 h the exception of pagination.

A Combined Wavelet and Data-Mining Based Intelligent Protection Scheme for Microgrid

Debi Prasad Mishra, Subhransu Ranjan Samantaray, *Senior Member, IEEE*, and Geza Joos, *Fellow, IEEE*

***Abstract*—This paper presents an intelligent protection scheme for microgrid using combined wavelet transform and decision tree. The process starts at retrieving current signals at the relaying point and preprocessing through wavelet transform to derive effective features such as change in energy, entropy, and standard deviation using wavelet coefficients. Once the features are extracted against faulted and unfaulted situations for each-phase, the data set is built to train the decision tree (DT), which is validated on the unseen data set for fault detection in the microgrid. Further, the fault classification task is carried out by including the wavelet based features derived from sequence components along with the features derived from the current signals. The new data set is used to build the DT for fault detection and classification. Both the DTs are extensively tested on a large data set of 3860 samples and the test results indicate that the proposed relaying scheme can effectively protect the microgrid against faulty situations, including wide variations in operating conditions.**

***Index Terms*—Consortium for Electric Reliability Technology Solutions microgrid, change in energy, data-mining, decision tree (DT), microgrid protection, wavelet entropy, wavelet transform.**

I. Introduction

MICROGRIDS are emerging as an important part of power distribution system due to significant developments in distributed generation (DG) technology. Microgrids might include various types of DG such as microturbine, diesel generator, battery storage, fuel cells, and renewable energy resources. They also have their communication, control system, and protection devices [1], [2]. The presence of DG close to the loads makes the power delivery more reliable as it reduces power transmission loss and helps to improve the energy efficiency of the power system. A microgrid has the ability to operate independently in islanded mode in case there is any disturbance in the utility grid such as voltage fluctuation, frequency deviation, etc. It can supply the critical loads while operating independently in islanded mode. However, the use of various types of DG for power generation presents challenges for operating, controlling, and protecting the microgrid.

Manuscript received December 22, 2014; revised June 9, 2015 and August 14, 2015; accepted September 30, 2015. Paper no. TSG-01259-2014.

D. P. Mishra and G. Joos are with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC H3A 2A7, Canada (e-mail: debi.mishra@mail.mcgill.ca; geza.joos@mcgill.ca).

S. R. Samantaray is with the School of Electrical Sciences, Indian Institute of Technology Bhubaneswar, Bhubaneswar 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/TSG.2015.2487501

More specifically the protection challenges arise due to the high penetration of DG with power electronics interfaces in the microgrid. The inertia of inverter interfaced DG units (IIDGs) is lower as compared to synchronous machines. A microgrid consisting of low inertia sources might have stability issues in case line faults are not cleared fast. IIDGs operate with different types of control depending on the operating mode of the microgrid and the type (intermittent or nonintermittent) of power source connected with it. Typically, the maximum current contribution capacity of IIDGs is around two times of per unit rated current [3]. So, the fault current level for the microgrid operating in islanded mode of operation is considerably smaller as compared to the grid-connected mode for any type of fault as the IIDGs have limited current carrying capacity. Fault protection becomes challenging using conventional high fault current approach. The power flow in a microgrid is dynamic and bidirectional as the microgrid has DG and also exchanges power with the utility grid [4]. Protection problem gets complicated when the microgrid switches between mesh and radial topologies.

A protection scheme for the microgrid must address the problems related to bidirectional power flow and different levels of fault current in islanded and grid-connected mode. It should be fast and accurate in order to maintain the stability within the microgrid and to protect sensitive loads [2], [5]. Various methods have been proposed to protect microgrids [5]–[13]. In [5], a communication (wireless local area network protocol) assisted relaying technique has been proposed using microprocessor based relay for a medium voltage grid with IIDGs only. Most of the concepts used within this relay was originally proposed in [6] which uses over current method and negative sequence current component for relaying system inside a low voltage microgrid. Protection strategy using zero sequence component for single-line-to-ground faults and negative sequence component of the line current for line-to-line faults was proposed for microgrid [9].

In [8], a differential relaying technique has been proposed for a medium voltage microgrid containing synchronous generators and IIDGs. This method has been tested for balanced and unbalanced faults including high impedance faults. However, it does not take islanded mode operation into account and the proposed method has not been tested on IIDGs. The strategies proposed in [5] and [12], are independent of location of DGs, microgrid operational mode, microgrid size, and fault current level. However, the protection strategy needs communication. In [10], a centralized communication assisted over current based relaying technique, using IEC-61850 protocol,

 h the exception of pagination.

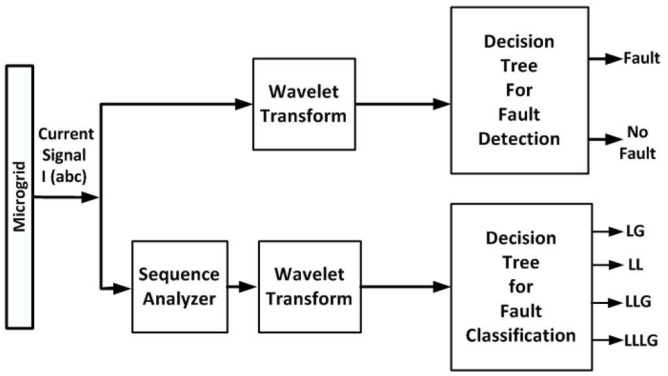


Fig. 1. Schematic of the proposed protection scheme.

has been proposed for the microgrid. It contains synchronous DGs and IIDGs in the test network. The centralized controller is connected to all DGs and relays in the network. The current value used in each relay is calculated dynamically using microgrid operational parameters and updated repeatedly.

A differential energy based protection strategy was proposed for the microgrid [13]. Balanced and unbalanced faults and high impedance faults were discussed in this paper. In [14], a protection scheme for only IIDG based low voltage microgrid was proposed and it uses current and voltage schemes to protect the grid. But, the relay settings need to be switched between grid-connected and islanded modes. In [15], a statistical classifier-based protection scheme has been proposed for an islanded microgrid. This paper concludes that the differential relaying technique is a better technique as compared to the technique using local measurements. It shows that the symmetrical components of differential current are better features for fault identification than the line current, voltage, and harmonic components of current. This strategy was tested for an islanded microgrid only.

Fig. 1 presents a schematic diagram of the proposed protection scheme using local current measurements to detect and classify faults by applying the wavelet transform and the decision tree (DT) model. Negative and zero sequence components of the measured current signal are calculated using the sequence analyzer. The protection scheme preprocesses the current data for one cycle using the wavelet transform and statistical features are computed. These features are used to build the DT models for fault detection and classification. The tripping signal is issued in case of a faulted condition to clear the fault. Section II describes the microgrid model used to test the protection scheme. Sections III and IV describe the wavelet based preprocessing and data-mining modeling process. Sections V and VI summarize the result and analysis the performance of various data-mining models.

