JAIDM
Journal of AI and Data Mining
Vol.1, No.1, 2013, 49-61.

Protection scheme of power transformer based on time–frequency analysis and KSIR-SSVM

M. Hajian*, A. Akbari Foroud

Department of Electrical and Computer Engineering, Semnan University, Semnan, Iran

Received 29 January 2013; accepted 24 February 2013
**Corresponding author: Mehdi.hajian.sem @gmail.com (M. Hajian)*

Abstract
The aim of this paper is to extend a hybrid protection plan for Power Transformer (PT) based on MRA-KSIR-SSVM. This paper offers a new scheme for protection of power transformers to distinguish internal faults from inrush currents. Some significant characteristics of differential currents in the real PT operating circumstances are extracted. Multi Resolution Analysis (MRA) is used as Time–Frequency Analysis (TFA) for decomposition of Contingency Transient Signals (CTSs), and the feature reduction is done by Kernel Sliced Inverse Regression (KSIR). Smooth Supported Vector Machine (SSVM) is utilized for classification. Integration KSIR and SSVM is tackled effectively and fast technique for accurate differentiation of the faulted and unfaulted conditions. The Particle Swarm Optimization (PSO) is used to obtain optimal parameters of the classifier. The proposed structure for Power Transformer Protection (PTP) provides a high operating accuracy for internal faults and inrush currents even in noisy conditions. The efficacy of the proposed scheme is tested by means of numerous inrush and internal fault currents. The achieved results are utilized to verify the suitability and the ability of the proposed scheme to make a distinction inrush current from internal fault. The assessment results illustrate that the proposed scheme presents an enhancement of distinguished inrush current from internal fault over the method to be compared without Dimension Reduction (DR).

Keywords: *Transformer Protection Scheme, Multi Resolution Analysis (MRA), Kernel Sliced Inverse Regression (KSIR), Smooth Supported Vector Machine (SSVM).*

1. Introduction
Power transformers are the most important component of power systems and play a vital role in any power system. In the power system security, protection of power transformers is a very challenging duty [1]. Identification of errors in the power transformer is considered the key to guarantee a reliable and continuous service for electric customers. The main weakness of the differential relay stems from its potential for false tripping caused by occurring inrush current phenomenon, which flows when the power transformer is energized [2]. Two important Contingency Transient Signals (CTSs) such as inrush current and internal fault that their mis-recognition might cause to mal-function of relays is presented.

The inrush currents of power transformers are non-stationary signals, due to flux saturation in the core during energization, high magnitude currents produced. The inrush currents are often occurred during the switching of power transformers. Magnitude of inrush current depends on the switching angles, switching time, magnitude and polarity of residual flux [2]. Also it depends on type and size of transformers.

Pattern recognition is an advanced field of research familiarly link to machine learning. As a part of this literature, classification concept endeavors for constructing classifiers that can find out an input pattern class. Many studies in the last decade have focused on developing classifiers that can learn from samples to execute recognition tasks [3]. The goal of the material presented in this paper is to introduce a new identification scheme for power transformer protection by signal processing and pattern recognition methods.

In transformer protection schemes area, the novelties presented in this paper can be summarized below.

1. We have used effective extracted features that increase percentage of the correct classification.
2. The purpose of this paper is to apply an effective dimension reduction method incorporate with classifier.
3. The proposed method uses the smooth supported vector machines for classification.

Time–frequency analysis [4] methods, such as Discrete Wavelet Transform (DWT) are required to attend to non-stationary behavior of the Contingency Signals (CSs) in order to express those in the time and frequency domain. The MRA is suitable for investigation of non-stationary signals, and unlike the spectral analysis, MRA represents a main advantage. It is powerful tool for time-frequency domain localization of transient contingencies (such as inrush current) [5].

One of the first and maybe the most accepted Dimension Reduction Method (DRM) is Slice Inverse Regression (SIR) [6]. Li (1991) [7] suggested SIR to find the Effective Dimension Reduction Directions (EDRD). The reduction techniques speed up the computation and boost the numerical stability.

SIR is a well-known DRM due to provide an effectual low-dimensional linear subspace. SIR can be comprehensive to nonlinear transform via the kernel approach. This study is investigated KSIR capability to combine with SSVM for classification. Numerical results indicate that KSIR is a useful kernel tool for nonlinear dimension reduction and it can combine with SSVM to structure a commanding tool for nonlinear data analysis.

SIR adopts the class information for evaluating the projection directions (unlike Principal Component Analysis (PCA)). Resembling the PCA, SIR is a technique based on the transformation of input features x to the effective features. Nevertheless, in contradiction of PCA, SIR provides the features by modeling the relation between input x and target variables *y* while maintaining the majority of the information in the input data. SIR can be observed as a PCA-like method applied on the random variable $E(\mathrm{x}|y)$ instead of on x. In other word, SIR directs to a generalized eigen-system, whilst PCA guides to an eigen-system [8]. The fundamental concept of the kernel SIR technique [9] is at first to preprocess the data pattern by some non-linear mapping and after that to apply the same linear SIR.

Compared with other methods of classification, SVMs have demonstrated outstanding potentials in coping with classification problems [10]. This paper tested the aptitude of SSVM in making a nonlinear separating surface. Support vector machine (SVM) has been proved to be a potent tool for fault detection. Smoothing processes [11] that are comprehensively used for resolving significant arithmetical programming challenges and applications are employed here to produce and resolve an unconstrained smooth redevelop of the support vector machine for pattern recognition by kernel-based algorithm. SSVM is resolved by a very fast Newton–Armijo approach and is developed to nonlinear separation surfaces using nonlinear kernel procedures [12]. The effectiveness of this scheme is expressed by widespread simulation of different operating conditions and faults in power transformers by PSCAD / EMTDC software. In this paper, two methods, SSVM and KSIR-SSVM, are investigated and compared. The proposed scheme is performed via the framework shown in Figure 1.

In this paper, MRA, KSIR and SSVM are jointly applied to differentiate internal faults from inrush currents. The schematic diagram for MRA-KSIR-SSVM method is presented in Figure 1. MRA has been applied for the TFA of fault signals for the distinguished inrush current and internal fault of power transformer using discrete wavelet coefficients (DWC). CTSs are broken down into frequency sub-bands by MRA. Then, a set of statistical features are extracted from these sub-bands to identification of CTSs. Furthermore, KSIR method is employed to decrease the dimensionality of features vector. Finally, the features vector is applied as an input to a Smooth Support Vector Machine (SSVM) with two discrete outputs: Inrush current or internal fault current.

KSIR can be employed for nonlinear dimension reduction. By combining with SSVM, we achieve a nonparametric and nonlinear classification. The effectual dimension reduction directions of the training data sets embedded in high-dimensional space is explored by KSIR. Next, the test data sets are projected onto these directions and SSVM as the classifier is more used to recognize the test data sets. PSO is adopted to optimize the parameters of SSVM. PSO is to find the optimal settings of parameters in SSVM. Compared to

GA, particle swarm optimization is powerful and easy to implement. PSO algorithm can select suitable parameters for SSVM classifier, which avoids over-fitting or under-fitting in the SSVM model occurred due to the improper determination of these parameters.

