Mining Partially Labeled Data from Edge Devices to Detect and Locate High Impedance Faults

Qiushi Cui and Yang Weng
School of Electrical, Computer and Energy Engineering
Arizona State University, Tempe, Arizona, USA
Email: {qiushi.cui, yang.weng}@asu.edu

***Abstract*—The security of active distribution systems is critical to grid modernization along with deep renewable penetration, where the protection plays a vital role. Among various security issues in protection, conventional protection clears only 17.5% of staged high impedance faults (HIFs) due to the limited electrical data utilization. For resolving this problem, a detection and location scheme based on μ-PMUs is presented to enhance data processing capability for HIF detection through machine learning and big data analytics. To detect HIFs with reduced cost on data labeling, we choose expectation-maximization (EM) algorithm for semi-supervised learning (SSL) since it is capable of expressing complex relationships between the observed and target variables by fitting Gaussian models. As one of the generative models, EM algorithm is compared with two discriminative models to highlight its detection performance. To make HIF location robust to HIF impedance variation, we adopt a probabilistic model embedding parameter learning into the physical line modeling. The location accuracy is validated at multiple locations of a distribution line. Numerical results show that the proposed EM algorithm greatly saves labeling cost and outperforms other SSL methods. Hardware-in-the-loop simulation proves a superior HIF location accuracy and detection time to complement the HIF's probabilistic model. With outstanding performance, we develop software for our utility partner to integrate the proposed scheme.**

I. Introduction

Societal production and living activities become more reliable on electricity as the grid modernization evolves. Consequently, power interruptions cause severe economical loss and even safety issues in modern grids. Take a recent one for example, mid-Manhattan experienced power outage on July 13, 2019, impacting approximately $72,000$ customers. A preliminary investigation from Consolidated Edison found the relay system failed to isolate a fault [1]. From a protection perspective, it is comparatively easy to detect a fault with low impedance. Nevertheless, conventional relays have difficulties in detection high impedance fault. For this purpose, various algorithms, are proposed, such as proportional relaying approach, impedance-based method, and PC-based fault locating and diagnosis algorithm [2], etc. However, these methods are unable to solve two fundamental problems in HIF detection, namely the measurement accuracy and information extraction capability. Due to this reason, [3] uses case studies to show that that conventional protection systems, even with new ideas, can only 17.5% of staged HIFs.

Notably, different edge devices, such as μ-PMUs are becoming available in the distribution grid, providing better measurement accuracy. With such measurement devices, HIF detection devices can provide high-precision and high-resolution measurements. For example, many utilities and campuses installed or plan to install μ-PMUs for distribution grid monitoring. Measurements from these devices can be easily be reused to capture peculiar characteristics of HIFs, e.g., phasor data concentrators (PDCs) can use data analytics algorithms with the data streams. Therefore, this paper proposes a architecture on highly accurate data-driven HIF detection and location, based on μ-PMUs, leading to significantly improved protection in distribution systems. Under such an architecture, a systematic process is developed to extract and select features, conduct semisupervised learnning (SSL), and perform probabilistic learning for situational awareness in fault identification.

Admittedly, fault detection schemes already used machine learning methods, but the past focus has been in the category of supervised learning. For example, Bayes classifier has been proposed to distinguish between fault cases in the Bayesian framework, following wavelet-transform-based data extraction. Decision tree algorithm was used for interpretations in the form of a white box, which has been a conventional approach to use neural network-structure to deal with inherently nonlinearity in the decision process. The major concern over such supervised-learning approach is that it is seldom the case that labelled records are readily available in utility database. Even worse, there are ambiguities between fault and non-fault events, even if all the labels are available. Finally, it is costly for utilities to pay labors to manually identify all the historical events without any error. So, we carefully design a SSL to greatly enlarge the information gain, while reduce uncertainties in HIF analysis.

In addition to detection, localizing the fault is also important. For localization, μ-PMUs are found to be quite helpful. For example, we can use the compensation theorem in circuit theory and PMU-based state estimation via analyzing the source of different power events. For our paper, we propose to improve further by considering the probabilistic result of the semi-supervised learning result as well as probability distribution over impedance in HIF. Specifically, we will suggest a probabilistic analysis by forming a moving-window total-least-square based on the probability distribution of the fault impedance values.

To test the effectiveness of using real time μ-PMU measurements, we use real-time simulator for validation. For example, real-time property of μ-PMUs are examined by establishing an experimental platform via a OPAL-RT real-time simulator. With such validation design, we observed a significantly improved HIF detection and location capability over the conventional methods by extensive simulations.

