Data Mining Model Based Differential Microgrid Fault Classification Using SVM Considering Voltage and Current Distortions

P. Venkata[*]*, V. Pandya, A. V. Sant*

Electrical Engineering Department, School of Technology, Pandit Deendayal Energy University, Gandhinatar, Gujarat, India

Abstract- *This paper reports support vector machine (SVM) based fault detection and classification in microgrid while considering distortions in voltages and currents, time and frequency series parameters, and differential parameters. For SVM-based fault classification, the data set is formed by analysing the operation of the standard IEC microgrid model, with and without grid interconnection, under different fault and non-fault scenarios. Fault scenarios also include different locations, resistances, and incident angles of fault. Whereas, for non-fault scenarios, the variation in load is considered. Voltages and currents from both ends of the distribution line (DL) are sampled at 1920 Hz. The time and frequency series parameters, total harmonic distortion (THD) in current and voltage, and differential parameters are determined. The SVM algorithm uses these parameters to detect and classify faults. The performance of this developed SVM based algorithm is compared with that of different machine learning algorithms. This comparative analysis reveals that SVM detects and classifies the faults on the microgrid with an accuracy of over 99.99%. The performance of the proposed method is also tested with 30 dB, 35 dB, and 40 dB noise in the generated data, which represent measurement errors.*

Keywords: Data Mining, Fault Identification and Classification, Microgrid Protection, Machine Learning, SVM.

1. INTRODUCTION

1.1. Background

Due to rapid urbanisation, industrialisation and rural electrification, the electric power demand is increasing day-by-day [1]. With the large-scale adoption of electric vehicles, this demand will further increase. This rapidly increasing energy demand, the depletion of fossil fuels, and growing concern for carbon emissions and global warming have led to increased grid penetration of renewable energy systems (RES) [2, 3]. With the consequent government policies for curbing carbon emissions, there has been a surge in the deployment of solar photovoltaic and wind energy-based RES [4]. Besides offering the merit of reduced carbon emissions, the integration of RESs with the existing power system offers advantages such as efficient transmission and distribution, reduction in environmental pollution and overall cost of generation, improved power reliability

and enhanced voltage profile [5].

Microgrids are receiving much attention as they facilitate the integration of RES and distributed generation. Microgrid involves the localised generation, management, control, and energy supply to local loads [6]. To increase the reliability of the microgrid, energy storage elements are also included. Various works of literature have reported the different concepts of microgrids, such as microgrid power management [7–9], microgrid control [10], and selection of the size and location of renewable energy sources used in microgrids [11]. With the microgrids being part of the distribution system, which is open to the atmosphere, there is more probability of experiencing different types of electrical faults such as a line to ground (LG) fault, line to line (LL) fault, double line to ground (LLG) fault, triple line (LLL) fault, and triple line to ground (LLLG) fault. Suppose these faults are not cleared within the shortest possible time. In that case, severe damage may be caused to the microgrid infrastructure and consumer loads. Furthermore, the reliability of the power supply is hindered. Hence, the protection of the microgrid is of paramount importance.

In traditional grids, the power flows unidirectionally from sources to the load in a radially distributed network. On the contrary, bidirectional power flow is

possible in microgrids. This bidirectional power flow and the variation of short circuit capability –caused due to the intermittent generation of RESs in the microgrid – make the protection of the microgrid more complicated.

1.2. Literature Review

Since the inception of microgrids, the research fraternity has paid due attention to the protection of microgrids. The previous literature has provided several solutions to this problem through different techniques. These protection techniques can be classified into the following six categories.

1. Optimisation technique based on conventional protection schemes [12–17]
2. Artificial intelligence (AI) based protection schemes [5, 18]
3. Protection schemes utilising signal processing and statistical parameters [19-22]
4. Protection schemes utilising signal processing and AI techniques[23–25]
5. Protection schemes utilising signal processing and machine learning techniques(MLT) [26–31]
6. Protection schemes based on miscellaneous methods [32–34]

Saad et al. have used optimisation techniques to find an appropriate protection coordination scheme with minimum operating time while guaranteeing reliability and security. This method does not talk about high impedance faults and the different modes of operation of microgrids [12]. Reference [14] proposes another method for maintaining proper protection coordination that uses dual setting directional over-current relays with distinct settings and low bandwidth communication. This method takes a long time to detect faults with high impedance [14]. A conventional and non-pilot-based protection scheme was developed for inverter-based microgrid protection by Lahiji et al. [16]. This method is network-dependent and requires the deployment of additional relays at the point of common coupling (PCC) and the location of the distributed generator (DG) [16]. A method is reported based on the phase difference between pre-fault and post-fault current components [17]. This method has been analysed for different microgrid configurations, except radial and interconnected [17].

For fault identification, Sadegh Jamali et al. have used wavelet packet transform to obtain features from the fault current and voltage waveforms [19]. The effect of measurement errors is also considered [19]. Reference [23] reported protection of radial network microgrid with a hybrid approach involving discrete wavelet transform and artificial neural network (ANN).

Similarly, Aljohani et al. have also proposed a hybrid approach involving Stockwell transform and MLT for detection, classification and location of LG faults only [26]. A fault index computed using the Wigner distribution, and Alienation indices was introduced in [20] for fault identification. This method was implemented for a microgrid involving a single DG [20]. A travelling wave-based protection scheme that utilises an initial current travelling wave and an improved mathematical morphology was also reported [15]. Dhivya et al. considered a fast recursive discrete Fourier transform (DFT) for efficient fundamental tracking of varying power system signals [24]. Another interesting approach involves the DFT of line currents to extract the necessary statistical features, which are then processed by deep neural networks to obtain the fault information [25]. This approach does not perform the accuracy and detailed classification [25]. Convolutional neural network-based fault classification is reported in [5], where no data pre-processing is needed. Reference [29] has presented logistic regression-based and AdaBoost-based fault classification with Hilbert transform-based feature extraction. An intelligent protection scheme using a combined wavelet transform and decision tree (DT) is also proposed [27]. Susmita Kar et al. had introduced a data-mining-based intelligent differential protection scheme for microgrids, wherein SVM and DT are used for fault detection [30].

