INL/JOU-22-69343-Revision-0

Ensemble voting-based fault classification and location identification for a distribution system with microgrids using smart meter measurements

October 2022

Md Maidul Islam, Alvi Newaz, Md Omar Faruque, Muhammad Usama Usman

Changing the World's Energy Future

Idaho National Laboratory

INL is a U.S. Department of Energy National Laboratory operated by Battelle Energy Alliance, LLC

INL/JOU-22-69343-Revision-0

Ensemble voting-based fault classification and location identification for a distribution system with microgrids using smart meter measurements

Md Maidul Islam, Alvi Newaz, Md Omar Faruque, Muhammad Usama Usman

October 2022

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-EE0004682

Received: 23 May 2022 | Revised: 5 August 2022 | Accepted: 20 September 2022

IET Smart Grid

DOI: 10.1049/stg2.12091

ORIGINAL RESEARCH

Ensemble voting-based fault classification and location identification for a distribution system with microgrids using smart meter measurements

Md Maidul Islam[1] | Muhammad Usama Usman[2] | Alvi Newaz[1] | Md Omar Faruque[1]

[1]Electrical & Computer Enginering, FSU, Tallahassee, Florida, USA

[2]Idaho National Lab, Idaho falls, Idaho, USA

Correspondence

Md Maidul Islam, Electrical & Computer Enginering, FSU, Tallahassee, USA.
Email: mi19b@fsu.edu

Funding information

Department of Electrical and Computer Engineering, FAMU-FSU College of Engineering

Abstract

This study presents an ensemble learning approach for fault classification and location identification in a smart distribution network containing photovoltaics (PV)-based microgrid. Lack of available data points and the unbalanced nature of the distribution system make fault handling a challenging task for utilities. The proposed method uses event-driven voltage data from smart meters to classify and locate faults. The ensemble voting classifier is composed of three base learners; random forest, k-nearest neighbours, and artificial neural network. The fault location (FL) task has been formulated as a classification problem where the fault type is classified in the first step and based on the fault type, the faulty bus is identified. The method is tested on IEEE-123 bus system modified with added PV-based microgrid along with dynamic loading conditions and varying fault resistances from 0 to 20 Ω for both unbalanced and balanced fault types. A further sensitivity analysis has been done to test the robustness of the proposed method under various noise levels and data loss errors in the smart meter measurements. The ensemble method shows improved performance and robustness compared to some previously proposed FL methods. Finally, the proposed method has been experimentally validated on a real-time simulation-based testbed using a state-of-the-art digital real-time simulator, industry standard DNP3 communication protocol and a cpu-based control centre running the FL algorithm.

1 | INTRODUCTION

Fast and accurate fault location identification (FLI) in a distribution system (DS) with microgrid is a difficult task for electric utilities due to i) lack of available data, ii) unbalanced network topology, iii) different loading conditions, and iv) microgrids with distributed energy resources (DER) integration. Recently, advanced metring infrastructure (AMI) and distribution phasor measurement unit (D-PMU and μPMU) have been deployed in distribution systems that are capable of collecting data at high resolution and transfer to the data centre [1]. The availability of more sophisticated signal information, high-speed communication, and more powerful computational capabilities of data processing devices have inspired utilities to shift from traditional FLI methods to data analytic or knowledge-based methods [1–3]. DPMUs or μPMUs can provide synchronised three-phase voltage and current phasor data at very high resolution (256/512 sample/cycle) and high accuracy (±0 01°) [2, 4]. Smart meters (SMs) are a key element in the AMI, which are capable of monitoring near real-time voltage, current and power data and can communicate with the utility data servers. SMs data resolution ($\frac{1}{3600}$ Hz to 10 Hz) is lower than µPMUs; however, the number of available μPMUs in the DS is very low because of the high installation cost. SMs have been widely deployed around the globe (around 96 million in 2016) [5].

Fault location identification in the DS can be divided into traditional methods, including impedance-based, traveling wave-based methods, and knowledge-based methods which are mainly artificial intelligence-based models applied on

monitored system data. Traditional FLI methods are not suitable for DS due to complex network configuration, implementation difficulties, and lack of accuracy [6]. Researchers have proposed modified impedance-based methods using voltage and current measurements at different buses to identify the fault location (FL) [7, 8], but their performance is compromised by unbalanced loads, fault resistance, fault type, and distributed generations. Multiple estimations of FL is a significant problem in impedance-based methods as it solves a set of quadratic equations to estimate the location. The authors in ref. [9] proposed an impedance-based method in combination with SM data to build low voltage zones and overcome the problem of multiple estimations. However, the method assumes a known homogeneous impedance throughout the network, which is uncommon for a DS. Performance of traveling-wave-based methods are free from operation mode changes but are more suitable for long distribution lines and require high sampling rate capable devices, data communication at a high rate which are rarely available in distribution networks (DN) [10, 11]. Implementation of traditional FLI methods like impedance-based methods require prior information of the system. There are very few numbers of monitoring devices available in a DN, which limits the observability of the system. The DN is primarily unbalanced, and the load varies throughout the day which makes parameter estimation very difficult and erroneous. Moreover, the DS is non-homogenous and line impedance is not the same throughout the network. All of these contribute to low accuracy and multiple estimations of the FL [12]. Travelling wave-based methods typically work better for voltages above 69 kV and require the installation of very high-resolution devices, making them hard to implement in a DN [13]. It also requires long lines which is not common in distribution systems. In addition, these methods are developed based on the traditional uni-directional power flow concept and do not consider fault current contribution from distributed resources. With a high amount of DER penetration, considering fault contribution from all DER and implementing them to form a model would become highly complicated as the number of state variables will increase significantly. A data-driven artificial intelligence (AI)-based approach on the other hand does not require prior system data or installation of any high-frequency devices. With the development of a better algorithm and different sensor-based devices, data-driven FLI methods can achieve high accuracy in FLI and can be implemented with a little or no cost in a smart DN.

Knowledge-based (KB) methods utilise AI-based models that can learn from historical data to solve complex problems. Several machine learning (ML) models like support vector machine (SVM), K-nearest neighbour (k-nearest neighbours (kNN)), fuzzy logic, artificial neural network (ANN), Deep neural network (DNN), and Graph neural network have been utilised for FL identification. Among the knowledge-based methods, ANN has been widely used in power system applications like event detection [14], fault applications [15], data integrity check, and distribution management [16]. Feedforward neural network combined with SVM to locate different types of faults in a practical DS using relay and breaker measurements at substations is presented in ref. [17]. Principal component analysis has been applied to the acquired data, and fault classification is done based on the reactances of their path, but it did not consider the effect of line capacitances. An FLI method based on ANN is presented in ref. [15] that shows reasonable accuracy for low impedance faults but is not tested for high impedance faults.

Fuzzy logic and kNN-based fault classification has been presented in refs. [18, 19], but the scope of these methods is limited only to fault classification. A hybrid model combining SVM, decision tree, and ANN for FL is presented in ref. [20]. This model utilises current and voltage phasors to classify the faulted phase and identify the faulted line but only applies to single-phase faults. Some researchers have used PMU data with deep learning methods to locate faults in the DS. In refs. [21, 22], the authors presented deep graph convolutional network (GCN) and stacked auto encoder-based methods, respectively to identify faults in distribution networks using PMU data. However, these methods assume the availability of synchronised PMU readings from a large number of buses, which is rare in the case of distribution networks. An AMI-based FLI method is presented in ref. [8] which needs pre-fault data from the SMs, which might not always be synchronous since SMs retrieve the data only periodically. The authors in ref. [23] presented an ensemble-based enhance FL method but their model was tested against very low impedance faults ranging from 0 to 5 Ω. Chen et al. [24] present a review of ML applications in power systems which lists several ML models and presents their comparative performance analysis in solving problems, such as fault detection, classification, and FLI.

The main challenges in deploying knowledge-based methods for DS applications are the scarcity of data and the unbalanced nature of the DN. Different models have their strengths and weaknesses, and performance varies with the system's diverse nature. An ensemble method combining several diverse ML models into one predictive model would be robust in different scenarios like noisy and lost data, fault resistance variation, and change in load or distribution feeder configuration. The main contribution of the paper is:

- Development of an ensemble voting classifier (EVC) utilising random forest (RF), kNN, and ANN models to classify fault types and identify FL using only voltage measurements from a limited number of data points.
- A modified AMI functionality has been proposed to gather event driven post fault data from existing small number of SMs with low data bandwidth requirement.
- The proposed algorithm and the AMI functionality has been implemented on a digital real simulator OPAL-RT [25], to demonstrate the performance and practical deployment potentiality in a real DS.

