Electric Power Systems Research 80 (2010) 743–752

Contents lists available at ScienceDirect

Electric Power Systems Research

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

Wavelet-based feature extraction and selection for classification of power system disturbances using support vector machines

Hüseyin Erişti[a], Ayşegül Uçar[b], Yakup Demir[b,*]

[a] Tunceli Vocational School, Tunceli University, Turkey
[b] Electrical and Electronics Engineering Department, Engineering Faculty, Firat University, Elazig, Turkey

ARTICLE INFO

Article history:
Received 30 January 2008
Received in revised form 23 January 2009
Accepted 25 September 2009
Available online 8 April 2010

Keywords:
Support vector machines
Classification
Wavelet transform
Feature selection technique
Power system disturbances

ABSTRACT

This paper presents a new approach for the classification of the power system disturbances using support vector machines (SVMs). The proposed approach is carried out at three serial stages. Firstly, the features to be form the SVM classifier are obtained by using the wavelet transform and a few different feature extraction techniques. Secondly, the features exposing the best classification accuracy of these features are selected by a feature selection technique called as sequential forward selection. Thirdly, the best appropriate input vector for SVM classifier is rummaged. The input vector is started with the first best feature and incrementally added the chosen features. After the addition of each feature, the performance of the SVM is evaluated. The kernel and penalty parameters of the SVM are determined by cross-validation. The parameter set that gives the smallest misclassification error is retained. Finally, both the noisy and noiseless signals are applied to the classifier given above stages. Experimental results indicate that the proposed classifier is robust and has more high classification accuracy with regard to the other approaches in the literature for this problem.

1. Introduction

Undesirable cases appearing in consequence of power system disturbances fairly affect electric utilities. These disturbances, such as momentary interruptions, voltage sag/swell, harmonic distortion and transients, resulting in switching, fault, lightning strike, etc. bring about poor power quality (PQ). In particular, PQ can be improved by providing the appropriate solutions if signals representing these disturbances are measured by the equipments located into power system sections and their causes are identified [1].

The basis steps of detecting and classifying the power system disturbances are the pre-processing, feature extraction and classification, respectively. The pre-processing and feature extraction of signals can be performed by direct techniques, such as the RMS value [2] of the raw samples, or transformation techniques, such as the Fourier transform (FT) [3] and the wavelet transform [4–7]. The wavelet transform (WT) is used as a powerful tool for the detection, localization and classification of PQ problems, extracting the distinctive information of the disturbance signals simultaneously in both the time and the frequency domains contrary to the FT. The primary characteristic of WT is that it decomposes a disturbance signal into localized components represented by a scaling and a translating parameter without energy aliasing between them. So, the information of different frequency contained in the disturbance signals is represented by each component individually and high correlation is obtained [8].

Recently, the synergy with the feature extraction techniques of artificial neural networks (ANNs), support vector machines (SVMs) and the other computational intelligence techniques have become popular for solving the problem about the power systems. Spline WT and radial based functional neural networks (RBFNN) have been integrated for PQ data compression in [9]. In [10], wavelet multiresolution analysis (WMRA) and nearest neighbors pattern recognition technique have been proposed for online disturbance classification of different PQ problems. In [11], both FT and WT have been used for extracting features at the proposed fuzzy expert system for classifying PQ disturbances. In [12], the input feature vectors of probabilistic neural networks (PNN), feedforward neural networks (FNN), and RBFNNs have been selected by using both S-transform and WT. In [13], the classification of the disturbances in a real time using ANNs and WT has been presented. ANNs have several important disadvantages such as determining a proper architecture problem, local optimum problem, bad convergence property, over-fit or under-fit problem, etc. although ANNs have very promising pattern recognition and nonlinear function approximation capabilities. On the other hand, SVM classifiers have been receiving a big interesting of power systems researchers because of producing single, optimum and automatic sparse solution by simultaneously minimizing both generaliza-

* Corresponding author. Tel.: +90 424 2370000; fax: +90 424 2415526.
E-mail addresses: ydemir@firat.edu.tr, huseyineristi@gmail.com (Y. Demir).

0378-7796/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.epsr.2009.09.021

tion and training error and separating data by the large margin at high dimensional space [14,15]. In [16], SVM and self-organizing learning array disturbances. In [17], a new approach is presented for impulse fault identification in transformers. In this paper, the WT and the SVM are used for extraction of features from the impulse test response of a transformer in time-frequency domain and classification the patterns inherent in the features extracted. As can (SOLAR) applying only energy feature extraction technique to WT decomposing coefficients have effectively classified the disturbances. As can be seen from these studies, the most of them have usually proposed to use of different classifier for identifying power system disturbances. Features applied as input to classifier have been also extracted using only one feature extraction technique such as mostly energy or entropy. Moreover, these features have not able to provide additional information about the characteristics of disturbances which especially included with noise. As a result, there is a big need to some additional improvements for increasing the classification performance, especially in a noisy environment.

In this paper, an effective classification algorithm based on WT and SVM is proposed for identifying power system disturbances. To build an effective classification algorithm, it is essential to choose robust and adequate features that can recognize the main characteristics of signal and reduce its data size. The main contribution of this paper is to present a novel scheme combined of feature extraction and feature selection for obtaining robust and adequate features of power system disturbances. In this schema, firstly, features are obtained by different extraction techniques to the wavelet coefficients of all decomposition levels of the disturbance signal. Then, by using sequential forward selection (SFS) technique, robust and adequate features are selected in feature set which is obtained from first stage. The experimental results showed that the proposed algorithm could classify with high accuracy both noiseless

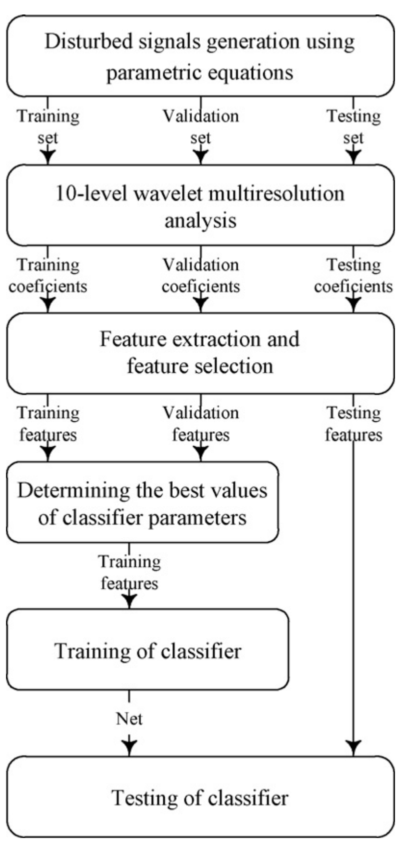


Fig. 1. The block diagram of the proposed classification approach.

and noisy disturbances without the need to apply a de-noising scheme.

The classification scheme of power system disturbances is outlined by the following steps shown in Fig. 1.

