EUROPEAN TRANSACTIONS ON ELECTRICAL POWER
*Euro. Trans. Electr. Power* (2009)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/etep.321

# High impedance fault detection in distribution feeders using extended kalman filter and support vector machine

S. R. Samantaray[1*,†] and P. K. Dash[2]

[1]*Department of Electrical Engineering, National Institute of Technology, Rourkela, Rourkela-769008, Orissa, India*
[2]*Center for Research in Electrical, Electronics and Computer Engineering, D-75, Maitrivihar, Bhubanesawr-751023, Orissa, India*

SUMMARY

The paper presents an intelligent technique for high impedance fault (HIF) detection using combined extended kalman filter (EKF) and support vector machine (SVM). The proposed approach uses magnitude and phase change of fundamental, 3rd, 5th, 7th, 11th and 13th harmonic component as feature inputs to the SVM. The Gaussian kernel based SVM is trained with input sets each consists of '12' features with corresponding target vector '1' for HIF detection and '−1' for non-HIF condition. The magnitude and phase change are estimated using EKF. The proposed approach is trained with 300 data sets and tested for 200 data sets including wide variations in operating conditions and provides excellent results in noisy environment. Thus, the proposed method is found to be fast, accurate, and robust for HIF detection in distribution feeders.


KEY WORDS: High impedance fault detection; Support vector machine; Distribution feeder; Extended kalman filter

## 1. INTRODUCTION

Faults on power distribution feeders are difficult to detect [1,2] using conventional over current, ground fault relays and some versions of distance relaying schemes. Diversity, uncertainties, selectivity, suitability and operational constraints introduce malfunction, limitations and detection errors in case of high impedance faults (HIF). This is notable when remote source loading, fault resistance non-linearity, capacitive line currents, mutual coupling, and back-feed effects are taken into consideration. HIF faults [3,4] are usually characterized by the ripple rich current harmonic content due to non-linearity and thus are abnormal events that frequently occur in distribution feeders. There are two types of HIFs: the active faults and the passive ones. Active faults are followed by electric arc and present currents below the threshold of the protection relays. Normally, these currents decay with time until the complete extinction of the arc [5]. The majority of the techniques used to detect active HIFs make use of signals generated by the electric arc (harmonic and non-harmonic components) [6–9]. However, the arc may vanish even before the detection system gathers enough information to confirm the fault. Passive faults do not present an electric arc. They are more hazardous to people since there is no indication of the energization condition of the conductor. Due to presence of low or no current in HIF, the conventional over-current protection system normally fails to detect the same. Thus, it is a challenging issue to detect the HIF and isolate the feeder.

A.-R. Sedighi *et. al*. [10] presented a combined wavelet transform and soft computing application to HIF classification. This work includes feature extraction using wavelet transform and then classification using soft computing methods. HIF detection using neo-fuzzy systems [11], uses an artificial neuron set, composed of 'neo-fuzzy' neurons, and is trained to recognize the standard responses. In another work, earth faults with high impedance earthing in electrical distribution networks are characterized [12]. In the occurrence of disturbances, the traces of phase currents, voltages, neutral currents, and voltages were recorded at two feeders at two substations. The study dealt with the clearing of earth faults, relation between short circuits and earth faults, arc extinction,

*Correspondence to: S. R. Samantaray, Department of Electrical Engineering, National Institute of Technology, Rourkela, Rourkela-769008, Orissa, India.
†E-mail: sbh_samant@yahoo.co.in

arcing fault characteristics, appearance of transients, and magnitudes of fault resistances. The above works finds limitations as wavelet transform is highly prone to noise and provides erroneous results even with noise of SNR 30 [13] dB. The fuzzy-neural networks are sensitive to system frequency-changes, and require large training sets and training time. Also the HIF detection technique [14] based on first fourier transform (FFT) and decision tree suffers when subjected noisy environment and subsequently in real time applications.

This paper presents a new approach for HIF detection using support vector machine (SVM). An SVM [15–19] is a relatively new machine learning method that optimizes model on training data by solving a quadratic program (QP). In essence, an SVM finds the maximal separating hyperplane in feature space. It is computationally efficient because the transformation to feature space need not be done explicitly because dot products in feature space can be represented by kernel functions. The SVM-based classification is a modern machine learning method that is rarely used in fault classification even if it has given superior results in various classification and pattern recognition problems such as in text categorization [20] or phoneme recognition [21]. Currently, there exist only a few publications that concentrate on developing fault diagnostic methods based on SVM techniques [22–25].

SVM has advantages over traditional approaches such as neural networks for the following reasons.

1. Good generalization performance – once it is presented with a training set, it is able to learn a rule, which can correctly classify a new object quite often.
2. Computational efficiency – it is efficient in terms of speed and complexity.
3. Robust in high dimensions – in general, dealing with high-dimensional data is difficult for a learning algorithm because of over-fitting. One of the major reasons for attracting much attention is that SVMs are more robust to this over-fitting than other algorithms.

Apart from above advantages over conventional techniques for classification, SVM is highly suitable for classifying overlapping data sets close to each other by constructing a suitable optimal hyperplane. Thus, in the proposed study SVM is selected for classifying HIF signals from non-HIF signals.

This paper proposes HIF detection using SVM based on statistical learning theory. The SVM uses amplitude and phase change of fundamental, 3rd, 5th, 7th, 11th, and 13th harmonic component of the HIF current (as High impedance faults generate high frequency [3] signals superimposed on the fundamental 50 Hz frequency), extracted using extended kalman filter (EKF), as feature inputs to result '1' for HIF detection and '−1' for non-HIF condition. The parameters of the SVM kernel function are selected after five-fold cross validation process. The SVM is tested for HIF detection with wide variations in operating conditions including load switching, capacitor switching, and transformer inrush currents.