Most AI algorithms for FLI present in the literature require either synchronised measurements from PMU, μPMU or measurements from a large number of points that are currently absent in DN. The proposed method can be deployed with existing SMs in the system and shows robust performance

compared with existing well-known methods and, as it is a data-driven approach, does not require prior knowledge of network topology. Hence, it does not require any additional significant investment.

2 | ENSEMBLE LEARNING FOR FAULT LOCATION IDENTIFICATION

In this paper, post fault SM voltage measurements from load points at the end of the branch and in the photovoltaics (PV) buses are utilised to identify the FL. Let the set of all buses in a DN is B and B_s is the set of buses with SMs. The cardinality of B_s is, $|B_s| = n_s$. The set of voltage measurement values after a fault is a row vector, x with cardinality $|x| = n_s \times p_s$, where p_s is the phase associated with the corresponding bus. The FL identification task has been formulated as a classification problem where, f_c is the fault classification ensemble model and $f_l = (f_1, f_2, \ldots\ldots, f_n)$ are the FLI ensemble models for each fault type. In the first step, $y_c = f_c(x)$ predicts the fault type, and then $y_l = f_l(x_p)$ predicts the FL where, $x_p \in x$ is a subset of x with the faulty phase measurements only.

Ensemble learning combines several ML algorithms to avoid poor decision-making probabilities and improve the robustness and overall model's performance. Dasarathy in ref. [26] presented the first idea of a composite model to enhance performance. A traditional ML model learns a hypothesis from training data, wherein in an ensemble method, a set of hypotheses is constructed from the base learners. The results from the base learners are combined by unweighted or weighted voting to reduce variance and bias of the base models and improve the overall performance. The ensemble approach introduces diversity and weakness of one algorithm is often covered with another algorithm, such that the combined accuracy increases or is at least better than the worst result scenario. The motivation behind the use of ensemble model is to improve accuracy with limited data. There are weaknesses in every model, however, different models have different strengths and weaknesses. A too large or small data set can hinder good classification results with a single learning algorithm. The decision boundary separating different classes may be too complex or may fall outside the chosen classifiers' space of function. The combination of several models would improve the overall accuracy and reduce the possibility of performing very poorly in a specific condition.

The ensemble method will produce a better result if the individual classifiers are sufficiently diverse and accurate. An accurate classifier produces an error rate lower than a random guess for new samples while, if multiple classifiers yield different errors on new samples, they are diverse in nature [27]. Figure 1 shows the ensemble learning process from the base models.

Let there be T number of classifiers $h_1, h_2, \ldots, h_T$ and the class label from a set of 'k' possible class labels ($b_1, b_2, \ldots, b_k$) $\in B$ is predicted by combining the h_i's. It is generally assumed that for an instance x, the outputs of the classifier h_is

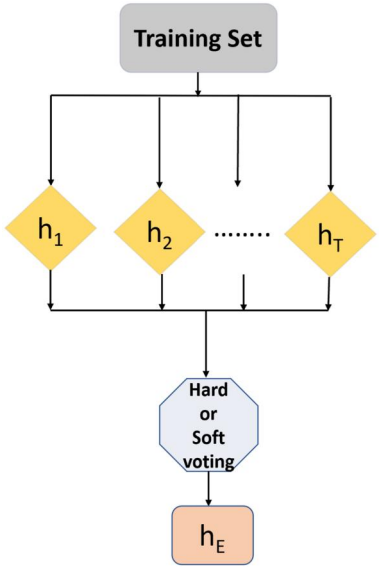

FIGURE 1 Ensemble learning process.

are given as an k dimensional label vector $\left(h_i^1(x), \ldots, h_i^k(x)\right)$, where $h_i^j(x)$ is the output of h_i for the class label b_j. There are two types of combination method hard voting or majority voting and soft voting. In majority voting, the final prediction of the ensemble model h_E is [28]:

$$H^j(x) = b_{\underset{j}{\text{argmax}} \sum_{i=1}^{T} h_i^j(x)} \tag{1}$$

In hard voting, the final output is the result produced by the majority of base models. In this method, all the individual base model results are counted, and the most frequent result is selected. In soft voting, individual probability outputs from all the base classifiers are averaged and based on this probability, the final output is chosen. In soft voting the final output for class b_j is given by:

$$H^j(x) = \frac{1}{T} \sum_{i=1}^{T} h_i^j(x) \tag{2}$$

In this paper, the individual classifiers have been chosen to ensure diverse output and sufficient accuracy from them. The base classifiers must be diverse in nature and sufficiently accurate to improve the overall accuracy of the ensemble method. If the base classifiers are not diversified enough in nature, then they would make the same mistakes, and combining them would not yield better results. Similarly, if each of the selected models is not accurate enough then the overall result may even become worse. The number of base classifiers is chosen odd in most cases to avoid the same number of votes in the case of hard voting. However, increasing the number of models would be recommended only if the overall accuracy increases. In our case, adding two more

base models did not bring significant improvement, so three base models were chosen. When choosing the base models, diversity and accuracy are the most important factors to be considered. The three base learners used in the proposed ensemble method are ANN, kNN, and RF models because of their higher accuracy and diversity. Each of these models is different in nature. Artificial neural network is a powerful classifier that learns complicated patterns of the data set. K-nearest neighbours is considered a lazy learning algorithm that is memoryless, each new sample is classified based on the distance calculated. K-nearest neighbours often performs badly for large data set and especially in cluster boundaries. Random forest is a combination of decision trees, which learns less complicated features of the data set easily and performs well for even large data sets [29, 30]. Details of the individual classifiers and ensemble methods are discussed in the following subsections.

2.1 | Artificial neural networks

Artificial neural network is a biologically inspired learning algorithm that imitates the learning procedure of the human brain or neurons and tries to imitate and solve complex problems based on cumulative learning. In a neural network model, there is an input layer, an output layer, and the layers in between are called the hidden layer. The neurons take the weighted output of other corresponding neurons in the previous layer as the input information in each layer. Each neuron computes the outcome as described by (3) and passes to the next neuron.

$$a_p = \sum_{q=1}^{n} w_{pq} x_{pq} + b_p \tag{3}$$

Here, x_{pq} is the signal value from input q to neuron p, w_{pq} is the associated weight for neuron ip and the input q, and b_p is the associated bias for neuron p. The output from each neuron is then fed into an activation function (AF) as in (4)

$$y_p = \varphi(a_p) \tag{4}$$

Here, φ is the AF, and y_p is the output of each neuron. In each layer, neurons receive output from the other connected units, adds the weighted inputs, and pass the output through an AF. The weight and bias values are updated to minimise the error. A supervised iterative method known as back-propagation [31] is used to train the ANN model.

2.2 | k-nearest neighbours

K-nearest neighbours is a supervised ML algorithm used for both classification and regression tasks. It is an instance-based or lazy learning algorithm, which means there are zero computational burdens in the training phase, and full effort is given in the prediction phase. In this method, weights are assigned to the contribution of neighbours based on similarity criterion so that the nearest neighbours contribute more significantly to the average than the distant neighbours. The performance of kNN depends on two hyperparameters: i) distance metric and ii) the number of neighbours, k. The most popular distance metric used in kNN implementation is the Euclidean method, defined as follows:

$$d(x, y) = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2} \tag{5}$$

Here, x and y are two inputs with n number of features. x_i and y_i are *ith* attributes of x and y respectively. A smaller value of d indicates similarity between the two samples. *K-nearest neighbours* of the sample x_i decide the final output by majority voting. The kNN algorithm is a widely used method for pattern classification tasks due to its competitive results and lazy learning approach (zero computational burden at training and full effort at prediction).

2.3 | Random forest

Random forest is a combination of weak decision trees to avoid correlation and build a stronger predictor. Breiman in ref. [30], presented the method of generating uncorrelated and variant trees, where split of bagged trees are generated based on a random number of observation samples and features. Bagged trees are generated based on a random number of observation samples and features. This helps to reduce the variance of the model, a common problem with the decision trees. The main hyperparameters associated with RF model are: the number of trees in the forest (B), feature number in each split (m), minimum sample number or leaf size ($l_{\min}$). The basic algorithm for implementing RF model is as follows [32]:

1. Sample data sets (B) with size N is generated.
2. For each sample data set, grow a RF tree (T_b) by iterating the following processes till leaf size ($l_{\min}$).
   - Select random m features from the feature set.
   - For each split point, the best feature is selected.
   - Based on the decision criteria, each node is split into two branch nodes.
3. Finally, generate the ensemble of trees $\{T_b\}^B$.

