да изнајдеме нивна малобројна комбинација која дава висока точност во класификацијата. Потоа, со употреба на таа комбинација, го истражуваме влијанието на различниот број влезни сигнали за учење на соодветниот класификатор со цел да се добие оптимален метод за класификација.

Клучни зборови: класификација; извлекување на карактеристики; квалитет на електрична енергија

1. INTRODUCTION

As a result of the widespread use of non-linear electrical devices and polluting loads, significant power quality (PQ) disturbances occur in energy systems. The increased interest in PQ analysis is due to the fact that equipment has become less tolerant

of voltage disturbances, production processes have become less tolerant of incorrect operation of the equipment and companies have become less tolerant of production stoppages [1].

PQ measurement and analysis are performed for a number of reasons, but the main one is finding the root cause of equipment malfunction and other

PQ problems. Finding the cause of PQ problem is in many cases the first step in solving and mitigating the problem in the power grid. Therefore, continuous, fast and accurate detection of voltage and current disturbances is of great importance.

The feature selection is the key component for designing an intelligent system for classification because it provides information that helps to recognize the faults responsible for the present disturbances. Practice has shown that even the best classifier will perform poorly if the features are not properly chosen. However, precise and faster feature extraction is still an open issue. Different feature vector construction methods for selection of “better” features among given feature set have been proposed. All these methods introduce additional mathematics and with that, increase the computational complexity of the classification algorithm. Therefore, in this work a research using the comparison “how useful the wavelet based features are with respect to each other” is presented, using all possible combinations of the features. The aim of the investigation is to show that high classification accuracy can be achieved using smaller number of features in the feature extraction process.

After choosing the “right” features, the next step is choosing the optimal number of input voltage signals for training the classifier, in order to trade-off between time complexity and improved performance. Hence, a research about the influence of the number of input voltage signals for training the corresponding classifiers is also made, where ten different numbers of samples from each class are used. Both investigations are made in case of 7 and 11 classes of voltage disturbances, performing support vector machine (SVM) [2], decision tree (DT) [3] and random forest (RF) [4] classification methods. For the aim of the investigations pure voltage signals and signals accompanied with 20 dB, 30 dB, 40 dB and 50 dB of white Gaussian noise are used.

The obtained results have shown that high classification accuracy can be achieved using smaller number of features in the feature vector. Applying the obtained optimal combination, it is shown that the influence of changing the number of training signals is the lowest when RF is used, while there are considerable variations in the accuracies when SVM and DT are used. Even more, it is shown that RF gives high accuracies in both cases of 7 and 11 PQ classes, whether the signal is pure or accompanied with noise. This paper represents an extended version of the research presented in the conference paper [4].

2. FEATURE VECTOR CONSTRUCTION

Several techniques for digital signal processing have been proposed in the area of PQ disturbances classification, such as short-time Fourier transform (STFT), fast working Fourier transform (FT), neural networks, S-transform (ST) and wavelet transform (WT) [6], [7]. Among all mentioned techniques, wavelet transform has been most extensively used in the last years. It is mostly due to the fact that wavelet transform provides not only accurate frequency information, but at the same time accurate time localization of the components, which makes wavelet transform suitable for detecting various deviations in voltage and current signals.

Wavelet transform is representation of a function by wavelets which are scaled and translated copies, known as “daughter wavelets”, of a finite-length oscillating waveform, known as “mother wavelet”. The continuous wavelet transform (CWT) of a signal $x(t)$ is given with the following expression:

$$\text{CWT}(a, b) = \int_{-\infty}^{\infty} x(t)\psi_{ab}(t)dt, \quad (1)$$

where $\psi(t)$ is the mother wavelet, defined as:

$$\psi_{ab}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-a}{a}\right), \quad a, b \in \mathfrak{R}; \quad a \neq 0. \quad (2)$$

In the equation, a is the scale parameter, defined as $|1/\text{frequency}|$ and corresponds to frequency information, while b is the translation parameter, related to the location of the wavelet function shifted through the signal, and corresponds to the time information. Since the transformation is achieved by dilating and translating the mother wavelet continuously, it generates substantial redundant information. Therefore, instead of continuous dilatation and translation, the mother wavelet can be dilated and translated discretely. That can be achieved by selecting the parameters $a = a_0^m$ and $b = nb_0a_0^m$, where a_0 and b_0 are fixed constants with values $a_0 > 1$ and $b_0 > 0$, $m, n \in \mathbb{Z}$, \mathbb{Z} is a set of integers. As a result discrete wavelet transform (DWT) is obtained, defined with the expression:

$$\text{DWT}[m, n] = \frac{1}{\sqrt{a_0^m}} \sum_{-\infty}^{\infty} f[k]\psi\left[\frac{k-nb_0a_0^m}{a_0^m}\right]. \quad (3)$$

With the choice of $a_0 = 2$ and $b_0 = 1$ a dyadic orthonormal WT is achieved and can be easily implemented using the Multi-Resolution Analysis (MRA) techniques [8], [9].

In this work as mother wavelet “Daubechies 4” is used, while the number of decomposition levels is chosen to be 6. The wavelet coefficients obtained by the wavelet analysis of the disturbance signal contain effective feature information. Hence, with the detail coefficients obtained at each decomposition level and the approximation coefficients from the last level, feature vector is constructed. Several techniques for feature extraction have been proposed, using the wavelet based features given in Table 1.

Table 1
Definition of wavelet based features

Labels	Features based on DWT
f_1 Energy	$E_i = \sum_{j=1}^N C_{ij} ^2$
f_2 Mean	$\mu_i = \frac{1}{N} \sum_{j=1}^N C_{ij}$
f_3 Standard deviation	$\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^N (C_{ij} - \mu_i)^2}$
f_4 Skewness	$SK_i = \sqrt{\frac{1}{6N} \sum_{j=1}^N \left(\frac{C_{ij} - \mu_i}{\sigma_i}\right)^3}$
f_5 Shannon entropy	$SE_i = - \sum_{j=1}^N C_{ij}^2 \log(C_{ij}^2)$
f_6 RMS	$rms_i = \sqrt{\frac{1}{N} \sum_{j=1}^N C_{ij}^2}$
f_7 Kurtosis	$KRT_i = \sqrt{\frac{N}{24} \left(\frac{1}{N} \sum_{j=1}^N \left(\frac{C_{ij} - \mu_i}{\sigma_i}\right)^4 - 3 \right)}$
f_8 Log-energy entropy	$LOE_i = \sum_{j=1}^N \log(C_{ij}^2)$
f_9 Norm entropy	$NE_i = \sum_{j=1}^N (C_{ij})^p, 1 \leq p$

For instance, in the researches presented in [10 – [13] feature extraction approaches are proposed using only one wavelet based feature, i.e. in [11] norm entropy is used, while in the others only energy is used. In [14] energy, standard deviation and Shannon entropy are used. In [15] all nine wavelet based features are used.