(1) In order to evaluate the performance of proposed algorithm, disturbance signals are generated by using the parametric equations because a better generalization performance can be obtained by collecting a few different signals belonging to same class. Taking into consideration normal and all disturbance classes, training, testing and validation sets are produced. (2) By applying WMRA to the signals belonging to these three sets, the decomposition coefficients are obtained. (3) Distinctive features for each signal are extracted by applying eleven different feature extraction methods to detail coefficients belonging to each level and 10th final level approximation coefficient. Then applying the feature selection technique called as sequential forward selection (SFS) to the feature set, the ones revealing the best classification performance of these features are taken aside for using at the classifier. (4) The parameters of SVM classifier are previously selected by cross-validation. By scanning the chosen parameter range, the parameters resulting in the lowest classification error on validation set are determined. (5) The SVM having the determined parameters and selected features is trained. (6) Selected test features are applied to the SVM and the classification process is succeeded.

The paper is organized as follows. WT and MRA are discussed in Section 2. Section 3 contains the proposed feature extraction and selection techniques. A short review to SVM classifiers is presented in Section 4. Experimental results and discussion are given in Sections 5 and 6 concludes this paper.

2. Wavelet transformation theory and multiresolution analysis

In this section, WT is introduced as the feature extraction tool for the proposed SVM classification system. WT can be classically accomplished in three different ways known as the wavelet series, continuous WT and discrete WT. The wavelet series map a function of continuous variable into a sequence of coefficients. In continuous WT, each wavelet is created by scaling and translating operations in a mother wavelet. The mother wavelet is an oscillate function with finite energy and zero average [5]. The continuous WT of a continuous time signal $x(t)$ is defined as

$$\mathrm{CWT}_{\psi}x(a,b)=\int_{-\infty}^{+\infty}x(t)\psi^{*}_{a,b}(t)dt,\quad a,b\in R,\quad a\neq 0 \tag{1}$$

where

$$\psi^{*}_{a,b}(t)=\frac{1}{\sqrt{a}}\psi^{*}\left(\frac{t-b}{a}\right) \tag{2}$$

$\Psi(t)$ is the mother wavelet. a and b are scaling and translating parameters, respectively. The discrete WT is discrete counterpart of continuous WT. In practical applications, the discrete WT of the sampled signal $x(k)$ is replaced by the continuous WT of $x(t)$ such that

$$\mathrm{DWT}_{\psi}x(m,n)=\sum\nolimits_{k}x(k)\psi^{*}_{m,n}(k) \tag{3}$$

where

$$\psi^{*}_{m,n}(k)=\frac{1}{\sqrt{a_0^m}}\psi^{*}\left(\frac{k-nb_0a_0^m}{a_0^m}\right) \tag{4}$$

m and n are scaling and sampling numbers, respectively.

There are many wavelet functions named as mother wavelets. The choice of mother wavelet is important because different types of mother wavelets have different properties. Several popular wavelet functions are *Haar*, *Morlet*, *Coiflet*, *Symlet* and *Daubechies*

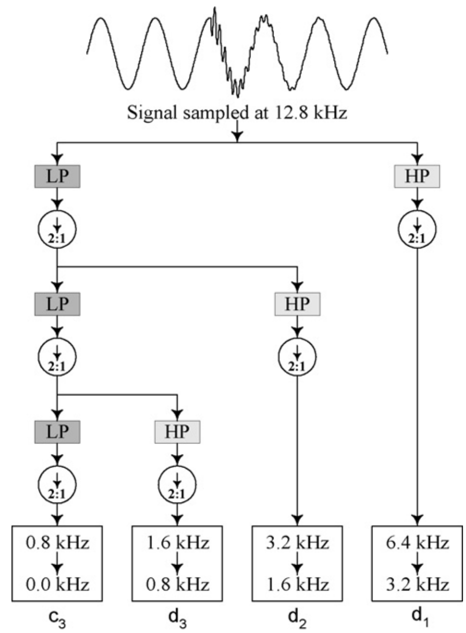


Fig. 2. WMRA decomposition.

wavelets. *Daubechies* wavelets are also well known and widely used in other applications. It is flexible as its order can be controlled to suit specific requirements. Among the different db*N* (*N*-order) wavelets, db4 is the most widely adopted wavelet in power quality applications [18].

The MRA decomposes the original signal into several other signals with different levels of resolution by means of high-pass filters (HP) and low-pass filters (LP) [19]. Fig. 2 shows a three level decomposition dividing a signal sampled at 12.8 kHz into four frequency bands. $\underset{2:1}{\downarrow}$ means down sampling with a factor of 2. c_3 is the approximation level with the lowest frequency band. d_1 to d_3 are respective detail or high frequency bands. The approximation and detail coefficients are given as

$$c_j(n) = \sum_k \mathrm{LP}(k-2n)c_{j-1}(k), \tag{5}$$

and

$$d_j(n) = \sum_k \mathrm{HP}(k-2n)c_{j-1}(k). \tag{6}$$

Thus, the signal is mapped by the following set of coefficients

$$x_{\text{signal}} = [c_j, d_j, d_{j-1}, \ldots, d_1]. \tag{7}$$

3. Dimensionality reduction

In the pattern classification problems, the dimensionality of the pattern representation at the network input is desired to keep small as possible to obtain higher classification accuracy and lower computational load and time. Therefore the feature extraction and feature selection techniques come into prominence for a classification system. Generally, the feature selection processing follows the feature extraction processing. Features are first extracted from the data and then some of the extracted features with low discrimination ability are discarded. On the other hand, depending on the application domain and the available specific training data, only one of the feature selection and feature extraction processes can be established [20].

3.1. Feature extraction

The detail coefficients and approximation coefficients are not directly used as the classifier inputs. In order to reduce the feature dimension, the feature extraction methods are generally implemented to these coefficients at each decomposition level. In this study, all of the method of mean, standard deviation, skewness, kurtosis, RMS, form factor, crest-factor, energy, Shannon-entropy, log-energy entropy and interquartile range given in Table 1 were used as the features extractors. All methods were individually applied to the detail coefficients of each level and the approximation coefficients at 10th level and the features were firstly extracted. Then the obtained features by using each feature extractor were asunder scaled to be having the same mean and standard deviation.

3.2. Feature selection

Another way to lessen the dimensionality of the feature space is to use feature selection methods. The feature selection methods realize the selection of the best subset of the input feature set. To less the large number of disturbance features in this study, SFS technique based on finding the set of features which minimizes the leave-one-out cross-validation (LOOCV) error of a 1-nearest-neighbor (1-NN) function was used. SFS technique operates by starting with an empty set of features and then repeatedly adding the feature that optimizes a given criterion until the set has the desired number of features. In LOOCV, each training example is temporarily deleted from the training data and the 1-NN learning algorithm is applied to predict the class of that example. The total number of classification errors is the LOOCV estimate of the error rate of the learning algorithm.

