Expert Systems with Applications 36 (2009) 10266–10273

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

The search for optimal feature set in power quality event classification

Serkan Gunal [a,*], Omer Nezih Gerek [b], Dogan Gokhan Ece [b], Rifat Edizkan [c]

[a] *Anadolu University, Department of Computer Engineering, Eskisehir, Turkiye*
[b] *Anadolu University, Department of Electrical and Electronics Engineering, Eskisehir, Turkiye*
[c] *Eskisehir Osmangazi University, Department of Electrical and Electronics Engineering, Eskisehir, Turkiye*

ARTICLE INFO

Keywords:
Feature selection
Classification
Power quality event

ABSTRACT

The significance of detection and classification of power quality (PQ) events that disturbs the voltage and/or current waveforms in the electrical power distribution networks is well known. Consequently, in spite of a large number of research reports in this area, the problem of PQ event classification remains to be an important engineering problem. Several feature construction, pattern recognition, analysis, and classification methods were proposed for this purpose. In spite of the extensive number of such alternatives, a research on the comparison of "how useful these features with respect to each other using specific classifiers" was omitted. In this work, a thorough analysis is carried out regarding the classification strengths of an ensemble of celebrated features. The feature items were selected from well-known tools such as spectral information, wavelet extrema across several decomposition levels, and local statistical variations of the waveform. The tests are repeated for classification of several types of real-life data acquired during line-to-ground arcing faults and voltage sags due to the induction motor starting under different load conditions. In order to avoid specificity in classifier strength determination, eight different approaches are applied, including the computationally costly "exhaustive search" together with the leave-one-out technique. To further avoid specificity of the feature for a given classifier, two classifiers (Bayes and SVM) are tested. As a result of these analyses, the more useful set among a wider set of features for each classifier is obtained. It is observed that classification accuracy improves by eliminating relatively useless feature items for both classifiers. Furthermore, the feature selection results somewhat change according to the classifier used. This observation shows that when a new analysis tool or a feature is developed and claimed to perform "better" than another, one should always indicate the matching classifier for the feature because that feature may prove comparably inefficient with other classifiers.

1. Introduction

The necessity of automatic and accurate detection and classification of PQ disturbances arises due to the increasing number of complicated manufacturing processes and delicate electronic devices for almost all purposes and places. Such disturbances, namely power quality (PQ) events, may interrupt working processes or even cause electronic devices to malfunction. The literature is rich in terms of proposals for detection and classification methods for PQ events (Anis Ibrahim & Morcos, 2002; Gerek, Ece, & Barkana, 2006; Hu, Zhu, & Ren, 2008). The classification of a PQ event is as important as its detection because a large class of events is due to the normal operation of power distribution networks, and these events should not cause nuisance tripping of protection equipment in the network. On the other hand, arcing faults (no matter how small the fault current is) correspond to relatively dangerous cases and they must be opened by protective equipment in order to avoid undesirable consequences such as fire in the wiring conduits and complete loss of a delicate load.

To prevent electronic devices from the undesirable effects of the PQ events, the protection system must respond quickly. This may only be accomplished by the careful selection of the classifier and the best matching feature items to that classifier. To further explain, the accuracy and the performance of a classifier depend upon the selected features. With a proper feature selection technique, it is possible to attain higher classification accuracies in low dimensional feature space. Therefore, the feature selection should be performed for the purpose of not only increase the classification accuracy, but also reduce the reaction time of the protection system.

Classical methods for classification of PQ events include two major steps that constitute a pattern recognition process:

- Feature construction from available observations.
- Application of a statistical/semantic/neural classifier to the constructed feature vectors.

* Corresponding author. Tel.: +90 222 3213550x6567; fax: +90 222 3239501.
E-mail address: serkangunal@anadolu.edu.tr (S. Gunal).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.01.051

The step of feature construction corresponds to the analysis of the voltage and/or current waveform, and it is the most thoroughly analyzed part of the PQ event classification problem. Typically, researchers in the area of power engineering apply one of the well-known signal processing analysis tools to the voltage waveform data, observe the visible changes at the output, and comment on the usefulness of the tool according to the observations. Few researchers proceed further to complete the classification process to an automated level by incorporating an available classifier and reporting classification accuracy figures (Angrisani, Daponte, & D'Apuzo, 2001; Gerek et al., 2006; Wang & Mamishev, 2004; Yang & Liao, 2001).

Despite the rich literature regarding the above analysis, comparison of available PQ event classifier performances has not gained its most deserved appreciation in related literature. The lack of a standard test corpus could be a reason for this situation. On the other hand, research reports usually lack comparisons between various feature items among the ones that might be introduced as novel or efficient. Furthermore, the case of utilizing the suggested features with several classifiers and cross comparisons has never been mentioned.

This paper investigates the efficiency characteristics of some well known feature items for PQ event analysis, and checks the variations in efficiency when they are used in combinations. The efficiency analyses are repeated for two well known classifier techniques, namely the Bayes classifier and the support vector machines (SVM). The efficiency tests were carried out by selecting combinations of feature items at various feature vector sizes. Since the efficiency determination methods are not unique in the literature, several available methods are applied for the selection of "better" features among the given set. The methods include

- Sequential forward selection (SFS).
- Sequential backward selection (SBS).
- Generalized sequential forward selection (GSFS).
- Generalized sequential backward selection (GSBS).
- Plus-*l* takeaway-*r* (PTA).
- Sequential forward floating selection (SFFS).
- Genetic algorithm (GA).

Due to the large number of such methods and because of obtaining different results using each method, a solid verification is performed by applying the exhaustive search (ES) method which spans the solutions for all possible combinations at all possible feature length sizes.