From the average of individual tree's output, classifier for a particular sample x is calculated.

In section V, the detailed hyper parameters for the base models and the implementation of ensemble model have been discussed.

2.4 | Ensemble voting classifier

The proposed EVC is a meta classifier that combines the above three base learners, that is, RF, kNN, and ANN. Figure 2

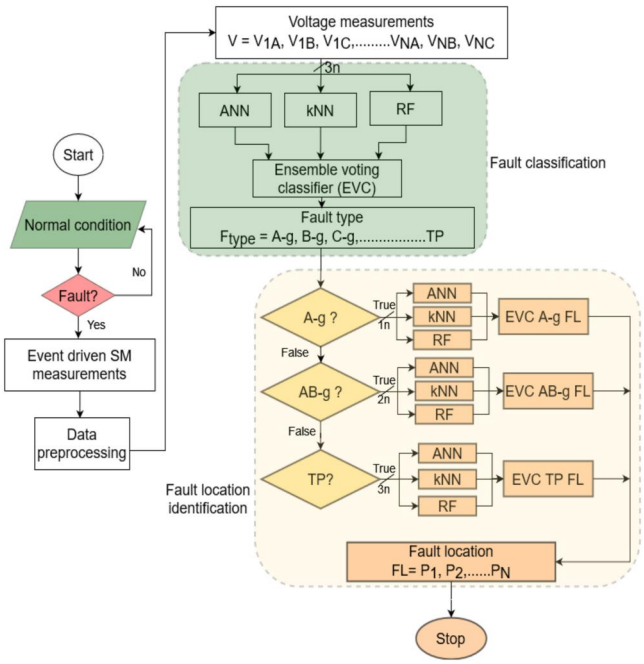

FIGURE 2 Flowchart for fault location identification (FLI) process.

shows the flowchart of the FLI process. The FLI process is divided into two steps, and in each step, the base learners are trained, and then the ensemble method performs the final prediction by majority voting (Hard and soft). The total procedure is described further here:

2.4.1 | Step 1: Fault Type Classification

In this step, fault classification is done using the fault on voltages from selected buses. SMs in the AMI are used to collect event-driven data during and after the fault. In the proposed method, only voltage magnitudes from SMs are used to perform fault classification. All different possible 11 types of faults have been placed on different buses. As in a DN, three-phase faults are rare; three-phase to ground and three-phase fault without ground samples are combined as three-phase faults (3P). So, there are in total 10 labels (A-g, B-g, C-g, AB-g, BC-g, CA-g, AB, BC, CA, 3P) to classify in the fault classification task as shown in Table 1. In a radial topology DN, SMs at the branch endpoint provides most relevant information for FLI [33].

The number of features for each fault incident is the number of SMs, n, multiplied by the corresponding phase number. In the fault type classification mode, all 67 voltage measurement data are used as input. One-hot encoding is used to convert the output of fault type from categorical to a numerical value. Each classifier produces an output vector, C, with different faults as shown in Table 1. The final output is predicted through majority voting using the prediction P from each classifier C.

TABLE 1 Fault types

Fault Type Code	Description
A-g, B-g, C-g	Single-line-to-ground fault (SLG)
AB-g, BC-g, CA-g	Double-line-to-ground fault (DLG)
AB, BA, CA	Line-to-line fault (LL)
ABC-g, ABC	Three phase fault (3P)

2.4.2 | Step 2: Fault location identification

After classifying the fault, base learners and the ensemble model are trained with the relevant voltage vectors for each fault type. From the voltage profile after a fault, it can be seen, the affected phase shows the maximum voltage variance in the event of a fault, and therefore only affected phase voltages have been utilised for the FL task. The size of the input vector fed into the models is affected phase number p, multiplied with the corresponding number of SMs. For an single-line-to-ground (SLG) (i.e. A-g) fault, the vector size will be $1n$, where one is the affected phase number and n is the number of SMs with phase A reading. Similarly, for a double line fault, the vector size will be $2n$, and for a three-phase fault, the vector size will be $3n$.

3 | PROPOSED MODIFIED ADVANCED METRING INFRASTRUCTURE

A microgrid consists of DER like photovoltaic plants, energy storage systems, and locally operating loads that can be treated as single entity [34]. Microgrids generally operate in both low and medium voltage levels ranging from 400 V to 69 KV [35]. Protection and identification of FL in a microgrid scenario are becoming a complex task with more DER penetration. There are three steps for faster fault clearance and service restoration: fault detection, classification, and FL identification. Fault detection is out of the scope of this paper, as techniques are already available in outage management software. The ensemble method presented in this paper can be combined with the existing detection method to ensure that the smallest possible part of the microgrid can be isolated in the event of a fault. Advanced metring infrastructure enables grid monitoring and reporting functionality. In the proposed method, SM [36] voltage measurement capability has been used to get the post fault voltage measurements from the metering points. Figure 3 shows the SM block and modified AMI infrastructure with DNP3 communication flowchart to harness post fault voltage data.

SMs in an AMI are capable of recording voltages, currents, and power data in near real-time resolution. However, constant data recording would require huge bandwidth and data storage capability. In this paper, a modified functionality based on DNP3 communication and SM has been proposed. DNP3 communication protocol is widely used for data communication in the power industry. DNP3 protocol is advantageous in

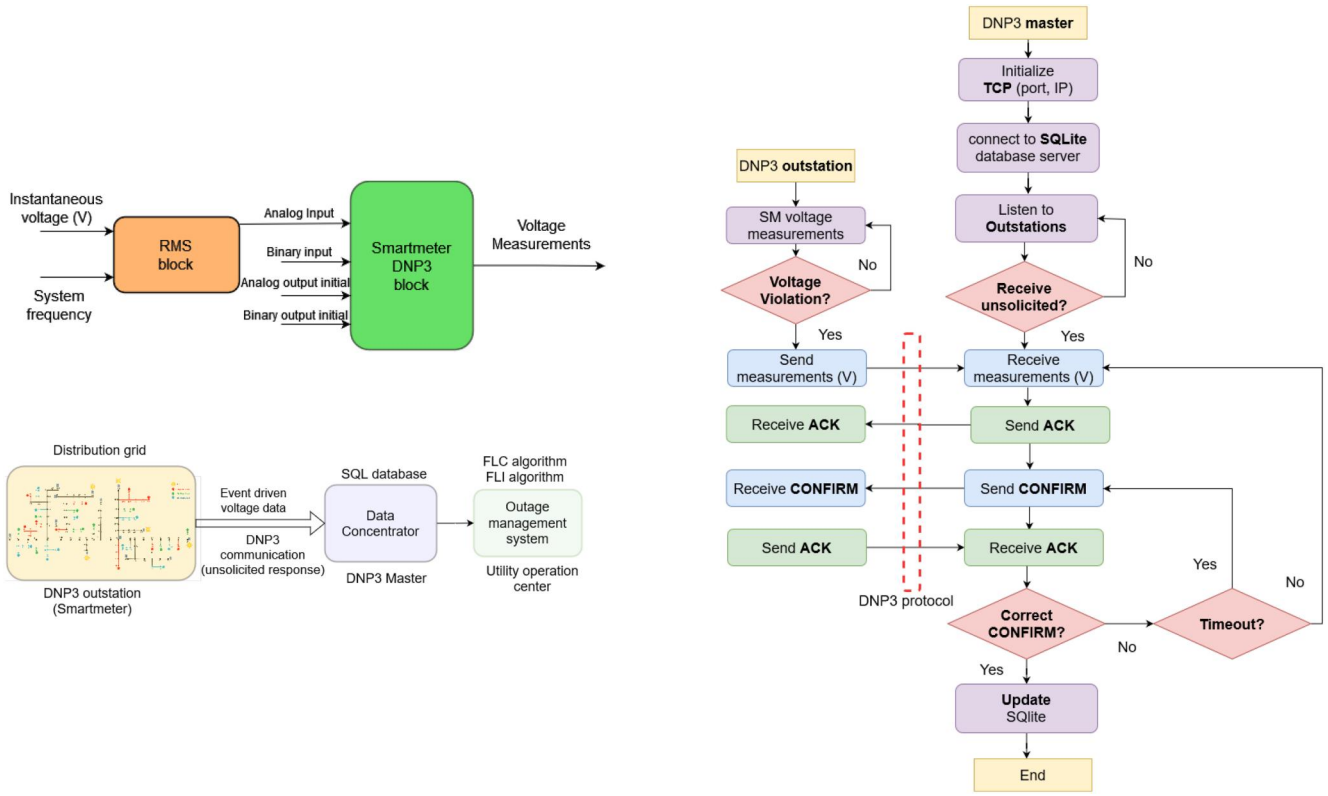

FIGURE 3 SM model overview and data transfer method (left), DNP3 communication flowchart (right).