4. Support vector machines for classification

SVM is a powerful tool for solving pattern classification problems [14,21]. Given the training data $(x_1, y_1), \ldots, (x_l, y_l)$, $\boldsymbol{x} \in R^M$, $y_i \in \{-1, +1\}$ for two class problem, SVM constructs the decision functions of form $sgn((\boldsymbol{w}^T\boldsymbol{x}_i) + w_0)$ by the maximum margin, where $\mathbf{w}$ is the normal vector of the separating hyperplane in the canonical form and w_0 is a bias term [14]. The distances of the point closest to the hyperplanes of both −1 and +1 are calculated as $1/||\boldsymbol{w}||$. The separating margin is defined to be $2/||\boldsymbol{w}||$.

In many practical cases in which data is corrupted by noise, the data may not be separable by a linear hyperplane. To allow to deviations from margin, the slack variables $\xi_i \geq 0$ are introduced:

$$y_i((\mathbf{w}^T\mathbf{x}_i) + w_0) \geq 1 - \xi_i, \quad i = 1, \ldots, \ell. \tag{8}$$

For the training data $\mathbf{x}_i$, if $0 < \xi_i < 1$, the data do not have the maximum margin but are still correctly classified. But if $\xi_i \geq 1$, the data are misclassified by the optimal hyperplane. Thus the separation margin is increased by leaving intramargin the noise points occurring to near the boundaries or outlier points or both, so that generalization performance is improved.

By accordingly, SVM constructs the constraint primal quadratic optimization problem that minimizes the training and generalization error by

$$\min_{\mathbf{w},\xi} \quad \Phi(\mathbf{w}, \xi) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{\ell}\xi_i \tag{9}$$

$$s.t. \quad y_i(\mathbf{w}^T\mathbf{x}_i + w_0) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \ldots, \ell, \tag{9a}$$

where C is a parameter which controls the penalty incurred by each misclassified point in the training set. Generally, larger C values generate SVM models with smaller margin and better training

Table 1
Formulations of feature extraction techniques.

Feature extraction techniques	Feature number (k)	Formulations of detail coefficients (i = 1,2,. . .,10)	Formulations for approximation coefficients
1-Mean	k = 1,. . .,11	$\text{Feat}_i = \mu_d = \frac{1}{N}\sum_{j=1}^{N} d_{ij}$	$\text{Feat}_{11} = \mu_c = \frac{1}{N}\sum_{j=1}^{N} c_{lj}$
2-Standard deviation	k = 12,. . .,22	$\text{Feat}_{i+11} = \sigma_d^2 = \frac{1}{N}\sum_{j=1}^{N}(d_{ij}-\mu_d)^2$	$\text{Feat}_{22} = \sigma_c^2 = \frac{1}{N}\sum_{j=1}^{N}(c_{lj}-\mu_c)^2$
3-Skewness	k = 23,. . .,33	$\text{Feat}_{i+22} = \sqrt{\frac{1}{6N}}\sum_{j=1}^{N}\left(\frac{d_{ij}-\mu_d}{\sigma_d}\right)^3$	$\text{Feat}_{33} = \sqrt{\frac{1}{6N}}\sum_{j=1}^{N}\left(\frac{c_{lj}-\mu_c}{\sigma_c}\right)^3$
4-Kurtosis	k = 34,. . .,44	$\text{Feat}_{i+33} = \sqrt{\frac{N}{24}}\left\{\frac{1}{N}\sum_{j=1}^{N}\left(\frac{d_{ij}-\mu_d}{\sigma_d}\right)^4 - 3\right\}$	$\text{Feat}_{44} = \sqrt{\frac{N}{24}}\left\{\frac{1}{N}\sum_{j=1}^{N}\left(\frac{c_{lj}-\mu_c}{\sigma_c}\right)^4 - 3\right\}$
5-RMS	k = 45,. . .,55	$\text{Feat}_{i+44} = rms_d = \sqrt{\frac{1}{N}\sum_{j=1}^{N} d_{ij}^2}$	$\text{Feat}_{55} = rms_c = \sqrt{\frac{1}{N}\sum_{j=1}^{N} c_{lj}^2}$
6-Form factor	k = 56,. . .,66	$\text{Feat}_{i+55} = \frac{\mu_d}{rms_d}$	$\text{Feat}_{66} = \frac{\mu_c}{rms_c}$
7-Crest-factor	k = 67,. . .,77	$\text{Feat}_{i+66} = \frac{\text{peak}}{rms_d}$	$\text{Feat}_{77} = \frac{\text{peak}}{rms_c}$
8-Energy	k = 78,. . .,88	$\text{Feat}_{i+77} = \sum_{j=1}^{N}\lvert d_{ij}\rvert^2$	$\text{Feat}_{88} = \sum_{j=1}^{N}\left\lvert c_{lj}\right\rvert^2$
9-Shannon-entropy	k = 89,. . .,99	$\text{Feat}_{i+88} = -\sum_{j=1}^{N} d_{ij}^2\log(d_{ij}^2)$	$\text{Feat}_{99} = -\sum_{j=1}^{N} c_{lj}^2\log(c_{lj}^2)$
10-Log-energy entropy	k = 100,. . .,110	$\text{Feat}_{i+99} = \sum_{j=1}^{N}\log(d_{ij}^2)$	$\text{Feat}_{110} = \sum_{j=1}^{N}\log(c_{lj}^2)$
11-Interquartile range	k = 111,. . .,121	$\text{Feat}_{i+110} = d_i(\%75\text{th}) - d_i(\%25\text{th})$	$\text{Feat}_{121} = c_j(\%75\text{th}) - c_j(\%25\text{th})$

accuracy as relatively smaller C values produce larger margin and better generalization accuracy.

To solve the primal problem in (9), Lagrange function is firstly formulated [22]. Then its derivatives with respect to the primal variables $\mathbf{w}$, ξ and w_0 are calculated and KKT conditions are satisfied [22]. Finally the obtained dual optimization problem is solved:

$$\max_{\alpha} \quad Q(\alpha) = -\frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell} y_i y_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j + \sum_{i=1}^{\ell}\alpha_i \tag{10}$$

$$s.t. \quad \sum_{i=1}^{\ell}\alpha_i y_i = 0 \quad 0 \le \alpha_i \le C, \quad i = 1, \ldots, \ell, \tag{10a}$$

where α_i is nonnegative Lagrange multipliers.

SVM maps the inputs $\mathbf{x}$ into some higher dimensional space by means of a nonlinear feature mapping $\varphi(\mathbf{x})$ for solving the classification problem separated by only highly complex decision boundaries in the input space. Thus the problem changes into linearly separable case at the feature space. If only scalar product $\boldsymbol{x}_i^T\boldsymbol{x}_j$ in (10) is replaced by the kernel function $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \varphi^T(\boldsymbol{x}_i)\varphi(\boldsymbol{x}_j)$ assumed to be symmetric and positive definite [14], the dual problem subject to constraints in (10a) is rewritten as

$$Q(\alpha) = -\frac{1}{2}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell} y_i y_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i=1}^{\ell}\alpha_i. \tag{11}$$

$K(\boldsymbol{x}, \boldsymbol{x}_i) = \boldsymbol{x}^T\boldsymbol{x}_i$ in (10) is a linear kernel. One of the most common kernels is Gaussian kernel defined as

$$K(\boldsymbol{x}, \boldsymbol{x}_i) = \exp(-\left\|\boldsymbol{x} - \boldsymbol{x}_i\right\|^2 / 2\sigma^2) \tag{12}$$

where σ is the width parameter of Gaussian function.