By inspecting the results of the search methods, it was observed that the strength of a feature item is not an isolated entity, but it rather depends on what kind of other feature items it is used with and what kind of a classifier is utilized. As an example, the wavelet extrema at a particular decomposition level was found to perform well when used alone, however, it was not found to be efficient when used in pair with another feature; i.e. two other feature items (say, skewness and variance extrema) were found more efficient if exactly "two" feature items should be used.

Although PQ event classification is a well-exploited feature generation application, the counter-intuitive behavior of feature combination according to a classifier is a novel observation.

2. Experimental PQ event generation

In this study, voltage waveforms of real-life PQ events were captured at a sampling rate of 20 kHz using the experimental system whose diagram is shown in Fig. 1.

The system is composed of a three-phase wye-connected 400-V, 50-Hz, 25-kVA, five-wire supply loaded with RL load bank and three-phase induction motors coupled with varying mechanical loads. The system also includes adjustable speed drives (ASD) controlling the induction motors for studying load generated harmonics. Experimental voltage sag events were obtained by starting mechanically loaded induction motors in a controlled way. During the motor starting experiments, mechanical loading was changed between 50% and 100% of rated load. Also, the instantaneous value of the supply voltage at the time of motor starting was naturally random. As a result, various voltage sag levels were obtained and acquired to be used in the proposed algorithm.

Arcing fault events were staged between a phase wire and the ground wire by stripping their insulation for a few millimeters, aligning the stripped parts, and placing several strands high gauge electrical wire between the stripped portions of the phase and ground wires. The arcing faults were initiated in a controlled fashion by turning on the switch connected in series to the phase wire. Once initiated, arcing fault experiments were recorded until the fault clears itself. Due to the randomness of the physical way of preparing the wire samples and randomness of the instantaneous

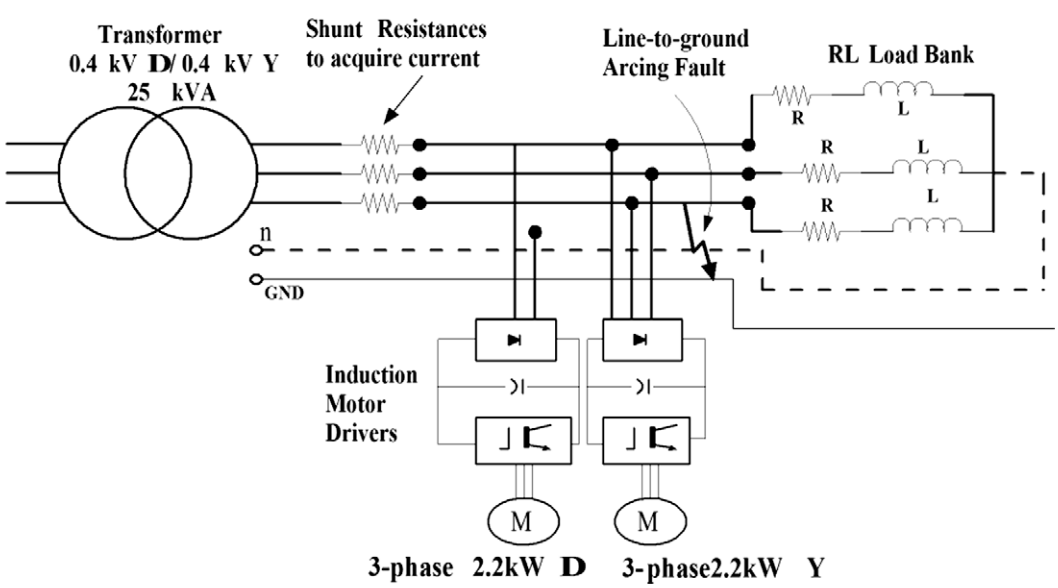

Fig. 1. Experimental system setup.

value of the voltage supplying the fault at the instant of fault initiation, a wide variety of fault sample records was obtained. As an example to the physical randomness of the process, it was observed that while some of the arcing faults restriked several times before clearing itself, others striked once and cleared quickly.

The data acquisition unit consists of an analog-to-digital converter (ADC) unit that was set to perform signal sampling at 20 kHz at each channel simultaneously from four different channels. The data acquisition system also includes programmable digital filters that can be adjusted to perform sharp frequency selective filtering operations in real time. The final classes of PQ events are constructed as:

(1) *Class 1:* Arcing fault with resistive, inductive, and ASD load.
(2) *Class 2:* Arcing fault with resistive and inductive load.
(3) *Class 3:* Motor startup with resistive and inductive load.
(4) *Class 4:* Motor start-up with resistive, inductive, and ASD load.

3. Feature vector construction

The critical stage for any pattern recognition problem is the extraction of discriminative features from the raw observation data. The combination of several scalar features forms the feature vector. The feature vector used in this work consists of scalars obtained from three major and classical methods; wavelets, spectrum analysis, and higher order statistical parameters. The system acquires voltage and current waveforms and their 50 Hz. notch filtered versions at a sampling rate of 20 kHz for each waveform. For local statistical parameter estimation, a window size as twice the fundamental period length (which corresponds to a size of 800 samples) is selected.

The overall feature vector used has a length of 19 (Gerek et al., 2006). The first eight numbers inside the feature vector correspond to the wavelet transform extrema for the four-level decomposition of voltage waveform using the Daubechies-4 (db4) orthogonal wavelet. These four levels depict time-frequency localized signatures at different frequency resolutions. The extrema are, therefore, the maximum and the minimum transform values around the instance of a PQ event. It was previously shown by several authors that the transform domain values exhibit high energy at or around PQ event instances. Usually, a simple thresholding of these coefficient magnitudes is enough to detect the existence of a PQ event. However, for classification between different classes of PQ events, the situation is not simple. In order to verify the validity of the arguments proposed in this work, several other wavelet types and much more number of decomposition levels are tested. Due to the similarity of the results and also due to the lack of space to present hundreds of tables within a single manuscript, only

Table 1
ES method with (a) Bayes and (b) SVM classifier.