II. System Studied

A new protection strategy using intelligent relaying technique has been proposed for a microgrid based on the topology of Consortium for Electric Reliability Technology Solutions (CERTS) microgrid model [16]. This microgrid is a part of the power distribution system and operates at 480 V, 60 Hz. It is connected to the utility grid using a step up

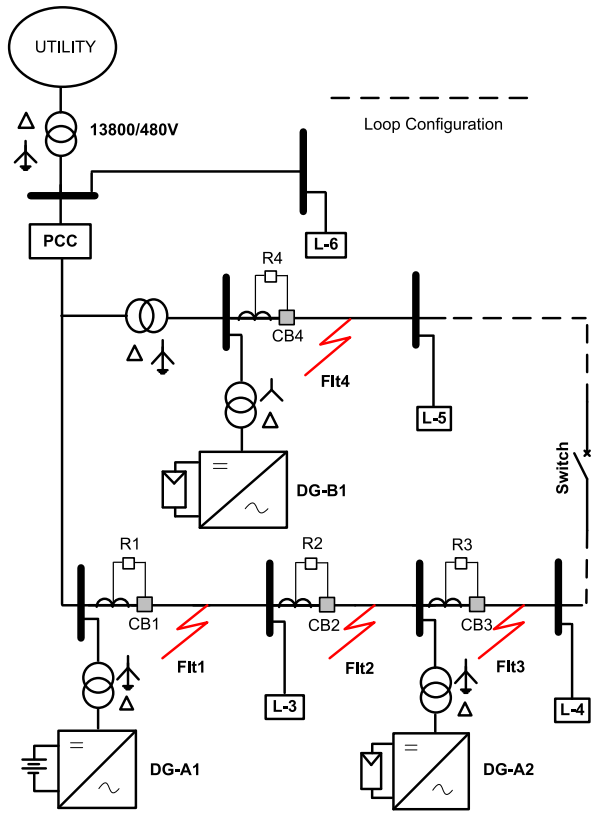


Fig. 2. CERTS microgrid single line diagram.

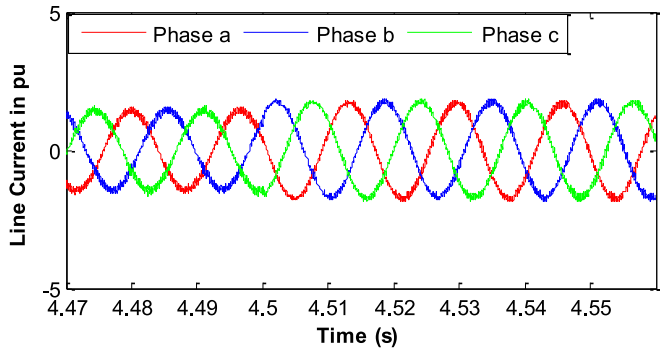
TABLE I
Loads Connected in the Microgrid

Load Name	Resistive Load in KW	Inductive Load in KVAR	Capacitive Load in KVAR
L-3	90	45	0
L-4	90	45	0
L-5	90	0	40
L-6	90	0	20

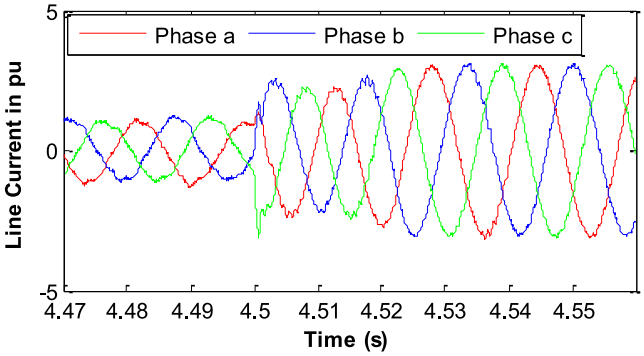
transformer 13 800/480 V. There are three DGs for supplying power within the microgrid. One battery-based storage source and two photovoltaic sources are present in this microgrid as per Fig. 2 (CB represents the circuit breaker and R represents the relay). DG-A2 and DG-B1 represent photovoltaic sources. DG-A1 is a battery storage based source and it is responsible for supplying additional power required during the islanded mode. Each source supplies 100 kW power with unity power factor when the microgrid operates in grid-connected mode. The switch at point of common coupling operates to island the microgrid. Information regarding the loads used for the simulation of the microgrid is shown in Table I. This grid was simulated using MATLAB/Simulink.

Each photovoltaic source is connected to the microgrid uses an inverter employing grid-frequency imposed current control method and the control method for these inverters remains unchanged whether the microgrid operates in grid-connected or islanded mode. The phase locked loop helps the current controlled IIDGs to follow the grid frequency. The current controlled voltage source inverter (VSI) was modelled using proportional-integral (PI) control in dq-frame [17]. The battery storage source is connected to an inverter which operates

 h the exception of pagination.



(a)



(b)

Fig. 3. (a) Fault current by DG-B1 in current controlled mode at Flt4 for an ABCG fault in islanded mode. (b) Fault current by DG-A1 in controlled frequency mode at Flt4 for an ABCG fault in islanded mode.

as a current controlled inverter in grid-connected mode, but the control strategy is switched to controlled frequency voltage source inverter when the microgrid operates as an islanded grid. The PI control in dq-frame was implemented for simulating controlled frequency VSI model with battery based power supply [22]. DG-A1 is responsible for controlling frequency and maintaining the voltage level in the microgrid when it operates in an islanded mode. It supplies all the reactive power needed in the islanded mode.

Maximum current contribution by the current controlled inverters has been restricted to 1.2 times of the rated current. The maximum current contributed by the controlled frequency VSI is 2.1 times of the per unit rated current of the inverter. In islanded mode, the fault current contribution by the two types of VSIs used for this paper is shown in Fig. 3(a) and (b) for ABCG fault at Flt4. The fault current contribution was measured at the bus where the sources are connected. It is clear that the fault current contribution by the current controlled inverter is much less than voltage controlled inverter. The fault starts at 4.5 s at Flt4.

The fault current from contributed by the current controlled inverter is 1.2 times of the prefault current. Similarly, the fault current from the frequency controlled inverter is approximately two times of the prefault current value.