In summary, this paper contributes to the following aspects.

978-1-7281-4930-1/19/$31.00 ©2019 IEEE

First, we propose an HIF detection and location scheme that leverages on big data obtained from the edge device of μ-PMU. Second, a generative EM-algorithm-based SSL model for HIF detection is designed iteratively. This model presents a higher binary classification accuracy than another two discriminative models under study. Second, the probabilistic characteristics of HIF impedance are modeled to assist HIF location. This model embeds the physical law of distribution lines into fault location method.

We organize the paper in the follows: Section II explain the proposed method on feature selection method for HIF detection, the EM-based SSL approach, and the probabilistic method for HIF location. We implement our method in Section III by integrating the three aforementioned approaches. Section IV shows experimental results followed by hardware implementation in Section V. Section VI summerize the paper.

II. HIF Detection and Location Scheme

The proposed HIF detection and location scheme utilizes SSL for HIF detection and a probabilistic model for location.

A. Feature Selection

This paper adopts the wrapper approach (refer to Section IV-D) as its feature evaluator to solve the binary classification problem in HIF detection. Unlike the filter approach, the training set in the wrapper approach goes through three steps before it is sent to the ultimate induction algorithm: feature selection search, feature evaluation, and induction algorithm.

In this paper, we propose to search feature by an engine called best-first search, due to its robust performance against the well known hill-climbing search engines. One of the major advantages in best-first search method is that it stops immediately after the performance starts to drop. However, it keeps a long list of all possible and evaluated attribute subsets and sorts according to performance measure with a goal of letting a prior configuration be reached again. Specifically, we search with greedy hill-climbing augmented with a backtracking facility in the attribute subset space. Algorithm 1 illustrate the process. Notably, we use a five-fold cross-validation (CV) for validation with an underline assumption: all folds are independent during training process. Additionally, we use standard deviation of the accuracy estimate to determine the number of repetition tines.

Algorithm 1 Best-first algorithm

1: Put the initial state on the *OPEN* list.
2: *CLOSED* list $\leftarrow \phi$, *BEST* $\leftarrow$ initial state.
3: Let $v = arg\ max_{w \in OPEN} f(w)$ (get the state from *OPEN* with maximal $f(w)$).
4: Remove v from *OPEN*, add v to *CLOSED*.
5: **if** $f(v) - \varepsilon > f(BEST)$ **then**
6: *BEST* $\leftarrow v$.
7: Expand v: apply all operators to v, giving v's children.
8: For each child not in the *CLOSED* or *CLOSED* list, evaluate and add to the *CLOSED* list.
9: **if** *BEST* changed in the last k expansions **then**
10: goto 3.
return *BEST*.

In short, we stop the search when we can not find a node with improvements during the past k expansions. Therefore, an enhanced node would have an accuracy estimation at least ϵ times higher than the best one found in the past. For consistancy and clearance, we use $k = 5$ and $\epsilon = 0.1\%$ in the rest of the paper.

B. Expectation-Maximization (EM) Algorithm in SSL for HIF Detection

After data cleaning and feature extraction, we can apply a supervised learning approach for HIF detection. However, the performance of supervised learning relies on the number of labeled HIF event in the past, which may not be sufficient. For this reason, we propose to employ semi-supervised learning to incorporate data from unseen events, therefore saving the cost of data labeling. Table I shows the comparison of required dataset among unsupervised learning, supervised learning, and SSL. As a highlight, SSL only requires a few labeled observations and can improve performance significantly by adding a large number of unlabeled observations, which is cheap to obtain.

TABLE I
Three machine learning categories for label availability.

Category	Input dataset [a]	Labeling
Unsupervised Learning	$X = [\mathbf{x_1}, \ldots, \mathbf{x_n}]^T$	$Y \in \emptyset$
Supervised Learning	$X = [\mathbf{x_1}, \ldots, \mathbf{x_n}]^T$	$Y \in \mathbb{R}^{n \times 1}$
Semi-Supervised Learning	$X = [\mathbf{x_1}, \ldots, \mathbf{x_l}, \ldots, \mathbf{x_{l+u}}]^T$ [b]	$Y \in \mathbb{R}^{l \times 1}$

[a] $\mathbf{x_i} = [x_1, x_2, \ldots, x_d]$, $i \in N$, $X \in \mathbb{R}^{n \times d}$, where n denotes the number of observations and d denotes the number of features.
[b] $l + u = n$, usually $l < u$, where l denotes the number of labeled observations and u denotes the number of unlabeled observations.