In the equations, the index $i = 1, 2, \dots, l$ corresponds to the level of decomposition, C_{ij} stands

from the detail coefficients obtained at each decomposition level and the approximation coefficients from the last level, while N is the number of these coefficients. In this work, one of the investigations is the influence of the wavelet based features and their combination on the voltage disturbance classification. The investigation is performed in case of 7 and 11 types of pure voltage signals and signals accompanied with white Gaussian noise. From the large set of results, in Section 4 are presented some of the most accurate.

According to the presented results, a feature combination that gives high classification accuracy and has lower number of features is chosen. This combination is further used for finding an optimally trained classification model.

3. CLASSIFICATION METHODS

Different methods for automated detection and classification of voltage and current disturbances have been proposed in the literature. The most commonly used are the SVM, DT and RF classifiers. Short theoretical overview of these classifiers is given in addition.

3.1. Support vector machine

SVM is a statistical learning technique based on structural risk minimization method. Its main objective is building a model with the use of training set, where each sample belongs to one of the possible classes. For M -dimension inputs $x_i (i = 1, 2, 3, \dots, N)$, N is the number of samples fitting to class 1 or class 2, with outputs $c_i = 1$ for class 1 and $c_i = -1$ for class 2. With the trained SVM, a prediction model can be achieved, which will separate new samples of the given classes using a hyperplane defined as:

$$f_0(x) = \omega \cdot x - b = \sum_{j=1}^M \omega_j \cdot x_j - b = 0. \quad (4)$$

In the definition ω is n -dimensional vector, while b is a parameter, and they define the position of the hyperplane. Their choice should be made in order to maximize the distance between the two parallel hyperplanes, f_1 and f_2 , which are as far apart as possible while still separating the data. These hyperplanes are defined as:

$$\begin{cases} f_1 : \omega \cdot x - b = 1, \\ f_2 : \omega \cdot x - b = -1 \end{cases} \quad (5)$$

The main issue of SVM classifier is choosing a hyperplane which can best separate the two classes. The optimal hyperplane can be determined by the following optimization problem:

$$\text{Minimize } \frac{1}{2} \|\omega\|^2 \quad (6)$$

subject to:

$$c_i(\omega \cdot x_i - b) \geq 1, \text{ for any } i = 1, \dots, N. \quad (7)$$

When the two classes cannot be linearly separated, a soft margin SVM is used [16]. This technique can choose a hyperplane which can still best separate the data. The optimization problem of the soft margin SVM can be expressed as following:

$$\text{Minimize } \frac{1}{2} \|\omega\|^2 \quad (8)$$

subject to:

$$c_i(\omega \cdot \phi(x_i) - b) \geq 1 - \xi_i, \xi_i \geq 0, \quad (9)$$

for any $i = 1, \dots, N$. Variable ξ_i measures the level of misclassification of the feature vectors x_i , while

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \quad (10)$$

is the kernel function, which projects the feature vector into a vector space where all of the features are linearly separable.

3.2. Decision tree

DT approach is widely used in classification based on the choice of features that maximize and fix data division [3]. These features are split into several branches recursively, until termination and classification are reached. Once the classifier is trained, testing new data sets is a simple process. The DT algorithm is constructed using the following expressions:

$$X_i = \{x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{im}\} \quad (11)$$

$$\bar{X} = \{X_1, X_2, \dots, X_m\}^T \quad (12)$$

$$S = \{S_1, S_2, \dots, S_i, \dots, S_m\} \quad (13)$$

In the expressions n is number of independent variables, m is variable observation number, X_i is i -th component of the n -dimension autonomous variables given as $x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{in}$, S is m -dimensional vector of variables forecasted from \bar{X} , and T is a transpose notation vector. For a given vector of observations, different DTs with different accuracy levels can be built. Thus, obtaining an optimal DT through the large space dimensions is challenging. However, there are suitable algorithms which can be

applied on DT and that reflect a trade-off between accuracy and complexity [17]. For instance, a properly chosen sized DT, T_{k0} , can be built on accordance with the following optimization problem:

$$\hat{E}(T_{k0}) = \min_k \{E(T_k)\}, \quad k = 1, 2, 3, \dots, K \quad (14)$$

$$\hat{E}(T) = \sum_{t \in \mathcal{T}} \{e(t)p(t)\} \quad (15)$$

In the equations, $\hat{E}(T)$ represents the level of error in the misclassification of the tree T_k , T_{k0} is the optimal DT that minimize the error of misclassification, T is a binary tree from the set of trees $\{T_1, T_2, \dots, T_k, t_1\}$, where k is the number of trees, t is a tree node with t_1 the root node, $e(t)$ is the re-substitution estimation of the misclassification error in node t , and $p(t)$ is the probability that any case drops into the node t .

3.3. Random forest

RF approach is a representation of tree predictors, such that each tree depends on the values of a random feature vector sampled independently and with same distribution for all trees in the forest. The RF classifier can be defined using the expression:

$$\{f(x, \delta_k), k = 1, \dots\}, \quad (16)$$

where $f(x, \delta_k)$ represents a meta classifier, x is the input vector, while δ_k are k random feature vectors and each of them determines the growth of a single DT. RF summarizes the classification result obtained from the different DTs in order to attain optimal classification result. Classification accuracy of the RF is measured by analyzing the generalization error. This error is an important index for measuring the extrapolation ability of the classifier [4]. For a set of classifiers $\{f_1(x), f_2(x), \dots, f_k(x)\}$ and training set obtained by random sampling of the original data set, (X, Y) , the margin function is defined with the following expression:

$$\begin{aligned} mg(X, Y) &= \\ &= av_k I(f_k(X) = Y) - \max_{j \neq Y} av_k I(f_k(X) = j). \end{aligned} \quad (17)$$

Therein, $I(\cdot)$ is an indicator function, $av_k(\cdot)$ represents an average value, Y is the correct classification of the feature vector, while j is the incorrect classification. The margin function is a measure for the extent to which the average number of votes for the correct class exceeds the average number of votes for any of the other classes. The rule says "The larger the margin, the better the performance of the classifier".

RF is an effective data mining based classifier, especially when dealing with massive amount of data. It has more advantages, compared to the other type of classifiers, such as less number of input parameters, less influence of the over-fitting problem and good anti-noise performance [18]. It can satisfy the demands of situations when real-time performance is highly needed.