III. Wavelet Transform

Wavelet transform represents the signal in time–frequency domain and provides the time localization of the signal [18], [19]. The wavelet transform is a powerful tool for the analysis of nonstationary signals as the frequency content of a nonstationary (transient) signal changes with time. The wavelet transform is a better tool than the Fourier transform as the latter is used for analyzing a signal in frequency domain only. Fourier transform does not give any information related to time–frequency domain. Short time Fourier transform (STFT) has limitation in resolution due to fixed window size. Wavelet transform is a better tool as compared to Fourier transform or STFT, providing both time and frequency information which is required to extract transient information from the nonstationary signals such as fault current.

A. Continuous Wavelet Transform

The continuous wavelet transform (CWT) can be defined as follows:

$$X_{a,b}(t) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\psi\left(\frac{t-b}{a}\right)dt \tag{1}$$

where, $x(t)$ is the signal and $\psi(t)$ is the mother wavelet. The translated and dilated version of the wavelet is given by

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)$$

where, a and b represent the dilated and translated parameters, respectively. The translation and dilation are responsible for time and frequency resolution.

B. Discrete Wavelet Transform

The practical application of wavelet transform is performed using the discrete wavelet transform [20]. Any time series signal $x(t)$ can be completely decomposed into approximations by using a scaling function $\phi_j(t)$ which is also known as the father wavelet. The signal can be decomposed into detailed coefficients using mother wavelet $\psi_j(t)$

$$\phi_{jk}(t) = 2^{-j/2}\phi\left(2^{-j}t - n\right) \tag{2}$$

$$\psi_{jk}(t) = 2^{-j/2}\psi\left(2^{-j}t - n\right) \tag{3}$$

where $n \in Z$, j and k are integers.

As j and k are integers and the basis function is scaled by a factor 2^j and translated by n units of time. The scaling function is associated with the low pass filters with filter coefficients $H = \{hn\}$ and the wavelet function is associated with the high pass filter with filter coefficient $G = \{gn\}$. It is important to note that G is reverse of H interleaved with sign changes. The two scale functions give rise to filters

$$\phi(t) = \sum_n h(n)\sqrt{2}\phi(2t-n) \tag{4}$$

$$\psi(t) = \sum_n g(n)\sqrt{2}\phi(2t-n). \tag{5}$$

A signal $x(t)$ having length 2^M, then there are maximum "M" levels of decomposition. The expression for the signal using wavelet transform is given as follows:

$$x(t) = \sum_{k=0}^{2^{M-j-1}} a_{jk}\phi_{jk}(t) + \sum_j \sum_{k=0}^{2^{M-j-1}} d_{jk}\psi_{jk}(t) \tag{6}$$

 h the exception of pagination.

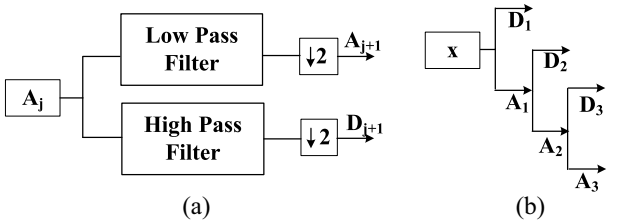


Fig. 4. (a) Single decomposition step. (b) Decomposition upto level 3 of a signal "x".

where a_{jk} and d_{jk} are known as the approximation and detail coefficients of the signal at level j, respectively. The signal is passed through high- and low-pass filters. Then the outputs of both the filters are down-sampled by a factor of 2. The approximation and detailed coefficients for level-1 A1 and D1 are obtained as shown in Fig. 4. This process is repeated for the approximation coefficient till the desired level of detailed coefficient is retrieved.

C. Feature Extraction

The proposed method preprocesses the current signal through the wavelet transform and extracts most effective statistical features. These features are used to build the data-mining model for fault detection and classification. Thus, the proposed scheme computes energy, entropy, and standard deviation of the wavelet coefficients which contains vital information during transient events such as faults. The details of the statistical features are as follows.

1) Energy of Signal: Using the wavelet coefficient, the formula for energy content of a signal can be written as follows [20]:

$$E_i = \sum_{j=1}^{n} |d_{ij}|^2 \tag{7}$$

where, $i = 1, 2, \ldots, I$ stands for the scale and $j = 1, 2, \ldots, n$ denotes the number of points used for each coefficient. So, change in energy using the wavelet coefficient shows the change in energy of a signal due to some disturbance. Change in energy is calculated by subtracting the energy of the signal between two successive cycles

$$\text{Change in Energy} = E_B - E_A \tag{8}$$

where, E_B = energy of the latest cycle and E_A = energy of last cycle.

2) Shannon Entropy: Shannon entropy is a measure of signal uncertainty (disorderness) and complexity. The uncertainty is associated with the signal represented by the states and probability. Power system faults are random events as the fault parameters such as fault location, time, fault impedance, network topology, angle of fault inception, etc. are unpredictable [21]. Moreover, the transient behavior of the fault signal is random in nature. Shannon entropy is used to map the correlative wavelet space into independent linearity space, and to indicate the uncertainty of the energy distribution in time–frequency domain [22]. The formula for calculating Shannon's entropy is given [23]

$$E_i = -\sum_{j=1}^{n} p_j \log p_j \tag{9}$$

where p_j is the energy probability distribution of the wavelet coefficients and $\sum_{j=1}^{n} p_j = 1$. $p_j \log p_j = 0$ if $p_j = 0$.

3) Standard Deviation: Standard deviation is a statistical measure of the disturbance present in a signal. It measures the dispersion of the data from the average. The standard deviation for a series of data points is calculated as follows:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{10}$$

where $\mu = (1/N)\sum_{i=1}^{N} x_i$ and x_i represents the ith data element of a data series and $i = 1, \ldots, N$; σ, μ represents standard deviation and mean of the data series.

TABLE II
FREQUENCY RANGE FOR EACH WAVELET COEFFICIENT

Wavelet Coefficient	Frequency Range in kHz
d1	1.66-3.33
d2	0.833-1.66
d3	**0.416-0.833**
d4	0.208-0.416
d5	0.104-0.208
d6	0.052-0.0104

D. Wavelet Based Feature Extraction

The line current data is sampled at 6.66 kHz. The measurement was taken using the per unit values. The base values used for per unit measurement were 480 V, 100 kVA. Then line current data is processed using the "haar" wavelet to generate various coefficients. Further, different features such as "change in energy," "Shannon entropy," and "standard deviation" are derived using the wavelet coefficients which are used as inputs to the data-mining model for the final decision.

1) Selection of Wavelet Coefficient "d3": There were six detailed coefficients ($d1$–$d6$) obtained after analyzing the input signal using the wavelet transform. Information related to the frequency range for all the detailed coefficients obtained using the wavelet transform has been mentioned in Table II. A comparison of the wavelet detailed coefficients of the fault and no-fault conditions are shown in Fig. 5. It is observed that the detailed coefficient at level-3 ($d3$) provides the most distinct deviation during fault condition. The $d3$ coefficient is more distinct as it contains 7th, 11th, and 13th harmonics which are pronounced during the fault.

