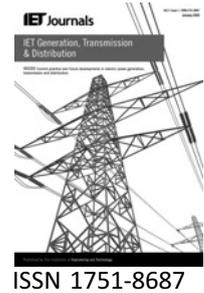


Published in IET Generation, Transmission & Distribution
 Received on 3rd July 2008
 Revised on 28th December 2008
 doi: 10.1049/iet-gtd.2008.0316



Decision tree-based fault zone identification and fault classification in flexible AC transmissions-based transmission line

S.R. Samantaray

*Department of Electrical Engineering, National Institute of Technology Rourkela, India
 E-mail: sbh_samant@yahoo.co.in*

Abstract: Transmission line distance relaying for flexible AC transmission lines (FACTS) including thyristor controlled series compensator (TCSC), STATCOM, SVC and unified power flow controller (UPFC) has been a very challenging task. A new approach for fault zone identification and fault classification for TCSC and UPFC line using decision tree (DT) is presented. One cycle post fault current and voltage samples from the fault inception are used as input vectors against target output '1' for fault after TCSC/UPFC and '0' for fault before TCSC/UPFC for fault zone identification. Similarly, the DT-based classification algorithm takes one cycle data from fault inception of three phase currents along with zero-sequence current and voltage, and constructs the optimal DT for classifying all ten types of shunt faults in the transmission line fault process. The algorithm is tested on simulated fault data with wide variations in operating parameters of the power system network including noisy environment. The results indicate that the proposed method can reliably identify the fault zone and classify faults in the FACTS-based transmission line in large power network.

1 Introduction

In the current open-access environment, transmission systems are being required to provide increased bulk power transfer capability and to accommodate a much wider range of possible generation patterns. This had led to an increased focus on transmission constraints and on the means by which much constraint can be alleviated. The flexible AC transmission lines (FACTS) [1] devices offer a versatile alternative to conventional reinforcement methods. The thyristor controlled series compensator (TCSC) [2] and the unified power flow controller (UPFC) [3] are important FACTS devices, which are used extensively for improving the utilisation of the existing transmission system. TCSC-based compensation possess thyristor controlled variable capacitor protected by metal oxide varistor (MOV) and an air-gap. However, the implementation of this technology changes the apparent line impedance, which is controlled by the firing angle of thyristors, and is accentuated by other factors including the MOV. The presence of the TCSC in fault loop not only affects the steady-state components but also the

transient components. The controllable reactance, the MOVs protecting the capacitors and the air-gaps operation make the protection decision more complex and therefore conventional relaying scheme based on fixed settings has its limitation.

The UPFC offers new horizons in terms of power system control, with the potential to independently control up to three power system parameters, for instance bus voltage, line active power and line reactive power. While the use of the UPFC improves the power transfer capability and stability of a power system, certain other problems emerge in the field of power system protection, in particular the transmission line protection [4–6], affecting greatly the reach of the distance relay.

Further, the current level may be of the same order at two different points of the transmission line (before and after the TCSC/UPFC) for the similar type of fault. Thus, fault classification and fault zone identification are very challenging tasks for a transmission line with the TCSC/UPFC. If the fault does not include the FACTS device, then the impedance calculation is like ordinary transmission line, and when the

fault includes the FACTS, then the impedance calculation accounts for the impedances introduced by the FACTS device. The line impedance is compared with the protective zone and if the line impedance is less than the relay setting, then the relay gives a signal to trip the circuit breaker. Thus, before impedance calculation, a more reliable and accurate fault zone identification approach is necessary, for safe and reliable operation of the distance relay. Thus the correct fault zone identification and fault classification in presence of these FACTS devices such as the TCSC and the UPFC, in the transmission line are critical tasks to be dealt with.

Different attempts have been made for fault zone identification and fault classification using wavelet transform, Kalman filtering approach and neural network [1, 7, 8]. The Kalman filtering approach finds its limitation, as fault resistance cannot be modelled and further it requires a number of different filters to accomplish the task. Back propagation neural network, radial basis function neural network, fuzzy neural network are employed for adaptive protection of such a line where the protection philosophy is viewed as a pattern classification problem [7, 8]. The networks generate the trip or block signals using a data window of voltages and currents at the relaying point. However, the above approaches are sensitive to system frequency changes, and require large training sets and training time and a large number of neurons.

Recently, wavelet transform combined with support vector machine (SVM) [9] has been applied for fault zone selection for TCSC-based lines. The above works find limitations as wavelet transform is highly prone to noise and provides erroneous results even with noise of signal-to-noise ratio (SNR) 30 dB [10]. Also the computational burden of SVM is higher compared to proposed decision tree (DT), which puts constraints for online realisation of SVM based relay for distance relaying.

Recently, DT has been found highly successful in applications such as online dynamic security assessment [11], transient stability [12] and islanding detection [13]. DT being rule-based is more transparent and human friendly compared to black-box solution such as neural network and other pattern classifiers. The proposed method is based on DT [14–18] for fault zone identification and fault classification for the TCSC/UPFC-based line. The one cycle post fault current and voltage samples are used as input to the DT against target outputs '0' for faults before the TCSC/UPFC and '1' for faults after the TCSC/UPFC. Similarly, the DT-based fault classification tree is built up for classifying ten types of shunt faults including the TCSC and the UPFC in the transmission line. The proposed technique is tested on wide variations in operating parameters including the noisy environment and found to be accurate and robust for classifying and identifying fault zone for the TCSC and the UPFC-based line. The following sections deal with the detailed system studied, methodology and testing results.

