IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 7, 2021, accepted January 28, 2021, date of publication February 2, 2021, date of current version February 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3056534

An Intelligent Data Mining-Based Fault Detection and Classification Strategy for Microgrid

SHAZIA BALOCH AND MANNAN SAEED MUHAMMAD, (Senior Member, IEEE)
Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, South Korea

Corresponding author: Mannan Saeed Muhammad (mannan@skku.edu)

ABSTRACT The specific characteristics and operations of microgrid cause protection problems due to high penetration of distributed energy resources. To resolve these issues, the proposed scheme employs the Hilbert transform and data mining approach to protect the microgrid. First, the Hilbert transform is used to preprocess the faulted voltage and current signals to extract the sensitive fault features. Then, the obtained data set of the extracted features is input to the logistic regression classifier for fault detection. Later, fault classification is done by training the AdaBoost classifier. In the proposed scheme, the simulation results for feature extractions are evaluated on a standard International Electrotechnical Commission (IEC) medium voltage microgrid, compatible with MATLAB/SIMULINK software environment, whereas, Python is used for training and testing of data mining model. The results are evaluated under grid-connected and islanded modes for both looped and radial configurations by simulating various fault and no-fault cases. The results show that the accuracy of the proposed logistic regression and AdaBoost classifier is higher when compared to decision tree, support vector machine, and random forest methods. The results further validate the robustness of the proposed method against the measurement noise.

INDEX TERMS AdaBoost classifier, data mining, fault protection, feature extraction, Hilbert transform, logistic regression.

I. INTRODUCTION

The exponential enhancement in power requirements, substantial reduction in conventional resources, and growing appeals for green power have become central issues of the traditional electricity generation system [1], [2]. These problems have attracted more attention in power systems, hence raising a need to provide alternative approaches to overcome these issues. Distributed Energy Resources (DERs) is an alternative solution where the sources are integrated at the distribution utilities, and are located close to the load to provide power to local customers. The advancement of DERs technologies introduced the concept of microgrid, which is a significant part of the power distribution system, and performed a pivoted role in addressing the issues of traditional power systems [3]–[6]. Essentially, microgrid is a low or medium voltage distributed system, interconnected with a cluster of low power generation units, loads, and energy storage devices. It can effectively operate in both preplanned islanded and grid-connected modes. It provides notable benefits by maintaining the accessibility of power by utilizing the islanded mode of operation during the main grid outages. Microgrid offers superior power quality by reducing the carbon emission, and provides low implementation costs by reducing the transmission lines. In short, it improves the overall efficiency and economic dispatch in the network. However, due to the penetration of DERs into microgrid, it suffers from significant technical issues like stability problems, frequency and voltage control, protective devices failure, and operation, but mainly concerning protections problems [7]–[9].

The inclusion of DERs modifies the existing conventional networks into active distribution networks and leads to power system protection problems. These problems are mainly due to the protection philosophy of the prevailing distribution systems, that are designed based on the assumption that conventional protection systems are radial and power flow is always unidirectional from the source to consumers. Therefore, the multi-looped and multi-generation active distribution systems make the distribution systems more complex [10], [11]. A major concern with the implementation of microgrid is that bidirectional power flows in either direction,

The associate editor coordinating the review of this manuscript and approving it for publication was Sathish Kumar.

depending on fault location. Also, microgrid may change topology due to grid or islanded, and meshed or radial configurations, resulting in various fault current levels [12], [13].

The contribution of fault current from the main utility is ten times larger as compared to small distributed generators (DGs) in the microgrid. Synchronous-based DGs contribute 5 to 6 times while the fault current is limited to twice of the rated current for Inverter Interfaced DGs (IIDGs). The variation in fault current mainly depends on the penetration rate, type, and location of DGs [11], [14]. In order to maintain the correct operation of microgrid, it should be fast enough to detect the faults for the protection of sensitive loads. However, the traditional over-current relays not only cause incorrect detection of a fault, but also create selectivity problems resulting in switching off of healthy phases. Therefore, the conventional over-current relays are inadequate for microgrid protection due to their preset conditions, and need to be reconsidered [15], [16]. Hence, it is challenging to design an efficient microgrid protection scheme [17].

Previous literature have reported several techniques to find a solution to this challenging problem. An adaptive relay technique proposed in [18], implements a fast-recursive discrete Fourier transform and fuzzy-logic decision-making module. In [19], the authors introduced an intelligent protection scheme for microgrid using a combined wavelet transform and decision tree (DT) by utilizing local current measurements for fault detection and classification with IIDGs. Another fault detection scheme based on deep neural networks and wavelet transform for microgrid protection was proposed in [20]. The authors in [21], employed an approach, focused on identifying and evaluating the faulted line section by implementing data mining and wavelet packet transform. A comprehensive communication assisted protection-based strategy for microgrid was reported in [22], with dual directional over-current relay. The authors in [23], presented a fault protection strategy based on an integrated impedance angle, using the phasor measurement unit. A non-pilot protection strategy was designed in [24], for symmetrical and asymmetrical faults for inverter-dominated microgrid. Another protection method for inverter-based microgrid was proposed in [25], using current-only polarity comparison to detect the fault direction. A combined machine learning and signal processing technique was introduced in [26], [27], for radial distribution grid for detection, classification and location of the fault. A Stockwell transform and machine learning-based hybrid technique was proposed in [28], for a modeled distribution feeder to a laboratory scale for the detection, location and classification of single-line-to-ground fault. Another Stockwell transform and machine learning method for power distribution grid was presented in [29], to locate and identify the fault section. In [30], the authors developed an intelligent fault protection scheme for microgrid using convolutional neural network by extracting the fault features internally. In [31], a three-stage protection scheme was proposed for dynamic security status detection in microgrid. The authors in [32], presented an autocorrelation-based scheme for microgrid by using squaring and low-pass filtering method.