The decision surface of SVM is obtained by using only the training data $\mathbf{x}_i$ with $\alpha_i \neq 0$ lying closest to the decision boundary called as support vector:

$$g(\mathbf{x}) = \mathbf{w}^T\varphi(\mathbf{x}_i) + w_0 = \sum_{\alpha_i \neq 0} y_i \alpha_i K(\mathbf{x}, \mathbf{x}_i) + w_0, \tag{13}$$

where threshold can be obtained averaging over unbounded named support vectors with $0 < \alpha_i < C$.

$$w_0 = \frac{1}{|U|}\sum_{i \in U} y_i - \mathbf{w}^T\varphi(\mathbf{x}_i), \tag{14}$$

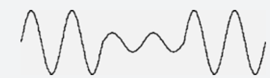
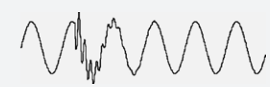
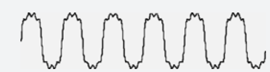
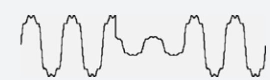
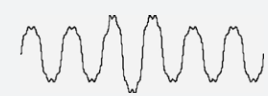
where U is the set of unbounded support vector indices.

Multi-class SVM classifiers can be obtained by combining two class SVMs. The most fundamental approaches are 1-against-r scheme (with r is the number of classes) and the 1-against-1 scheme. 1-against-r scheme users $(r-1)$ two classifiers: each machine is trained as classifier for one class against all other classes. 1-against-1 scheme constructs a multi-class classifier i.e. $(r(r-1)/2)$ two class machines are constructed. Each machine is trained as a classifier for one class against other class. In order to classify test data, pair wise competition between all the machines is performed; each winner competes against another winner until a single winner remains. This final winner determines the class of the test data [14,23].

5. Experimental results

An experimental study was carried out to test the robustness and effectiveness of the proposed classification algorithm. The recognition problem of the power system disturbances was considered as the classification problem with eight classes consisting of normal signal and the disturbance signals called as sag, swell, outage, oscillatory transient, harmonics, sag with harmonic and swell

Table 2
Disturbance signal modelling.

Disturbance types	Equation	Parameter's variation	An example of signal window
Normal (C_1)	$x(t) = A(\sin(Wt))$	–	
Sag (C_2)	$x(t) = A(1 - \alpha_T(u(t-t_1)\ldots - u(t-t_2)))\sin(Wt)$	$P_1 = \left(t_1 < t2; u(t) = \begin{cases} 1, t \geq 0 \\ 0, t < 0 \end{cases}\right)$ $P_2 = (T \leq t_2 - t_1 \leq 9T)$ $P_3 = (0.1 \leq \alpha_T \leq 0.9)$	
Swell (C_3)	$x(t) = A(1 + \alpha_T(u(t-t_1)\ldots - u(t-t_2)))\sin(Wt)$	$P_1; P_2; P_4 = (0.1 \leq \alpha_T \leq 0.8)$	
Outage (C_4)	$x(t) = A(1 - \alpha_T(u(t-t_1)\ldots - u(t-t_2)))\sin(Wt)$	$P_1; P_2; P_5 = (0.9 \leq \alpha_T \leq 1)$	
Oscillatory transient (C_5)	$x(t) = A(\sin(Wt)\ldots + \alpha_{osc}.\exp(-(t-t_b)/\tau_{osc})\ldots\sin(W_{osc}(t-t_b))$	$P_6 = 0.1 \leq \alpha_{osc} \leq 0.8$ $P_7 = 0.008\ \text{s} \leq \tau \leq 0.04\,\text{s}$ $P_8 = 100\,\text{Hz} \leq f_{osc} \leq 400\,\text{Hz}$	
Harmonic (C_6)	$x(t) = A(\alpha_1 \sin(Wt) + \alpha_3 \sin(3Wt)\ldots + \alpha_5 \sin(5Wt) + \alpha_7 \sin(7Wt))$	$P_9 = (0.05 \leq \alpha_{3,5} \leq 0.15)$ $P_{10} = (0.05 \leq \alpha_7 \leq 0.15)$ $P_{11} = \sum \alpha_i^2 = 1$	
Sag with harmonic (C_7)	$x(t) = A(1 - \alpha_T(u(t-t_1) - u(t-t_2)))\ldots(\alpha_1\ \sin(Wt) + \alpha_3 \sin(3Wt) + \alpha_5 \sin(5Wt))$	$P_1; P_2; P_3; P_9; P_{11}$	
Swell with harmonic (C_8)	$x(t) = A(1 + \alpha_T(u(t-t_1) - u(t-t_2)))\ldots(\alpha_1\ \sin(Wt) + \alpha_3 \sin(3Wt) + \alpha_5 \sin(5Wt))$	$P_1; P_2; P_4; P_9; P_{11}$	

with harmonic given in Table 2. The simulation data was generated using MATLAB and the parametric equations in [15,16,24].

Two hundred disturbance signals of each class were randomly generated for training, validating and testing at interval of their control parameters. These signals were sampled at 256 points/cycle and generated for a total of 4096 points which contain the disturbances. Then, in order to analyze the proposed method in different noisy environments, same signals were added by Gaussian white noises being signal to noise ratio 20 and 30 dB, respectively. The mother wavelet used was db4. The power system disturbance signals were analyzed with a 10-level multiresolution decomposition. SVMlight algorithm that is fast training time was used [25]. For SVMlight, the pattern recognition toolbox was used [26]. One against one method was preferred as multi-class classifier.

5.1. Feature selection and training stage of the algorithm

Features obtained from single feature extraction technique are not sufficient to address all possible power system disturbances, as implemented in the literature for automatic classification of power system disturbances [27]. Therefore, this paper focuses using selections of features obtained from different feature extraction techniques. The feature selection process is fulfilled in the training stage of the algorithm. Besides, in the training stage, SVM classifier parameters were determined by using the cross-validation method.

To solve the above defined classification problem, firstly the decomposition coefficients relating to the normal and disturbance signals were obtained by applying WMRA. Two hundred and twenty one features for each signal were extracted by applying 11 different feature extraction methods to detail coefficients belonging to each levels and 10th final level approximation coefficient. Secondly, SFS technique was applied to 121 features. The features resulting with the best classification performance of these features were set off. Thirdly the proposed SVM classifier was constructed by the best proper parameters.