4. RESULTS FROM INVESTIGATIONS

The generation of the PQ signals is performed according to the mathematical definitions of nine different PQ disturbances, given in Table 3, using Matlab 8.5. Each signal is sampled with sampling frequency of 3.2 KHz (64 samples/cycle) and from every signal 10 cycles are included, such that each signal is consisted of 640 samples. The fundamental frequency of the signals is assumed to be 50 Hz. Both of the investigations were made is in case of 7 classes of voltage disturbances given in Table 2, which are denoted with the class labels from D1 to D7, and 11 classes given in the same table, which

are denoted with the class labels from D1 to D11. The classes are named as PQ classes.

Table 2

Class labels of the voltage disturbances

PQ classes	Class labels
Normal	D1
Swell	D2
Sag	D3
Harmonic	D4
Interruption	D5
Sag with harmonic	D6
Swell with harmonic	D7
Flicker	D8
Oscillatory transient	D9
Notch	D10
Spike	D11

Table 3

Mathematical definitions for the used voltage disturbances

Disturbance	Model	Parameters
Normal	$x(t) = \sin(\omega t)$	$f = 50 \text{ Hz}, \omega = 2\pi f$
Sag	$x(t) = [1 - \alpha(u(t - t_1) - u(t - t_2))] \sin(\omega t)$	$0.1 \leq \alpha \leq 0.9, T \leq t_2 - t_1 \leq 9T$
Swell	$x(t) = [1 + \alpha(u(t - t_1) - u(t - t_2))] \sin(\omega t)$	$0.1 \leq \alpha \leq 0.8, T \leq t_2 - t_1 \leq 9T$
Interruption/ Outage	$x(t) = [1 - \alpha(u(t - t_1) - u(t - t_2))] \sin(\omega t)$	$0.9 \leq \alpha \leq 1, T \leq t_2 - t_1 \leq 9T$
Flicker	$x(t) = [1 + \alpha \sin(2\pi\beta t)] \sin(\omega t)$	$0.1 \leq \alpha \leq 0.2, 5 \text{ Hz} \leq \beta \leq 20 \text{ Hz}$
Oscillatory transient	$x(t) = \sin(\omega t) + \alpha \exp(-(t - t_1)\tau)(u(t - t_1) - u(t - t_2)) \sin(2\pi f_n t)$	$0.1 \leq \alpha \leq 0.8, 8\text{ms} \leq \tau \leq 40\text{ms}$ $0.5T_n \leq t_2 - t_1 \leq 3T_n,$ $300 \text{ Hz} \leq f_n \leq 900 \text{ Hz}$
Harmonics	$x(t) = \alpha_1 \sin(\omega t) + \alpha_3 \sin(3\omega t) + \alpha_5 \sin(5\omega t) + \alpha_7 \sin(7\omega t)$	$0.05 \leq \alpha_3, \alpha_5, \alpha_7 \leq 0.15, \sum \alpha_i^2 = 1$
Notch	$x(t) = \sin(\omega t) - \text{sgn}(\sin(\omega t)) \left\{ \sum_{k=0}^9 k [u(t - (t_1 + 0.2n)) - u(t - (t_2 + 0.2n))] \right\}$	$0.1 \leq k \leq 0.4, 0 \leq t_1, t_2 \leq 0.5T$ $0.01T \leq t_2 - t_1 \leq 0.05T$
Spike	$x(t) = \sin(\omega t) + \text{sgn}(\sin(\omega t)) \left\{ \sum_{k=0}^9 k [u(t - (t_1 + 0.2n)) - u(t - (t_2 + 0.2n))] \right\}$	$0.1 \leq k \leq 0.4, 0 \leq t_1, t_2 \leq 0.5T$ $0.01T \leq t_2 - t_1 \leq 0.05T$

In order to extend our investigation and to achieve some level of non-stationarity, which is a case in the real power systems, the pure voltage signals are synthesized in noisy environments by

adding evenly distributed white Gaussian noise with signal-to-noise ratio values of 20 dB, 30 dB, 40 dB and 50dB. The obtained results from the investigations are presented in the following subsections.

4.1. Effectiveness of the wavelet based features and their combinations

In the investigation of the effectiveness of the wavelet based features, 1000 signals from each PQ class are used. Its aim was choosing an optimal feature combination used in the feature extraction process, for each classifier.

4.1.1. Support vector machine

SVM classification approach is performed using libSVM implementation for Matlab [19]. Some of the most accurate classification results obtained in case of 7 and 11 PQ classes are given in Table 4 and Table 5, and are graphically presented in Figure 1 and Figure 2, respectively.

From the presented results it is evident that in case of 7 classes after using a combination of two

features up to nine features the classification accuracies are almost identical, weather the PQ signal is pure or accompanied with noise. Accordingly, the combinations of two features gives high accuracy with less number of features and represent optimal feature combination. Hence, the optimal wavelet based feature combination for 7 PQ classes is the combination “f1, f8” or “f5, f8”. In the same way, an optimal feature combination for 11 PQ classes is chosen. It is the combination of three features, i.e. “f1, f5, f7”, “f1, f5, f9” or “f5, f7, f8”.

Using these features additional investigation is made. In Figure 3 and Figure 4 is shown how only these combinations affect the classification accuracy, for different noise levels, in case of 7 and 11 PQ classes, respectively. Due to the fact that in measurements the level of the noise is not fixed, there is a need of feature combination that gives nearly the highest accuracy for each level.

Table 4

On the choice of feature combination using SVM classifier in case of 7 PQ classes

N°	20 dB noise		30 dB noise		40 dB noise		50 dB noise		Pure signal	
	Features (f)	Acc. (%)								
1	5	95.44	5	97.24	5	97.44	5	97.43	5	97.49
2	1,8	98.96	5,8	99.41	5,8	99.51	1,8	99.57	1,8	99.76
3	1,5,8	99.01	1,5,8	99.56	1,5,8	99.59	1,7,8	99.60	1,5,8	99.80
4	1,5,7,8	99.24	1,5,8,9	99.64	1,5,7,8	99.73	1,5,8,9	99.67	1,5,7,8	99.87
5	1,5,6,8,9	99.20	1,5,6,7,8	99.63	1,2,5,6,8	99.70	1,5,6,8,9	99.70	1,5,6,8,9	99.89
6	1,2,3,5,7,8	99.19	1,2,5,6,7,8	99.71	1,2,3,5,7,8	99.74	1,3,5,6,7,8	99.71	1,2,6,7,8,9	99.91
7	1,3,4,5,6,8,9	99.19	1,2,3,5,6,7,8	99.66	1,2,6,5,6,7,8	99.71	1,2,3,5,6,7,8	99.67	1,2,4,5,6,8,9	99.91
8	1,2,3,5,6,7,8,9	99.11	1,2,3,4,5,6,7,8	99.50	1,2,3,5,6,7,8,9	99.67	1,2,3,5,6,7,8,9	99.64	1,2,3,5,6,7,8,9	99.91
9	All nine	99.19	All nine	99.47	All nine	99.56	All nine	99.63	All nine	99.79