The aforementioned protection strategies, employed different protection aspects to solve the protection problems in microgrid. However, each of the scheme has some limitations. Most of the protection strategies have considered only inverter-interfaced DGs [19], [24]. Some schemes have high fault detection time [19], [20]. Adaptive methods suffers from high computational burden caused by the complex fault calculations for relay settings [18]. Therefore, to overcome these limitations, this manuscript proposes a new scheme for microgrid protection using the logistic regression and AdaBoost classifier-based data mining model, for fault detection as well as classification. Standard deviation of seven electrical parameters at fault are computed to build the data mining model to train the logistic regression for fault detection, and AdaBoost classifier for fault classification. MATLAB/SIMULINK is used for features extractions, whereas Python is used for training and testing of data mining model. The contributions of this research are:

- Ascertain a simpler and efficient data mining-based fault detection and classification approach for microgrid.
- Collect the data by extracting the standard deviation of seven different fault features.
- Build the data mining model from the collected data set.
- Train the logistic regression and AdaBoost classifier for fault detection and classification respectively.
- Investigate the capability of microgrid for grid-connected and islanded modes of operation with looped and radial configurations.

The remaining of the paper is composed of four sections as follows. The proposed protection strategy and methodology is addressed in section II. Section III deals with a detailed discussion of the microgrid test system under study. Section IV is concerned with the findings and results of the proposed scheme, mainly focusing on fault detection and classification. Finally, the paper is concluded in section V.

II. PROPOSED FEATURES EXTRACTION AND DATA MINING STRATEGY

In the proposed scheme standard deviation of seven most effective and sensitive fault features are computed to build the data mining model. The envelope of autocorrelation function (ACF) and variance of autocorrelation function (VACF), containing vital information of the fault transient, are extracted through Hilbert transform. The transient energy is obtained by implementing the delta filter. Sequence analyzer is used to extract the negative sequence components of the active power of the measured current. Total Harmonic Distortion (THD) of current and voltage are also extracted. These features are obtained by simulating various fault and no-fault events. Fault events are obtained with a wide variation in fault resistance, changing the fault type and fault location. However, the no-fault cases are obtained by sudden load variations and capacitor switching. A summary of the

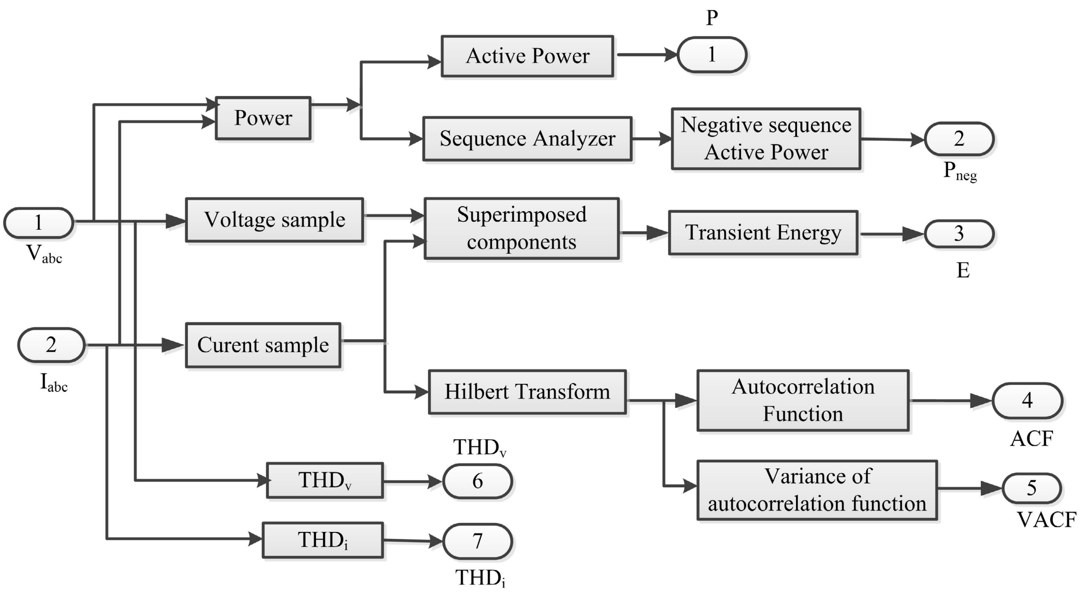

FIGURE 1. Schematic of proposed scheme feature extraction.

TABLE 1. Fault events simulating conditions.

Parameters	Fault Events	Counts
Operating mode	Grid-connected or islanded	2
Topology	Radial or loop	2
Fault types	SLG, LL, LLG, LLLG	4
Fault resistance (Ω)	0.01-100	10
Fault line	$DL_1, DL_2, DL_3, DL_4, DL_5$	5
Total fault cases	800	

TABLE 2. No-fault events simulating conditions.

Parameter	Counts
Operating modes (grid-connected or islanded)	2
Topology (radial or loop)	2
Capacitor switching at PCC and load buses	6
Sudden load changes	6
Total no-fault cases	144

different fault and no-fault events is shown in Tables 1 and 2. Once all the features are extracted, they are used to build the data mining model to train the logistic regression and AdaBoost classifier respectively. At first, data mining model is built to train the logistic regression classifier to detect the fault. After the fault detection, AdaBoost classifier is trained to classify the fault. The time considered for fault detection and fault classification are taken to be half cycle and one cycle respectively. The schematic of the proposed strategy for feature extraction is shown in Fig. 1.

A. HILBERT TRANSFORM

The proposed scheme uses the Hilbert transform to determine the envelope of each phase current signal. It is a signal processing technique utilized in communication systems for current signal analysis. It is also widely used in protection schemes to detect the faults in microgrid [33]. It produces an imaginary signal $\hat{x}(t)$ of an original signal $x(t)$ with a 90° phase shift [34], [35]. Hilbert transform is defined as the convolution of a signal with the Hilbert transform operator. The time domain Hilbert transform operator can be expressed as:

$$H(t) = \frac{1}{\pi t}, \quad -\infty < t < +\infty \tag{1}$$

because the Hilbert transform operator is not an integrable function, therefore it is obtained by using the Cauchy integral as:

$$\begin{aligned} \hat{x}(t) &= x(t) \otimes \frac{1}{\pi t} \\ &= \frac{1}{\pi} p \int_{+\infty}^{-\infty} \frac{x(\tau)}{t-\tau}, \end{aligned} \tag{2}$$

where, p is the integral principal value and ⊗ represents convolution.

This method is simple and takes less time for fault detection in microgrid as compared to other signal processing techniques [36].