Besides these methods, the communication-free dual setting-based method for microgrid [35], inverter control-based method [36], a combination of Hilbert Huang Transform and extreme learning machine [31], and traditional method for relay settings [37] are very recently reported for microgrid protection. Adaptive protection for grid-connected and islanded microgrids is proposed in [38]. The adaptive protection needs the updated relay setting when the power output of RES changes. Baloch et al. have developed a new protection strategy based on the auto-correlation of three-phase current envelops for grid-connected and islanded microgrids [22]. Reference [39] reported a novel discrete wavelet Transform (DWT) based probabilistic generative model for protecting the microgrid, which has been tested for a relatively less number of the total fault and non-fault scenarios [39]. A protection method based on a multi-resolution analysis of the DWT and a Taguchi-based ANN is presented by Hong et al. This work concentrates only on the static switch present at the PCC of a grid-connected microgrid [18]. Another work proposed a method based on the corresponding

sum of the current samples obtained after every half-cycle [21]. Nsengiyaremye et al. proposed a new method based on a differential scheme using low-cost communication systems [40]. The islanded mode or grid-connected mode of the microgrid is not studied in the system. The effect of noise is also not considered in the method [40]. Another work suggests a new microgrid protection scheme based on Teager–Kaiser energy operator approach. This method uses the change in the value of the sum of squares of three-phase currents to detect the fault. Selecting the values of the indices is rather difficult in dynamic variations in the microgrid operation [41].

Table 1. Limitations of existing literature in microgrid protection

Reference	Limitation
[18]	Faults in utility networks are only considered in the study, and faults in the microgrid are not considered
[21]	Threshold selection is rather difficult
[22]	Quantification of results in terms of regularly used accuracy, precision, and F1-score is not presented.
[29]	Limited fault and non-fault scenarios are considered
[30]	The presence of noise in the data-set is not considered in the analysis
[31]	A limited number of fault and non-fault scenarios were considered in the study
[35]	Discussion on the variations in short circuit rating due to the intermittent nature of renewable energy sources is not presented
[36]	This work focuses only on the islanded microgrid, and this scheme requires the injection of fifth harmonic currents into the system to have a higher fault current similar to the traditional system.
[37]	Only islanded mode is considered
[38]	Adaptive protection needs an updated relay setting when the generation changes
[39]	A limited number of fault and non-fault scenarios were considered in the study
[40]	A microgrid is analysed in this work. However, whether the analysis is related to grid-tied or islanded mode is not mentioned.
[41]	The proposed method requires the selection of indices, which is rather difficult under the dynamic variations in the microgrid operation

The above-mentioned protection strategies employed different protection ideas to solve the protection problems in the microgrid. However, each of the schemes has some limitations. Most of the protection strategies have not considered the effect of measurement noise in their work. Some schemes suffer from higher fault detection times. Some methods have used fewer fault and non-fault scenarios, while others have focused only on one mode of the microgrid. The adaptive methods experience a high computational burden due to the complex fault calculations for relay settings. Therefore, this manuscript proposes a new scheme for microgrid protection using the Support Vector Machine (SVM) [42] classifier-based data mining model for fault detection and fault classification to overcome these limitations. Time series, frequency series, THD, and derived differential parameters are

computed to build a data mining model to train the SVM-based fault detection and classification models. PSCAD/EMTDC [43] is used for basic data generation, whereas Python [44] is used for feature extractions, training, and testing the data mining model. Table 1 states the scope for improvement in the earlier reported literature. The scope for improvement is treated as the research gap, and an SVM classifier is developed to overcome the limitations associated with the earlier reported work. The contribution of the proposed SVM classifier is given in the next sub-section.

1.3. Major contributions of this research

- A simpler and more efficient data mining-based fault detection and classification method for the microgrid is proposed in this paper.
- Three-phase line currents and three-phase line to ground voltages on both ends of the distribution line (DL) are sampled based on which the time series, frequency series, THD, and differential features of voltages and currents are computed.
- A data mining model is developed for the collected data set. The mined data set is further utilised for training the SVM classifier to detect and classify faults.
- The performance of the developed SVM classifier is investigated with different measurement errors by adding noise with 30 dB SNR, 35 dB SNR, and 40 dB SNR. This is done to demonstrate effective classification even under measurement noise.
- Due to the extensive use of parameters, such as time series, frequency series, THD and differential parameters, the proposed data mining based SVM scheme provides more accurate detection and classification of faults, which can further enhance the protection of microgrid against different faults.
- The developed SVM model can easily adapt to multiple operating modes (grid-connected and standalone modes) and network topologies (radial and interconnected configurations) without any need for change in the algorithm. Additionally, the developed SVM model can handle power dynamics associated with conventional synchronous generators (small hydropower plants and diesel generators, etc.) and inverter integrated DG (Solar PV, Wind, etc.) can be handled.

The remaining part of this paper is organised as follows: Section 2 explains the methodology employed for developing the proposed SVM model. Section 3 describes the simulation of the IEC microgrid in PSCAD/ EMTDC and the process of data generation by

simulating fault and non-fault scenarios. Section 4 describes the implementation and the training procedure of the proposed model. Section 5 provides the results and discussions. Finally, section 6 reveals the conclusions and recommendations.

2. THE METHODOLOGY USED FOR THE PROTECTION OF MICROGRID

2.1. Proposed method

This work proposes a data mining model based on differential microgrid fault classification using SVM considering voltage and current distortions. Noise is also considered in the analysis presented. This robust scheme for the protection of microgrid can effectively detect and classify faults irrespective of its mode of operation (islanded mode or grid-connected mode) and configuration (radial configuration or interconnected configuration). The input features for the developed SVM model are the time series, frequency series, THD parameters, and differential parameters. The data of the latest thirty-two samples of three-phase line currents and three-phase line to ground voltages on both ends of the DL to be protected are required to compute these features. A moving window of thirty-two samples stores the required voltages and currents to compute these input parameters. The sampling frequency is selected such that the data over an entire cycle can fit into the moving window. The data in the moving window is used to calculate the above-mentioned features. The calculated features are applied to the trained model at a frequency equal to the sampling frequency of 32 samples per power cycle, i.e. 1920 Hz. Figure 1 shows the methodology adopted in this paper for tripping the circuit breaker after the fault. An extensive data set consisting of fault, and non-fault scenarios is needed to train the proposed method. The IEC microgrid model considered in this study has four DGs and five DLs. The data is generated by simulating different types of faults at a location on a particular DL by varying fault resistances and fault inception angles. The data is gathered for different fault locations on DL while considering various fault resistances and fault inception angles. This process is similarly repeated for faults in all other DL. The fault data is also generated for different modes and configurations of the microgrid. The non-faults are also simulated by varying the loads available in the microgrid in all the different modes and configurations. The following sub-section discusses the mathematical formulations for computing the time and frequency series, THD and differential parameters.
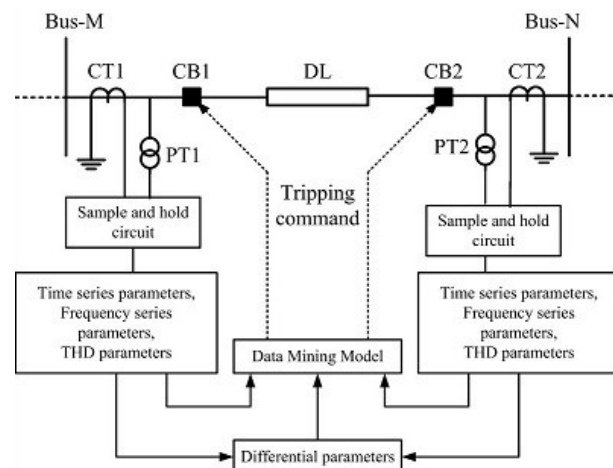

