Real-time AI-based Fault Detection and Localization in Power Electronics Dominated Grids

[1]Matthew Baker, *Student Member*, *IEEE*, [1]Mohammad B. Shadmand, *Senior Member*, *IEEE*,
[2]Arslan Munir, *Senior Member, IEEE*
Department of Electrical & Computer Engineering, University of Illinois Chicago, IL, USA
Department of Computer Science, Kansas State University, Manhattan, KS, USA
mbaker36@uic.edu, shadmand@uic.edu, amunir@ksu.edu

***Abstract*— This paper presents a real-time fault detection and classification in power electronics dominated grids (PEDG). The challenges in detection and localization of faults in active distribution networks are addressed by the proposed approach. The proposed approach is based on a long short-term memory (LSTM) neural network to detect and localize faults based on measurements at the point of common coupling of distributed energy resources (DERs) within the network. The proposed scheme is implementable at the grid-edge in active distribution networks for real-time detection, classification, and localization using DERs as a grid probing tool to enhance the situational awareness of futuristic PEDG. This work includes a detailed theoretical analysis and case studies that evaluate the performance of the proposed LSTM-based fault detection and localization in active distribution networks. A comprehensive database is created for the training process, and the network operates with optimized hyperparameters. The proposed method is examined for a modified IEEE 14-bus network dominated by DERs. The results demonstrate promising performance and very fast (i.e., within one line cycle) fault detection and localization that enhances the situational awareness of the system.**

***Index Terms*— Modern Power Systems, Distributed Energy Systems, Power Electronics Dominated Grid, Microgrid, Long Short-term Memory, Artificial Neural Networks, Line-Line faults, Anomaly Classification,**

I. Introduction

The increasing penetration of distributed energy resources (DERs) is transforming power systems into a new concept commonly referred to as power electronics dominated grid (PEDG) [1]. In this energy paradigm shift, DERs are introducing a new active distribution network which has various vulnerabilities as well as features that are not fully studied and revealed yet. In traditional power systems, various types of faults may occur that can cause catastrophic blackouts and endanger the safety of power grid devices. In addition to these classical faults, PEDG is introducing a new surface of faults and anomalies as well. These new surfaces of faults include power electronics failures, and cyber layer failures/anomalies, etc. Thus, it is crucial for resilient operation of PEDG to identify, classify, and localize these types of faults at an early stage to quickly clear faults and protect the rest of the grid. Detection and localization of these faults in real-time is a challenging task that is not fully addressed in literature.

Detecting faults and anomalies in PEDGs can be done through a variety of fashions. Grid operators aim to incorporate fault classifications, which provide actionable intelligence to mitigate the impact of the fault after it occurs. Two common detection techniques are model-based solutions and data-driven solutions. Model based solutions find faults based on predictions, tracking, and deviations from a model of the system and the system itself [2, 3]. This contrasts with data-driven systems which begin with data collected from a world system or a robust digital twin which replicates the system to high accuracy [4-8]. Both of these systems could potentially classify and identify anomalies within a PEDG as well as system faults within the grid [9-12]. While both approaches have a potential for successful implementation, data-driven schemes are considered for this work. This is because they may prove to be more adaptable to changing topologies and specific local controllers than model-based solutions which can deviate significantly as power system topologies are altered and no longer fit the model.

When considering data-driven approaches for the PEDG, artificial intelligence (AI) and neural network (NN) based anomaly detection schemes are especially promising. NN approaches have been successful in determining changing grid conditions, multi-classification of faults and various other useful detection and correction schemes in PEDGs[13]. While classical algorithms, both simple and complex, are capable of detecting faults and anomalies, classification anomalies must be created for specific classifications. An algorithm for one fault cannot be used for another. NNs have an advantage for multi-class distinction for power system faults. A single NN-based detection network can classify wider ranges of faults with potentially unrelated transients.