The penalty parameter C and Gaussian kernel parameter σ, of SVM classifier were selected by cross-validation method. For the cross-validation, the pre-determined range of C parameter was taken as 100, 500, 1000, 5000, 10,000, 50,000. The range of kernel parameter was taken as [0.1–15] by steps of 0.1. Classification accuracy of SVM on previously determined validation set was examined. The parameter values producing the lowest classification error on the validation set were determined.

The results of feature selection process applying to the 121 training features were shown in Table 3. In this table, only the first best 20 ones of the selected features were tabulated for both noiseless and noisy signals. The classification results of selected features were evaluated using the SVM classifier. For this purpose, the SVM classifier was trained with training set and tested with validation set. At the beginning, the selected first feature was entered to the input of SVM classifier. As an example, at the proposed classifier for pure signal, the first best feature was determined as the feature 100th. Then the other features were incrementally added to the input set. The results were tabulated in Tables 4–6. The highest accuracies together with the lowest feature dimensions were displayed in bold for each table. The highest accuracies were obtained by 14 features for pure signals, by 15 features for the signals with 30 dB noise and by 15 features for the signals with 20 dB noise.

5.2. Testing stage of the algorithm

In the testing stage of the algorithm, firstly, the decomposition coefficients of the testing set were obtained by applying WMRA. Secondly, the feature vector of noiseless and noisy signals was created by using formulations of the selected features. Then classification results were obtained both noiseless and noisy signals.

Table 3
The first best 20 ones of the selected features.

Feat. order	For Pure Signals		For Signals with 30 dB noise		For Signals with 40 dB noise	
	Feat.Num.	Feature Name	*Feat.Num.*	Feature Name	*Feat.Num.*	Feature Name
1	100	1th level of Log-energy Entropy	8	8th level of Mean	51	7th level of RMS
2	74	8th level of Crest Factor	93	5th level of Shannon Entropy	93	5th level of Shannon Entropy
3	102	3rd level of Log-energy Entropy	41	8th level of Kurtosis	106	7th level of Log-energy Entropy
4	84	7th level of Energy	37	4th level of Kurtosis	16	5th level of Standard Deviation
5	96	8th level of Shannon Entropy	51	7th level of RMS	8	8th level of Mean
6	85	8th level of Energy	74	8th level of Crest Factor	63	8th level of Form Factor
7	101	2nd level of Log-energy Entropy	40	7th level of Kurtosis	52	8th level of RMS
8	103	4th level of Log-energy Entropy	15	4th level of Standard Deviation	81	4th level of Energy
9	18	7th level of Standard Deviation	73	7th level of Crest Factor	84	7th level of Energy
10	95	7th level of Shannon Entropy	52	8th level of RMS	18	7th level of Standard Deviation
11	40	7th level of Kurtosis	84	7th level of Energy	49	5th level of RMS
12	106	7th level of Log-energy Entropy	48	4th level of RMS	85	8th level of Energy
13	41	8th level of Kurtosis	106	7th level of Log-energy Entropy	95	7th level of Shannon Entropy
14	108	9th level of Log-energy Entropy	95	7th level of Shannon Entropy	96	8th level of Shannon Entropy
15	52	8th level of RMS	16	5th level of Standard Deviation	92	4th level of Shannon Entropy
16	8	8th level of Mean	18	7th level of Standard Deviation	82	5th level of Energy
17	104	5th level of Log-energy Entropy	82	5th level of Energy	19	8th level of Standard Deviation
18	51	4th level of RMS	63	8th level of Form Factor	73	7th level of Crest Factor
19	19	8th level of Standard Deviation	96	8th level of Shannon Entropy	30	8th level of Skewness
20	112	2nd level of Interquartile Range	81	4th level of Energy	41	8th level of Kurtosis

Table 4
The classification results of selected features for pure signals.

Dimension	Accuracy (%)	Selected feature
1	69.37	100
2	99.63	100, 74
3	99.69	100, 74, 102,
4	99.75	100, 74, 102, 84,
5	99.75	100, 74, 102, 84, 96
6	99.75	100, 74, 102, 84, 96, 85
7	99.75	100, 74, 102, 84, 96, 85, 101
8	99.75	100, 74, 102, 84, 96, 85, 101, 103
9	99.75	100, 74, 102, 84, 96, 85, 101, 103, 18
10	99.75	100, 74, 102, 84, 96, 85, 101, 103, 18, 95
11	99.69	100, 74, 102, 84, 96, 85, 101, 103, 18, 95, 40
12	99.69	100, 74, 102, 84, 96, 85, 101, 103, 18, 95, 40, 106
13	99.63	100, 74, 102, 84, 96, 85, 101, 103, 18, 95, 40, 106, 41
14	**99.81**	**100, 74, 102, 84, 96, 85, 101, 103, 18, 95, 40, 106, 41,108**
15	99.75	100, 74, 102, 84, 96, 85, 101, 103, 18, 95, 40, 106, 41,108, 52
16	99.69	100, 74, 102, 84, 96, 85, 101, 103, 18, 95, 40, 106, 41,108, 52, 8
17	99.69	100, 74, 102, 84, 96, 85, 101, 103, 18, 95, 40, 106, 41,108, 52, 8, 104
18	99.69	100, 74, 102, 84, 96, 85, 101, 103, 18, 95, 40, 106, 41,108, 52, 8, 104, 51
19	99.75	100, 74, 102, 84, 96, 85, 101, 103, 18, 95, 40, 106, 41,108, 52, 8, 104, 51, 19
20	99.75	100, 74, 102, 84, 96, 85, 101, 103, 18, 95, 40, 106, 41,108, 52, 8, 104, 51, 19, 112

Table 5
The classification results of selected features for the signals with 30 dB noise.

Dimension	Accuracy (%)	Selected feature
1	40.81	8
2	90.75	8, 93
3	95.69	8, 93, 41
4	98.25	8, 93, 41, 37
5	98.19	8, 93, 41, 37, 51
6	98.25	8, 93, 41, 37, 51, 74
7	98.25	8, 93, 41, 37, 51, 74, 40
8	98.50	8, 93, 41, 37, 51, 74, 40, 15
9	98.81	8, 93, 41, 37, 51, 74, 40, 15, 73
10	98.94	8, 93, 41, 37, 51, 74, 40, 15, 73, 52
11	98.87	8, 93, 41, 37, 51, 74, 40, 15, 73, 52, 84
12	98.81	8, 93, 41, 37, 51, 74, 40, 15, 73, 52, 84, 48
13	99.25	8, 93, 41, 37, 51, 74, 40, 15, 73, 52, 84, 48, 106
14	99.19	8, 93, 41, 37, 51, 74, 40, 15, 73, 52, 84, 48, 106, 95
15	**99.31**	**8, 93, 41, 37, 51, 74, 40, 15, 73, 52, 84, 48, 106, 95, 16**
16	99.31	8, 93, 41, 37, 51, 74, 40, 15, 73, 52, 84, 48, 106, 95, 16, 18
17	99.31	8, 93, 41, 37, 51, 74, 40, 15, 73, 52, 84, 48, 106, 95, 16, 18, 82
18	99.31	8, 93, 41, 37, 51, 74, 40, 15, 73, 52, 84, 48, 106, 95, 16, 18, 82, 63
19	99.37	8, 93, 41, 37, 51, 74, 40, 15, 73, 52, 84, 48, 106, 95, 16, 18, 82, 63, 96
20	99.37	8, 93, 41, 37, 51, 74, 40, 15, 73, 52, 84, 48, 106, 95, 16, 18, 82, 63, 96, 81

Table 6
The classification results of selected features for the signals with 20 dB noise.

