

Physics-Informed Learning for High Impedance Faults Detection

Wenting Li

Center for Non-Linear Studies (CNLS)
Los Alamos National Laboratory
Los Alamos, NM
wenting@lanl.gov

Deepjyoti Deka

Theoretical Division
Los Alamos National Laboratory
Los Alamos, NM
deepjyoti@lanl.gov

Abstract—High impedance faults (HIFs) in distribution grids may cause wildfires and threaten human lives. Conventional protection relays at substations fail to detect more than 10% HIFs since over-currents are low and the signatures of HIFs are local. With more μ PMU being installed in the distribution system, high-resolution μ PMU datasets provide the opportunity of detecting HIFs from multiple points. Still, the main obstacle in applying the μ PMU datasets is the lack of labels. To address this issue, we construct a physics-informed convolutional auto-encoder (PICAe) to detect HIFs without labeled HIFs for training. The significance of our PICAe is a physical regularization, derived from the elliptical trajectory of voltages-current characteristics, to distinguish HIFs from other abnormal events even in highly noisy situations. We formulate a system-wide detection framework that merges multiple nodes' local detection results to improve the detection accuracy and reliability. The proposed approaches are validated in the IEEE 34-node test feeder simulated through PSCAD/EMTDC. Our PICAe outperforms the existing works in various scenarios and is robust to different observability and noise.

Index Terms—High impedance faults Detection, Convolutional neural networks, Auto-encoder, μ PMU, Physics informed,

I. INTRODUCTION

Energized conductors hitting the high impedance ground surfaces, usually accompanied by arc flashing, have led to most HIFs [1]. People are concerned with HIFs, as they are one of the main causes/initiators of destructive wildfires and threaten public safety. Diversity of physical models have well described the process of HIFs of randomness and nonlinearity [2]. However, more than 10% detection failures of HIFs have been reported [3] using voltages or currents measured by devices at relays or breakers [4]. Conventional over-current protection systems often neglect HIFs due to the low fault current [1, 3]. This problem is exacerbated in distribution grids as measurements are not ubiquitous, and signatures of HIFs are local and do not propagate much in the grid. In recent years, there has been growing interest in detecting HIFs in distribution grids accurately when more μ PMUs being installed.

The existing data-driven HIF detection methods usually separate HIFs from others by supervised classification with

various features, based on time-domain, frequency-domain, and time-frequency domain measurements [1, 3, 5, 6]. However, these methods are either not robust to noise and low harmonics rates or require a sufficient number of labeled datasets¹ to learn the features.

To address these issues, we propose a novel and practical HIF neural-network-based detector for distribution grids with limited measurement availability that uses only normal data *and no labeled faults* for training. Neural networks have achieved great success in computer vision, natural language processing, and health care [7]. While applications with labeled data are many, success with partially labeled or even completely unlabeled datasets has been demonstrated with satisfactory accuracy and efficiency [8, 9]. One label-free model is Auto-encoder (AE) [7], a neural network architecture consisting of an encoder and a decoder to learn the features and reconstruct the data. Various AE derivatives have been proposed for specific applications [8, 9]. However, such pure data-driven applications take the risk of violating the physical rules that govern the cyber-physical systems such as power grids. Hence, our method can overcome these issues by judiciously using constraints related to the *physics-informed dynamics* in regular operation during the detector training.

We are inspired by the recent attempts of embedding physical laws in neural networks or statistical machine learning for power flow calculation, state estimation, topology learning [10, 11, 12], and power system monitoring [13, 14]. Outside of power grids, [15] reveals promising progress in regulating the learned parameters of neural networks with physical laws as priors. These physics-based promotions improve both interpretability as well as the model's computational efficiency.

Contribution: We propose a physics-informed learning framework to detect HIFs, on the conditions of a limited number of measured nodes and scarce labeled faults for training. Explicitly, relying on the fact that elliptical curves can model the trajectory of normal voltage-current with time, we construct a Convolutional Auto-Encoder (CAE) to represent the voltage time-series data during normal operations (no faults). Additionally, we constrain its output with the physics-regulated (PR) elliptical characteristics of voltages and cur-

The authors acknowledge the support from the Department of Energy through the Advanced Grid Modeling (AGM) Program, and the Center for Non Linear Studies (CNLS) at Los Alamos National Laboratory.

¹Labeled datasets denote the types of the recorded datasets are tagged.

rents. Furthermore, as HIF's signatures are local, we establish a low-communication central scheme that merges the observed nodes' local decisions to augment the detection robustness and reliability. We validate the proposed methods in the IEEE 34 node benchmark system [16] simulated by Power Systems Computer-Aided Design (PSCAD) [17]. We demonstrate our detector's high performance even when systems are not fully observed and interpret the physics-informed regularization's advantages to distinguish HIFs from others. Moreover, we show that PICAE outperforms existing schemes on HIF detection in multiple noisy scenarios.

