A Data-Mining Model for Protection of FACTS-Based Transmission Line

S. R. Samantaray, *Senior Member, IEEE*

***Abstract*—This paper presents a data-mining model for fault-zone identification of a flexible ac transmission systems (FACTS)-based transmission line including a thyristor-controlled series compensator (TCSC) and unified power-flow controller (UPFC), using ensemble decision trees. Given the randomness in the ensemble of decision trees stacked inside the random forests model, it provides effective decision on fault-zone identification. Half-cycle postfault current and voltage samples from the fault inception are used as an input vector against target output "1" for the fault after TCSC/UPFC and "−1" for the fault before TCSC/UPFC for fault-zone identification. The algorithm is tested on simulated fault data with wide variations in operating parameters of the power system network, including noisy environment providing a reliability measure of 99% with faster response time (3/4th cycle from fault inception). The results of the presented approach using the RF model indicate reliable identification of the fault zone in FACTS-based transmission lines.**

***Index Terms*—Distance relaying, fault-zone identification, random forests (RFs), support vector machine (SVM), thyristor-controlled series compensator (TCSC), unified power-flow controller (UPFC).**

I. Introduction

INCREASED demand of bulk power transfer in the modern power network has led to an increased focus on transmission constraints and alleviation. Flexible ac transmission systems (FACTS) [1] devices offer a versatile alternative to conventional reinforcement methods. Among them, the thyristor-controlled series compensator (TCSC) [2] and unified power-flow controller (UPFC) [3] are important FACTS devices, which are used extensively for improving the utilization of the existing transmission system. The presence of TCSC in fault loop not only affects the steady-state components but also the transient components. The controllable reactance, the metal–oxide varistors (MOVs) protecting the capacitors, and the air-gap operation make the protection decision more complex and, therefore, the conventional relaying scheme based on fixed settings finds limitations. On the other hand, UPFC offers new horizons in terms of power system control. While the use of UPFC improves the power transfer capability and stability of a power system, certain other problems emerge in transmission-line protection [4]–[6], greatly affecting the reach of the distance relay.

In the FACTS-based transmission line, if the fault does not include FACTS device, then the impedance calculation is like an ordinary transmission line, and when the fault includes FACTS, then the impedance calculation accounts for the impedances introduced by 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 issues a signal to trip the circuit breaker (CB). Further, for similar types of faults, the current level may be of the same order at two different points of the transmission line, (before and after TCSC/UPFC). Thus, before the apparent impedance to the fault point is computed, a more reliable and accurate fault-zone identification technique is necessary for safe and reliable operation of the distance relay. The correct fault-zone identification in presence of the FACTS devices, such as TCSC and UPFC in the transmission line, is one of the critical tasks to be dealt with.

Recent techniques based on neural networks [7], [8], find limitations, since they require a large number of neurons to model the structure of the network involving large training sets and training time. Recently, a hybrid technique using a wavelet transform combined with support vector machine (SVM) [9], [10] has been proposed for fault-zone identification in the TCSC line. The aforementioned work finds limitations since the wavelet transform is highly prone to noise and provides erroneous results even with a signal-to-noise ratio (SNR) of 30 dB [12]. Also, the computational time of SVM is higher compared to the proposed ensemble DTs-based data-mining model, which puts constraints on the online realization of SVM-based relays for distance relaying applications, where speed and accuracy are prime considerations. Thus, there is a strong motivation to build up an accurate and faster data-mining model for fault-zone identification in FACTS-based transmission lines.

The proposed research is based on a data-mining model known as ensemble decision trees [13]–[18], also known as random forests (RFs), for fault-zone assessment in a FACTS (TCSC/UPFC)-based transmission line. Half-cycle postfault current and voltage samples (time-domain data samples) are used as inputs to the RF against target outputs "−1" for faults before TCSC/UPFC and "1" for faults after TCSC/UPFC. The RF is trained to build a data-mining model with an extensive data set derived from a series of fault simulations. The proposed technique is tested on wide variations in operating parameters in the power system network, including a noisy environment and, was found to be accurate and robust for fault-zone identification in TCSC/UPFC-based transmission lines. The following sections deal with RFs, systems studied, results, analysis, discussion, and conclusions.

Manuscript received July 20, 2011; revised May 28, 2012, October 13, 2012, and December 12, 2012; accepted January 18, 2013. Date of publication February 22, 2013; date of current version March 21, 2013. This work was supported by the Department of Science and Technology, Government of India (SR/FTP/ETA-78/2010). Paper no. TPWRD-00614-2011.

