

# Application of support vector machines for fault diagnosis in power transmission system

B. Ravikumar, D. Thukaram and H.P. Khincha

**Abstract:** Post-fault studies of recent major power failures around the world reveal that mal-operation and/or improper co-ordination of protection system were responsible to some extent. When a major power disturbance occurs, protection and control action are required to stop the power system degradation, restore the system to a normal state and minimise the impact of the disturbance. However, this has indicated the need for improving protection co-ordination by additional post-fault and corrective studies using intelligent/knowledge-based systems. A process to obtain knowledge-base using support vector machines (SVMs) is presented for ready post-fault diagnosis purpose. SVMs are used as Intelligence tool to identify the faulted line that is emanating and finding the distance from the substation. Also, SVMs are compared with radial basis function neural networks in datasets corresponding to different fault on transmission system. Classification and regression accuracies are reported for both strategies. The approach is particularly important for post-fault diagnosis of any mal-operation of relays following a disturbance in the neighbouring line connected to the same substation. This may help to improve the fault monitoring/diagnosis process, thus assuring secure operation of the power systems. To validate the proposed approach, results on IEEE 39-Bus New England system are presented for illustration purpose.

## 1 Introduction

Recent studies of significant power system disturbances reported by North American Electric Reliability Council (NERC) indicate that protective relays are involved, one way or another, in 75% of the major disturbances [1] and the most troublesome ones are backup protection relays. With their limited view of the interconnected network based on their locally measured inputs, conventional backup protection relays generally take actions to protect a localised region of the network without considering the impact on the whole network.

On 2 July, 1996, a fault occurred on a 345 kV line in southern Idaho, which tripped correctly, followed by an incorrect trip of a parallel line and an overload trip of a third line. System instability resulted in four islands being formed, and system frequency dropped to 58.75 Hz and 3000 MW of load was dropped by under frequency load shedding [2–5]. All of these events involved, in part, relay failures that were not detected during normal operation, calibration or maintenance [1, 6–8].

Transmission line relaying involves detection, classification and location of transmission line faults. Fast detections of faults enable quick isolation of the faulty line from service and hence, protecting it from the harmful effects of fault. Classification of faults means identification of the type of fault, and this information is required for the fault location and assessing the extent of repair work to be carried out. Accurate fault location is necessary to facilitate

quick repair and restoration of the line and to improve the reliability and availability of the power supply.

When a major power system disturbance occurs, protection and control actions are required to stop the power system degradation, restore the system to a normal state and minimise the impact of the disturbance. The modern energy management system is supported by supervisory control and data acquisition (SCADA) software with numerous power system analysis tools such as state estimation, power flow, optimal power flow, security analysis, transient stability analysis and midterm to long-term stability analysis; and with optimisation techniques such as linear and nonlinear programming. The available time for running these application programs is the limiting factor in applying these tools during an emergency, and a trade-off with accuracy is required. Further, propagation of a major disturbance is difficult to incorporate into a suitable numerical algorithm, and heuristic procedures may be required. The experienced and well-trained operator can recognise the situation and react properly given sufficient time, but often not reliably or quickly enough [9–11].

In this paper, a process to obtain knowledge-base using support vector machines (SVMs) is presented for ready post-fault diagnosis purpose. SVMs are used as Intelligence tool for identifying the faulted line that is emanating from a substation and finding the distance from the substation. The approach described in this paper uses phasor values of the line voltages and currents after the fault has been detected. The approach is particularly important for post-fault diagnosis of any mal-operation of relays following a disturbance in the neighbouring line connected to the same substation. This may help to improve the fault monitoring/diagnosis process and coordination of the protective relays, thus assuring secure operation of the power systems. Our approach, based on SVMs, exploits the first part of this goal. For comparison, classifier and regression

tools based on the Radial Basis Function Neural Networks (RBFNNs) were also investigated. The RBFNNs and SVM networks are introduced and considered as an appropriate tool for pattern recognition problems. Results on IEEE-39 bus New England system are presented to illustrate the proposed method.

## 2 Proposed approach

In this paper, a method to obtain knowledge-base using SVMs is presented for ready post-fault diagnosis purpose. The method employs wide-area information system and intelligent technique to minimise the impact of an event on a network. There are two ways in which a wide-area post-fault diagnostic security system can ensure system stability and improve the reliability and availability of power supply:

1. Exact identification of faulted line at substation/Energy Control Centre (ECC) level and
2. Precise location of a fault for easy fault isolation.

When fault occurs on a system, current and voltage magnitudes change at many places in the network. Suppose we are interested in monitoring one of the substations in the power system, then the possible diagnosis are ‘where is the fault?, either it belongs to one of the transmission lines connected to the same substation or in other lines not related to the substation under diagnosis’. Of course, following a fault on one of the transmission lines, relay will operate to clear the fault, but this may cause tripping of other lines where fault has not occurred. In this paper, SVMs are used as an intelligent tool (as shown in Fig. 1) to allocate the coordination between relays in such a way that ‘Fault is occurred in other zonal area of the network, not in the lines emanating from the substation premises’. This approach is purely related to substation level, can be helpful post fault diagnosis of relay operations and so on. Even though the method appears to be substation level, can be designed for a wide-area protection in large networks using the proposed approach at an ECC as shown in Fig. 2.

Block description of the proposed method at a substation using the SVMs is illustrated in Fig. 3. The three phase voltages and currents measured at the substation during fault are supplied as an input to the SVMs and the expected output will be identifying the faulted line and fault location on the transmission line. The relation between the measurements and targets are build-up by SVM, which relates them

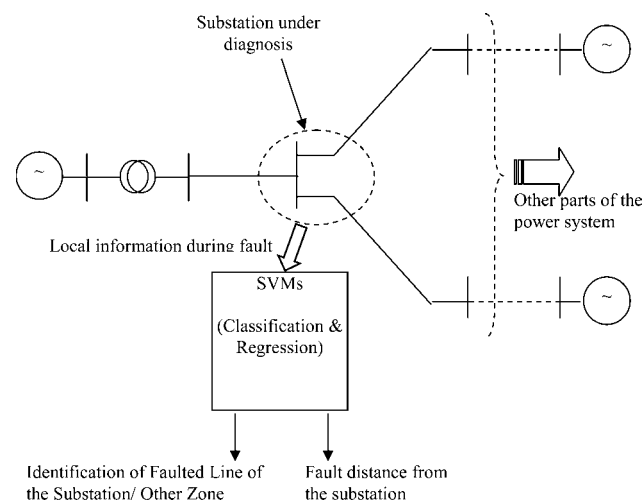


Fig. 1 Implementation of the proposed method at substation

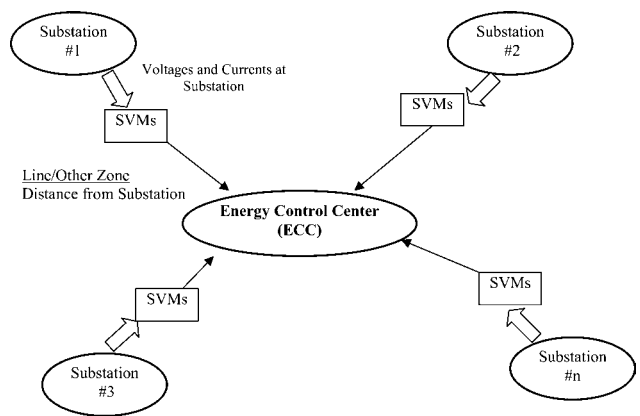


Fig. 2 Wide-area protection design for large networks

only in a condition when fault occurs in one of the transmission lines emanating from the substation. The variation of data points in the input space leads to the difficulty in mapping such a space by the classifier. In the proposed approach identification of faulted line among the transmission lines emanated from the substation is a ‘classification’ problem. This is solved as a classification problem using the techniques such as support vector (SV) classifiers, which have well-established advantages over other methods. Once the fault is on one of its (substation’s) transmissions lines, identifying the location of the fault is a ‘regression’ problem (function approximation), which is solved by  $\nu$ -support vector regression ( $\nu$ -SVR) approach [12].