The remaining part of this paper is organized as follows: Section II introduces the physical rules of HIFs; based on these rules, we construct a physics-informed convolutional auto-encoder (PICAE) to detect HIFs in Section III; the detection framework of local and central determination are presented in Section IV; numerical experiments in Section V show the detection performance of the proposed approaches, in comparison with some existing works in different scenarios. Section VI concludes the main results.

II. BACKGROUND OF THE PHYSICAL MODEL FOR HIF

HIF is a nonlinear, random event that is often unnoticeable by over-current relays or fuses. In the last decades, various arc models have been utilized to describe the stable or dynamic HIF process [1, 2, 18]. Two-parallel diodes and a voltage source model accurately represent the dynamic re-striking and quenching process of arcs during HIF *at the fault point*. [1, 5].

A. Modeling of HIF Process

Let $v(t)$ be the single phase voltage at the time t that interacts with the two DC voltage sources $V_p > 0, V_n < 0$, and variable resistances $R_p \neq R_n$ in the down and up lines.

$$v(t) = \begin{cases} V_p + i_p(t)R_p & \text{if } v(t) > V_p \\ V_n - i_n(t)R_n & \text{if } v(t) < V_n \\ v(t-1) & \text{else} \end{cases} \quad (1)$$

When $v(t) > V_p$, the diode D_p is switch on to allow fault current i_p to flow through, and when $v(t) < V_n$, the diode D_n is switch on to let i_n flow in. These structures mimic the re-striking process of arcs; otherwise, no currents flow through the HIF circuit and the voltages of the fault point keep the same with the previous phase voltage $v(t-1)$, which represents the quenching of arcs. Note that the re-striking and quenching process will *cycle and last* for seconds or even longer [19]. This process is random and nonlinear since the impedance R_n, R_p are randomly varying.

B. Physical Laws of HIFs

On normal conditions, it is demonstrated that the trajectories of voltages and currents are rotated ellipses for resistance-inductive or resistance-capacitive linear circuits, and are circles if resistance is zero [5]. Let phase voltages and currents be $v(t) = V_0 \cos(\omega t), c(t) = C_0 \cos(\omega t - \phi)$ with a phase angle

ϕ , then we can fit them into the standard parametric format of a rotated ellipse equation as follows:

$$\left(\frac{v(t)}{\alpha_1} + \frac{c(t)}{\alpha_2}\right)^2 + \left(\frac{v(t)}{\alpha_3} - \frac{c(t)}{\alpha_4}\right)^2 = 1 \quad (2)$$

where $\alpha_1 = 2V_0 \cos(\phi/2), \alpha_2 = 2C_0 \cos(\phi/2), \alpha_3 = 2V_0 \sin(\phi/2), \alpha_4 = 2C_0 \sin(\phi/2)$, where α_i are determined by line impedance and system power flow.

Once HIF occurs, parameters $\alpha_1, \dots, \alpha_4$ are immediately altered, but as the circuit is not open, *the trajectory is still approximate elliptical with different parameters as R_n, R_p vary*.

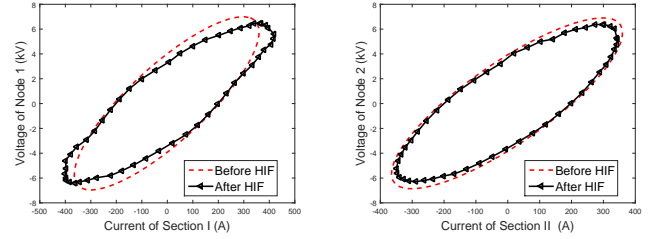


Fig. 1: The trajectories of voltages and currents at node 1 and 2 in the four node test feeder system [20], where the line section I is between node 1 and 2 and section II is between node 2 and a transformer. Red curves are the voltages-current trajectories on normal conditions while the black ones are after HIF event

Four-node test feeder example: We illustrate the physical property of HIFs in the four-node test feeder [20] simulated by PSCAD/EMTDC [17]. When a HIF occurs near node 1, the trajectories of voltages and currents at node 1 and 2 are impacted.

Fig. 1 compares the trajectories before and after HIF event. It is clear that the black trajectory deviates from the red one to formulate another approximate ellipse, and the deviation is more serious when the node is closer to the HIF.

As HIF's unique feature is the approximate elliptical trajectory of voltages and currents, varying from node to node, we present in the next section our detector that regulates the learning process in training by the elliptical trajectory without relying on sparsely available and expensive labels.