In the machine learning field, there are various dedicated approaches for combining labelled and unlabelled datasets. The two most popular approaches are based on self-training and co-training. For self-training, it uses predictions to train itself on its own whereas co-training uses two classifiers to train each other with the most appropriate prediction labels. This paper involves the use of self-training since HIF detection as HIF detection does not need two classifiers in co-training. To perform such training, it is necessary to select appropriate discriminative or generative probabilistic models.

The discriminative model relies on conditional probability distribution $P(\mathbf{x_i}|\mathbf{y_i})$ of features $\mathbf{x_i}$ and labels $\mathbf{y_i}$. However, the problem of discriminative model is that it cannot generally express complex relationships between the observed and target variables. We, therefore, choose the EM method in generative models by fitting Gaussian mixture models. A mixture model assumes that the probability of $\mathbf{x_i}$ is given by

$$p(\mathbf{x_i}) = \sum_{c=1}^{k} p(\mathbf{x_i}, \mathbf{y_i} = c) = \sum_{c=1}^{k} p(\mathbf{y_i} = c) p(\mathbf{x_i}|\mathbf{y_i} = c), \quad (1)$$

where c is viewed as a set of k clusters of the data, and $\mathbf{y_i}$ is the cluster membership of $\mathbf{x_i}$. $(\mathbf{x_i}|\mathbf{y_i})$ usually has a simple form like a Gaussian distribution. But we do not know the $\mathbf{y_i}$ values when labels are partially available. Therefore, by viewing $\mathbf{y_i}$

as hidden values, we deploy the EM method for HIF data classification. The goal of the proposed SSL method for HIF is to classify any $x \in X$ achieving $\hat{y} = \arg\max_{y \in Y} p(\mathbf{x}, \mathbf{y}|\theta)$. The expected log-likelihood parameterized by θ is

$$Q(\theta|\theta^t) = \sum_{i=1}^{n} \log p(\mathbf{y_i}, \mathbf{x_i}|\theta^t) + \sum_{i=1}^{t} \sum_{\mathbf{y_i} \in \{0,1\}} r_i^{\mathbf{y_i}} \log p(\mathbf{y_i}, \mathbf{x_i}|\theta). \quad (2)$$

We define the following

$$r_i^0 = p(\tilde{\mathbf{y}}^i = 0|\tilde{\mathbf{x}}^i, \theta^t) = \frac{p(\tilde{\mathbf{y}}^i = 0, \tilde{\mathbf{x}}^i|\theta^t)}{\sum_{\mathbf{y} \in \{0,1\}} p(\mathbf{y}, \tilde{\mathbf{x}}^i|\theta^t)}, \quad (3)$$

$$r_i^1 = p(\tilde{\mathbf{y}}^i = 1|\tilde{\mathbf{x}}^i, \theta^t) = \frac{p(\tilde{\mathbf{y}}^i = 1, \tilde{\mathbf{x}}^i|\theta^t)}{\sum_{\mathbf{y} \in \{0,1\}} p(\mathbf{y}, \tilde{\mathbf{x}}^i|\theta^t)}, \quad (4)$$

where $\tilde{x}$ and $\tilde{y}$ denote the unlabeled data and labels respectively. Thus, the EM method proposed for semi-supervised learning alternates between the following two steps:

1) E-step: Compute probabilities r_i^0 and r_i^1 for all the unlabeled examples i based on the current θ^t.
2) M-step: Maximize the expected complete-data log-likelihood (equation (2)), which is a weighted version of the complete-data log-likelihood.

C. A Probabilistic Model for HIF Location

After fault detection, we need to locate the fault. The location system (Fig. 1) firstly calculates fault data classification accuracy $\lambda_i (i = 1, \cdots, n$, assuming there are n μ-PMUs) based on each μ-PMU's measurement. Secondly, we compute the fault location using our proposed probabilistic model that outputs the calculated fault location. This probabilistic model will be elaborated later.

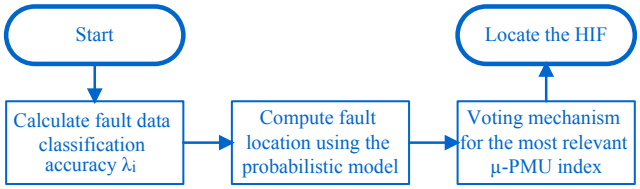

Fig. 1. The flowchart for the HIF location method.