2) Features Using Wavelet Coefficient d3: Fifteen different features were calculated using the wavelet coefficient $d3$. The wavelet coefficients of the line currents and sequence components were used in the features calculation process. All the features are independent of each other. The description of each feature is as follows.

1) *F1*—change in energy of phase A (post fault–prefault for one cycle).

 h the exception of pagination.

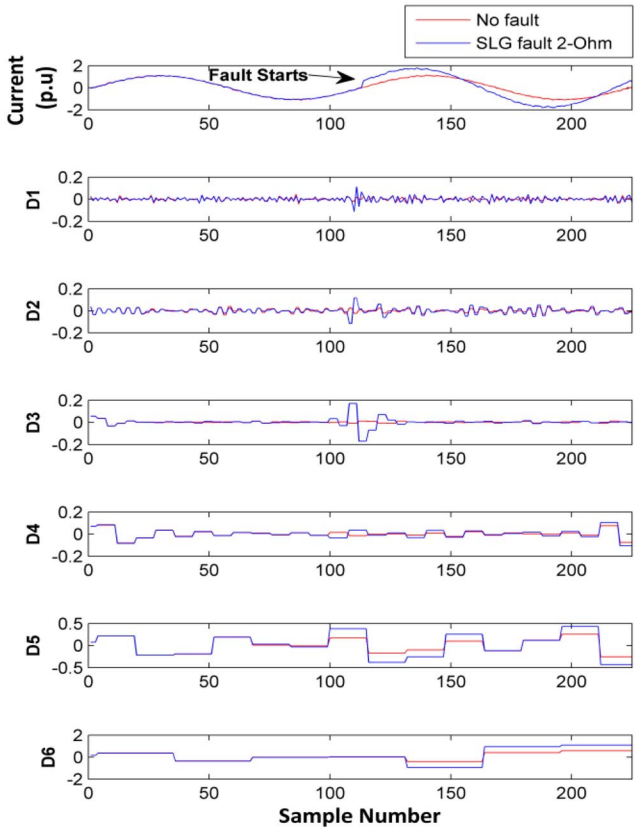


Fig. 5. Comparison between the wavelet transform coefficients generated for the two fault currents.

2) *F2*—change in energy of phase B (post fault–prefault for one cycle).
3) *F3*—change in energy of phase C (post fault–prefault for one cycle).
4) *F4*—entropy of phase A one cycle post fault.
5) *F5*—entropy of phase B one cycle post fault.
6) *F6*—entropy of phase C one cycle post fault.
7) *F7*—standard deviation of phase A after the fault occurs for one cycle.
8) *F8*—standard deviation of phase B after the fault occurs for one cycle.
9) *F9*—standard deviation of phase C after the fault occurs for one cycle.
10) *F10*—change in energy of negative sequence component (post fault–pre fault for one cycle).
11) *F11*—entropy of negative sequence after the fault occurs for one cycle.
12) *F12*—standard deviation of negative sequence after the fault occurs for one cycle.
13) *F13*—change in energy of zero sequence components.
14) *F14*—entropy of zero sequence after the fault occurs for one cycle.
15) *F15*—standard deviation of zero sequence after the fault occurs for one cycle.

IV. Data-Mining Models for Decision Making

Data mining models aim to create a comprehensible structure by analyzing a dataset which can be further used for predicting or understanding the system behavior. Data mining is a tool to model the relationship or dependencies between system variables of a dataset containing data pertaining to the system. The models created by data mining are used for prediction and description of system outputs [24]–[26]. Predictive data mining produces a system model as per the input dataset and this model can be used for classification, estimation, prediction, etc. Descriptive data mining aims to discover patterns and relationships in large data sets. The data mining method used for the design of intelligent relay falls into the predictive category i.e., classification. There are different methods for data mining such as DT, random forest (RF), support vector machine, and neural networks etc. The proposed study uses two effective data-mining models. The transparent data-mining model called as DT and a black-box model known as RF. These two models belong to the same family. However, there is a tradeoff between transparency and accuracy between two data-mining models which are discussed in Section V.

The objective of the intelligent relay proposed in this paper is to detect and classify faults. The data mining method would be used to build a classifier which will predict the output value for a set of input features. To build a classifier model, a dataset consisting of system data is assigned to different type of groups as defined by the user and fed to a data mining algorithm to produce a classifier model. Each group of input data is known as a class. The next few paragraphs discuss regarding two data mining methods used for classification.

DT is a highly popular and efficient way to build classifiers. The DT algorithms are based upon supervised learning algorithms. It resembles an inverted tree structure and works on a series of conditional (if-then) logic. A DT consists of "decision node" where the input attributes are evaluated. The first node is referred as "root node." The "branches" originating from the nodes correspond to the outcomes of the decision node. The branches may lead to a class which is known as "leaf" node or to another decision node which is the root for another sub tree within the DT. Thus, every "nonleaf" or a decision node has two branches and a leaf node has no branch. Any node in the DT except leaf node and root node is known as "internal node." DT is also known as binary tree if every nonleaf node has two branches. A path starting from root node and ending at a leaf node represents the set of evaluations done to identifying a class. These sets of evaluations can be grouped to create a classification rule. DTs can be converted to a set of classification rules. DTs are easy to implement and transparent in nature as it shows the possible alternatives for any decision node.

RF is an ensemble learning method to increase the classification accuracy [25]–[27]. It is created by combining many classifiers. It uses a collection of DTs acting on same set of input attributes. The DTs used for the RF method are independent of each other. The individual classifiers vote for a class and the most popular class is selected. The output of a single tree might be noisy. Therefore, the RF method counts on an average decision using more number of DTs.

 h the exception of pagination.

TABLE III
PARAMETERS FOR SIMULATING FAULT CONDITIONS

Mode of Operation	Grid Connected	Islanded
Topology	2	2
Types of Faults	10	10
Fault Resistance	3	3
Fault Location	4	4
Fault Inception Angle	4	4
DG Penetration	3	1
Total Conditions	2880	960

TABLE IV
PARAMETERS FOR SIMULATING NO FAULT CONDITIONS

Mode of Operation	Grid Connected	Islanded
Topology	2	2
Types of Load	2	2
Measurement Locations	4	4
DG Penetration	3	1
Total Conditions	48	16

V. PERFORMANCE RESULTS AND ANALYSIS

A. Data Count Generation

Fault data was generated by simulating different types of conditions such as by varying the fault resistance, switching between different fault locations, changing fault current inception angle, and various levels of DG penetration within different topologies of microgrid. The faults were simulated at the middle of the distribution line between two buses and the current is measured using the current transformer (CT). A summary of the fault conditions for the proposed study is shown in Table III. All the conditions for the fault study are explained as follows.