2 Studied system

2.1 TCSC-based line

A 230 kV, 50 Hz power system is illustrated in Fig. 1. In this system, a TCSC is located at mid-point of the transmission line (distributed model), used for the distance protection study. The power system consists of two sources, TCSC and associated components and a 300 km transmission line. The transmission line has zero sequence impedance $Z_0 = 96.45 + j335.26 \Omega$ and positive sequence impedance $Z_1 = 9.78 + j110.23 \Omega$. Here $E_S = 230 \text{ kV}$ and $E_R = 230 \angle \delta \text{ kV}$. The TCSC is designed to provide compensation varying from minimum 30% to maximum 40%. All the components are modelled using the PSCAD (EMTDC) subroutines.

The sampling frequency is 1.0 kHz at 50 Hz base frequency. The MOV consists of a number of zinc oxide discs electrically connected in series and parallel. The purpose of the MOV is to prevent the voltage across the capacitor from rising to levels that will damage the capacitor. This is most likely to happen when a fault occurs at a point on the compensated line which minimises the impedance of the fault loop. When instantaneous voltage across the capacitor approaches a dangerous level, the MOV begins to draw a significant proportion of the line current thereby limiting the voltage across the capacitor at that level. This action alters the impedance in the series path and hence the fault-loop impedance. In the event that the MOV remains in conduction long enough to raise its temperature (energy) to a dangerous level, an air-gap is triggered to short out both the MOV and the capacitor, again changing the fault-loop impedance. The operation of the MOV can be within the first-half cycle of fault and depending on the severity of the fault, it may continue to operate until the air-gap is triggered cycles later. This is precisely the time when a digital relay makes protection decision. Further, a bypass switch in parallel with the gap automatically closes for abnormal system conditions that cause prolonged current flow through the gap. Fig. 2a shows a typical series capacitor arrangement for one phase of a transmission line and the typical voltage–current characteristic of an MOV is shown in Fig. 2b.

The small inductance in the arrangement limits the current through the air-gap or switch circuit. The TCSC is designed such that it provides 30% compensation at 180° (minimum)

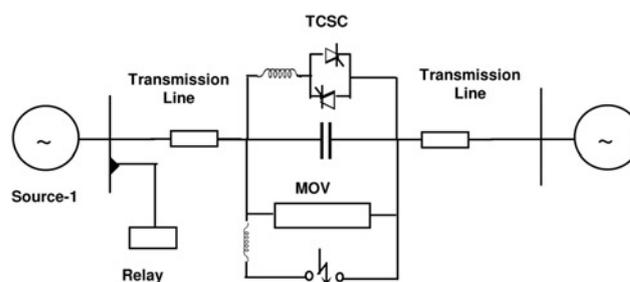


Figure 1 Transmission system with TCSC

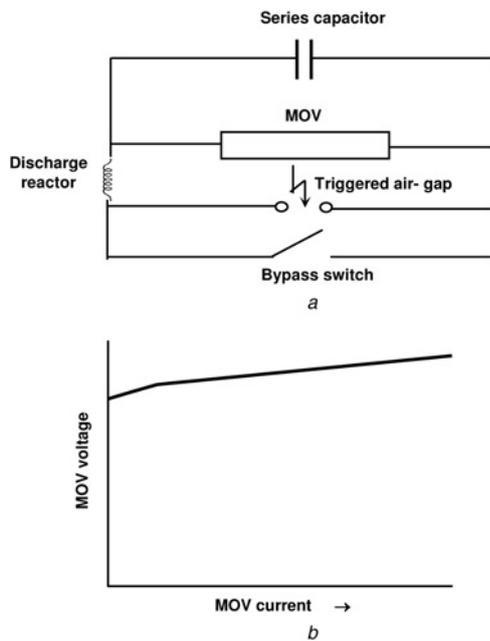


Figure 2 Typical series capacitor arrangement and the typical voltage–current characteristic of an MOV

a MOV protected series capacitor
b MOV characteristic

and 40% compensation at 150° (maximum) firing angle and in this study the firing angle is varied within this range as shown in Fig. 3.

The TCSC is placed at 50% of the transmission line with 300 km line length, which is 150 km from relaying end. The simulation for all ten types of shunt faults (L-G, LL-G, LL and LLL) are made on the transmission line with different fault resistance, source impedance and incident angles at different fault locations with varying the firing angle from 150° to 180° with (after) and without including (before) TCSC.

2.2 UPFC-based line

The schematic diagram for the studied system is shown in Fig. 4. The network having two areas connected by the

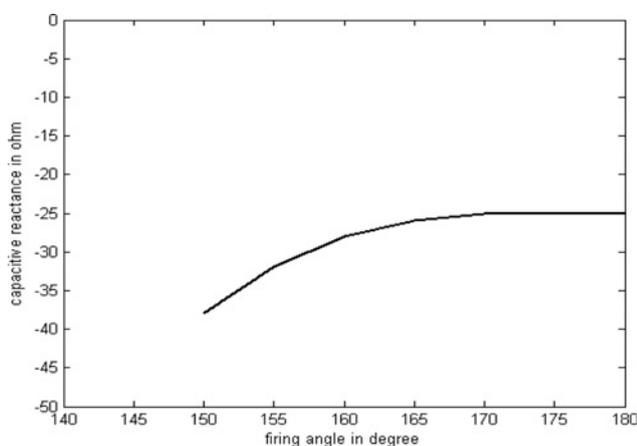


Figure 3 Variation of capacitive reactance with firing angle

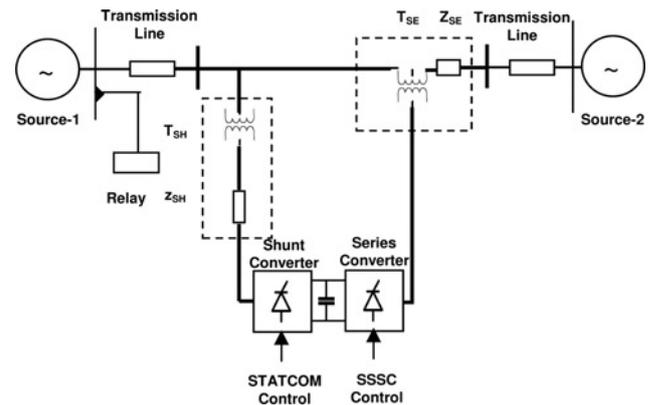


Figure 4 Transmission line with UPFC

transmission line of 230 kV and the transmission line (distributed model) parameters are same as mentioned for TCSC line. The UPFC is placed at 50% (mid point) of the line with series injected voltage varying from 0 to 15%. The relaying point is as shown in the Fig. 4, where data is retrieved for different fault conditions (L-G, LL-G, LL and LLL). The sampling rate is 1.0 kHz at 50 Hz base frequency. The system is simulated using PSCAD (EMTDC) package.