B. FEATURE EXTRACTION

To build the data mining model for the proposed protection strategy, the standard deviation of the following electrical parameters is considered:

1) AUTOCORRELATION FUNCTION (ACF)

The envelope of ACF is considered as one of the sensitive features and it is computed through the Hilbert transform. It substantially changes with fault, and provides hidden fault information in the microgrid. When a fault occurs, the envelope of current signal changes significantly, indicating a clear fault existence [32], [37]. The ACF of the current signal can be computed by (3):

$$r_k = \frac{\sum_{i=1}^{M-k}(x_i - \overline{x})(x_{i+k} - \overline{x})}{\sum_{i=1}^{M}(x_i - \overline{x})^2}, \tag{3}$$

where, r_k is ACF between two data points, x_i is the value of original data set, x_{i+k} is the value of shifted data set, and $\overline{x}$ is the mean of original data set.

The input current for three-phase-to-ground fault is shown in Fig. 2, whereas, the distorted envelope of the current signal is shown in Fig. 3 which demonstrates the behavior of ACF under the fault situation.

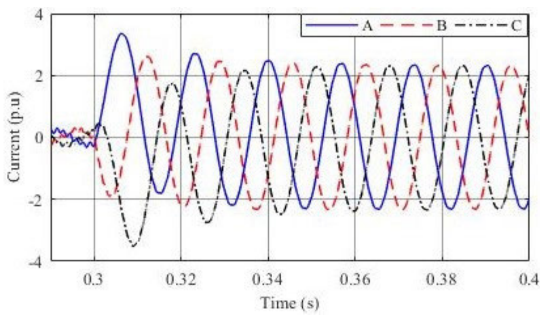

FIGURE 2. Behavior of input current under three-phase-to-ground fault.

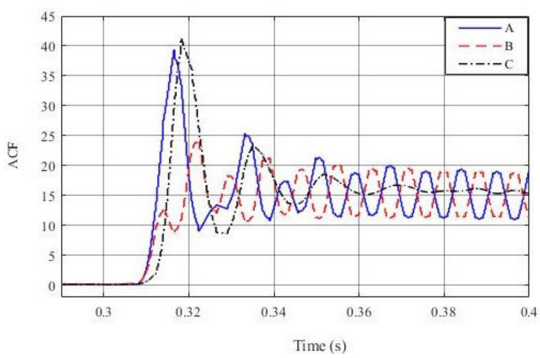

FIGURE 3. Behavior of autocorrelation function at fault.

2) VARIANCE OF AUTOCORRELATION FUNCTION (VACF)

After fault is detected by the ACF, the fault existence is verified by the VACF. When a fault occurs, the VACF increases and confirms the fault in the microgrid [38], as shown in Fig. 4.

3) TOTAL HARMONIC DISTORTION (THD) OF VOLTAGE AND CURRENT

THD is an indicator of signal distortion. It increases the harmonics and is considered as a sensitive feature against the fault. It is the ratio of total harmonics to the fundamental frequency, and can be voltage or current harmonics [39], [40]. Equations (4) and (5) are used to compute the THD of voltage

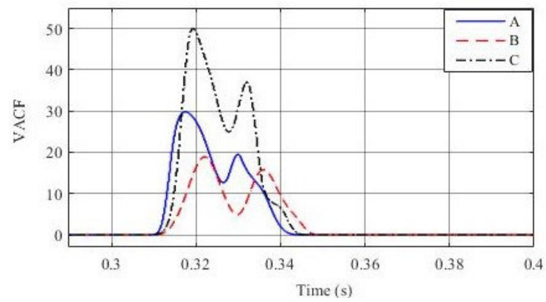

FIGURE 4. Behavior of variance of autocorrelation function of three-phase-to-ground fault.

and current respectively.

$$THD_v = \frac{\sqrt{V_2^2 + V_3^2 + V_4^2 + \ldots . V_n^2}}{V_1}, \tag{4}$$

$$THD_i = \frac{\sqrt{I_2^2 + I_3^2 + I_4^2 + \ldots . I_n^2}}{I_1}, \tag{5}$$

where, THD_v and THD_i represent the Total Harmonic Distortion for voltage and current respectively.

The THD of the fault current is shown in Fig. 5, and it can be seen that when a fault occurs the current harmonics increase.

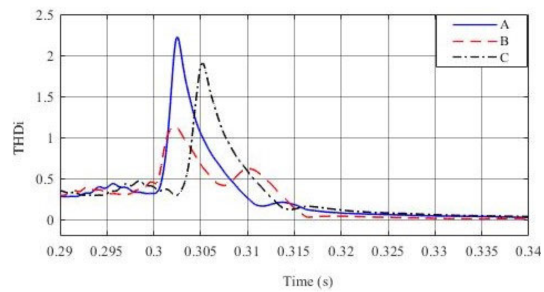

FIGURE 5. Behavior of Total Harmonic Distortion.

4) TRANSIENT ENERGY

Transient energy is computed in three steps. At first, the three-phase signal is converted into a modal signal. In the second step, delta filter is employed to extract the superimposed modal voltage and superimposed modal current as in (6):

$$\Delta x(t) = x_{\beta F}(t) - x_{\alpha F}(t), \tag{6}$$

where, $\Delta x(t)$ represents the superimposed modal signal for voltage or current, and $x_{\alpha F}(t)$ and $x_{\beta F}(t)$ represent the pre-fault and post-fault signals respectively.

Once the superimposed modal current and voltage are obtained, their product is used to calculate the transient power as per (7):

$$\Delta P = \Delta V \times \Delta I, \tag{7}$$

where, ΔP, ΔV, and ΔI represent the transient power, superimposed modal voltage and current respectively.

Next the integration of the transient power over one cycle gives the transient energy, as:

$$\Delta E = \int_0^T \Delta P dt, \tag{8}$$

where, ΔE is the transient energy.

5) ACTIVE POWER

The power dissipated in the system during the fault appears as I^2 losses that indicate the fault in the system. It can be calculated from (9) as:

$$P = VI \cos \phi, \tag{9}$$

where, P represents the active power and ϕ is the phase angle between the voltage and current.

6) NEGATIVE-SEQUENCE ACTIVE POWER

Like other features, negative-sequence active power is also very sensitive to a fault. Whenever a fault occurs, negative sequence components change significantly and show transients in the system. The negative sequence component behavior under fault situation is shown in Fig. 6.