Thirdly, a voting mechanism is introduced to determine which μ-PMU provides the most reliable data for location. We view each μ-PMU as a voter. Each voter v_i $(i = 1, \cdots, n)$ has a ranking of the vector of candidates $C_i = [c_1, \cdots, c_m]^T$, where each element in C has a binary value indicating the voting of protection $1, \cdots, m$. In this application, the voted faulty zone is translated from the calculated fault location obtained from last step. To refine the ranking, with a slight abuse of notation, we define:

$$\text{idx}^* = \max_{j=1}^{n} \lambda_i (\max_i \sum_{i=1}^{n} C_i). \quad (5)$$

where idx^* is the index of the most relevant μ-PMU. It is obtained by firstly finding the index of the highest voted μ-PMU(s) and then the one with highest classification accuracy.

The probabilistic model for HIF Location is discussed here. The HIF is modelled with the extensively used anti-parallel dc-source model, as shown in Fig. 2b. The details of the same can be found in [4]. In the beginning, a relationship is established between the one-terminal measurement and fault location using a constant-impedance constant-DC-source HIF model. It is difficult to represent the random phenomenon of fault impedance during arcing using the HIF model in [5] as constant impedance and DC source are assumed. The conventional solutions such as in [5] come with limitations on measurement location, device, and accuracy. To overcome the aforementioned issues, the μ-PMU environments come equipped with HIF location systems that are capable of introducing randomness.

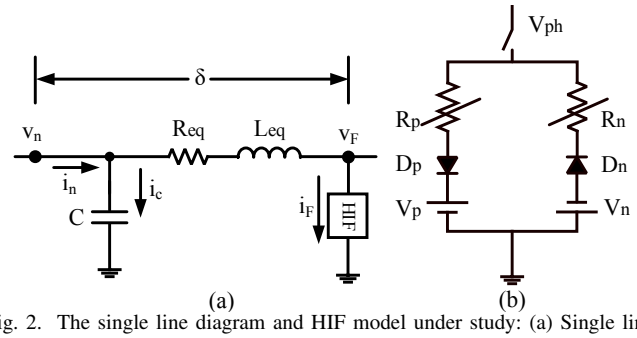

Fig. 2. The single line diagram and HIF model under study: (a) Single line model for the HIF location analysis [5]; (b) Two anti-diode HIF model [6].

Fig. 2a presents the one terminal system's single line diagram. On the left of this figure, v_n is the measured voltage of the n^{th} μ-PMU. Assuming a fault is occurring at the distance δ, then the line impedance, $R_{eq} + j\omega L_{eq}$, seen by the n^{th} μ-PMU is calculated as the multiplication of δ and the per-unit-length impedance. We merge the two shunt capacitance into one on the left, with the value of C. It draws i_C current. In this paper, the per-unit-length resistance, inductance, and capacitance are denoted by R, L, and C.

Based on Fig. 2b, the HIF voltage is computed as follows:

$$v_F = \begin{cases} i_F R_p + V_p, & i_F \geq 0, \\ i_F R_n - V_n, & i_F < 0, \end{cases} \quad (6)$$

Line capacitance has a substantial contribution to the fault location error in underground cables, but for overhead lines, we can ignore the impact of the shunt capacitors without jeopardizing the results: $i_C = \delta C \frac{\mathrm{d}V_n}{\mathrm{d}t} \approx 0$. Here, we use one equation to represent the above two conditions in (6) and apply KVL for the circuit in Fig. 2a:

$$v_n = \delta(Ri_n + L\frac{\mathrm{d}i_n}{\mathrm{d}t}) + R_F i_F + V_{DC}, \quad (7)$$

where R_p, R_n, and V_p, $(-V_n)$ are the positive and negative cycle values of R_F and V_{DC}.

The way of predicting the fault current (i_F) originates from [5]. We firstly estimate the fault distance, then employ the least square method to calculate δ, R_F, and V_{DC} of both positive and negative cycles. According to 7, we have

$$\begin{cases} R_p = c_{p0} + c_{p1}\delta, & i_F > 0, \\ R_n = c_{n0} + c_{n1}\delta, & i_F < 0, \end{cases} \quad (8)$$

where $c_{p0} = \frac{v_n - V_p}{i_F}$, $c_{p1} = -\frac{1}{i_F}(Ri_n + L\frac{\mathrm{d}i_n}{\mathrm{d}t})$, $c_{n0} = -\frac{v_n + V_n}{i_F}$, and $c_{n1} = \frac{1}{i_F}(Ri_n + L\frac{\mathrm{d}i_n}{\mathrm{d}t})$.