Fig. 1. The methodology adopted in this paper for tripping the circuit breakers after the fault

The currents and voltages are continuously fed at 32 samples per power cycle. As shown in Figure 2, the acquired samples of three-phase current and voltages at both the ends of DL are stored in a moving window having a length of thirty-two. This window is updated at each sampling instant. Based on the stored data, the time series, frequency series, differential, and THD parameters are calculated recursively using full-cycle data of the latest thirty-two data samples. The calculated parameters are applied as features to the already trained MLT-based fault identification and classification models. From the eleven possible outcomes (No-fault/AG/BG/CG/ABG/BCG/ACG/ABCG/AB/AC/BC), the model identifies one based on the input features. The identified outcome is provided as the output of the model.

2.2. Feature extraction

Extraction of features starts with the measurement and sampling of the three-phase currents and voltages on both sides of DL. Let the DL be connected between the bus M and N of the microgrid. The three-phase instantaneous currents and instantaneous voltages measured at bus M can be denoted as i_{aM}, i_{bM}, i_{cM}, v_{aM}, v_{bM}, v_{cM}, respectively. Similarly, the three-phase instantaneous currents and instantaneous voltages measured at bus N can be denoted as i_{aN}, i_{bN}, i_{cN}, v_{aN}, v_{bN}, and v_{cN}, respectively. The extraction of different input parameters for training and testing of the proposed SVM method is shown in Figure 3.

2.2.1. Time series parameters

Let S be the moving window consisting of the most recent thirty-two sampled values of the instantaneous values of current and voltages as shown in (1), where x denotes the sampled value, and the sub-script depends on the number of sample. The time-series parameters such as the minimum, maximum, average, peak-to-peak

and rms values are calculated based on S as shown in Eqns. (2)-(6).

$$S = \{x_1, x_2, x_3, \ldots\ldots\ldots, x_{31}, x_{32}\} \tag{1}$$

$$X_{min} = \min\{S\} \tag{2}$$

$$X_{max} = \max\{S\} \tag{3}$$

$$X_{pp} = |X_{max} - X_{min}| \tag{4}$$

$$X_{avg} = \frac{\sum_{i=1}^{32} x_i}{32} \tag{5}$$

$$X_{rms} = \sqrt{\frac{\sum_{i=1}^{32} x_i^2}{32}} \tag{6}$$

2.2.2. Frequency Series Parameters and THD Parameters

The frequency series parameters involve computation of the peak magnitudes of the fundamental and the nine most dominant harmonics using DFT. With DFT, the peak magnitude of the *ac* component with frequency *k* times the fundamental magnitude can be expressed as in (7), where, *k* is an integer representing the ratio of the frequency of the *ac* signal to the fundamental, *N* is the number of samples in one cycle (i.e. 32 in this case), *i* is the index corresponding to a particular sample, *and j* is the imaginary operator.

$$X_k = \frac{2}{N} \sum_{i=0}^{N-1} x_i \, e^{-\frac{j2\pi}{N}ik} \tag{7}$$

Based on Eq. (7), the THD parameters can be calculated as

$$X_{THD} = \frac{1}{X_1} \sqrt{\sum_{I=2,3,4}^{10} X_I^2} \tag{8}$$

All the time series, frequency series, and THD parameters are calculated individually for all the three-phase voltages and currents on both sides of DL.

2.2.3. Differential Parameters

The differential parameters are calculated using all the voltages and currents at bus *M* and *N*. By using three-phase voltages and three-phase currents at bus *M* (say sending end bus), the three-phase active power (P_M), three-phase reactive power (Q_M), positive sequence voltage (V_{1M}), negative sequence voltage (V_{2M}), zero-sequence voltage (V_{0M}), positive sequence current (I_{1M}), negative sequence current (I_{2M}) and zero sequence current (I_{0M}) are calculated. Similarly, by using three-phase voltages and currents at bus-N (say Receiving end), P_N, Q_N, V_{1N}, V_{2N}, V_{0N}, I_{1N}, I_{2N}, I_{0N} are calculated. Then the differential parameters are calculated by taking the absolute difference between the respective parameter determined based on the measurements at bus *M* and *N*. Figure 3 shows the block diagram representation of the computation of time series, frequency series and differential parameters.

2.3. Relation between SVM and THD along with other parameters

The developed SVM model utilises THD of voltage and

current as one of the input features. Nowadays, there is a tremendous increment in the usage of non-linear loads. These non-linear loads draw distorted currents from the grid, which leads to current and voltage harmonics in the distribution network. Traditionally, the short circuit analysis neglects the comparison of pre-fault and fault currents as the system mainly consists of synchronous machines.
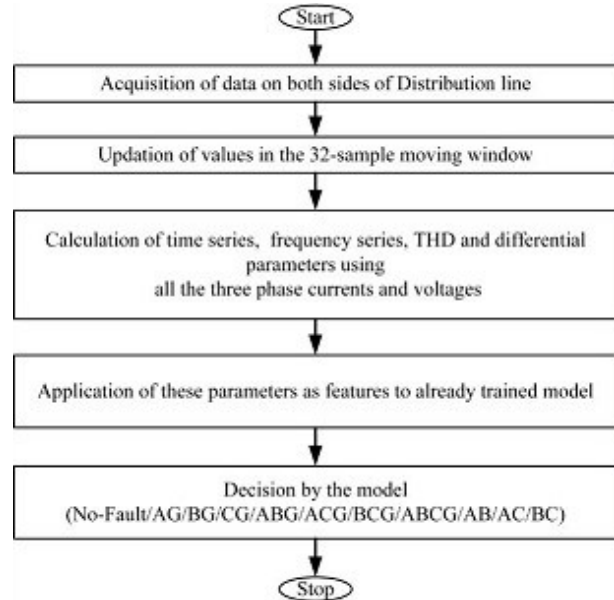