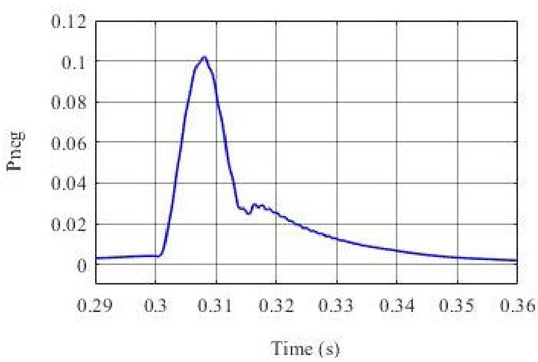

FIGURE 6. Negative-sequence active power at fault.

C. DATA MINING-BASED CLASSIFIERS

Data mining is the most important tool in machine learning, and is used to evaluate the dependencies between the system variables. It is applied to analyze and generate a more comprehensive configuration of a data set to predict the performance of a system, by creating a system model as per the input features data set. This model can be further used for prediction, classification, and estimation purposes. There are various data mining methods such as random forest (RF), decision tree (DT), neural network, and support vector machines (SVM), that have been adopted to detect and classify the faults in power system [41], [42].

In this study, two efficient data mining methods, logistic regression and AdaBoost classifier are employed to compute the accuracy of the proposed fault detection and classification strategy. A brief discussion of both classifiers is given below:

TABLE 3. Loads data of the proposed test system.

Load	P (MW)	Q (MVAR)
L_1	3.0	1.0
L_2	3.0	1.0
L_3	4.0	1.5
L_4	1.0	0.75
L_5	1.0	0.75
L_6	1.0	0.5

1) LOGISTICS REGRESSION

Logistic regression is based on regression analysis and is used to detect the fault in the microgrid for the proposed scheme. It is employed to investigate the relationship between numerous independent variables [43], [44]. It predicts the output of an estimated expected value from a categorical dependent binary variable [45]. It can be computed as per (10):

$$y = \frac{\exp[s_0 + s_1 x]}{1 + \exp[s_0 + s_1 x]}, \tag{10}$$

where, the predicted output is represented by y, s_0 is the intercept term, x is the single input, and s_1 is the coefficient of x.

2) AdaBoost CLASSIFIER

AdaBoost stands for adaptive boosting, and used in the proposed scheme to classify the fault in the microgrid. It is a strong ensemble, designed for classification problems. Yoav Freund and Robert Schapire in 1995/1996 introduced the first boosting method [46], [47]. It joins several poor accuracy models to compensate the weaknesses of predecessors to obtain a strong classifier. Several algorithmic and theoretical features make it exceptionally attractive because it is simpler and easier to train the weak classifiers that failed to provide the accurate prediction. AdaBoost increases the weight of misclassified data points by training the data sample in each iteration and ensures the accurate predictions [48], [49]. The weight of M weak classifiers is computed by using (11):

$$F(x) = \text{sign}(\sum_{m=1}^{M} \theta_m f_m(x)), \tag{11}$$

where, m_{th} weak classifier is defined by f_m, θ_m is the corresponding weight, and M is the combination of weak classifiers.

III. TEST SYSTEM UNDER STUDY

The microgrid test system, developed in the MATLAB/SIMULINK environment is shown in Fig. 7. The microgrid distribution network is interconnected at the PCC to a 25kV, 15MVA, and 60Hz grid with a switch, that changes the modes between grid-connected and islanded. The sampling frequency considered is 3.6kHz. It comprises of two 3MVA (DER_1, and DER_3) and one 2MVA (DGR_2) IIDGs, and

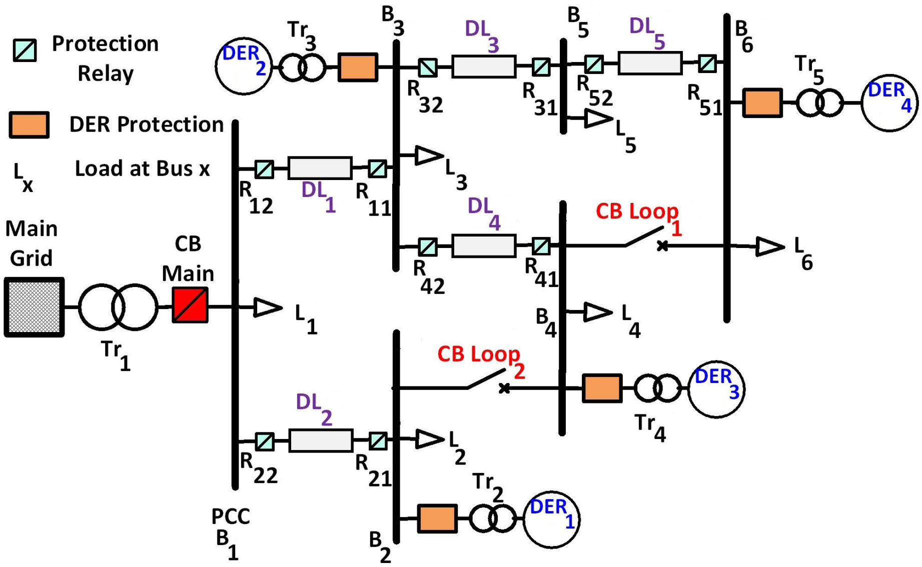

FIGURE 7. Test system of microgrid.

TABLE 4. Parameter values of logistic regression classifier.

Parameter	Value
C	100
Random state	20
Solver	Lbfgs

TABLE 5. Confusion matrix for fault and no-fault cases.

Total cases 283	No-fault	Fault
No-fault	43	0
Fault	2	238

TABLE 6. Comparison of the proposed scheme with DT and SVM.

Data mining method	Accuracy %	Precision %	Recall %
DT	97.87	97.91	97.33
SVM	98.59	98.33	100
Proposed scheme	99.29	99.16	100

TABLE 7. Parameter values of AdaBoost classifier.

Parameter	Value
Max depth	4
n-estimator	80
Learning rate	0.01
Random state	12

one 7MVA synchronous-based DG (DER_4). To switch the microgrid between looped and radial configurations, Circuit Breaker CB loop_1 and Circuit Breaker CB loop_2 are used. It is composed of five distributed sections (DL_1, DL_2, DL_3, DL_4, DL_5), each with a line length of 20km. There are six loads connected to each bus with the values shown in Table 3. The microgrid and the main grid are interconnected through a 120/25kV Dyn transformer, whereas all DG sources are connected through a 0.630/25kV transformer.