When considering NN approaches, the exact type of NN used is an important decision. Many types of NN have shown potential in power systems, including feedforward NNs [14], convolutional NNs [15, 16], and graph NNs [17]. Deep NNs are especially useful for processing time series data, as instantaneous predictions fall short of making connections between subsequent data entries. The long short-term memory (LSTM) system is a deep NN topology that has shown potential in PEDGs. The LSTM network utilizes update and forget gates to "remember" previous data states [18]. This has allowed for proper classification in PEDG in [13], and the incorporation of additional anomaly classifications would further improve the utility of the scheme.

In PEDGs, detection and correction of anomalies take a different form compared to the traditional power system. Anomaly detection in PEDGs should include data specific to

transients of the power electronics grid, be robust to system uncertainties, designed to classify multiple anomalies, and detect anomalies in real time. A research gap exists for networks capable of addressing all these concerns simultaneously. Line-line fault detection techniques for PEDGs require unique solutions to address this literature gap. The following concerns thus are goals of the proposed network:

1) fast detection of faults (ideally in less than one line cycle) in bilateral power flow grids to allow for fast mitigation and improved resiliency,

2) localization of faults to assist in physical correction and fault clearing,

3) fault detection framework which allows for easy "plug and play" into other anomaly classification techniques,

4) scalability for PEDGs with changing topologies without complete reconfiguring of the system when DERs are connected to or islanded from grid.

To meet these concerns, this work proposes a novel data-driven method for transient and incipient fault location detection to integrate in a multi-class neural network detection system. The main contribution of this paper is to realize a real-time AI-based fault detection and localization in active distribution network of PEDGs using the DERs and their existing sensors as a probing apparatus at the grid-edge. This approach aims to make classifications from the voltage and current data of point of common coupling (PCC) of DERs. The proposed approach can be integrated within the hierarchical control structure of PEDG for optimal fault clearing mechanism. Thus, the proposed approach enhances the situational awareness of the active distribution networks in PEDG using the existing infrastructure of DERs. This addresses the shortcomings of the existing protection schemes in traditional power grid given its real-time feature and localization scheme to enhance the situational awareness and resiliency of the grid as well as minimizing the power outage in post-fault.

The remainder of this paper discusses the training data collection, training process, and verification case studies to support the LSTM local classification principle.

II. Data Collection And Classification

The collection of training data is a fundamental aspect of any NN-based approach. Quality training data must incorporate both normal and anomalous conditions of the system under study. The system in this work is a modified IEEE 14 bus system dominated by DERs as demonstrated in Fig. 1.

The data collection process consists of several batches simulating a line-to-line fault in one of the 14 lines of the PEDG. The training data is collected using a 14-bus system in MATLAB/Simulink. By using simulation data, NNs can take advantage of the fact that faults and other anomalies can be emulated in a simulated environment as a digital twin of real network without doing any damage to a physical grid or with only relying on limited real-data which may impact the performance of NNs. For the training data collected in this work a fault is triggered in each of the 14 lines of the system. Data is collected for two seconds of simulation time, half of which is a no-fault condition, and the other half consists of the faulty data. Each bus independently collects its voltage and current data for its own LSTM detection network. Thus, while each local neural network is training, only data obtained from the local bus of the total training data is considered. This reduces the simulation time on a per bus basis. All data is eventually considered, as a total of 14 local LSTM networks are created. The input voltages and current become an input