1) Different types of balanced and unbalanced faults were simulated. They can be referred as AG, BG, CG, AB, BC, CA, ABG, BCG, CAG, and ABCG (ten types).
2) Faults resistances 0.01, 0.5, and 2 Ω were used for all the faults simulated (three cases).
3) Faults are simulated in the grid at location Flt1, Flt2, Flt3, and Flt4 as per Fig. 2 (all the faults are simulated at the mid-point on the distribution line between two adjacent buses).
4) DG penetration levels used for the simulation are 100%, 40%, and 10% for simulating every fault with microgrid operating in grid-connected mode. In islanded mode, the IIDGs operate at 100% of its capacity.
5) Fault inception angles used for the simulation study are as follows: a) 0°; b) 45°; c) 90°, and d) 180°(four cases).
6) Topology: radial and loop.

The line current measurements for no fault conditions were taken under heavy load and light loading conditions within the microgrid. Three loads within the microgrid were changed for adjusting the total microgrid load. Those three loads were loads 3–5. During the heavy load condition within the microgrid, the microgrid load is: 2 × [90 kw, 45 kvar (inductive load)] + [90 kw, 40 kvar (capacitive load)]. The description for lightly loaded condition is: 2 × [20 kw, 5 kvar (inductive load)] + [20 kw, 5 kvar (capacitive load)]. The DG penetration was varied in exactly same manner as it was done for faulted case. Microgrid operates with both radial and loop topologies for these simulations. Table IV presents a summary of the conditions for the no-fault situations.

The line current signals were collected and processed using the wavelet transform. Fifteen different features which were generated using the detail coefficient from the wavelet transform will be used by the data mining algorithms to build the DT. We have used "R" programming software to build the DT for the intelligent relay [28]. The intelligent relay uses one DT model for fault detection and another one for fault classification.

Nine features related to line currents were given as inputs along with the class to build a DT model for fault detection. The classes used to build detection DT are fault and no-fault. For the training a DT model to classify a fault, 15 features related to the line current were given as inputs along with the class of fault. Ten types of faults mentioned for this paper were classified into four classes: 1) the single line to ground faults were bundled into one class "LG"; 2) the line–line faults were bundled into class "LL"; 3) the line-to-line-to ground faults were classified as class "LLG"; and 4) three-phase-to-ground fault were classified as class "LLLG." The dataset having features related to fault data contains 3840 rows for fault measurement and 16 columns which include 15 columns for features and one column for class.

TABLE V
FEATURES USED FOR TRAINING THE DTS

Model Objective	No of Input Features	Name of the Input Features
Fault Detection	9	F1, F2, F3, F4, F5, F6, F7, F8, F9
Fault Classification	15	F1, F2, F3, F4, F5, F6, F7, F8, F9, F10, F11, F12, F13, F14, F15

TABLE VI
ERROR MATRIX (%) FOR FAULT DETECTION DT VERSUS RF

		Predicted Class using DT		Predicted Class using RF	
		Fault	No-Fault	Fault	No-Fault
Actual Class	Fault	96	1	97	0
	No-Fault	2	1	1	2

TABLE VII
SUMMARY FOR THE FAULT DETECTION

Data Mining Method	Decision Tree	Random Forest
Accuracy %	97	99
Misdetection %	1	0
False Alarm %	2	1

 h the exception of pagination.

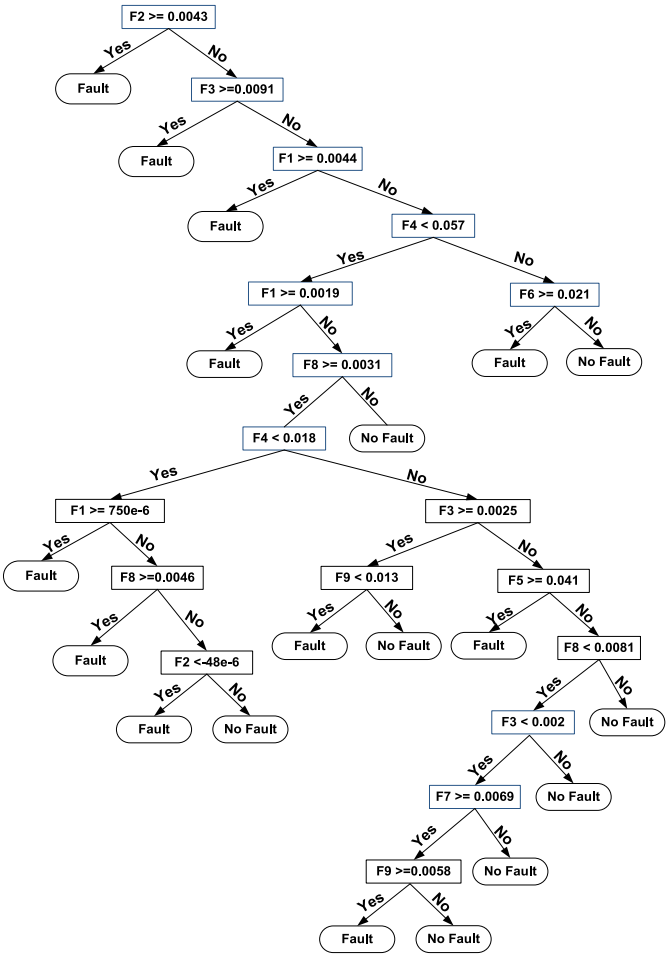


Fig. 6. DT for fault detection.

B. DT for Fault Detection

The datasets used for building DT for fault detection and classification are also used for building the RF model. Each dataset used for building data mining models is partitioned into two parts. First part of the dataset consists of 70% of the data and was used for training the DT. The second part consists of rest 30% of the data and it was used for testing the DT model. The DT was trained for fault detection using all the nine input features (F1–F9 for all conditions) related to phase current components which are grouped into a fault class and a no-fault class. Fig. 6 shows the trained DT model for fault detection. Tables VI and VII provide the performance matrix for fault detection. It is found that the DT provides 96% of predicted class against 97% resulted using the RF. While observing the overall accuracy, it is found that DT results 97% for fault detection as compared to 99% by RF.