The basic components of the UPFC are two voltage source inverters (VSIs) sharing a common DC storage capacitor, and connected to the system through coupling transformers. One VSI is connected in shunt to the transmission system via a shunt transformer, while the other one is connected in series through a series transformer. The UPFC consists of two 48 pulse VSIs, which are connected through 2000 μF common DC capacitance. The STATCOM is connected to the power system through 400/20 kV shunt transformer and injects reactive power to the transmission system to regulate the voltage at the connecting point. Another inverter SSSC connects into the power system through 20/60 kV series transformer to inject almost sinusoidal voltage of variable magnitude and phase angle to regulate the power flow through the transmission line.

The UPFC control system is divided into two parts, STATCOM control and SSSC control. The STATCOM is controlled to operate the VSI for reactive power generation to the connecting point voltage to V_{ref} . The voltage at the connecting points are sent to the phase-locked-loop (PLL) to calculate the reference angle, which is synchronised to the reference phase voltage. The currents are decomposed into direct and quadrature component I_d and I_q by $d-q$ transformation using PLL angle as reference. The magnitude of the positive sequence component of the connecting point voltage is compared with V_{ref} and the error is passed through the PI controller to generate $I_{q\text{ref}}$. The reactive part of the shunt current is compared with $I_{q\text{ref}}$ and the error is passed through the PI controller to obtain the relative phase angle of the inverter voltage with respect to the reference phase voltage. This phase angle and the PLL signal are fed to

the STATCOM firing circuit to generate the desired pulse for the VSI.

The series injected voltage is determined by the closed-loop control system to ensure that the desired active and reactive power flow occurs despite power system changes. The desired P_{ref} and Q_{ref} are compared with the measured active and reactive power flow in the transmission line, and the error is passed through the PI controller to derive the direct and quadrature component of the series inverter voltage, V_{dref} and V_{qref} . Thus the series injected voltage and phase angle can be found out from the rectangular to polar conversion of the V_{dref} and V_{qref} . The dead angle (found out from the inverter voltage and DC link voltage), phase angle and the PLL signal are fed to the firing circuit to generate the desired pulse for the SSSC VSI.

3 Decision tree

DT [14–18] is the data mining classification algorithm used for high-dimension pattern classification. The mathematical representation of the DT algorithm is built on the following definitions

$$\begin{aligned} \bar{X} &= \{X_1, X_2, \dots, X_m\}^T \\ X_i &= \{x_1, x_2, \dots, x_{ij}, \dots, x_{in}\} \\ S &= \{S_1, S_2, \dots, S_i, \dots, S_m\}^T \end{aligned} \quad (1)$$

where m is the number of available observations (cases), n the number of independent variables (features), S the m -dimensional vector of the categorical (dependent) variable to be predicted from \bar{X} , X_i the i th component vector of n dimensional independent variables, $x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{in}$ the independent variables (predictors) of the pattern vector X_i and T the vector transpose notation.

The goal of DT data mining is to predict S based on observing \bar{X} . As many DTs with various accuracy levels can be constructed from a given \bar{X} [14], finding the optimal tree is difficult in practice because of the large size of the search space. However, powerful algorithms have been developed [15] to construct DTs with a reasonable trade-off between accuracy and complexity. These algorithms use a strategy that grow a DT by making a series of locally optimum decisions about which feature (system parameter) to use for partitioning the data set \bar{X} . The right-sized (or optimal) DT T_{k0} is then constructed according to the following optimisation problem [14]

$$\hat{R}(T_{k0}) = \min_k \{ \hat{R}(T_k) \}, \quad k = 1, 2, \dots, K \quad (2)$$

$$\hat{R}(T) = \sum_{t \in T} \{ r(t) \cdot p(t) \} \quad (3)$$

where $\hat{R}(T_k)$ is the misclassification error rate of the tree T_k , T_{k0} the optimal DT model that minimises the misclassification error $\hat{R}(T_k)$, T a binary tree $\in \{T_1, T_2, \dots, T_K, t_1\}$, k is the tree index number, t is a node in a tree, with t_1 the root node, $r(t)$ is the resubstitution estimate of the misclassification error of a case in node t and $p(t)$ the probability that any case falls into node t .

Any binary DT T is a collection of nested binary partitions, denoted here by a quintuple (n, q, m, n^L, n^R) , represented in the following recursive form [14]

$$T = \{(n, q, m, n^L, n^R), T^L, T^R\} \quad (4)$$

where n denotes a decision node label for the partition, q the feature axis, m the threshold value used for the partition and n^L and n^R the node labels for the partition of the left and right sets, respectively. Here T^L and T^R denote the subtrees defined on the left and right sets of a partition. In summary, (4) defines the DT T in terms of the pattern lattice L created by partitioning the features plane. The equation states that the lattice L can be binary partitioned on the feature axis q into mutually exclusive left and right sets, as also depicted in Fig. 5 for the two-dimensional binary classification boundaries shown in Fig. 6. The left set