Dimension	Accuracy (%)	Selected feature
1	45.69	51
2	84.50	51, 93
3	91.06	51, 93, 106
4	96.25	51, 93, 106, 16
5	96.87	51, 93, 106, 16, 8
6	97.00	51, 93, 106, 16, 8, 63
7	97.00	51, 93, 106, 16, 8, 63, 52
8	97.50	51, 93, 106, 16, 8, 63, 52, 81
9	97.44	51, 93, 106, 16, 8, 63, 52, 81, 84
10	97.38	51, 93, 106, 16, 8, 63, 52, 81, 84, 18
11	97.56	51, 93, 106, 16, 8, 63, 52, 81, 84, 18, 49
12	97.44	51, 93, 106, 16, 8, 63, 52, 81, 84, 18, 49, 85
13	97.50	51, 93, 106, 16, 8, 63, 52, 81, 84, 18, 49, 85, 95
14	97.38	51, 93, 106, 16, 8, 63, 52, 81, 84, 18, 49, 85, 95, 96
15	**97.69**	**51, 93, 106, 16, 8, 63, 52, 81, 84, 18, 49, 85, 95, 96, 92**
16	97.56	51, 93, 106, 16, 8, 63, 52, 81, 84, 18, 49, 85, 95, 96, 92, 82
17	97.62	51, 93, 106, 16, 8, 63, 52, 81, 84, 18, 49, 85, 95, 96, 92, 82, 19
18	97.69	51, 93, 106, 16, 8, 63, 52, 81, 84, 18, 49, 85, 95, 96, 92, 82, 19, 73
19	97.96	51, 93, 106, 16, 8, 63, 52, 81, 84, 18, 49, 85, 95, 96, 92, 82, 19, 73, 30
20	97.96	51, 93, 106, 16, 8, 63, 52, 81, 84, 18, 49, 85, 95, 96, 92, 82, 19, 73, 30, 41

Table 7
Classification results of the noisy and noiseless signals.

	For Pure Signals							
Class	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
C_1	200	0	0	0	0	0	0	0
C_2	0	192	0	8	0	0	0	0
C_3	0	0	200	0	0	0	0	0
C_4	0	1	0	199	0	0	0	0
C_5	0	0	0	0	200	0	0	0
C_6	0	0	0	0	0	200	0	0
C_7	0	0	0	0	0	0	200	0
C_8	0	0	0	0	0	0	0	200
	Overall accuracy: Testing = % 99.43 Training = %99.87							
	For Signals with 30 dB							
Class	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
C_1	200	0	0	0	0	0	0	0
C_2	0	191	0	9	0	0	0	0
C_3	0	0	200	0	0	0	0	0
C_4	0	3	0	197	0	0	0	0
C_5	1	0	0	0	199	0	0	0
C_6	0	0	0	0	0	200	0	0
C_7	0	1	0	0	0	2	197	0
C_8	0	0	1	0	0	0	0	199
	Overall accuracy: Testing = % 98.93 Training = %99.75							
	For Signals with 20 dB							
Class	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
C_1	200	0	0	0	0	0	0	0
C_2	0	185	0	14	0	0	1	0
C_3	0	0	200	0	0	0	0	0
C_4	0	2	0	198	0	0	0	0
C_5	6	0	0	0	194	0	0	0
C_6	1	0	0	0	2	197	0	0
C_7	0	2	0	2	0	2	194	0
C_8	0	0	0	0	0	4	0	196
	Overall accuracy: Testing = % 97.75 Training = %98.69							

The classification results of the generated SVM classifier are illustrated in Table 7. The correct classified and misclassified feature numbers at the testing test are tabulated at diagonal axes and the others, respectively. In noiseless signals, sag and outage classes result in misclassified features because of magnitude sag > 10% and magnitude outage < 10% while the other classes are exactly correctly classified. This case appears increasingly for the noisy signals. 30 dB noise exhibits peak noise magnitude of approximately 3.5% of the original signal [28]. Since the peak noise magnitude of 20 dB is greater than that of 30 dB, the classification correctness of 20 dB is slightly less that of 30 dB.

In Fig. 3, the accuracies of the test features belonging to each feature extraction technique are given. The used feature extraction techniques are numbered with 1 of 11 numerals and the chosen features are represented with 0 numeral. By means of these graphics, it can be said that log energy technique is the best accuracy one for noiseless signals but is not robust against with noise, on the other hand, crest-factor, RMS and energy techniques have not only a rather good accuracy but also robust with noise.

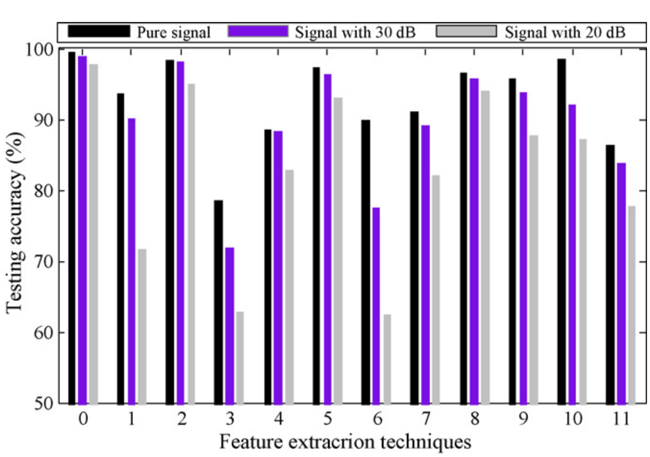


Fig. 3. Classification results for feature extraction techniques.

Fig. 4 shows two-dimensional projections of the training set for best two features of signals. In Fig. 4a, it is seen that the feature set consisting of 1st level log-energy feature (feature 100) and 8th level crest-factor feature (feature 74) depicts a linearly separable case. By feature 100, especially the classes with harmonic components (harmonic, sag with harmonic, swell with harmonic) are exactly separated than the others. By feature 74, three classes are exactly separated among self-classes. It is seen from Table 4 that the classification accuracy of the validation set was obtained as 99.63% by using these two features. This is a promising result.

The chosen best features for the signals with 30 dB noise were 8th level mean feature (feature 8) and 5th level Shannon-entropy feature (feature 93). On the other hand, the chosen best features for the signals with 20 dB noise were 7th level RMS feature (feature 51) and 5th level Shannon-entropy feature (feature 93). In Fig. 4b and c, it is pictured these features. It is seen that sag and outage classes are difficult to linearly separate form each other and that the other classes corrupts into normal signal. It is seen from Tables 5 and 6 when these two features for the signals with 30 and 20 dB noises were used, the classification accuracies of the validation set were obtained as 90.75 and 84.50%, respectively. These results are very reasonable for noisy signals.