C. DT for Fault Classification

All the 15 features used to train the fault classification DT model have been shown in Table V. The DT model generated for fault classification uses five input features (F10–F14) and four output classes (LG, LL, LLG, and LLLG). The training and testing datasets were created using the rule followed for building the DT for detection. The features selected by the trained DT are the features related to symmetrical components. Those features were F10–F14. Fig. 7 represents the trained DT model used for fault classification. Table VIII compares the performance of the DT with RF for the fault classification. It is observed that the accuracy is 85% in case of DT compared to 94% that of RF. Table IX depicts the summary of classification of faults.

TABLE VIII
ERROR MATRIX (%) FOR FAULT CLASSIFICATION (DT VERSUS RF)

		Predicted Class by DT				Predicted Class by RF			
		LG	LL	LLG	LLLG	LG	LL	LLG	LLLG
Actual Class	LG	28	0	3	0	28	0	2	0
	LL	0	29	0	0	0	29	0	0
	LLG	12	0	18	0	4	0	26	0
	LLLG	0	0	0	10	0	0	0	11

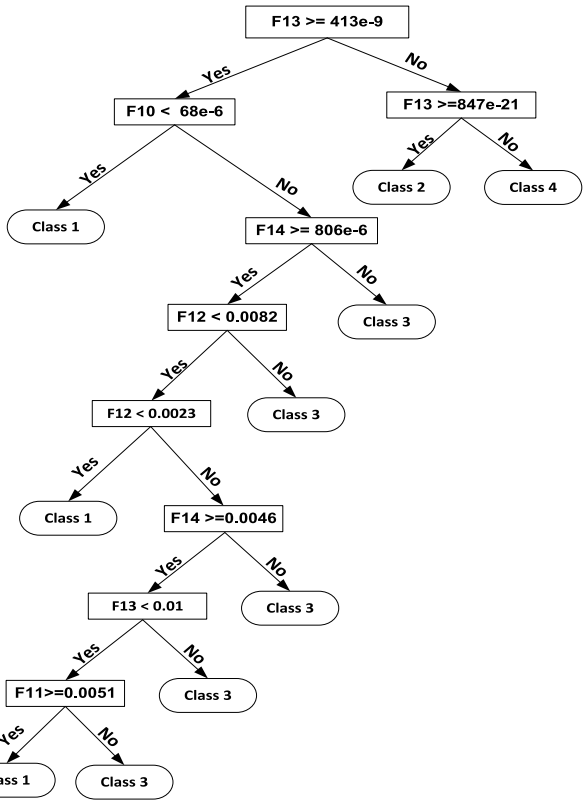


Fig. 7. DT for fault classification.

TABLE IX
SUMMARY FOR CLASSIFICATION OF FAULTS

Data Mining Method	Accuracy in %	Misdetection in %
Decision Tree	85	15
Random Forest	94	6

D. Comparison With Existing Relay

A comparison between the performance of the existing relays and the proposed intelligent relay is done for all the conditions shown in Table III. The threshold setting for the

 h the exception of pagination.

TABLE X
OVER CURRENT RELAY VERSUS INTELLIGENT RELAY FOR FAULT DETECTION

Data Mining Method	Accuracy %	Misdetection %	False Alarm %
Decision Tree	97	1	2
Random Forest	99	0	1
Over Current Relay	56	44	-
Differential Relay	96	4	-

over current relay was set at 1.5 times of the maximum load current value in the fully loaded microgrid and it detects a fault when the fault current exceeds the specified threshold. The performance of the over current relay substantially degraded as the accuracy stays at 56% as depicted in Table X. Further, the current differential relay has been compared which considers differential current between the both ends of the feeder subjected to fault. The threshold for differential relay is set according to the standard setting [4] considering CT mismatch error and other interferences. Though the accuracy of the current differential relay is high at 96%, however, the differential relay requires mandatory communication infrastructure and thereby increasing the investment substantially. Also, the reliability of the protection measure depends upon the reliability of the communication system.

It is observed from Table X that the instantaneous over current relay reports 44% misdetection yielding a low level of security to the microgrid followed by 4% that of differential relay. This problem can be alleviated by using the proposed intelligent relays as the fault detection accuracy improves significantly. The misdetection reduces to 1% for the DT based relay and 0% percent for the RF based relay. False alarm in case of the DT is 2% as compared to 1% for the RF model. It should be noted that misdetection is dangerous as this is a fault case and detected as no-fault and thus must be as low as possible to improve the reliability. While false alarm predicts no-fault as fault which is not that dangerous, however, may trigger unnecessary action. Thus, both the indices must be as minimal as possible for an effective relaying scheme.

During the fault classification, problems arise when LG faults are classified as LLG and vice-versa. DT model classifies 3% LG faults as LLG faults and 12% LLG faults as LG faults. The RF based relay has 9% higher fault classification accuracy than the DT based relay as per Table IX. The proposed data-mining based relays are able to classify the faults involving ground and faults not involving ground accurately. As the proposed intelligent relaying scheme is based on distance relaying concept, thus the complete protection unit including CT, relay, and CB are shown at one end in each line (Fig. 2). However, similar set of protection units is also required to be placed at other end of the line for providing protection from both ends.

The proposed intelligent relays use one cycle post fault inception data and perform the required computations. Thus, the response time including wavelet based preprocessing, feature computation time, and the DT processing takes one and half cycle (for fault detection) from fault inception which is well accepted for relaying task. The response time for fault classification is two and half cycles from fault inception as one cycle more is required for sequence analyzer. It is observed that the RF based relay outperforms the DT based model for fault detection and classification. However, the RF being a black box data-mining technique faces serious challenges for real implementation on digital signal processor or field programmable gate array boards while the DT being transparent techniques finds easy implementation.

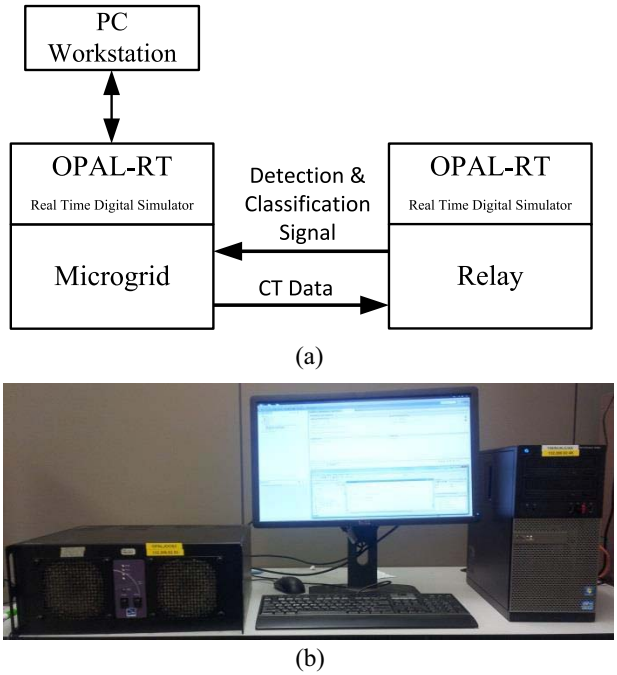


Fig. 8. (a) Schematic of the OPAL-RT test bed. (b) OPAL-RT module with work station.