IV. RESULTS AND DISCUSSIONS

For performance validation and verification, various indices were used. Accuracy, precision, and recall are three statistical metrics, exploited to evaluate the performance of proposed fault detection and classification strategy. These performance measurement metrics are defined as follows:

1) *Accuracy* computes the reliability between the predicted events versus the actual events for both fault and no-fault events simultaneously. It can be defined as follows:

$$\frac{\text{Total } (\hat{F} + \hat{\hat{F}})}{\text{Total } (F + \overline{F})}, \tag{12}$$

where, $\hat{F}$ and $\hat{\hat{F}}$ represent predicted fault and no-fault events, whereas, F and $\overline{F}$ show the actual fault and no-fault events.

2) *Precision* is one of the crucial statistical metrics employed to precisely give the relationship between predicted fault events and actual fault events to assess the reliability of fault protection relay. It is given as follows:

$$\frac{\text{Total } \hat{F}}{\text{Total } F}, \tag{13}$$

where, $\hat{F}$ represents the predicted fault events and F denotes the actual fault events.

3) *Recall* reveals the total number of events which are no-fault but predicted as fault events and considered as misdetection, given in (14):

$$\frac{\text{Total } \hat{\hat{F}}}{\text{Total } \overline{F}}, \tag{14}$$

where, $\hat{\hat{F}}$ represents the predicted no-fault events and $\overline{F}$ represents the actual no-fault events.

TABLE 8. Confusion matrix for fault classification.

Fault type	SLG	LLG	LL	LLLG	Accuracy per fault type %
SLG	60	0	0	0	100
LLG	0	59	1	0	98.33
LL	0	1	59	0	98.33
LLLG	0	1	0	59	98.33
Overall accuracy	98.75%				

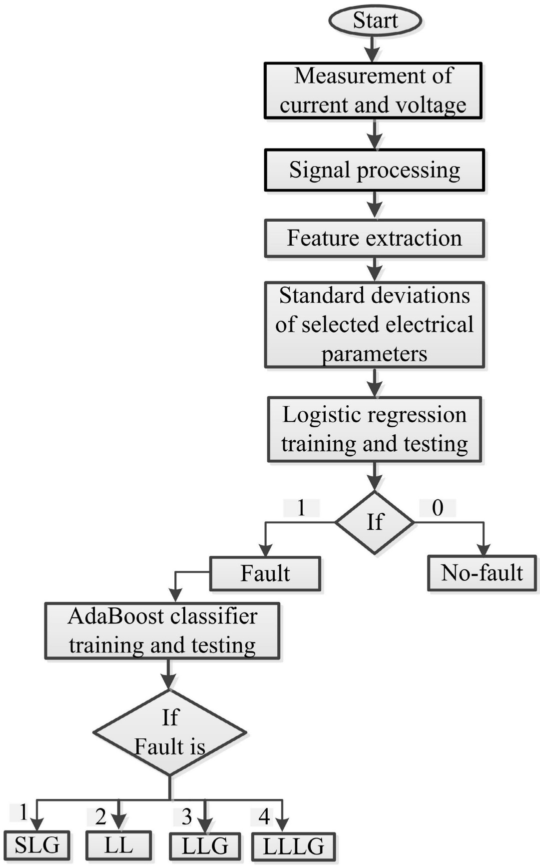

FIGURE 8. Flow chart of proposed fault detection and classification strategy.

A. DATA MINING MODEL FOR FAULT DETECTION

The complete fault detection and classification strategy of the proposed protection scheme is given in Fig. 8. While building the data mining model for the proposed scheme, the obtained data set is divided into two parts. The first part comprises of 70% of the data set and is used to build the data mining model. Once the model is trained, the remaining 30% unseen data set is used to carry out the testing to assess the performance. The training data belongs to a particular class i.e., 1 for fault and 0 for no-fault events. The training is performed with a logistic regression classifier to differentiate between the fault and no-fault events. The parameter values of logistic regression for fault detection are shown in Table 4. The confusion matrix for fault and no-fault cases is shown in Table 5. The proposed study contains a feature set, with a total number of 944 cases, with both fault and no-fault events. The total number of fault cases are 800, which are derived from the looped and radial configuration networks for grid-connected and islanded modes. The no-fault cases are 144, obtained by the capacitor switching and load variations. Table 6 illustrates the performance of the proposed scheme with DT and SVM. It is observed that the proposed scheme has an accuracy of 99.29% (with 99.16% precision, and 100% recall) for fault detection, as compared to DT which has an accuracy of 97.87% (with 97.91% precision and 97.33% recall), whereas SVM has accuracy of 98.59% (with 98.33% precision and 100% recall). Similarly, Fig. 9 shows the graphical representation of accuracy comparison of the proposed scheme with DT and SVM.

TABLE 9. Comparison of the proposed scheme with DT, SVM, and RF.

Data mining method	Accuracy in %
DT	94.16
SVM	96.66
Random Forest	97.91
Proposed scheme	98.75

B. DATA MINING MODEL FOR FAULT CLASSIFICATION

After fault detection, AdaBoost classifier is used to classify the fault. The model generated for fault classification uses four classes, single-line-to-ground faults (SLG), line-to-line faults (LL), line-to-line-to-ground faults (LLG), and three-phase-to-ground faults (LLLG). The data mining model for fault classification is created in the same manner as the logistic regression by assigning four different values to each fault type i.e., 1 for SLG, 2 for LL, 3 for LLG, and 4 for LLLG faults. Table 7 indicates the parameter values of AdaBoost classifier. Table 8 shows the confusion matrix of the proposed scheme for fault classification. The results analysis show that the accuracy for SLG fault is 100%, however; for LL, LLG, and LLLG faults is 98.33%. Table 9 shows the performance of the proposed scheme with DT, SVM, and RF. It can be seen that the accuracy of AdaBoost classifier is 98.75% in comparison of 94.16% for DT, 96.66% for SVM, and 97.91% for RF. Similarly, Fig. 10, shows the graphical representation of AdaBoost classifier accuracy with other methods. Further analysis shows that during the fault classification, problem arises when LLG fault is classified as LL. Similarly, LL and LLLG faults are classified as LLG faults, which can be considered as misdetection by the proposed scheme.

A comparison of the proposed scheme with other existing techniques is given in Table 10.