III. PHYSICS-INFORMED CONVOLUTIONAL AUTO-ENCODER (PICAE)

The configuration of our PICAE is shown in Fig. 2. Given time series of voltages a matrices $V_i \in R^{T \times N}, i = 1, \dots, m$ as inputs, where m, T, N are the number of measured nodes, the length of the moving window, and the number of windows. According to the physical laws, the elliptical regularization structure in PICAE constrains the weights in the convolution and deconvolution layer. In training, the normal voltages V_i are reconstructed by the encoder, and then regulated by the corresponding line currents to obey the elliptical trajectory. Note that the elliptical regularization is *only employed during offline training*. In testing, only the measured voltages are needed to detect the occurrence of HIFs.

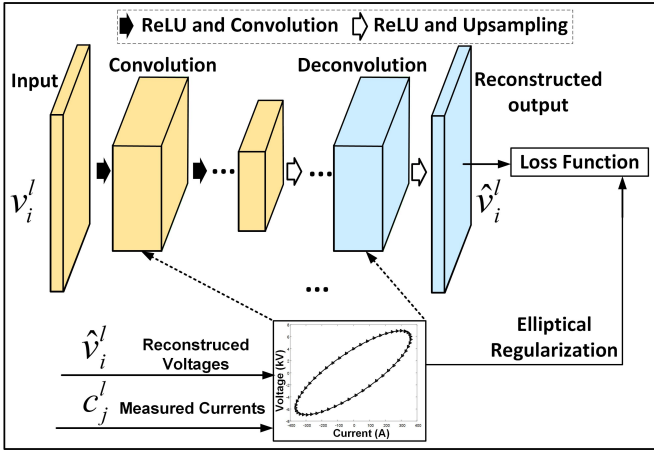


Fig. 2: Physics-Informed Convolutional Auto-encoder

A. The Encoder and Decoder of PICAE

The encoder is a convolutional neural network with decreasing size of the latter layer than that of the previous layer. The s th convolutional layer down-samples the input g^s with filters W^s and bias matrices B^s to reduce the dimensions and then goes through the nonlinear activation function of the Rectified Linear Units (ReLU) to enter into the next layer.

$$g^{s+1} = \max(0, g^s \otimes W^s + B^s), s = 1, \dots, S \quad (3)$$

where \otimes denotes the convolution operation, and $g^1 = v_i^l \in R^T$, the l th column of V_i . The decoder has the symmetric structure with the encoder, which improves the reconstruction accuracy [21]. Here “symmetric” emphasizes the same sizes of the outputs of the deconvolution layer with that of the mirrored convolution layer. The h th deconvolution layer up-samples the inputs f^h with the filters \bar{W}^h and the bias \bar{B}^h through deconvolutional and ReLU operations.

$$f^{h+1} = \max(0, f^h * \bar{W}^h + \bar{B}^h), h = 1, \dots, S \quad (4)$$

where $*$ denotes the deconvolution operation. The final output f^{S+1} is the reconstructed voltages \hat{v}_i^l .

B. Physical Regularization of PICAE

The regularization item acts as prior knowledge that direct the trained model to follow the latent physical rules mentioned in Section II-B, to enhance the robustness against noise and other abnormal events. Our regularization encodes the rotated elliptical trajectory of the nodal voltages against currents. Let time series v_i be the voltage of the i th node in one window, and $c_j \in R^T$ be the current on line connecting i to a neighboring node $j \in \mathcal{N}(i)$. Let $Z_i = [v_i \odot v_i, v_i \odot c_j, c_j \odot c_j, v_i, c_j] \in R^{T \times 5}$, where \odot denotes the entry-wise product. Assuming normal conditions during the T samples, the entries of voltages and currents measurements v_i, c_j ideally follow an elliptical trajectory with five parameters $\beta = [a, b, c, d, e]^T$, expressed as [22]:

$$Z_i \beta + \mathbf{f} = \mathbf{0} \quad (5)$$

where $\mathbf{f}, \mathbf{0} \in R^T$ are an all one and all zero vectors, respectively. The five unknown parameters in β can be estimated by the following least square method, given sufficient number of voltages and currents measurements ($T \geq 5$):

$$\beta^* = \arg \min_{\beta} \frac{1}{2} \|Z_i \beta + \mathbf{f}\|_2^2 = -(Z_i^T Z_i)^{-1} Z_i^T \mathbf{f} \quad (6)$$

Remark: If no clean historical data-sets are present to compute β^* through (6), we can approximate β through power flow analysis. Specifically, as the equations of (2) and (5) are equivalent, β in (5) can be estimated by the corresponding V_0, C_0, ϕ in (2) [22], obtained by power flow analysis on steady states [23].