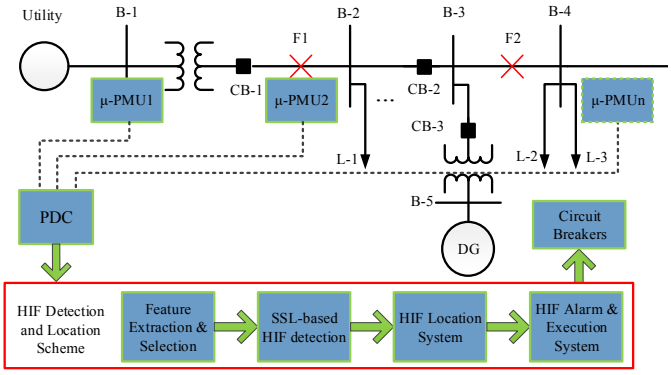

Fig. 3. The proposed HIF detection and location scheme with four function blocks: feature extraction and selection, SSL-based HIF detection, HIF location system, and HIF alarm & execution.

III. Implementation Scheme

To implement our proposed scheme, we assume μ-PMUs are available in the system with full observability. The synchronized phasor data are sent to phasor data concentrators (PDCs), as shown in Fig. 3. The red box shows the proposed scheme with four functions: feature extraction and selection, SSL-based HIF detection, HIF location system, and HIF alarm & execution. The method for feature extraction and selection is supplied in [4]. This section elaborates on the remaining three functions.

A. Classification Between HIF and non-HIF

From the machine learning perspective, it is a classification issue regarding differentiating between fault and non-fault data. The collected data from μ-PMUs is firstly feature engineered and then classified by the proposed EM-based SSL method. There is one binary classification model for each μ-PMU, since the local measurement of μ-PMUs differ from each other. The detailed method of classifying the HIF and non-HIF data can be referred to as Section II-B.

B. HIF Location Function

If the location of the HIF is identified, line dispatchers can efficiently clear the fault. However, the biggest challenge here is the determination of the varying HIF impedance. The random and irregular characteristics of the HIF are usually associated with contact objects such as soil, plant, and sand, as well as the moisture and temperature of the air. Also, the low fault current is beyond the capacity of the legacy protective elements. To solve this problem, two probability models are proposed to recognize the nature of the varying HIF impedance.

1) Normal distribution of the HIF impedance: The first model explores the normal distribution of the HIF impedance R_F. We have $R_F \sim \mathcal{N}(\mu, \sigma^2)$. The fault location, therefore, follows

$$\delta \sim \mathcal{N}(\frac{\mu - c_0}{c_1}, (\frac{\sigma}{c_1})^2), \tag{9}$$

where c_0 and c_1 are the normal distribution parameters. Consequently, the confidence interval of the fault location estimation can be easily quantified.

2) Uniform distribution of the HIF impedance: In the second model, we adopt a uniform distribution model to capture the fault impedance range. Given the HIF impedance range of (R_{min}, R_{max}), we can derive the following fault location equation:

$$(\frac{R_{min} - c_0}{c_1}, \frac{R_{max} - c_0}{c_1}), \tag{10}$$

where c_0 and c_1 are the uniform distribution parameters.

C. High Impedance Fault Alarm and Execution System

With the help of the HIF location function, an estimated fault location can be calculated through the two proposed impedance models. Since they are probability models, the fault location is a predicted range. Although a precise location cannot be determined, the location range can greatly helpline dispatchers to eliminate unnecessary search. The HIF signal and its location information are transmitted to the system operator. Associated alarming or tripping signals can be immediately sent to the execution system, which controls the circuit breakers or switches.

IV. Experiment Results

A. Benchmark System

The McGill Electric Energy Systems Laboratory has developed the benchmark system showed in Fig. 4, which is a simulation of the distribution feeder as a common rural community feeder. This feeder is rated for 25 kV which is obtained through utility step down transformer rated for 120 kV : 25 kV. In the upcoming section, the hardware experiment platform is discussed in details with the real-time benchmark system simulation.

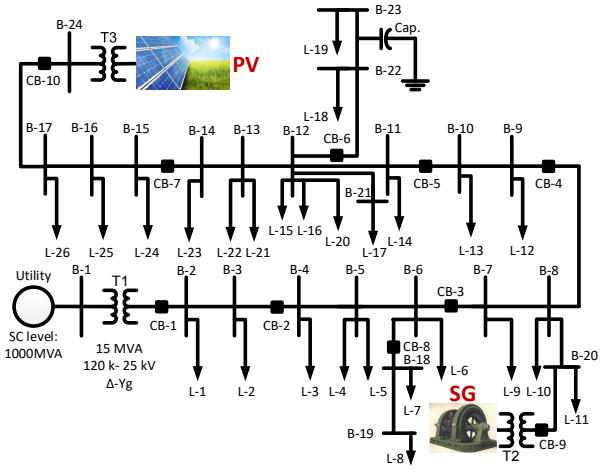

Fig. 4. Benchmark system [7].