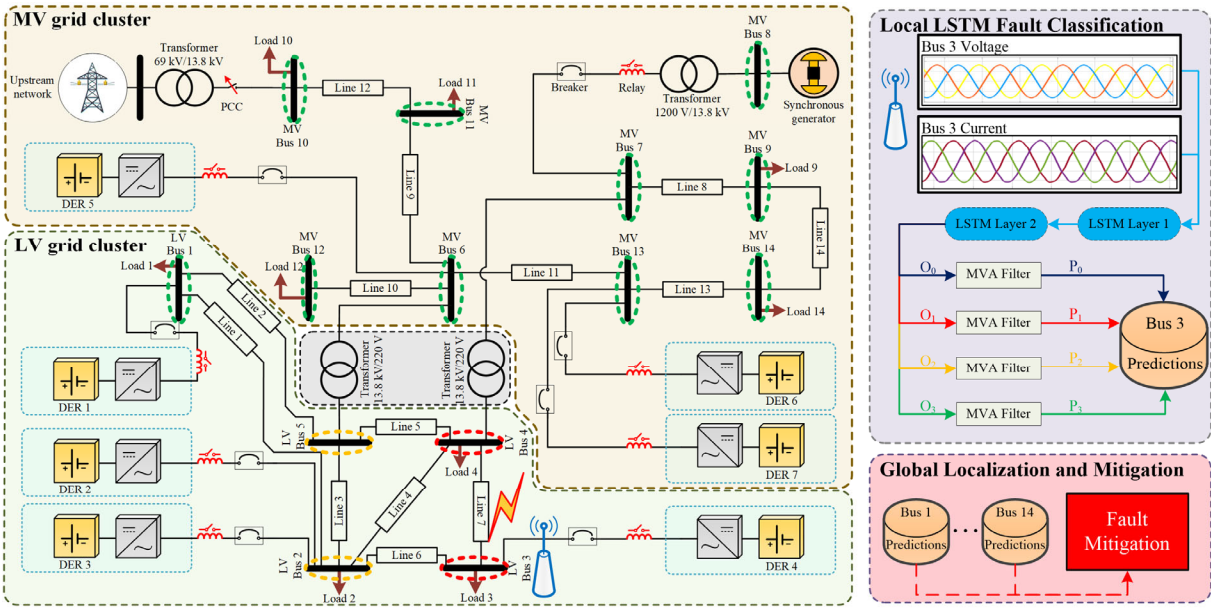

Fig. 1 The modified 14-bus system considered in this work. In this figure, a line-line fault has occurred in Line 7. This results in a Class 1 (Red) classification in Buses 3 and 4, a Class 2 (Orange) classification in Buses 2 and 5, and a Class 3 (Green) classification in all remaining buses. The LSTM classification network for Bus 3 is visualized.

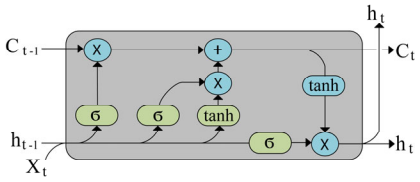

Fig 2. The fundamental LSTM module

sequence for the LSTM training process. The training data is partitioned into smaller sections for LSTM training, and in this work each LSTM training section consists of 500 data points sampled at 100kHz. Thus, the sample size fed into the LSTM has a period of 5ms. The total data collected for training is divided into three groups: 70% of the data is considered training data, 20% is validation data, and 10% is testing data.

For training of the LSTM based neural network, the PCC voltage and current measurements of DERs are used as the network inputs. Thus, we are using the DERs sensors readings for the primary local controller for the proposed AI-based fault location classification. The primary local controller of DER's communicates with the supervisory global controller for coordination in PEDG. The voltage and current data from each bus is used to detect and classify line-line faults. These data are accessible in the supervisory global control layer of PEDG for localizing the transient and incipient faults. Once the data from local controller has made a prediction concerning a fault, the global controller makes the ultimate detection and mitigation decisions. Thus, the 14-bus system in this work consists of 14 local LSTM networks with one global controller. Data is iteratively collected for each bus during a transient fault in each of the 14 lines in Fig. 1, as well as during no fault. The 3-phase current and voltage data are divided into 5ms periods, representing the 6 features of the sequential input to the LSTM. The data is classified into one of four classes: Class 0: No fault in the system, Class 1: Fault in a line connected to the bus, Class 2: Fault in a line of an adjacent bus, Class 3: Fault in a line more than one bus away from the bus. The description of Fig. 1 visualizes these classes for a fault in Line 7.

III. Implementation of the Proposed LSTM Network

The NN used in the study is the LSTM topology. LSTM is used to incorporate previous data from the network to make sequential estimations of classification. LSTMs, a category of recurrent NNs, can incorporate memory of recent events. This is opposed to feed forward NNs, which rely solely on instantaneous data. This is especially useful in fault detection as it allows for interpretation of short-term transients and previous operating conditions as opposed to being solely reliant on instantaneous values.