The author is with the School of Electrical Sciences, Indian Institute of Technology Bhubaneswar, Orissa 751013, India (e-mail: sbh_samant@yahoo.co.in).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRD.2013.2242205

II. Random Forests

RFs are a large combination of de-correlated tree predictors such that each tree depends on the values of a random vector sampled independently. Individual trees are noisy and unstable, but since when grown sufficiently deep, they have relatively low bias. Therefore, they are ideal candidates for ensemble growing since they can capture complex interactions, while fully benefitting from aggregation-based variance reduction. Using a random selection of features to split each node and resampling (with replacement) the training set to grow each tree yields error rates that are de-correlated and more robust with respect to noise. The generalization error of forests converges to a limit since the number of trees in the forest increases.

The basic idea of most ensemble tree growing procedures is that for the kth tree ($k \leq n_{\text{tree}}$, the number of trees in the ensemble) a random vector Φ_k is generated, independent of the past random vectors $\Phi_k, \ldots \Phi_{k-1}$, but with the same distribution, and a single tree is grown using the training set S and the set of attributes in Φ_k, resulting in a classifier $T_k(x, \Phi_k)$ where x is an input vector. In random split selection, Φ consists of a number n_{try} of independent random integers where $n_{\text{try}} < n_a$ is the number of attributes in.

An RF consists of a collection of tree-structured classifiers $\{T_k(x, \Phi_k), k = 1, \ldots, n_{\text{tree}}\}$, where $\{\Phi_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x. An algorithmic view of the RF growing process is summarized as follows [13]:

1) For $k = 1$ to n_{tree}:
   a) Draw a boostrap sample S* of size N from the training data S
   b) Grow a random forest tree $T_k(x, \Phi_k)$ to the boostrapped data, by recursively repeating the steps below for each terminal noted of the tree, until the no other split is possible (unpurned tree of maximal depth):
      i) Select n_{try} variables from the n_a features
      ii) Pick the best variable/split-point among the n_{try}
      iii) Split the node into two daughter nodes
2) Output the ensemble of trees $\{T_k(x, \Phi_k), k = 1, \ldots, n_{\text{tree}}\}$.

A traditional decision tree essentially represents an explicit decision boundary, and an instance E is classified into class c if E falls into the decision area (a leaf in the decision tree) corresponding to c [16]. The class probability p(c|E) is typically estimated by the fraction of instances of class c in the leaf into which E falls. This probability estimate is very crude when the tree is pruned because all of the instances falling into the same leaf have the same class probability. More accurate probability estimates require unpruned trees [19], which are the backbone of the RFs. Stated otherwise, the RF predictor has the additional advantage of providing a stability or instability level of the event through probability-based ranking. Assuming that the probability estimates from individual trees are random variables, each with variance σ^2, the average variance is σ^2/n_{tree}, which confirms that the RF leads seamlessly to improved probability estimates [13].

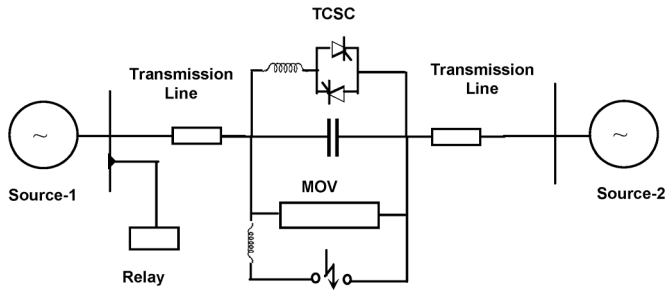

Fig. 1. Transmission line with TCSC.

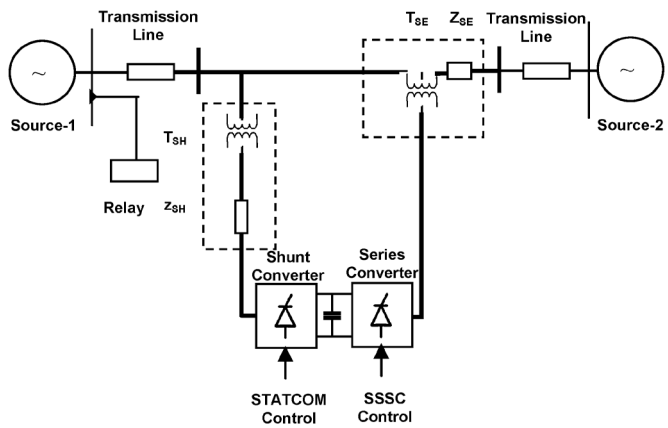

Fig. 2. Transmission line with UPFC.