B. HIF Detection Performance

The SSL algorithm as discussed above is tested with shorted out feature group from the 14,850 HIF and non-HIF event characteristics of the benchmark system such as HIF with unbalanced impedance, HIF with different fault location, HIF with various fault types, load and capacitor switching, load variation and so on. First, it is required to define SSL principle, therefore we identify a couple of common features such as – I_2 and $\theta_{V_2} - \theta_{V_0}$ – which are helpful to identify the difficulties associated with the classification feature grouping and to illustrate feature groping process of the unlabeled data

TABLE II
BINARY CLASSIFICATION RESULTS USING SEMI-SUPERVISED LEARNING METHODS.

% of labeled data	K-nearest-neighbors method			Information-theoretic method [8]			EM method		
	precision	recall	F1 score	precision	recall	F1 score	precision	recall	F1 score
6.25	0.996	0.996	0.996	0.868	0.969	0.915	1.000	1.000	1.000
12.5	0.997	0.997	0.997	0.855	0.980	0.913	1.000	1.000	1.000
25	0.997	0.997	0.997	0.900	0.955	0.927	1.000	1.000	1.000
50	0.998	0.998	0.998	0.998	0.996	0.997	1.000	1.000	1.000

for the same task. As shown in the left of Fig. 5, only twenty-five percent of the total available data is used for the training purpose at first.

Two notations are used to discriminate the non-fault events and fault events such as solid circle and solid square respectively which are concentrated at the particular mark in the graph. Therefore, we require to generate a feasible algorithm to large scale binary feature grouping. On the other hand, there are other notations such as circle and square, used to show two other types of events to illuminate the forecast groups which help to see through the discussed approach picks up through the training data and generates forecast about the unidentified data-set. In addition to this, Fig. 5 clearly shows that forecast results are nearby the labeled data even it is difficult to identify multi-dimension data.

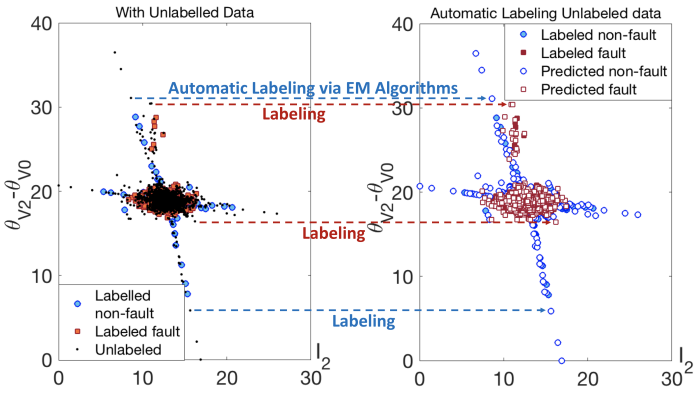

Fig. 5. Visualization of training and testing data in the SSL-based method. We utilize two typical features of I_2 and $\theta_{V_2} - \theta_{V_0}$ to visualize the complexity of the classification task in a two-dimension plot and to depict the SSL classification process of the unlabeled data. The corresponding data is rescaled and standardized using the following equation: $Y = \frac{X - X_{min}}{\sigma}$, where X is the vector of the original data, Y is the rescaled and standardized vector, X_{min} is the minimum data value in the corresponding vector, and σ is the standard deviation of X. In this figure, 25% data are labeled.

To further investigate the accuracy of the SSL-based method, we demonstrate the performance of the generative EM method in comparison with two discriminative models. Table II reveals and quantifies the improvement in fault detection as the percent of labeled data grows. To maintain a fair comparison, we randomly select the labeled data, which means that the quality of data during SSL varies case to case. Despite the performance oscillation due to the goodness of data, the precision, recall and F1 score have a trend to rising as the percent of labeled data increases. Through horizontal comparison, it is easy to see that EM method outperforms k-nearest-neighbors and information-theoretic method.

C. HIF Location

Here, μ-PMUs are used to take measurements from the source side and six main lines such as 2-3, 4-5, 7-8, 9-10, 11-12, 15-16. These measurements are utilized to simulate and analyze 18 HIF events. And, exercises results show that HIF events are found on these lines as mentioned above.