Dimension	Accuracy (%)	Selected features
(a)		
1	50.00	18
2	73.33	1, 13
3	78.33	1, 11, 13
4	81.67	1, 8, 11, 13
5	82.50	1, 8, 11, 12, 13
6	83.33	1, 3, 11, 13, 14, 18
7	83.33	1, 7, 11, 12, 13, 18, 19
8	**85.83**	1, 7, 8, 11, 12, 13, 18, 19
9	85.83	1, 3, 7, 8, 11, 13, 14, 18, 19
10	85.00	1, 3, 6, 7, 8, 11, 12, 14, 18, 19
11	85.00	1, 3, 6, 7, 8, 11, 12, 14, 15, 18, 19
12	85.00	1, 2, 3, 6, 7, 8, 10, 11, 12, 14, 18, 19
13	85.83	1, 2, 3, 6, 7, 8, 10, 11, 12, 13, 14, 15, 18
14	83.33	1, 2, 3, 5, 6, 7, 8, 11, 12, 14, 15, 16, 18, 19
15	81.67	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 18
16	80.00	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 17, 18
17	78.33	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 17, 18, 19
18	72.50	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19
19	67.50	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19
(b)		
1	57.50	16
2	70.83	13, 16
3	74.17	14, 16, 18
4	75.83	3, 4, 12, 15
5	75.83	3, 4, 12, 15, 19
6	77.50	3, 12, 14, 15, 16, 18
7	**78.33**	2, 3, 12, 14, 15, 16, 18
8	–	
9	–	**Requires**
10	–	**Too high**
11	–	**Processing time !**
12	–	
13	78.33	1, 2, 3, 5, 6, 7, 8, 11, 12, 14, 15, 18, 19
14	77.50	1, 2, 3, 4, 5, 6, 7, 8, 12, 13, 15, 16, 17, 18
15	77.50	1, 2, 3, 4, 5, 6, 7, 8, 9, 12, 13, 15, 16, 17, 18
16	77.50	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 15, 16, 17, 18
17	75.83	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 15, 16, 17, 18, 19
18	74.17	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19
19	71.67	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19

Table 2
SFS method with (a) Bayes and (b) SVM classifier.

Dimension	Accuracy (%)	Selected features
(a)		
1	50.00	18
2	65.83	17, 18
3	68.33	8, 17, 18
4	70.83	4, 8, 17, 18
5	73.33	4, 8, 14, 17, 18
6	71.67	4, 7, 8, 14, 17, 18
7	73.33	4, 7, 8, 14, 15, 17, 18
8	74.17	3, 4, 7, 8, 14, 15, 17, 18
9	75.83	1, 3, 4, 7, 8, 14, 15, 17, 18
10	78.33	1, 3, 4, 7, 8, 14, 15, 16, 17, 18
11	78.33	1, 3, 4, 6, 7, 8, 14, 15, 16, 17, 18
12	**79.17**	1, 3, 4, 6, 7, 8, 11, 14, 15, 16, 17, 18
13	79.17	1, 3, 4, 6, 7, 8, 11, 14, 15, 16, 17, 18, 19
14	77.50	1, 3, 4, 6, 7, 8, 11, 12, 14, 15, 16, 17, 18, 19
15	77.50	1, 3, 4, 5, 6, 7, 8, 11, 12, 14, 15, 16, 17, 18, 19
16	75.83	1, 2, 3, 4, 5, 6, 7, 8, 11, 12, 14, 15, 16, 17, 18, 19
17	73.33	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 14, 15, 16, 17, 18, 19
18	70.00	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19
19	67.50	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19
(b)		
1	57.50	16
2	70.83	13, 16
3	72.50	3, 13, 16
4	73.33	3, 7, 13, 16
5	73.33	2, 3, 7, 13, 16
6	73.33	2, 3, 7, 8, 13, 16
7	74.17	2, 3, 7, 8, 12, 13, 16
8	74.17	2, 3, 7, 8, 9, 12, 13, 16
9	74.17	2, 3, 7, 8, 9, 11, 12, 13, 16
10	73.33	2, 3, 7, 8, 9, 11, 12, 13, 16, 17
11	72.50	2, 3, 7, 8, 9, 11, 12, 13, 14, 16, 17
12	75.83	1, 2, 3, 7, 8, 9, 11, 12, 13, 14, 16, 17
13	75.83	1, 2, 3, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17
14	**76.67**	1, 2, 3, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17
15	75.83	1, 2, 3, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17
16	75.00	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17
17	73.33	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 19
18	72.50	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19
19	71.67	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19

the case of Db4 wavelet with four decomposition levels are exemplified in the account of wavelet coefficients.

Consequently, the first 8 coefficients of the considered feature vector are constructed as:

$$v_i = \begin{cases} \max\{d_i(n), t_1 - k < n < t_1 + k\}, & i = 1,2,3,4, \\ \min\{d_{i-5}(nt), t_1 - k < n < t_1 + k\}, & i = 5,6,7,8, \end{cases} \quad (1)$$

where d_i corresponds to the ith detail decomposition level of the wavelet transform, t_1 corresponds to the time instance of the PQ event, and k is a time-window around the vicinity of the event in which the extrema are calculated.

After constructing the first 8 coefficients from wavelets, the ninth coefficient of the feature vector was selected according to a classical spectral analysis. The signal energy corresponding to the exact line frequency (50 Hz) is evaluated and proportioned to the remaining spectral energy at all other frequencies. The reciprocal of this proportion is considered as the ninth feature coefficient, which can be expressed as:

$$v_9 = \frac{\int_{-\infty}^{2\pi 50^-} \Phi_v(\omega)d\omega + \int_{2\pi 50^+}^{\infty} \Phi_v(\omega)d\omega}{\Phi_v(\omega)|_{\omega=2\pi 50}}, \quad (2)$$

where $\Phi_v(\omega)$ is the power spectral density of the voltage waveform, $v(t)$.