The fundamental LSTM cell is shown in Fig. 2. These gates operate via the formulation given by (1)-(4) as first presented in [18]. These formulations include the Forget, Update, and Output gates. An explanation of each is enumerated below. The first gate is the Forget gate **F** of (1),

$$\mathbf{F} = C[t-1] \circ \sigma(W_F X[t] + R_F h[t-1] + b_F) \quad (1)$$

Table. I: LSTM Training Parameters

Parameter	Value
Maximum Epochs	50
Hidden Layer 1 Size	25
Hidden Layer 2 Size	10
Mini Batch size	64
Validation Frequency	200
Gradient Threshold	1
Optimizer	Adam
Data sample length	500
Dropout Rate	0.1

Table. II: LSTM Training Accuracy

Bus	Validation Accuracy
1	93.19%
2	82.50%
3	79.58%
4	85.42%
5	77.22%
6	78.33%
7	76.94%
8	89.86%
9	81.39%
10	85.42%
11	81.25%
12	91.53%
13	78.06%
14	80.00%

where $X[t]$ is the NN input sequence and $h[t\text{-}1]$ is the hidden state. The variable $[t]$ represents the current state, and $[t\text{-}1]$ the previous state. The Update gate **U** is defined by (2),

$$\mathbf{U} = C[t] = \mathbf{F} + \sigma(W_1 X[t] + R_1 h[t-1] + b_1) \circ \tanh(W_U X[t] + R_U h[t-1] + b_U) \quad (2)$$

The Update gate is used to update future states and predictions based upon new information, combining it with previous observations. Via the update gate, the information which passed **F** is thus updated based on new information.

The final gate is the Output gate **O** given by (3) which is used to determine the hidden state h[t] given by (4),

$$\mathbf{O} = \sigma(W_o X[t] + R_o h[t-1] + b_o) \quad (3)$$

$$h[t] = \mathbf{O} \circ \tanh(\mathbf{U}) \quad (4)$$

O influences the output of the cell for determining the hidden state of the cell. This finally produces the cell outputs $C[t]$ and $h[t]$. With these gates, the "long short-term memory" cell is complete. $C[t]$ represents the long-term influence and $h[t]$ the short-term influence of the cell's memory. The LSTM then produces four outputs, this is the probability the time series data belongs to each of the four classes.

The training tunes the weights of the LSTM network. The weights are defined as: the hidden input weights W_x, the hidden state weights R_x, and the biases b_x. The subscript x denotes the function and gate of the corresponding parameter. Values of W_x, R_x, and b_x are calculated for each gate. In this paper, the structure of each LSTM network is identical, consisting of the same 8 layers as detailed in Section IV. The dropout layer has a rate of 0.2.

Once training occurs for the LSTM, the weights and biases are tuned to account for the previous data and allow for proper categorization. The input size of the NN depends on the number of data points considered, and the output is determined by the number of classifications needed for the system. Additionally, the LSTM system has tunable hyperparameters which affect the size of the network, the network accuracy, the computational burden of the network, and the time needed to train the network. Hyperparameter tuning can impact the accuracy and processing power of the LSTM.

IV. Training Process

From the collected training data, the LSTM training process is designed for high accuracy and fast detection. Each bus undergoes a similar training process for classification of faults. The data for a specific bus is used to train an LSTM network with 6 features and 4 classes. The LSTM network for each system has identical hyperparameters. The first layer of hidden units in the LSTM has 25 units and the second hidden layer has 10 units. These hidden units model the LSTM "memory".

The selection of the hyperparameters for the LSTM affects the accuracy, training time, and computational burden of the LSTM. While a comprehensive search for optimal hyperparameters is beyond the scope of this work, hidden unit size of 50x25 has shown previous success [19], and thus these hyperparameters were selected for this work. Other training properties such as the number of epochs, sample size, validation frequency, etc., were selected to minimize the training time without sacrificing accuracy, and are given in Table I.