As per [5] location error is Location error $= \frac{\text{estimated distance} - \text{actual distance}}{\text{line length}} \times 100\%$. To test the HIF location system, many HIFs are generated on the line connected between bus 2 and 3 in Fig. 4, of the distribution feeder and results are shown in Fig. 6. The estimated error using linear least square estimation (LSE) method in [5] is compared with these results mentioned above. Here, the line under test in the benchmark system is $4.167\,\text{km}$ long. Therefore, the proposed method has more data points than the method tested in [5], which is $0.6\,\text{km}$ long only. And, due to small line length, fault impedance plays a dominant role in the calculation of the estimation error while the fault is near the measurement unit but the fault impedance does not deviate from the location estimation error.

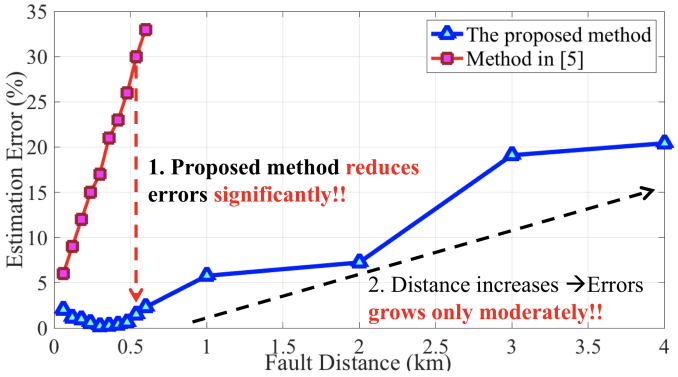

Fig. 6. Fault location error comparison. HIF impedance follows the uniform distribution in the proposed method. Line length in the proposed method: $4.167\,\text{km}$. Line length in [5]: $0.6\,\text{km}$.

Moreover, HIF location estimation approach and results discussed here illustrate the remarkable improvement compared to the HIFs location estimation method and results. For example, it is difficult to locate HIF in [5] due to measurement devices and the approach limitations if the line is $4\,\text{km}$ long. But, if the line is shorter than $1\,\text{km}$, then error rate is smaller than 6% compare to [5]. Here, the focus of discussion aims to eliminate the difficulties associated with the identifying HIF location. After the HIF detection, utility management makes a public announcement to illuminate potential danger and initiate an operation to discover for the ground touch conductor as a conventional approach [3]. Here, it is required to conduct an operation to discover the downed conductor for only 20%

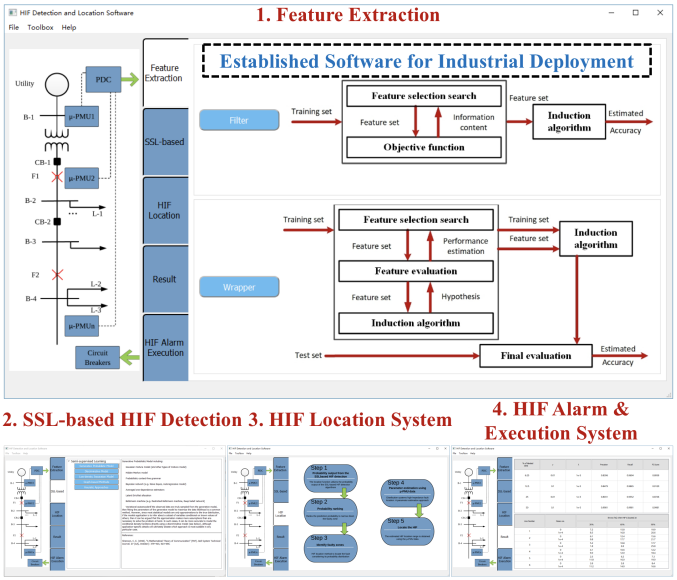

Fig. 7. Our self-develop software with HIF detection and location scheme.

of the line to make a safe environment as per the approach discussed above.

D. Commercial Software

To commercialize the HIF Detection and Location software, we create a GUI application with Python cross-platform GUI toolkit PyQt. Its interface is shown in Fig. 7, whose functionality includes Feature Extraction, SSL-based HIF Detection, HIF Location System, and HIF Alarm and Execution System. We have commercialized this software. Our industrial partners also test the software.