Although the RF is a relatively young data-mining tool, people [20]–[22] have started recognizing its strengths: 1) it is simple and easy to use; 2) very high accuracy; 3) its relatively robust to outliers and noise; 4) it gives useful internal estimates of error, strength, and correlation; 5) not overfitting if selecting a large number of trees; and 6) insensitive to the choice of split.

III. System Studied

A. TCSC- and UPFC-Based Line

A 400-kV, 50-Hz power system is illustrated in Fig. 1. The power system consists of two sources: TCSC [10] and associated components, and a 300-km transmission line. The transmission line has zero-sequence impedance $Z_0 = 105.65 + j356.71\ \text{ohm}$ and positive-sequence impedance $Z_1 = 10.52 + j115.45\ \text{ohm}$. $E_s = 400$ kV and $E_R = 400\angle\delta$ kV.

The TCSC is designed such that it provides 30% compensation at 180° (minimum) and 40% compensation at 150° (maximum) firing angle, and in this study, the firing angle is varied within this range. The TCSC is placed at 30%, 50% and 80% of the transmission line to assess the impact of TCSC placement on the performance of the developed data-mining model. Similarly, UPFC [11] is placed at 30%, 50%, and 80% of the line (Fig. 2) with variations in series injected voltage and phase angle. The simulation includes all ten types of shunt faults (L-G: Line-Ground, LL-G: Line-Line-Ground, LL: Line-Line, and LLL: Line-Line-Line) with different fault

TABLE I
SUMMARY COUNT OF THE COMPLETE DATABASE GENERATED

Types of line	Numbers of cases generated
TCSC	**38400**
UPFC	**43200**

resistance, source impedance, inception angles, fault locations and firing angles. The systems studied are simulated using PSCAD (EMTDC) subroutines with a sampling rate of 1.0 kHz at 50-Hz base frequency.

B. Summary Count of the Data Sets Generated

Initially, the fault current and voltage signal samples are collected at the relaying point. Half cycles postfault current and voltage samples after fault inception are fed to the RFs as an input vector against the target output of "1" for the fault after TCSC/UPFC and "−1" for the fault before TCSC/UPFC. Thus, there are six sets of inputs against one output. The inputs are half cycle signal samples of three phase currents and three phase voltages after fault inception. Based on the sampling frequency of 1.0 kHz, half cycle contains 10 samples and thus the input vector (one case) contains 60 data points $(10 * 3$ for currents $+$ $10 * 3$ for voltage signals) against one target output.

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

- variations in fault resistance $(\mathrm{R_f})$ from 0 to 200 Ω
- variations in source impedance (Zs) by 30% from normal value;
- variations in fault location: 25%, 45%, 55%, and 85% of the line;
- variations in fault inception angle (FIA): 30°, 45°, 60°;
- different types of fault: a-g (a-phase to ground), b-g (b-phase to ground), c-g(c-phase to ground), ab-g (a-b-phase to ground), bc-g (b-c-phase to ground), ca-g (c-a-phase to ground), a-b (a-phase to b-phase), b-c (b-phase to c-phase), c-a (c-phase to a-phase), a-b-c (a-phase to b-phase to c-phase);
- reverse power flow;
- sudden load change;
- TCSC firing angle (α) changed from 180° (minimum compensation) to 150° (maximum compensation);
- UPFC series injected voltage (Vse) varied at 5%, 10%, and 15% of the normal voltage;
- UPFC voltage phase angle (θ_{se}) varied from 0°–360°;
- FACTS device (TCSC/UPFC) at different locations (30%, 50%, 80%) in the transmission line.

Total simulations carried for TCSC line are $5\ (\mathrm{R_f}) \times 3\ (\mathrm{Zs}) \times 4\ (\mathrm{FIA}) \times 10$ (Types of fault)$\times 2$ (Reverse power flow) $\times$ 2 (load change)$\times 4$ (firing angles) $\times$ 4 (locations) $=$ 38400. The total fault simulations carried out for UPFC line are $5\ (\mathrm{R_f}) \times 2\ (\mathrm{Zs}) \times 3\ (\mathrm{FIA}) \times 10$ (Types of fault)$\times 2$(Reverse power flow) $\times$ 2 (load change)$\times 3\ (\mathrm{Vse}) \times 3$ (phase angles) $\times 4$ (locations) $=$ 43200. The summary count of the detailed data sets is given in Table I. The proposed RF-based protection scheme for fault-zone identification is shown in Fig. 3.