The system is trained to detect faults from the training data. To reduce training time during hyperparameter tuning, the data is processed to include a balanced dataset approximately 10% of the size of all the collected data. All networks trained have an identical layer structure. It consists of 8 layers in the following order: 1) Sequence Input, 2) LSTM Layer 1, 3) Dropout, 4) LSTM Layer 2, 5) Dropout, 6) Fully Connected, 7) Softmax, 8) Classification Output.

The accuracy of the LSTM of each bus is given in Table II.

The output of the LSTM, $\boldsymbol{O_c}$, is given by (3). In this work, $\boldsymbol{O_c}$ is modified via a moving average filter given by (5) to mitigate false positives,

$$\mathbf{P}_c = \sum_{\tau=0}^{L} \frac{O_c\left[t-1\right]}{L} \tag{5}$$

where c is a Class 0, 1, 2, or 3, $\mathbf{P_c}$ is the prediction of the Class c, L is the number of samples considered in the filter, $\boldsymbol{O_c}$ is the LSTM output of class c for the given bus, and t is the simulation time. In this work, we have taken L = 500. $\mathbf{P_c}$, a number from 0 through 1, is compared to a threshold constant $\mathbf{P_t}$ to prevent unnecessary mitigation. $\mathbf{P_t}$ can be determined for each LSTM based on its accuracy, or the same values for each local LSTM. In this work, we have taken $\mathbf{P_t}$ = 0.51 for all

Table. III: LSTM Metrics

	Class 3			Class 2			Class 1			Class 0		
Bus	Precision	Recall	F_1 Score	Precision	Recall	F_1 Score	Precision	Recall	F_1 Score	Precision	Recall	F_1 Score
1	0.995	0.995	0.995	0.980	1.000	0.990	0.775	0.886	0.827	0.775	0.886	0.827
2	0.854	0.980	0.913	0.892	0.917	0.904	0.450	0.439	0.444	0.450	0.439	0.444
3	0.854	0.983	0.914	0.660	0.532	0.589	0.900	0.857	0.878	0.900	0.857	0.878
4	0.867	0.989	0.924	0.611	0.468	0.530	0.771	0.841	0.804	0.771	0.841	0.804
5	0.872	0.926	0.898	0.600	0.646	0.622	0.653	0.914	0.762	0.653	0.914	0.762
6	0.791	0.923	0.852	0.542	0.456	0.495	0.947	1.000	0.973	0.947	1.000	0.973
7	0.748	0.932	0.830	0.913	0.457	0.609	0.588	0.500	0.540	0.588	0.500	0.540
8	0.933	0.915	0.924	0.955	1.000	0.977	1.000	0.833	0.909	1.000	0.833	0.909
9	0.831	0.972	0.896	0.895	0.596	0.716	0.936	0.957	0.946	0.936	0.957	0.946
10	0.798	0.962	0.872	1.000	0.522	0.686	0.771	0.871	0.818	0.771	0.871	0.818
11	0.761	0.950	0.845	0.914	0.500	0.646	0.627	0.865	0.727	0.627	0.865	0.727
12	0.883	0.982	0.930	1.000	1.000	1.000	0.956	0.977	0.966	0.956	0.977	0.966
13	0.812	0.858	0.834	0.804	0.976	0.882	0.348	0.235	0.281	0.348	0.235	0.281
14	0.816	0.863	0.839	0.545	0.387	0.453	0.971	0.895	0.931	0.775	0.782	0.778

The data is classified into one of the four classes: Class 0 denotes no fault in the system; Class 1 denotes a fault in a line connected to the bus; Class 2 denotes a fault in a line of an adjacent bus; and Class 3 denotes a fault in a line more than one bus away from the bus.

LSTMs. When $\mathbf{P_c} > \mathbf{P_t}$, the local data for LSTM alerts the global controller for localization of transient and incipient faults. The global controller utilizes LSTM predictions from all 14-buses to make the final classification. In this work, the global controller triggers a final classification when it receives a Class 1 detection from adjoined buses.