A review of the literature for distinguished fault from inrush current in power transformer is presented in the next section. Multi-resolution analysis definition and formulations are presented in section 3, and the extracted features are offered in section 4. The dimension reduction and concept of SIR and KSIR are explained in section 5, The SSVM concept and formulations are introduced in section 6 for classification of proposed protection scheme. In sections 7 and 8, modeling and simulation of various operating conditions of power system and numerical results are presented, respectively. Conclusion of the whole study is provided in section

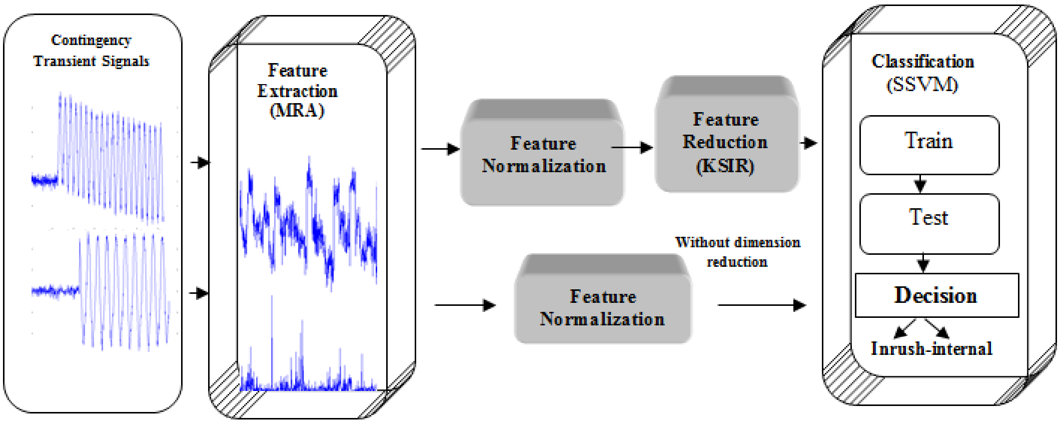

Figure 1. Framework of the proposed scheme

2. Related work

Dimension reduction has been a concentration of several main headlines of research in the statistical machine learning literature [13]. The supervised dimension reduction methods focus on the classification targets, including Linear Discriminant Analysis (LDA) [14], and Kernel Linear Discriminant Analysis (KLDA). The unsupervised dimension reduction techniques imagine that y is unidentified. PCA and kernel PCA [15], GPLVM [16] and nonlinear locality-preserving manifold learning [17] belong to this category of methods that do not control recognized target values. KSIR as a nonlinear feature reduction has been proven essential as a preprocessing stage for classification problems [18].

Many methods have been employed to distinguish inrush from internal faults in transformers. Most of them rely on an index and fixed threshold [19]. Some research has focused on to restrain tripping command of power transformer when an inrush current creates in the transformer windings. The common procedure employed to avoid false trips when inrush currents flow in windings is harmonic restraint relay [20]. These procedures have problems, when 2nd order harmonic component makes in different operating conditions the magnitude exceeds the predefined threshold. This may be due to resonant conditions of power system, presence of a shunt capacitor and nonlinear loads or saturation of transformer [1, 2].

The different methods have been proposed to distinguish fault from inrush current based on measuring the voltage and current waveforms [21, 22]. The need to utilize voltage of transformers and increased protection strategy computational burden are disadvantage of these methods. Some other methods detect based on measuring of the time between the respective peaks of differential current [23]. Recently, several new protective schemes have been proposed to deal with the foregoing problem in power transformer protection. Most of them have focused on the applications of intelligent techniques. Neural network is applied to discriminate the internal fault current and inrush current in [24]. The MRA technique is a dominant tool for power system transient analysis. Some protective schemes were used with signal processing and machine learning techniques. [1,19]. Conventional Neural Networks

(NNs) are complicated to construct in order to require to determination a suitable number of hidden neurons. The support vector machine (SVM) is a machine-learning tool, which is generally used to data patterns [25]. Integration of WT and artificial neural fuzzy system (ANFIS) has been proposed in [26] to differentiate the faulted from unfaulted conditions.

3. Wavelet-based multi-resolution analysis

Mallat (1989) presents the Multi-resolution analysis (MRA) that it can decompose the signal into various scales of orthogonal signal component. By means of MRA, the fault signals are divided into multi-scale signals. Multi Resolution Analysis (MRA) and Quadrature Mirror Filters (QMF) are also important for evaluating the discrete wavelet decomposition. In multi-resolution strategy, a disturbance signal is decomposed into several sub-signals which have specified harmonic components. A QMF consists of two filters. One gives the average (Low Pass Filter (LPF)), while the other gives details (High - Pass Filter (HPF)). These filters are related to each other in such a way as to be able to perfectly reconstruct a signal from the decomposed components. In this strategy, the approximation sub-signal $S_j(t)$ and the detailed sub-signal $D_j(t)$, which correspond to the components of signal $x(t)$ at different scales, formulated as follows [5,27]:

$$S_J(t) = \sum_k S_{J,K} \phi_{J,K}(t) J, K \in I \quad (1)$$

$$D_J(t) = \sum_k d_{j,K} \Psi_{j,K}(t) j, K \in I \quad (2)$$

The $D_j(t)$ contents an approximate frequency bound of $[f_s/2^{j+1} - f_s/2^j]$ Hz and the $S_J(t)$ contents an approximate frequency bound of $[0 - f_S/2^{J+1}]$ Hz, f_s is the sampling frequency. Therefore, the better scales of $D_j(t)$ mainly capture the detailed (high-frequency) feature of $x(t)$, while the larger scales of $D_j(t)$ and $S_j(t)$ mainly reveal the whole-view (low-frequency) feature of $x(t)$. After that, the original signal $x(t)$ can be recovered in terms of these sub-signals with diverse resolutions as follow:

$$x(t) \approx S_J(t) + D_J(t) + D_{J-1}(t) + ... + D_1(t) \quad (3)$$

MRA is normally based on Daubechies orthogonal wavelet basis. The choice of mother wavelet is important because different types of mother wavelets have different properties. This rigidly depends on the nature of the application. The most used mother wavelet in PTP diagnosis is the Daubechies wavelet with a four-coefficient filter (db4). Therefore, in this paper, the mother wavelet db4 is used, and shows that db4 has a good performance of detection of CTSs. The wavelet coefficients obtained by MRA can yield good information of CTSs. The fourth-order Daubechies' wavelet (Daublet4) was utilized as mother wavelet function to perform MRA on the CTSs.