## 3 Brief review on SVMs and RBFNNs

### 3.1 SVM Classification

SVMs are a new learning-by-example paradigm spanning a broad range of classification, regression and density estimation problems. Recently, a number of practical applications of SVMs have been reported [13–17]. They were first introduced by Vapnik [12] on the basis of structural risk minimisation principle [12] and were described in more detail by Schölkopf *et al.* [18]. The roots of this

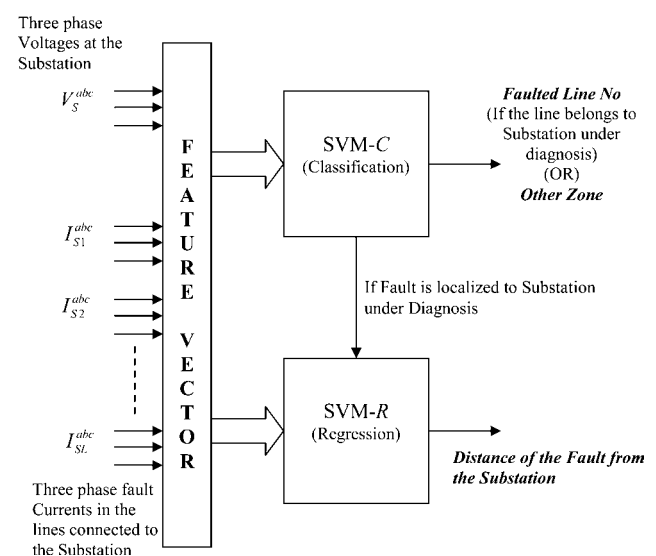


Fig. 3 Block description of the proposed method at substation; the input feature vector space is of  $(3 + 3 * L)$  dimension (consists of three phase voltages at the substation and three phase currents in the  $L$  number of lines emanating from the substation)

approach, the so-called SV methods construct the optimal separating hyperplane for pattern recognition. The SV technique was generalised for nonlinear separating surfaces in [18], and was further extended for constructing decision rules in the non-separable case. In the case when there are no separating hyperplane, the goal of SVM is to maximise the margin and to minimise the number of errors. The solution is a trade-off between the largest margin and the lowest number of errors. A brief review of support vector classification [19, 20–23] is presented in this section. Computing this hyperplane is equivalent to solve the following optimisation problem [12]

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \left( \sum_{i=1}^l \xi_i \right) \\ \text{s.t.} \quad & \mathbf{y}_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \forall i \end{aligned} \quad (1)$$

where  $\mathbf{x}_i$  is the  $i$ th example and  $\mathbf{y}_i$  the class label value which is either +1 or -1. This problem is computationally solved using the solution of its dual form

$$\begin{aligned} \max L_D = \quad & \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j \mathbf{y}_i \mathbf{y}_j (\mathbf{x}_i^T \mathbf{x}_j) \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq C \quad \forall i, \quad \sum_i \alpha_i \mathbf{y}_i = 0 \end{aligned} \quad (2)$$

### 3.2 Multi class classification

Basically, the two types of approaches usually followed for the extension from the binary two-class problem to  $n$  classes are 1. to modify the design of the SVMs to incorporate the multi-class learning in the quadratic solving algorithm [24] and 2. to combine several binary classifiers [25]. In the second case, methods like ‘one-against-all’ and ‘one-against-one’ have been proposed where typically a multi-class classifier is constructed by combining binary classifiers. In this paper ‘one-against-one’ method is used for multi-class classification, because of its less training time over ‘one-against-all’.

### 3.3 $\nu$ -Support vector regression

SVR and the  $\epsilon$ -insensitive loss function were introduced by Vapnik [13], where SVR was formulated in a similar way to SVM classification. Given a set of training samples  $\{(x_1, y_1), \dots, (x_l, y_l)\} \in \mathbf{R}^N \times \mathbf{R}$ , we want to learn a regression function  $f(x) = \mathbf{w}^T \cdot x + b$ ,  $\mathbf{w}$ ,  $\mathbf{x} \in \mathbf{R}^N$ ,  $b \in \mathbf{R}$  that can best predict unseen data  $x$  drawn from the same distribution as the training data. The regression function  $f(x)$  can be solved through an optimisation problem as presented in [13]. Parameter  $\epsilon$  ( $\epsilon \geq 0$ ) in the  $\epsilon$ -insensitive loss function controls the accuracy of the regressor. But the parameter  $\epsilon$  needs to be specified in advance. Smola *et al.* [26] presented a comprehensive tutorial on SVR, where the kernel trick, algorithms and variants of SVR were discussed. To facilitate tuning, SVR with automatic accuracy control in  $\nu$ -SVR was introduced by assembling  $\epsilon$  into the objective function [27].

### 3.4 Kernel choice

The use of kernel methods [28] provides a powerful way of obtaining nonlinear algorithms capable of handling non-separable datasets in the original input space. Different types of kernels used to train the SVMs are linear kernel  $\mathbf{Q}_{ij} = K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \cdot \mathbf{x}_j$ , polynomial kernel

$\mathbf{Q}_{ij} = (\gamma(\mathbf{x}_i \cdot \mathbf{x}_j) + r)^{\text{Degree}}$  with  $\gamma > 0$ , radial basis function (RBF) kernel  $\mathbf{Q}_{ij} = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$  where  $\gamma > 0$ , related with the kernel width and sigmoid kernel  $\mathbf{Q}_{ij} = \tanh(\gamma(\mathbf{x}_i \cdot \mathbf{x}_j) + r)$  where  $\gamma$  and  $r$  are kernel parameters.

The RBF kernel nonlinearly maps samples into a higher dimensional space, so it, unlike the linear kernel, can handle the case when the relation between class labels and attributes is nonlinear. Furthermore, the linear kernel is a special case of RBF as Keerthi *et al.* [29] show that the linear kernel with a penalty parameter  $\tilde{C}$  has the same performance as the RBF kernel with some parameters  $(C, \gamma)$ . In addition, the sigmoid kernel behaves like RBF for certain parameters [30]. The second reason is the number of hyper-parameters which influences the complexity of model selection. The polynomial kernel has more hyper-parameters than the RBF kernel. Finally, the RBF kernel has less numerical difficulties. Moreover, we must note that the sigmoid kernel is not valid (i.e. not the inner product of two vectors) under some parameters [12]. In this paper, we suggest that RBF kernel is a reasonable first choice for SVM training.

### 3.5 SVM model selection

In any predictive learning task, such as classification/regression, an appropriate representation of examples as well as the model and parameter estimation method should be selected to obtain a high level of performance of the learning machine. To obtain a good performance, some parameters in SVMs have to be chosen carefully. These parameters include:

- The regularisation parameter  $C$ , which determines the trade-off between minimising the training error and minimising model complexity;
- Kernel function  $K(\mathbf{x}_i, \mathbf{x}_j)$  and
- Parameter  $\gamma$  ( $\gamma = 1/2\sigma^2$ ) or  $d$  of the kernel function that implicitly defines the nonlinear mapping from input space to some high-dimensional feature space.

### 3.6 Radial basis function neural networks (RBFNNs)

The RBFNN [31, 32] is a feed forward neural network with an input layer, a nonlinear hidden layer and a linear output layer, realises a mapping function  $f: \mathbf{x} \rightarrow f(\mathbf{x})$ ; where  $\mathbf{x}$  is an  $N$ -dimensional vector and  $f(\mathbf{x})$  is the scalar output, obtained to generate the decision function. The RBFNN output is given by  $f(\mathbf{x}) = \sum_{i=1}^m \mathbf{w}_i \phi_i(\mathbf{x}, c_i) + b$ , where  $m$  is the number of hidden layer neurons,  $\mathbf{w}_i$  and  $\mathbf{b}$  are the weights and bias vectors. The hidden nodes are the RBF units  $\phi_i = \exp(-\|\mathbf{x} - c_i\|^2/2\sigma^2)$ ,  $\phi: \mathbf{R}^N \rightarrow \mathbf{R}$ , where  $c_i \in \mathbf{R}^N$  and  $\sigma$  are the RBF centres and width, respectively. The performance of RBFNNs depends on the choice of the values of the centres. Training of RBFNNs involves selecting the centres  $c_i$ , estimating the weights  $\mathbf{w}_i$  and bias  $\mathbf{b}$  that connect the hidden and the output layers.

## 4 SVM training

Transmission lines are frequently subjected to wide-variety of faults. Different types of faults can occur including phase faults among two or more different conductors or ground faults including one or more conductors to ground types. The commonly occurred faults are 3-Ph, Line-to-Ground (SLG), Line-to-Line (LL) and Double Line-to-Ground

(LLG). And the fault can occur at any point on the line. In this paper, the training patterns for the SVM are generated by creating the above four types of faults with varying fault locations on the system lines.

Digital computer tool such as Electro Magnetic Transients Program (EMTP) [33,34] is universally accepted as industry standard for evaluating electromagnetic transients. An EMTP developed by D. Thukaram at Indian Institute of Science, Bangalore, India, which has all features like frequency dependence, saturation of reactors, lighting arrestors, nonlinearity of various components and so on, has been extensively used in planning studies of Indian Extra High Voltage system. We have used this EMTP program to simulate different faults.