V. Results and Discussion

A. Testing Data Verification

With the LSTM network fully trained, verification of the network is necessary to ensure it is fully capable of detecting faults in the IEEE 14 bus system. The first verification is done through the testing data gathered during the data collection process. During training, the validation data is used to create the validation accuracy seen in Table II. However, in order to ensure that the LSTM classification system can properly classify data beyond that employed during training, the testing data must be utilized. The first verification of the proposed LSTM model is demonstrated in the confusion matrix shown in Fig. 3. The confusion matrix shows how the trained LSTM predicts fault classes for a bus with the testing data. It compares the predicted class on the x-axis and the true class on the y-axis. The confusion matrix exhaustively shows the number of classifications for all possible scenarios; 16 or 2^4 scenarios are possible because of the four classes. Fig. 3 shows the confusion matrix for the Bus 1 local LSTM classification. To conserve space, the confusion matrices of other 13 buses are not shown.

In addition to the confusion matrix, all 14 local LSTM networks corresponding to 14 buses are verified using the precision, recall, and F_1 score of each bus. These metrics are tabulated for each class for the 14 buses as shown in Table III. The precision metric highlights the ability of the network to only make correct classifications, the recall metric highlights the network's ability to predict all occurrences of a fault, and the F_1 score is a combination of both precision and recall metrics as given in (6).

$$F_1 score = \frac{2 \times precision \times recall}{precision + recall} \tag{6}$$

The F_1 score can range from 0 to 1, where scores closer to 1 are superior. Thus, as seen from the metrics in Table III, the local LSTM method exhibits great performance in terms of all the metrics across nearly all buses and classes.

B. Case Study Verification

A verification case study is performed in MATLAB/Simulink that validates the proposed LSTM-based fault location detection system and further validates its feasibility in a PEDG using DERs as grid-probing tool. The verification is completed using the 14-bus system displayed in Fig. 1. In this study, a fault in Phase A-B in Line 1 of the system occurs at 11 s. The local LSTM of Bus 1 operates during the study, as real time voltage and current data is provided to the detection network. During each sampling

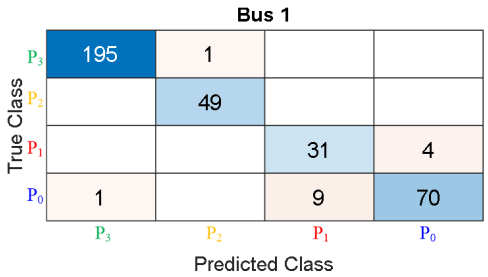

Fig. 3 Confusion matrix of the LSTM detection of faults in Bus 1. The diagonals indicate correct predictions of the training data, while off-diagonals are incorrect predictions

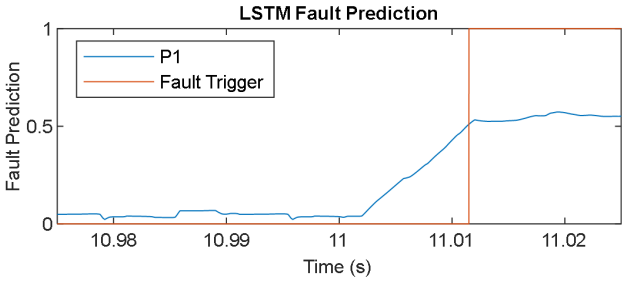

Fig. 4 Prediction of a Fault in a line connected to Bus 1.

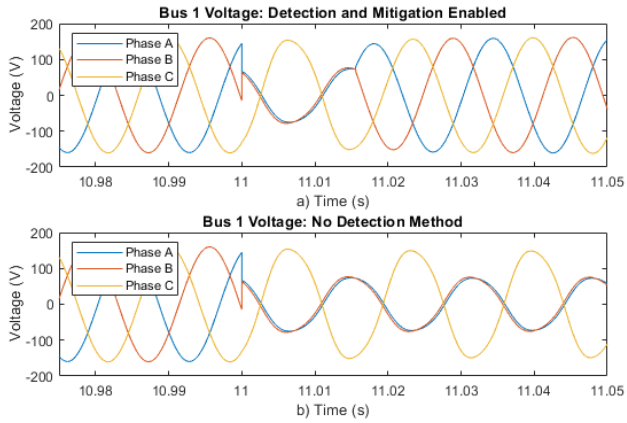

Fig. 5 Bus 1 voltage a) with the proposed LSTM and fault mitigation mechanism, b) without the proposed fault detection and mitigation system.

instance, the LSTM predicts the probability of each of the four fault classes. The prediction of each class is regulated by the supervisory controller to decide when a fault has occurred.