As mentioned above, MRA permits decomposing signal into approximations (i.e. low frequency coefficients) and details (i.e. high frequency coefficients) by using a filter bank arranged of both LPF and HPF. The filtering procedure can be repeated, in order that CS is decomposed into lower resolution components and individual details. This is namely the MRA tree [27]. Figure 2 shows the MRA tree with three levels. In fact, the high frequency coefficients are ignored. So, a signal can be first decomposed into an approximation a1 and a detail d1 (that is the level 1of the decomposition). Subsequently, a1 can be decomposed into an approximation a2 and a detail d2 (that is the level 2 of the decomposition) and so on. Taking into account n levels of decomposition, the reconstruction process consents to recovering the initial signal, summing the n details d1, d2,…,dn and the approximation an of level n [5, 27]. (According to Figure 2)

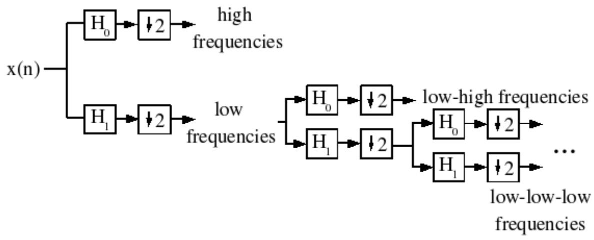

Figure 2. Three decomposed levels of MRA [27]

4. Extracted features

Feature extraction is the most important part of the intelligent system as a pattern recognition scheme. In the literature, the signal processing techniques are available for analyzing protection of power transformer signals. Some examples are Fast Fourier transforms method fractal-based method time–frequency ambiguity plane method, short time power and correlation transform method, and wavelet transforms method [28].

A signal is said to be stationary if its frequency or spectral contents are not changing with respect to time. Fourier Transform (FT) can be applied to the stationary signals. A non-stationary signal is one whose frequency changes over time. One of the major types of non-stationary signals has been identified, transient signals. Like inrush current, plenty of signals may contain non-stationary or transitory characteristics [27]. Various methods have been introduced and used to study non-stationary signals and both spectral as well as localized information has been obtained. The FT and Short Time Fourier Transform (STFT) view a signal in terms of finite time series since STFT uses a window of fixed width; they are unable to provide effective frequency relative resolution. On the other hand, the MRA is effective in providing time localized information as the information given is in both the time and frequency domains [28].

Essentially, the fault diagnosis is a pattern recognition problem, for which the key step is to extract useful fault features from vibration signals through some suitable signal processing methods [5]. In other words, the CTSs can be detected on the comparison of the extracted features from them.

Appropriate features are extracted by spectral data and statistical indicators of DWT coefficients outputs. One of the important characteristics of the selected features, are their severability for different classes. The suggested integrated feature extraction strategy has several desirable characteristics:

1. It includes the minimum number of important and effective features necessary to achieve high performance of the classification of inrush and internal fault.
2. It creates a separable feature vector.

Most useful features must be first extracted from coefficients in order to more effectively recognize the type of CTSs. For the recognition of faults, four-level MRA are applied on the measured current signals. Eight statistical features are extracted from the coefficients of different bands generated using MAR, i.e. features number from 1 to 8 are shown in Table 1.

In this paper, some statistical methods, such as mean, standard deviation, energy, shannon-entropy, log-entropy, threshold-entropy presented in Table 1 have been used as the features extractors. These features are obtained based on the practical experiments.

Table 1. Extracted features for detecting scheme

Feature number	Description
1	Standard deviation of level 2 of detail
2	Minimum value of absolute of level 5 of approximation
3	Mean of average of absolute of all level of details
4	Mean of disturbances energy (all level).
5	Energy of level 3 of detail.
6	Mean of Shannon entropy of all level of details
7	Mean of log energy of all level of details
8	Mean of threshold entropy of all level of details

Extracted feature vector should have properties, such as a variety from class to class. Also, it should not have correlation with other features. Figure 3 shows the normalized value of extracted features from inrush current and internal fault for noisy condition. As shown in Figure 3 the values of the features of each class (inrush current and internal fault) are different. So the features for any classes have different behavior. Therefore, these features create a separable feature vector. The features vector is fed to SSVM for classification.

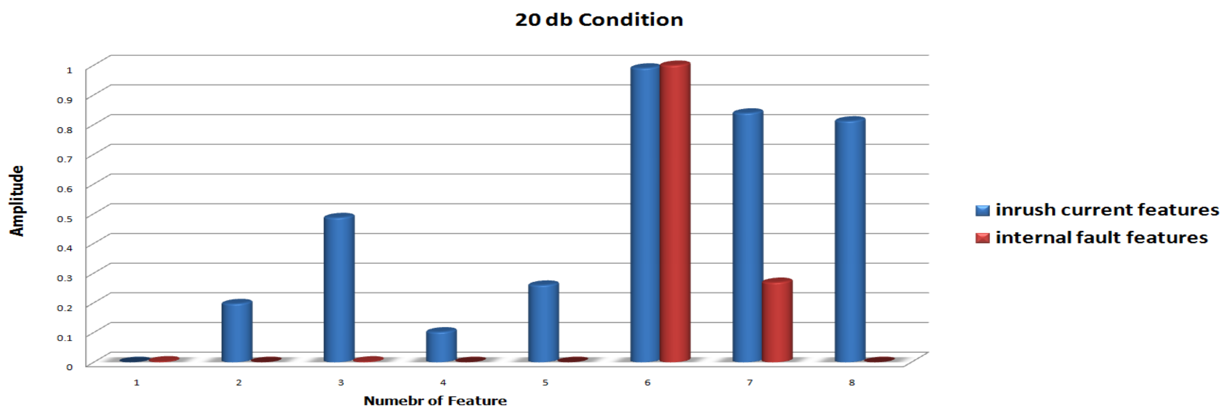

Figure 3. The values of the features of inrush current and internal fault in 20 dB noise condition.

5. Dimension reduction

Dimension reduction is a main popular strategy in analyzing multivariate input data due to visualization of the patterns of data. Also, it is a key material in data mining and machine learning. The main motivation applying dimension reduction methods are: Improve visualization, remove redundancy of data, data compression, reduce computational time and enhancement of accuracy [9]. The dimension reduction offered by KSIR can be applied as a preprocess for classification. The aim of reduction hypothesis is the high-dimensional data is projected to a lower dimensional subspace without the loss of information for separability among classes [29].

SIR finds the directions of maximum variance, with P data points collapsed into K slice means using the affinity in classes. This reduction technique will speed up the computation and increase the numerical stability.

In this paper, we employed a hybrid SIR method using a kernel machine which we call kernel SIR. The method of SIR employ to explore the EDRD from the training data embedded in high-dimensional space. The test data are then projected onto these directions and the classifier is further applied to classify the test data. In the kernel extension of SIR, input data is mapped to the Hilbert space induced from the kernel function. It means that KSIR actually finds an effective projection direction in the kernel feature space. The results show that KSIR-SSVM is an effective classification method in the structural risk minimization, non-linear characteristics, avoiding the over-fitting and strong generalization ability.

5.1. SIR method

This paper proposed an effective data-analytic tool, SIR, for reducing the dimension of the input features. The aim of below formulation is to get cut down the dimension of the input data from P to a smaller number K without losing any information. The hypothetical properties of SIR can be presented below [9].

$$\xi = \psi(\mu_1^T X, \ldots, \mu_K^T X, \varepsilon) \quad , \quad X \in R^P \tag{4}$$

Where the μ_K is the unknown row vector and ε is the error. For effectively reducing the dimension, we need only to estimate the EDRD generated by the μ_K. In fact, the μ_K itself is not identifiable without an exact structural shape on f. If the distribution of x has been standardized to have the zero mean and the identity covariance, the $E(X|\xi)$, generates the EDRD.