The approaches described in this paper use phasor values of the line voltages and currents. The simplest method of forming phasor values from sampled data values is to use the discrete fourier transform (DFT) evaluated at power system frequency. The result from the DFT will be in terms of real and imaginary parts that can then be readily converted to phasor form where necessary. The output of DFT contains phasor values of currents and voltages of all the three phases. When the fault is detected, the present fundamental values of currents and voltages are frozen in a buffer for six cycles (or till the circuit breaker operates). The phasor quantities of current and voltage extracted in this way are presented to SVMs.

The input pattern consists of phase voltages ( $v^a$ ,  $v^b$  and  $v^c$ ) and phase currents ( $i^a$ ,  $i^b$  and  $i^c$ ) in all the lines available at the substation during fault. Dimension of the input pattern depends on the number of lines connected by the substation. For example, substation 'S' is under diagnosis and is connected by  $L$  number of lines to other parts of the power system, then the input vector dimension is  $(3 + 3 * L)$ . For classification problem, the decision function is mapping from  $R^{(3+3*L)}$  to  $\{1, 2, \dots, L, L+1\}$  (where  $1-L$  are class labels for the  $L$  'lines connected by the substation' and  $(L+1)$  is the class label for the 'lines not connected by the substation (Other Zone)' and for regression problems, the decision function maps from  $R^{(3+3*L)}$  to  $R$  (% value of fault distance from the substation for localised faults).

#### 4.1 Design issues in SVM-C (classification)

Our primary goal is 'exact identification of the faulted line'. To achieve this goal, a SVM-C model as shown in Fig. 4 is developed. It has two phases of classification. In Phase 1, the SVM-1 identifies the type of the fault (3-Ph/SLG/LL/LLG, whose class labels are 1, 2, 3 and 4, respectively). In Phase 2, on the basis of the fault type obtained in Phase 1, one of the SVMs in Phase 2 (SVM-2 to SVM-5) is selected

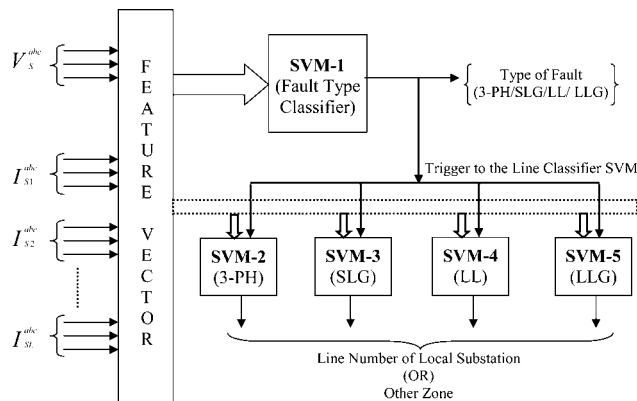


Fig. 4 Architecture of SVM-C model for fault line identification

to identify the faulted line/area. The output from Phase 2 is either the 'line number' (in case the fault is on one of the lines connected to substation under diagnosis) or 'Other Zone'.

#### 4.2 Design issues in SVM-R (regression)

Our next goal of finding the fault distance from the substation is initiated when the SVM-C (shown in Figs. 3 and 4) declares that the fault is on one of the substation's line. In this regression problem of fault identification, SVM-R uses the fault type information obtained from SVM-C and the same local information available at the substation during the fault. As shown in Fig. 5, the design of the SVM-R consists of four  $v$ -SVR blocks. One of these blocks is triggered based on the fault type identified by the SVM-C. To train each one of these SVR blocks, patterns are generated for each fault type at different locations on the lines with fault resistance varied.

At this point, it is important to notify that all the SVM blocks presented in Figs. 4 and 5 are re-established with RBFNNs and results are presented for the comparison purpose.

### 5 System studies

The proposed strategy is explained with the IEEE-39 bus New England system. The single line diagram of the system is shown in Fig. 6. The system has 10 generators, 12 transformers, 34 transmission lines. The total system can be monitored for faults by monitoring multiple substations such that all the lines in the system are covered. In this paper, the substation bus number 16 is selected for monitoring. The Bus 16 is connected by five transmission lines to other substations numbered 15, 17, 19, 21 and 24 buses. The transmission lines 16-15, 16-17, 16-19, 16-21 and 16-24 are labelled as Class 1, 2, 3, 4 and 5, respectively. The other lines that are not connected to the substation 16, for example, 26-28, 15-14, 9-39 and so on are collectively labelled as class 6.

#### 5.1 Training and test patterns

Training patterns are extracted from EMTP simulations (as explained in Section 4) run for the four types of faults at different locations on the five transmission lines (for the five classes) and on some selected lines from other parts of the system (represent the Class 6). The selected lines for

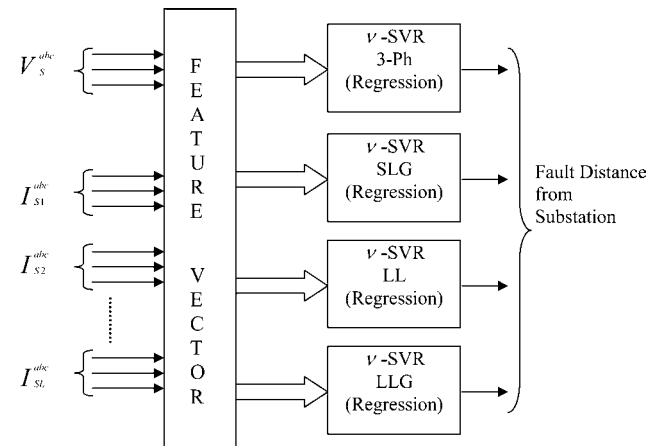
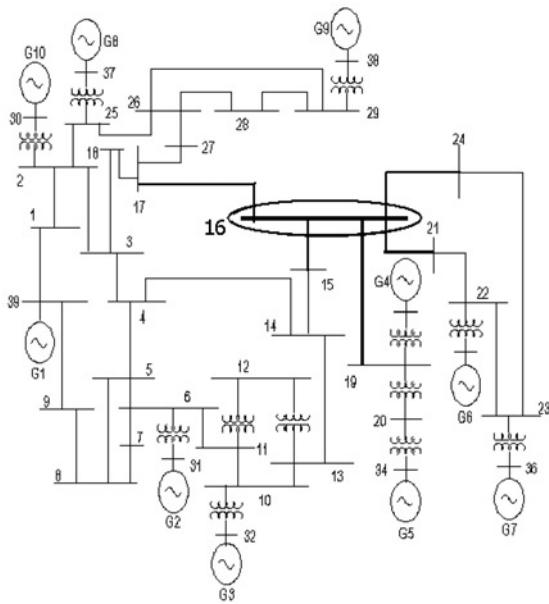


Fig. 5 Block description of the  $v$ -SVR for identification of the fault location on the line



**Fig. 6** Single line diagram of IEEE-39 bus New-England system; Bus No. 16 is selected for fault diagnosis

Class 6 are 21–22, 22–23, 23–24, 17–18, 17–27, 3–18, 2–3, 7–8, 13–10 and 15–14. The faults are created at distances of 5, 15, 25, . . . , 95% of their overall transmission length. During the simulation, the fault resistance values are varied over the values 0, 2, 5, 10, 20, 50 and 100  $\Omega$ . Over all, the training patterns are generated for 4 types of faults on 15 transmission lines over 10 locations with varying 7 impedance values. For each type of fault, the number of training patterns generated is  $15 \times 10 \times 7 = 1050$  patterns.

Test patterns are extracted from EMTP simulations (as explained in Section 4) by creating the four types of faults with six fault resistance values (that are not used in generating the training patterns), on the five substation's lines and on 10 lines from other parts of the system. During simulations by EMTP, the shorting resistance values are varied over the six values of 3, 7, 15, 30, 40 and 60  $\Omega$ . The selected lines for generating test patterns are line 3–4, 13–14, 26–27, 21–22, 22–23, 23–24, 17–18, 17–27, 3–18 and 15–14. In this case, the faults are created at 20, 40, 60 and 80% distances of their overall transmission line length. The number of test patterns generated is 15 lines  $\times$  6 fault resistances  $\times$  4 locations = 360 patterns per each fault type. The details of simulations carried out using EMTP for generating the training and test patterns are given in Table 1.

The input patterns (training and test patterns) are normalised to  $[-1, +1]$  before inputting to the SVM module. For the normal scaling method, if the maximum and minimum

values of the  $i$ th attributes are  $M_i$  and  $m_i$ , respectively, then, scaling to  $[-1, +1]$  means  $x^1 = 2(x - m_i) / (M_i - m_i) - 1$ .

To obtain a feel of the pattern scattering of data points in the input space, the variable reduction or dimensionality reduction helps to show the input space in 3D. The dimensionality reduction is carried out using the principal component analysis (PCA) technique [35,36]. A low-dimensional feature vector is a critical element in any pattern recognition problem as the number of data examples needed for training grows explosively with the dimension of the problem. The PCA, which is a feature reduction technique, is fast and characterised by minimal loss of information.