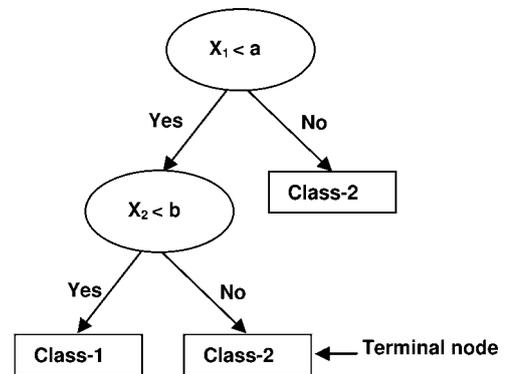


Figure 5 DT for binary classification

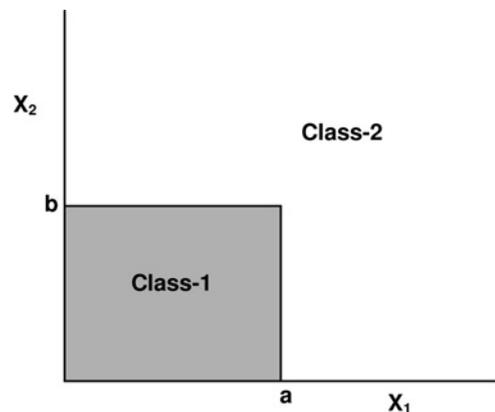


Figure 6 Two-dimensional binary classification

includes lattice elements with feature q values smaller than the threshold value while the right set includes lattice elements with feature q values exceeding the threshold value. The CART software package is used in the proposed study to develop the DT for identifying the fault zone for the TCSC and the UPFC-based transmission line.

4 DT-based fault zone identification

Initially the fault current and voltage information are collected at the relaying point. One cycle post fault current and voltage samples after fault inception are fed to the DT as input vector against target output of '1' for fault after the TCSC/UPFC and '0' for fault before TCSC/UPFC. Thus there are six sets of inputs against one output. The inputs are $ia = [ia0, ia1, ia2, \dots, ia19]$, $ib = [ib0, ib1, ib2, \dots, ib19]$, $ic = [ic0, ic1, ic2, \dots, ic19]$, $va = [va0, va1, va2, \dots, va19]$, $vb = [vb0, vb1, vb2, \dots, vb19]$, $vc = [vc0, vc1, vc2, \dots, vc19]$. As the sampling frequency is 1.0 kHz, one cycle contains 20

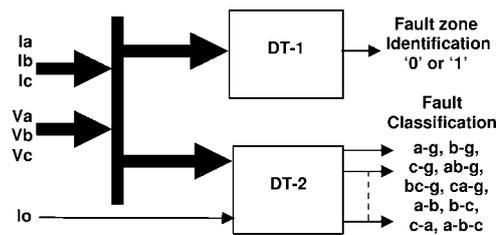


Figure 7 Proposed DT-based protection scheme

samples. Thus the input vector (one set) contains 120 data points for one target output.

The data sets are generated under various operating conditions of the power system network as follows.

- Variations in fault resistance (R_f) from 0 to 200 Ω
- Variations in source impedance (Z_s) by 30% from normal value

Table 1 Fault zone identification for TCSC and UPFC line

Faults	DT output
'ag'-type fault at 10% of the line, $R_f = 0 \Omega$, FIA = 30° , $\alpha = 180^\circ$ (TCSC)	0
'ag'-type fault at 30% of the line, $R_f = 20 \Omega$, FIA = 45° , $\alpha = 165^\circ$ (TCSC)	0
'bg'-type fault at 30% of the line, $R_f = 50 \Omega$, FIA = 60° , Vse = 5%, $\theta = 45^\circ$ (UPFC)	0
'cg'-type fault at 30% of the line, $R_f = 75 \Omega$, FIA = 90° , Vse = 7%, $\theta = 60^\circ$ (UPFC)	0
'ab'-type fault at 40% of the line, $R_f = 100 \Omega$, FIA = 30° , $\alpha = 160^\circ$ (TCSC)	0
'bc'-type fault at 40% of the line, $R_f = 120 \Omega$, FIA = 45° , Vse = 10%, $\theta = 90^\circ$ (UPFC)	0
'ca'-type fault at 40% of the line, $R_f = 150 \Omega$, FIA = 45° , Vse = 12%, $\theta = 120^\circ$ (UPFC)	0
'ag'-type fault at 48% of the line, $R_f = 200 \Omega$, FIA = 90° , $\alpha = 155^\circ$ (TCSC)	0
'ag'-type fault at 48% of the line, $R_f = 200 \Omega$, FIA = 30° , Vse = 15%, $\theta = 150^\circ$ (UPFC)	0
'ag'-type fault at 52% of the line, $R_f = 50 \Omega$, FIA = 45° , $\alpha = 165^\circ$ (TCSC)	1
'ag'-type fault at 52% of the line, $R_f = 120 \Omega$, FIA = 60° , Vse = 5%, $\theta = 180^\circ$ (UPFC)	1
'abg'-type fault at 60% of the line, $R_f = 150 \Omega$, FIA = 90° , $\alpha = 165^\circ$ (TCSC)	1
'bcg'-type fault at 60% of the line, $R_f = 200 \Omega$, FIA = 30° , Vse = 7%, $\theta = 90^\circ$ (UPFC)	1
'ag'-type fault at 70% of the line, $R_f = 50 \Omega$, FIA = 45° , $\alpha = 155^\circ$ (TCSC)	1
'ab'-type fault at 80% of the line, $R_f = 100 \Omega$, FIA = 90° , $\alpha = 175^\circ$ (TCSC)	1
'abcg'-type fault at 90% of the line, $R_f = 120 \Omega$, FIA = 45° , Vse = 9%, $\theta = 270^\circ$, $Z_s = 130\%$ (UPFC)	1
'ag'-type fault at 70% of the line with fault resistance 100 Ω , $R_f = 150 \Omega$, FIA = 90° , $\alpha = 165^\circ$ (TCSC)	1
'ag'-type fault at 70% of the line with fault resistance 100 Ω , $R_f = 0 \Omega$, FIA = 30° , Vse = 12%, $\theta = 210^\circ$ (UPFC)	1
'abg'-type fault at 70% of the line with fault resistance 100 Ω , $R_f = 150 \Omega$, FIA = 90° , $\alpha = 150^\circ$, $Z_s = 130\%$ (TCSC)	1
'abc'-type fault at 90% of the line with fault resistance 100 Ω , $R_f = 200 \Omega$, FIA = 90° , Vse = 15%, $\theta = 360^\circ$, $Z_s = 130\%$ (UPFC)	1

