Unsupervised Anomaly Detection and Diagnosis in Power Electronic Networks: Informative Leverage and Multivariate Functional Clustering Approaches

Shushan Wu, Luyang Fang, Jinan Zhang, T.N. Sriram, Stephen J. Coshatt, Feraidoon Zahiri, Alan Mantooth, Jin Ye, Wenxuan Zhong, Ping Ma, WenZhan Song

***Abstract*—We propose a novel unsupervised anomaly detection and diagnosis algorithm in power electronic networks. Since most anomaly detection and diagnosis algorithms in the literature are based on supervised methods that can hardly be generalized to broader scenarios, we propose unsupervised algorithms. Our algorithm extracts the Time-Frequency Domain (TFD) features from the three-phase currents and three-phase voltages of the point of coupling (PCC) nodes to detect anomalies and distinguish anomaly types, cyber-attacks and physical faults. To detect anomalies through TFD features, we propose a novel Informative Leveraging for Anomaly Detection (ILAD) algorithm. The proposed unsupervised ILAD algorithm automatically extracts noise-reduced anomalous signals, achieving more accurate anomaly detection results than other score based methods.**

To assign anomaly types for anomaly diagnosis, we apply a novel Multivariate Functional Principal Component Analysis (MFPCA) clustering method. Unlike deep learning methods, the MFPCA clustering method does not require labels for training and provides more accurate results than other deep embedding-based clustering approaches. Furthermore, it is even comparable to supervised algorithms in both offline and online experiments. To the best of our knowledge, the proposed unsupervised framework accomplishing anomaly detection and anomaly diagnosis tasks is the first of its kind in power electronic networks.

***Index Terms*—Anomaly Detection, Anomaly Diagnosis, Leverage Score, Multivariate Principal Component Analysis based Clustering, Power Electronic Networks.**

I. Introduction

In smart grids, power electronics are the fundamental building blocks. The expansion of the Distributed Energy Resources (DERs), such as Photovoltaic (PV) farms and wind farms, has particularly become a major opportunity and challenge for smart grids. The interconnection of power electronics in cyber networks allows coordinated control for better energy efficiency and resilience in smart grids. On the other hand, the cyber-network connectivity among power electronics also exposes them to cyber threats. In addition, the physical faults due to the deterioration of the equipment, e.g., converter, also threaten the safety and security of smart grids.

A catastrophic failure of power electronic networks due to a malicious cyber-attack or an accidental physical fault would cause degradation of equipment and substantial economic loss. Early detection and diagnosis of the anomalies are essential for timely maintenance and recovery of power electronic networks [1], [2].

To the best of our knowledge, for anomaly detection, limited studies have been conducted using the information embedded in electrical signals for cyber-threat in cyber-physical systems [3]. Thus, exploring physical signals to advance cyberspace security and trust is essential. While there are a plethora of potential cyber-attacks, this study utilizes attacks that directly affect the operation of the electrical machine and its components [4], [5]. In addition to malicious cyber-attacks, electrical equipment may also suffer from accidental physical faults, of which open-circuit and short-circuit faults are common in electrical systems [1]. Thus, there is an urgent need to propose an anomaly detection method suitable for both cyber-attacks and physical faults.

Anomaly detection of cyber-attacks and physical faults alone is not enough to mitigate their impact on power electronic networks. Anomaly diagnosis is also a crucial follow-up step in identifying the root causes of the failure. That said, false identification of the root cause might lead to severe operational failure while performing mitigation strategies in the power electronic networks [6], [7]. Thus, an ideal system should be able to distinguish attacks from faults. To solve this critical problem, we are motivated to develop effective anomaly diagnosis methods to distinguish cyber-attacks from physical faults in the power electronic network.

In the anomaly detection literature, several supervised algorithms [8], [9], [10] have been developed to identify anomalies using deep learning methods. These methods use two different pathways to solve the anomaly detection problem. One pathway is to train the Autoencoder, Autoregressive Integrated Moving Average (ARIMA) model to reconstruct the distribution of normal data. If the reconstruction error of testing data is much larger than that of the normal data, the detector would immediately raise a flag, declaring the start of an anomaly. Such methods use labeled normal data to train and assume that the training and testing data have the same distribution. This assumption limits the extent of applications, especially

This research is partially supported by U.S. National Science Foundation under grants DMS-1903226, DMS-1925066, DMS-2124493, SaTC-2019311 and ECCS-1946057, the U.S. National Institute of Health under grant R01GM122080, the U.S. Department of Energy DE-EE0009026.