In general, the dimensionality of the dataset is reduced to the number of significant Eigen values of the covariance matrix. Although these dimensions do not have any physical meaning in the fault monitoring and location problem, that is depict the most important aspect of the function approximation problem, that is, the way in which the data are scattered in the reduced dimensional space.

For 3-Ph fault, Figs. 7a–g show the data points in 3D space along the three major eigenvectors. Fig. 7a shows the scattering of the data points for all the six classes. Fig. 7b is showing the spread of data points for faults on the substation lines. Figs. 7c–g shows the scattering of data points in 3D space for faults on individual transmission lines (classes from 1 to 5 respectively) connected by the Bus 16.

## 5.2 Model design

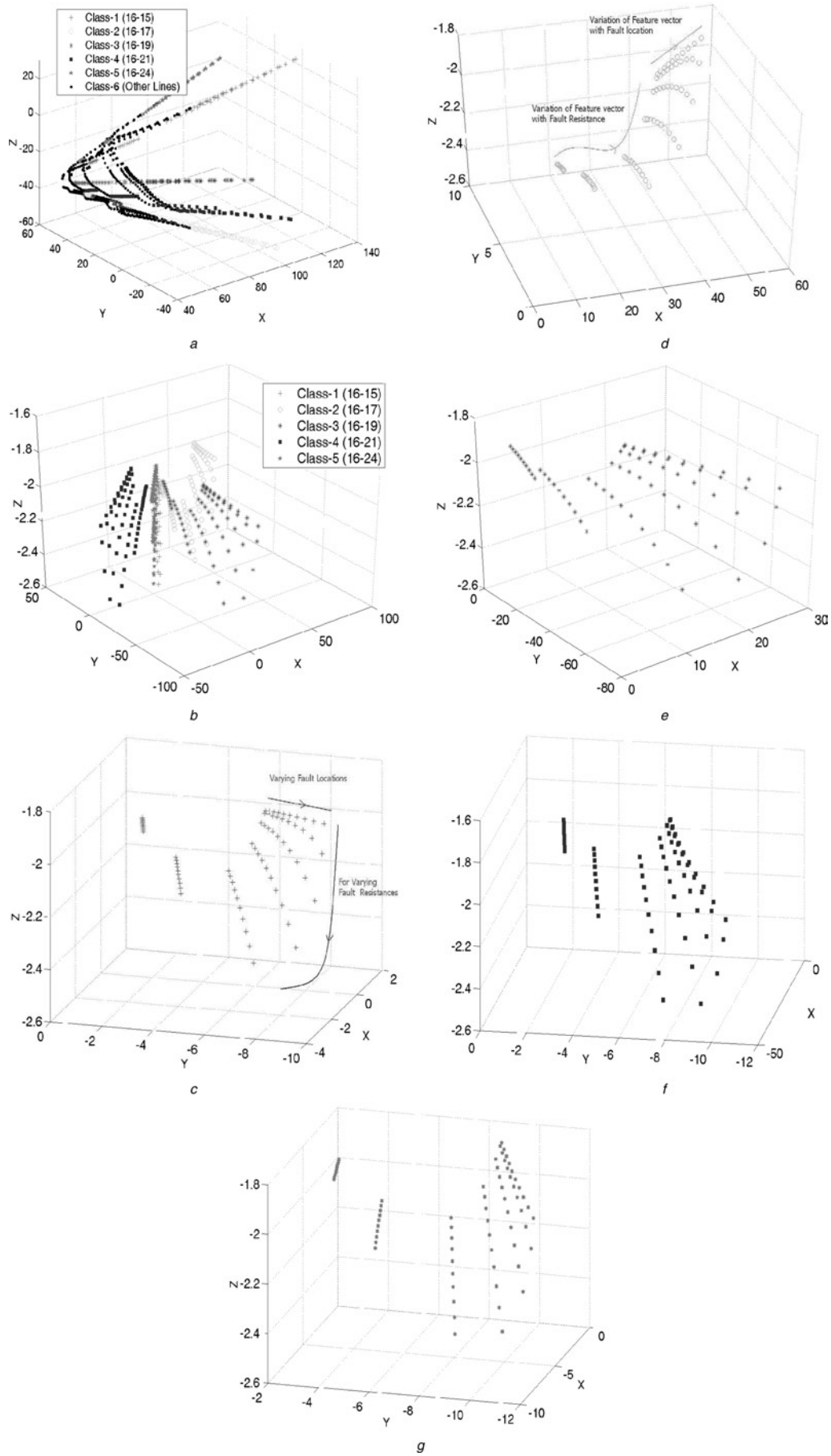
Training of SVM requires selecting the cost function (C) and kernel function parameters, which influence the ensuring model performance. In our simulations, we have considered the RBF as kernel function. RBF kernel is advantageous in complex non-separable classification problems due to its ability of nonlinear input mapping. So proper selection of the parameter  $\gamma\{\gamma = 1/(2\sigma^2)$ , where  $\sigma$ : kernel width} is important in selecting a good RBF kernel. In this paper, LIBSVM [37,38] is used for training the SVMs in classification and regression modules.

## 5.3 Parameter selection

**5.3.1 Parameter selection for RBFNNs:** The various RBFNNs required for the classification of faulted line section and fault type have been trained with the training data. Selection of proper values of spread is important in designing an RBFNN. The spread determines the width of the RBFs. The spread should be larger than the minimum distance and smaller than the maximum distance between the input vectors [39]. After a number of simulations, the observed optimal values of spreads for different RBFNNs, RBFNN-1 to RBFNN-5 are 0.5, 1.1, 0.5, 0.9 and 0.707,

**Table 1: Details of training and test patterns generation**

Line	Class label	Fault distances from substation	
		Training	Testing
16–15	1	at distances of 5, 15, . . . , 95% of	at distances of 20, 40, 60 and 80
16–17	2	their total line length and the	of their total line length and
16–19	3	fault resistance values are	the fault resistance values are
16–21	4	varied over the values 0, 2, 5,	varied over the six values of 3,
16–24	5	10, 20, 50 and 100 $\Omega$ .	7, 15, 30, 40 and 60 $\Omega$ .
lines not connected by Bus 16	6		



**Fig. 7** Principal component analysis applied to the input training patterns of 3-Ph faults

Dimension is reduced from 18 to 3 for visualisation purpose

*a* 3-Ph fault input patterns for all the six classes

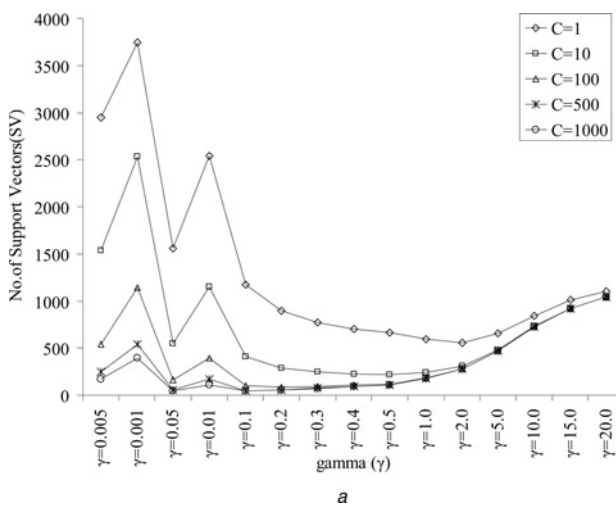
*b* Input patterns for 3-Ph faults on the five lines connected by Bus 16

*c–g* Input patterns for 3-Ph faults on the lines 16–15 (Class 1), 16–17 (2), 16–19 (3), 16–21 (4) and 16–24 (5), respectively

**Table 2: Hidden neurons and spreads relating to different RBFNNs at various schemes**

Block name	Number of hidden neurons	Spread of RBFNN	% Testing accuracy
RBFNN-1	750	1.0	99.12
	650	0.5	99.46
	550	1.05	99.31
RBFNN-2	300	0.9	97.57
	275	1.1	98.7
	325	0.707	95.15
RBFNN-3	260	0.5	98.7
	275	0.95	98.54
	300	1.05	94.24
RBFNN-4	225	0.707	97.57
	250	0.9	98.17
	275	1.1	97.1
RBFNN-5	260	1.0	95.6
	250	0.707	99.46
	275	1.1	93.10

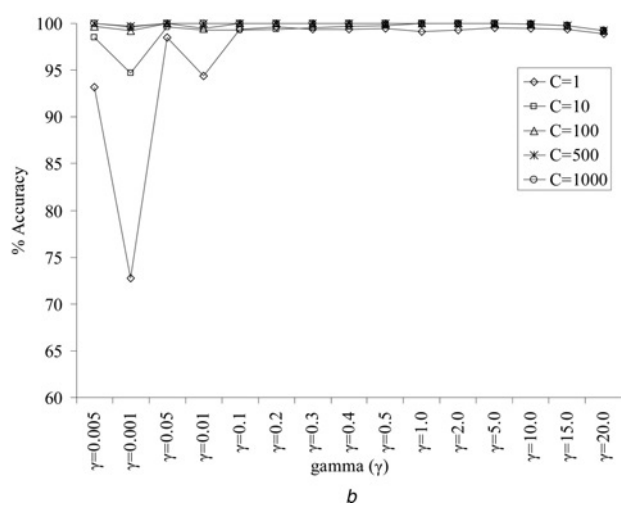
respectively. After training is over, each of the RBFNNs is tested with test data created by varying distances and resistances. Some significant test results for various RBFNNs are represented in Table 2. In summary, RBFNN classifiers with highest accuracy corresponding to best configuration and will be considered for further comparison.