Fig. 2. The proposed moving window of thirty-two samples for calculating the features applied to the already trained model
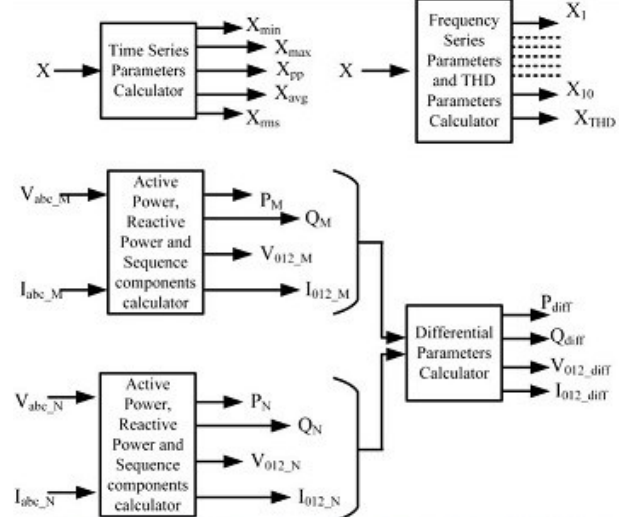

Fig. 3. Block diagram for the calculation of time series, frequency series and differential parameters

Nevertheless, in the inverter-based renewable energy systems, the fault current magnitude is 1.5 to 2 times the rated current. Hence, the pre-fault current compared to the fault current cannot be neglected. Moreover, the decaying *dc* component in the fault current will produce *dc* offset, and the non-linear resistance of the arc will produce harmonics in the fault currents, which lead to distortions in the fault voltage. The harmonic analysis of the healthy currents and voltages under the normal

operating condition will differ from those under the fault state. Along with these THD parameters, the time series, frequency series and differential parameters also have different values for fault and non-fault conditions. SVM classifies the data based on the hyper plane. The higher the difference between the values of the parameters among the different classes, the higher the possibility of forming the hyperplane that can classify the different faults.

3. THE SYSTEM UNDER STUDY AND DATA GENERATION

3.1. Test System
To simulate the fault and non-fault scenarios, an IEC microgrid [27], shown in Figure 4, is studied in this paper. The microgrid consists of six buses, four DGs, five DLs, six loads, and four transformers. The microgrid is connected with the power system grid at PCC through a circuit breaker. If the circuit breaker at the PCC is closed, then the micro-grid operates in grid-connected mode; otherwise, it operates in island mode. The resulting microgrid configuration is radial if the Loop-1 circuit breaker and Loop-2 circuit breaker are open. The resulting microgrid configuration is interconnected if these circuit breakers are closed. The total load on the microgrid is rated at 22 MW and 10 MVAR. In Figure 3, the length of each DL is assumed to be 30 km. The microgrid is connected to the main utility grid through a 120/25 kV, 60 Hz transformer. The utility is considered to have a short circuit MVA rating of 1000 MVA.

3.2. Data Generation Required for Model Training and Testing
Quality of data is of utmost importance for MLT or any other AI method. The data required for successful fault identification and classification in this study is generated by simulating all possible faults in the microgrid, shown in Figure 4, for (i) interconnected microgrid connected to the grid, (ii) radial microgrid connected to the grid, (iii) interconnected microgrid operating in islanded mode, and (iv) radial microgrid operating in islanded mode. Data for each of these four cases is obtained by simulating the respective microgrid configuration without any fault. Then each of the four cases is simulated with A phase to Ground fault (AG), B phase to Ground (BG), C phase to Ground (CG), A phase to B phase to Ground (ABG), A phase to C phase to Ground (ACG), B phase to C phase to Ground (BCG), A phase to B phase to C phase to Ground (ABCG), A phase to B phase (AB), A phase to C phase (AC) and B phase to C phase (BC).

Different faults are analysed with the location of fault occurrence on DL-1 is considered 3km from Bus-1. Also, while simulating each type of fault at a particular location, the fault resistance is varied from 0.01 Ω to 100 Ω. Then for each value of fault resistance, the fault inception angle is changed from 0^0 deg to 180^0 deg in steps of 22.5^0 deg. This procedure is repeated for the fault location of 6km, 9km, 12km, 15km, 18 km, 21km, 24km, and 27km distances from Bus-1 on the DL1. A similar procedure is followed for DL-2, DL-3, DL-4, and DL-5. Thus, for the four cases, this procedure is repeated. This results in data being collected for 210600 fault scenarios. This is summarised in Table 2. Similarly, the non-fault scenarios are also simulated by changing the load (both active and reactive powers) on each bus from 80% to 120 % of the nominal load in steps of 5%. The total non-fault scenarios considered for data generation are 21220. The details of the load variation while simulating the non-fault scenarios are provided in Table 3. Thus, an extensive data set is available to train the SVM classifier. For each fault and non-fault scenario, the data is collected and made available over eight power cycles, i.e., 0.133 sec. The selected sampling frequency of 1920Hz results in the acquisition of 32 samples over a power cycle.

4. SYSTEM IMPLEMENTATION AND TRAINING OF THE PROPOSED MODEL
The fault and non-fault scenarios in the IEC micro gird were simulated in PSCAD/EMTDC software. The generated data from this software is saved as EMTDC output files. All these EMTDC output files are combined and converted to 'CSV' files with the help of MS-EXCEL. These CSV files are exported to Python software for further signal processing and noise analysis and the application of different MLT. All the simulations, signal processing, noise analysis, training, and testing of different MLT are performed in a laptop PC having an Intel core i3-4005U CPU@1.70GHz processor with 4GB RAM.