Therefore, any supposition of ξ includes only the summary statistic S(X) that is of much lower dimension than the original data X. Linear techniques for dimension reduction concentrate on linear summaries of the data, that is,

$$S(X) = \left[\mu_1^T X, \ldots, \mu_K^T X\right] \tag{5}$$

The K-dimensional subspace, $s = span\{\mu_1, \ldots, \mu_k\}$ is mentioned as the EDRD space in [7] since S summaries all the information, we require to identify the ξ. The obtained result in [7] is that under some soft conditions the EDRD $\{\mu_j\}_{j=1}^{K}$ according to the eigenvectors of the matrix, that covariance matrix is denoted by Γ.

$$W = [\Gamma(X)]^{-1}\Gamma[E(X|\xi)] \tag{6}$$

Consequently, the EDRD or subspace can be calculated by means of an eigen-analysis of matrix W, that these directions display the major distinction in the class means relative to the within-class variance and are better for classification [9]. In other words, the output of the KSIR method permits for the evaluation of nonlinear EDRD.

5.2. Kernel process

Kernel-based algorithm [30], a non-linear transformation can map the input feature space into a high-dimensional feature space. Basically, the classification is more probable to be linearly resolved in high-dimensional space (according to Figure 4).

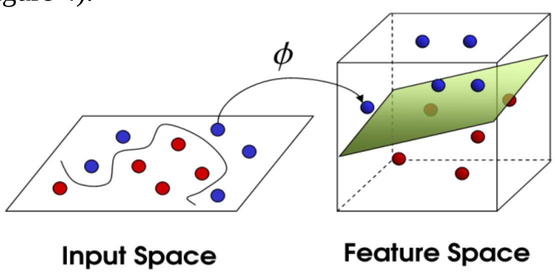

Figure 4. Concept of non-linear kernel function

This kernel function corresponds to the inner product of $\phi(x_i)$ and $\phi(x_j)$ where ϕ is a definite mapping from input space R^d to a high-dimensional feature space *F*. In the other words, to train the classifier, only the kernel is required and no explicit knowledge of $\varphi(x_i)$ is needed.

$$\varphi : x \to \varphi(x), \quad k\left(x_i, x_j\right) = \varphi\left(x_i\right)^T . \varphi(x_j) \tag{7}$$

5.3. KSIR method

When the variation of data pattern is nonlinear, KSIR based visualization process has better performance than SIR. Due to overcoming several constraints, KSIR finds the linear subspace that best represents data. KSIR is a generalizing method based on linear SIR that converts into nonlinear case by the kernel approach [9]. The concept of KSIR is to transfer the original input x_i into a high-dimensional feature space $\varphi(x_i)$ firstly by kernel method and then by computing the linear SIR in $\varphi(x_i)$. The linear SIR in $\varphi(x_i)$ corresponds to a nonlinear SIR in x_i . Similarly to SIR, KSIR estimates $\Gamma[E(\Phi(x)|\xi)]$ by slicing the output.

6. Proposed classification

Classification should be considered as a significant component for the design of intelligent systems based on pattern recognition techniques. An important example of the general discriminated classifiers is the support vector machine (SVM) [31]. Support Vector Machines [32] have been widely applied to pattern classification problems [19], non-linear regressions and clustering. In years of contemporary, linear or nonlinear kernels SVMs have become one of the most hoping machine learning methods for classification. The original learning form of SVM goes ahead to a quadratic program (QP), which is a convex constrained optimization problem and therefore has a unique answer. Compared with other machine learning methods, such as the neural networks (NN), that is a great advantage [30, 32]. SVM was suggested by Vapnik [32]. After that, studies have been worked to use the SVM tool in many sciences.

6.1. Smooth support vector machine

The classification by support vector machine is defined as discovering the weights and bias parameters (separating surface) in order to maximize the margin while ensuring that the training data are well classified. This can be expressed as the QP optimization problem [32]. The optimal surface in the sense of machine learning is a balanced behavior between over fitting and under fitting. Concept of optimal surface is based on maximizing the margin shown in Figure 5.

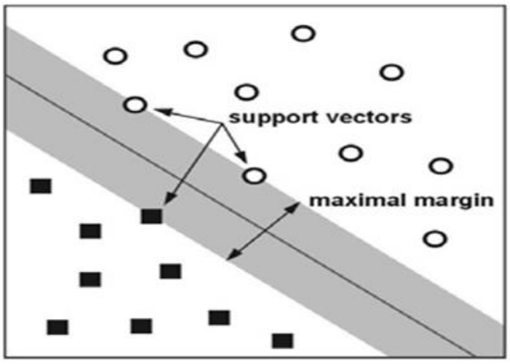

Figure 5. Concept of optimal surface based on maximize the margin

Due to convenient discussion, we first describe the linear SVM. Given a training dataset $S = \{(x_1, y_1), ..., (x_n, y_n)\} \subseteq R^d \times R$, where $x_i \in R^d$ is the input data and $y_i \in \{-1,1\}$ is the corresponding class tag, the classical SVM separating surface is modeled by resolving a convex optimization that can be defined as a mathematical problem as follows:

$$\min_{(\omega,b,\xi)\in R^{d+1+n}} \left\{ C\sum_{i=1}^{n} \xi_i + \frac{1}{2}\|\omega\|_2^2 \right. \tag{8}$$

With non-eqilibrium constraints:

$$s.t.\{ y_i\left(\omega^{T} x_i + b\right) + \xi_i \geq 1$$
$$\xi_i \geq 0,\ for i = 1, 2, ..., n, \tag{9}$$

Where *C* is a cost and positive parameter that controls the trade of between the error of training process and maximizing the margin, which is got by minimizing $\|\omega\|_2^2$.

An alternative smoothing strategy has been recommended and resolved by a fast Newton–Armijo approach that converges globally and quadratically. In general, finding the global optimum of a function can be a very difficult duty. However, for a particular type of optimization problems acknowledged as convex optimization problems, and many cases can discover the global solution. A main difficulty with non-convex formulations is that the global optimal answer cannot be effectively calculated, and the behavior of a local solution is hard to evaluate. In practice, convex relaxation (such as SVM for classification) has been accepted to remedy the problem. Newton–Armijo method is applied for discovering a convex function minimum that needs neither strong convexity nor even smoothness merits on the whole space [33]. It guarantees to find global convergence when the function is not strongly convex. The smooth formulation with a nonlinear kernel retains the strong convexity and twice differentiability and as a result can be applied Newton–Armijo method to solve it [12].

The key difference between smoothing approach and that of the classical SVM [32] is that solving a linear system instead of resolving a QP as the case with the classical SVM. Furthermore, the result can be indicated that smoothing approach converges globally to the unique solution [12].