**5.3.2 Trail and error selection for SVMs:** The various SVM classifiers are trained with several values of  $C$  and  $\gamma$  to guess the combination of parameters that might be the best for a ‘good’ model. The good model is expressed through the evaluation of performance accuracy. The SVMs (SVM-1 to SVM-5) are trained with different values of penalty terms  $C$  and  $\gamma$ . The range of  $C$  chosen for training is [1, 1000] and  $\gamma$  values ranges between [0.005, 20.0]. The stopping tolerance for solving the optimisation problem is set to 0.00001. Unbalance in the input data is taken care by selecting different penalty terms for different classes of data.

In a first series of experiments, we run the classifier with several values of  $C$  and  $\gamma$  somehow trying to guess which combination of parameters might be the best for a ‘good’ model. For each combination of  $C$  and  $\gamma$ , the network undergoes learning and retrieval modes. During learning mode, the network is trained with training data and is tested on the test data during retrieval mode. The test data for validation are generated as explained in Section 4 using EMTP. Fig. 8 shows the results for SVM-1 of  $a$  the variation of number of SVs and  $b$  testing accuracy as a function of kernel  $\gamma$  both parameterised with  $C$ . In this section, we have collected the best model parameters optimised for ‘number of SVs’ and ‘testing accuracy’. We compare, in Table 3, the number of iterations, the number of SVs and the testing accuracy of the learning machine with  $C$  set to 1 and 500.

Practical issues: Larger  $C$  corresponds to less number of SVs as well as higher testing accuracy although over-fitting cannot thus be avoided. Further explanation is required for these results taking into account both  $C$  and  $\gamma$  parameters.

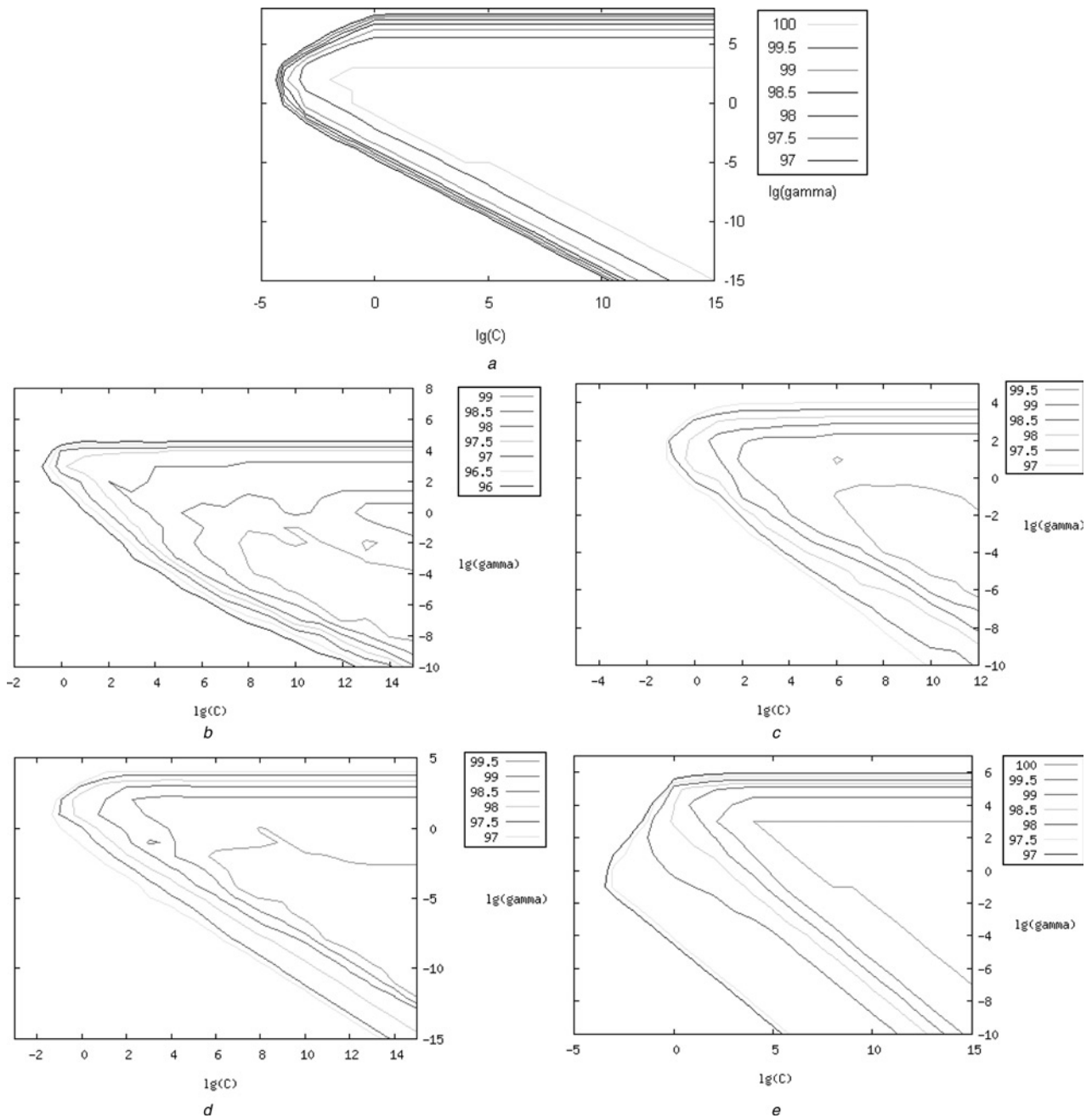


**Fig. 8 Model selection with trail and error using  $C$ ,  $\gamma$  for SVM-1 (fault type) classifier**

$a$   $\gamma$  against number of support vectors  
 $b$   $\gamma$  against % testing accuracy

**Table 3: Details of number of iterations, number of SVs and testing accuracy of different classifiers for  $C = 1, 500$**

Classifier name	$C = 1$			$C = 500$		
	No. of iterations	No. of SVs	% Testing accuracy	No. of iterations	No. of SVs	% Testing accuracy
SVM-1	20 684	558	99.31	9826	108	100
SVM-2	3586	289	98.18	7687	80	100
SVM-3	4454	264	99.54	5568	64	100
SVM-4	3020	232	99.84	7448	61	100
SVM-5	1681	151	99.54	2626	57	100



**Fig. 9** Interactive grid search graphic contours for the SVMs after a five-cross validation

*a* Parameter selection using interactive grid search for classifier SVM-1; the grid search is on for  $C = 2^{-5}, 2^{-4}, \dots, 2^{15}$  and  $\gamma = 2^{-15}, 2^{-14.5}, 2^{-14.0}, \dots, 2^{5.0}$

*b–e*: Similarly for classifiers SVM-2(SVM 3-Ph) to SVM-5(SVM LLG)

In fact, minimising  $(1/2) \|\mathbf{w}\|^2$  in (1) corresponds to maximise the margin  $2/\|\mathbf{w}\|^2$  between two classes of data. For non-separable data, the penalty term  $C \sum_{i=1}^l \xi_i$  is able to reduce the training errors in the working data set.