Training: Given N data samples $v_i^l, c_j^l, l = 1, \dots, N$ of normal operation, the loss function of PICAE for node i is:

$$\mathcal{L}(\Theta) = \frac{1}{N} \sum_{l=1}^N [\|v_i^l - \hat{v}_i^l(\Theta)\|_2^2 + \lambda_r \|\hat{Z}_i \beta^* + \mathbf{f}\|_2^2] \quad (7)$$

Here the first term $\|v_i^l - \hat{v}_i^l(\Theta)\|_2^2$ denotes the mean square errors between the original and reconstructed voltages $\hat{v}_i^l(\Theta)$ with parameters Θ . The second item is the regularization, which uses the estimated β^* to ensure that \hat{v}_i^l follows the elliptical trajectory via $\hat{Z}_i = [\hat{v}_i^l \odot \hat{v}_i^l, \hat{v}_i^l \odot c_j^l, c_j^l \odot c_j^l, \hat{v}_i^l, c_j^l]$. Considering the impact of topological changes in realistic power grids on β , λ_r is set to be a relatively small value to allow some variations of the regularization term $\|\hat{Z}_i \beta^* + \mathbf{f}\|_2^2$, and the β needs to be updated if the trajectory of voltages and currents changes significantly. The training also produces the average reconstructed error $\epsilon_i = \frac{1}{N} \sum_{l=1}^N \|v_i^l - \hat{v}_i^l\|_2^2$ during normal conditions. The training steps are listed in Algorithm 1. In testing, we use the trained PICAE on online voltage $v_i^{l'}$

Algorithm 1 Training of local PICAE

- 1: Input: N training datasets v_i^l, c_j^l , maximum iterations k_{\max} .
- 2: Compute β^* by (6) with v_i^l, c_j^l ; $k \leftarrow 0$.
- 3: **while** $k < k_{\max}$ and early stop is not reached **do**
- 4: Optimize Θ of PICAE by minimizing $\mathcal{L}(\Theta)$ in (7).
- 5: **end while**
- 6: Output: trained PICAE, $\epsilon_i = \frac{1}{N} \sum_{l=1}^N \|v_i^l - \hat{v}_i^l\|_2^2$ on normal conditions.

to reconstruct voltages, and determine the confidence score $\gamma_i = \epsilon_i / \epsilon_i$, the relative error compared to testing, where $\epsilon_i = \|v_i^{l'} - \hat{v}_i^{l'}\|_2^2$ is the mean square reconstructed error of the testing data $v_i^{l'}$. We compare γ_i with two predefined thresholds ξ_1, ξ_2 to distinguish HIFs from other events. The threshold $\xi_1 = \frac{\max_l \|v_i^l - \hat{v}_i^l\|_2^2}{\epsilon_i}$ to discern the normal and abnormal events based on the results of training in Algorithm 1. Then if γ_i of the testing data $v_i^{l'}$ is lower than ξ_1 , the algorithm treats the testing data as normal since the PICAE can well represent normal voltages with a small reconstruction error; another threshold ξ_2 is defined with the maximum confidence score computed by validation data samples of a few HIF events. As HIFs follow the elliptical trajectory, explained in Section II, the

Table I: The variation range of parameters of HIF model

$R_p(\Omega)$	$R_n(\Omega)$	$V_p(kV)$	$V_n(kV)$
600 ~ 1400	600 ~ 1400	5 ~ 6	7 ~ 8

reconstruction errors of HIFs are smaller than those of events, such as capacitor switching, disobeying the elliptical trajectory, which is further explained in Section V-C in details. Thus the primary function of ξ_2 is to distinguish HIFs from other abnormal events. The detailed steps are listed in Algorithm 2.

Algorithm 2 HIF detection through Local PICAE

- 1: Input: Online testing dataset in moving windows $v_i^{l'}, l' = 1, \dots, N'$, averaged reconstruction error ϵ_i for normal voltages of node i , two thresholds ξ_1, ξ_2 .
- 2: Input $v_i^{l'}$ into trained PICAE to reconstruct $\hat{v}_i^{l'}$.
- 3: $\epsilon_i \leftarrow \|v_i^{l'} - \hat{v}_i^{l'}\|_2^2$. Confidence score $\gamma_i \leftarrow \epsilon_i / \epsilon_i$
- 4: **if** $\gamma_i < \xi_1$ **then**
- 5: Output: normal conditions
- 6: **else if** $\gamma_i > \xi_2$ **then**
- 7: Output: Other abnormal events are detected
- 8: **else**
- 9: Output: HIF events are detected
- 10: **end if**