On the contrary of the basic SVM of (8), a smooth support vector machine (SSVM) minimizes the square of the slack vector ξ with weight $\frac{C}{2}$.

Besides, SSVM appends on the term $\frac{b^2}{2}$ to the objective to be minimized results in the following optimization problem:

$$\min_{(\omega,b,\xi)\in R^{d+1+n}}\left\{\frac{C}{2}\sum_{i=1}^{n}\xi_i^2+\frac{1}{2}\left(\|\omega\|_2^2+b^2\right)\right. \tag{10}$$

$$\begin{aligned} s.t.\{y_i\left(\omega^{\mathrm{T}}\mathrm{X}_i+b\right)+\xi_i &\geq 1 \\ \xi_i \geq 0,\ for i &= 1,2,\ldots,n, \end{aligned} \tag{11}$$

At a solution of (10), ξ is given by $\xi_i=\{1-y_i(\omega^{\mathrm{T}}\mathrm{X}_i+b)\}+$ for all i where the *plus* function x+ is defined as $\mathrm{X}_+=\max\{0,x\}$. Thus, we can replace ξ_i in (2) by $\{1-y_i(\omega^{\mathrm{T}}\mathrm{X}_i+b)\}+$. This will transfer the problem (10) into an unconstrained minimum optimization below:

$$\min_{(\omega,b)\in R^{d+1}}\left\{\frac{C}{2}\sum_{i=1}^{n}\{1-y_i(\omega^{\mathrm{T}}\mathrm{x}_i+b)\}_+^2+\frac{1}{2}\left(\|\omega\|_2^2+b^2\right)\right. \tag{12}$$

This modulation decreases the variables from $d+1+n$ to $d+1$.

However, the minimization of objective function is not twice differentiable which precludes the employ of a fast Newton method. In SSVM, the plus function X_+ is estimated using a smooth p-function $p(x,\alpha)=x+\frac{1}{\alpha}\log(1+e^{-\alpha.x}),\alpha\succ 0$. By replacing the plus function with a very accurate smooth approximation *p*-function provides the smooth support vector machine strategy [34]:

$$\min_{(\omega,b)\in R^{d+1}}\left\{\frac{C}{2}\sum_{i=1}^{n}p\{1-y_i(\omega^{\mathrm{T}}\mathrm{X}_i+b),\alpha\}^2+\frac{1}{2}\left(\|\omega\|_2^2+b^2\right)\right. \tag{13}$$

Where $\alpha\succ 0$ is the smooth parameter. The objective function in (13) is robustly convex and infinitely differentiable. Hence, it has a unique solution and can be resolved by fast Newton-Armijo approach.

For the nonlinear application, this modulation can be developed to the nonlinear SVM by the kernel trick scheme as follows:

$$\min_{(u,b)\in R^{d+1}}\left\{\frac{C}{2}\sum_{i=1}^{n}p([1-y_i\{\sum_{j=1}^{n}u_j\mathrm{K}(\mathrm{X}_i,\mathrm{X}_j)+b\}],\alpha\}^2+\frac{1}{2}\left(\|u\|_2^2+b^2\right)\right. \tag{14}$$

Where $k(x_i,x_j)$ is a kernel function.

The non-linear SSVM can be formulated in matrix form as follows:

$$\sum_{u_j\neq 0}u_j\mathrm{K}(A_j^{\mathrm{T}},\mathrm{X})+b=\mathrm{K}(\mathrm{X},A^{\mathrm{T}})u+b \tag{15}$$

Where $A=[x_1^{\mathrm{T}};\ldots;x_n^{\mathrm{T}}]$, $\mathrm{A}_j=x_j^{\mathrm{T}}$. The coefficient u_j is terminated by resolving an optimization problem (15) and the data points with corresponding non-zero coefficients are named support vectors. It is frequently suitable to have fewer support vectors [34].

7. Power system under study and simulation

PSCAD/EMTDC software is a graphical interface industry standard simulation instrument for simulation the transient behavior of electrical grids. The experiments of this paper are implemented through digital time-domain simulation studies in the PSCAD/EMTDC 4.2.1 software environment. To have comprehensive study, different contingency signals of inrush currents and internal faults are simulated. More than 2000 samples are generated.

For evaluation of the proposed method, a part of power system including a 160 MVA 230/63 kV real three-phase transformer is modeled (Figure 6). Different conditions of internal fault and inrush currents that may be occurred are simulated. Currents of primary and secondary sides can be saved in PSCAD/EMTDC software via output channels. By using these measured currents of both sides of power transformer, differential currents are calculated after three main stages i.e. zero sequence elimination, vector group adaptation and CT mismatch ratio correction. The sampling frequency is set 2.5 kHz. So each cycle contains 50 samples.