Table 2 Confusion matrix for fault zone identification for TCSC line

	Actual	
50% training and 50% testing (classification accuracy 96.0%)		
Predicted	0	1
0	2015	136
1	200	6049
60% training and 40% testing (classification accuracy 97.0%)		
0	1568	100
1	102	4950
70% training and 30% testing (classification accuracy 98.8%)		
0	1287	30
1	31	3692
80% training and 20% testing (classification accuracy 97.2%)		
0	1058	40
1	55	2207
90% training and 10% testing (classification accuracy 97.0%)		
0	659	21
1	31	970

Table 3 Confusion matrix for fault zone identification for UPFC line

	Actual	
50% training and 50% testing (classification accuracy 96.6%)		
Predicted	0	1
0	4057	209
1	226	8308
60% training and 40% testing (classification accuracy 97.1%)		
0	3076	140
1	157	6867
70% training and 30% testing (classification accuracy 98.5%)		
0	2030	56
1	60	5850
80% training and 20% testing (classification accuracy 97.2%)		
0	1650	70
1	74	3326
90% training and 10% testing (classification accuracy 96.5%)		
0	824	41
1	497	1646

- Variations in fault location: 10, 30, 55, 70 and 90% of the line
- Variations in inception angle (FIA): 30°, 45°, 60°, 90°
- Different types of fault: $a-g$, $b-g$, $c-g$, $ab-g$, $bc-g$, $ca-g$, $a-b$, $b-c$, $c-a$, $a-b-c$
- Reverse power flow
- Sudden load change
- TCSC firing angle (α) changed from 180° (minimum compensation) to 150° (maximum compensation)
- UPFC series injected voltage (V_{se}) varied for 0–15% of the normal voltage
- UPFC voltage phase angle (θ) varied form 0–360°.

Total simulations carried out for TCSC line are $5 (R_f) \times 3 (Z_s) \times 4 (FIA) \times 10$ (types of fault) $\times 2$ (reverse power flow) $\times 2$ (load change) $\times 7$ (firing angles) = 16 800. The total fault simulations carried out for the UPFC line are $5 (R_f) \times 2 (Z_s) \times 4 (FIA) \times 10$ (types of fault) $\times 2$ (reverse power flow) $\times 2$ (load change) $\times 4 (V_{se}) \times 4$ (phase angles) = 25 600. The DT is trained and tested for different sets of data such as 50–50, 60–40, 70–30, 80–20 and 90–10% for training and testing data sets, respectively. The above process is done to acquire the information of the training pattern on testing data sets. The proposed protection scheme is shown in Fig. 7. DT-1 is used for fault zone identification and DT-2 is used for fault classification.

Table 1 depicts the fault zone identification for both the TCSC and the UPFC line for different operating conditions. The DT-based algorithm provides '0' for fault before the TCSC/UPFC and '1' for faults after TCSC/UPFC. For 'a-g' fault at 48% of the line, $R_f = 200 \Omega$, $FIA = 90^\circ$, $\alpha = 155^\circ$

Table 4 Comparison between DT and SVM for fault zone identifications

TCSC line				
Predicted	DT (70% train and 30% test)		SVM (70% train and 30% test)	
	0	1	0	1
0	1287	30	1076	156
1	31	3692	167	3641
accuracy, %	98.8		93.6	
processing time, s	0.88		5.32	
DT (80% train and 20% test)				
0	1058	40	987	120
1	55	2207	136	2117
accuracy, %	97.2		92.4	
processing time, s	0.88		5.32	
UPFC line				
Predicted	DT (70% train and 30% test)		SVM (70% train and 30% test)	
	0	1	0	1
0	2030	56	2067	220
1	60	5850	226	5167
accuracy, %	98.5		94.2	
processing time, s	0.96		6.98	
DT (80% train and 20% test)				
0	1650	70	1567	196
1	74	3326	199	3158
accuracy, %	97.2		92.3	
processing time, s	0.96		6.98	

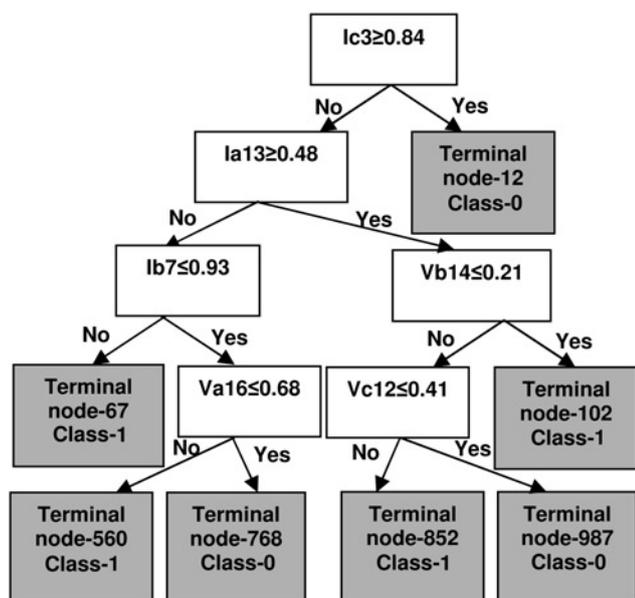


Figure 8 DT-based fault zone identification tree for TCSC and UPFC line

(TCSC), DT provides output '0' showing the fault occurrence before the TCSC. Similarly, for 'a-g' fault at 52% of the line, $R_f = 120 \Omega$, $FIA = 60^\circ$, $V_{se} = 5\%$, $\theta = 180^\circ$ (UPFC), DT provides output '1' showing fault occurrence after the UPFC.