## 2. EXTENDED KALMAN FILTER (EKF) FOR HARMONIC ESTIMATION

The EKF is a nonlinear time domain stochastic estimator that provides an efficient estimation of the harmonic components of fault currents during a high impedance fault, characterized by the ripple rich current harmonic content due to non-linearity. Due to the nature of the filter, the Kalman gain is independent to the measurements, thus as the filter approaches steady state, it becomes less sensitive to parameter variations and begins to lose its ability of tracking time varying parameters. For optimum filtering results the *a priori* knowledge of the process noise covariance matrix **Q** and measurement noise covariance matrix **R** are required. However, in actual practice these matrices are unknown and hence by trial and error these matrices are chosen for optimal tracking results. Besides during dynamic changes that occur during a fault, both **Q** and **R** are updated by formulas given in the following section.

The advantage of EKF is that it is a recursive means to estimate the state of a process, in a way that minimizes the mean of the squared error. The filter is very powerful in several aspects and supports estimations of past, present, and even future states, and it can do so even when the precise nature of the modeled system is unknown. The FFT method, which is non-recursive technique, cannot handle signals with partial disturbances (noise), nor can these methods be applied to nonuniformly sampled signals.

Let the discrete signal which contains fundamental and harmonics along with a decaying DC component is represented by the model (such components are generated during a high impedance fault) given below

$$Z_k = A_1 \sin(kwT_s + \varphi) + A_2 \sin(3kwT_s + \varphi) + A_3 \sin(5kwT_s + \varphi) + A_4 \sin(7kwT_s + \varphi) + A_5 \sin(11kwT_s + \varphi) + A_6 \sin(13kwT_s + \varphi_{13}) + A_0 e^{-\alpha kT_s} \quad (1)$$



DOI: 10.1002/etep

The discrete signal can be represented in state space as

$$\mathbf{x}_{k+1} = \mathbf{F}_k \, \mathbf{x}_k \tag{2}$$

Where

$$\begin{aligned}
\mathbf{x}_k(1) &= A_1 \cos\phi, \mathbf{x}_k(2) = A_1 \sin\phi, \\
\mathbf{x}_k(3) &= A_2 \cos\phi, \mathbf{x}_k(4) = A_2 \sin\phi, \\
\mathbf{x}_k(5) &= A_3 \cos\phi, \mathbf{x}_k(6) = A_3 \sin\phi, \\
\mathbf{x}_k(7) &= A_4 \cos\phi, \mathbf{x}_k(8) = A_4 \sin\phi, \\
\mathbf{x}_k(9) &= A_5 \cos\phi, \mathbf{x}_k(10) = A_5 \sin\phi, \\
\mathbf{x}_k(11) &= A_6 \cos\phi, \mathbf{x}_k(12) = A_6 \sin\phi \\
\mathbf{x}_k(13) &= \mathrm{e}^{-T_S}, \mathbf{x}_k(14) = A_0 \mathrm{e}^{-kT_S}
\end{aligned} \tag{3}$$

and the state transition matrix is given by

$$\mathbf{F}_k = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & e^{-kT_s}
\end{bmatrix} \tag{4}$$

The observation matrix is given by

$$\begin{aligned}
\mathbf{G}_\mathrm{k} = [&\sin(kwT_\mathrm{s})\cos(kwT_\mathrm{s}), \\
&\sin(3kwT_\mathrm{s})\cos(3kwT_\mathrm{s}), \\
&\sin(5kwT_\mathrm{s})\cos(5kwT_\mathrm{s}), \\
&\sin(7kwT_\mathrm{s})\cos(7kwT_\mathrm{s}), \\
&\sin(11\mathrm{kw}T_\mathrm{s})\cos(11\mathrm{kw}T_\mathrm{s}), \\
&\sin(13kwT_\mathrm{s})\cos(13kwT_\mathrm{s}), 0, 1]
\end{aligned} \tag{5}$$

Literalizing the above system, the AEKF algorithm is obtained as follows

$$\widehat{\mathbf{x}}_{k/k} = \widehat{\mathbf{x}}_{k/k} + K_k(\mathbf{Z}_k - \mathbf{H}_k \mathbf{x}_{k/k-1}) \tag{6}$$

$$\mathbf{Z}_k = \mathbf{H}_k \, \mathbf{x}_k \tag{7}$$

$$\text{Where } \mathbf{H}_k = \left.\frac{\partial \mathbf{G}}{\mathrm{d}x}\right|_{k,k-1} = \begin{bmatrix}
x(1)kT_\mathrm{s}\cos(wkT_\mathrm{s}) - x(2)kT_\mathrm{s}\sin(wkT_\mathrm{s}) \\
x(3)kT_s\cos(3wkT_s) - x(4)kT_\mathrm{s}\sin(3wkT_\mathrm{s}) \\
x(5)kT_s\cos(5wkT_s) - x(6)kT_\mathrm{s}\sin(5wkT_s) \\
x(7)kT_s\cos(7wkT_\mathrm{s}) - x(8)kT_\mathrm{s}\sin(7wkT_s) \\
x(9)kT_s\cos(11wkT_\mathrm{s}) - x(10)kT_\mathrm{s}\sin(11wkT_s) \\
x(11)kT_\mathrm{s}\cos(13wkT_s) - x(12)kT_s\sin(13wkT_s) \\
0 \\
1
\end{bmatrix} \tag{8}$$




DOI: 10.1002/etep

The Kalman filter gain $K_k$ is obtained as

$$K_k = \widehat{P}_{K/K-1}\mathbf{H}_k^{\mathrm{T}}\left(\mathbf{H}_k\widehat{P}_{k/k-1}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1} \tag{9}$$

$$\widehat{P}_{k/k} = P_{k/k-1} - K_k\mathbf{H}_k\widehat{P}_{k/k-1} \tag{10}$$

$$\widehat{P}_{k+1/k} = P_{k/k} + \mathbf{Q} \tag{11}$$

where $\mathbf{Q}$ is the covariance matrix and $\mathbf{R}$ is the measurement noise covariance.