The supervisory controller in this study takes the moving average of each prediction, as described by (5). For determining a fault in this case study, the output prediction $\mathbf{P_1}$ for Bus 1 is pivotal. $\mathbf{P_1}$, the output of the moving average filter of Bus 1, is displayed in Fig. 4. After t=11s, the local detection LSTM predicts with higher certainty that a fault has occurred at Bus 1. Finally, the LSTM prediction $\mathbf{P_1}$ exceeds the 0.51 trigger threshold. This trigger occurs at 11.49 ms after the fault. Thus, the trigger meets the goal of detection within a 60 Hz line cycle of the fault, where the deadline for fault detection/triggering is 16.67ms. The local detection network communicates a fault trigger to the supervisory controller, which instantaneously trips breakers in Line 1. It is to be noted

that we can reduce the sample window size of the moving average filter in (5), and further improve the timeliness of the proposed fault trigger scheme at the expense of somewhat increased false positive rates.

The mitigation effects are demonstrated in Fig. 5. Here, the timescale is expanded to display the voltage at Bus 1 with and without the proposed LSTM detection. Fig. 5a demonstrates the impact of the detection system, as the proposed real-time detection shows no significant impact on the system one line cycle after the fault. Fig. 5b demonstrates the system without the local detection network. Without the proposed network, the system remains unbalanced which ultimately will trigger the protection relays at various locations. As the supervisory controller is not able to quickly locate where the fault occurred without a fault detection and localization system, the fault can introduce demand-supply stability challenges due to the loss of DERs generation without identifying the fault location to minimize the power outage. Thus, the simulated case study verifies that the proposed LSTM network has accurately been able to determine fault detection and localization which informs the supervisory controller of exactly where the fault occurred. This enables the supervisory controller to make the best mitigation decision possible to the entire PEDG.

VI. Conclusion

This paper presents a real-time AI-based fault detection and localization technique in PEDG. The training process, data collection, network testing verification, and a verification case study is provided for a real-time location detection of faults in a PEDG. The training data is collected from a MATLAB/Simulink model of an IEEE 14 bus system specifically tuned to be a realistic representation of a PEDG. The proposed LSTM network is trained for each bus in an IEEE 14-bus grid. The local fault detection system operates with the three-phase voltage and current data obtained at a local bus in a PEDG. The voltage and current measurements are inputs to an LSTM network, which predicts one of the four identified classes. Each class represents the fault status of the bus as a part of a larger system: three classes indicate a fault exists in the system and predicts the distance between the bus and the fault, while the fourth class indicates there is no fault in the PEDG. The process of training the network from this data is detailed. After the training of each local LSTM, the network is verified through two separate processes. First, the testing data verifies the precision and recall of each network and each class. Second, a case study verifies the online operation of the detection network by creating a fault in Line 1 of the PEDG. The case study results reveal that the local LSTM at Bus 1 can detect the fault and quickly trigger breakers in less than one line-cycle. Thus, the local LSTM line-line fault detection system is capable of detecting faults and can be trusted for operating on individual buses. It communicates fault triggers to a supervisory controller for proper mitigation at each bus of a PEDG. The local controller scheme improves scalability as it accommodates changes in the number of DERs in a PEDG.

References

[1] Q. Peng, et al , "On the Stability of Power Electronics-Dominated Systems: Challenges and Potential Solutions," *IEEE Transactions on Industry Applications,* vol. 55, no. 6, pp. 7657-7670, 2019.

[2] K. Manandhar, et al, "Detection of Faults and Attacks Including False Data Injection Attack in Smart Grid Using Kalman Filter," *IEEE Trans. on Control of Network Systems,* vol. 1, no. 4, pp. 370-379, 2014.