this scenario because of its event-based communication capability. This mode of operation is called the unsolicited response. Here, the outstations (SMs) communicate with the master (control centre) if a predefined event occurs. In our experiment, we set this event as voltage threshold violation:

$$0.95pu \leq V \leq 1.05pu \tag{6}$$

If any of the SMs observe voltage threshold violation, all the SM data are collected and sent to the utility data centre through DNP3 communication as a voltage array: V_1, V_2, V_3, ……V_n, where the ensemble algorithm predicts the type and location of the fault. This significantly reduces the bandwidth requirement in the AMI infrastructure. A more detailed description of the communication process and AMI infrastructure are discussed in section VI.

Once the fault type and location have been predicted by the proposed ensemble method, in the case of single-phase fault, only affected phase tripping is possible, while the circuit breaker at the interface point can isolate the grid by three-phase tripping if needed. Figure 4 shows a distribution feeder including microgrid with PV system and separate load in a DS. If the proposed methodology identifies a fault in bus 84, then it would be sufficient for the operators to isolate buses 84 and 85. If the faulty node is bus 77, then three-phase tripping of the breaker at bus 76 is necessary. Microgrid protection and reliability can be enhanced by combining the ML algorithm with the existing relay system. Knowledge of fault type and the faulty bus would help isolate the affected smallest possible section and restore operation quickly.

4 | CASE STUDY AND OFF-LINE VERIFICATION OF THE PROPOSED ALGORITHM

For off-line verification of the ensemble method, the 4.16 KV, IEEE 123 distribution test feeder has been modified and simulated in OpenDSS [37]. The test feeder is shown in Figure 4 and details of the original model can be found in ref. [38]. The IEEE 123 bus system has in total 128 buses with spot loads at 85 of them. The following bus pairs are associated with normally closed switches: (149, 150r), (18, 135), (13, 152), (60, 160 (r)), (61, 610), and (97, 197). There are 5 open switches (250, 251), (450, 451), (54, 94), (151, 300), and (300, 350). Regulators are installed at buses 9, 25, and 160 to maintain the voltage level. For simplicity and real-time simulation purposes, closed switches have been considered a short line, while open switches have been considered open circuits. 5 PV plants have been placed at buses 56, 95, 151, 300, and 450 as shown in Table 2.

There is a microgrid included in the system involving PV at buses 300 and 450 and it has islanded operation capability. Photovoltaics systems have been modelled as a current source with a simple interface inverter in OpenDSS. More details on PV modelling and control strategy can be found in ref. [15]. In practical feeders, PV systems with an inverter are connected to

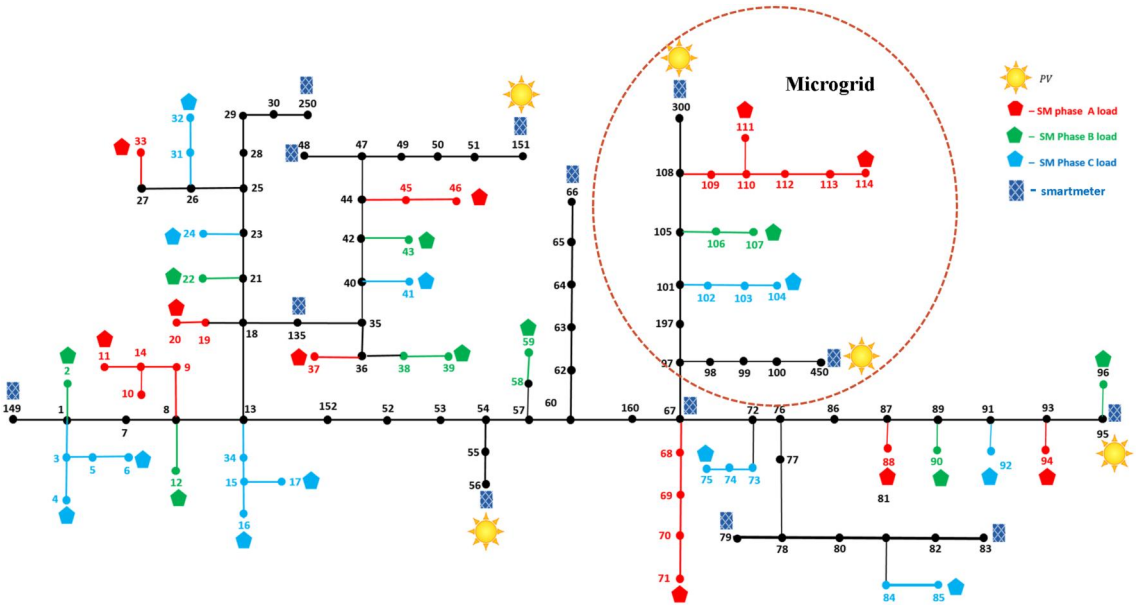

FIGURE 4 Modified IEEE-123 Bus System with added photovoltaics (PV)-based microgrids and SMs.

TABLE 2 photovoltaics (PV) specifications

PV location	Power (kW)	Voltage (kV)	Vmin (pu)	Controller strategy
300	350	0.48	0.8	Constant PQ
151	500			
56	500			
450	350			
95	350			

the grid with a back to back converter, which controls power sharing in normal mode and limits fault current during a fault event [39]. Thus, the microgrid contributes a fixed amount of current during a fault, and typically it is up to 1.25 p.u. The PV systems have been modelled to inject constant current if bus voltage reaches a minimum value, $V_{\min}$ from the grid side of the PV [40] plants in a fault event. Voltage data after a fault event are extracted from the metring locations using OpenDSS-Python com interface. Here, measurements from 43 SMs have been used, where 12 are on three-phase buses, and the rest are on single phase. SMs have been installed at the end of branches and with the five PV plants for collecting most significant information for FLI [33].

4.1 | Offline simulation of fault cases

The modified IEEE 123 model has been simulated in OpenDSS for different fault cases, as discussed in the previous section. Figure 5 provides an overview of the data generation, model training, and testing process. Training and testing data sets have been generated with the following process:

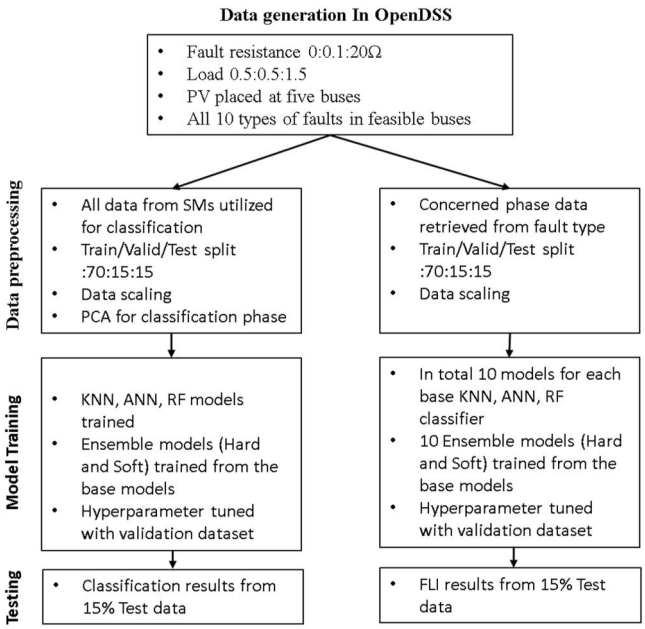

FIGURE 5 Offline data generation and model training process.

- For three-phase nodes, all types of faults have been placed where only SLG fault is considered in single-phase nodes.
- Fault resistance has been varied from 0 to 20 Ω.
- Unbalanced and dynamic loading has been explored by varying loading conditions from 0.5 times to 1.5 times the normal load.