Table 2. Variation of Parameters for simulation of fault scenarios

| Parameter	Possible values	Count
Operating Mode	Grid Connected mode, Islanded mode	2
Configuration	Radial, Loop or Interconnected	2
Fault line	DL1, DL2, DL3, DL4, DL5	5
Fault Location from sending end (km)	3, 6, 9, 12, 15, 18, 21, 24, 27	9
Fault Inception Angle (degree)	0, 22.5, 45, 67.5, 90, 112.5, 135, 157.5, 180	9
Fault Resistance	0.01 Ω, 0.1 Ω, 1 Ω, 10 Ω, 20 Ω, 30 Ω, 40 Ω, 50 Ω, 60 Ω, 70 Ω, 80 Ω, 90 Ω, 100 Ω	13
Fault Type	No-fault, AG, BG, CG, ABG, ACG, BCG, ABCG, AB, AC and BC	11
Total fault scenarios (2 X 2 X 5 X 9 X 9 X 13 X 10) = 210600		
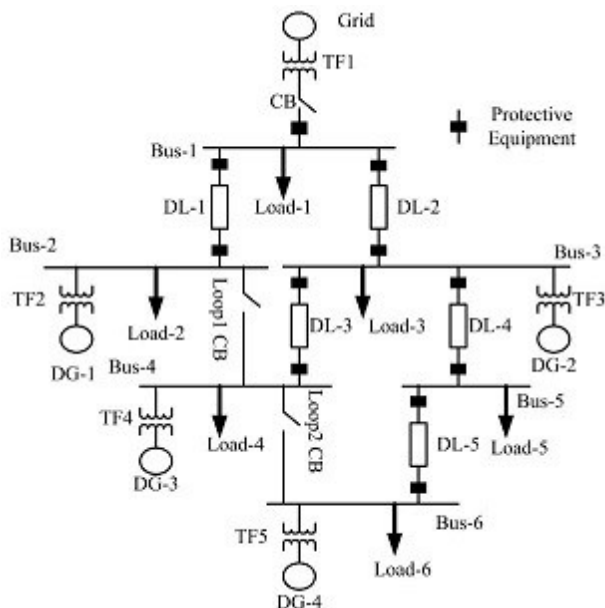
Fig. 4. IEC microgrid used for simulation of the fault and non-fault scenarios to create the feature set for training of the proposed model [30]
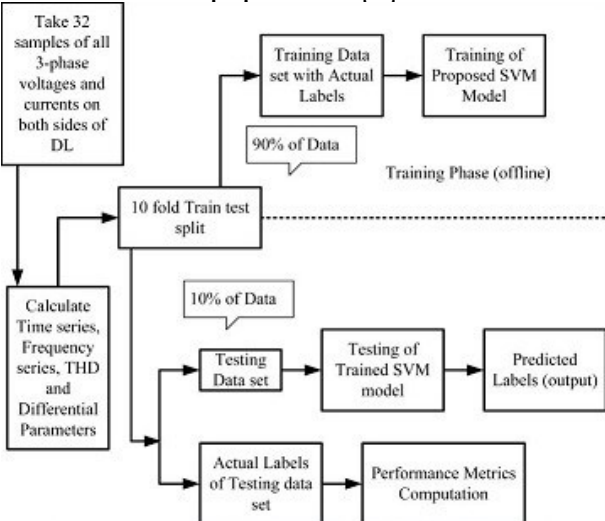

Fig. 5. The method of training and testing the proposed technique in one fold of k-fold cross-validation

Table 3. Variation of parameters for simulation of non-fault scenarios operation

Parameter	Possible values	Count
Operating Mode	Grid-Connected mode, Islanded mode	2
Configuration	Radial, Loop or Interconnected	2
Simultaneous load change in all the loads w. r. t base load	+5%, +10%, +15%, +20%, -5%, -10%, -15%, -20%	8
No. of distribution lines	DL1, DL2, DL3, DL4, DL5	5

Non-fault scenarios due to load variation (2 x 2 x 8 x 5) = 160
Non-fault scenarios from Table 1 (2 X 2 X 5 X 9 X 9 X 13 X 1) = 21060
total non-fault scenarios = 21220

Instead of the usual train/test split, a ten-fold cross-validation method is applied in this study to validate the identification and classification capabilities of proposed method. This ten-fold cross-validation method divides the entire data set (all differential, time-series, frequency series, and THD parameters) into ten random equal parts.

In the first fold of validation, as shown in Figure 5, the first part of the data is reserved for testing, and the remaining nine parts (i.e., 2 to 9) are used for MLT training. In the second fold of validation, the second part of the data is reserved for testing. The remaining nine parts (i.e., 1, 3 to 9) are used to train the MLTs. This process is repeated until all parts of the data are utilised for the testing, with the remaining parts used as training data.The properly trained model can be installed in the computers of the microgrid operator or microgrid control center.

5. RESULTS AND DISCUSSIONS

In this work, eleven classes (No-fault, AG, BG, CG, ABG, ACG, BCG, ABCG, AB, AC, and BC) are considered. The classification capabilities of the proposed method are expressed using classification metrics called accuracy, precision, F1-score, and recall. Calculation of these metrics requires the quantitative values of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The confusion matrix can obtain quantitative TP, TN, FP, and FN values in a multi-class classification problem. TP is the total number of scenarios or instances in which the estimated or predicted class is the same as the actual class. FP is the total number of instances in which a class, C1 is predicted as one of the remaining C–1 classes, where C is the total number of classes. FN is the total number of instances in which the classes that belong to C–1 classes are predicated as class C1. TN is the total number of instances in which the classes belonging to C–1 classes are predicted as classes belonging to C–1 classes. As already mentioned, the evaluation is carried out using a ten-fold cross-validation procedure. The evaluation metrics range from 0 to 1.0, where 0.0 indicates the worst performance and 1.0 indicates the best possible performance.

The output parameters of the confusion metrics are TP, TN, FP and FN, which are utilised for computing accuracy, F-1 precision, and recall, as shown in (9)-(12). Here, accuracy is the measure of how many positive and negative predictions are correctly classified by the algorithm. Accuracy is defined as the ratio of total correct predictions to the total number of predictions. F1-score is the single metric that quantifies the effect of both precision and recall. F1- score is defined as the harmonic mean of recall and precision. Precision measures how many positive predictions are actually

positive. Precision is defined as the ratio of the total number of correctly predicted positives to the total number of predicted positives. Recall measures how many positive data points are correctly predicted. The recall is the ratio of the total number of correctly predicted positives to the total number of actual positives.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (9)$$

$$F1\ score = \frac{2X(Precision\ X\ Recall)}{Precision+Recall} \qquad (10)$$

$$Precision = \frac{TP}{TP+FP} \qquad (11)$$

$$Recall = \frac{TP}{TP+FN} \qquad (12)$$

5.1. Fault Classification without any noise by the proposed SVM method

This subsection discusses kernel function and hyperparameters used in the SVM, and the classification results for the actual generated data using the proposed method. SVM is a popular classification method used for classifying different classes present in the data [42]. SVM method uses the concept of a hyperplane to draw the boundary between the classes [42]. Hyperplane provides the maximum possible gap between the border candidates of each class [42]. Besides binary classification problems, the SVM algorithm can also classify the data with multiple classes by converting the multi-class problem into a binary classification problem. The outcome of this conversion is the division of multiple classes into two classes, M1 and M2. For C1, C2, ….., CN classes, M1 represents C1 and M2 represents C2 to CN after conversion.