To improve the performance of the EKF, the measurement error covariance is updated in the following manner. The expression for $\mathbf{R}$ is obtained as the error between observed and estimated values of $x_k$ as

$$\mathbf{R} = \left(\mathbf{z}_k - \mathbf{H}_k\widehat{\mathbf{x}}_k\right)^{\mathrm{T}}\left(\mathbf{z}_k - \mathbf{H}_k\widehat{\mathbf{x}}_k\right) \tag{12}$$

The error covariance $R$ is recursively updated as

$$\mathbf{R}_k = \lambda_k\mathbf{R}_{k-1} + (1-\lambda_k)e_k^2 \tag{13}$$

where $\lambda_k$ is forgetting factor given by

$$\lambda_k = \frac{1}{1 + |\mathbf{R}(k)/\mathbf{R}_0|} \tag{14}$$

where $\mathbf{R}_0$ is the initial error covariance $\mathbf{R}$

Further, the model error covariance matrix $\mathbf{Q}$ is adapted by using a covariance function ce as

$$\mathrm{ce}_k = \lambda_q * \mathrm{ce}_{k-1} + (1-\lambda_q) * e_k * e_{k-1} \tag{15}$$

If $\mathrm{ce}(k) > \mathrm{ce}_{\mathrm{th}}$, $\mathbf{Q} = \mathbf{Q}_1$, and $\mathrm{ce}(k) < \mathrm{ce}_{\mathrm{th}}$, $\mathbf{Q} = \mathbf{Q}_0$
where $\mathbf{Q}_0$ is the model error covariance and $\mathbf{Q}_1$ is a new value of $\mathbf{Q}$ and $\mathbf{Q}_1 > \alpha\mathbf{Q}_0, \alpha > 1$, and $\mathrm{ce}_{\mathrm{th}}$ is the threshold value of error covariance.

Another possible adaptation $\mathbf{Q}_k$ is obtained as

$$\mathbf{Q}_k = (1-\alpha_q) * \mathbf{Q}_{k-1} + \alpha_q * P_{k-1} * P_{k-1}^{\mathrm{T}} * \mathrm{ce}_k \tag{16}$$

## 3. SUPPROT VECTOR MACHINE FOR CLASSIFICATION

The SVM is firmly grounded in the framework of statistical learning theory, which characterizes the properties of learning machines enabling them to generalize well to unseen data. In SVM, original input space is mapped into a high-dimensional dot product space called a feature space, and in the feature space the optimal hyperplane is determined to maximize the generalization ability of the classifier. SVMs have the potential to handle very large feature spaces, because training of SVM is carried out so that the dimension of classified vectors does not have as distinct an influence on the performance of SVM as it has on the performance of conventional classifiers. That is why it is noticed to be especially efficient in large classification problems. Also, SVM-based classifiers are claimed to have good generalization properties compared to conventional classifiers, because in training the SVM classifier, the so-called structural misclassification risk is to be minimized, whereas traditional classifiers are usually trained so that the empirical risk is minimized. SVM is compared to the RBF neural network in an industrial fault classification task [23], and it has been found to give better generalization.

Considering the n-dimensional input $x_i$ ($i = 1 \ldots \ldots M$, $M$ is the number of samples) belong to class I or II and associated labels be $y_i = 1$ for Class I and $y_i = -1$ for Class II, respectively. For linearly separable data, we can determine a hyperplane $f(x) = 0$ that



DOI: 10.1002/etep

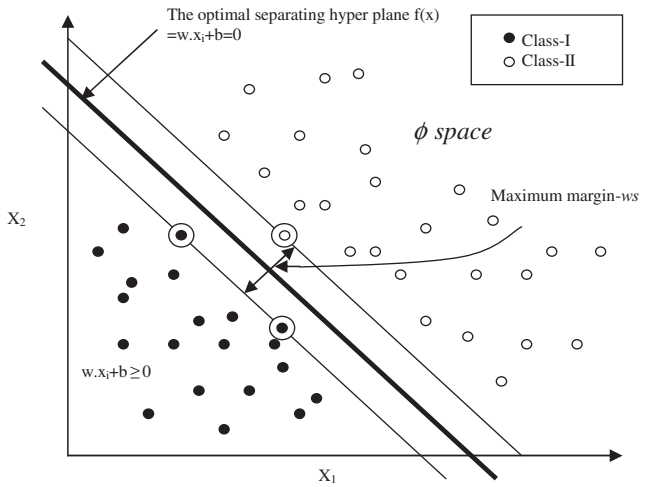


Figure 1. $f(x)$ as a separating hyperplane lying in a high-dimensional space. Support vectors are inside the circles.

separates the data

$$f(x) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + b = \sum_{k=1}^{\mathrm{n}} \mathbf{w}_k \mathbf{x}_k + b = 0 \tag{17}$$

where '$\mathbf{w}$' is an $n$-dimensional vector and '$b$' is a scalar. The vector '$\mathbf{w}$' and the scalar '$b$' determine the position of the separating hyperplane. Function sign ($f(x)$) is also called the decision function. A distinctly separating hyperplane satisfies the constraints $f(x_i) \geq 1$ if $y_i = +1$ and results in

$$y_i f(x_i) = y_i(\mathbf{w}^{\mathrm{T}} x_i + b) \geq 1 \quad \text{for } i = 1 \ldots\ldots M \tag{18}$$