The remaining ten coefficients are obtained from the extrema of statistical parameters (second, third, and fourth order) of 50 Hz. notch filtered voltage waveforms. Besides the central cumulants for orders 3 and 4, normalized parameters, i.e. the skewness and kurtosis, are also added to the ensemble. Since the extremas have a minimum and a maximum value, these 5 parameters produce a total of 10 coefficients. These vector elements can be compactly expressed as

$$\begin{aligned} v_{10,11} &= \max \text{ and } \min\{c_2\} = \max \text{ and } \min\{\sigma_x^2\}, \\ v_{12,13} &= \max \text{ and } \min\{\widehat{S}\}, \\ v_{14,15} &= \max \text{ and } \min\{\widehat{K}\}, \\ v_{16,17} &= \max \text{ and } \min\{c_3\}, \\ v_{18,19} &= \max \text{ and } \min\{c_4\}. \end{aligned} \quad (3)$$

In these equations, for a signal $x[n]$ of length N and mean, $\hat{m}_x$, the variance is calculated as

$$\sigma_x^2 = \frac{\sum_{i=1}^{N}(x[i] - \hat{m}_x)^2}{N-1} \quad (4)$$

the third central cumulant is

$$c_3 = \frac{\sum_{i=1}^{N}(x[i] - \hat{m}_x)^3}{N-1} \quad (5)$$

and the fourth central cumulant is

$$c_4 = \frac{\sum_{i=1}^{N}(x[i] - \hat{m}_x)^4}{N-1}. \quad (6)$$

The skewness and kurtosis terms are calculated as below

$$\widehat{S} = \frac{\sum_{i=1}^{N}(x[i] - \widehat{m}_x)^3}{(N-1)\sigma_x^3}, \quad (7)$$

$$\widehat{K} = \frac{\sum_{i=1}^{N}(x[i] - \widehat{m}_x)^4}{(N-1)\sigma_x^4} - 3. \quad (8)$$

During all the calculations, it was observed that it is reasonable to keep the parameter estimation window size (N) a small integer multiple (i.e. twice) of the fundamental period length since this is statistically long enough to accurately estimate statistical parameters, and short enough to accurately resolve the time localization.

Table 3
SBS method with (a) Bayes and (b) SVM classifier.

Dimension	Accuracy (%)	Selected features
(a)		
1	30.00	14
2	66.67	13, 14
3	75.00	11, 13, 14
4	75.83	10, 11, 13, 14
5	78.33	3, 10, 11, 13, 14
6	79.17	3, 10, 11, 13, 14, 15
7	82.50	2, 3, 10, 11, 13, 14, 15
8	80.83	2, 3, 6, 10, 11, 13, 14, 15
9	81.67	2, 3, 5, 6, 10, 11, 13, 14, 15
10	82.50	2, 3, 4, 5, 6, 10, 11, 13, 14, 15
11	82.50	2, 3, 4, 5, 6, 10, 11, 13, 14, 15, 19
12	**83.33**	2, 3, 4, 5, 6, 8, 10, 11, 13, 14, 15, 19
13	81.67	2, 3, 4, 5, 6, 7, 8, 10, 11, 13, 14, 15, 19
14	80.00	2, 3, 4, 5, 6, 7, 8, 10, 11, 13, 14, 15, 16, 19
15	77.50	2, 3, 4, 5, 6, 7, 8, 10, 11, 13, 14, 15, 16, 17, 19
16	75.83	2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 19
17	74.17	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19
18	72.50	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19
19	67.50	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19
(b)		
1	50.00	14
2	65.83	14, 18
3	74.17	14, 16, 18
4	73.33	14, 15, 16, 18
5	72.50	12, 14, 15, 16, 18
6	77.50	3, 12, 14, 15, 16, 18
7	76.67	3, 8, 12, 14, 15, 16, 18
8	**78.33**	3, 8, 10, 12, 14, 15, 16, 18
9	77.50	2, 3, 8, 10, 12, 14, 15, 16, 18
10	75.83	2, 3, 5, 8, 10, 12, 14, 15, 16, 18
11	75.00	2, 3, 5, 8, 10, 12, 14, 15, 16, 18, 19
12	73.33	2, 3, 5, 8, 10, 12, 14, 15, 16, 17, 18, 19
13	74.17	2, 3, 5, 8, 10, 12, 13, 14, 15, 16, 17, 18, 19
14	74.17	2, 3, 5, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19
15	75.00	2, 3, 5, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19
16	75.00	2, 3, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19
17	74.17	2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19
18	74.17	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19
19	71.67	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19

Table 4
GSFS method with (a) Bayes and (b) SVM classifier.

Dimension	Accuracy (%)	Selected features
(a)		
2	73.33	1, 13
4	81.67	1, 8, 11, 13
6	**82.50**	1, 5, 8, 11, 12, 13
8	80.00	1, 3, 5, 8, 9, 11, 12, 13
10	80.00	1, 2, 3, 5, 7, 8, 9, 11, 12, 13
12	76.67	1, 2, 3, 5, 7, 8, 9, 10, 11, 12, 13, 18
14	75.00	1, 2, 3, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 18
16	76.67	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 18
18	71.67	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19
19	67.50	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19
(b)		
2	70.83	13, 16
4	73.33	3, 7, 13, 16
6	75.00	3, 7, 13, 14, 16, 17
8	**76.67**	3, 4, 7, 9, 13, 14, 16, 17
10	75.83	3, 4, 7, 8, 9, 13, 14, 15, 16, 17
12	75.83	1, 2, 3, 4, 7, 8, 9, 13, 14, 15, 16, 17
14	75.00	1, 2, 3, 4, 5, 7, 8, 9, 11, 13, 14, 15, 16, 17
16	75.00	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17
18	72.50	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18
19	71.67	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19

4. Feature selection methods

In this work, several methods are tried for the selection of useful and elimination of useless feature identifiers among the described 19-dimensional feature vector. The work gives useful information about

- Which features are, indeed, better identifiers for the engineering problem of PQ event classification?
- Which feature selection strategy is more accurate for the particular case?
- Is there a pronounced relation between the useful features and the desired feature size?
- Is there a pronounced relation between the useful features and the utilized classification method?