4.2 | Fault type classification

For fault classification, the data set comprises of voltage vectors for all 10 types of faults at the feasible nodes. Training, validation, and testing data are generated by splitting the data set in the ratio of 70:15:15. The model specifications are:

1. RF: The number of trees is 100, minimal number of samples per leaf is 1, while minimal number of samples required for a split has been set to 3.
2. kNN: Scikit-learn library module has been used. Leaf size is 18, and the number of neighbour is set to 3.
3. ANN: 3 hidden layers with 90 neurons in each layer have been used. Adaptive Moment Estimation (Adam) is used as the AF. Regularisation and Max iteration limitation have been applied.
4. Ensemble method: Unweighted ensemble voting for both Soft and Hard methods have been considered.

The hyperparameters are tunable parameters of a ML model defined by the developer before the training process. The methods of selecting hyperparameters can be chosen either through empirical estimation or through complex algorithms like GridSearchCV, RandomizedSearchCV, and evolutionary optimisation. In this experiment, the hyperparameters of RF and kNN have been selected through grid search, while the hyperparameters for ANN were selected empirically by the trial-and-error method. The voltage data are in p.u. measurements to accommodate different voltage levels in the DN. Data curation and standard scaling have been done before forwarding to the model input. As the voltage measurements should always be positive, any negative or too large data (inconsistent values) have been replaced by 0. Principal component analysis was performed before the classification part to reduce the dimensionality before passing to the EVC classifier, but in FL stage, Principal component analysis was not applied, and all corresponding affected phase voltages were used as input feature. Table 3 shows the fault classification performances of the base and ensemble models. Figure 6 shows a detailed report of the classification performance for the ensemble method. The confusion matrix shows that the ensemble classifier shows 100% accuracy for all the fault types except AB-g and BC-g faults. After analysing misclassified instances, the wrong predictions were instances within 15–20 Ω fault resistance range, and within 15 Ω fault resistance, the ensemble method perfectly classifies all fault types. The confusion matrix also reports the precision and F1 score of the classification task.

TABLE 3 Offline fault type classification results

Fault type	Fault resistance (Ω)	Classifier	Accuracy (%)
SL-g	0–20	RF	91.20
		kNN	98.25
		ANN	98.38
		Ensemble (H,S)	100 (H), 99.61 (S)
DL-g	0–20	RF	91.47
		kNN	95.28
		ANN	99.02
		Ensemble (H,S)	99.34 (H), 98.25 (S)
LL	0–20	RF	91.10
		kNN	98.01
		ANN	98.95
		Ensemble (H,S)	100 (H), 99.18 (S)
3P	0–20	RF	92.49
		kNN	98.25
		ANN	98.88
		Ensemble (H,S)	100 (H), 99.25 (S)

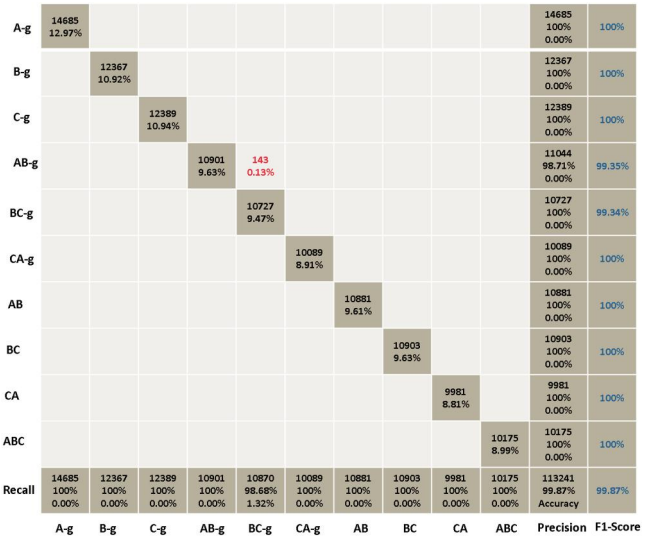

FIGURE 6 Offline fault classification Confusion matrix (Ensemble method (Hard)).

4.3 | Fault location identification

The FL task has been designed as a classification task to find out the faulty bus. The input data contains voltage information of only the faulted phases. The training data is divided into (70:15:15) training, validation, and testing sets. For FL identification, separate models are trained for each fault type. Table 4 shows the performance of the three base classifiers and EVC. K-nearest neighbours and ANN showed better performance than RF, however, EVC showed better performance than the base models. In hard voting, 97.65% has been achieved which

TABLE 4 Offline fault location identification (FLI) results for k-nearest neighbours (kNN), random forest (RF), artificial neural network (ANN), and Ensemble methods

Fault type	Fault resistance (Ω)	Classifier type	Accuracy (%)
A-g	0–20	ANN	95.80
		kNN	97.49
		RF	95.15
		Ensemble (H,S)	97.65 (H), 97.51 (S)
B-g	0–20	ANN	96.69
		kNN	96.28
		RF	93.72
		Ensemble (H,S)	96.29 (H), 96.24 (S)
C-g	0–20	ANN	97.06
		kNN	97.79
		RF	95.27
		Ensemble	97.91 (H), 97.79 (S)
AB-g	0–20	ANN	91.38
		kNN	92.25
		RF	91.2
		Ensemble	92.7 (H), 93.10 (S)
BC-g	0–20	ANN	93.24
		kNN	93.68
		RF	90.15
		Ensemble	92.60 (H), 93.30 (S)
CA-g	0–20	ANN	91.94
		kNN	98.17
		RF	96.95
		Ensemble	98.20 (H), 98.12 (S)
AB	0–20	ANN	91.36
		kNN	93.24
		RF	92.18
		Ensemble	93.30 (H), 93.60 (S)
BC	0–20	ANN	92.18
		kNN	92.65
		RF	92.21
		Ensemble	92.80 (H), 93.10 (S)
CA	0–20	ANN	95.10
		kNN	98.49
		RF	98.24
		Ensemble	98.70 (H), 98.62 (S)
3P	0–20	ANN	97.80
		kNN	93.80
		RF	96.76
		Ensemble	97.80 (H), 97.80 (S)

signifies predicting the faulty node correctly by at least two classifiers 97.65% times. In soft voting, 97.51% accuracy means the average of the class probabilities suggests correct faulty node 97.51% times.

5 | CASE STUDY UNDER ADVERSE SCENARIOS

In this section, the performance of the ensemble method for FL in the IEEE 123 benchmark system under different adverse scenarios has been reported. Robustness of the model in adverse scenarios such as noise, missing or miscalibrated SM data due to a communication error, or faulty SM has been explored. More specifically, the following modifications in measurements have been done to observe performance in real-world scenarios. Here, in evaluating different scenarios random errors are considered, so the results are not unique and accuracy may vary slightly according to the respective cases. In case of ensemble methods all the results have been simulated three times and average is recorded in the concerned tables.

5.1 | Impact of noise

In the proposed methodology, SM voltage data has been used, typically the accuracy class of SM measurements is $\pm 0.1\% - \pm 0.5\%$ [41]. In order to perform sensitivity analysis, Gaussian noise with zero-mean and standard deviation, σ, where $\sigma = 0.5\%$, 1%, 5%, and 10% has been multiplied with random measurements. The noisy SM data is then fed to the ensemble model. Table 5 lists the performance of ensemble classifiers under different noisy conditions.

5.2 | Random data loss of SM

Due to communication error or faulty SM, some data are assumed to be missing. In the algorithm, if any SM data is absent or some irrational data like negative value appears, it is replaced with zero. Here to analyse such scenarios, randomly, some SM data have been set to zero to mimic such scenarios. Cases I, II, III, and IV in Table 6 depicts situations where 10%, 15%, 20%, and 25% SM readings are missing. The worst results happen when SM data closest to the faulty bus are lost.

5.3 | Impact of reconfiguration of distribution network

The configuration of DN may change to balance the loads. Such scenarios may be generated by changing the normally closed/open switches in the network. The IEEE 123 distribution test feeder has six normally closed and five normally open switches. Some of these switch statuses were modified to achieve new network configuration:

1. Configuration I: Close normally open switch between bus (151, 300), and (54, 94), open normally closed switches between bus (18, 135) and bus (13, 152).
2. Configuration II: Close normally open switch between bus (151, 300), and (300, 350), open normally closed switches between bus (97, 197) and bus (18, 135).