5.3. Discussion

In this paper, a discussion about the performance of proposed algorithm is provided; the discussion considers the following crite-

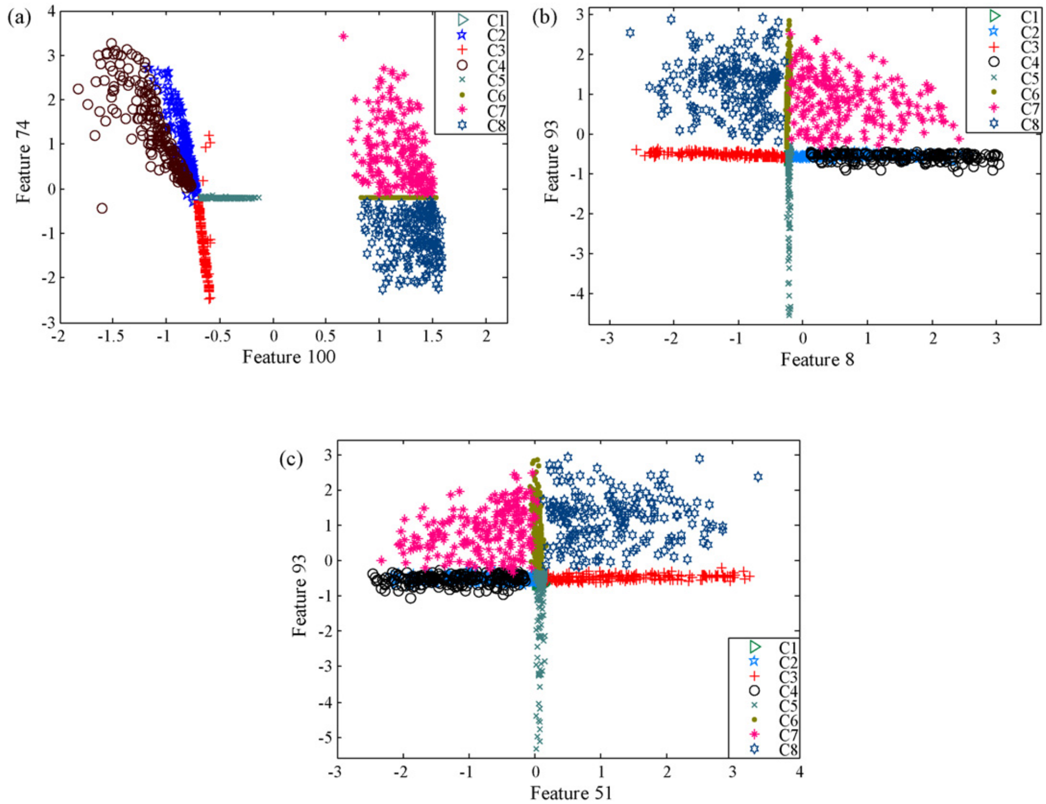


Fig. 4. Two-dimensional projections of selected best two features for (a) pure signal, (b) signal with 30 dB noise and (c) signal with 20 dB noise.

ria: the comparison of classification accuracy, computational effort and practical realization.

Taking into consideration the given results in Tables 4–6, the performance of the proposed classifier can be evaluated comparatively with the obtained results in [16], [24] and [29]. In these papers, the disturbance signals were generated by same parametric equations used in our paper. Besides, only one feature extraction technique such as energy technique and entropy technique was used in the classification algorithm.

In [16], the proposed algorithm is consisted of the WT, the energy feature extraction technique and the SOLAR and the SVM classifier. The test accuracies for pure signal and signals with 30 dB and 20 dB noises are obtained as 94.93%, about 92% and about 93%, respectively.

In [24], the feature vector of classification algorithm is created by using the WT, the energy feature extraction technique and the decision tree. The test accuracy for pure signal was obtained as 90.4%.

In [29], the wavelet packet transform, the entropy feature extraction technique and the weighted SVM are used to classify each type of disturbance. The classification result of the algorithm for pure signal was obtained as 98.4%.

On the other hand, in this paper, feature vector is obtained from applying SFS technique and eleven different feature extraction methods. Thus, the distinctive features having high classification performance are applied the SVM classifier. The test accuracies are obtained as 99.43%, 98.63% and 97.75% for pure signal and signals with 30 dB and 20 dB noises respectively. Also, when the obtained results by each feature extraction technique in Fig. 3 are checked over, it is clearly seen that the proposed approach in this paper classifiers effectively the power system disturbances with different type and have a robust structure.

In the proposed algorithm, while data size of a disturbance is 4096, its data size in the classifier input is reduced to 14 for noiseless and 15 noisy signals. Furthermore, it can reduce memory space and increase computation speed of the SVM classifier.

A wavelet with small support width is fast to compute. Besides, more levels of decomposition will increase the computational cost [16]. We choose db4 wavelet and scan the decomposition levels from 1 to 10. The numbers of selected features at frequency band levels are illustrated in Table 8. As it can be seen from Table 5, there are more features at level 4, level 5 and levels 7–8 that are main frequency levels, The db4 wavelet with short wavelet filter length and the 10-level of decomposition provide a great potential for disturbance classification in terms of both nice classification results and small computational cost. Moreover, in order to further decrease the computation cost, it can be chosen the db2 wavelet with smaller width and 6-level of decomposition for sampling rate with 3200 Hz. But, these parameters decrease the classification performance. Since the feature selection and cross-validation processes are

Table 8
Feature numbers at the frequency band levels.

Level	1	2	3	4	5	6	7	8	9	10	Approx.
Frequency division (Hz)	6400	3200	1600	800	400	200	100	50	25	12.5	6.25
	3200	1600	800	400	200	100	50	25	12.5	6.25	0
Pure signal	1	1	1	1			5	4	1		
With 30 dB				3	2		6	5			
With 20 dB				2	3		5	5			

implemented the training stage of the algorithm, they increase only the training computation cost. The approach of hardware prototyping for the realization of the proposed algorithm has great advantages because SVM and WMRA is computationally simple, accurate and exhibits a good balance of flexibility, speed, size and design cycle time. As a result of this discussion, the proposed algorithm has the great potential for the future development of fully automated monitoring systems with online classification capabilities.

6. Conclusions

This paper considers a new approach to classifying the power system disturbances, relying on the SVM classifier, WT, and SFS. As different form the well-known studies using only one or two feature extraction techniques in the literature, the distinctive features of all signals are determined by applying eleven feature extraction techniques to the WT decomposition coefficients at each level relating to the normal and distorted signal. The only a few features producing the best classification accuracy of obtained features are selected by SFS technique. By accordingly the proposed classifier needs less memory space and less computing time at both the training and testing processes. Moreover, the experimental results show that the proposed classifier has effectively a classification capability the power system disturbances under the conditions with both noisy and noiseless.