IV. CENTRALIZED HIF DETECTION FRAMEWORK FOR PARTIALLY OBSERVED SYSTEMS

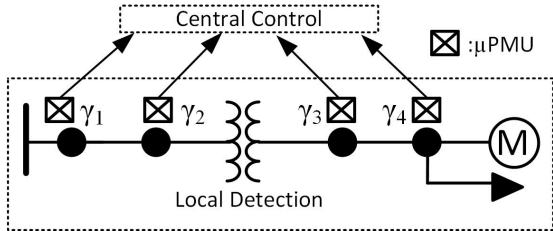


Fig. 3: The configuration of the proposed detection framework. γ_i is the confidence score of the i th measured node for HIF detection

While Algorithm 2 is implemented for each observed node independently, we design a system-wise detection framework in Fig. 3 combining all the local detectors for the partially observed systems. The computed γ_i at each observed node can be communicated to a central detector (Distribution system operator), which decides HIF occurrence using $\max \gamma_i$. Note that we avoid high communication overhead by not relying on the entire voltage sequence to the central detector. The high γ_i scores can also provide auxiliary information about the possible location of the HIF since we observe that the nearby node voltages reveal a relatively high confidence score.

V. NUMERICAL EXPERIMENTS

We validate our approaches in the IEEE 34-node in Fig. 4 with a voltage level of 24.9 kV test feeder [16] modeled by PSCAD/EMTDC[17]. The parameters R_p, R_n, V_p, V_n of HIF

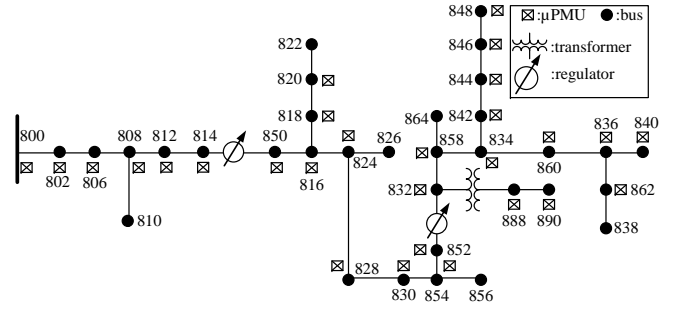


Fig. 4: 34 node testing feeder [16]

models vary in the ranges of Table I randomly at every 1K Hz [1, 6]. We record wave-forms of node voltages and line currents with 512 samples per cycle, and the interval between any two consecutive windows is around four million-second (ms) or $\tau = 128$ samples. Training datasets are composed of $N = 325$ windows of node voltages and line currents. Total 286 testing events in various situations include: 100 HIF events occurring on different branches with varying resistance and DC voltages; 42 different loads switching near the node 890 at various time instants, 54 capacitor switching near the node 844 with the reactive powers in the range of 0.5 to 5 MVA; the remaining 90 normal events with varying initial conditions. We also generate another 10% of testing HIFs events with different random parameters as the validation data for the model selection described in Section III-B. We apply the range normalization to augment the data-sets [24]. The designed PICAE has the symmetrical two convolution-layer structure with the filter W^s size of 5. The number of filters of the two layers change from 32 to 1 to generate hidden variables in a low-dimension subspace. We train the PICAE using the Adam optimizer [25] with a learning rate of 0.0001 and batch of size 12. The maximum iteration $k_{\max} = 1500$ and $\lambda_r = 200$ in (7). Note that we present our major results here, but more extensive explanations and experimental results are in the supplement materials [26].

A. Performance Metrics

We evaluate the detection performance with three criteria: **Precision, Recall and F1 score** [27]. A high “precision” demonstrates that the detector has a low mistake rate of identifying non-HIF as HIF events. A large “recall” value means that the detector has a strong capability to recognize HIF events from others. “F1 score” is a weighted average of precision and recall, and comprehensively evaluates the capability of the detector.

B. Detection Performance with Partial Measurements

To investigate the detection performance for the distribution system without full observability, we show the detection performance when only 24% to 6% (or 8 to 2) nodes are measured. We compare the detection performance when the placement of the measured nodes are “random” (the averaged performance after 100 times of uniformly random selection)

and “selected” (determined by the algorithm in [26]). The “recall” degrades for the low measured ratio, because some abnormal events, such as capacitor switching, are far away from the measured nodes that their signatures are not fully captured, but the proper placement of μ PMU improves the performance by 1%~15%. Table II reveals that the detection performance can be more than 95% when more than 24% nodes are measured.