TABLE 5 Fault location (FL) accuracy (in %) under noisy conditions

	Model					
	Ensemble (Soft)			Ensemble (Hard)		
Fault type	1(Φ)	2(Φ)	3(Φ)	1(Φ)	2(Φ)	3(Φ)
$\sigma = 0.5\%$	96.27	93.31	96.63	96.34	93.24	96.34
$\sigma = 1\%$	96.14	93.12	96.55	96.23	93.27	95.76
$\sigma = 5\%$	95.74	92.74	96.55	93.44	92.87	95.76
$\sigma = 10\%$	90.78	87.29	89.55	90.39	87.50	90.07

TABLE 6 Fault location (FL) accuracy (in %) under random data loss cases

	Model					
	Ensemble (Soft)			Ensemble (Hard)		
Fault type	1(Φ)	2(Φ)	3(Φ)	1(Φ)	2(Φ)	3(Φ)
Case I	97.29	93.59	96.37	97.18	94.01	96.22
Case II	96.27	93.31	96.08	96.34	93.24	95.22
Case III	89.67	83.70	85.02	90.80	87.44	88.10
Case IV	80.67	83.70	81.02	81.08	83.44	80.10

TABLE 7 Fault location (FL) accuracy (in %) under network reconfiguration change

	Model					
	Ensemble (soft)			Ensemble (hard)		
Fault type	1(Φ)	2(Φ)	3(Φ)	1(Φ)	2(Φ)	3(Φ)
Configuration I	82.45	85.79	87.17	81.52	84.31	86.42
Configuration II	76.27	81.31	85.23	80.34	81.24	84.32

For both configuration, all different types of faults have been placed at feasible buses and 10 samples are collected. The samples are tested with the ensemble model. Table 7 shows ensemble model performances in reconfiguration scenarios.

Although, samples with changed configuration were not included in the training process, ensemble model produced around 80% accuracy in locating faults.

5.4 | Islanded mode operation

The analysis presented so far considers grid-connected mode operation of microgrids, where the main fault current contributor is the utility grid and DERs. However, microgrids have the capability to operate in islanded mode, where the microgrid is disconnected from the main grid and DER solely supports the loads in the microgrid according to their capacity. There can be several scenarios in this mode and inverter response may vary according to their control strategy [39]. To analyse such a scenario we simulated islanded mode operation of microgrid one in Figure 4 by operating a breaker in the point of common coupling (PCC) and placing all different type of faults upstream of the microgrid in random buses with fault resistance 0.01–5 Ω. The microgrid operates in the islanded mode if the voltage of PCC goes below operating range [42]. Performance in such scenario is presented in Figure 7. The ensemble models performed better than the individual models.

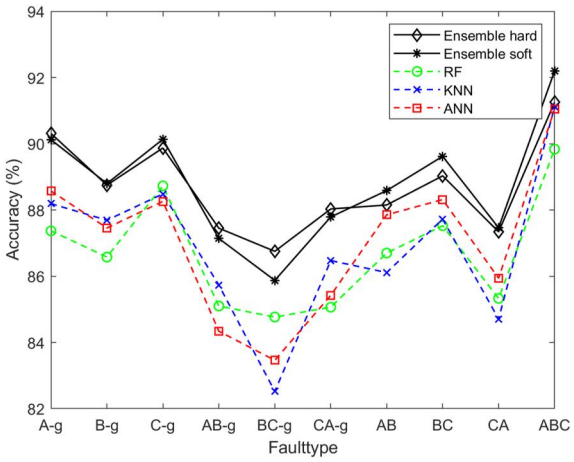

FIGURE 7 Performance in islanded operation of microgrid.

6 | EXPERIMENTAL SETUP FOR REAL-TIME VALIDATION

In this section, the proposed algorithm was tested with an experimental setup using a digital real-time simulator (DRTS). In the beginning, the ensemble approach was tested offline by using offline simulation tools and collecting data for training and testing the proposed method. In the offline implementation, cases of data loss, data delay or jitters are not present as data is processed first asynchronously and then sent to the algorithm without keeping any track of time. However, in real-time validation, data is sent in real-time from the power system (synchronously) to the ensemble algorithm using DNP3 communication backbone. This provides a more realistic environment to see how the approach performs. To validate the proposed method with real-time signals, the DS, PV-based DERs, and the DNP3 communication system with modified functionality have been developed in the OPAL-RT simulator [25]. OPAL-RT is a DRTS which performs parallel computation with multiple multicore processors. Figure 8 shows the simulation setup for real-time validation. The simulated SMs use the IEEE 1815 [43] protocol for transferring data to the control centre where the algorithm is running to classify and locate the fault. The setup comprises of three parts:

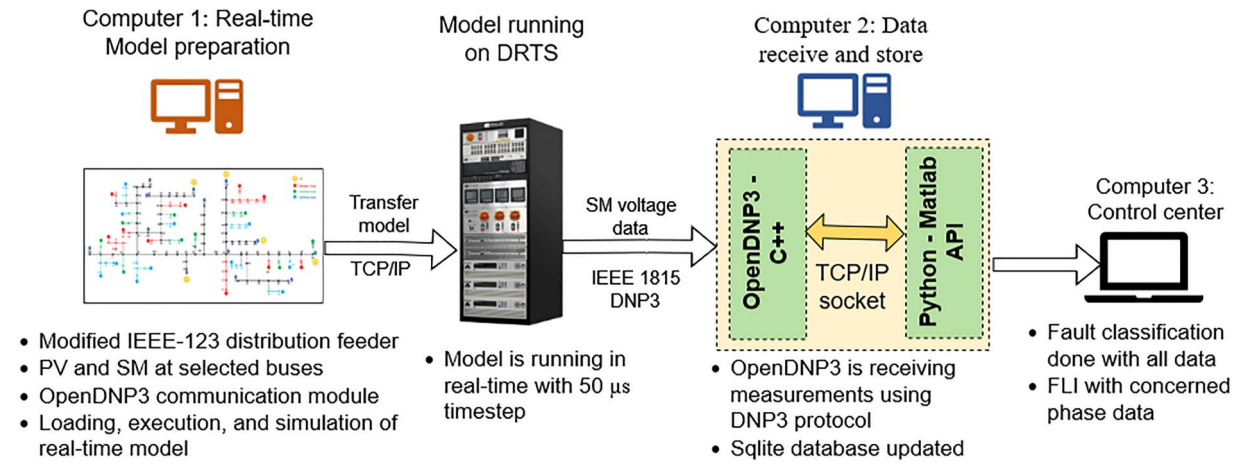

FIGURE 8 Real-Time testbed Setup using digital real-time simulator (DRTS) for experimental validation of fault location (FL) algorithm.

1. **Real-Time Distribution Feeder Model**: Consists of IEEE 123 test feeder model with PV system model and SMs designed to run in a real-time simulation. The transient system has been implemented in Opal-RT Simulink platform using Simscape power system libraries. It represents a DS with microgrids and AMI. Different faults have been placed at random points to get the fault response of the system. To avoid a large state space matrix that eventually causes overrun, 5 state-space nodal solvers (SSN) have been used. These SSN blocks work as a decoupling tool to reduce a large state space matrix to smaller ones to ensure easier and faster computation in real-time.
2. **Communication Network**: Consists of a communication layer similar to a wide-area monitoring system (WAMS). An open-source implementation of DNP3 known as openDNP3 [44] has been modified to establish communication with the data centre. The modified AMI collects event driven data from selected SMs and sends to the datacenter.
3. **Database and Data Streaming**: SQLite [45] has been used here to store the data incoming from the SMs. The database is updated when the fault is detected and fed to the algorithm.