Therefore the margin is an indicator of the generalisation accuracy. In the absence of a method to compute the best trade-off between the regularisation term and the training errors, the balance sought by the SVMs technique is hard

**Table 4: Optimised parameters obtained grid search analysis for different SVMs with RBF kernels**

Classifier name	$C$	$\gamma$	No. of iterations	No. of SVs	% Training accuracy	% Testing accuracy
SVM-1	4.0	0.1250	2098	559	100	99.55
SVM-2 3-Ph	1024	0.0625	12 035	63	99.33	100
SVM-3 SLG	4096	0.0156	29 518	62	99.81	100
SVM-4 LL	256	0.125	9613	64	99.81	100
SVM-5 LLG	256	0.5	5153	66	100	100



**Table 5: Optimised parameters for fault locating RBFNNs**

Regression block	Spread	Number of HN
RBFNN-3-Ph	0.95	140
RBFNN-SLG	0.9	180
RBFNN-LL	1.05	175
RBFNN-LLG	0.707	175

to find. Thus, a larger  $C$  corresponds to assign a higher penalty of training errors and clearly over-fitting occurs. On the other hand, when the kernel parameter  $\gamma$  becomes higher, the greater the variety of the decision boundaries that can be formed, originate a more complex model. The added flexibility decreases initially the generalisation error as the model can better fit the data.

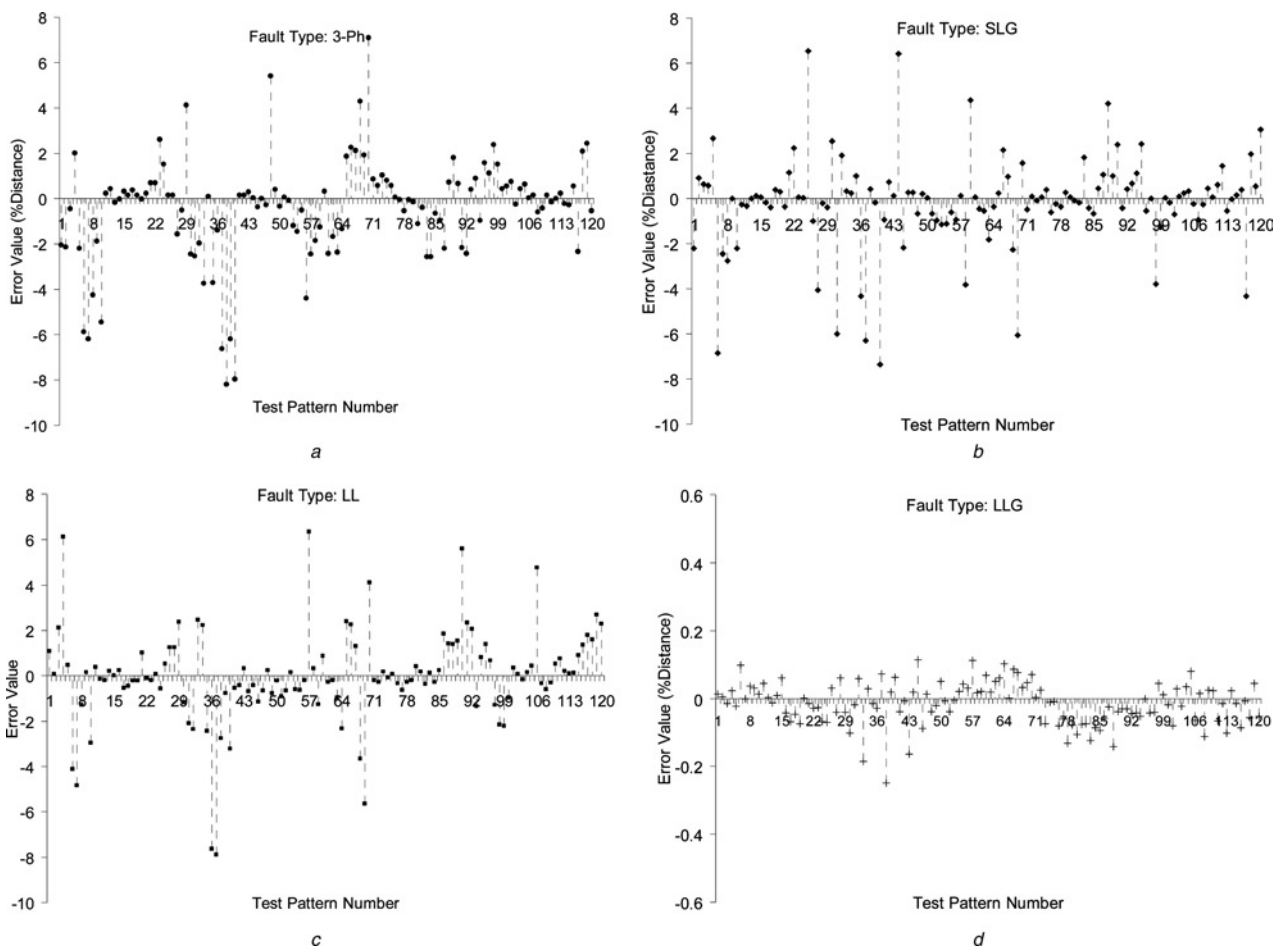
**5.3.3 Interactive grid search selection for SVMs:**

Choosing the best parameters can be time consuming if a systematic approach is not used and/or the problem knowledge do not aid for proper selection. Therefore an interactive grid search model selection has been accomplished for each one of the SVMs and the generalised accuracy evaluated on the training data. Fig. 9 portrays the generalisation graphic contours for the SVMs after a five-cross validation, thus, reducing the search space. The efficient heuristic way of searching points in that space with small

generalisation errors will lead to a good understanding of the hyper-parameter space [29]. We can then do a refined search of the  $(C, \gamma)$  pairs for proper model selection. Fig. 9a shows the parameter selection using interactive grid search for classifier SVM-1. The grid search is on for  $C = 2^{-5}, 2^{-4}, \dots, 2^{15}$  and for  $\gamma = 2^{-15}, 2^{-14.5}, 2^{-14.0}, \dots, 2^{5.0}$ . The cross validation accuracy resulted for SVM-1 is 100% on the training data with extracted model parameters of  $C = 4.0$  and  $\gamma = 0.125$ . Once the SVM-1 is learned with these parameters, all parameters of the trained SVM-1 have been frozen and then used in retrieval mode for testing the capabilities of the system on the data not used in learning. The test data samples have been extracted using the EMTP as explained in Section 4 and Table 1. The results obtained in terms of % testing accuracy (No. of samples correct classified\*100/total number of samples presented) are given in Table 4.

The above process of grid search using trained data is retrieval mode of operation on test data are repeated for SVM-2 to SVM-5. The extracted model parameters with their training and testing accuracies are given in Table 4. Table 4 illustrates the results obtained choosing the pair of parameters conveys to the learning model with the smallest capacity and, thus, the highest generalisation.

If we compare with the RBFNNs, we may conclude that SVMs are more accurate and allow better generalisation than the former. Besides, in the SVMs technique there is no heuristic choice for model design as it is required in the RBFNNs. Also, the different solution method comparison in both techniques; the Quadratic Programming



**Fig. 10** Error values in terms of % distances from actual fault locations are simulated for different fault types with RBFNNs a–d Error value plots from RBFNN-3-Ph, RBFNN-SLG, RBFNN-LL and RBFNN-LLG, respectively, when optimised parameters are used

**Table 6: Testing accuracy in identifying the fault location in terms of MSE from  $\nu$ -SVR with different  $C$ ,  $\nu$  and  $\gamma$  values**

$\nu$ -SVR 3-Ph		$\nu$ -SVR SLG		$\nu$ -SVR LL		$\nu$ -SVR LLG	
{kernel, $C$ , $\nu$ , $\gamma$ }	MSE	{kernel, $C$ , $\nu$ , $\gamma$ }	MSE	{kernel, $C$ , $\nu$ , $\gamma$ }	MSE	{kernel, $C$ , $\nu$ , $\gamma$ }	MSE
$r$ , 100, 0.5, 0.1	$12.38 \times 10^{-4}$	$r$ , 100, 0.5, 0.1	$3.91 \times 10^{-4}$	$r$ , 10, 0.5, 0.05	$3.91 \times 10^{-4}$	$r$ , 100, 0.1, 0.5	$2.4 \times 10^{-07}$
$r$ , 1000, 0.5, 0.05	$13.32 \times 10^{-4}$	$p(3)$ , 10, 0.5, 0.5	$3.97 \times 10^{-4}$	$p(3)$ , 10, 0.1, 0.5	$3.92 \times 10^{-4}$	$r$ , 1000, 0.01, 0.5	$3.27 \times 10^{-07}$
$r$ , 1000, 0.5, 0.1	$24.23 \times 10^{-4}$	$r$ , 1000, 0.5, 0.05	$4.42 \times 10^{-4}$	$p(3)$ , 1000, 0.1, 0.1	$4.04 \times 10^{-4}$	$p(3)$ , 1, 0.5, 1.0	$4.11 \times 10^{-07}$
$P(3)$ , 1000, 0.5, 0.5	$47.46 \times 10^{-4}$	$p(3)$ , 1.0, 0.5, 1.0	$4.439 \times 10^{-4}$	$p(3)$ , 10, 0.5, 0.1	$4.05 \times 10^{-4}$	$p(3)$ , 1000, 0.5, 0.1	$4.2 \times 10^{-07}$