Table II: Detection Performance with different μ PMU placement algorithms when system is partial observed with $\xi_1 = 2, \xi_2 = 350$

Measured Ratio	Precision	Recall	F1 score
24% (Selected)	100.0%	100.0%	100.0%
24% (Random)	100.0%	98.1%	99.0%
12% (Selected)	100.0%	98.0%	99.0%
12% (Random)	100.0%	87.7%	93.1%
6% (Selected)	100.0%	92.0%	95.8%
6% (Random)	100.0%	68.5%	80.3%

C. The Effectiveness of the Regularization

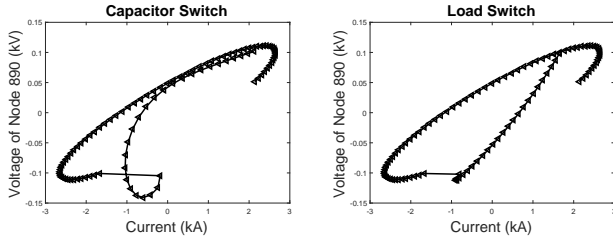


Fig. 5: The characteristics curves of voltages currents in one cycle at the node 890 when a capacitor bank switch or a load switch occurs respectively

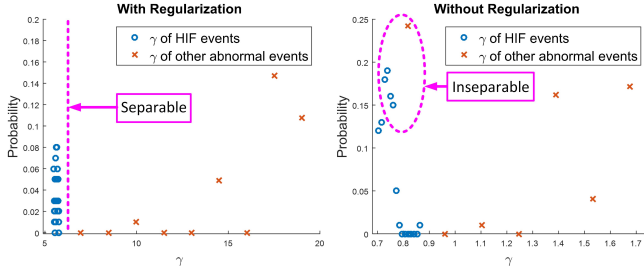


Fig. 6: Probability of γ for different noisy abnormal events when SNR is 50dB detected by PICAE with (left) and without (right) regularization item

Fig. 5 displays the trajectories of voltages and currents in one cycle after capacitor switching and load switching occur respectively. It is evident that these trajectories deviate from the original ellipse dramatically. On the contrary, Fig. 1 indicates that the trajectories for HIFs still follow certain ellipses. As a result, the reconstruction errors of PICAE for the capacitor switching and loads switching are significant compared with that of HIFs. Hence, the reconstruction errors themselves distinguish HIFs from other abnormal events due to the *elliptical regularization item*.

We discover this separability of PICAE becomes even more evident in noisy situations. According to the practical noise

level of PMUs [28], we corrupt the training and testing datasets by Gaussian noise of signal-noise-ratio (SNR) ranging from 30 dB to 90 dB and train the PICAEs with ($\lambda_r \neq 0$) and without ($\lambda_r = 0$) the regularization item. Fig. 6 statistically depicts the probabilistic distribution of γ 's, which generally reflect the variations of reconstruction errors, of various testing events in noisy situations. The γ 's of the HIFs become separable from those of the other abnormal events when using the PICAE with the elliptical regularization. On the contrary, the HIFs and non-HIFs are not separable if the PICAE is trained without the regularization.

D. Comparison with Existing Works

Table III: Detection F1 score of the PICAE for node 832 when SNRs are from 30dB to 90dB

SNR (dB)	PICAE	AE	PCA	ER
30dB	92.9%	81.5%	43.2%	39.5%
50dB	97.1%	81.3%	72.2%	62.9%
70dB	97.6%	83.0%	76.1%	64.7%
90dB	100.0%	83.3%	76.6%	64.7%

We compare the detection performance of the local PICAE with three existing unsupervised methods: auto-encoder (AE), principle component analysis (PCA), and Ellipse regression (ER) [21, 27, 29]. The structure of AE is similar to PICAE but without the physical regularization. We implement the PCA by the truncated singular value decomposition (SVD), and the number of principle components is selected by $r^* = \arg \min_r \frac{\sum_{n=1}^r \sigma_n}{\sum_{n=1}^T \sigma_n} \geq \tau$, where σ_n 's are the decreasing singular values of voltages V_i , and $\tau = 0.99$. ER represents the training data using the elliptical equation (5), through a linear regression method [29]. The performance of these three methods for normal and abnormal events are employed in the same way of Algorithm 2 to detect HIFs.

We summarize the F1 score of these four methods when SNR changes from 30 dB to 90 dB in Table III. PICAE is more robust to noise than others, achieving up to 17% higher F1 scores above all. The improvement profits from two attributes of PICAE. First, the convolutional autoencoder reconstructs normal events with high accuracy. Second, the physical regularization term enables a more considerable distinction between HIFs and other non-HIFs even in noisy situations. Note that when the SNR as low as 30 dB, we increase $\lambda_r = 440$ to improve the detection performance.