The optimal separating hyperplane decides the maximum margin, the maximum distance between the plane and the nearest data. An example of the optimal separating hyperplane of two datasets is presented in Figure 1. From the geometry, the geometrical margin is found to be $||\mathbf{w}||^{-2}$. Taking into account the noise with slack variables $\xi_{\mathrm{i}}$ and error penalty $C$, the optimal hyperplane can be found by solving the following convex quadratic optimization problem,

$$\begin{aligned} &\text{minimize} \\ &\qquad \tfrac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{\mathrm{M}} \xi_i \\ &\text{subject to} \\ &\qquad y_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b) \geq 1 - \xi_i, \quad \text{for } i = 1.....M \\ &\qquad \xi_i \geq 0, \text{for all } i \end{aligned} \tag{19}$$

where $\xi_i$ is measuring the distance between the margin and the examples $x_i$ lying on the wrong side of the margin. The calculations can be simplified by converting the problem with Kuhn–Tucker conditions into the equivalent Lagrange dual problem, which will be

$$\begin{aligned} &\text{maximize} \\ &\qquad W(\alpha) = \sum_{i=1}^{\mathrm{M}} \alpha_i - \frac{1}{2}\sum_{i,k=0}^{\mathrm{M}} \alpha_i \alpha_k y_i y_k x_i^{\mathrm{T}} x_k \\ &\text{subject to} \\ &\qquad \sum_{i=1}^{\mathrm{M}} y_i \alpha_i = 0, C \geq \alpha_i \geq 0, i = 1...., M \end{aligned} \tag{20}$$

The number of variables of the dual problem is the number of training data. Let us denote the optimal solution of the dual problem with $\alpha^*$ and $\mathbf{w}^*$. According to the Karush–Kuhn–Tucker theorem, the inequality condition in (18) holds for the training input–output



DOI: 10.1002/etep

(feature and label) pair $x_i$, $y_i$ only if the associated $\alpha^*$ is not 0. In this case, the training example $x_i$ is a support vector (SV). Usually, the number of SVs is considerably lower than the number of training samples making SVM computationally very efficient. The value of the optimal bias $b^*$ is found from the geometry

$$b^* = -\frac{1}{2}\sum_{\text{SVs}} y_i \alpha_i^* \left(s_1^* x_i + s_1^{\text{T}} x_i\right) \tag{21}$$

where $\mathbf{s}_1$ and $\mathbf{s}_2$ are arbitrary support vectors (SVs) for class I and II, respectively. Only the samples associated with the SVs are summed, because the other elements of optimal Lagrange multiplier $\alpha^*$ are equal to zero.

The final decision function will be given by

$$f(x) = \sum_{\text{SVs}} \alpha_i y_i x_i^{\text{T}} x + b^* \tag{22}$$

Then unknown data example '$\mathbf{x}$' is classified as follows:

$$\mathbf{x} \in \begin{cases} \text{Class} - \text{I} & \text{if } f(x) \geq 0 \\ \text{Class} - \text{II} & \text{Otherwise} \end{cases} \tag{23}$$

SVM can also be used in nonlinear classification tasks with application of kernel functions. The data to be classified is mapped onto a high-dimensional feature space, where the linear classification is possible. Using a nonlinear vector function $\phi(x) = (\phi_1(x) \cdots \phi_{\text{m}}(x)), \text{m} >> \text{n}$ to map the 'n'-dimensional input vector '$\mathbf{x}$' into the 'm' dimensional feature space, the linear decision function in dual form is given by

$$f(x) = \sum_{\text{SVs}} \alpha_i y_i \phi^{\text{T}}(x_i)\phi(x) \tag{24}$$

Working in the high-dimensional feature space enables the expression of complex functions, but it also generates problems. Computational problems occur due to the large vectors and the danger of overfitting also exists due to the high dimensionality. The latter problem is solved with application of the maximal margin classifier, and so-called kernels give solution to the first problem. Notice that in (24) as well as in the optimization problem (19), the data occur only in inner products. A function that returns a dot product of the feature space mappings of original data points is called a kernel, $K(x, z) = \phi^{\text{T}}(x)\ \phi(z)$. Applying a kernel function, the learning in the feature space does not require explicit evaluation of $\phi$. Using a kernel function, the decision function will be

$$f(x) = \sum_{\text{SVs}} \alpha_i{}^* y_i K(x_i\, x) \tag{25}$$

and the unknown data example is classified as before. The values of $K(x_i, x_j)$ over all training samples $i, j = 1 \ldots M$ from the kernel matrix, which is a central structure in the kernel theory. Mercer's theorem [26] states that any symmetric positive-definite matrix can be regarded as a kernel matrix. The radial basis function (RBF) machines have the inner product kernel which maps the data from the original input space into a potentially higher dimensional feature space where linear methods may then be used. The RBF kernel which returns the dot product of two arguments '$x$' and '$z$' is given as

$$K(x, z) = \exp\left\{-\frac{|x - z|^2}{2\sigma^2}\right\} \tag{26}$$

Where '$\sigma$' is the width of the Gaussian function.

## 4. SYSTEM STUDIED

The system studied in the proposed research is shown in Figure 2a. The feeders are from 138/25 kV substation transformer which is connected from a 138 kV transmission line of 100 km line length. The loads (lagging pf of 0.8) and shunt capacitors are also connected as shown in the figure. The HIF faults are created on the distribution feeder as shown in the figure. The HIF model is