Figure 9 shows the experimental setup for ensemble model validation at centre for advanced power system (Centre for Advanced Power Systems), FSU. The modified IEEE123 feeder with microgrid is simulated in the RT-LAB/SIMULINK environment in computer 1 and uploaded to the DRTS, where real-time simulation using a cluster of processors is performed. The grid is monitored by SMs present at different buses at the end of each branch. The SMs use DNP3 (IEEE1815 standard) to communicate with the data centre in the event of a fault. DNP3 is operated in unsolicited response mode, which means data will be transmitted if a predefined event happens and there is a change in the monitored data set. This significantly reduces the transmission bandwidth requirement, and as only fault-driven voltage data is required for the proposed model, it is sufficient for deploying the algorithm. In computer 2, the data concentrator comprises a

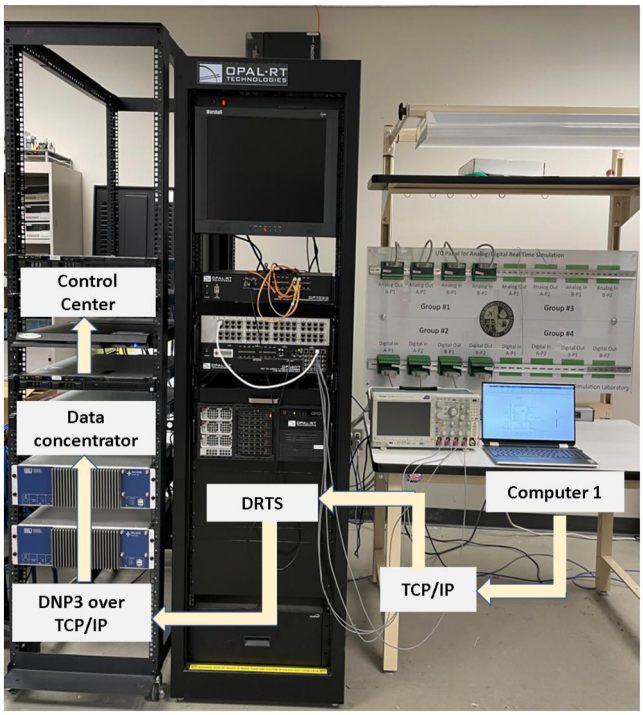

FIGURE 9 Experimental setup in the Centre for Advanced Power Systems (CAPS).

database system (SQLite) and a python-based module that enables the algorithm to pull data from the database. Computer 3 represents the control centre that runs the ML models for fault type and location identification. For validating the proposed method, test samples are generated by placing faults in the real-time model running on the DRTS. A 50 μs simulation timestep has been used for this purpose. The offline data was generated by OpenDSS which is a frequency domain electric simulation tool, whereas OPAL-RT performs electromagnetic transient analysis with 50 μs time-step which depict real-time simulation of the system. For offline case studies, faults were placed at every bus with varying fault resistance and

TABLE 8 Real-time fault location (FL) identification results for faults on the nodes

Fault type	Fault resistance	Fault placement node	Node identified by offline prediction	Node identified by real-time prediction
A-g	1	19	19	19
	15	23	23	21 (adjacent)
	20	33	33	33
B-g	1	12	12	12
	2	25	25	25
	15	44	44	42 (adjacent)
C-g	1.5	5	5	5
	2.5	17	17	17
	3.5	29	29	29
AB-g	7	36	36	34
	8.5	47	47	47
	15.5	53	53	53
BC-g	15	13	13	13
	0.5	50	50	50
	4.5	64	64	64
CA-g	5	101	101	101
	1	95	95	95
	2	60	60	60
AB	5	98	98	98
	20	99	97 (adjacent)	98 (adjacent)
	1	100	100	100
BC	1.5	54	54	54
	2.5	55	55	55
	3.5	56	56	56
CA	15	101	102 (adjacent)	104
	5	102	102	102
	20	108	300	108
3P	0.5	18	18	18
	1.5	82	82	82
	2.5	108	108	108
3P	1	72	72	72
	20	89	89	90 (adjacent)
	10	135	135	135

loading condition and overall performances have been presented in corresponding tables. However, in real-time simulation conducting such an extensive case study analysis is time exhaustive. So, for real-time performance analysis a handful of cases based on experiences from offline studies covering 0–20 Ω fault resistances have been performed. Few instances of Simulink-generated data were included in the training data set to ensure the consistency of data distribution. The results are summarised in Table 8. We can see that FL identification is accurate in low impedance conditions, but the algorithm provided a few misclassified results when the fault resistance is 15 Ω or higher. As we analysed further into these misclassified cases, the faulty bus identified were found to be adjacent buses as can be seen in Table 8.

7 | DISCUSSION

The ensemble model presented in this paper utilises a data-driven, yet physics aware approach to classify and locate

faults in distribution systems. A testbed setup with communication method and AMI enhancement has been discussed. In the previous sections, the performance of the ensemble model has been compared with three base classifiers ANN, kNN, and RF. The ensemble model presented enhanced performance compared to each of the base models in different scenarios. Although the ensemble model is subject to increased complexity resulting in higher training and testing time than individual models, the gain in accuracy compensates for this. To achieve faster service restoration, accurate FLI is a more important metric for utilities than faster but less accurate operation. Another thing to consider for practical implementation of data analytic-based methods is the fault data. A large number of real fault data for varying resistance levels, loads, and fault types may be difficult to acquire. The practical data may be combined with synthetic simulated data to tune the hyper-parameters before implementation in a real system. Such an implementation technique for AI-based event analysis has been investigated in ref. [46]. Researchers in Oak ridge national lab combined real waveform data from Electric Power Board of Chattanooga and simulated data from Matlab to implement a convolutional neural network-based event classifier. The study proposes simulated data to be the starting point and refine the simulation model based on real field samples. As more field data are available, these are included in training process and simultaneously the simulation data are adjusted to create more realistic data. Such cycle is repeated until sufficient field data are collected to train the AI model. In comparison to other AI-based techniques present in literature, the proposed method has the following advantages:

- GCN proposed in ref. [21] achieved overall FLI accuracy 97.13%, whereas ensemble model achieved 95.91% accuracy. The GCN model however uses PMU data (voltage and current magnitude and angle measurements) from all existing buses, whereas the ensemble model only utilises voltage measurements from a limited number of buses.
- The stacked ensemble model presented in ref. [23] only considered faults up to 5 Ω, whereas this study considered faults up to 20 Ω.
- In comparison to stacked auto encoder-based FL identification method presented in ref. [22], this method has been tested with PV-based DER and a comparatively larger distribution feeder.

In addition to these, we have presented a detailed communication network setup and tested in a DRTS to further validate the model. From the results in Table 8, we see the model accuracy decreases in high fault resistance cases. As fault resistance increases, the voltage drop becomes minimal for low fault current, and only voltage measurement may be insufficient for FLI in high impedance scenarios. In such cases, including more features like angle measurements [47] can increase accuracy; however, this will require enhancement in the WAMS system with more capable devices.

8 | CONCLUSION

This paper presents an EVC for FL applications in the DN. SMs data from a selected few buses have been utilised for fault classification and location identification in a DN. A modified AMI structure has been presented with an open-source implementation of the DNP3 communication protocol to harness post fault data for FLI applications. It can be implemented in an existing AMI infrastructure requiring low bandwidth and without significant investment. As a data-driven approach, the ensemble model does not require prior system information. The proposed ensemble method has been verified using a modified IEEE-123 bus system and shows better accuracy than individual classifiers including ANN, kNN, and RF. The proposed methodology showed robust performance in multiple fault scenarios with noisy or lost SM data, load change, topology modifications, and varying fault resistances up to 20 Ω. The proposed method thus can improve situational awareness of smart distribution networks and enable faster service restoration by identifying fault types and faulty nodes with a data analytic approach.

AUTHOR CONTRIBUTION

Md Maidul Islam: Conceptualisation, Methodology, Software, Validation, Visualisation, Writing – original draft. **Muhammad Usama Usman:** Conceptualisation, Methodology, Writing – review & editing. **Alvi Newaz:** Methodology, Software, Validation. **Md Omar Faruque:** Supervision, Writing – review & editing.

ACKNOWLEDGEMENT

This research was funded by the Department of Electrical and Computer Engineering, FAMU-FSU College of Engineering and authored in part by Battelle Energy Alliance, LLC with the U.S. Department of Energy.