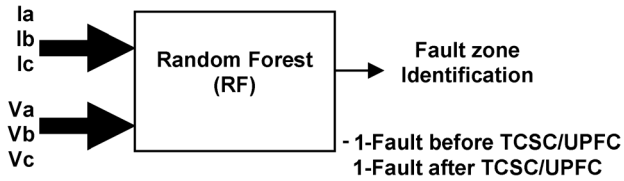

Fig. 3. Proposed RF-based fault-zone identification scheme.

TABLE II
COMPARISON OF THE CONFUSION MATRIX BETWEEN RF AND SVM FOR THE TCSC AND UPFC LINE FOR FAULT-ZONE IDENTIFICATION

TCSC Line				
	RF (70 % Train and 30 % Test)		SVM (70 % Train and 30 % Test)	
	Actual			
Predicted	**-1**	**1**	**-1**	**1**
-1	5732	28	5396	483
1	30	5730	209	5432
Accuracy (%)	**99.50 %**		**93.99 %**	
UPFC Line				
	RF (70 % Train and 30 % Test)		SVM (70 % Train and 30 % Test)	
	Actual			
Predicted	**-1**	**1**	**-1**	**1**
-1	6448	54	6059	346
1	50	6408	575	5980
Accuracy (%)	**99.19 %**		**92.89 %**	

IV. RESULTS

In the proposed study, an extensive data set is generated to train and test the RF-based data-mining model for developing an accurate and robust classifier for fault-zone identification in FACTS-based transmission line. The total data sets generated for TCSC and UPFC lines are 38400 and 43200, respectively. The proposed approach uses 26880 (70% for TCSC) and 30240 (70% for UPFC) cases for training, and the rest 30% for testing in each case, which is the most generalized training-testing ratio for data-mining algorithms. The RF is trained for 100 generations of trees to build the accurate model for fault-zone identification.

The results obtained using RF has been compared with SVM for similar applications [9], [10]. The same time-domain data set is used to train and test the SVM, without preprocessing through the wavelet transform. In this study, the Gaussian kernel is used for the SVM implementation and the values of SVM parameters, such as width of Gaussian function, bound on Lagrangian multipliers, and the conditioning parameter are the same (after cross validation) as considered in [10]. Table II depicts the confusion matrix generated using RF and SVM for fault-zone identification. "−1" corresponds to the fault before TCSC/UPFC and "1" for faults after TCSC/UPFC (placed at 50% of the line). The confusion matrix provides the results of the predicted class *versus* the actual class for the test data sets (30%). The accuracy for TCSC line with RF is 99.50% compared to 93.99 using SVM. Similar observations are made for UPFC, where the fault-zone identification accuracies are 99.19% and 92.89, with RF and SVM, respectively.

TABLE III
FAULT-ZONE IDENTIFICATION FOR DIFFERENT FAULT SITUATIONS FOR THE TCSC AND UPFC LINE USING RF

Types of Fault	No of cases	Mid-detection	Accuracy (%)
TCSC			
L-G	3456	14	99.59
LL-G	3456	18	99.47
LL	3456	19	99.45
LLL	1152	07	99.39
Overall	**11520**	**58**	**99.50**
UPFC			
L-G	3888	26	99.33
LL-G	3888	31	99.20
LL	3888	37	99.04
LLL	1296	10	99.22
Overall	**12960**	**104**	**99.19**

TABLE IV
CLASSIFICATION ACCURACIES AT DIFFERENT FIRING ANGLES-α (COMPENSATION LEVELS) FOR TCSC LINE USING RF

Types of Faults	Accuracy (%)			
	$\alpha=150^0$	$\alpha=160^0$	$\alpha=170^0$	$\alpha=180^0$
L-G	99.45	99.76	99.72	99.39
LL-G	99.32	99.89	99.57	99.87
LL	99.05	99.54	99.49	99.38
LLL	99.37	99.78	99.51	98.68

TABLE V
CLASSIFICATION ACCURACIES AT DIFFERENT SERIES-INJECTED VOLTAGE (Vse) FOR THE UPFC LINE USING RF

Types of Faults	Accuracy (%)		
	Vse=5%	Vse=10%	Vse=15%
L-G	99.52	99.39	99.58
LL-G	99.43	99.29	99.65
LL	99.09	99.11	99.45
LLL	99.28	99.39	99.49

Table III provides the complete statistics of fault-zone identification for different types of faults, such as line-ground (L-G), line-line-ground (LL-G), line-line (LL), and line-line-line (LLL) faults. The fault-zone identification accuracies are above 99% for each category of fault situations. The misclassification cases are only 58 and 104 for TCSC and UPFC (placed at 50% of the line), respectively.