Table 2 provides the confusion matrix, which shows the predicted against actual conditions during testing for fault zone identification for TCSC line. The DT provides the confusion matrix only on testing data sets. Thus for examples, 70–30% data sets mean, the confusion matrix provides classification results on 30% of total data sets. The results for different training and testing data sets are

found out for generalising the perfect training and testing data sets for best results. For 50–50% training and testing data sets, 8064 (6049 + 2015) cases are classified, where as 336 (200 + 136) cases are misclassified, providing classification accuracy 96.0%. Similar results are provided for 60–40, 70–30, 80–20 and 90–10% training and testing data sets. Table 3 provides confusion matrix for fault zone identification of the UPFC line for different training and testing data sets. It is observed that DT provides maximum classification accuracy of 98.8% for TCSC line and 98.5% for the UPFC line with 70–30% training data sets, compared to all other training and testing data sets. It is also observed that the classification accuracy increases up to 70–30% data sets and decreases after that in case of 80–20 and 90–10% data sets. This occurs due to over fitting of data sets for the classification algorithm.

Comparisons between the existing SVM algorithm and proposed DT algorithm are given in Table 4 for both the TCSC and the UPFC-based line. The confusion matrix for both the TCSC and the UPFC line is given with the same training and testing data sets for comparison. The classification accuracy for SVM and DT are 93.6 and 98.8%, respectively, for 70–30% training and testing data sets. A similar observation is made with the UPFC line also. The offline comparison of their computational requirements on a PC desktop equipped with a Pentium IV – 3 GHz CPU, shows that the processing time of SVM is 5.32 s compared to 0.88 s with DT for TCSC line for same data sets. Similarly, for the UPFC line, the processing time is 6.98 s for SVM and 0.96 s for DT with same data sets. Thus the computational burden of SVM is higher compared to DT. The DT-based fault zone identification tree is shown in Fig. 8.

Table 5 Confusion matrix for fault classification using DT for 70% training and 30% testing case (TCSC line)

Predicted (types of fault)	Actual (types of fault)									
	1(a-g)	2(b-g)	3(c-g)	4(ab-g)	5(bc-g)	6(ca-g)	7(a-b)	8(b-c)	9(c-a)	10(abc)
1(a-g)	496	3	0	2	0	0	1	1	1	0
2(b-g)	0	498	1	2	1	0	0	2	0	0
3(c-g)	0	0	492	3	4	0	2	1	2	0
4(ab-g)	0	0	0	488	5	3	1	7	0	1
5(bc-g)	3	1	0	0	490	3	0	5	2	0
6(ca-g)	0	4	0	3	0	492	3	2	0	0
7(a-b)	1	7	0	2	0	3	490	0	1	0
8(b-c)	0	3	5	0	0	1	0	495	0	0
9(c-a)	0	0	1	3	0	1	2	0	497	0
10(a-b-c)	1	2	0	3	0	1	0	0	0	497

Table 6 Confusion matrix for fault classification using DT for 70% training and 30% testing case (UPFC line)

Predicted (types of fault)	Actual (types of fault)									
	1(<i>a-g</i>)	2(<i>b-g</i>)	3(<i>c-g</i>)	4(<i>ab-g</i>)	5(<i>bc-g</i>)	6(<i>ca-g</i>)	7(<i>a-b</i>)	8(<i>b-c</i>)	9(<i>c-a</i>)	10(<i>abc</i>)
1(<i>a-g</i>)	756	2	5	1	0	3	0	0	1	0
2(<i>b-g</i>)	0	752	4	1	5	0	2	0	3	1
3(<i>c-g</i>)	1	0	760	1	2	0	1	0	1	0
4(<i>ab-g</i>)	2	1	0	751	1	6	0	6	0	1
5(<i>bc-g</i>)	0	5	0	0	755	2	3	0	2	1
6(<i>ca-g</i>)	4	3	0	2	6	752	0	1	0	0
7(<i>a-b</i>)	5	0	3	1	3	0	751	1	4	0
8(<i>b-c</i>)	6	0	7	1	2	4	0	748	3	1
9(<i>c-a</i>)	3	4	1	0	3	0	5	0	752	0
10(<i>a-b-c</i>)	0	3	5	0	1	1	0	0	0	758

5 DT for fault classification

For fault classification, one cycle fault current and voltage signal samples along with the zero-sequence currents are used as an input vector against different categories of the faults as target output for the DT-2. The inputs are $ia = [ia0, ia1, ia2, \dots, ia19]$, $ib = [ib0, ib1, ib2, \dots, ib19]$, $ic = [ic0, ic1, ic2, \dots, ic19]$, $io = [io1, io2, io3, \dots, io19]$, $va = [va0, va1, va2, \dots, va19]$, $vb = [vb0, vb1, vb2, \dots, vb19]$, $vc = [vc0, vc1, vc2, \dots, vc19]$. Thus the input vector (one set) contains 140 data points for one target output (one category of fault). Different faults classes are categorised as 1(*a-g*), 2(*b-g*), 3(*c-g*), 4(*ab-g*), 5(*bc-g*), 6(*ca-g*), 7(*a-b*), 8(*b-c*), 9(*c-a*) and 10(*a-b-c*) as target output. The fault simulations are carried out with various operating conditions as given in the previous section. Total fault cases simulated for TCSC line is 16 800 and for the UPFC line is 25 600. As 70–30% training and testing data set is the generalised one for better classification accuracy (as found in the fault zone identification section), this ratio is considered for fault classification task.