In SVM, the kernel draws a clear boundary between the classes by projecting the features into a different plane. This work uses the radial basis function (RBF) as the kernel function because of its localised nature and finite response along the x-axis. The RBF kernel used in this work is as given as

$$K(x_i,\ x_j) = \exp(-\gamma||x_i - x_j||^2), \gamma > 0 \qquad (13)$$

Where, $K(x_i,\ x_j)$ is the output of the kernel, and γ as the hyperparameter considered as 0.1.

As mentioned earlier, ten-fold cross-validation is used for the performance evaluation of the proposed method. The performance of the proposed method for each fold and the average performance over the ten folds are stated in Table 4. The data set used for the performance computation of the proposed SVM model does not consider any noise. It is clear from Table 4, that the folds 5, 6, 7, and 9 can classify the faults with 100 % accuracy. The accuracy for the remaining folds is at least 99.99871%. From this figure, it can be inferred

that the SVM can classify faults accurately. The overall accuracy of the proposed method is found to be 99.9948%, which is highly desirable for fault classification.

5.2. Fault Classification Results with distorted data (with noise) by the proposed SVM method

In order to test the robustness of the proposed method under noisy measurements, white noise is added to the measured voltage and current samples. With these distorted voltages and currents, the features are calculated in the same manner as that for the data without any noise. Noise with Gaussian distribution and SNR noise of 30 dB, 35 dB, and 40 dB is added to the actual measured data and then performance evaluation is carried out.

Table 4. Fold wise classification performance indices for fault classification by using SVM for data set without any noise

Fold	Accuracy	F1 Score	Precision	Recall
1	0.999957	0.999957	0.999957	0.999957
2	0.999871	0.999871	0.999871	0.999871
3	0.999957	0.999957	0.999957	0.999957
4	0.999957	0.999957	0.999957	0.999957
5	1	1	1	1
6	1	1	1	1
7	1	1	1	1
8	0.999784	0.999784	0.999785	0.999784
9	1	1	1	1
10	0.999957	0.999957	0.999957	0.999957
Avg	0.999948	0.999948	0.999948	0.999948

Table 5. Fold wise classification performance indices for fault classification by using SVM for data set with 30 dB noise

Fold	Accuracy	F1 Score	Precision	Recall
1	0.999957	0.999957	0.999957	0.999957
2	0.999914	0.999914	0.999914	0.999914
3	0.999957	0.999957	0.999957	0.999957
4	1	1	1	1
5	0.999957	0.999957	0.999957	0.999957
6	1	1	1	1
7	0.999784	0.999784	0.999785	0.999784
8	0.999827	0.999827	0.999828	0.999827
9	0.999871	0.999871	0.999871	0.999871
10	0.999871	0.999871	0.999871	0.999871
Avg	0.999914	0.999914	0.999914	0.999914

Table 6. Fold wise classification performance indices for fault classification by using SVM for data set with 35 dB noise

Fold	Accuracy	F1 Score	Precision	Recall
1	0.999914	0.999914	0.999914	0.999914
2	0.999827	0.999827	0.999828	0.999827
3	0.999957	0.999957	0.999957	0.999957
4	0.999914	0.999914	0.999914	0.999914
5	1	1	1	1
6	0.999871	0.999871	0.999871	0.999871
7	0.999914	0.999914	0.999914	0.999914
8	0.999827	0.999827	0.999828	0.999827
9	1	1	1	1
10	0.999957	0.999957	0.999957	0.999957
Avg	0.999918	0.999918	0.999918	0.999918

The performance of the proposed method for each fold and the average performance over the ten folds while considering noise with SNR of 30 dB, 35 dB and 40 dB in the input data are stated in Tables 5, 6 and 7, respectively. From these three tables, it is observed that

only two out of 10 folds produced 100 % accuracy in each case. The overall accuracy of the prosed method in the most polluted environment, i.e., 30 dB noise level, is 99.9914 %. With 35 dB noise levels and 40 dB noise levels, the overall accuracies for the proposed method are 99.9918 % and 99.9935 %, respectively. This demonstrates that the proposed SVM model can classify the faults even when the input data is corrupted with noise.

Table 7. Fold wise classification performance indices for fault classification by using SVM for data set with 40 dB noise

Fold	Accuracy	F1 Score	Precision	Recall
1	0.999914	0.999914	0.999914	0.999914
2	1	1	1	1
3	0.999957	0.999957	0.999957	0.999957
4	0.999914	0.999914	0.999914	0.999914
5	1	1	1	1
6	0.999957	0.999957	0.999957	0.999957
7	0.999914	0.999914	0.999914	0.999914
8	0.999827	0.999827	0.999828	0.999827
9	0.999914	0.999914	0.999914	0.999914
10	0.999957	0.999957	0.999957	0.999957
Avg	0.999935	0.999935	0.999935	0.999935

Table 8. Comparison of the performance of the fault classification model of the proposed method with other popular algorithms

Noise	Accuracy	F1 Score	Precision	Recall	Algorithm
No noise	0.783151	0.777388	0.784194	0.783151	ABC
30 dB	0.730791	0.72989	0.740443	0.730791	ABC
35 dB	0.641575	0.621429	0.646285	0.641575	ABC
40 dB	0.860042	0.859768	0.871698	0.860042	ABC
No noise	0.999236	0.999236	0.999237	0.999236	DT
30 dB	0.987922	0.987923	0.98793	0.987922	DT
35 dB	0.991153	0.991153	0.99116	0.991153	DT
40 dB	0.995203	0.995203	0.995206	0.995203	DT
No noise	0.700556	0.700022	0.793125	0.700556	GNB
30 dB	0.678134	0.67385	0.772852	0.678134	GNB
35 dB	0.688879	0.682995	0.775827	0.688879	GNB
40 dB	0.686369	0.682698	0.7766	0.686369	GNB
No noise	0.999638	0.999638	0.999638	0.999638	KNN
30 dB	0.873354	0.875816	0.888708	0.873354	KNN
35 dB	0.953408	0.953541	0.955323	0.953408	KNN
40 dB	0.988832	0.988841	0.988973	0.988832	KNN
No noise	0.666539	0.645958	0.714193	0.666539	RF
30 dB	0.755655	0.749317	0.764528	0.755655	RF
35 dB	0.760241	0.753039	0.770742	0.760241	RF
40 dB	0.789975	0.786445	0.803602	0.789975	RF

5.3. Comparison of the performance of the proposed method with other Machine Learning Algorithms

Table 8 presents the evaluation of different classification algorithms in terms of the average of ten-fold values of performance indices for the data set without any noise and data set with 30, 35 and 40 dB noise levels. The classification algorithms considered for this evaluation are such as Adaptive Boosting Classifier (ABC), Decision Tree (DT), Gaussian Naive Bayes (GNB), k-Nearest Neighbours (kNN), Random Forest (RF) and SVM. Out of the stated algorithms, the DT algorithm's performance is somewhat near to the performance of SVM with a classification accuracy of at least 98.7922%, which is reported for the noise level of 30 dB. ABC algorithm performs well in 40 dB noise data

with an accuracy of 86.0042%. However, its performance is worse in the 35 dB noise level with an accuracy of 64.1575%.