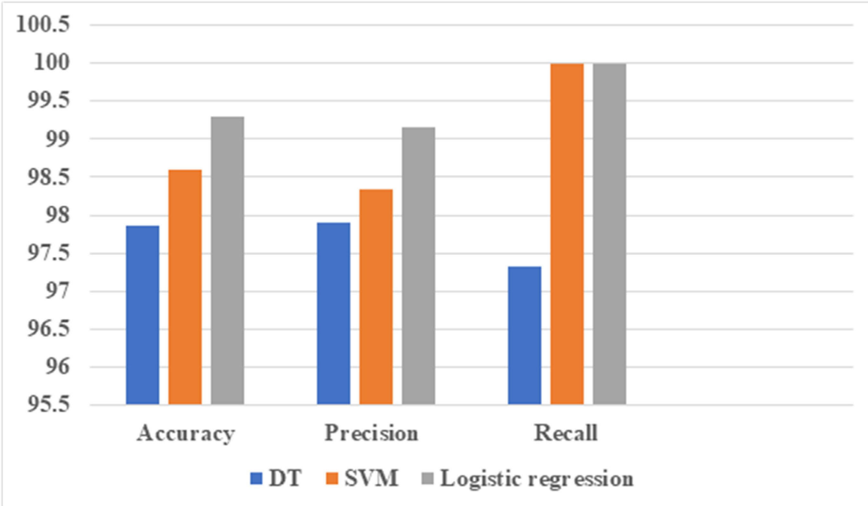

FIGURE 9. Graphical representation of proposed scheme compared with DT and SVM.

TABLE 10. Comparison of the proposed scheme with existing techniques.

S.No	Characteristics	Existing schemes	Drawbacks	Proposed schemes
1.	**Operating modes**	[07]	Islanded only	Both modes
2.	**Simplicity**	[27]	Complex (due to dual setting)	Simple
3.	**Compatution issues**	[21]	Causes computation burden	No computation burden
4.	**DGs type**	[19,24]	IIDGs only	IIDGs and synchronous-based DGs
5.	**Configuration**	[26]	Radial	Radial and looped configuration
6.	**Accuracy**	Data mining method	Fault Detection	Fault Classification
		DNNs [20]	99.31	97.92
		Decision tree [42]	99.08	Not considered
		Proposed method	99.29	98.75

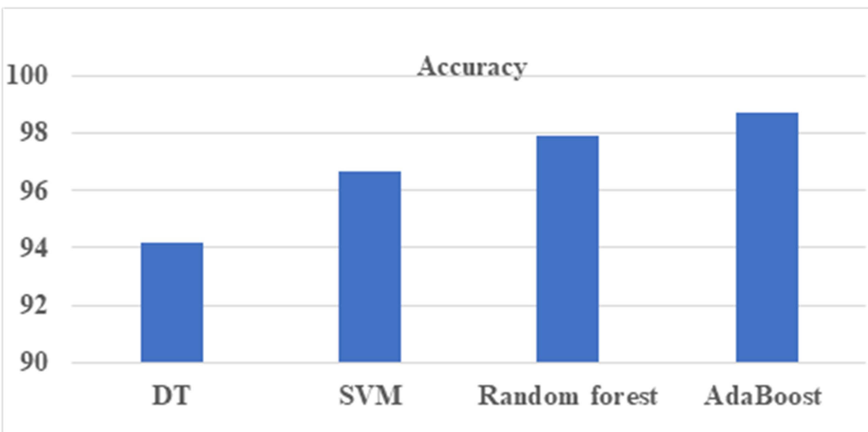

FIGURE 10. Graphical representation of proposed scheme accuracy with DT, SVM, and RF.

C. IMPACT OF NOISE

The robustness of the proposed scheme was also evaluated under the measurement noise. The three-phase voltage measurement was distorted to conduct the simulations by adding the white Gaussian noise with 30dB and 40dB signal-to-noise ratio (SNR) [20]. Logistic regression was trained and tested with the distorted data, with the same features. Table 11 summarizes the performance of the scheme under noise. It can be seen that the presence of 40dB noise did not have a significant impact on the performance of the proposed scheme, but deviated from 99.29% to 97% with 30dB. However, the

TABLE 11. Effect of noise on fault detection.

SNR	Accuracy %	Precision %	Recall %
30dB	97	97.22	95.83
40dB	98.33	100	96

performance of SVM and DT had a significant impact with the accuracies between $97-95\%$ under measurement noise.

V. CONCLUSION

This scheme is designed to evaluate fault detection and classification strategy for the protection of microgrid by implementing the Hilbert transform and data mining tools. Seven sensitive fault features are extracted by varying the fault resistance, fault type, fault location, capacitor switching and load variations for both fault and no-fault events. Later, they are used to build the data mining model to train the logistic regression and AdaBoost classifier to detect and classify the fault in the microgrid. The system is tested on a standard IEC medium voltage microgrid for grid-connected and islanded modes of operation for both looped and radial configurations. The results demonstrated that logistic regression classifier has good performance with an accuracy of 99.29% as compared to SVM and DT with accuracy rates of 98.59% and 97.87% respectively. The results also show that the accuracy of AdaBoost classifier for fault classification is 98.75% as compared to 94.16%, 96.66%, and 97.91% for DT, SVM and RF respectively. Based on these results it can be concluded that the proposed scheme provides high accuracy and reliable measures for microgrid protection.

REFERENCES

[1] M. Manohar, E. Koley, and S. Ghosh, "Reliable protection scheme for PV integrated microgrid using an ensemble classifier approach with real-time validation," *IET Sci., Meas. Technol.*, vol. 12, no. 2, pp. 200–208, Mar. 2018.

[2] M. R. Alam, M. T. A. Begum, and K. M. Muttaqi, "Assessing the performance of ROCOF relay for anti-islanding protection of distributed generation under subcritical region of power imbalance," *IEEE Trans. Ind. Appl.*, vol. 55, no. 5, pp. 5395–5405, Sep. 2019.

[3] A. Mohamed, B. Younes, T. Lamhamdi, H. El Moussaoui, and H. El Markhi, "Fault location and isolation technique in smart distribution systems with distributed generation," in *Proc. 1st Int. Conf. Innov. Res. Appl. Sci., Eng. Technol. (IRASET)*, Apr. 2020, pp. 1–5.