Shushan Wu, Luyang Fang, T.N. Sriram, Wenxuan Zhong, and Ping Ma are with Department of Statistics, University of Georgia. shushanwu@uga.edu, Luyang.Fang@uga.edu, tn@stat.uga.edu, wenxuan@uga.edu, pingma@uga.edu

Jinan Zhang, Stephen J. Coshatt, Jin Ye, WenZhan Song are with College of Engineering, University of Georgia. jinan.zhang@uga.edu, stephen.coshatt@uga.edu, jin.ye@uga.edu, wsong@uga.edu

Feraidoon Zahiri is with Robins Air Force Base. feraidoon.zahiri@us.af.mil

Alan Mantooth is with the Department of Electrical Engineering, University Arkansas, Fayetteville, AR 72701

when the load of the system increases, causing false-positive alarms. The second pathway approaches anomaly detection as a binary classification task, which uses binary labels (normal and abnormal) and waveforms to train a classification model [11], [12] by Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) network. However, such black-box deep learning models cannot be adapted to detect new types of anomalies.

Anomaly diagnosis, which distinguishes different anomaly types, is a classification or clustering task, depending on whether labels are used in training. Most available approaches focus on supervised learning methods [6], [11], [12] that use a support-vector-based algorithm or deep learning framework for multi-classification to distinguish different types of anomalies. Unsupervised methods are more potent in applications since they do not need label information during training and can discover new clusters if novel types of anomalies occur. However, to the best of our knowledge, the literature still lacks unsupervised approaches for anomaly diagnosis, especially those based on multi-dimensional features in power electronic networks.

Furthermore, most unsupervised anomaly detection and diagnosis methods are offline, which necessitates making the decisions based on all the data across time. In real scenarios, this poses certain difficulties in implementing methods for steaming data. Thus, there is an urgent need to develop an online framework for anomaly detection and diagnosis tasks.

To overcome the aforementioned challenges, we propose an unsupervised, data-driven Informative Leveraging for Anomaly Detection (ILAD) algorithm, combined with a Multivariate Functional Principal Component Analysis (MFPCA) clustering algorithm to distinguish between cyber-attacks and physical faults for anomaly diagnosis. For streaming six-dimensional waveform data, we sequentially process them window by window. Specifically, we first extract Time-Frequency Domain (TFD) features from the waveform data to combine the time and frequency domain information. We then model the multivariate time series as a Vector Autoregressive (VAR) model since the TFD features in a normal state are stationary, and compute leverage scores [2], [13] for each time window. Unfortunately, the leverage scores incorporate noise, making it hard to distinguish between the normal period and anomalous period of the TFD features. Therefore, we select significant singular values of the lag matrix associated with the VAR model by the permutation test, thereby obtaining informative leverage scores that more accurately identify anomalies. Instead of heuristically setting thresholds to detect anomalies [14], we use a data-driven change point detection algorithm to automatically and sequentially raise flags of the starts and ends of anomalies. Both offline and online experiments show that the ILAD algorithm achieves high accuracy.

After performing anomaly detection, we assign the identified anomalous windows the type of anomalies based on our MFPCA clustering algorithm. To this end, we project the multivariate time series onto a feature space spanned by eigenfunctions and approximate the densities of the TFD features by the product of the densities of the principal component scores. By projecting onto a lower-dimensional feature space, we are able to assign a label to each anomalous time window by maximizing the likelihood of the TFD features. Features characterizing two anomaly types on their respective principal components are extracted. While the pattern of different anomaly types is hard to detect in a multivariate context, projection to a lower-dimensional space not only enables us to distinguish between them but also do so more accurately.

To evaluate the performance of the proposed ILAD and MFPCA clustering algorithms, we use a PV farm as a study case and generate a variety of electric waveform data under both offline and online scenarios. The offline dataset has 43 cases, of which 25 are cyber-attacks, and 18 are physical faults. Both the starts and ends of the anomalies need to be identified for the subsequent processing step. The proposed offline ILAD (off-ILAD) successfully identifies the anomalies of 42 cases. The accuracy of the anomaly diagnosis task is about 0.94, which is a competitive number and comparable to the accuracy of the classification task. The real-time dataset is simulated using NI-device connected to the OPAL-RT. Of the two simulated scenarios, one ends with a short circuit fault and the other end with an open circuit fault. For the online anomaly detection, the proposed online ILAD (on-ILAD) algorithm achieves higher accuracy compared with other change point detection algorithms. For online testing of anomaly diagnosis, we assign the streaming time window to a closer cluster and obtain more accurate results compared with other deep embedding based clustering methods.