Before proceeding with the observations, the feature selection methods used in the study are briefly described below.

4.1. Exhaustive search (ES)

In this selection method, $\binom{N}{d}$ possible feature combinations are analyzed to obtain optimal d dimensional feature subset out of N dimensional full feature set. Although this method guarantees to reach the optimal solution, required processing time is quite high even for moderate number of features.

4.2. Sequential forward selection (SFS)

SFS was first proposed in (Whitney, 1971). It operates in bottom-to-top manner. The selection procedure starts with an empty set initially. Then, at each step, the feature maximizing the criterion function is added to the current set. This operation continues until the desired number of features is selected. The nesting effect is present such that a feature added into the set in a step can not be removed in the subsequent steps. As a consequence, SFS method can offer only suboptimal result.

4.3. Sequential backward selection (SBS)

SBS method proposed in (Marill & Green, 1963) works in a top-to-bottom manner. It is the reverse case of SFS method. Initially, complete feature set is considered. At each step, single feature is removed from the current set so that the criterion function is maximized for the remaining features within the set. Removal operation continues until the desired number of features is obtained. The nesting effect is present in this method as in SFS. Once a feature is eliminated from the set, it can not enter into the set in the subsequent steps. Thus, SBS offers suboptimal solution.

4.4. Generalized sequential forward selection (GSFS)

In generalized version of SFS, instead of single feature, n features are added to the current feature set at each step (Kittler, 1978). The nesting effect is still present.

4.5. Generalized sequential backward selection (GSBS)

In generalized form of SBS (GSBS), instead of single feature, n features are removed from the current feature set at each step (Kittler, 1978). The nesting effect is present here, too.

4.6. Plus-l takeaway-r (PTA)

The nesting effect present in SFS and SBS can be partly avoided by moving in the reverse direction of selection for certain number of steps. With this purpose, at each step, l features are selected using SFS and then r features are removed with SBS. This method

Table 5
GSBS method with (a) Bayes and (b) SVM classifier.

Dimension	Accuracy (%)	Selected features
(a)		
1	25.00	12
3	75.83	7, 11, 12
5	77.50	1, 7, 8, 11, 12
7	81.67	1, 7, 8, 11, 12, 18, 19
9	**85.00**	1, 3, 7, 8, 11, 12, 14, 18, 19
11	85.00	1, 3, 6, 7, 8, 11, 12, 14, 15, 18, 19
13	84.17	1, 2, 3, 6, 7, 8, 11, 12, 13, 14, 15, 18, 19
15	81.67	1, 2, 3, 4, 5, 6, 7, 8, 11, 12, 13, 14, 15, 18, 19
17	78.33	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 17, 18, 19
19	67.50	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19
(b)		
1	48.33	17
3	70.00	4, 13, 17
5	73.33	1, 3, 4, 13, 17
7	73.33	1, 3, 4, 8, 13, 15, 17
9	75.00	1, 3, 4, 6, 8, 12, 13, 15, 17
11	76.67	1, 3, 4, 6, 8, 12, 13, 15, 16, 17, 18
13	**77.50**	1, 2, 3, 4, 5, 6, 8, 12, 13, 15, 16, 17, 18
15	77.50	1, 2, 3, 4, 5, 6, 8, 9, 11, 12, 13, 15, 16, 17, 18
17	75.83	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 15, 16, 17, 18, 19
19	71.67	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19

Table 6
PTA method with (a) Bayes and (b) SVM classifier.

Dimension	Accuracy (%)	Selected features
(a)		
1	48.33	16
2	73.33	1, 13
3	78.33	1, 11, 13
4	78.33	1, 11, 13, 14
5	79.17	2, 11, 13, 14, 15
6	83.33	1, 3, 11, 13, 14, 18
7	80.83	1, 3, 11, 13, 14, 15, 18
8	81.67	1, 3, 11, 12, 13, 14, 15, 18
9	80.83	1, 2, 3, 11, 13, 14, 15, 18, 19
10	**85.00**	2, 3, 5, 6, 11, 13, 14, 15, 18, 19
11	83.33	2, 3, 4, 5, 6, 11, 13, 14, 15, 18, 19
12	84.17	2, 3, 4, 5, 6, 8, 11, 13, 14, 15, 18, 19
13	83.33	2, 3, 4, 5, 6, 8, 10, 11, 13, 14, 15, 18, 19
14	83.33	1, 2, 3, 4, 5, 6, 7, 8, 11, 13, 14, 15, 18, 19
15	81.67	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 13, 14, 15, 18, 19
16	80.00	1, 2, 3, 4, 5, 6, 7, 8, 11, 12, 13, 14, 15, 17, 18, 19
17	78.33	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 17, 18, 19
18	71.67	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19
19	67.50	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19
(b)		
1	57.50	16
2	70.83	3, 16
3	72.50	3, 13, 16
4	73.33	13, 14, 16, 17
5	73.33	1, 3, 4, 13, 17
6	75.83	1, 3, 4, 13, 16, 17
7	75.83	1, 3, 4, 9, 13, 16, 17
8	75.00	1, 2, 3, 4, 6, 13, 16, 17
9	76.67	1, 2, 3, 4, 13, 15, 16, 17, 18
10	76.67	1, 2, 3, 4, 9, 13, 15, 16, 17, 18
11	76.67	1, 2, 3, 4, 8, 12, 13, 15, 16, 17, 18
12	**77.50**	1, 2, 3, 4, 6, 8, 12, 13, 15, 16, 17, 18
13	77.50	1, 2, 3, 4, 6, 7, 8, 12, 13, 15, 16, 17, 18
14	77.50	1, 2, 3, 4, 5, 6, 7, 8, 12, 13, 15, 16, 17, 18
15	77.50	1, 2, 3, 4, 5, 6, 7, 8, 9, 12, 13, 15, 16, 17, 18
16	77.50	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 15, 16, 17, 18
17	75.83	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 15, 16, 17, 18, 19
18	74.17	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19
19	71.67	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19

is called as PTA (Stearns, 1976). Although the nesting effect is reduced with respect to SFS and SBS, PTA still provides suboptimal results.