DOI: 10.1002/etep

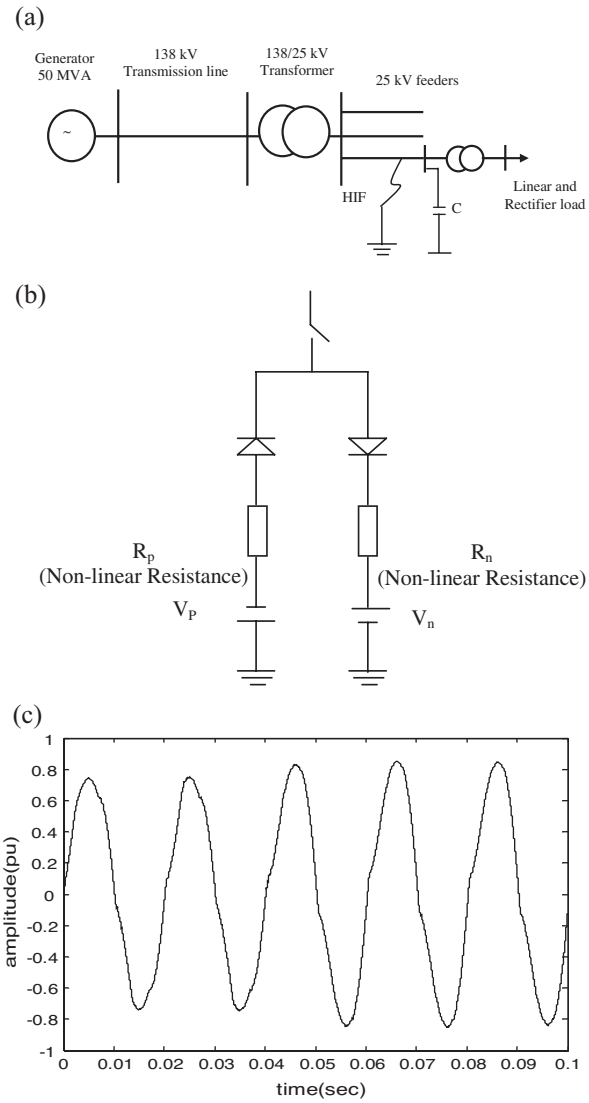


Figure 2. a) System studied. b) High-impedance fault model. c) Typical HIF current.

developed using anti-parallel diodes with non-linear resistance and DC source connected together for each phase as shown in Figure 2b. Both linear and rectifier loads are connected for load switching purpose. The simulation models are developed using Power System Blokset (SIMULINK) and the sampling rate chosen is 960 Hz on a 60 Hz base frequency. The typical HIF fault current is shown in Figure 2c.

The different simulation conditions taken into considerations are as follows

- Three-phase load change from 20–60%, 30–70%, 60–110%, 20–110% in forward and reverse way.
- One-phase load change 30–70%, 20–50%, 40–80%, 40–110%, 20–100% in forward and reverse way.
- Transformer enrgization at different timings in the cycle slot (16 instances in one cycle).
- The above changes are made with change in infinite source phase angle of 0 to 120°, with a span of 10°.
- Shunt capacitors are switched on and off.
- The above changes are made under varying conditions in the HIF model by varying the DC source voltages by −25–25% in a step of 5%. The central voltage also varies from 2000 to 10 000 V with a step of 1000 V.



DOI: 10.1002/etep

From the above operating conditions 500 test cases are simulated, out of which 300 cases are used for training the SVM and rest 200 cases are used for testing purpose.

The proposed technique is also tested with a standard mesh type distribution network as shown in Figure 4, supplied from two separate three-phase sources through transmission line and transformers. The transmission lines are 138 kV and the transformers are 50 MVA supplying at138/25 kV to the distribution network. The distribution feeders (pi sections of 20 km each) work at 25 kV and connected with shunt capacitors, linear loads, and 2 MVA 6-pulse rectifier load (non-linear load). The resistance, inductance and capacitance of positive and zero sequence of transmission lines are $R_1 = 0.01273$ ohm/km; $X_1 = 0.9337$ mH/km; $C_1 = 0.0012$ $\mu$F/km, and $R_0 = 0.3864$ ohm/km; $X_0 = 4.1264$ mH/km; $C_0 = 0.0075$ $\mu$F/km, respectively. The resistance, inductance and capacitance of distribution lines (pi-section) are $R_1 = 0.2568$ ohm/km; $X_1 = 2.0$ mH/km; $C_1 = 0.0086$, respectively. The total impedance percentage of the transformers is 6.75% and the frequency of the system is 60 Hz. Total number of HIF and non-HIF conditions are simulated on distribution feeders are 500, out of which 300 are used for training and 200 for testing purpose (100 for HIF and 100 for non-HIF).

## 5. PROPOSED HIF DETECTION SCHEME

### *5.1. Harmonic component extraction using EKF*

The proposed scheme is shown as in Figure 3a. The HIF current is processed through EKF and required components are extracted. In this case magnitude and phase change of fundamental, 3rd, 5th, 7th, 11th, and 13th harmonic component are estimated using EKF. In the proposed study, the **Q** (Covariance matrix) and **R** (Measurement noise covariance) are selected as 0.001 and 0.05 respectively for estimating harmonic components. As seen in Figure 3b, the harmonics are estimated within almost one cycle (16 samples). Thus, the estimated values of magnitude and phase change at the end of one cycle (17th sample) are considered as inputs to the SVM. The estimated harmonic components are in per unit (pu) system and shown with respect to samples. Thus, there are 12 features selected to be used as input features against one target output for one case. There are 500 cases simulated under various operating conditions of the distribution network. Out of which 300 cases are used to train and 200 cases are used to test the designed SVM. In the training set, 100 HIF conditions and 200 are non-HIF conditions are taken in to consideration. Like wise in the testing data set, 100 cases are of HIF conditions and 100 cases are of non-HIF conditions are considered.

### *5.2. HIF detection using SVM*

In the proposed study, Gaussian kernel based SVM is used for designing the HIF detector. The bound on the Lagrangian multipliers '*C*' is selected 5.0 and the conditioning parameter for QP method, lambda is chosen as $1.0 * \text{e-}5$. The value of width of the Gaussian function '$\sigma$' is selected as 1.5. All the above parameters are selected after cross validation process [27,28]. In this study five-fold cross validation technique is used to get the final parameter selection. In the training 300 data sets are divided into 5 sub sets each having 60 data sets for cross validation process. Thus, it is 5-fold cross validation process used in the proposed study for parameter selection.