Table 5

On the choice of feature combination using SVM classifier in case of 7 PQ classes

N°	20 dB noise		30 dB noise		40 dB noise		50 dB noise		Pure signal	
	Features (f)	Acc. (%)								
1	1	77.18	1	84.36	1	86.09	5	85.76	5	94.85
2	1,5	84.15	1,5	89.21	1,5	91.85	5,7	92.49	5,8	97.24
3	1,5,9	85.12	1,5,9	90.35	1,5,9	92.65	1,5,7	94.04	5,7,8	97.99
4	1,4,5,9	85.12	1,4,5,9	90.34	1,5,7,8	92.88	1,5,7,9	94.91	1,5,7,8	98.13
5	1,4,5,7,9	85.39	1,4,5,7,8	90.48	1,4,5,7,8	92.79	1,5,6,7,9	94.64	1,4,5,7,8	98.09
6	1,2,5,6,7,9	85.55	1,3,4,5,6,9	90.47	1,2,5,6,7,9	93.00	1,4,5,6,7,9	94.75	1,3,4,5,7,8	98.12
7	1,3,4,5,6,7,9	85.51	1,3,4,5,6,7,9	90.56	1,2,3,5,7,8,9	93.58	1,2,3,4,5,6,7	94.56	1,2,3,5,7,8,9	98.12
8	1,2,3,4,5,6,7,9	85.54	1,2,3,4,5,6,7,9	90.59	1,2,3,4,5,6,7,9	93.32	1,2,3,4,5,6,7,9	94.91	1,3,4,5,6,7,8,9	98.04
9	All nine	83.30	All nine	90.63	All nine	93.19	All nine	93.98	All nine	98.11

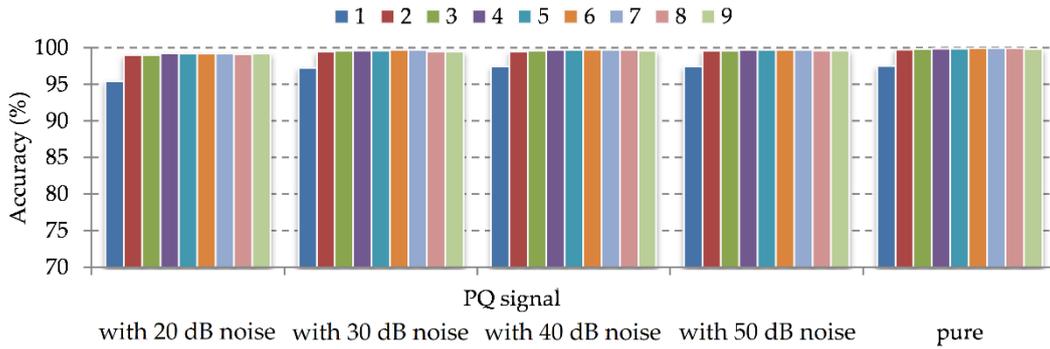


Fig. 1. Effectiveness of the wavelet based features in case of SVM classifier and 7 PQ classes

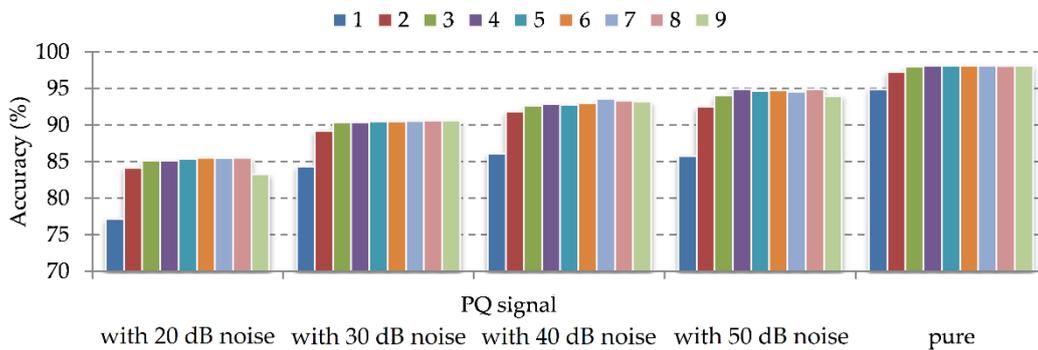


Fig. 2. Effectiveness of the wavelet based features in case of SVM classifier and 11 PQ classes

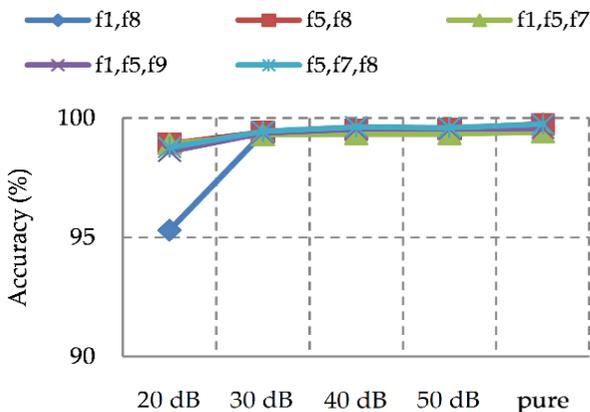


Fig. 3. Effectiveness of the optimal feature combination using SVM classifier at different noise levels, in case of 7 PQ classes

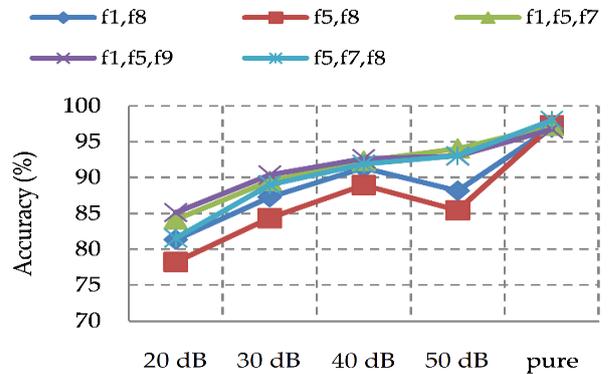


Fig. 4. Effectiveness of the optimal feature combination using SVM classifier at different noise levels, in case of 11 PQ classes

According to the given figures, in case of 7 PQ classes the combination of Shannon entropy and log-energy entropy, (“f5, f8”) is optimal combination. For 11 PQ classes the feature combination energy, Shannon entropy and kurtosis (“f1, f5, f7”) and combination energy, Shannon entropy and norm entropy (“f1, f5, f9”) give almost identical accuracies. Hence, these are the optimal combinations.