The novelty and contribution of our work are summarized as follows.

1) To the best of our knowledge, our algorithm is one of the first unsupervised data-driven anomaly detection and diagnosis algorithms utilizing Time-Frequency Domain (TFD) features in power electronic networks.
2) We propose a novel Informative Leveraging for Anomaly Detection (ILAD) algorithm to remove random noise in the original leverage score and amplify the changes due to anomalies.
3) Our algorithm utilizes a data-driven change point detection method to raise a flag if the informative leverage score increases drastically instead of heuristically using a threshold [14] to flag anomalies. Thus, the proposed ILAD algorithm is more robust to new anomalies.
4) We apply a novel Multivariate Functional Principal Component Analysis (MFPCA) clustering algorithm to the power electronic network, which projects the TFD features onto the lower-dimensional spaces spanned by eigenfunctions. Thus, the MFPCA clustering algorithm extracts features differentiating cyber-attacks and physical faults.
5) Our algorithm can be used in online scenarios to detect anomalies and diagnose the types of anomalies for TFD features in each time window.

II. Power Electronic Network and Attack Models

A. A General Power Electronic Network Model

As the number of DERs grows, a power electronics network for converting renewable energy sources into smart grids

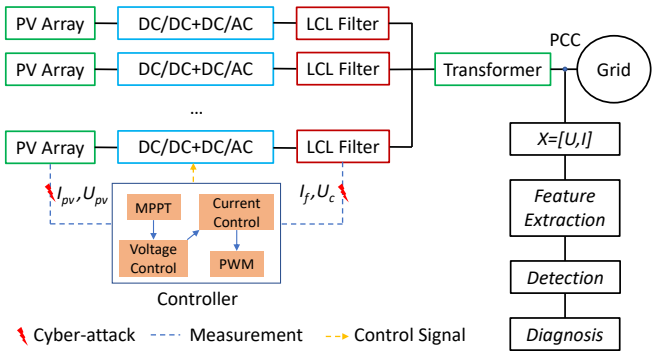

Fig. 1: Schematic diagram of the power electronic converter-enabled PV farm. I_{pv}, U_{pv}, I_f, and U_c are PV array current, PV array voltage, inductance current in the LCL, and capacitor voltage in the LCL, respectively.

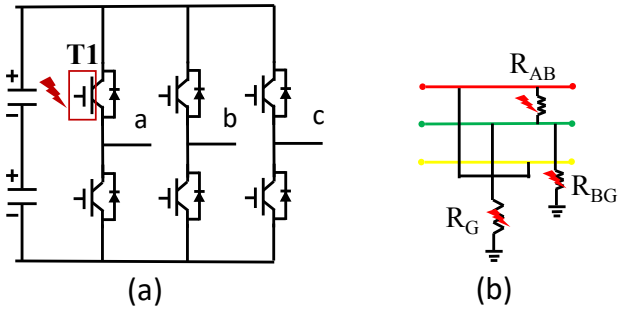

Fig. 2: (a) Open-circuit fault in the PV converter; (b) Short-circuit faults in the LCL.

is gradually taking shape. Figure 1 shows a typical power electronic network in a PV farm. To study the impact of cyber-attacks and physical faults, a high-fidelity PV farm is modeled. In the first stage, maximum power point tracking (MPPT) is designed to generate the maximum power of the PV array. In the second stage, voltage and current control are designed to maintain DC-link voltage and convert the power from the PV array to the power grid. Then, the LCL of each PV inverter is designed to filter out high-order harmonics in inductance current, which is expressed as follows:

$$\dot{x} = Ax + Bu, \tag{1}$$

where $x = [I_{fa}, I_{fb}, I_{fc}]^T$, and $I_{f\{\cdot\}}$ is one phase inverter-side inductance current in the LCL, $u = [U_{ka}, U_{kb}, U_{kc}, U_{ga}, U_{gb}, U_{gc}]^T$, $U_{k\{\cdot\}}$ and $U_{g\{\cdot\}}$ are one phase inverter-side voltage and grid-side voltage in the LCL, respectively,