The results from the proposed approach for radial distribution feeder is depicted in Table I. In testing, 200 cases are considered, out of which 100 are HIF and rest 100 are non-HIF conditions. It is found that the proposed SVM is able to classify 98% in case of HIF testing and 100% for non-HIF testing cases. Also the SVM is tested in noisy conditions with SNR up to 20 dB. In case of noisy data, the classification rates are 96 and 99% for HIF and non-HIF testing conditions, respectively. The results for HIF detection for the mesh distribution network are depicted on Table III. The HIF detection rate is 96% and for non-HIF is 100. For noisy conditions with SNR 20 dB, the classification rates for HIF and non-HIF are 94 and 98%, respectively. Thus, the proposed scheme is able to detect HIF for a standard mesh distribution network with possible HIF and non-HIF conditions.

Similar comparisons are made with other existing techniques such as FFT with Decision tree (FFT + DT) and Wavelet Transform with Radial Basis Function Neural Network (RBFNN) (WAVELET + RBFNN) as shown in Table II. While comparing the results with the above techniques it is found that the earlier approaches suffer in case of data sets with and without noise. In all cases 100 HIF cases are considered for testing purposes. It is found that EKF + SVM provide 98% classification accuracy, while FFT + DT and WAVELET + RBFNN provide 92% and 89% classification accuracy, respectively. In case of noisy environment, the



DOI: 10.1002/etep

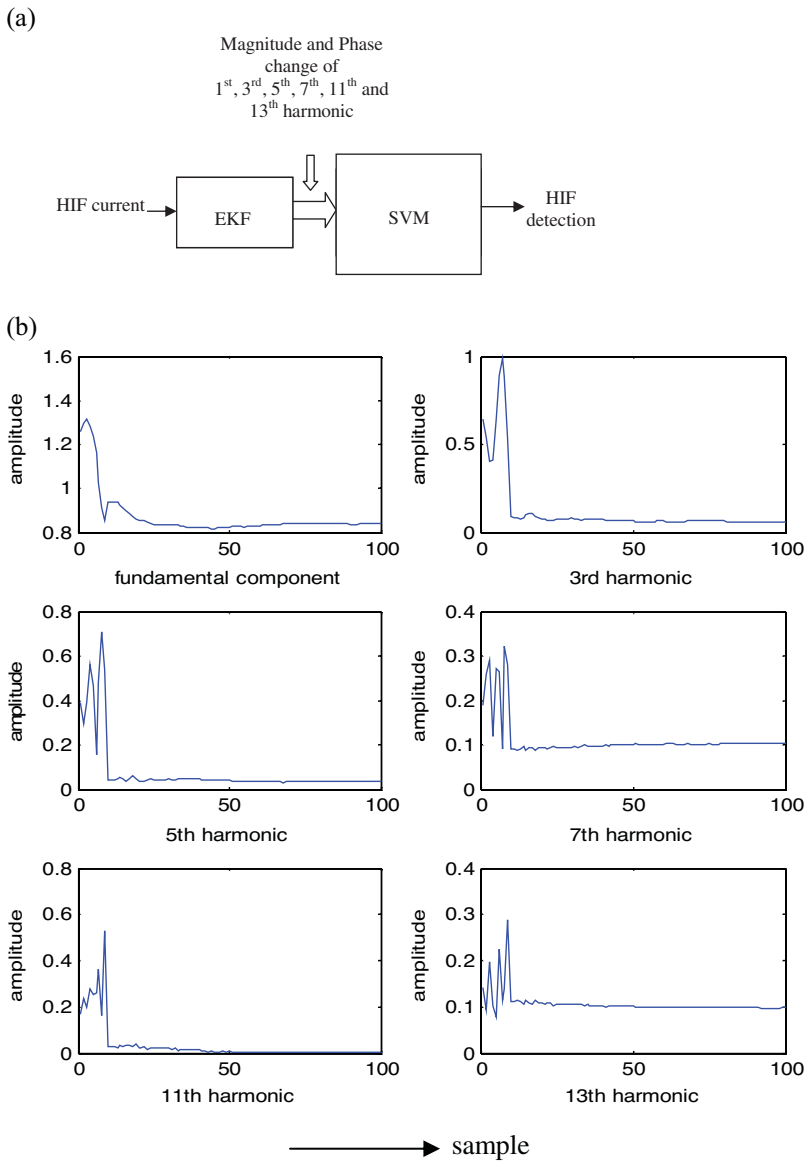


Figure 3. a) Proposed HIF detection scheme. b) Harmonic components from Extended Kalman Filter.

classification accuracy suffers to a larger extent for other two approaches. For testing data with SNR up to 20 dB, the proposed EKF + SVM provides 96% classification accuracy, while FFT + DT and WAVELET + RBFNN provide 87% and 82% classification accuracy, respectively. Thus, it is observed that the performance of other approaches is degraded heavily in presence of noise.

In another comparison, performance of FFT + SVM and EKF + RBFNN is also found out, and compared with FFT + DT and EKF + SVM, respectively. While comparing, it is found that, EKF + RBFNN provides HIF detection rate of 95% compared to 98% by EKF + SVM. Similarly, FFT + SVM provide HIF detection rate of 93% compared to 92% by FFT + DT. Similar observations are made for noisy environment. The comparison results of HIF detection for mesh distribution network are depicted in Table IV. It is found from the above observation that EKF and SVM provide better results in feature extraction and classification compared to existing conventional techniques.