Table 9. Performance comparison of the proposed fault classification model with methods reported in the literature

Parameter	Proposed Method	The method reported in Reference					
		[25]	[27]	[29]	[30]	[31]	[45]
Grid-connected mode	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Islanded mode	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Radial configuration	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Interconnected Configuration	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Consideration of noise Levels	30, 35 and 40 dB	30, 35 and 40 dB	NA	30 and 40 dB	NA	NA	NA
Considered Classes for classification	NF, AG, BG, CG, ABG, ACG, BCG, ABCG, AB, AC and BC	NF, AG, BG, CG, ABG, ACG, BCG, ABCG, AB, AC and BC	LG, LLG, LL, LLLG	NF, LG, LLG, LL, LLLG, LLL	NF, LG, LLG, LL, LLLG, LLL	NF, AG, BG, CG, ABG, ACG, BCG, ABCG, AB, AC and BC	NF, AG, BG, CG, ABG, ACG, BCG, ABCG, AB, AC and BC
Average Accuracy (percent)	99.99	99.31	99	97.92	99.45	95.02	95.63
Number of fault and non-fault scenarios	231820	47168	3904	944	1435	794	352

The performance of the GNB algorithm is inferior in the case of data with a 30 dB noise level with an accuracy of 67.8134 %. Even in the case of data without any noise, also it has an accuracy of only 70.0556 %. kNN performs relatively well in the case of data without any noise and data with 35 and 40 dB noise levels with classification accuracies of 99.9638 %, 95.3408 %, and 98.8832 %, respectively. Nevertheless, in the case of data with a 30 dB noise level, its performance is rather poor, with a classification accuracy of 87.3354 %. The worst performer is the RF algorithm because it offers a classification accuracy of 66.6539%, even when the data without any noise is considered.

5.4. Comparison of the performance of the proposed method with existing literature

The qualitative comparison of the proposed fault classification and detection scheme with the earlier reported schemes is shown in Table 9, where NF represents no-fault scenario.

The proposed SVM-based fault classification and detection scheme has been analysed for grid-tied and islanded mode, input data set with and without noise, and all possible types of fault and non-fault scenarios. The developed SVM algorithm can accurately detect and classify the fault regardless of the operating scenario. The earlier reported schemes either concentrate on a particular type of microgrid

configuration or type/s of faults only. The accuracy of the classification is higher than the earlier reported schemes. Moreover, the number of fault and non-fault scenarios considered in the study is also higher than the existing schemes.

6. CONCLUSIONS

The proposed method uses a differential, time series, frequency series, and THD parameters for fault identification and classification for microgrid protection. The proposed method is tested on a standard IEC microgrid model. All possible faults are simulated with the variation of fault resistance, fault inception angle, and fault location along the distribution line. Non-fault cases are also simulated with a variation in loading from −80% to +120% of rated values. The fault and non-fault scenarios are implemented in both grid-connected microgrids with radial and interconnected configurations and islanded microgrids with radial and interconnected configurations. The results demonstrated that the SVM-based fault classification model has an accuracy of 99.9918% in the case of data with 30 dB noise case- the most polluted data set. The classification accuracy with the proposed algorithm is far better than that achieved with ABC, DT, GNB, kNN, and RF algorithms irrespective of whether the fault has occurred, type of fault, fault location, fault resistance, or angle of inception. The proposed SVM classifier outperforms the other algorithms with 30 dB, 35 dB, and 40 dB noise levels. This is also demonstrated through the comparative analysis presented.

REFERENCES

[1] T. Ahmad and D. Zhang, "A critical review of comparative global historical energy consumption and future demand: The story told so far", *Energy Rep.*, vol. 6, pp. 1973–1991, 2020.

[2] K. J. Warner and G. A. Jones, "The 21st century coal question: China, India, development, and climate change", *Atmosphere*, vol. 10, pp. 476, 2019.

[3] A. V. Sant, V. Khadkikar, W. Xiao, H. Zeineldin and A Al-Hinai, "Adaptive control of grid connected photo voltaic inverter for maximum VA utilisation", *IECON 2013-39th Annual Conf. IEEE Ind. Electron. Soc.*, 2013.

[4] R. Kempener, E. Assoumou, A. Chiodi, U. Ciorba, M. Gaeta, D. Gielen, H. Hamasaki, A. Kanudia, T. Kober, M. Labriet, et al, "A global renewable energy roadmap: Comparing energy systems models with irena's remap 2030 project", *Inf. Energy Climate Policies Using Energy Syst. Models*, 2015.

[5] P. Rai, N. D. Londhe and R. Raj, "Fault classification in power system distribution network integrated with distributed generators using CNN", *Electr. Power Syst. Res.*, vol. 192, pp. 106914, 2021.

[6] H. S. Bhalja, A.V. Sant, A. Markana and B. R. Bhalja, "Microgrid with five-level diode clamped inverter based hybrid generation system", *2019 IEEE Int. Conf. Electr.*

Comput. Communication Tech., 2019.

[7] A. H. Einaddin, A. S. Yazdankhah and R Kazemzadeh, "Power management in a utility connected micro-grid with multiple renewable energy sources", *J. Oper. Autom. Power Eng.*, vol. 5, pp. 1–10, 2017.

[8] M. Zolfaghari, G. B. Gharehpetian and M. Abedi, "A repetitive control-based approach for power sharing among boost converters in dc microgrids", *J. Oper. Autom. Power Eng.*, vol. 7, pp. 168–175, 2019.