Table IV depicts the classification accuracies of RF for the TCSC line (placed at 50% of the line) at different firing angles which in turn decide the compensation level. In the proposed study, firing angle $\alpha = 150^\circ$ corresponds to 40% (maximum) and $\alpha = 180^\circ$ corresponds to 30% (minimum) compensation. It is observed that at different compensation levels, the fault-zone identification accuracies of RF are more than 99%.

Table V depicts the classification accuracies at different series-injected voltage (Vse) in the case of UPFC line (UPFC placed at 50% of the line). It is found that the fault-zone identification accuracies are above 99% at different series-injected voltages of 5%, 10%, and 15%. Similar observation are made for the UPFC line with a different series-injected voltage phase angle (θ_{se}), and the results are highly promising (Table VI).

The performance comparisons between RF and SVM are depicted in Table VII. The classification accuracies are above 99%

TABLE VI
CLASSIFICATION ACCURACIES AT DIFFERENT SERIES-INJECTED VOLTAGE PHASE ANGLE (θ_{se}) FOR THE UPFC LINE USING RF

Types of Faults	Accuracy (%)		
	$\theta_{se}=90^0$	$\theta_{se}=210^0$	$\theta_{se}=360^0$
L-G	99.87	99.67	99.87
LL-G	99.64	99.32	99.65
LL	99.35	99.15	99.24
LLL	99.52	99.65	99.78

TABLE VII
COMPARISON BETWEEN RF AND SVM FOR DIFFERENT FAULT SITUATIONS

Types of Fault	**RF Accuracy (%)**	**SVM Accuracy (%)**
TCSC		
L-G	99.59	96.12
LL-G	99.47	96.34
LL	99.45	89.14
LLL	99.39	94.29
Overall	**99.50**	**93.99**
UPFC		
L-G	99.33	95.28
LL-G	99.20	94.82
LL	99.04	87.98
LLL	99.22	93.59
Overall	**99.19**	**92.89**

TABLE VIII
COMPARISON BETWEEN RF AND SVM FOR DIFFERENT FAULT SITUATIONS WITH AN SNR OF 20 dB (TCSC AND UPFC AT 50% OF THE LINE)

Types of Fault	**RF Accuracy (%)**	**SVM Accuracy (%)**
TCSC		
L-G	99.12	93.42
LL-G	99.08	94.21
LL	99.00	86.23
LLL	99.04	91.22
Overall	**99.06**	**91.27**
UPFC		
L-G	99.08	92.54
LL-G	99.03	91.76
LL	99.00	84.87
LLL	99.01	90.47
Overall	**99.03**	**89.91**

in the case of RF compared to 93% (around) that of SVM. Also, the performance of RF and SVM is assessed for the data sets with noise of a signal-to-noise ratio (SNR) of 20 dB (white Gaussian noise) as depicted in Table VIII. It is observed that the overall performance accuracy of SVM is highly degraded in the noisy environment touching 92.89%.

Table IX depicts the performance comparisons between RF and SVM for combined data sets (TCSC+UPFC). In this comparison, the data sets considered are 81600 (38400 + 43200), and 70% data are used for training and the remainder are 30% for testing. It is observed that the overall accuracy and reliability of RF are above 99% compared to 93% for that of SVM. Thus, the RF provides substantially improved results compared to SVM, leading to a more generalized classifier for fault-zone identification in FACTS-based transmission lines.

To test the impact of location of FACTS devices in the transmission lines on the performance of the RF-based predictor for

TABLE IX
COMPARISON BETWEEN RF AND SVM FOR COMBINED DATA SETS TCSC+UPFC (TCSC AND UPFC PLACED AT 50% OF THE LINE)

Fault situation	RF		SVM	
	Accuracy (%)	Reliability (%)	Accuracy (%)	Reliability (%)
L-G	99.07	99.24	95.05	94.52
LL-G	99.01	99.56	96.12	93.47
L-L	99.00	99.61	88.32	91.68
LLL	99.12	99.19	95.15	92.87
Overall	**99.05**	**99.40**	**93.66**	**93.13**

TABLE X
PERFORMANCE OF THE RF AT DIFFERENT LOCATIONS OF THE FACTS DEVICE IN TRANSMISSION LINES

Types of Fault	(30 %) Accuracy (%)	(50 %) Accuracy (%)	(80 %) Accuracy (%)
TCSC			
L-G	99.11	99.12	99.15
LL-G	99.08	99.08	99.02
LL	99.06	99.00	99.11
LLL	99.15	99.04	99.07
Overall	**99.10**	**99.06**	**99.08**
UPFC			
L-G	99.06	99.08	99.04
LL-G	99.09	99.03	99.01
LL	99.05	99.00	99.07
LLL	99.08	99.01	99.04
Overall	**99.07**	**99.03**	**99.04**