VI. REAL-TIME IMPLEMENTATION

A real time software in loop (SIL) test bed was built to validate the proposed relaying technique using OPAL-RT digital simulator. The OPAL-RT is a real time digital simulation platform which consists of multiple multicore processors used for parallel processing. It uses RT-LAB software environment and is integrated with MATLAB/Simulink. The results from the real time digital environment is accessed using a PC connected to the OPAL-RT server. Power system models need simulations of many power sources, network elements, and switching events. The power system is broken in different subsystems on RT-LAB. Every subsystem is allocated one core to perform real time simulation. Therefore, parallel processing capability enables OPAL-RT to perform the real time simulation of the power system model. A schematic of the OPAL-RT implementation test bed and the OPAL-RT set up are shown in Fig. 8(a) and (b), respectively.

The real time simulator implements the same microgrid as shown in Section II. The relay was implemented as a separate subsystem of the microgrid on the OPAL-RT platform. It operates as SIL mode on this test bed. The sampling rate selected for this implementation is 6.66 kHz. All the sampled current

 h the exception of pagination.

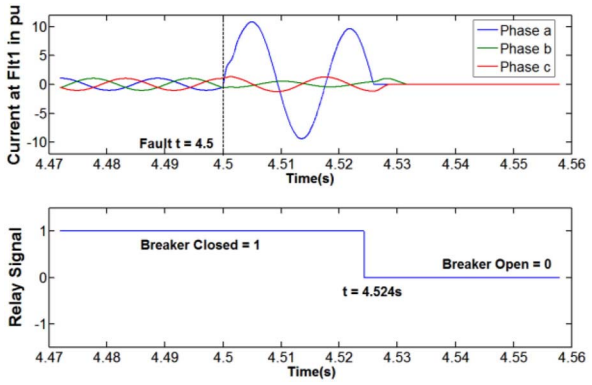


Fig. 9. Current and relay tripping signal for an AG fault (0.1 Ω fault resistance) at Flt1 on OPAL-RT.

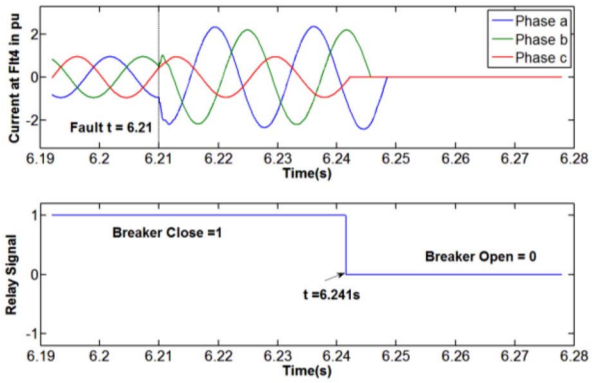


Fig. 10. Current and relay tripping signal for an AB-G fault (2 Ω fault resistance) at Flt1 on OPAL-RT.

signal data for one cycle is then transferred to the relay unit at the end of every cycle. This data is processed to generate features F1–F14 which are evaluated by the DT to produce the relay tripping signal. The relay tripping signal triggers the CB associated with the relay in the microgrid to clear the fault.

The performance testing for AG fault (0.1 Ω fault resistance) at Flt1 at time = 4.5 s was carried out on the OPAL-RT platform. It is found out that the relay tripping signal is issued at $t = 4.524$ s and thus, the response time is within two cycles (24 ms) from the fault inception. The CB1 associated with the relay operates to disconnect the fault section. The pre and post fault current with relay tripping signal is shown in Fig. 9. Similar observation is made for ABG fault situation (higher fault resistance) at Flt4 as shown in Fig. 10. The performance on the OPAL-RT platform clearly indicates on the ability of the proposed intelligent relaying scheme in providing accurate and reliable protection measure for microgrids with DG penetration.

VII. Conclusion

The proposed research develops a new protection scheme for microgrids with IIDGs using the wavelet transform and data-mining model. Initially the current is preprocessed to extract most effective statistical features which contain the transient information. Further, the wavelet based features are used to build the data-mining models for final relaying decision. The proposed DT model provides significantly improved performance over existing over current relays. Even though another data-mining model RF provides similar performance like the DT, however, being a black-box solution, the RF faces implementation difficulty as compared to the transparent data-mining model DT. The most important issue is the use of time–frequency information for building the data-mining models which improves the dependability and reliability of the relay. As the wavelet transform is very fast, thus the response time for fault detection and classification falls within 1.5 cycles and 2.5 cycles, respectively, from fault inception which is well accepted for the protection task. The test results of the validation on OPAL-RT platform indicate the potential ability of the proposed intelligent relay for microgrid protection.

References

[1] R. H. Lasseter *et al.*, "The CERTS microgrid concept," White Paper for Transmission Reliability Program Office of Power Technologies, U.S. Department of Energy, April 2002.

[2] H. J. Laaksonen, "Protection principles for future microgrids," *IEEE Trans. Power Electron.*, vol. 25, no. 12, pp. 2910–2918, Dec. 2010.

[3] J. Keller and B. D. Kroposki, *Understanding Fault Characteristics of Inverter-Based Distributed Energy Resources*, Nat. Renew. Energy Lab., Golden, CO, USA, Tech. Rep. NREL/TP-550-46698, 2010.

[4] M. Dewadasa, A. Ghosh, and G. Ledwich, "Protection of microgrids using differential relays," in *Proc. 21st Aust. Univ. Power Eng. Conf. (AUPEC)*, Brisbane, QLD, Australia, 2011, pp. 1–6.

[5] M. A. Zamani, A. Yazdani, and T. S. Sidhu, "A communication-assisted protection strategy for inverter-based medium-voltage microgrids," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 2088–2099, Dec. 2012.

[6] M. A. Zamani, T. S. Sidhu, and A. Yazdani, "A protection strategy and microprocessor-based relay for low-voltage microgrids," *IEEE Trans. Power Del.*, vol. 26, no. 3, pp. 1873–1883, Jul. 2011.

[7] P. Mahat, C. Zhe, B. Bak-Jensen, and C. L. Bak, "A simple adaptive overcurrent protection of distribution systems with distributed generation," *IEEE Trans. Smart Grid*, vol. 2, no. 3, pp. 428–437, Sep. 2011.