V. Hardware Experiment Platform

To simulate and prove the discussed HIF detection and location method, a hardware experiment setup has been prepared by the authors as per shown in Fig. 8. This model consists of a real-time 25 kV feeder and a number of μ-PMUs to process CTs and PTs analog output of the simulator via amplifiers. 3-ϕ current and voltage signals of the specific bus are processed by the μ-PMUs and feature data is sent to the PDC. Furthermore, discussed HIF detection and location algorithm is embedded in the PC with the all programmed functions. Here, HIF location and detection scheme efficiency are tested on six lines as shown (named after from bus - to bus) in Fig. 4, and results associated with testing are stated in Table III. In addition to this, experiments to recognize HIF and its time from available measurements for the HIF at 30%, 60%, and 90% of the length of the line location under research.

There are two μ-PMUs installed in the simulation to collect data. And, the algorithm processing and HIF clearing time are not critical in this method due to reasons discussed in section IV-D about sending line dispatchers to search. Plus, this method is a backup protection scheme to detect the HIFs that legacy protective relays cannot detect. While some of the existing and undeveloped work has already shown response time below 200 ms. Moreover, this system is not fully developed in terms of the backloop of a tripping signal to the appropriate circuit breaker. Therefore, the interface between the HIF detection algorithm and the circuit breaker is our future work.

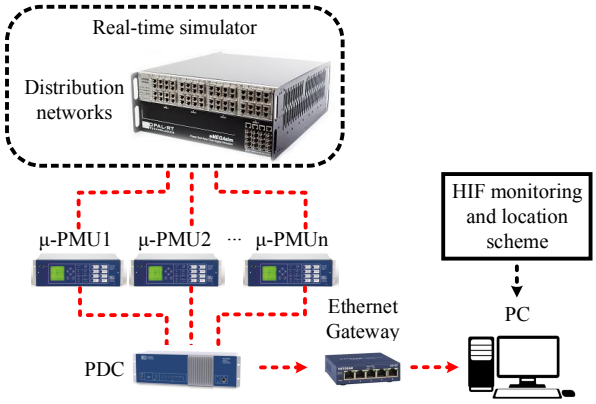

Fig. 8. The hardware experiment platform. The red dotted lines indicate physical connection among real-time simulator, μ-PMUs, PDC, Ethernet gateway and PC.

TABLE III
Hardware Experiment Results of the HIF Detection Time.

Location of line	2-3	4-5	7-8	9-10	11-12	15-16
Avg. detection time (ms)	469	482	488	514	491	470

VI. Conclusions

Edge devices such as μ-PMUs are generating a large amount of data everyday. This paper proposes to utilize μ-PMUs for HIF detection and location. An EM method on semi-supervised learning is proposed to recognize HIFs, avoiding the cost of high volumes HIF data labeling. After identifying the faulty range with a voting mechanism and the impedance probability models, HIF location can be predicted accordingly. The results of the fault location function show a small estimation error with the help of μ-PMUs, in comparison with previous work. Furthermore, the HIF detection time obtained from hardware-in-the-loop simulation demonstrates a promising result.

References

[1] NBC News. (2019) Terrorism, cyberattack ruled out as cause of manhattan power outage. [Online]. Available: https://www.nbcnews.com/news/us-news/power-outage-strikes-midtown-manhattan-n1029636

[2] D. C. Yu and S. H. Khan, "An adaptive high and low impedance fault detection method," *Power Delivery, IEEE Transactions on*, vol. 9, no. 4, pp. 1812–1821, Oct 1994.

[3] J. Tengdin, R. Westfall *et al.*, "High impedance fault detection technology report," *PSRC Working Group D15*, 1996.

[4] Q. Cui, K. El-Arroudi, and Y. Weng, "A feature selection method for high impedance fault detection," *IEEE Transactions on Power Delivery*, vol. 34, no. 3, pp. 1203–1215, June 2019.

[5] L. U. Iurinic, A. R. Herrera-Orozco, R. G. Ferraz, and A. S. Bretas, "Distribution systems high-impedance fault location: A parameter estimation approach," *Power Delivery, IEEE Transactions on*, vol. 31, no. 4, pp. 1806–1814, Aug 2016.

[6] S. Gautam and S. M. Brahma, "Detection of high impedance fault in power distribution systems using mathematical morphology," *Power Systems, IEEE Transactions on*, vol. 28, no. 2, pp. 1226–1234, 2013.

[7] D. Zhuang, "Real time testing of intelligent relays for synchronous distributed generation islanding detection," Master's thesis, McGill University, Dec 2012.

[8] G. Niu, W. Jitkrittum, B. Dai, H. Hachiya, and M. Sugiyama, "Squared-loss mutual information regularization: A novel information-theoretic approach to semi-supervised learning," in *International Conference on Machine Learning*, 2013, pp. 10–18.