4.1.2. Decision tree

The same procedure as in subsection 4.1.1 is performed using DT classification method. The highest results obtained for 7 and 11 PQ classes of voltage disturbances are graphically shown in Figure 5 and Figure 6, respectively. According to the results, the most frequent and least numerous feature combination that gives high accuracies for 7 PQ classes is a combination of two features, i.e. “f5, f8”,

while for 11 classes is the combination of two features, i.e. “f5, f7”, or three features, i.e. “f5, f7, f8” or “f5, f7, f9”. The effective veness of only these

features in case of 7 and 11 PQ classes can be noticed from Figures 7 and 8.

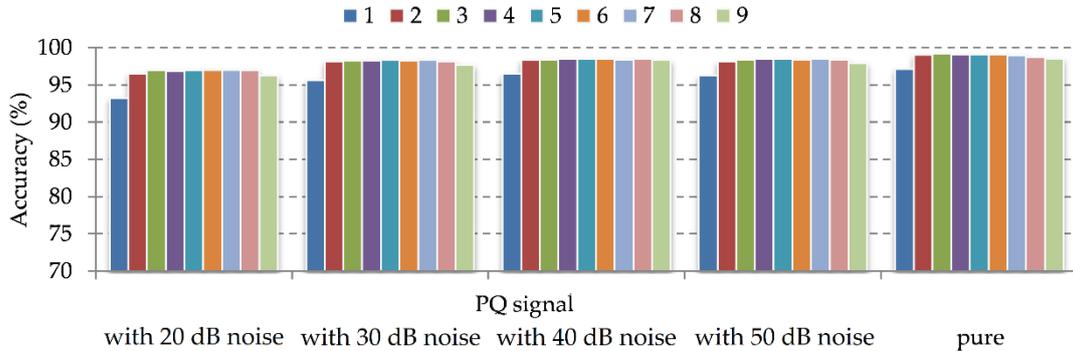


Fig. 5. Effectiveness of the wavelet based features in case of DT classifier and 7 PQ classes

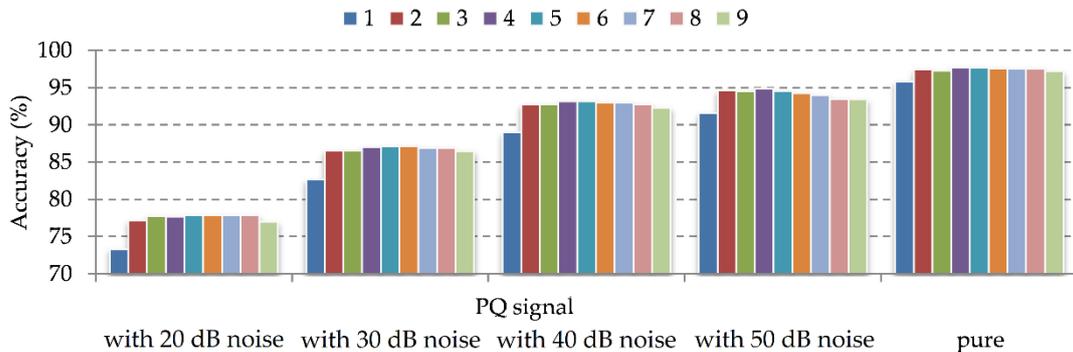


Fig. 6. Effectiveness of the wavelet based features in case of DT classifier and 11 PQ classes

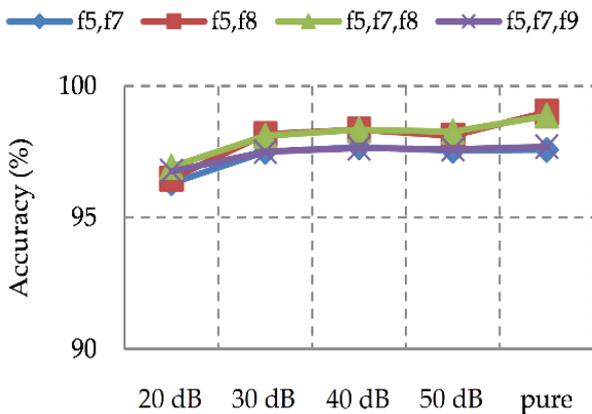


Fig. 7. Effectiveness of the optimal feature combination using DT classifier at different noise levels, in case of 7 PQ classes

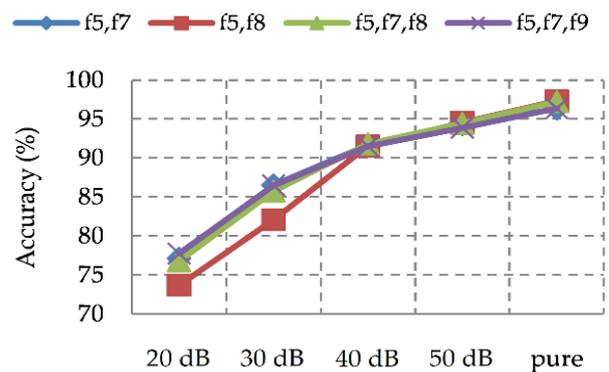


Fig. 8. Effectiveness of the optimal feature combination using DT classifier at different noise levels, in case of 11 PQ classes

From the graphs it is obvious that for 7 classes the combination of Shannon entropy and log-energy entropy (“f5, f8”) is a general optimal combination. For 11 classes the accuracies when two features and

three features are almost equal. Hence, the combination of Shannon entropy and kurtosis is general optimal combination for 11 PQ classes when DT classifier is used.

4.1.3. Random forest

RF classifier is constructed using one hundred trees in the forest. The classification results obtained for 7 and 11 PQ classes are presented in Figure 9 and Figure 10, respectively. According to the presented results, the combinations of two features, i.e. “f1, f8” or “f5, f8”, and three features, i.e. “f3, f5, f7” or “f5, f7, f8”, are used for additional investigation, graphically presented in Figure 11 and Figure 12. From the graphs it is evident that in both cases of 7 and 11 PQ classes the obtained accuracy curves

are almost overlapping. Since the accuracies obtained for the combination “f5, f8” are slightly higher than the accuracies obtained for “f1, f8”, the combination consisted of Shannon entropy and log-entropy is the general optimal combination for the both cases of PQ classes when RF classifier is used.

According to the presented results for the three classifiers, it can be concluded that the most frequent and less numerous general optimal combination is the combination of Shannon entropy and log-entropy, “f5, f8”. Hence, this combination is used for further investigation.