VI. CONCLUSIONS

HIF, potentially causing wildfires in the western U.S., is a significant concern in the industry. Existing data-driven algorithms can detect HIFs with high accuracy when a sufficient number of labeled datasets are provided. Rather than relying on the expensive labeled datasets, our PICAE exploits the unique voltage-current characteristic curves of HIFs as regularization in training. The regularization improves the capability of PICAE to separate HIFs from other events, even in highly noisy situations. Furthermore, a low-communication system-wide detection framework is proposed to improve detection

accuracy and reliability, especially when systems have low observability. PICA-E demonstrates superior performances in different noisy scenarios than existing works. An interesting avenue for future work is to unify the location and detection algorithms to enable follow-up control actions.

REFERENCES

- [1] S. Gautam and S. M. Brahma, "Detection of high impedance fault in power distribution systems using mathematical morphology," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1226–1234, 2012.
- [2] A. Mamishev, B. D. Russell, and C. L. Benner, "Analysis of high impedance faults using fractal techniques," *IEEE Trans. Power Syst.*, vol. 11, no. 1, pp. 435–440, 1996.
- [3] A. Ghaderi, H. A. Mohammadpour, H. L. Ginn, and Y.-J. Shin, "High-impedance fault detection in the distribution network using the time-frequency-based algorithm," *IEEE Trans. Power Del.*, vol. 30, no. 3, pp. 1260–1268, 2014.
- [4] M. Adamiak, C. Wester, M. Thakur, and C. Jensen, "High impedance fault detection on distribution feeders," *GE Industrial solutions*, 2006.
- [5] B. Wang, J. Geng, and X. Dong, "High-impedance fault detection based on nonlinear voltage–current characteristic profile identification," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3783–3791, 2016.
- [6] S. Chakraborty and S. Das, "Application of smart meters in high impedance fault detection on distribution systems," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3465–3473, 2018.
- [7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
- [8] Y. Pu, Z. Gan, R. Henao, X. Yuan, C. Li, A. Stevens, and L. Carin, "Variational autoencoder for deep learning of images, labels and captions," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 2352–2360.
- [9] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 2802–2810.
- [10] A. S. Zamzam and N. D. Sidiropoulos, "Physics-aware neural networks for distribution system state estimation," *IEEE Trans. Power Syst.*, 2020.
- [11] S. Park, D. Deka, and M. Chertkov, "Exact topology and parameter estimation in distribution grids with minimal observability," in *2018 Power Systems Computation Conference (PSCC)*. IEEE, 2018, pp. 1–6.
- [12] S. Talukdar, D. Deka, H. Doddi, D. Materassi, M. Chertkov, and M. V. Salapaka, "Physics informed topology learning in networks of linear dynamical systems," *Automatica*, vol. 112, p. 108705, 2020.
- [13] W. Li, M. Wang, and J. H. Chow, "Real-time event identification through low-dimensional subspace characterization of high-dimensional synchrophasor data," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 4937–4947, Jan. 2018.
- [14] W. Li, D. Deka, M. Chertkov, and M. Wang, "Real-time faulted line localization and pmu placement in power systems through convolutional neural networks," *IEEE Trans. Power Syst.*, vol. 34, no. 6, pp. 4640–4651, Nov. 2019.
- [15] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations," *arXiv preprint arXiv:1711.10561*, 2017.
- [16] N. Mwakabuta and A. Sekar, "Comparative study of the IEEE 34 node test feeder under practical simplifications," in *2007 39th North American Power Symposium*. IEEE, 2007, pp. 484–491.
- [17] *PSCAD/EMTDC manual version*, 4th ed., Manitoba H. V. D. C. Research centre Inc., 211 Commerce Drive Winnipeg, Manitoba Canada R3P 1A3, 2018.
- [18] S. Shiller, "High impedance fault arcing on sandy soil in 15kv distribution feeders: contributions to the evaluation of the low frequency spectrum," *IEEE Trans. Power Del.*, vol. 5, no. 2, 1990.
- [19] J. J. Theron, A. Pal, and A. Varghese, "Tutorial on high impedance fault detection," in *2018 71st Annual Conference for Protective Relay Engineers (CPRE)*. IEEE, 2018, pp. 1–23.
- [20] W. H. Kersting, "Radial distribution test feeders," *IEEE Trans. Power Syst.*, vol. 6, no. 3, pp. 975–985, 1991.
- [21] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [22] R. Halir and J. Flusser, "Numerically stable direct least squares fitting of ellipses," in *Proc. 6th International Conference in Central Europe on Computer Graphics and Visualization*. WSCG, vol. 98. Citeseer, 1998, pp. 125–132.
- [23] Y. Liu, Y. Wang, N. Zhang, D. Lu, and C. Kang, "A data-driven approach to linearize power flow equations considering measurement noise," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2576–2587, 2019.
- [24] M. J. Zaki and W. Meira, *Data mining and analysis: fundamental concepts and algorithms*. Cambridge University Press, 2014.
- [25] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, 2014, pp. 1–15.
- [26] W. Li and D. Deka, "Physics regulated neural network for high impedance faults detection," *arXiv preprint arXiv:2008.02364*, 2020.
- [27] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.
- [28] S. B. S. J. R. Michael Brown, Milan Biswal and H. Cao, "Characterizing and quantifying noise in PMU data," in *Proc. IEEE Power and Energy Society General Meeting*, 2016, pp. 1–5.
- [29] C. Robert, *Machine learning, a probabilistic perspective*. United Kingdom: Taylor & Francis, 2014.
- [30] D. Deka and S. Vishwanath, "PMU placement and error control using belief propagation," in *Proc. IEEE Int.*