[8] E. Sortomme, S. S. Venkata, and J. Mitra, "Microgrid protection using communication-assisted digital relays," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Minneapolis, MN, USA, 2010, p. 1.

[9] H. Nikkhajoei and R. H. Lasseter, "Microgrid protection," in *Proc. IEEE Power Eng. Soc. Gen. Meeting*, Tampa, FL, USA, 2007, pp. 1–6.

[10] T. S. Ustun, C. Ozansoy, and A. Zayegh, "Modeling of a centralized microgrid protection system and distributed energy resources according to IEC 61850-7-420," *IEEE Trans. Power Del.*, vol. 27, no. 3, pp. 1560–1567, Aug. 2012.

[11] T. S. Ustun, C. Ozansoy, and A. Ustun, "Fault current coefficient and time delay assignment for microgrid protection system with central protection unit," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 598–606, May 2013.

[12] M. A. Zamani, T. S. Sidhu, and A. Yazdani, "A communication-based strategy for protection of microgrids with looped configuration," *Elect. Power Syst. Res.*, vol. 104, pp. 52–61, Nov. 2013.

[13] S. R. Samantaray, G. Joos, and I. Kamwa, "Differential energy based microgrid protection against fault conditions," in *Proc. IEEE PES Innov. Smart Grid Technol. (ISGT)*, Washington, DC, USA, 2012, pp. 1–7.

[14] M. A. Haj-ahmed and M. S. Illindala, "The influence of inverter-based DGs and their controllers on distribution network protection," in *Proc. IEEE Ind. Appl. Soc. Annu. Meeting*, Lake Buena Vista, FL, USA, 2013, pp. 1–9.

[15] E. Casagrande, W. W. Lee, H. H. Zeineldin, and N. H. Kan'an, "Data mining approach to fault detection for isolated inverter-based microgrids," *IET Gener. Transm. Distrib.*, vol. 7, no. 7, pp. 745–754, Jul. 2013.

[16] P. Piagi and R. H. Lasseter, "Autonomous control of microgrids," in *Proc. IEEE Power Eng. Soc. Gen. Meeting*, Montreal, QC, Canada, 2006, p. 8.

[17] A. Yazdani and R. Iravani, *Voltage-Sourced Converters in Power Systems: Modeling, Control, and Applications*. Hoboken, NJ, USA: Wiley, 2010.

 h the exception of pagination.

[18] S. G. Mallat, *A Wavelet Tour of Signal Processing*. San Diego, CA, USA: Academic press, 1999.
[19] W. Li, A. Monti, and F. Ponci, "Fault detection and classification in medium voltage DC shipboard power systems with wavelets and artificial neural networks," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 11, pp. 2651–2665, Nov. 2014.
[20] W. Gao and J. Ning, "Wavelet-based disturbance analysis for power system wide-area monitoring," *IEEE Trans. Smart Grid*, vol. 2, no. 1, pp. 121–130, Mar. 2011.
[21] A. G. Phadke and J. S. Thorp, *Computer Relaying for Power Systems*. Hoboken, NJ, USA: Wiley, 2009.
[22] A. Samui and S. R. Samantaray, "Wavelet singular entropy-based islanding detection in distributed generation," *IEEE Trans. Power Del.*, vol. 28, no. 1, pp. 411–418, Jan. 2013.
[23] Z. Liu, Z. Han, Y. Zhang, and Q. Zhang, "Multiwavelet packet entropy and its application in transmission line fault recognition and classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 11, pp. 2043–2052, Nov. 2014.
[24] S. R. Samantaray, "A data-mining model for protection of FACTS-based transmission line," *IEEE Trans. Power Del.*, vol. 28, no. 2, pp. 612–618, Apr. 2013.
[25] I. Kamwa, S. R. Samantaray, and G. Joos, "On the accuracy versus transparency trade-off of data-mining models for fast-response PMU-based catastrophe predictors," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 152–161, Mar. 2012.
[26] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Amsterdam, The Netherlands: Elsevier, 2011.
[27] S. R. Samantaray, I. Kamwa, and G. Joos, "Ensemble decision trees for phasor measurement unit-based wide-area security assessment in the operations time frame," *IET Gener. Transm. Distrib.*, vol. 4, no. 12, pp. 1334–1348, Dec. 2010.
[28] G. J. Williams, "Rattle: A data mining GUI for R," *Graham J Williams, The R Journal*, vol. 1, no. 2, pp. 45–55, Dec. 2009.

Debi Prasad Mishra received the Master's degree in electrical engineering from McGill University, Montreal, QC, Canada, in 2015, where he is currently pursuing the Ph.D. degree.

His current research interests include developing intelligent and adaptive relaying schemes for microgrids using combined signal processing and data-mining framework.

Subhransu Ranjan Samantaray (M'08–SM'10) received the B.Tech. degree in electrical engineering from the University College of Engineering Burla, Burla, India, in 1999, and the Ph.D. degree in power system engineering from the Department of Electronics and Communication Engineering, National Institute of Technology at Rourkela, Rourkela, India, in 2007.

He is an Assistant Professor with the School of Electrical Sciences, Indian Institute of Technology Bhubaneswar, Bhubaneswar, India. He was a Postdoctoral Research Fellow and a Visiting Professor with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC, Canada. His current research interests include adaptive and intelligent protection to transmission and distribution grids, phasor measurement unit and wide-area monitoring, microgrid planning, and dynamic security assessment in large power networks.

Dr. Samantaray was a 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, the 2012 IEEE Power Engineering Society Technical Committee Prize Paper Award, the 2013 Excellence in Reviewing Award, the Elsevier Science Award, and the 2015 National Academy of Sciences, India-SCOPUS Young Scientists Award. He is an Editor of *IET Generation, Transmission, and Distribution*; the *Canadian Journal of Electrical and Computer Engineering*; and *Electric Power Components and Systems*. He is a Member of the National Academy of Sciences, India.

Geza Joos (M'82–SM'89–F'06) received the M.Eng. and Ph.D. degrees in electrical engineering from McGill University, Montreal, QC, Canada, in 1974 and 1987.

He was with ABB, Zürich, Switzerland; École de Technologie Supérieure, Montreal; and Concordia University, Montreal. He has been a Professor with McGill University, since 2001, and holds a Canada Research Chair in power electronics applied to power systems. He is involved on a regular basis in consulting activities in power electronics and power systems. His current research interests include fundamental and applied research related to the application of high-power electronics to power conversion (including distributed generation and wind energy) and to power systems.

Mr. Joos is active in a number of IEEE Industry Applications Society committees, as well as IEEE Power Engineering Society working groups and CIGRE working groups. He is a Fellow of the Canadian Academy of Engineering and the Engineering Institute of Canada.