[3] W. Ao, Y. Song, and C. Wen, "Adaptive cyber-physical system attack detection and reconstruction with application to power systems," *IET Control Theory & Applications,* vol. 10, no. 12, pp. 1458-1468, 2016.

[4] A. Abdullah, "Ultrafast transmission line fault detection using a DWT-based ANN," *IEEE Transactions on Industry Applications,* vol. 54, no. 2, pp. 1182-1193, 2017.

[5] M. Ozay, et al "Machine learning methods for attack detection in the smart grid," *IEEE trans. on neural net. and lear. sys.,* vol. 27, no. 8, pp. 1773-1786, 2015.

[6] J. Yan, H. He, X. Zhong, and Y. Tang, "Q-learning-based vulnerability analysis of smart grid against sequential topology attacks," *IEEE Trans. on Information Forensics and Security,* vol. 12, no. 1, pp. 200-210, 2016.

[7] S. A. Foroutan and F. R. Salmasi, "Detection of false data injection attacks against state estimation in smart grids based on a mixture Gaussian distribution learning method," *IET Cyber-Physical Systems: Theory & Applications,* vol. 2, no. 4, pp. 161-171, 2017.

[8] F. Li *et al.*, "Detection and Identification of Cyber and Physical Attacks on Distribution Power Grids With PVs: An Online High-Dimensional Data-Driven Approach," *IEEE Journal of Emerging and Selected Topics in Power Electronics,* vol. 10, no. 1, pp. 1282-1291, 2022.

[9] B. Lu and S. K. Sharma, "A Literature Review of IGBT Fault Diagnostic and Protection Methods for Power Inverters," *IEEE Transactions on Industry Applications,* vol. 45, no. 5, pp. 1770-1777, 2009.

[10] A. M. S. Mendes and A. J. M. Cardoso, "Voltage source inverter fault diagnosis in variable speed AC drives, by the average current Park's vector approach," in *IEEE IEMDC*, 9-12 May 1999 1999, pp. 704-706.

[11] K. Rothenhagen and F. W. Fuchs, "Performance of diagnosis methods for IGBT open circuit faults in three phase voltage source inverters for AC variable speed drives," in *2005 European Conference on Power Electronics and Applications*, 11-14 Sept. 2005 2005, pp. 10 pp.-P.7.

[12] R. Peuget, S. Courtine, and J. Rognon, "Fault detection and isolation on a PWM inverter by knowledge-based model," *IEEE Transactions on Industry Applications,* vol. 34, no. 6, pp. 1318-1326, 1998.

[13] M. Baker, A. Y. Fard, H. Althuwaini, and M. B. Shadmand, "Real-Time AI-Based Anomaly Detection and Classification in Power Electronics Dominated Grids," *IEEE Journal of Emerging and Selected Topics in Industrial Electronics,* vol. 4, no. 2, pp. 549-559, 2023.

[14] M. W. Baker, et al, "Artificial Intelligence based Anomaly Detection and Classification for Grid-Interactive Cascaded Multilevel Inverters," in *2022 SGRE*, 20-22 Mar. 2022, pp. 1-6.

[15] L. Zhang, et al, "A Fault Diagnosis Method of Power Transformer Based on Cost Sensitive One-Dimensional Convolution Neural Network," in *ACPEE*, 4-7 June 2020 2020, pp. 1824-1828.

[16] X. Qu, B. Duan, Q. Yin, M. Shen, and Y. Yan, "Deep Convolution Neural Network Based Fault Detection and Identification for Modular Multilevel Converters," in *IEEE PESGM*, 2018, pp. 1-5.

[17] H. Sun, et al, "Distribution Fault Location Using Graph Neural Network with Both Node and Link Attributes," in *IEEE ISGT Europe*, 18-21 Oct. 2021 2021, pp. 1-6.

[18] S. Hochreiter and J. Schmidhuber, "Long Short-term Memory," *Neural computation,* vol. 9, pp. 1735-80, 12/01 1997.

[19] M. Baker and M. B. Shadmand, "An LSTM-based Anomaly Classification Framework for Power Electronics Dominated Grids," in *IEEE PECI*, 2-3 March 2023 2023, pp. 1-7.