4.7. Sequential forward floating selection (SFFS)

SFFS is like a dynamic version PTA. In this selection method, l and r parameters float in each step where they are constant in case of PTA (Pudil, Novovicova, & Kittler, 1994). Thus, in each step of selection, different number of features can be added to or removed from the set until a better criterion value is attained. This flexible structure causes the feature dimension to float at each step.

4.8. Genetic algorithm (GA)

Genetic algorithm is a probabilistic search method inspired by the biological evolution process (Goldberg, 1989). The principle of GA is the survival of the fittest solutions among a population of potential solutions for a given problem. Thus, new generations produced by the surviving solutions are expected to provide better approximations to the optimum solution. The solutions correspond to chromosomes that are encoded with an appropriate alphabet. The fitness value of each chromosome is determined by a fitness function. New generations are obtained using genetic operators, crossover and mutation, with certain probabilities on the fittest members of the population. Initial population can be randomly or manually defined. Population size, number of generations, probability of crossover and mutation are defined empirically.

In genetic selection, chromosome length is equal to the dimension of full feature set. The chromosomes are encoded with $\{0, 1\}$ binary alphabet. In a chromosome, the indices represented with "1" indicate the selected features while "0" indicates the unselected ones. For example, a chromosome defined as

$$\{1 \quad 0 \quad 1 \quad 0 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0 \quad 1\} \tag{9}$$

specifies that the features with index 1, 3, 5, 6, and 10 are selected while the others are unselected. The fitness value corresponding to a chromosome is usually defined as the classification accuracy obtained with the selected features. (Huang & Wang, 2006; Siedlecki & Sklansky, 1989; Yang & Honavar, 1998) are some examples of genetic feature selection studies in the literature.

5. Experimental study

The feature selection methods described in previous section are separately applied to the 19-dimensional initial feature vector. The results are first tabulated in Tables 1–8 to provide a complete set of experimental observations in their raw forms. The tables are captioned by the utilized feature selection method with the abbreviations indicated above. For each method, two popular classifiers are applied: Bayesian classifier and support vector machine (SVM). The selection results for the two classifiers are tabulated immediately after one another. In order to show the effect of including (or excluding) a feature, the classification accuracy is also presented where the highest accuracies together with the lowest feature dimensions are displayed in bold for each table.

5.1. Observations about features

Despite the sparse look of the information presented in the tables, some critical observations are possible regarding the success of feature selection methods and the significance of certain features.

The first important observation is regarding the "difference" between the optimization results obtained from Bayes and SVM methods. The results indicate that the choice of classifier has a direct impact on the feature subset obtained. Use of another classifier could also provide different results than ones obtained herein. Therefore, it can be argued that there is no unique solution to the problem of finding the best feature subset for PQ event classification.

The unified consideration of the two classifiers, however, also has clear indications and results about some of the features. As an example, 8D feature subset $\{1, 7, 8, 11, 12, 13, 18, 19\}$ obtained from ES offers the highest classification accuracy (85.83%) with Bayes classifier. Among these eight features, only the three are wavelets while the remaining five features are related to second and fourth order cumulant and skewness. Similarly, 7D feature subset $\{2, 3, 12, 14, 15, 16, 18\}$ provides the highest accuracy (78.33%) in case of SVM classifier. Although, the content of this subset is not identical to one obtained with Bayes classifier, dom-

Table 7
SFFS method with (a) Bayes and (b) SVM classifier.

Dimension	Accuracy (%)	Selected features
(a)		
1	50.00	18
2	70.00	4, 15
3	75.00	4, 11, 12
4	76.67	4, 5, 11, 12
5	78.33	3, 4, 13, 14, 15
6	80.00	3, 4, 11, 13, 14, 15
7	80.83	2, 4, 5, 6, 8, 11, 12
8	80.83	2, 3, 4, 5, 6, 8, 11, 12
9	81.67	2, 3, 4, 5, 6, 8, 11, 12, 13
10	82.50	2, 3, 4, 5, 6, 8, 11, 13, 14, 15
11	83.33	2, 3, 4, 5, 6, 8, 11, 12, 13, 14, 15
12	**84.17**	2, 3, 4, 5, 6, 10, 11, 13, 14, 15, 18, 19
13	83.33	2, 3, 4, 5, 6, 7, 10, 11, 13, 14, 15, 18, 19
14	83.33	1, 2, 3, 4, 5, 6, 7, 8, 11, 13, 14, 15, 18, 19
15	81.67	1, 2, 3, 4, 5, 6, 7, 8, 11, 12, 13, 14, 15, 18, 19
16	80.00	1, 2, 3, 4, 5, 6, 7, 8, 11, 12, 13, 14, 15, 17, 18, 19
17	78.33	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 17, 18, 19
18	72.50	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19
19	67.50	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19
(b)		
1	57.50	16
2	70.83	13, 16
3	72.50	3, 13, 16
4	73.33	3, 7, 13, 16
5	75.00	8, 13, 14, 16, 17
6	76.67	3, 4, 13, 14, 16, 17
7	76.67	3, 4, 11, 13, 14, 16, 17
8	76.67	3, 4, 11, 13, 14, 15, 16, 17
9	76.67	3, 4, 9, 11, 13, 14, 15, 16, 17
10	76.67	3, 4, 9, 11, 13, 14, 15, 16, 17, 19
11	76.67	1, 2, 3, 4, 9, 10, 13, 14, 15, 16, 17
12	**77.50**	1, 2, 3, 4, 6, 8, 12, 13, 15, 16, 17, 18
13	77.50	1, 2, 3, 4, 5, 6, 8, 12, 13, 15, 16, 17, 18
14	77.50	1, 2, 3, 4, 5, 6, 8, 11, 12, 13, 15, 16, 17, 18
15	77.50	1, 2, 3, 4, 5, 6, 7, 8, 11, 12, 13, 15, 16, 17, 18
16	77.50	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 15, 16, 17, 18
17	75.83	1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 15, 16, 17, 18, 19
18	74.17	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19
19	71.67	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19