DOI: 10.1002/etep

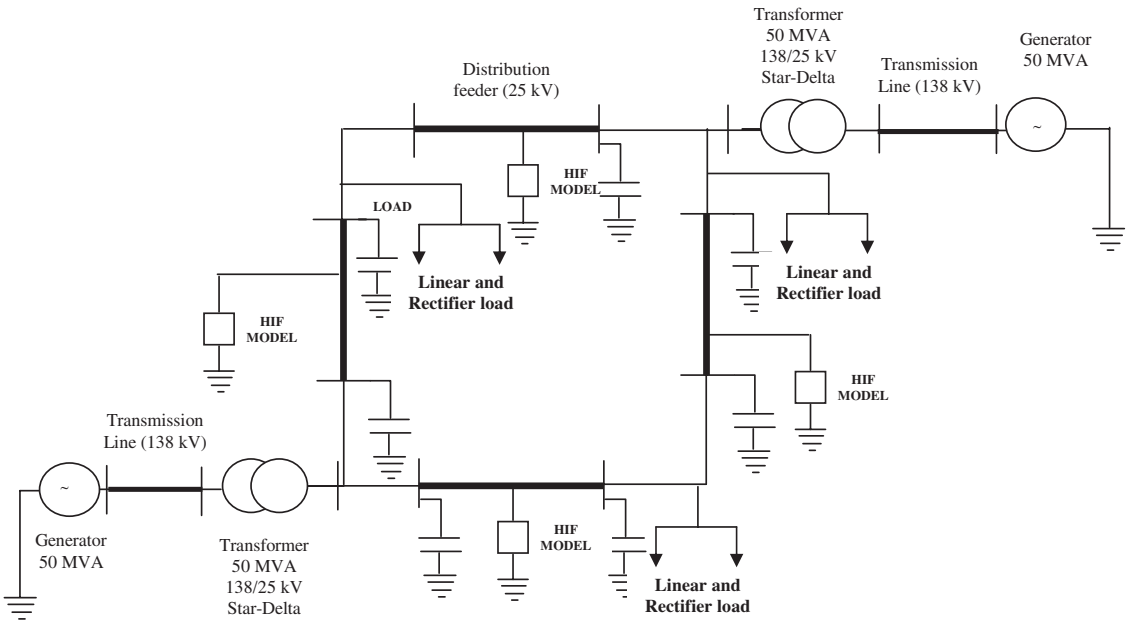


Figure 4. 25 kV meshed power distribution network.

Table I. Testing Results for radial distribution network.

| Actual class | No. of cases | No. of cases for HIF detection Class-I (1) | No. of cases for non-HIF detection Class-II (−1) | No. of misclassification | Classification rate % |
|---|---|---|---|---|---|
| | | | Data from simulation without noise | | |
| I | 100 | 98 | 2 | 2 | 98 |
| II | 100 | 0 | 100 | 0 | 100 |
| | | | Data from simulation with SNR 20 dB | | |
| I | 100 | 96 | 4 | 4 | 96 |
| II | 100 | 1 | 99 | 1 | 99 |

Table II. Performance comparison with the existing methods.

| Methods | No. of cases | No. of cases for HIF detection | No. of misclassification | Classification rate % |
|---|---|---|---|---|
| | | Data from simulation without noise | | |
| EKF + SVM | 100 | 98 | 2 | 98 |
| EKF + RBFNN | 100 | 95 | 5 | 95 |
| FFT + DT | 100 | 92 | 8 | 92 |
| FFT + SVM | 100 | 93 | 7 | 93 |
| WAVELET + RBFNN | 100 | 89 | 11 | 89 |
| | | Data from simulation with SNR 20 dB | | |
| EKF + SVM | 100 | 96 | 4 | 96 |
| EKF + RBFNN | 100 | 93 | 7 | 93 |
| FFT + DT | 100 | 87 | 13 | 87 |
| FFT + SVM | 100 | 89 | 11 | 89 |
| WAVELET + RBFNN | 100 | 82 | 18 | 82 |



DOI: 10.1002/etep

Table III. Testing Results for meshed distribution network.

| Actual class | No. of cases | No. of cases for HIF detection Class-I (1) | No of cases for non-HIF detection Class-II (−1) | No. of misclassification | Classification rate % |
|---|---|---|---|---|---|
| | | | Data from simulation without noise | | |
| I | 100 | 96 | 4 | 4 | 96 |
| II | 100 | 0 | 100 | 0 | 100 |
| | | | Data from simulation with SNR 20 dB | | |
| I | 100 | 94 | 6 | 6 | 94 |
| II | 100 | 2 | 98 | 2 | 98 |

Table IV. Performance comparison with the existing methods for meshed distribution network.

| Methods | No. of cases | No. of cases for HIF detection | No. of misclassification | Classification rate % |
|---|---|---|---|---|
| | | Data from simulation without noise | | |
| EKF + SVM | 100 | 96 | 4 | 96 |
| EKF + RBFNN | 100 | 94 | 6 | 94 |
| FFT + DT | 100 | 90 | 10 | 90 |
| FFT + SVM | 100 | 91 | 9 | 91 |
| WAVELET + RBFNN | 100 | 87 | 13 | 87 |
| | | Data from simulation with SNR 20 dB | | |
| EKF + SVM | 100 | 94 | 6 | 94 |
| EKF + RBFNN | 100 | 92 | 8 | 92 |
| FFT + DT | 100 | 86 | 14 | 86 |
| FFT + SVM | 100 | 87 | 13 | 87 |
| WAVELET + RBFNN | 100 | 80 | 20 | 80 |

## 6. CONCLUSIONS

The proposed study provides a new technique for HIF detection using EKF and Support Vector Machine. The magnitude and phase change are extracted using EKF, which are used as input features to the designed SVM. The results obtained from the proposed technique were compared with the existing techniques and provides excellent results under noisy conditions also. The SVM based technique has been tested under wide variations in network operating conditions including load change, in rush and switching conditions and found to be accurate and robust for HIF detection, and thus can be extended for HIF detection in large power distribution network.