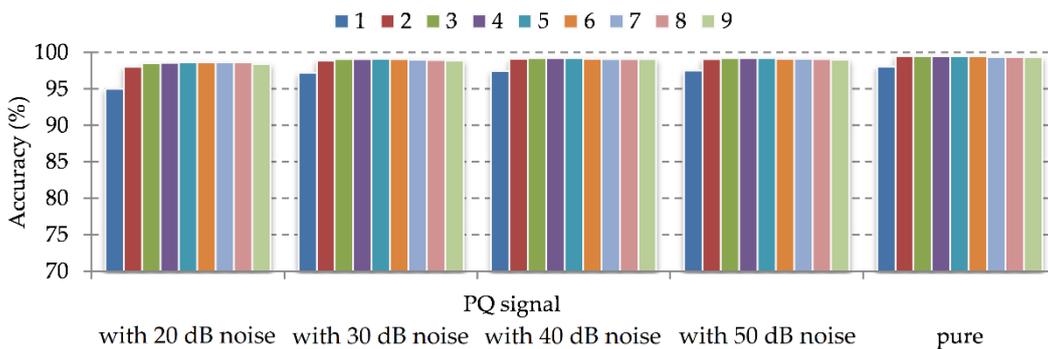


Fig. 9. Effectiveness of the wavelet based features in case of RF classifier and 7 PQ classes

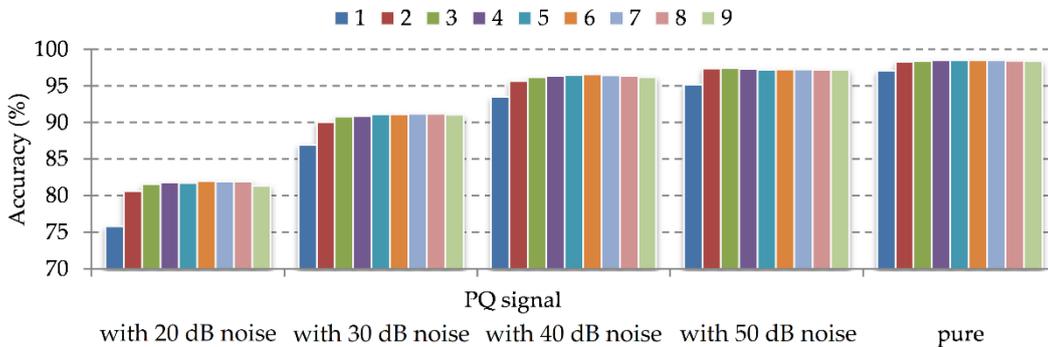


Fig. 10. Effectiveness of the wavelet based features in case of DT classifier and 11 PQ classes

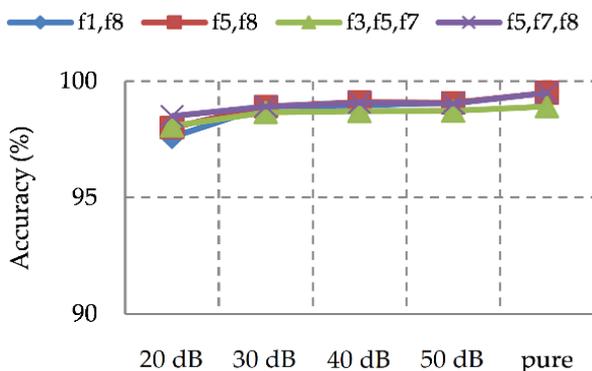


Fig. 11. Effectiveness of the optimal feature combination using RF classifier at different noise levels, in case of 7 PQ classes

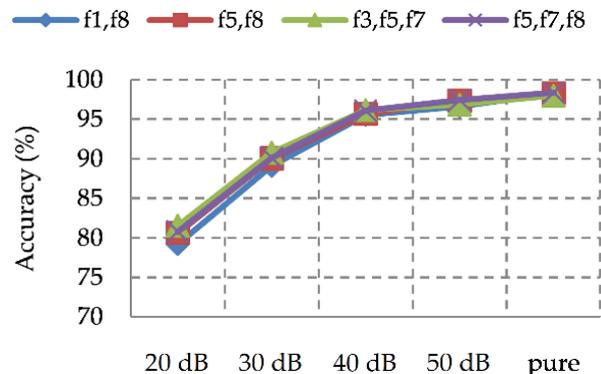


Fig. 12. Effectiveness of the optimal feature combination using RF classifier at different noise levels, in case of 11 PQ classes

With the obtained results it can be noticed that the most effective wavelet based features are those based on entropy. Entropy, as a higher order invariant statistics, is commonly used in information theory and in the area of signal processing [20]. It provides valuable information for analyzing non-stationary signals. Entropy is used as a measure for disorder in the system. Hence, it characterizes the uncertainty in information source that behaves in the voltage and current signal and that property makes it the most effective feature.

4.2. On the choice of the number of training signals for optimal PQ disturbances classification

For the aim of this investigation ten different numbers of voltage signals, from 200 to 2000 per class, with a step of 200 signals are used. The obtained results in case of 7 and 11 PQ classes are presented in the following subsections.

4.2.1. In case of 7 PQ classes

The classification accuracies obtained in case of 7 PQ classes when SVM, DT and RF classifiers are used are shown in Figure 13, Figure 14 and Figure 15, respectively. There is a need of a trained model that gives high accuracies for every noise level.

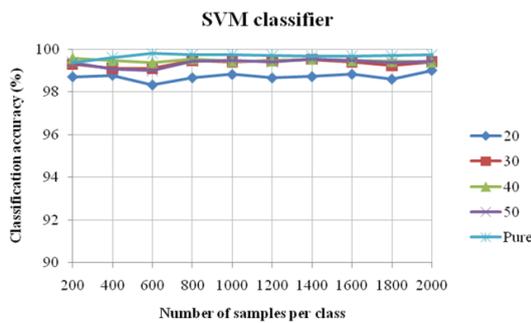


Fig. 13. On the choice of optimal number of power quality signals for training the SVM classifier in case of 7 PQ classes

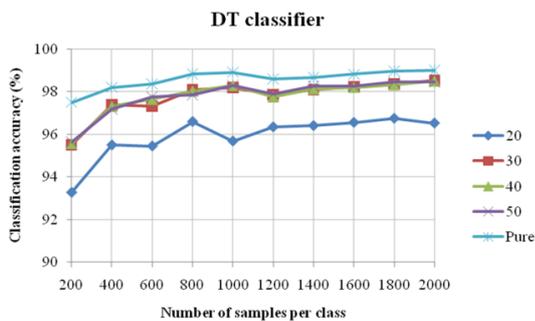


Fig. 14. On the choice of optimal number of power quality signals for training the DT classifier in case of 7 PQ classes

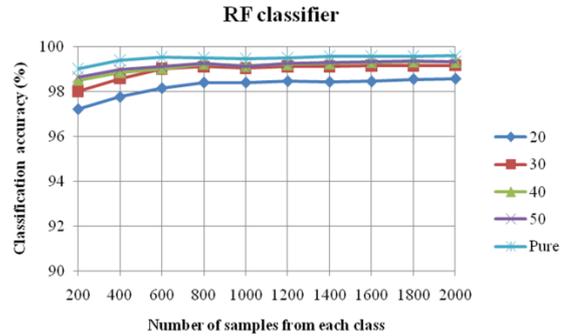


Fig. 15. On the choice of optimal number of power quality signals for training the RF classifier in case of 7 PQ classes

The properly trained classifier should reach maximum accuracy that could be reached for given features, and at the same time the process of training should last as short as it is possible. In accordance with this time-accuracy tradeoff, and in accordance with the presented results, it is evident that the number of 1000 signals per class is an optimal number of input signals for training the SVM, DT and RF classifiers.