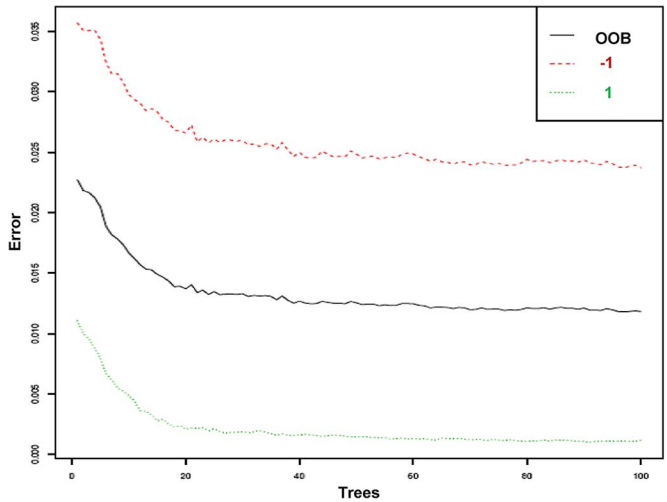

Fig. 4. Convergence characteristic of the RF during training for the TCSC line placed at 50% of the line.

TABLE XI
RELIABILITY MEASURE DERIVED FROM RF AND SVM

	Reliability (%)	
Fault situation	RF	SVM
TCSC	**99.51** (28)	**91.83** (483)
	99.47 (30)	**96.27** (209)
UPFC	**99.16** (54)	**94.53** (346)
	99.23 (50)	**91.33** (575)

fault-zone identification, the same has been tested for fault situations with FACTS devices (UPFC/TCSC) placed at different locations of the transmission line (30%, 50%, and 80%). Table X depicts the performance accuracy of RF for fault-zone identification at different locations of the TCSC/UPFC in the transmission line. It is observed that RF is able to identify the fault zone with accuracy more than 99%, taking all three situations into consideration.

The convergence characteristic of RF is represented by the out-of-bag prediction (OOB) error [13] and the advantage of OOB error is that more realistic estimate of the error rate can be obtained. If we feed the random forest with only 70% of the original data and keep the rest for testing, giving that each tree is trained on two third of the data only, it turns out that only 50% of the data are actually seen by a given random forest tree at learning stage. The RF is trained for 100 generations of trees and the convergence characteristic is shown in Fig. 4. There are three characteristic curves in Fig. 4, showing the red line for "−1" (faults before FACTS), green for "1" (faults after FACTS) and black for "OOB" error during training. It is observed that after 50 tree generations, the error becomes minimum and almost constant. However, 100 tree generations are considered for more reliable and accurate decision making.

V. ANALYSIS AND DISCUSSION

While comparing the performance with SVM [9], [10], it is observed that the classification accuracies are above 99% in case of RF compared to 93% of SVM, considering both TCSC and UPFC, together. In earlier studies, the performance accuracies have not been assessed in noisy environment, which is one of the critical issues in power system applications. The classification results in noisy environment (SNR 20 dB) are given in Table VIII and it is observed that the fault-zone identification accuracy of RF is above 99%, where as the overall performance of SVM is degraded to 90% (average), which shows the noise immunity of RF over SVM.

Reliability is one of the important measures of the classification results for fault-zone identification. In this analysis, reliability is defined as follows:

- **Reliability for faults after FACTS**: Number of cases predicted for faults after FACTS ("1")/(total number of actual cases for faults after FACTS);
- **Reliability for faults before FACTS**: Number of cases predicted for faults before FACTS ("−1")/(total number of actual cases for faults before FACTS).

In the aforementioned reliability analysis, total numbers of actual cases is equal to the numbers of cases predicted + numbers of cases mis-detetected for faults before or after FACTS. The complete statistics of the reliability is depicted in Table X. In this study, "1" corresponds to faults after TCSC/UPFC and "−1" corresponds to faults before TCSC/UPFC. The bold figures in Table XI depict the reliability in percentage and the figures in brackets show the numbers of misdetection to other classes. It is observed that 28 cases have been misdetected from class 1 to −1, thus the resulting reliability of 99.51% in case of RF for faults after TCSC. Similar situation with SVM provides 483 cases misdetection, providing a reliability of 91.83%. For faults before TCSC, the reliability becomes 99.47% with RF

TABLE XII
COMPARISON OF COMPUTATIONAL TIME BETWEEN RF AND SVM FOR TRAINING DATA SETS

Data sets	Processing time (RF), s	Processing time (SVM), s
TCSC	12.1	25.3
UPFC	12.5	28.9
TCSC+UPFC	18.8	48.5

TABLE XIII
COMPARISON OF PROCESSING TIME AND RESPONSE TIME BETWEEN RF AND SVM FOR EACH FAULT CASE

	RF	SVM
Processing time in ms	5	14
Response time in ms	15	24

compared to 96.27% for that of SVM. Similar reliability measures are observed for faults before and after UPFC in the transmission line.