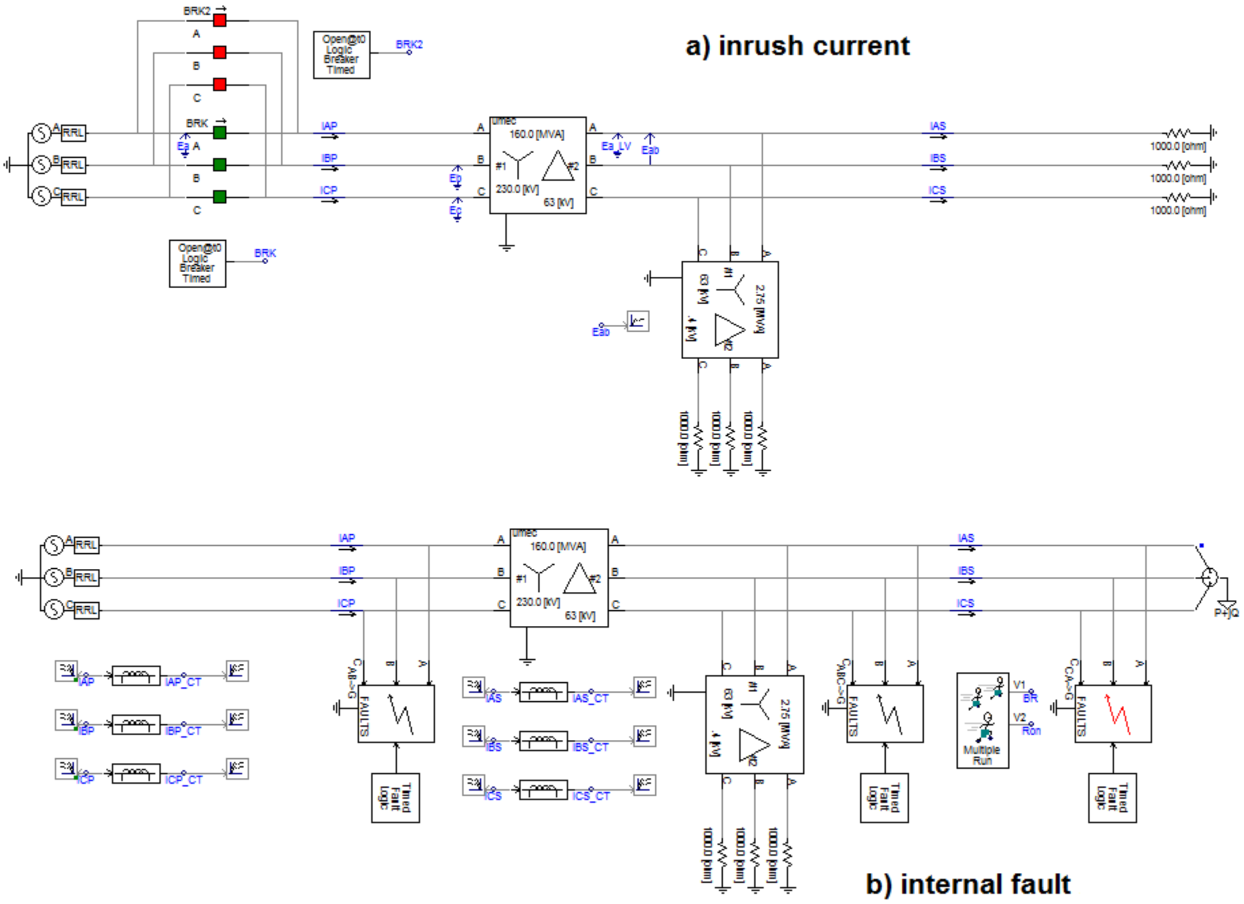

Figure 6. View of power system model, a) model of inrush current, b) model of internal fault current.

Moreover, the UMEC [35, 36] model is used for modeling of power transformer. This model considers the magnetic coupling between windings of different phases in addition to the coupling between windings of the same phase. Also, it uses the magnetization curve of transformer for considering the non-linearity of core characteristics. Thus, more accurate differential currents are obtained, specifically when transformer cores enter to saturation region.

Besides, the Jiles–Atherton [37] model is applied for precise simulation of CTs' behavior. This model can yield better representation of B-H curve using a modified Langevin function and a new parameter. In the simulation process, different parameters that affect the CT saturation are considered. Remenance flux, connected burden of secondary side and magnetization curve of CT core are main parameters, which have been considered in simulation conditions.

For simulation of transformer energization, remenance flux and switching time are considered as effective parameters on magnitude of inrush currents. Usually, the maximum remenance flux in each leg is lower than 80% of the peak flux. This residual flux can be modeled by inserting a DC current source in parallel of each winding (Figure 6). Besides, the switching time interval for energization of transformer has been set 1 msec. Moreover, the similar situation occurs when transformers are energized in parallel or nearby an already energized transformer. This event that the transformers draw causes large current from the source is known as the "sympathetic inrush". Moreover, terminal faults are also simulated considering different effective parameters on fault currents like fault occurrence instant, fault resistance and fault type. For different fault types, fault resistances are considered 0, 5 and 10 ohms and the time interval of fault occurrence is set to 1 msec.

Also, the practical CTSs in an electrical power network consist of noise. Therefore, the proposed scheme has to be scrutinized under noisy condition. Due to test the sensitivity of the proposed scheme against noisy condition, white Gaussian noise with the signal-to-noise ratio (SNR) 20 dB is added with CTSs and operated with MRA for computation of features. The value of the SNR is described as follows:

$$SNR(dB) = 10 * \log\left(\frac{P_s}{P_n}\right) \tag{16}$$

Where P_s is the power of signal and P_n is that of the noise.

The SSVM classifier is trained with training set and tested. Arbitrary parameters C, ε and kernel mainly affect on the accuracy and performance of the SSVM classifier. In this paper, PSO algorithm [38] has been used for solving the problem of choosing optimal parameters of SSVM. In order to get adequate gamma (ε) in radial basis kernel function and C parameters, a heuristic search was performed.

8. Numerical results and analysis

After obtaining differential currents, the MRA is applied for extraction of features. For normalization purpose, feature vectors are transformed to the range [0 1]. Then, the KSIR method is applied for reduction of features dimensions. Afterward, selected features are used for training and testing process. This study presents a strategy for improving SSVM performance in two aspects: Feature reduction and parameter optimization. The obtained classification accuracy shows the performance of the proposed algorithm.

In this section, we describe a sensitivity analysis of the SSVM parameters. Then, we determine the optimal parameters and employ them for pattern recognition. Finally, the results are presented.

8.1. The optimal SSVM model by PSO

For the optimization, a particle swarm optimization algorithm is proposed to improve the generalization performance of the recognizer. In this module, the SSVM classifier design is optimized by searching for the best value of the parameters that tune its kernel function parameter. Inappropriate choice of the parameters can lead to over-fitting or under-fitting. In the study, PSO is used to determine the SSVM parameters (C, ε). The obtained values of classifier parameters are given in Table 2. These values yield maximum classification accuracy.

Particle swarms have two primary operators: Velocity update and position update. During each generation, each particle is accelerated toward its previously best position and towards its best global position. At each iteration, a new velocity value for each particle is calculated based on its current velocity, the distance from its previous best position and the distance from its best global position. The new velocity value is then used to calculate the next position of the particle [38].

Figures. 7 and 8 show the evaluated parameters of the kernel function of the SSVM classifier for MRA-KSIR-SSVM structure obtained by PSO for different runs. According to results (Figures. 7 and 8), the best classification for MRA-KSIR-SSVM structure is obtained in proper pairs of C = 0.3 and ε = 0.01 (see Table 2).

In this section, we mainly compare the two strategy of protection of power transformer. The concept of using pattern recognition and signal processing device for the protection of power transformer, by MRA-KSIR-SSVM scheme through training from simulation data to properly classify future data is presented. The chief goal of this study is to separate the two classes of CTSs by using MRA-KSIR-SSVM scheme. As it can be seen in Table 2, the obtained results indicate a very good classification performance and the proposed scheme showing the high robustness against noise. Also, the implementation of KSIR algorithm is concluded the fast computation and numerical stability.

As a result, the proposed method (MRA-KSIR-SSVM) can achieve better identification of internal fault and inrush current than MRA-SSVM. This comparison result implies that the proposed KSIR-SSVM strategy obviously outperforms the SSVM strategy in diagnosing different of internal fault and inrush current.

We have demonstrated the applicability of the proposed technique to the practical condition. A serious problem of practical recognition system is its low classifying speed. In the SVM classifier, the speed depends on the number of support vectors. The training points that are nearest to the separating function are called support vectors [32]. We have devised a method to overcome this problem. The KSIR-SSVM, proposed in this paper, is a new method to reduce the number of support vectors (SVs) (see Table 2). The advantages of the approach lies in the fact that the smaller the input dataset is, the fewer SVs would yield and that it would require less CPU time and memory. Experiment results demonstrate that KSIR-SSVM method can control the tradeoff between the classifying speed and the performance of SSVM. The employment of the KSIR-SSVM method to reduce the number of SVs decreases the expectation value of the probability of committing an error on a test example and enhances SSVM's generalization capability.