4.2.2. In case of 11 PQ classes

The classification accuracies obtained for SVM, DT and RF classifiers, in case of 11 PQ classes, are presented in Figure 16, Figure 17 and Figure 18, respectively.

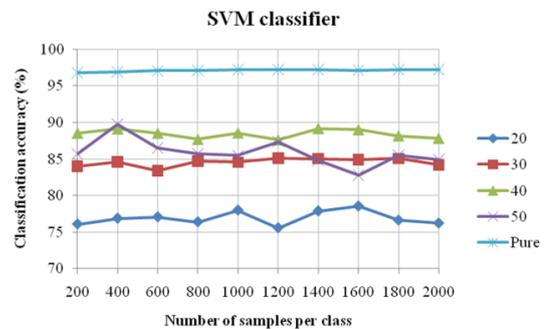


Fig. 16. On the choice of optimal number of power quality signals for training the SVM classifier in case of 11 PQ classes

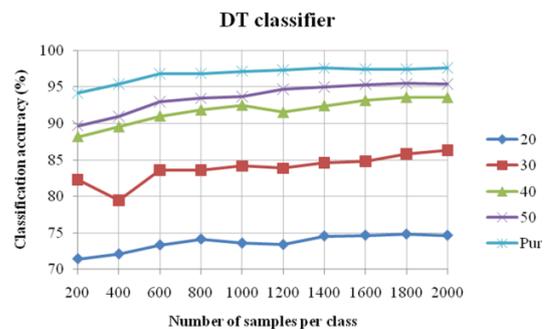


Fig. 17. On the choice of optimal number of power quality signals for training the DT classifier in case of 11 PQ classes

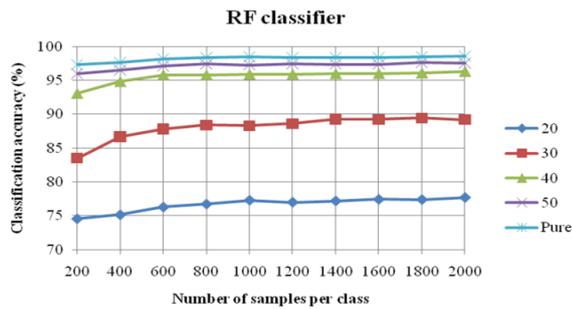


Fig. 18. On the choice of optimal number of power quality signals for training the RF classifier in case of 11 PQ classes

From the presented results it is noticeable that for SVM classifier 400 samples per class is an optimal number of training samples. Although, there are significant variations in the accuracies as this number is changed. For DT it is shown that 1400 samples per class give nearly the highest accuracies for each noise level. The conclusion for RF classifier is the same as that for 7 PQ classes, 1000 samples per class is an optimal number of samples for training this classifier. From the figures it is also evident that the RF classifier provides most stable results. Namely, there are no significant changes in its classification accuracy as the number of input signals is

changed, whether the signal is pure or accompanied with noise, which is not a case with the other classifiers.

4.2.3. Additional comparison of the obtained classification results

Additional comparison of the classification results obtained for 7 and 11 PQ classes is given in Figure 19 and Figure 20, respectively. As it could be seen from the figures, for 7 PQ classes in most of the cases the accuracies of RF are significantly higher than the accuracies of DT and slightly lower than the accuracies of SVM, while for 11 PQ classes RF gives the highest accuracies. Despite the fact that for SVM the optimal number of training samples per class is 400, its accuracy for 30 dB, 40 dB and 50 dB of noise is significantly lower than the accuracy of the RF classification method. In accordance with that, and considering the fact that RF provides most stable results for both number of classes, it can be concluded that RF classification model, trained with 1000 voltage signals per class, represents an optimal method for classification of PQ disturbances.