It is observed from the aforementioned analysis that the reliability of fault-zone identification is much higher in RF case (more than 99%) compared to SVM. The reliability directly shows the misdetection from the desired class to another class. It is also found out from the analysis, that the reliability is highly uneven in case of SVM (91.83%, 96.27% for TCSC and 94.53%, 91.33% for UPFC). This shows that the misdetection rates for faults before and after TCSC/UPFC are not even with SVM. However, RF provides the reliability measure in all cases by more than 99%, and the misdetection cases are evenly distributed over each class. This shows the capability of RF to detect the fault zone with even and minimal misdetection rates for faults before and after TCSC/UPFC. Further, the testing results on combined data sets (TCSC+UPFC) are highly promising and, thus, result in a more generalized fault-zone identifier for the TCSC/UPFC-based transmission line using RF.

The computational time is measured in terms of processing time on a Core2-duo, 4-GB RAM desktop during training. The computational time of SVM is 25.3 s compared to 12.1 s of RF for the TCSC line during training. Similarly, the processing times are 28.9 and 12.5 s for SVM and RF, respectively, for the UPFC line. The computational time is very important since the data-mining model will be retrained time to time to include new contingencies if at all. A similar observation is made for mixed data sets as given in Table XII. For a particular case of faulty situation (each pattern on the test case), the response time of RF from fault inception is 15 ms (10 + 5 ms) and that of SVM is 24 ms (10 + 14 ms) for fault-zone identification. The response time includes half-cycle (10 ms) data samples from fault inception, used as inputs to RF or SVM and the processing time of RF or SVM for each test case, as depicted in Table XIII. From the aforementioned observation, it is found that the computational time as well as the response time of RF is lower compared to SVM for the fault-zone identification task.

VI. CONCLUSIONS

The proposed technique provides a data-mining model, such as ensemble decision trees (RFs), for fault-zone identification in a FACTS-based transmission line with an accuracy and reliability of more than 99%. RF, the data-mining algorithm, was found to be faster (3/4 cycles) and accurate compared to the existing machine-learning technique, such as SVM for fault-zone identification. The results indicate that the ensemble trees is highly effective and reliable in identifying the fault zone in the FACTS-based transmission line which triggers the next cascaded algorithms, such as an apparent impedance calculation for issuing the tripping signal in distance relaying.

REFERENCES

[1] A. A. Girgis, A. A. Sallam, and A. Karim El-din, "An adaptive protection scheme foe advanced series compensated (ASC) transmission line," *IEEE Trans. Power Del.*, vol. 13, no. 2, pp. 414–420, Apr. 1998.

[2] M. Noroozian, L. Angquist, M. Ghandhari, and G. Anderson, "Improving power system dynamics by series connected FACTS devices," *IEEE Trans. Power Del.*, vol. 12, no. 4, pp. 1635–1641, Oct. 1997.

[3] L. Gyugyi, "Unified power flow concept for flexible ac transmission systems," *Proc. Inst. Elect. Eng. C*, vol. 139, no. 4, pp. 323–332, 1992.

[4] X. Zhou, H. Wang, R. K. Aggarwal, and P. Beaumont, "Performance of evaluation of a distance relay as applied to a transmission system with UPFC," *IEEE Trans. Power Del.*, vol. 21, no. 3, pp. 1137–1147, Jul. 2006.

[5] K. EI-Arroudi, G. Joos, and D. T. Mcgills, "Operation of impedance protection relays with the STATCOM," *IEEE Trasn. Power Del.*, vol. 17, no. 2, pp. 381–387, Apr. 2002.

[6] M. Khederzadeh and T. S. Sidhu, "Impact of TCSC on the protection of transmission lines," *IEEE Trans. Power Del.*, vol. 21, no. 1, pp. 80–87, Jan. 2006.

[7] Y. H. Song, Q. Y. Xuan, and A. T. Johns, "Protection of scheme for EHV transmission systems with thyristor controlled series compensation using radial basis function neural networks," *Electric Mach. Power Syst.*, vol. 25, pp. 553–565, 1997.