The confusion matrix obtained for fault classification is depicted in Table 5 for the TCSC-based line with 70–30% training and testing data set. The confusion matrix provides results for testing data sets for 5040 cases (30% of 16 800). In Table 5, the predicted against the actual class are given for all ten types of faults. In case of class 1(*a-g*), 496 cases are classified where as eight misclassified (with other classes). For class 2(*b-g*), 498 cases are classified and six are misclassified. Similar results are depicted for other types of fault cases.

Table 6 depicts fault classification results for UPFC line with 70–30% training and testing data sets. The confusion matrix provides results on 7680 fault cases (30% of

25 600). For class 1(*a-g*), 756 cases are classified and 12 misclassified. For class 2(*b-g*), 752 cases are classified and 16 cases are misclassified. Fault classification results for other types of faults are depicted in the table.

The comparison of classification accuracies between the existing SVM and proposed DT for different fault cases are depicted in Table 7. It is found that, for *L-G* (line to ground – *a-g*, *b-g*, *c-g*) faults, the classification accuracy is 98.28 and 94.31% for DT and SVM, respectively, for the TCSC line. For other fault types such as *LL-G*, *L-L*, *L-L-L* faults, the classification accuracies are higher in case of DT (around 98%) compared to SVM (around

Table 7 Performance comparison between DT and SVM for different types of faults for 70% training and 30% testing data sets

TCSC line		
Types of faults	SVM (classification accuracy), %	DT (classification accuracy), %
<i>L-G</i>	94.31	98.28
<i>LL-G</i>	94.42	97.22
<i>L-L</i>	93.51	98.01
<i>L-L-L</i>	94.54	98.61
UPFC line		
<i>L-G</i>	93.52	98.43
<i>LL-G</i>	94.13	98.00
<i>L-L</i>	92.81	97.69
<i>L-L-L</i>	93.64	98.70

94%). Similar observations are made for all types of faults for the UPFC-based line.

The comparison between SVM and DT is made for different types of faults with mixed data from the TCSC and the UPFC line for fault zone identification as depicted in Table 8. It is observed that the classification accuracy for L-G fault is 98.7% with DT compared 94.1% in case of SVM. It is observed that the classification accuracy with DT is higher compared to SVM for all other types of faults also.

The processing time of SVM is 15.67 s compared to 1.12 s of DT for TCSC line data sets. Similarly, for UPFC line, the processing time is 18.76 s for SVM and 1.23 s for DT with same data sets. Similar observation is made for mixed data sets as given in the Table 9. Thus it is observed that the computational burden of SVM is higher compared to DT for fault classification task.

Table 8 Performance comparison between DT and SVM for different types of fault for 70% training and 30% testing data sets with mixed TCSC and UPFC data

Types of faults	SVM (classification accuracy), %	DT (classification accuracy), %
L-G	94.1	98.7
LL-G	93.8	98.4
L-L	92.8	97.2
L-L-L	94.1	98.4

Table 9 Processing time comparison between DT and SVM for different types of faults for 70–30% training and testing data sets

Types of faults	Processing time (SVM), s	Processing time (DT), s
TCSC	15.67	1.12
UPFC	18.76	1.23
TCSC + UPFC	23.21	1.56

The DT is also tested for data with SNR 20 dB and the classification accuracies for different faults are depicted in Table 10. It is found that the classification accuracy is marginally affected for DT compared to SVM with noise. This ensures robustness of the proposed DT algorithm. The DT-based fault classification tree is shown in Fig. 9.

Table 10 Performance comparison between DT and SVM for different types of fault for 70% training and 30% testing data sets with mixed UPFC and TCSC data with SNR 20 dB

Types of faults	SVM (classification accuracy), %	DT (classification accuracy), %
L-G	90.7	97.9
LL-G	89.8	97.4
L-L	88.9	96.8
L-L-L	91.1	97.7

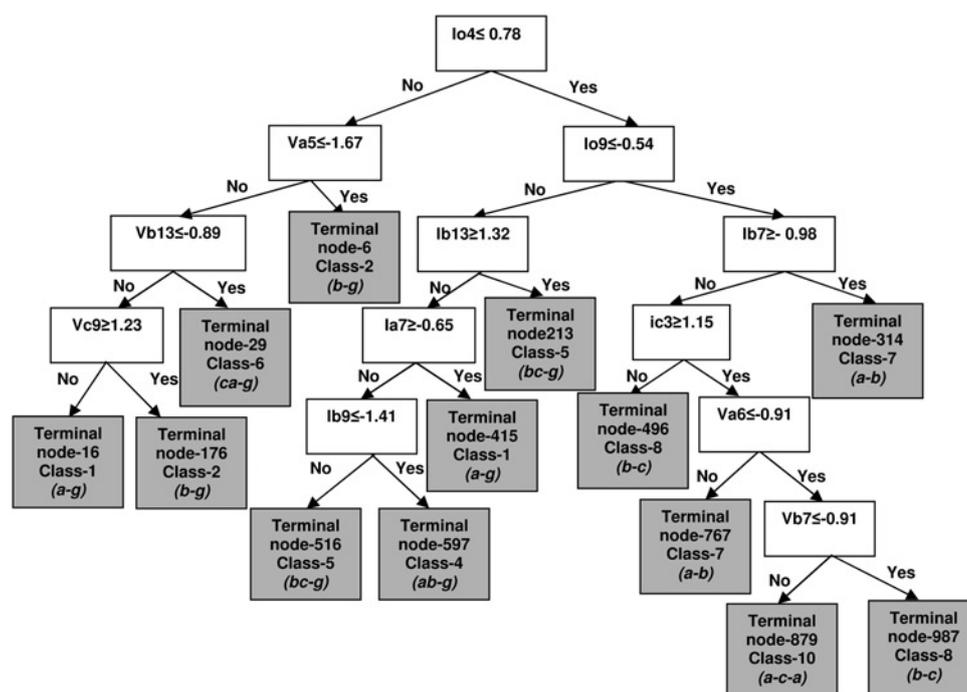


Figure 9 DT-based fault classification tree for TCSC and UPFC line

6 Discussion

The previous sections provide a new approach for fault zone identification and fault classification for the TCSC and UPFC-based transmission lines using DT. One cycle post fault current and voltage signals are used to build up a classification tree for fault zone identification and fault classification. The proposed DT-based approach is compared with the existing SVM approach for the same. It is found that the classification accuracy is higher in case of DT compared to SVM for different combinations of data sets and fault types.