Table 8
GA method with (a) Bayes and (b) SVM classifier.

Dimension	Accuracy (%)	Selected features
(a)		
13	**85.83**	1, 2, 3, 6, 7, 8, 10, 11, 12, 13, 14, 15, 18
(b)		
12	**77.50**	1, 2, 3, 8, 9, 11, 12, 13, 14, 15, 16, 17

inant features are again the statistical parameters as in Bayes. Based on this outcome, the statistical parameters are arguably more discriminative than the wavelet based features.

When the results of Bayes classification are separately examined for the features selected by ES, it can be easily observed that the classification accuracy constantly improves up to 8D feature subset. During this improvement, dominant features present in the subsets are {1,11,13} corresponding to wavelet, second order cumulant and skewness features, respectively. 3D subset containing these features offers a considerable improvement in accuracy from 50% to approximately 79%. After this point, addition of new features to the subset does not have a significant effect considering the increase in processing time of higher dimensional data. After nine dimensions, new features even degrade the classification accuracy.

Another important point revealed is that even if an individual feature alone seems important, that feature may not be helpful at all when used together with other features. The opposite case may also be true. As an example, although the feature #18 is the best in 1D feature subset, it is not included in 2, 3, 4 and 5D best feature subsets. Consequently, correlations among features are proven to be critical when searching the best set.

5.2. Observations about feature selection methods

If the size of initial feature set is large, exhaustive search may not be feasible due to processing time considerations. In that case, suboptimal selection algorithms should be preferred. However, none of these algorithms guarantee that the best feature set is obtained. This study also compares the outcome of exhaustive search and the other feature selection techniques in terms of the classification accuracies and feature similarities. Thus, the answer to the question "which suboptimal selection method can offer a comparable performance with the exhaustive search" is obtained for the case of PQ event classification.

A weighted score is constructed to evaluate the accuracy of the feature selection method with more weight being given in favor of better accuracy at lower dimensions (Gunal & Edizkan, 2008) according to the following equation:

$$\text{Score} = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{\dim_{\text{total}}}{\dim_i}\right)R_i, \tag{10}$$

where k is the total number of trials, $\dim_i$ is the feature dimension at the ith trial, $\dim_{\text{total}}$ is the total feature dimension and R_i is the average recognition rate achieved at the ith trial. The scores of the tested feature selection methods for both classifiers are given in Table 9. According to the weighted scores obtained by the feature dimension and classification accuracy information, SFFS and PTA methods provide the closest results to the exhaustive search.

In addition to the efficiency of the selection method at lower dimensions, the "similarity" of the selection results between the selection method and the exhaustive search is investigated using the straightforward similarity ratio. This ratio is defined as the percentage of the selected features matching to ones selected by the exhaustive search for regarding feature dimension. The similarity ratio with respect to ES can be formulated as

$$\text{Similarity} = \frac{1}{\dim_{\text{total}}}\sum_{i=1}^{d} m_i m_i = \begin{cases} 0 & f^i_{\text{method}} \notin f_{ES}, \\ 1 & f^i_{\text{method}} \in f_{ES}, \end{cases} \tag{11}$$

where d is the dimension at which the similarity is computed, f^i_{method} is the ith feature of d-dimensional feature subset selected by particular selection method, f_{ES} is d-dimensional feature subset selected by ES method, and $\dim_{\text{total}}$ is the total feature dimension.

The overall similarity ratios for the Bayes and the SVM classifiers are presented in Table 10. According to the average similarities, the most similar feature subsets to the exhaustive search are provided by GSFS method for both classifiers. Although basic SFS and SBS methods are much faster than the others, they could not get close to the exhaustive search in terms of neither accuracy nor similarity due to nesting effect.

Table 9
Weighted scores of the selection methods.

Method	Bayes score	SVM score	Average score
ES	254.25	310.32	282.29
SFS	233.03	243.82	238.42
SBS	223.66	236.20	229.93
GSFS(2)	214.52	204.40	209.46
GSBS(2)	217.92	249.80	233.86
PTA(1,2)	249.19	246.40	247.79
SFFS	246.73	247.20	246.97
GA	125.44	122.71	124.08

Table 10
Similarity ratios (%) of the suboptimal feature selection methods for (a) Bayes and (b) SVM classifier.