## 7. LIST OF SYMBOLS

| | |
|---|---|
| F(x) | Optimal separating hyperplane |
| $\xi_i$ | Slack variable |
| $C$ | Error penalty |
| $b^*$ | Optimal bias |
| $K(x, z)$ | Kernel function |
| $\mathbf{F}_k$ | State transition matrix |
| $\mathbf{G}_k$ | Observation matrix |
| $K_k$ | Kalman filter gain |
| $\mathbf{Q}$ | Covariance matrix |
| $\mathbf{R}$ | Measurement noise covariance |
| $\lambda_k$ | Forgetting factor |



DOI: 10.1002/etep

## REFERENCES


1. Johns AT. Correspondence on microprocessor based algorithm for high-resistance earth-fault distance protection. *Proceedings of Institute of Electrical Engineering*, Part C 1985; **132**:94–95.
2. Yang Q, Morrison I. Microprocessor-based algorithm for high-resistance earth-fault distance protection. *Proceedings of Institute of Electrical Engineering*, Part C 1983; **130**:306–310.
3. Aucoin BM, Russell BD. Distribution high impedance fault detection using high frequency current components. *IEEE Transactions Power Applied Systems* 1982; **101**(6):1596–1606.
4. Sharaf AM, E1-Sharkawy RM, Talaat HEA, Badr MAL. Novel alpha-transform distance relaying scheme. *Proc. IEEE-CCECE Conf*., Calgary, Canada 1996.
5. Aucoin M, Russell BD. Detection of distribution high impedance faults using burst noise signals near 60 Hz. *IEEE Transaction on Power Delivery* 1987; **2**(2):342–348.
6. Sultan AF, Swift GW, Fedirchuk DJ. Detection of high impedance arcing faults using a multilayer perceptron. *IEEE Transaction on Power Delivery* 1992; **7**(4):1871–1877.
7. Sultan AF, Swift GW, Fedirchuk DJ. Detection arcing downed wires using fault current flicker and half cycle asymmetry. *IEEE Transaction on Power Delivery* 1994; **9**(1):461–470.
8. Russell BD, Chinchali RP. Signal processing algorithm for detecting arcing faults on power distribution feeders. *IEEE Transactions on Power Delivery* 1989; **4**(1):132–140.
9. Russell BD, Chinchali RP, Kim CJ. Behaviour of low frequency spectra during arcing fault and switching events. *IEEE Transaction on Power Delivery* 1988; **3**(4):1485–1492.
10. Sedighi A-R, Haghifam M-R, Malik OP. Soft computing applications in high impedance fault detection in distribution systems. *Electric Power System Research* 2006; **76**(1–3):136–144.
11. Jota PRS, Jota FG. Fuzzy detection of high impedance faults in radial distribution feeders. *Electric Power Systems Research* 1999; **49**:169–174.
12. Hanninen S, Lehtonen M. Characteristics of earth faults in electrical distribution networks with high impedance earthing. *Electric Power Systems Research* 1998; **44**:155–161.
13. Yang H-T, Liao C-C. A de-noising scheme for enhancing Wavelet-based power quality monitoring system. *IEEE Transaction on Power Delivery* 2001; **16**(3):353–360.
14. Sheng Y, Rovnyak M. Decision tree based methodology for high impedance fault detection. *IEEE Transaction on Power Delivery* 2004; **19**(2):533–536.
15. Chapelle O, Vapnik VN, Bousquet O, Mukherjee S. Choosing multiple parameters for support vector machines. *Machine Learning.* 2002; **46**(1–3):131–159.
16. Burges CJC. A tutorial on support vector machines for pattern recognition. *Data Mining Knowledge Discovery* 1998; **2**(2):121–167.
17. Burges CJC. Geometry and invariance in kernel based methods. In *Advances in Kernel Methods—Support Vector Learning*, Schoelkopf B, Burges CJC, Smola AJ (eds). The MIT Press: Cam-Cambridge, MA, 1999; pp. 89–116.
18. Chapelle O, Vapnik VN. Model selection for support vector machines. In *Advances in Neural Information Processing Systems*, vol. 12, Solla S, Leen TK, Muller K-R (eds). The MIT Press: Cambridge, MA, 2000; pp. 230–236.
19. Collobert R, Bangio S. SVMTorch: a support vector machine for large-scale regression and classifcation problems. *Journal of Machine Learning Research* 2001; **1**:143–160.
20. Joachims T. Text categorization with support vector machines: Learning with many relevant features. Technical report, University of Dortmund, 1997.
21. Salomon J. Support vector machines for phoneme classification. M.Sc Thesis, University of Edinburgh, 2001.
22. Salat R, Osowski S. Accurate fault location in the power transmission line using support vector machine approach. *IEEE Transaction on Power Delivery* 2004; **19**(2):979–986.
23. Thukaram D, Khincha HP, Vijaynarasimha HP. Artificial neural network and support vector machine approach for locating faults in radial distribution systems. *IEEE Transaction on Power Delivery.* 2005; **20**(2):710–721.
24. Parikh UB, Das B, Maheshwari RP. Combined Wavelet-SVM Technique for Fault Zone Detection in a Series Compensated Transmission Line. *IEEE Trans. On Power Delivery* (To appear—October 2008).
25. Ribeiro B. Support vector machines and RBF neural networks for fault detection and diagnosis. *Proceedings of the 8th international conference on neural information processing*, paper 191. 2001.
26. Mercer J. Functions of positive and negative type and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society, London* 1909; **A 209**:415–446.
27. Kearns M, Ron D. Algorithmic stability and sanity-check bounds for leave-one-out cross validation. *Proc. 10th Conf. Comput. Learning theory*, New York: ACM, pp. 152–162 1997.
28. Cawley GC, Talbot NLC. Fast exact leave-one-out cross-validation of sparse least-squares support vector machines. *Neural Networks, Elsevier* 2004; **17**:1467–1475.




DOI: 10.1002/etep