$r$ , RBF kernel,  $p(d)$ , polynomial kernel with degree  $d$

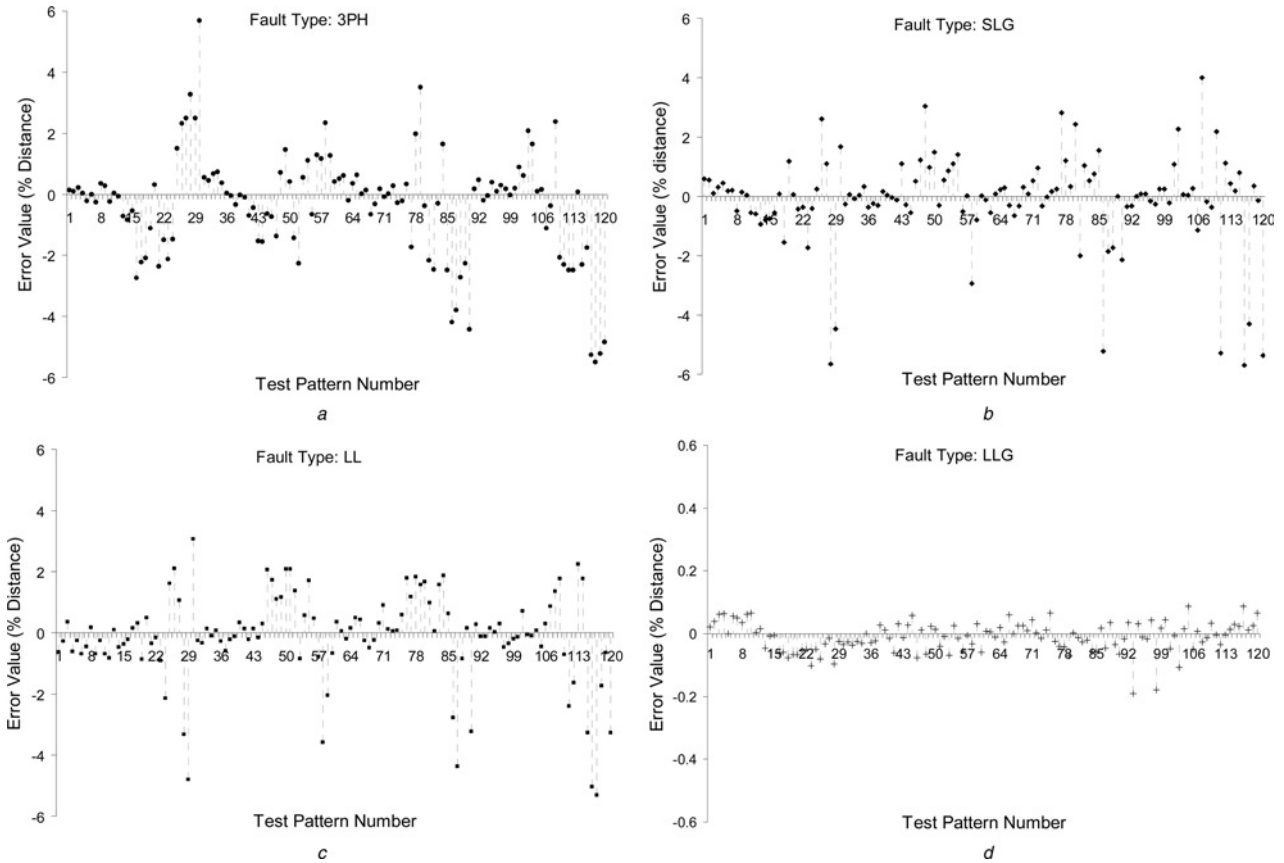
problem in SVMs formulation is simpler to solve, conveying to a technique without local minima.

#### 5.4 Identifying location of faults

**5.4.1 Using RBFNNs:** Once the faulted line section with fault type is identified, our second goal is to identify the location of the fault from the substation. Fault location requires more accuracy as compared to faulted line section identification. The various RBFNNs are trained with the training data of 350 samples. After carrying a number of simulation studies, Table 5 shows the best parameters from the results obtained. After training phase is over, each of the RBFNNs are tested for error values. The test data consist of 120 samples and the error values in terms of %distances  $\{(Distance\ obtained\ from\ RBFNN - Real\ distance\ of\ fault) \times 100 / Total\ line\ length\}$  from actual fault locations simulated for different fault types are shown in Fig. 10. Fig. 10a shows the error values obtained from RBFNN-3-Ph with spread 0.95 and having 140 neurons in hidden layer. Similarly, Figs. 10b–d show

the error deviations from the target values for the regression blocks RBFNN-SLG [spread = 0.9, No. of hidden neurons (HN) = 180], RBFNN-LL (spread = 1.05, No. of HN = 175) and RBFNN-LLG (spread = 0.707, No. of HN = 175), respectively. The maximum and minimum values of error from the 3-Ph, SLG, LL and LLG fault type RBFNNs locators are (8.20%, 0.0121%), (7.36%, 0.0014%), (7.89%, 0.0135%) and (0.25%,  $2.0 \times 10^{-4}$ ), respectively. There are 19 test patterns (out of 120 test patterns) have crossed the 2.5% error values for the 3-Ph faults. The number of test patterns that crossed 2.5% error values for SLG, LL and LLG faults types 18, 15 and 0, respectively.

**5.4.2 Using  $\nu$ -SVR:** Once the faulted line is identified from the SVM classifiers, approximate identification of the fault location on the transmission line is important. This is achieved by training the corresponding fault type  $\nu$ -SVR. To model the  $\nu$ -SVRs, two types of kernels, polynomial and RBF, are chosen. A series of experiments were run on the  $\nu$ -SVR blocks with several values of  $C$ ,  $\gamma$  and  $\nu$  to



**Fig. 11** Error values in terms % distances plotted for different fault types with  $\nu$ -SVRs using RBF kernel

a–d Error value plots from  $\nu$ -SVR 3-Ph,  $\nu$ -SVR SLG,  $\nu$ -SVR LL and  $\nu$ -SVR LLG, respectively when optimised parameters are used

guess which combination of parameters might be the best. For each combination, the network undergoes learning and retrieval modes. During learning mode, the network is trained on the training data and tested on the test data in retrieval mode. The value of the penalty parameter range is trailed between 1 and 1000. Aforementioned,  $\gamma$  is an important parameter for both the kernel functions. The range of  $\gamma$  is varied as [0.001, 20]. After a series of experiments, the best four results obtained for the four types of  $\nu$ -SVR are given in Table 6. It gives the information related to the variation of testing accuracy in terms of Mean Squared Error (MSE) for different combinations of the  $\nu$ -SVR parameters. From the results,  $\nu$ -SVR with RBF kernel are more accurate in identifying location of fault as compared to the RBFNNs.

Fig. 11 shows the error values in terms of % distances, simulated for different fault types using RBF kernel. The error values have been calculated using the expression  $\{(\text{Distance obtained from } \nu\text{-SVR} - \text{Real distance of fault}) \times 100 / \text{Total line length}\}$ . Fig. 11a shows the error values resulted from  $\nu$ -SVR for 3-Ph faults with the parameter combination  $C = 100.0$ ,  $\nu = 0.5$  and  $\gamma = 0.1$ . The maximum and minimum error value in case of 3-Ph faults is noted to be 5.68 and 0.125%, respectively. Similarly, Figs. 11b–d show the error values obtained from the regression blocks  $\nu$ -SVR SLG ( $C = 100.0$ ,  $\nu = 0.5$  and  $\gamma = 0.1$ ),  $\nu$ -SVR LL ( $C = 10.0$ ,  $\nu = 0.5$  and  $\gamma = 0.05$ ) and  $\nu$ -SVR LLG ( $C = 100.0$ ,  $\nu = 0.1$  and  $\gamma = 0.50$ ), respectively. The maximum and minimum error value from the SLG, LL and LLG fault type  $\nu$ -SVR locators are (5.69, 0.001), (5.31, 0.01%) and (0.19, 0.0001%), respectively. The maximum errors may happen for the testing data at conditions highly different from the training data. There are 12 test patterns (out of 120 test patterns) have crossed the 2.5% error values for the 3-Ph faults. A number of test patterns that crossed 2.5% error values for SLG, LL and LLG faults types are 12, 11 and 0, respectively. From the results,  $\nu$ -SVR with RBF kernel is accurate enough in identifying location of fault.