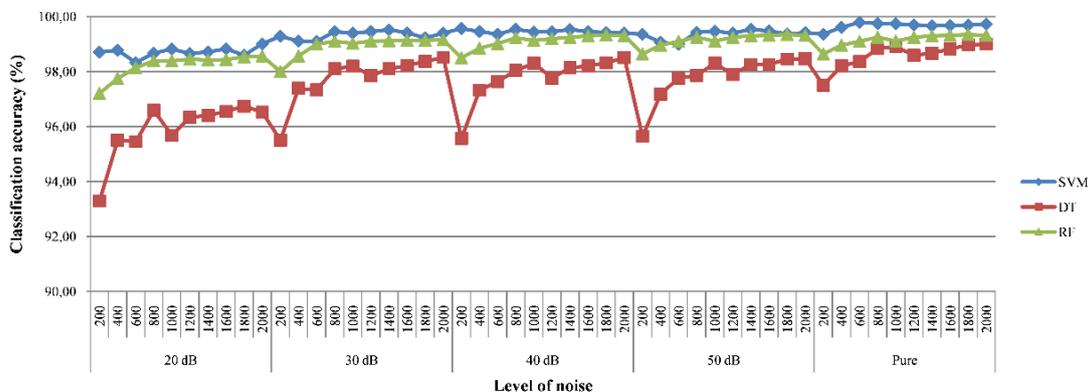


Fig. 19. Comparison of the classification results for 7 PQ classes

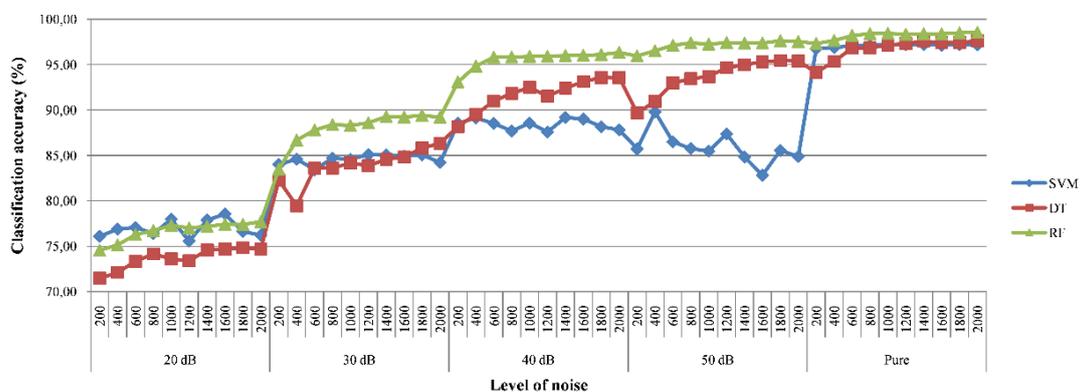


Fig. 20. Comparison of the classification results for 11 PQ classes

5. CONCLUSION

The increased number of polluting loads demands higher quality of power in the generation, transmission and distribution systems. Hence, voltage and currents abnormalities should be monitored and analyzed continuously. The analysis must be able to detect and classify PQ disturbances, where the main issue is how to select the “right” features from a given feature set. Thus, in this work we have investigated the effectiveness of the different wavelet based features and their combinations, as frequently used features in the classification of voltage disturbances. The effectiveness of the features was experimentally tested using three classification methods, SVM, DT and RF, in case of 7 and 11 PQ classes and in presence of different noise levels. The investigation has shown that the effectiveness of a given feature differs depending on the other features it is used with, the noise level present in the signal, the number of PQ classes included in the classification and the type of the used classification method. Related to the noise, in PQ measurements the level of noise is not fixed. In accordance with that, there is a need of feature combination that is less numerous and gives higher or nearly higher accuracies for all of the considered noise levels. With the experimental testing it was shown that in most of the cases when using two wavelet based features, instead of high numerous combination, there is no significant impact on the classification accuracy, while the computational efficiency of the used classification algorithm will be significantly increased. However, it can be noted that if it is expected to have a major presence of noise in the voltage signals, the combination of three features is more reliable choice in case of 11 PQ classes, while the combination of two features will remain to be sufficient for classification of 7 PQ classes with high accuracy.

Furthermore, with the use of the obtained most frequent and less numerous feature combination we investigated the influence of the different number of input voltage signals for training each of the three previously mentioned classifiers. The investigation was also made for 7 and 11 classes of PQ disturbances, in case of pure signals and signals accompanied with different levels of white Gaussian noise. Results obtained from the investigation showed that the change in the number of training signals has the least impact on the classification accuracy of the RF classifier, while in most of the cases there are significant variations in the accuracies of the SVM and DT classifiers. It was also shown that RF gives high accuracies for both numbers of 7 and 11 PQ classes,

regardless of the noise level present in the signal. Thus, RF trained with 1000 samples per class can be considered as an optimal method for classification of voltage disturbances. The aim of our future work will be applying the conclusions obtained in this paper for real-time classification of real voltage signals with high accuracy.

REFERENCES

- [1] Bollen, M. H. J. and Gu, I. Y. H.: *Signal Processing of Power Quality Disturbances*, Wiley – IEEE Press, 2006.
- [2] Janik, P. and Lobos, T.: Automated classification of power-quality disturbances using SVM and RBF networks, *IEEE Trans. Power Del.*, Vol. **21**, No. 3, pp. 1663–1669 (2006).
- [3] Quinlan, J. R.: Induction of decision trees, *Machine Learning*, Vol. **1**, No. 1, pp. 81–106 (1986).
- [4] Breiman, L.: Random Forests, *Machine Learning*, Vol. **45**, No. 1, pp. 5–32 (2001).
- [5] Markovska, M., Tačovski D.: The effectiveness of wavelet based features on power quality disturbances classification in noisy environment, *Proc. of IEEE ICHQP*, pp. 1–6, 2018.
- [6] Decanini, J. G., Tonelli-Neto, M. S., Malange, F. C., Minussi, C. R.: Detection and classification of voltage disturbances using a fuzzy-artmap-wavelet network, *Electric Power Systems Research*, Vol. **81**, No. 12, pp. 2057–2065 (2011).
- [7] Gaing, Z. L.: Wavelet-based neural network for power disturbance recognition and classification, *Power Delivery, IEEE Transaction*, Vol. 2001, No. 4, pp. 1560–1568 (2004).
- [8] Vaidyanathan, P.: *Multirate Systems and Filter Banks*, Prentice-Hall, 1993.
- [9] Strang, G., Nguyen, T.: *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Cambridge, MA, 1996.
- [10] Milčevski, A., Tašovski, D.: Classification of power quality disturbances using wavelet transform and SVM decision tree, *Proc. of IEEE EPQU 2011*, pp. 1–5, 2011.
- [11] Uyara, M., Yildirim, S., Gencoglu, M.T.: An effective wavelet-based feature extraction method for classification of power quality disturbance signals, *Journal of Elec. Power Syst. Res.*, Vol. **78**, No. 10, pp. 1747–1755 (2008).
- [12] Galil, T. K. A., Kamel, M., Youssef, A. M., Saadany, E. F. E., Salama, M. M. A.: Power quality disturbance classification using the inductive inference approach, *IEEE Trans. Power Delivery*, Vol. **19**, No. 4, pp. 1812–1818 (2004).
- [13] He, H., Starzyk, J. A.: A self-organizing learning array system for power quality classification based on wavelet transform, *IEEE Trans. Power Delivery*, Vol. **21**, No. 1, pp. 286–295 (2006).
- [14] Upadhyaya, S., Mohanty, S.: Localization and classification of power quality disturbances using maximal overlap discrete wavelet transform and data mining based classifiers, *Elsevier, IFAC – Papers On Line*, Vol. **49**, No. 1, pp. 437–442 (2016).

- [15] Eristi, H., Yıldırım, Ö., Eristi, B., Demir, Y.: 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.*, Vol. **49**, pp. 95–100 (2013).
- [16] Cortes C., Vapnik, V.: Support-vector network, *Machine Learning*, Vol. **20**, No. 3, pp. 273–297 (1995).
- [17] Ray, P. K., Mohanty, S. R., Kishor, N., Catalao, J. P. S.: Optimal feature and decision tree-based classification of power quality disturbances in distributed generation systems, *IEEE Trans. on Sustainable Energy*, Vol. **5**, No. 1, pp. 200–208 (2014).
- [18] Li, T., Ni, B., Wu, X., Gao, Q., Li, Q. and Sun, D.: On random hyper-class random forest for visual classification, *Neurocomputing*, Vol. **172**, pp. 281–289 (2016).
- [19] A library for support vector machine. <https://www.csie.ntu.edu.tw/~cjlin/libsvm>. (Dec. 2016).
- [20] Shannon, C. E.: A mathematical theory for communication, *Bell Syst. Tech. J.*, Vol. **27**, No. 3, pp. 379–423 (1948).