In addition, the obtained results show that proposed approach can give rise to new algorithms for the analysis of the various large data sets about power or biological systems having high dimensional thanks to the proposed feature extracting and selecting procedures.

Appendix A. List of symbols

Wavelets

t	time
k	integer-valued variable
$x(t)$	continuous time signal
$x(k)$	discrete time signal
$\Psi(t)$	mother wavelet function
a	scaling parameter
b	translating parameter
m	scaling number
n	sampling number
j	decomposition level
c_j	approximation coefficient
d_j	detail coefficient
N	detail (or approximation) coefficient number
LP	low-pass filter
HP	high-pass filter

Support vector machines

x_i	ith M-dimensional training data
M	number of input variables
y_i	class label 1 or $\subseteq$1 for input $\mathbf{x}_i$
l	number of training data
$\mathbf{w}_i$	coefficient vector of the ith hyperplane
$\|\|\mathbf{w}\|\|$	Euclidean norm of vector $\mathbf{w}$
$\mathbf{w}^T$	transpose of matrix $\mathbf{w}$
w_0	bias term of the hyperplane
ξ_i	slack variable associated with $\mathbf{x}_i$
C	penalty parameter
α_i	Lagrange multiplier for $\mathbf{x}_i$
$K(\mathbf{x}, \mathbf{x}_i)$	kernel function
$\varphi(\mathbf{x})$	mapping function from $\mathbf{x}$ to the feature space
σ	width parameter for Gaussian kernel
U	the set of unbounded support vector indices
$g(\mathbf{x})$	decision surface

References

[1] Z.L. Gaing, Wavelet-based neural network for power disturbance recognition and classification, IEEE Trans. Power Deliv. 19 (4) (2004) 1560–1568.
[2] P.G.V. Axelberg, I.Y. Gu, M.H.J. Bollen, Support vector machine for classification of voltage disturbances, IEEE Trans. Power Deliv. 22 (3) (2007) 1297–1303.
[3] G.T. Heydt, P.S. Fjeldt, C.C. Liu, D. Pierce, L. Tu, G. Hensley, Applications of the windowed FFT to electric power quality assessment, IEEE Trans. Power Deliv. 14 (4) (1999) 1411–1416.
[4] D.C. Robertson, O.I. Camps, J.S. Mayer, W.B. Gish, Wavelets and electromagnetic power system transients, IEEE Trans. Power Deliv. 11 (2) (1996) 1050–1058.
[5] A.M. Gaouda, M.M.A. Salama, M.R. Sultan, A.Y. Chikhani, Power quality detection and classification using wavelet multiresolution signal decomposition, IEEE Trans. Power Deliv. 14 (4) (1999) 1469–1476.
[6] S.J. Huang, C.T. Hsieh, C.L. Huang, Application of wavelets to classify power system disturbances, Electr. Power Syst. Res. 47 (2) (1998) 87–93.
[7] C.H. Lin, M.C. Tsao, Power quality detection with classification enhancible wavelet-probabilistic network in a power system, IEE Proc. Gen. Transm. Distrib. 152 (6) (2005) 969–976.
[8] D.G. Cesar, V.G. Valdomiro, O.P. Gabriel, Automatic power quality disturbances detection and classification based on discrete wavelet transform and artificial intelligence, TDC06, IEEE/PES Latin Am. (2006) 1–6.
[9] S.K. Meher, A.K. Pradhan, G. Panda, An integrated data compression scheme for power quality events using spline wavelet and neural network, Electr. Power Syst. Res. 69 (2–3) (2004) 213–220.
[10] A.M. Gaouda, S.H. Kanoum, M.M.A. Salama, On-line disturbance classification using nearest neighbor rule, Electr. Power Syst. Res. 57 (1) (2001) 1–8.
[11] Y. Liao, J.B. Lee, A fuzzy-expert system for classification power quality disturbances, Electr. Power Energy Syst. 26 (3) (2004) 199–205.
[12] P.K. Dash, M. Nayak, M.R. Senepati, I.W.C. Lee, Mining for similarities in time series data using wavelet-based feature vectors and neural networks, Eng. Appl. Artif. Intell. 20 (2) (2007) 185–201.
[13] I. Monedero, C. Leon, J. Ropero, A. Garcia, J.M. Elena, J.C. Montano, Classification of electrical disturbances in real time using neural networks, IEEE Trans. Power Deliv. 22 (3) (2007) 1288–1296.
[14] V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.
[15] P. Janik, T. Lobos, Automated classification of power quality disturbances using SVM and RBF Networks, IEEE Trans. Power Deliv. 21 (3) (2006) 1663–1669.
[16] H. He, J.A. Starzyk, A self-organizing learning array system for power quality classification based on wavelet transform, IEEE Trans. Power Deliv. 21 (1) (2006) 286–295.
[17] C. Koley, P. Purkait, S. Chakravorti, Wavelet-aided SVM tool for impulse fault identification in transformers, IEEE Trans. Power Deliv. 21 (3) (2006) 1283–1290.
[18] S. Chen, H.Y. Zhu, Wavelet transform for processing power quality disturbances, EURASIP J. Adv. Signal Process. 47695 (2007) 20.
[19] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (7) (1989) 674–693.
[20] A.K. Jain, R.P.W. Duin, J. Mao, Statistical pattern recognition: a review, IEEE Trans. Pattern Anal. Mach. Intell. 22 (1) (2000) 4–37.
[21] C. Cortes, V.N. Vapnik, Support vector networks, Mach. Learn. 20 (1995) 273–297.
[22] D.P. Bertsekas, Nonlinear Programming, second ed., Athena Scientific, 1999.
[23] S. Abe, Support Vector Machines for Pattern Classification, Springer-Verlag, 2005.
[24] T.K. Abdel-Galil, M. Kamel, A.M. Youssef, E.F. El-Saadany, M.M.A. Salama, Power quality disturbance classification using inductive inference approach, IEEE Trans. Power Deliv. 19 (4) (2004) 1812–1818.
[25] T. Joachims, Making Large-scale SVM Learning Practical, Advances in Kernel Methods—Support Vector Learning, MIT Press, Massachusetts, 1999, pp. 169–184.
[26] V. Franc, V. Hlavac, Statistical Pattern Recognition Toolbox for Matlab, 2004.
[27] A.M. Gargoom, N. Ertugrul, W.L. Soong, A comparative study on effective signal processing tools for optimum feature selection in automatic power quality events clustering, in: In Industry Appl. Conf. 40th IAS Annual Meeting, vol. 1, 2005, pp. 52–58.
[28] P.K. Dash, M.V. Chilukuri, Hybrid S-transform and Kalman filtering approach for detection and measurement of short duration disturbances in power networks, IEEE Trans. Inst. Meas. 53 (2) (2004) 588–596.
[29] G.S. Hu, F.F. Zhu, Z. Ren, Power quality disturbance identification using wavelet packet energy entropy and weighted support vector machines, Expert Syst. with Appl. 35 (1–2) (2008) 143–149.