[8] Y. H. Song, A. T. Johns, and Q. Y. Xuan, "Artificial neural network based protection scheme for controllable series-compensated EHV transmission lines," *Proc. Inst. Elect. Eng., Gen. Transm. Distrib.*, vol. 143, no. 6, pp. 535–540, 1996.

[9] U. B. Parikh, B. Das, and R. P. Maheshwari, "Combined wavelet-SVM technique for fault zone detection in a series compensated transmission line," *IEEE Trans. Power Del.*, vol. 23, no. 4, pp. 1789–1794, Oct. 2008.

[10] P. K. Dash, S. R. Samantaray, and G. Panda, "Fault classification and section identification of an advanced series-compensated transmission line using support vector machine," *IEEE Trans. Power Del.*, vol. 22, no. 1, pp. 67–73, Jan. 2007.

[11] S. R. Samantaray, L. N. Tripathy, and P. K. Dash, "Differential equation-based fault locator for unified power flow controller-based transmission line using synchronised phasor measurements," *IET Gen., Transm. Distrib.*, vol. 3, no. 1, pp. 86–98, 2009.

[12] H.-T. Yang and C.-C. Liao, "A de-noising scheme for enhancing Wavelet-based power quality monitoring system," *IEEE. Trans. Power Del.*, vol. 16, no. 3, pp. 353–360, Jul. 2001.

[13] L. Breiman, "Random forests," *Mach. Learn.* vol. 45, pp. 5–32, 2001. [Online]. Available: http://www.stat.berkeley.edu/users/breiman/RandomForests/

[14] A. Liaw and M. Wiener, "Classification and regression by random forest in R," *R News* vol. 2, no. 3, pp. 18–22, Dec. 2002. [Online]. Available: http://www.r-project.org/

[15] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. Berlin, Germany: Springer-Verlag, 2009, p. 745.

[16] F. J. Provost and P. Domingos, "Tree induction for probability-based ranking," *Mach. Learn.*, vol. 52, no. 3, pp. 199–215, 2003.

[17] R. E. Banfield, L. O. Hall, K. W. Bowyer, and W. P. Kegelmeyer, "A comparison of decision tree ensemble creation techniques," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 173–180, Nov. 2007.

[18] D. S. Siroky, "Navigating random forests and related advances in algorithmic modeling," *Stat. Rev.* vol. 3, pp. 147–163, 2009. [Online]. Available: http://projecteuclid.org/DPubS?service=UI&version=1.0&verb=Display&handle=euclid.ssu

[19] H. Liang, H. Zhang, and Y. Yan, "Decision trees for probability estimation: An empirical study," in *Proc. 18th IEEE Int. Conf. Tools with Artif. Intell.*, 2006, pp. 1–9.

[20] C. Strobl, J. Malley, and G. Tutz, An introduction to recursive partitioning: rationale, application and characteristics of classification and regression trees, Dept. Statistics, Univ. Munich, Munich, Germany, Tech. Rep. No. 55, Apr. 2009. [Online]. Available: http://epub.ub.uni-muenchen.de/10589/1/partitioning.pdf, on line

[21] T. Hastie, R. Tibshirani, and J. Friedman*, The Elements of Statistical Learning*, 2nd ed. Berlin, Germany: Springer-Verlag, 2009, p. 745.

[22] D. S. Siroky, "Navigating random forests and related advances in algorithmic modeling," *Stat. Rev.* vol. 3, pp. 147–163, 2009. [Online]. Available: http://projecteuclid.org/DPubS?service=UI&version=1.0&verb=Display&handle=euclid.ssu

S. R. Samantaray (M'08–SM'10) received the B.Tech. degree in electrical engineering from UCE Burla, India, in 1999 and the Ph.D. degree in power system engineering from the National Institute of Technology, Rourkela, India, in 2007.

Currently, he is Assistant Professor in the School of Electrical Sciences, Indian Institute of Technology Bhubaneswar, Orissa, India. He visited the Department of Electrical and Computer Engineering, McGill University, Montréal, QC, Canada, as a Postdoctoral Research Fellow and Visiting Professor. His major research interests include intelligent protection for transmission systems (including flexible ac transmission systms) and microgrid protection with distributed generation and dynamic security assessment in large power networks.

Prof. Samantaray is the recipient of the 2007 Orissa Bigyan Academy Young Scientists Award, the 2008 Indian National Academy of Engineering Best Ph.D. Thesis Award, the 2009 Institute of Engineers (India) Young Engineers Award, the 2010 Samanta Chandra Sekhar Award, and the 2012 IEEE PES Technical Committee Prize Paper Award. Dr. Samantaray is Editor of *IET, Generation, Transmission & Distribution* and *Electric Power Components and Systems*.