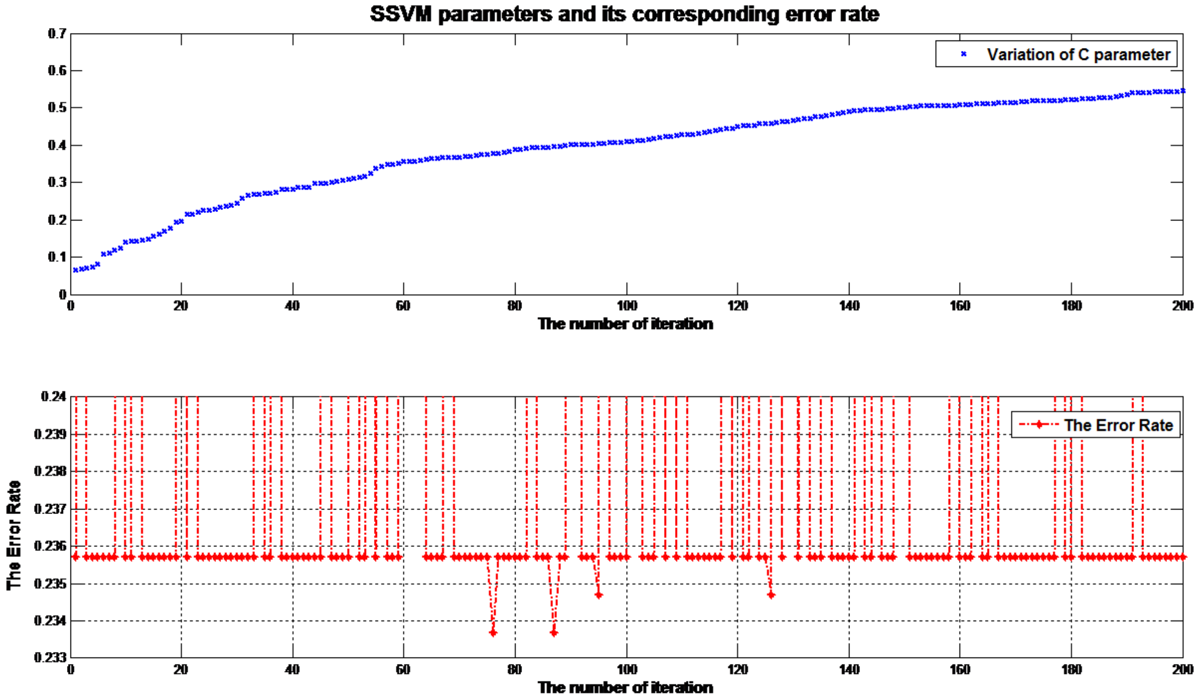

Figure 7. Sensitivity analysis of C parameter (Error rate with different C)

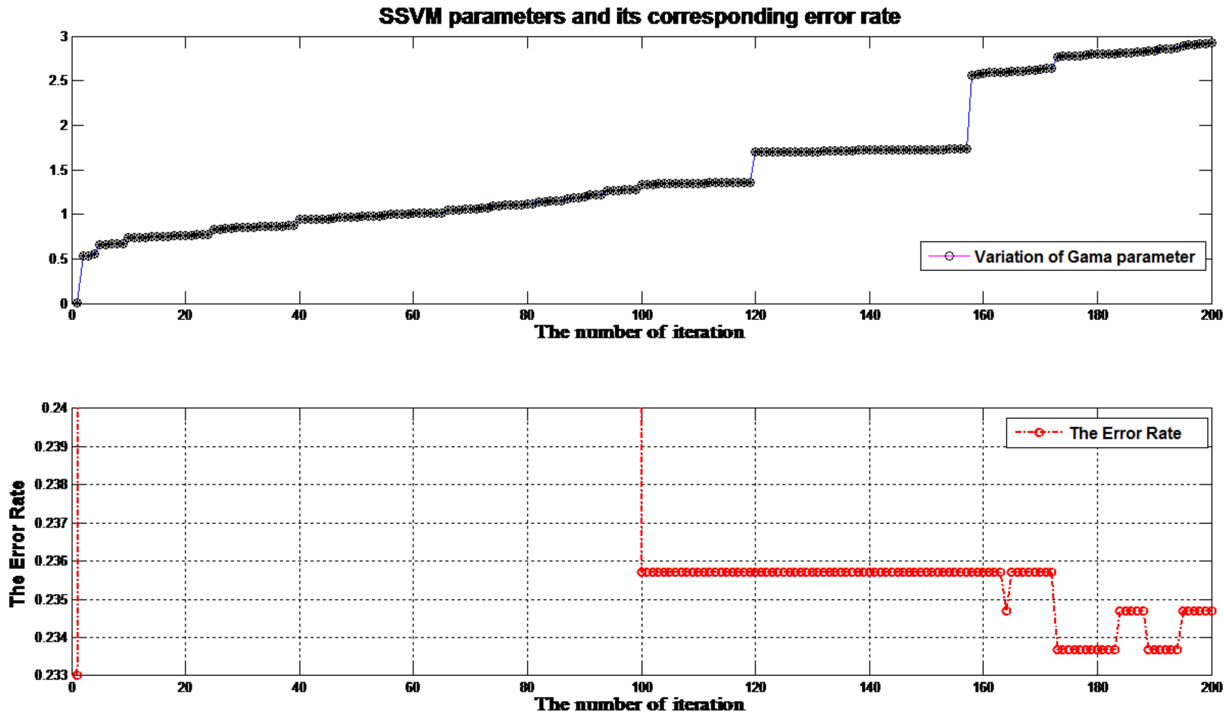

Figure 8. Sensitivity analysis of gama parameter (Error rate with different gama)

Table 2. Identification results of proposed scheme for conditions including internal fault and inrush current.

Methods	Dataset input	(C, $\mathcal{E}$)	#SVs	The error rate of testing	CPU Sec.
KSIR-SSVM	20 dB	(0.3,0.01)	41	0.233	0.0123
SSVM	20 dB	(1,0.4654)	143	0.285	0.0488

9. Conclusion

This paper has proposed and evaluated MRA-KSIR-SSVM scheme to distinguish between internal faults and inrush currents in power transformer protection. Required CSs have been attained through many simulations in PSCAD/EMTDC software. MRA has been applied to extract features. This paper is explored KSIR ability to combine with SSVM for the protection scheme. The joint KSIR and SSVM are able to efficiently decrease the feature vectors, speed up the convergence in the training of SSVM and obtain higher identification accuracy. PSO algorithm is chosen as an optimization technique to optimize the input feature subset selection and the SSVM parameters setting simultaneously. This technique will improve the SSVM performance. The KSIR–SSVM method could be a promising tool for the protection of power transformers. Also, the results show first, comparing SSVM with KSIR-SSVM obtains better generalization performance. Second, the stability of protection scheme in presence of the inrush current, internal faults signals, robust fault detection, and even in presence of a bad noisy condition may assure secure and correct performance of protection system of transformer.

References

[1] Samantaray, S.R. and Dash, P.K. (2011). Decision Tree based discrimination between inrush currents and internal faults in power transformer.Electrical Power and Energy Systems. 33, 1043–1048.