Dimension	SFS	SBS	GSFS	GSBS	PTA	SFFS	GA
(a)							
1	100.00	0.00	–	0.00	0.00	100.00	–
2	0.00	50.00	100.00	–	100.00	0.00	–
3	0.00	66.67	–	33.33	100.00	33.33	–
4	25.00	50.00	100.00	–	75.00	25.00	–
5	20.00	40.00	–	80.00	40.00	20.00	–
6	33.33	66.67	50.00	–	100.00	66.67	–
7	28.57	28.57	–	85.71	57.14	28.57	–
8	37.50	25.00	62.50	–	62.50	37.50	–
9	55.56	44.44	–	88.89	77.78	44.44	–
10	60.00	40.00	60.00	–	60.00	50.00	–
11	72.73	54.55	–	100.00	63.64	63.64	–
12	66.67	66.67	75.00	–	66.67	66.67	–
13	69.23	76.92	–	92.31	76.92	76.92	100.00
14	85.71	78.57	78.57	–	85.71	85.71	–
15	80.00	80.00	–	93.33	93.33	93.33	–
16	87.50	87.50	93.75	–	93.75	93.75	–
17	94.12	88.24	–	100.00	100.00	100.00	–
18	94.44	100.00	94.44	–	94.44	100.00	–
19	100.00	100.00	100.00	100.00	100.00	100.00	–
Average	58.44	60.20	81.43	77.36	76.15	62.40	100.00
(b)							
1	100.00	0.00	–	0.00	100.00	100.00	–
2	100.00	0.00	100.00	–	50.00	100.00	–
3	33.33	100.00	–	0.00	33.33	33.33	–
4	25.00	25.00	25.00	–	0.00	25.00	–
5	20.00	40.00	–	40.00	40.00	0.00	–
6	33.33	100.00	50.00	–	33.33	50.00	–
7	57.14	85.71	–	28.57	28.57	42.85	–
8	–	–	–	–	–	–	–
9	–	–	–	–	–	–	–
10	–	–	–	–	–	–	–
11	–	–	–	–	–	–	–
12	–	–	–	–	–	–	–
13	69.23	69.23	–	69.23	69.23	69.23	–
14	85.71	78.57	78.57	–	100.00	92.86	–
15	86.67	80.00	–	93.33	100.00	93.33	–
16	93.75	81.25	93.75	–	100.00	100.00	–
17	94.12	88.24	–	100.00	100.00	100.00	–
18	94.44	94.44	94.44	–	100.00	100.00	–
19	100.00	100.00	100.00	100.00	100.00	100.00	–
Average	70.91	67.32	77.40	53.89	68.18	71.90	–

6. Conclusions

Despite an extensive number of research reports about the PQ event classification, there is still a lack of analysis regarding how well the constructed or proposed features perform given a particular classifier. This work starts with a relatively large feature vector (containing features from wavelets, harmonics, and local statistics), and selects the more useful feature subsets using several feature selection methods. The experiments are repeated for Bayes and SVM classifiers. The results indicate that the selected features are strongly related to the number of desired features and the utilized classifier. Furthermore, there is hardly any significant relation whether a feature that is selected for a vector of N elements should also be selected for a vector of $N+1$ elements, or not.

The work carried out here also contains an information about what sorts of feature selection methods should be adopted in case of computational restrictions – which are almost always the case due to limited computer resources and extensive data sets. The observations are in par with the reports in this field (Kudo & Sklansky, 2000), which state that SFFS and PTA methods are good compromises in terms of good accuracy and fair computational load with respect to ES.

Incorporation of several other features is possible as extensions of this work. However, the observed arbitrary feature efficiency behavior indicates the necessity of a thorough analysis of any new feature that may be possibly introduced in the field of PQ event classification, and the attraction of attention to this matter remains the main scope of this work.

References

Angrisani, L., Daponte, P., & D'Apuzo, M. (2001). Wavelet network-based detection and classification of transients. *IEEE Transactions on Instrumentation and Measurement, 50*(5), 1425–1435.

Anis Ibrahim, W. R., & Morcos, M. M. (2002). Artificial intelligence and advanced mathematical tools for power quality applications: A survey. *IEEE Transactions on Power Delivery, 17*(2), 668–673.

Gerek, O. N., Ece, D. G., & Barkana, A. (2006). Covariance analysis of voltage waveform signature for power-quality event classification. *IEEE Transactions on Power Delivery, 21*(4), 2022–2031.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Reading, MA: Addison-Wesley.

Gunal, S., & Edizkan, R. (2008). Subspace based feature selection for pattern recognition. *Information Sciences, 178*, 3716–3726.

Huang, C.-L., & Wang, C.-J. (2006). A GA-based feature selection and parameters optimization for support vector machines. *Expert Systems with Applications, 31*, 231–240.

Hu, G.-S., Zhu, F.-F., & Ren, Z. (2008). Power quality disturbance identification using wavelet packet energy entropy and weighted support vector machines. *Expert Systems with Applications, 35*, 143–149.

Kittler, J. (1978). Feature set search algorithms. In C. H. Chert (Ed.), *Pattern recognition and signal processing* (pp. 41–60). Mphen aan den Rijn, Netherlands: Sijthoff and Noordhoff.

Kudo, M., & Sklansky, J. (2000). Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition, 33*, 25–41.

Marill, T., & Green, D. M. (1963). On the effectiveness of receptors in recognition system. *IEEE Transactions on Information Theory, 9*, 11–17.

Pudil, P., Novovicova, J., & Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognition Letters, 15*, 1119–1125.

Siedlecki, W., & Sklansky, J. (1989). A note on genetic algorithms for large-scale feature selection. *Pattern Recognition Letters, 10*, 335–347.

Stearns, S. D. (1976). On selecting features for pattern classifiers. In *Proceedings of the third international conference on pattern recognition* (pp. 71–75). Coronado, CA.

Wang, M., & Mamishev, A. V. (2004). Classification of power quality events using optimal time-frequency representations – Part 1: Theory. *IEEE Transactions on Power Delivery, 19*(3), 1488–1495.

Whitney, A. W. (1971). A direct method of nonparametric measurement selection. *IEEE Transactions on Computers, 20*, 1100–1103.

Yang, J., & Honavar, V. (1998). Feature subset selection using a genetic algorithm. *IEEE Intelligent Systems, 13*(2), 44–49.

Yang, H.-T., & Liao, C.-C. (2001). A de-noising scheme for enhancing wavelet-based power quality monitoring system. *IEEE Transactions on Power Delivery, 16*(3), 353–360.