CONFLICT OF INTEREST

The authors declare that they have no conflict to declare.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

ORCID

Md Maidul Islam https://orcid.org/0000-0002-6245-3609

REFERENCES

1. Jiang, Y., et al.: Outage management of distribution systems incorporating information from smart meters. IEEE Trans. Power Syst. 31(5), 4144–4154 (2016). https://doi.org/10.1109/tpwrs.2015.2503341
2. Usman, M., Faruque, M.: Applications of synchrophasor technologies in power systems. J. Mod. Power Syst. Clean Energy. x, 1-1(2), 211–226 (2018). https://doi.org/10.1007/s40565-018-0455-8
3. Wang, Y., et al.: Review of smart meter data analytics: applications, methodologies, and challenges. IEEE Trans. Smart Grid. 10(3), 3125–3148 (2018). https://doi.org/10.1109/tsg.2018.2818167
4. AbdelRaheem, M., Hassan, M., Selim, H.: A lightweight sampling time error correction technique for micro phasor measurement units. IEEE Trans. Instrum. Meas. 71, 1–8 (2022). https://doi.org/10.1109/tim.2022.3179009

5. Reinhardt, A., Pereira, L.: Energy data analytics for smart meter data. Energies. 14(17), 5376 (2021). https://doi.org/10.3390/en14175376
6. Saha, M., Izykowski, J., Rosolowski, E.: Fault Location on Power Networks. Springer-VerlagLondon (2010)
7. Krishnathevar, R., Ngu, E.: Generalized impedance-based fault location for distribution systems. IEEE Trans. Power Deliv. 27(1), 449–451 (2012). https://doi.org/10.1109/tpwrd.2011.2170773
8. Trindade, F., Freitas, W., Vieira, J.: Fault location in distribution systems based on smart feeder meters. IEEE Trans. Power Deliv. 29(1), 251–260 (2014). https://doi.org/10.1109/tpwrd.2013.2272057
9. Trindade, F., Freitas, W.: Low voltage zones to support fault location in distribution systems with smart meters. IEEE Trans. Smart Grid. 8(6), 2765–2774 (2017). https://doi.org/10.1109/tsg.2016.2538268
10. Thomas, D., Carvalho, R., Pereira, E.: Fault location in distribution systems based on traveling waves. In: Power Tech Conference Proceedings, 2003 IEEE Bologna, vol. 2, pp. 5 (2003)
11. Hamidi, R., Livani, H., Rezaiesarlak, R.: Traveling-wave detection technique using short-time matrix pencil method. IEEE Trans. Power Deliv. 32, 2565–2574 (2017)
12. Das, S., et al.: Impedance-based fault location in transmission networks: theory and application. IEEE Access. 2, 537–557 (2014). https://doi.org/10.1109/access.2014.2323353
13. Das, S., Santoso, S., Ananthan, S.: Fault Location on Transmission and Distribution Lines: Principles and Applications. John Wiley & Sons (2021)
14. Postolache, O., et al.: An ANN fault detection procedure applied in virtual measurement systems case. Instrumentation and Measurement Technology Conference, IMTC/98. Conference Proceedings. IEEE.1998 1 pp. 257–260 (1998)
15. Usama, M., Ospina, J., Faruque, M.: Fault Classification and Location Identification in a Smart Distribution Network Using ANN. Power Engineering Society General Meeting, 2018, pp. 1–5. IEEE (2018)
16. Cramer, M., Goergens, P., Schnettler, A.: Bad data detection and handling in distribution grid state estimation using artificial neural networks. PowerTech, 2015 IEEE Eindhoven, 1–6 (2015)
17. Thukaram, D., Khincha, H., Vijaynarasimha, H.: Artificial neural network and support vector machine approach for locating faults in radial distribution systems. IEEE Trans. Power Deliv. 20(2), 710–721 (2005). https://doi.org/10.1109/tpwrd.2005.844307
18. Youssef, O.: Combined fuzzy-logic wavelet-based fault classification technique for power system relaying. IEEE Trans. Power Deliv. 19(2), 582–589 (2004). https://doi.org/10.1109/tpwrd.2004.826386
19. Costa, F., Souza, B., Brito, N.: Real-time classification of transmission line faults based on maximal overlap discrete wavelet transform. PES T&D 2012, 1–8 (2012)
20. Maruf, H., et al.: Locating faults in distribution systems in the presence of distributed generation using machine learning techniques. In: 2018 9th IEEE International Symposium on Power Electronics for Distributed Generation Systems (PEDG), pp. 1–6 (2018)
21. Chen, K., et al.: fault location in power distribution systems via deep graph convolutional networks. IEEE J. Sel. Area. Commun. 38(1), 119–131 (2020). https://doi.org/10.1109/jsac.2019.2951964
22. Luo, G., et al.: Stacked auto-encoder-based fault location in distribution network. IEEE Access. 8, 28043–28053 (2020). https://doi.org/10.1109/access.2020.2971582
23. Ghaemi, A., et al.: Accuracy enhance of fault classification and location in a smart distribution network based on stacked ensemble learning. Elec. Power Syst. Res. 205, 107766 (2022). https://doi.org/10.1016/j.epsr.2021.107766
24. Chen, K., Huang, C., He, J.: Fault detection, classification and location for transmission lines and distribution systems: a review on the methods. High Volt. 1, 25–33 (2016). https://doi.org/10.1049/hve.2016.0005
25. OPAL-RT OPAL-RT technologies. (https://www.opal-rt.com/)
26. Dasarathy, B., Sheela, B.: A composite classifier system design: concepts and methodology. Proc. IEEE. 67(5), 708–713 (1979). https://doi.org/10.1109/proc.1979.11321
27. Dietterich, T.: Ensemble methods in machine learning. International Workshop On Multiple Classifier Systems, 1–15 (2000)
28. Zhou, Z.: Ensemble methods: foundations and algorithms. CRC press (2012)
29. Bishop, C., Nasrabadi, N.: Pattern Recognition and Machine Learning. Springer (2006)
30. Breiman, L.: Random forests. Mach. Learn. 45(1), 5–32 (2001). https://doi.org/10.1023/a:1010933404324
31. Buduma, N., Locascio, N.: Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms. O'Reilly Media, Inc. (2017)
32. Abuella, M., Chowdhury, B.: Random forest ensemble of support vector regression models for solar power forecasting. In: Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2017 IEEE, pp. 1–5 (2017)
33. Trindade, F., Freitas, W.: Low voltage zones to support fault location in distribution systems with smart meters. IEEE Trans. Smart Grid. 8(6), 2765–2774 (2017). https://doi.org/10.1109/tsg.2016.2538268
34. Farrokhabadi, M., et al.: Microgrid stability definitions, analysis, and examples. IEEE Trans. Power Syst. 35(1), 13–29 (2020). https://doi.org/10.1109/tpwrs.2019.2925703
35. Hatziargyriou, N., et al.: Microgrids. IEEE Power Energy Mag. 5(4), 78–94 (2007). https://doi.org/10.1109/mpae.2007.376583
36. Jiang, J., Yu, L.: Design of a new three-phase multi-rate watt-hour meter based on singlechip. In: 2009 International Conference on Computational Intelligence and Software Engineering, pp. 1–4 (2009)
37. OpenDSS: EPRI distribution system simulator. https://sourceforge.net/projects/electricdss/
38. Kersting, W.: Radial distribution test feeders. In: 2001 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.01CH37194), vol. 2, pp. 908–912 (2001)
39. Samet, H., Azhdari, E., Ghanbari, T.: Comprehensive study on different possible operations of multiple grid connected microgrids. IEEE Trans. Smart Grid. 9(2), 1434–1441 (2018). https://doi.org/10.1109/tsg.2016.2591883
40. Keller, J., Kroposki, B.: Understanding Fault Characteristics of Inverter-Based Distributed Energy Resources. National Renewable Energy Laboratory (2010)
41. Irwin, L.: A high accuracy standard for electricity meters. IEEE PES T&D 2010, 1–3 (2010)
42. Prabakar, K., Singh, A., Tombari, C.: IEEE 1547-2018 based interoperable PV inverter with advanced grid-support functions. In: 2019 IEEE 46th Photovoltaic Specialists Conference (PVSC), pp. 2072–2077 (2019)
43. IEEE Standard for Exchanging Information Between Networks Implementing IEC 61850 and IEEE Std 1815(TM) [Distributed Network Protocol (DNP3)]: IEEE Std 1815.1-2015 (Incorporates IEEE Std 1815.1-2015/Cor 1-2016). pp. 1–358 (2016)
44. OpenDNP3 documentation. (https://dnp3.github.io/)
45. SQLite Homepage. https://www.sqlite.org/index.html
46. Sticht, C.: Power system waveform classification using time-frequency and CNN. (2022,1). https://www.osti.gov/biblio/1841478
47. Cui, Q., El-Arroudi, K., Weng, Y.: A feature selection method for high impedance fault detection. IEEE Trans. Power Deliv. 34(3), 1203–1215 (2019). https://doi.org/10.1109/tpwrd.2019.2901634

How to cite this article: Islam, M.M., et al.: Ensemble voting-based fault classification and location identification for a distribution system with microgrids using smart meter measurements. IET Smart Grid. 1–14 (2022). https://doi.org/10.1049/stg2.12091