[4] A. Sharma and B. K. Panigrahi, "Phase fault protection scheme for reliable operation of microgrids," *IEEE Trans. Ind. Appl.*, vol. 54, no. 3, pp. 2646–2655, May 2018.

[5] A. Hooshyar and R. Iravani, "Microgrid protection," *Proc. IEEE*, vol. 105, no. 7, pp. 1332–1353, Jul. 2017.

[6] S. Ram Ola, A. Saraswat, S. K. Goyal, V. Sharma, B. Khan, O. P. Mahela, H. Haes Alhelou, and P. Siano, "Alienation coefficient and wigner distribution function based protection scheme for hybrid power system network with renewable energy penetration," *Energies*, vol. 13, no. 5, p. 1120, Mar. 2020.

[7] Z. Chen, X. Pei, M. Yang, L. Peng, and P. Shi, "A novel protection scheme for inverter-interfaced microgrid (IIM) operated in islanded mode," *IEEE Trans. Power Electron.*, vol. 33, no. 9, pp. 7684–7697, Sep. 2018.

[8] R. Escudero, J. Noel, J. Elizondo, and J. Kirtley, "Microgrid fault detection based on wavelet transformation and Park's vector approach," *Electr. Power Syst. Res.*, vol. 152, pp. 401–410, Nov. 2017.

[9] M. Faisal, M. A. Hannan, P. Jern Ker, A. Hussain, M. B. Mansor, and F. Blaabjerg, "Review of energy storage system technologies in microgrid applications: Issues and challenges," *IEEE Access*, vol. 6, pp. 35143–35164, 2018.

[10] P. Mahat, Z. Chen, 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.

[11] S. Beheshtaein, R. Cuzner, M. Savaghebi, and J. M. Guerrero, "Review on microgrids protection," *IET Gener., Transmiss. Distrib.*, vol. 13, no. 6, pp. 743–759, 2019.

[12] A. Kulshrestha, O. P. Mahela, M. K. Gupta, N. Gupta, N. Patel, T. Senjyu, M. S. S. Danish, and M. Khosravy, "A hybrid fault recognition algorithm using stockwell transform and wigner distribution function for power system network with solar energy penetration," *Energies*, vol. 13, no. 14, p. 3519, Jul. 2020.

[13] S. Ram Ola, A. Saraswat, S. K. Goyal, S. K. Jhajharia, B. Khan, O. P. Mahela, H. Haes Alhelou, and P. Siano, "A protection scheme for a power system with solar energy penetration," *Appl. Sci.*, vol. 10, no. 4, p. 1516, Feb. 2020.

[14] N. El-Naily, S. M. Saad, T. Hussein, and F. A. Mohamed, "A novel constraint and non-standard characteristics for optimal over-current relays coordination to enhance microgrid protection scheme," *IET Gener., Transmiss. Distrib.*, vol. 13, no. 6, pp. 780–793, Mar. 2019.

[15] X. Li, A. Dysko, and G. M. Burt, "Traveling wave-based protection scheme for inverter-dominated microgrid using mathematical morphology," *IEEE Trans. Smart Grid*, vol. 5, no. 5, pp. 2211–2218, Sep. 2014.

[16] I. Xyngi and M. Popov, "An intelligent algorithm for the protection of smart power systems," *IEEE Trans. Smart Grid*, vol. 4, no. 3, pp. 1541–1548, Sep. 2013.

[17] F. Mumtaz and I. S. Bayram, "Planning, operation, and protection of microgrids: An overview," *Energy Procedia*, vol. 107, pp. 94–100, Feb. 2017.

[18] D. S. Kumar, D. Srinivasan, and T. Reindl, "A fast and scalable protection scheme for distribution networks with distributed generation," *IEEE Trans. Power Del.*, vol. 31, no. 1, pp. 67–75, Feb. 2016.

[19] D. P. Mishra, S. R. Samantaray, and G. Joos, "A combined wavelet and data-mining based intelligent protection scheme for microgrid," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2295–2304, Sep. 2016.

[20] J. J. Q. Yu, Y. Hou, A. Y. S. Lam, and V. O. K. Li, "Intelligent fault detection scheme for microgrids with wavelet-based deep neural networks," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1694–1703, Mar. 2019.

[21] S. Jamali, S. Ranjbar, and A. Bahmanyar, "Identification of faulted line section in microgrids using data mining method based on feature discretisation," *Int. Trans. Electr. Energy Syst.*, vol. 30, no. 6, Jun. 2020, Art. no. e12353.

[22] H. M. Sharaf, H. H. Zeineldin, and E. El-Saadany, "Protection coordination for microgrids with grid-connected and islanded capabilities using communication assisted dual setting directional overcurrent relays," *IEEE Trans. Smart Grid*, vol. 9, no. 1, pp. 143–151, Jan. 2018.

[23] N. K. Sharma and S. R. Samantaray, "PMU assisted integrated impedance angle-based microgrid protection scheme," *IEEE Trans. Power Del.*, vol. 35, no. 1, pp. 183–193, Feb. 2020.

[24] H. Lahiji, F. Badrkhani Ajaei, and R. E. Boudreau, "Non-pilot protection of the Inverter- dominated microgrid," *IEEE Access*, vol. 7, pp. 142190–142202, 2019.

[25] B. Wang and L. Jing, "A protection method for inverter-based microgrid using current-only polarity comparison," *J. Modern Power Syst. Clean Energy*, vol. 8, no. 3, pp. 446–453, 2020.

[26] Y. D. Mamuya, Y.-D. Lee, J.-W. Shen, M. Shafiullah, and C.-C. Kuo, "Application of machine learning for fault classification and location in a radial distribution grid," *Appl. Sci.*, vol. 10, no. 14, p. 4965, Jul. 2020.

[27] M. Shafiullah, M. A. Abido, and Z. Al-Hamouz, "Wavelet-based extreme learning machine for distribution grid fault location," *IET Gener., Transmiss. Distrib.*, vol. 11, no. 17, pp. 4256–4263, Nov. 2017.

[28] A. Aljohani, T. Sheikhoon, A. Fataa, M. Shafiullah, and M. A. Abido, "Design and implementation of an intelligent single line to ground fault locator for distribution feeders," in *Proc. Int. Conf. Control, Autom. Diagnosis (ICCAD)*, Jul. 2019, pp. 1–6.