Conf. Smart Grid Communications, 2011, pp. 552–557.

- [31] K. G. Nagananda, “Electrical structure-based PMU placement in electric power systems,” *arXiv preprint arXiv:1309.1300*, 2013.
- [32] M. Soltanolkotabi *et al.*, “Robust subspace clustering,” *Ann. Stat.*, vol. 42, no. 2, pp. 669–699, April 2014.

APPENDIX

In the realistic setting where μ PMUs and the corresponding PICAEs are sparsely placed in the distribution grid, the centralized HIF detector’s performance depends on the placement of the μ PMUs. We now discuss a μ PMU placement algorithm to maximize the detection performance using a limited number of K observed nodes.

A. μ PMU Placement Algorithm

The placement of μ PMU is crucial because the signatures of HIFs are local and only revealed by nearby μ PMUs. Conventional PMU or μ PMU placement algorithms determine PMU placement by solving a set cover problem [30],[31], that ensures that each bus is within one-hop of a PMU, or at least one terminal bus of a line has a PMU. In settings where the number of PMUs is too small to ensure complete observability, we present an alternate placement approach that maximizes the recorded PMU data diversity to improve detection.

The intuition comes from the empirical observation that grid segments/edges have distinctive voltages-curves at different parts of the network. By collecting measurements from nodes with different voltage dynamics, we are able to model the diversity of features. We measure the distinction of the voltages v_i and v_j by the subspace angle $\delta_{i,j}$ [32],

$$\delta_{i,j} = \begin{cases} \arccos\left(\frac{\cos(v_i, v_j)}{\|v_i\|_2 \|v_j\|_2}\right) & \text{if } (i, j) \in \mathcal{E} \\ 0 & \text{else} \end{cases} \quad (8)$$

where we only compare the dissimilarity of nodes i and j if $(i, j) \in \mathcal{E}$. With the measured distinction $\delta_{i,j}$, we determine a set \mathcal{S} of at most K non-adjacent μ PMU locations that maximizes the total dissimilarity $\sum_{i \in \mathcal{S}, j \in \mathcal{N}(i)} \delta_{i,j}$. Algorithm 3 provides a greedy approach to determine locations to maximize the total dissimilarity. The performance improvements due to our placement strategy is described with other numerical experiments in the next section.

Algorithm 3 μ PMU Placement

- 1: Input: $K, \delta_{i,j}, i, j = 1, \dots, m$
 - 2: $\mathcal{S} \leftarrow \emptyset, \Delta_i = \sum_{j \in \mathcal{N}(i)} \delta_{i,j}$.
 - 3: **while** $|\mathcal{S}| < K$ and Δ_i is not a all-zero vector **do**
 - 4: $\mathcal{S} \leftarrow \mathcal{S} \cup i^*, \Delta_j \leftarrow 0, \forall j \in \mathcal{N}(i^*)$, where $i^* = \arg \max_i \Delta_i$
 - 5: **end while**
 - 6: Output: \mathcal{S}
-

Table IV: Detection Performance of local PICAЕ at node 832 for Different Low Sampling Rate

f (kHz)	15.36	7.68	3.84	1.92	0.96
T	256	128	64	32	16
Precision (%)	100.0%	100.0%	100.0%	95.2%	94.3%
Recall (%)	100.0%	100.0%	100.0%	100.0%	100.0%
F1 Score (%)	100.0%	100.0%	100.0%	97.6%	97.1%

B. Robustness to Low Sampling rates

We downsample the datasets and demonstrate the robustness of PICAЕ to low sampling rates in Table IV, which is one of the concerns in the industry. When T , the number of samples per cycle, changes from 256 to 16, F1 score of the PICAЕ is higher than 90%, indicating the same PICAЕ tolerates lower sampling rates without obvious reduction of accuracy. Moreover, the structure of PICAЕ adapts to inputs of various sampling rate and does not require redesigning of the filters and bias matrices.