## 6 Conclusions

In this paper, we have discussed the application of the SVMs and  $\nu$ -SVRs compared to RBFNNs to post-fault diagnosis in power transmission system. The basic idea of the SVMs is to determine the structure of the classifier by minimising the bounds of the training error and generalisation error.

Results indicate that, in general, the test accuracy of RBFNNs is a little worse than that of the SVMs in cases of the fault type and faulted line section identification. In case of locating the faults,  $\nu$ -SVRs are more accurate over RBFNNs. Within the chosen SVMs framework, we show that over-fitting can be avoided by a interactive grid selection search on parameter's space. Regarding the implementation issues, SVMs are considerably faster than the training of RBFNNs, even for large-size problems, requiring less heuristics and, thus, being preferable. In addition, SVMs attain sparse solutions in the sense that most of the coefficients are set to zero as a result of the optimisation problem. This property is also computationally attractive. Besides, the expected ratio of the number of SVs and the number of training points bounds the generalisation error. Our results demonstrate that the SVMs have the potential to obtain a reliable post-fault diagnostic system to assist operators to make correct decisions.

Briefly, the use of SVMs as a powerful tool for substation diagnosis and fault location identification is presented.

SVMs with RBF kernel are able to learn the underlying concept between voltages and currents with the faulted transmission line and fault location. From the results, we conclude that, although measurements observed during fault in practical power systems are very limited, they contain significant information about the substation status and location of fault. They can be processed by an approach as described in this paper to obtain an efficient knowledge based fault diagnosis system.

## 7 References

- 1 Phadke, A.G., and Thorp, J.S.: 'Expose hidden failures to prevent cascading outages', *Comput. Appl. Power IEEE*, 1996, pp. 20–23
- 2 [ftp://www.nerc.com/pub/sys/all\\_updl/docs/blackout/BlackoutTable.pdf](ftp://www.nerc.com/pub/sys/all_updl/docs/blackout/BlackoutTable.pdf)
- 3 [ftp://www.nerc.com/pub/sys/all\\_updl/docs/blackout/ch7-10.pdf](ftp://www.nerc.com/pub/sys/all_updl/docs/blackout/ch7-10.pdf)
- 4 <http://www.eecs.wsu.edu/~mani/blackouts.PDF>
- 5 [http://www.columbia.edu/itc/sociology/watts/w3233/client\\_edit/power\\_outages.html](http://www.columbia.edu/itc/sociology/watts/w3233/client_edit/power_outages.html)
- 6 Thorp, J.S., and Phadke, A.G.: 'Protecting power systems in the post-restructuring era', *Comput. Appl. Power IEEE*, 1999, **12**, (1), pp. 33–37
- 7 Horowitz, S.H., and Phadke, A.G.: 'Boosting immunity to blackouts', *Power Energy Mag. IEEE*, 2003, **1**, (5), pp. 47–53
- 8 Phadke, A.G., and Horowitz, S.H.: 'Adaptive relaying', *Comput. Appl. Power IEEE*, 1990, **3**, (3), pp. 47–51
- 9 Karlsson, D., Morten, H., and Lindhal, S.: 'Wide area system monitoring and control', *IEEE Power Energy Syst.*, 2004, **2**, pp. 68–76
- 10 Begovic, M., Novosel, D., Karlsson, D., Henville, C., and Michel, G.: 'Wide-area protection and emergency control', *Proc. IEEE*, 2005, **93**, (5), pp. 876–891
- 11 Thukaram, D., Khincha, H.P., and Vijaynarasimha, H.P.: 'Artificial neural network and support vector machine approach for locating faults in radial distribution systems', *Trans. Power Deliv. IEEE*, 2005, **20**, (2), pp. 710–721
- 12 Vapnik, V.N.: 'The nature of statistical learning theory' (Springer-Verlag, New York, 1995)
- 13 Salat, R., and Osowski, S.: 'Accurate fault location in the power transmission line using support vector machine approach', *Power Syst., IEEE Trans.*, 2004, **19**, (2), pp. 979–986
- 14 Janik, P., and Lobos, T.: 'Automated classification of power-quality disturbances using SVM and RBF networks', *Power Deliv. IEEE Trans.*, 2006, **21**, (3), pp. 1663–1669
- 15 Janik, P., Lobos, T., and Schegner, P.: Eighth IEE International Conf. on Developments in Power System Protection, 5–8 April 2004, vol. 2, pp. 768–771
- 16 Moulin, L.S., da Silva, A.P.A., El-Sharkawi, M.A., and Marks, R.J. II: 'Support vector machines for transient stability analysis of large-scale power systems', *IEEE Trans. Power Syst.*, 2004, **19**, (2), pp. 818–825
- 17 Moulin, L.S., da Silva, A.P.A., El-Sharkawi, M.A., and Marks, R.J. II: 'Neural networks and support vector machines applied to power systems transient stability analysis', *Int. J. Eng. Intell. Syst.*, 2001, **9**, (4), pp. 205–211
- 18 Schölkopf, B., Burges, C., and Smola, A.: 'Advances in kernel methods—support vector learning' (MIT Press, Cambridge, MA, 1999)
- 19 Vapnik, V.: 'Statistical learning theory' (Wiley, New York, NY, 1998)
- 20 Sastry, P.S.: 'An introduction to support vector machine, chapter from computing and information sciences: recent trends' (Narosa Publishing House, New Delhi, 2003)
- 21 Burges, C.J.C.: 'A tutorial on support vector machines for pattern recognition', *Data Mining Knowl. Discov.*, 1998, **2**, (2), pp. 955–974
- 22 Joachims, T.: 'Making large-scale SVM learning practical. Advances in Kernel Methods – Support Vector Learning' (MIT Press, Cambridge, MA, 1998)
- 23 Platt, J.: 'Fast training of support vector machines using sequential minimal optimization, advance in kernel methods: support vector learning' (MIT Press, Cambridge, MA, 1999), pp. 185–208, available at: <http://www.research.microsoft.com/users/jplatt/smo-book.pdf>
- 24 Weston, J., and Watkins, C.: 'Multi-class support vector machines'. Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, Egham, TW20 0EX, UK, 1998, available at: <http://citeseer.ist.psu.edu/article/weston98multiclass.html>
- 25 Platt, J., Cristianini, N., and Shawe-Taylor, J.: 'Large margin dags for multiclass classification' 'Advances in neural information processing systems' (MIT Press, Cambridge, MA, 2000, 12th edn.)

- 26 Smola, A., and Schölkopf, B.: 'A tutorial on support vector regression', NeuroCOLT2 Tech Report, NC2-TR-1998-030, 1998
- 27 Schölkopf, B., Smola, A., Williamson, R.C., and Bartlett, P.L.: 'New support vector algorithms', *Neural Comput.*, 2000, **12**, pp. 1207–1245
- 28 Campbell, C.: 'Kernel methods: a survey of current techniques', *Neuro Comput.*, 2002, **48**, pp. 63–84
- 29 Keerthi, S.S., and Lin, C.-J.: 'Asymptotic behaviors of support vector machines with Gaussian kernel', *Neural Comput.*, 2003, **15**, (7), pp. 1667–1689
- 30 Lin, H.-T., and Lin, C.-J.: 'A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods'. Technical report, Department of Computer Science and Information Engineering, National Taiwan University, 2003, available at: <http://www.csie.ntu.edu.tw/~cjlin/papers/tanh.pdf>
- 31 Buhmann, M.D., and Ablowitz, M.J.: 'Radial basis functions: theory and implementations' (Cambridge University, 2003)
- 32 Haykin, S.: 'Neural networks: a comprehensive foundation' (Prentice-Hall, Upper Saddle River, NJ, 1999, 2nd edn.)
- 33 Dommel, H.W.: 'Techniques for analyzing electromagnetic transients', *IEEE Comput. Appl. Power*, 1997, **10**, pp. 18–21
- 34 Dommel, H.W.: 'Digital computer solution of electromagnetic transient in single and multiphase networks', *IEEE Trans.*, 1969, **PAS-88**, pp. 388–398
- 35 Raychaudhuri, S., Stuart, J.M., and Altman, R.B.: 'Principal components analysis to summarize microarray experiments: application to sporulation time series', *Pac. Symp. Biocomput.*, 2000, pp. 455–66 available at: <http://citeseer.ist.psu.edu/raychaudhuri00principal.html>
- 36 [http://csnet.otago.ac.nz/cosc453/student\\_tutorials/principal\\_components.pdf](http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf)
- 37 Chang, C.-C., and Lin, C.-J.: 'LIBSVM: a library for support vector machines', 2001, available at: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- 38 <http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf>
- 39 Kim, K.H., and Park, J.K.: 'Application of hierarchical neural networks to fault diagnosis of power systems', *Int. J. Elect. Power Energy Syst.*, 1993, **15**, (2), pp. 65–70