[29] M. Shafiullah, M. Abido, and T. Abdel-Fattah, "Distribution grids fault location employing ST based optimized machine learning approach," *Energies*, vol. 11, no. 9, p. 2328, Sep. 2018.

[30] S. B. A. Bukhari, C. Kim, K. K. Mehmood, R. Haider, and M. Saeed Uz Zaman, "Convolutional neural network-based intelligent protection strategy for microgrids," *IET Gener., Transmiss. Distrib.*, vol. 14, no. 7, pp. 1177–1185, Apr. 2020.

[31] S. Teimourzadeh, F. Aminifar, M. Davarpanah, and M. Shahidehpour, "Adaptive protection for preserving microgrid security," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 592–600, Jan. 2019.

[32] S. Baloch, S. Z. Jamali, K. K. Mehmood, S. B. A. Bukhari, M. S. Uz Zaman, A. Hussain, and C.-H. Kim, "Microgrid protection strategy based on the autocorrelation of current envelopes using the squaring and low-pass filtering method," *Energies*, vol. 13, no. 9, p. 2350, May 2020.

[33] O. P. Mahela, J. Sharma, B. Kumar, B. Khan, and H. H. Alhelou, "An algorithm for the protection of distribution feeder using stockwell and Hilbert transforms supported features," *CSEE J. Power Energy Syst.*, 2020.

[34] Y. Shmaliy, *Continuous-Time Signals*, vol. 129. Dordrecht, The Netherlands: Springer, 2006.

[35] S. A. Tretter, *Communication System Design Using DSP Algorithms: With Laboratory Experiments for the TMS320C6713TM DSK*. Springer, 2008.

[36] G. S. Yogee, O. P. Mahela, K. D. Kansal, B. Khan, R. Mahla, H. Haes Alhelou, and P. Siano, "An algorithm for recognition of fault conditions in the utility grid with renewable energy penetration," *Energies*, vol. 13, no. 9, p. 2383, May 2020.

[37] T. Ghanbari, "Autocorrelation function-based technique for stator turn-fault detection of induction motor," *IET Sci., Meas. Technol.*, vol. 10, no. 2, pp. 100–110, Mar. 2016.

[38] R. Haider, C. H. Kim, T. Ghanbari, S. B. A. Bukhari, M. S. Zaman, S. Baloch, and Y. S. Oh, "Passive islanding detection scheme based on autocorrelation function of modal current envelope for photovoltaic units," *IET Gener., Transmiss. Distrib.*, vol. 12, no. 3, pp. 726–736, Feb. 2018.

[39] A. Danandeh, H. Seyedi, and E. Babaei, "Islanding detection using combined algorithm based on rate of change of reactive power and current THD techniques," in *Proc. Asia–Pacific Power Energy Eng. Conf.*, Mar. 2012, pp. 1–4.

[40] T. S S Senarathna and K. T M U. Hemapala, "Review of adaptive protection methods for microgrids," *AIMS Energy*, vol. 7, no. 5, pp. 557–578, 2019.

[41] M. Manohar, E. Koley, and S. Ghosh, "Enhancing the reliability of protection scheme for PV integrated microgrid by discriminating between array faults and symmetrical line faults using sparse auto encoder," *IET Renew. Power Gener.*, vol. 13, no. 2, pp. 308–317, Feb. 2019.

[42] S. Kar, S. R. Samantaray, and M. D. Zadeh, "Data-mining model based intelligent differential microgrid protection scheme," *IEEE Syst. J.*, vol. 11, no. 2, pp. 1161–1169, Jun. 2017.

[43] T. Liu, Z. Li, Y. Tang, D. Yang, S. Jin, and J. Guan, "The application of the machine learning method in electromyographic data," *IEEE Access*, vol. 8, pp. 9196–9208, 2020.

[44] J. Ran, G. Zhang, T. Zheng, and W. Wang, "Logistic regression analysis on learning behavior and learning effect based on SPOC data," in *Proc. 13th Int. Conf. Comput. Sci. Edu. (ICCSE)*, Aug. 2018, pp. 1–5.

[45] M. Perez-Ortiz, P. A. Gutierrez, and C. Hervas-Martinez, "Projection-based ensemble learning for ordinal regression," *IEEE Trans. Cybern.*, vol. 44, no. 5, pp. 681–694, May 2014.

[46] A. J. Ferreira and M. A. Figueiredo, "Boosting algorithms: A review of methods, theory, and applications," in *Ensemble Machine Learning*. 2012, pp. 35–85.

[47] Y. Freund, R. Schapire, and N. Abe, "A short introduction to boosting," *J.-Jpn. Soc. Artif. Intell.*, vol. 14, nos. 771–780, p. 1612, 1999.

[48] L. Chen. (2019). *Basic Ensemble Learning (Random Forest, AdaBoost, Gradient Boosting)-Step by Step Explaineds*. [Online]. Available: https://towardsdatascience.com/basic-ensemble-learning-random-forest-adaboost-gradient-boosting-step-by-step-explained-95d49d1e2725

[49] J. Kim, J. Lee, C. Lee, E. Park, J. Kim, H. Kim, J. Lee, and H. Jeong, "Optimal feature selection for pedestrian detection based on logistic regression analysis," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2013, pp. 239–242.

SHAZIA BALOCH received the B.E. degree in electrical engineering from the Balochistan University of Engineering and Technology at Khuzdar, Khuzdar, Pakistan, in 2010, and the M.E. degree in electrical power engineering from the Mehran University of Engineering and Technology, Jamshoro, Pakistan, in 2016. She is currently pursuing the Ph.D. degree in electrical and computer engineering with Sungkyunkwan University, Suwon, South Korea. Her research interests include power electronics, data mining, and microgrid protection.

MANNAN SAEED MUHAMMAD (Senior Member, IEEE) received the B.S. degree in mechatronics engineering from the National University of Science and Technology, Rawalpindi, Pakistan, in 2003, and the M.S. and Ph.D. degrees from the Gwangju Institute of Science and Technology, Gwangju, South Korea, in 2006 and 2012, respectively, all in mechatronics engineering. From 2012 to 2015, he worked as an Assistant Professor with Hanyang University, Seoul. Since 2016, he has been with Sungkyunkwan University, Suwon, South Korea. His research interests include machine learning, and signal and image processing.