[9] G.R. Aghajani and I. Heydari, "Energy management in microgrids containing electric vehicles and renewable energy sources considering demand response", *J. Oper. Autom. Power Eng.*, vol. 9, pp. 34–48, 2021.

[10] F. S. Zavareh, E. Rokrok, J. Soltani and M.R. Shahkarami, "Adaptive sliding mode control of multi-dg, multi-bus grid connected microgrid", *J. Oper. Autom. Power Eng.*, vol. 7, pp. 65–77, 2019.

[11] S. Ghobadpour, M. Gandomkar and J. Nikoukar, "Multi-objective function optimisation for locating and sizing of distributed generation sources in radial distribution networks with fuse and recloser protection". *J. Oper. Autom. Power Eng.*, vol. 9, pp. 266–273, 2021.

[12] 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. Transm. Distrib.*, vol. 13, pp. 780–793, 2019.

[13] S. M. Saad, N. El-Naily and F. A. Mohamed, "A new constraint considering maximum psm of industrial over-current relays to enhance the performance of the optimisation techniques for microgrid protection schemes", *Sustain. Cities Soc.*, vol. 44, pp. 445–457, 2019.

[14] 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, pp. 143–151, 2016.

[15] I. Xyngi and M. Popov, "An intelligent algorithm for the protection of smart power systems", *IEEE Trans. Smart Grid*, vol. 4, pp. 1541–1548, 2013.

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

[17] 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, pp. 446–53, 2019.

[18] Y. Hong and M. T. A. M. Cabatac, "Fault detection, classification, and location by static switch in microgrids using wavelet transform and taguchi-based artificial neural network", *IEEE Syst. J.,* vol. 14(2), pp. 2725–2735, 2019.

[19] 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(6), pp. e12353, 2020.

[20] S. R. Ola, A. Saraswat, S. K. Goyal, S.K. Jhajharia, B. Khan, O. P. Mahela, H. H. Alhelou and Pierluigi Siano, "A protection scheme for a power system with solar energy penetration", *Appl. Sci.*, vol. 10(4), pp. 1516, 2020.

[21] G. K. Rao, T. Gangwar and S. Sarangi, "Advanced relaying for dg-penetrated distribution system", *Arabian J. Sci. Eng.*, vol. 46(10), pp. 9649–9661, 2021.

[22] S. Baloch, S. Z. Jamali, K. K. Mehmood, S. B. A.

Bukhari, M. S. U. 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. 3(9), pp. 2350, 2020.

[23] 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(14), pp. 4965, 2020.

[24] 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(1), pp. 67–75, 2015.

[25] J.Q. James, 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(2), pp. 1694–1703, 2017.

[26] 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", *2019 Int. Conf. Control, Autom. Diagnosis (ICCAD)*, 2019.

[27] D. P. Mishra, S. R. Samantaray and Geza Joos, "A combined wavelet and data-mining based intelligent protection scheme for microgrid", *IEEE Trans. Smart Grid*, vol. 7(5), pp. 2295–2304, 2015.

[28] M. Shafiullah, M. A. Abido and Z. A.L. Hamouz, "Wavelet-based extreme learning machine for distribution grid fault location", *IET Gener. Transm. Distrib.*, vol. 11(17), pp. 4256–4263, 2017.

[29] S. Baloch and M. S. Muhammad, "An intelligent data mining-based fault detection and classification strategy for microgrid", *IEEE Access*, vol. 9, pp. 22470–22479, 2021.

[30] S. Kar, S.R. Samantaray and M. D. Zadeh, "Data-mining model based intelligent differential microgrid protection scheme", *IEEE Syst. J.*, vol. 11(2), pp. 1161–1169, 2015.

[31] M. Mishra and P. K. Rout, "Detection and classification of micro-grid faults based on HHT and machine learning techniques", *IET Gener. Transm. Distrib.*, vol. 12(2), pp. 388–397, 2018.

[32] 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(5), pp. 2211–2218, 2014.

[33] Y. S. Oh, C.H. Kim, G.H. Gwon, C.H. Noh, S. B. A. Bukhari, R. Haider and T. Gush, "Fault detection scheme based on mathematical morphology in last mile radial low voltage dc distribution networks", *Int. J. Electr. Power Energy Syst.*, vol. 106, pp. 520–527, 2019.

[34] H. A. A. Hassan, A. Reiman, G. F. Reed, Z.H. Mao and B. M. Grainger, "Model-based fault detection of inverter-based microgrids and a mathematical framework to analyse and avoid nuisance tripping and blinding scenarios", *Energies*, vol. 11(8), pp. 2152, 2018.

[35] H. Beder, B. Mohandes, M. S. E. Moursi, E.A. Badran, and M. M. E. Saadawi, "A new communication-free dual setting protection coordination of microgrid", *IEEE Trans. Power Del.*, vol. 36(4), pp. 2446–2458, 2020.

[36] 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(9), pp. 7684–7697, 2017.

[37] H. F. Habib, M. M. Esfahani and O. Mohammed, "Development of protection scheme for active distribution systems with penetration of distributed generation", *SoutheastCon,* 2018.

[38] R. Jain, D. L. Lubkeman and S. M. Lukic," Dynamic adaptive protection for distribution systems in grid-connected and islanded modes", *IEEE Trans, Power Del.*, vol. 34(1), pp. 281–289, 2018.

[39] S.R. Fahim, S. K. Sarker, S.M. Muyeen, M. Sheikh, R. Islam, and S.K.Das, "Microgrid fault detection and classification: Machine learning based approach, comparison, and reviews", *Energies*, vol. 13(13), pp. 3460, 2020.

[40] J. Nsengiyaremye, B. C. Pal and M. M. Begovic, "Microgrid protection using low-cost communication systems" *IEEE Trans. Power Del.*, vol. 35(4), pp. 2011–2020, 2019.

[41] M. A. Jarrahi, H. Samet and T. Ghanbari, "Novel change detection and fault classification scheme for ac microgrids", *IEEE Syst. J.*, vol. 14(3), pp. 3987–3998, 2020.

[42] W. S. Noble, "What is a support vector machine?", Nature *biotechnology*, vol. 24(12), pp. 1565–1567, 2006.

[43] Manitoba HVDC Research Centre. PSCAD/EMTDC, version: 4.5.0.0, 2012.

[44] G.V. Rossum. Python, version: 3.0, 2008.

[45] T. S. Abdelgayed, W. G. Morsi and T. S. Sidhu. "A new approach for fault classification in microgrids using optimal wavelet functions matching pursuit". *IEEE Trans. Smart Grid*, vol. 9(5), pp.4838–4846, 2017.