[2] Sedighi, A.R. and Haghifam, M.R. (2005). Detection of inrush current in distribution transformer using wavelet transform. Electrical Power and Energy Systems. 27, 361–370.

[3] Saxena, D., Verma, K.S. and Singh, S.N. (2010). Power quality event classification: an overview and key issues. International Journal of Engineering, Science and Technology. 2(3), 186-199.

[4] Hlawatsch, F. and Auger, F. (2008). Time-Frequency Analysis: Concepts and Methods. UK: ISTE/Wiley, London.

[5] Hossam Eldin, A.A. and Refaey, M.A. (2011). A novel algorithm for discrimination between inrush current and internal faults in power transformer differential protection based on discrete wavelet transform. Electric Power Systems Research. 81, 19–24.

[6] Scrucca, L. (2011). Model-based SIR for dimension reduction. Computational Statistics and Data Analysis. 55, 3010–3026.

[7] Li, K.C. (1991). Sliced inverse regression for dimension reduction (with discussion). Journal of the American Statistical Association. 86, 316–34.

[8] Wu, H.M. and Lu, H.H.S. (2007). Iterative sliced inverse regression for segmentation of ultrasound and MR images. Pattern Recognition. 40, 3492 – 3502.

[9] Yeh, Y.R., Huang, S.Y. and Lee, Y.J. (2009). Nonlinear Dimension Reduction with Kernel Sliced Inverse Regression. IEEE Transactions on Knowledge and Data Engineering. 21(11).

[10] Wang, Y., Chen, S. and Xue, H. (2011). Support Vector Machine incorporated with feature discrimination. Expert Systems with Applications. 38, 12506–12513.

[11] Chen, C. and Mangasarian, O.L.A. (1996). Class of Smoothing Functions for Nonlinear and Mixed Complementarity Problems. Computational Optimization and Applications. 5(2), 97-138.

[12] Lee, Y. and Mangasarian, O.L.(2001). SSVM: A Smooth Support Vector Machine for Classification. Computational Optimization and Applications. 20, 5–22.

[13] Duintjer Tebbens, J. and Schlesinger, P. (2007). Improving implementation of linear discriminant analysis for the high dimension/small sample size problem. Computer Statistic Data Analysis.

[14] Yang, J. and Yang, J.-Y. (2003). Why can LDA be performed in PCA transformed space?. Pattern Recognition. 36 (2), 563–566.

[15] Delalleau, Y. O., Le Roux, N., Paiement, J-F., Vincent, P. and Ouimet, M. (2004). Learning eigen functions links spectral embedding and Kernel PCA. Neural Computation. 16(10), 2197–2219.

[16] Lawrence, N.D. and Quiñonero Candela, J. (2006). Local distance preservation in the GP-LVM through back constraints.proceedings. of the 23rd International Conference on Machine Learning. 513-520.

[17] Roweis, S.T. and Saul, L.K. (2000). Nonlinear Dimensionality Reduction by Locally Linear

Embedding. Science. 290(22).

[18] Wu, H.M. (2008). Kernel Sliced Inverse Regression with Applications on Classification. Computational and Graphical Statistics. 17(3),590-610.

[19] Tripathy, M. (2010). Power transformer differential protection using neural network Principal Component Analysis and Radial Basis Function Neural Network. Simulation Modelling Practice and Theory. 18, 600–611.

[20] Verma, H.K. and Kakoti, G.C. (1990). Algorithm for harmonic restraint differential relaying based on the discrete Hartley transform. Electric Power Systems Research. 18 (2), 125–129.

[21] Liu, P., Malik, O. P., Chen, C., Hope, G.S. and Guo, Y. (1992). Improved operation of differential protection of power transformers for internal faults. IEEE Transaction on Power Delivery. 7(4), 1912-1919.

[22] Inagaki, K. and Higaki, M. (1998). Digital protection method for power transformers based on an equivalent circuit composed of inverse inductance. IEEE Transaction on Power Delivery. 4(4),1501-1510.

[23] Morsi, W.G. and El-Hawary, M.E. (2009). Wavelet Packet Transform-Based Power Quality Indices for Balanced and Unbalanced Three-Phase Systems under Stationary or Nonstationary Operating Conditions. IEEE Transaction on Power Delivery. 24(4).

[24] Guzman, A., Zocholl, S., Benmouyal, G. and Altuve, H.J. (2001). Acurrent based solution for transformer differential protection-part I: Problem statement. IEEE Transaction on Power Delivery. 16(5), 485-491.

[25] Chong, Z., Chong-Xun, Z. and Xiao-Lin, Y. (2009). Automatic recognition of cognitive fatigue from physiological indices by using wavelet packet transform and kernel learning algorithms. Expert Systems with Applications. 36,4664–4671.

[26] Monsef, H. and Lotfifard, S. (2007). Internal fault current identification based on wavelet transform in power transformers. Electric Power Systems Research.77, 1637–1645.

[27] Uyar, M., Yildirima, S. and Gencoglu, M.T. (2008). An effective wavelet-based feature extraction method for classification of power quality disturbance signals. Electric Power Systems Research. 78, 1747–1755.

[28] Saini, M.K. and Kapoor, R. (2012). Classification of power quality events – A review. Electrical Power and Energy Systems. 43, 11–19.

[29] Jimenez, L.O. and Landgrebe, D.A. (1997). Supervised classification in high dimensional space: Geometrical, statistical, and asymptotical properties of multivariate data. IEEE Transactions on Systems, Man and Cybernetics. 28(1), 39–54.

[30] Liang, Z. and FuLi, Y. (2009). Incremental support vector machine learning in the primal and applications. Neurocomputing. 72, 2249–2258.

[31] Ayadi, M.E., Kamel, M.S. and Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern Recognition. 44, 572–587.

[32] Vapnik, V. (1995). The nature of statistical learning theory. Springer. New York.

[33] Zhang, T. (2010). Analysis of Multi-stage Convex Relaxation for Sparse Regularization. Journal of Machine Learning Research. 11, 1081-1107.

[34] Chang, C.C., Chien, L.J. and Lee, Y.J. (2011). A novel framework for multi-class classification via ternary smooth support vector machine. Pattern Recognition. 44, 1235–1244.

[35] Woodford, D. (2000). Introduction to PSCAD/EMTDC V3, Manitoba. Canada: Manitoba HVDC Research Center Inc.

[36] Enright, W., Nayak, O.B., Irwin, G.D. and Arrillaga, J. (1997). An electromagnetic transients model of multi-limb transformers using normalized core concept. Proceedings of the International Conference on Power Systems Transients (IPST '97), Seattle, WA, 93–98, 22–26 June 1997.

[37] Annakkage, U. D., McLaren, P.G., Dirks, E., Jayasinghe, R.P. and Parker, A.D. (2000). A current transformer model based on the Jiles–Atherton theory of ferromagnetic hysteresis. IEEE Transactions on Power Delivery. 15(1), 57–61.

[38] Liu, L., Zhuang, Y. and Xue-yong, L. (2011). Tax forecasting theory and model based on SVM optimized by PSO. Expert Systems with Applications. 38 ,116–120.