As for the computational burden, there is a large gap between SVM and DT algorithm, but the computational burden is higher for SVM compared to DT for fault zone identification and fault classification task. This raises concerns about the practical implementation of SVM-based decision making on low-cost DSPs. By contrast, the DT method executes even faster than neural networks and definitely has an edge in real-time application. The classification accuracy is also higher in case of DT compared to SVM for the sample data sets taken for the proposed study including the noisy environment, which ensures robustness of the proposed DT algorithm.

7 Conclusions

The proposed approach provides a new technique for fault zone identification and fault classification for transmission line employing the TCSC and the UPFC using DT. The proposed DT is found to be better compared to existing SVM with respect to accuracy and computational burden. The DT is tested with data sets with wide variations in operating conditions of the power system network including the noisy environment and provides accurate results. The robustness and accuracy of the proposed DT show potential of the proposed method for protection of the FACTS line in large power network. Dealing with current transformer nonlinearity, harmonic and interharmonic resonances and so on will also require a much larger data base, although the principles presented in this paper will still hold.

8 References

- [1] GIRGIS A.A., SALLAM A.A., KARIM EL-DIN A.: 'An adaptive protection scheme for advanced series compensated (ASC) transmission line', *IEEE Trans. Power Deliv.*, 1998, **13**, pp. 414–420
- [2] NOROOZIAN M., ANGQUIST L., GHANDHARI M., ANDERSON G.: 'Improving power system dynamics by series connected FACTS devices', *IEEE Trans. Power Deliv.*, 1997, **12**, (4), pp. 1635–1641
- [3] GYUGYI L.: 'Unified power flow concept for flexible AC transmission systems', *IEE Proc. C*, 1992, **139**, (4), pp. 323–332
- [4] ZHOU X., WANG H., AGGARWAL R.K., BEAUMONT P.: 'Performance of evaluation of a distance relay as applied to a transmission system with UPFC', *IEEE Trans. Power Deliv.*, 2006, **21**, (3), pp. 1137–1147
- [5] EI-ARROUDI K., JOOS G., MCGILLS D.T.: 'Operation of impedance relays with the STATCOM', *IEEE Trans. Power Deliv.*, 2002, **17**, (2), pp. 381–387
- [6] KHEDERZADEH M., SIDHU T.S.: 'Impact of TCSC on the protection of transmission lines', *IEEE Trans. Power Deliv.*, 2006, **21**, (1), pp. 80–87
- [7] SONG Y.H., XUAN Q.Y., JOHNS A.T.: 'Protection of scheme for EHV transmission systems with thyristor controlled series compensation using radial basis function neural networks', *Electr. Mach. Power Syst.*, 1997, **25**, pp. 553–565
- [8] SONG Y.H., JOHNS A.T., XUAN Q.Y.: 'Artificial neural network based protection scheme for controllable series-compensated EHV transmission lines', *IEE Proc., Gener. Transm. Distrib.*, 1996, **143**, (6), pp. 535–540
- [9] PARIKH U.B., DAS B., MAHESHWARI R.P.: 'Combined wavelet-SVM technique for fault zone detection in a series compensated transmission line', *IEEE Trans. Power Deliv.*, 2008, **23**, (4), pp. 1789–1794
- [10] YANG H.-T., LIAO C.-C.: 'A de-noising scheme for enhancing wavelet-based power quality monitoring system', *IEEE Trans. Power Deliv.*, 2001, **16**, (3), pp. 353–360
- [11] SUN K., LIKHATE S., VITTAL V., KOLLURI V.S., MANDAL S.: 'An online dynamic security assessment scheme using phasor measurements and decision trees', *IEEE Trans. Power Syst.*, 2007, **22**, (4), pp. 1935–1943
- [12] WEHENKEL L., PAVELLA M., EUXIBIE E., HEILBRONN B.: 'Decision tree based transient stability method a case study', *IEEE Trans. Power Syst.*, 1994, **9**, (1), pp. 459–469
- [13] EI-ARROUDI K., JOOS G., KAMWA I., MCGILLS D.T.: 'Intelligent-based approach to islanding detection in distributed generation', *IEEE Trans. Power Deliv.*, 2007, **22**, (2), pp. 828–835
- [14] BREIMAN L., FRIEDMAN J.H., OLSEN R.A., STONE C.J.: 'Classification and regression trees' (Chapman & Hall Wadsworth, Inc., New York, 1984)
- [15] SAFAVIAN S.R., LANDGREBE D.: 'A survey of decision tree classifier methodology', *IEEE Trans. Syst. Man Cybern.*, 1991, **21**, (3), pp. 660–674

[16] ROKACH L., MAIMON O.: 'Top-down induction of decision trees classifiers – a survey', *IEEE Trans. Syst. Man Cybern. C*, 2005, **35**, (4), pp. 476–487

[17] HESS K.R., ABBRUZZESE M.C., LENZI R., RABER M.N., ABBRUZZESE J.L.: 'Classification and regression tree analysis of 1000

consecutive patients with unknown primary carcinoma', *Clin. Cancer Res.*, 1999, **5**, pp. 3403–3410

[18] KASS G.V.: 'An exploratory technique for investigating large quantities of categorical data', *Appl. Statist.*, 1980, **29**, (2), pp. 119–127