$$A = \begin{bmatrix} \frac{-R}{L} & 0 & 0 \\ 0 & \frac{-R}{L} & 0 \\ 0 & 0 & \frac{-R}{L} \end{bmatrix} \quad B = \begin{bmatrix} \frac{1}{L} & 0 & 0 & \frac{-1}{L} & 0 & 0 \\ 0 & \frac{1}{L} & 0 & 0 & \frac{-1}{L} & 0 \\ 0 & 0 & \frac{1}{L} & 0 & 0 & \frac{-1}{L} \end{bmatrix} \tag{2}$$

where R and L are the resistance and inductance.

B. Cyber-attack Model

As discussed in many studies [15], [16], [17], cyber-attacks could destroy the operation of PV farms by compromising sensor measurement. In this paper, we assume that the attacker manipulates the measured data or injects false data into the sensor. The cyber-attack can be expressed as

$$Y_A(t) = \alpha Y_o(t) + \beta \tag{3}$$

where Y_A is the compromised data vector that is eventually the input of the controller, Y_o is the original measurement including I_{pv}, U_{pv}, I_f, and U_c, α is a multiplicative factor, and β is the false data injection.

C. Physical Fault

Besides cyber-attacks, physical faults also threaten PV farms. Figure 2 shows two types of physical faults that are modeled in PV farms, including open-circuit fault and short circuit fault. The open-circuit fault occurs in each switch of the PV converter, which leads to the open transistor. Figure 2(a) shows the open-circuit fault occurs in $T1$. Short circuit fault leads to a heavy current which further creates overheating or destroys the equipment. As shown in Figure 2(b), short-circuit faults are modeled, including three-phase short-circuit fault, single phase-to-ground fault, and phase-to-phase short circuit fault. The R_{AB}, R_{GB}, and R_G are the fault resistance.

III. Problem Statement

In a power network, our data consists of observations of the waveform in the PCC node in many cases. For case i at time t, let $X_i(t) = [I_{ia}(t), I_{ib}(t), I_{ic}(t), U_{ia}(t), U_{ib}(t), U_{ic}(t)]$ denotes a multivariate time series consisting of three-phase current $I = (I_a, I_b, I_c)$ and three-phase voltage $U = (U_a, U_b, U_c)$. By combining information both in the time domain and frequency domain, we utilize the TFD features proposed in [18]. We denote the nine-dimensional TFD features, a multivariate time series, by $\overline{\mathbf{X}}(t) = \left[\overline{X}^1(t), \ldots, \overline{X}^\ell(t), \ldots, \overline{X}^9(t)\right]$, where the i-th case is denoted as $\overline{\mathbf{X}}_i(t)$. Figure 3 shows an example of waveform data for three-phase voltages (bottom panel), three-phase currents (middle panel), and nine-dimensional TFD features (bottom panel) respectively, for one case. Based on this multivariate time series, we have two goals. The first is to find the starting point t_{k+1} and ending point t_{k+T} of an anomaly in the multivariate time series $\overline{\mathbf{X}}_i(t)$ and consider the anomalous portion of the series, $\left[\overline{\mathbf{X}}_i(t_{k+1}), ..., \overline{\mathbf{X}}_i(t_{k+T})\right]$. Given this anomalous series, the second goal is to assign an anomaly type (*cyber-attack* or *physical fault*) to each detected anomalous time period.

A. Anomaly Detection Problem

Based on the extracted TFD feature vector from one case, we aim to find a change point that shows a large change in the pattern of the data. We assume that there are n cases in total and the i-th case of the TFD feature $\overline{\mathbf{X}}_i(t)$ under normal conditions is generated by the model, $\overline{\mathbf{X}}_i(t) = \eta_i(t) + \epsilon_i(t)$, where $t = 1, .., t_k$, and $i = 1, ..., n$. If there is an abrupt change at time point t_{k+1}, then the TFD feature vector would be assumed to have the form: $\overline{\mathbf{X}}_i(t) = \alpha_i \eta_i(t) + \epsilon_i(t)$, for some real number α_i and $t = t_{k+1}, ..., t_{k+T}$ for some T, where α_i denotes the rate of change. That is, there would be a significant change in some dimensions of the TFD feature

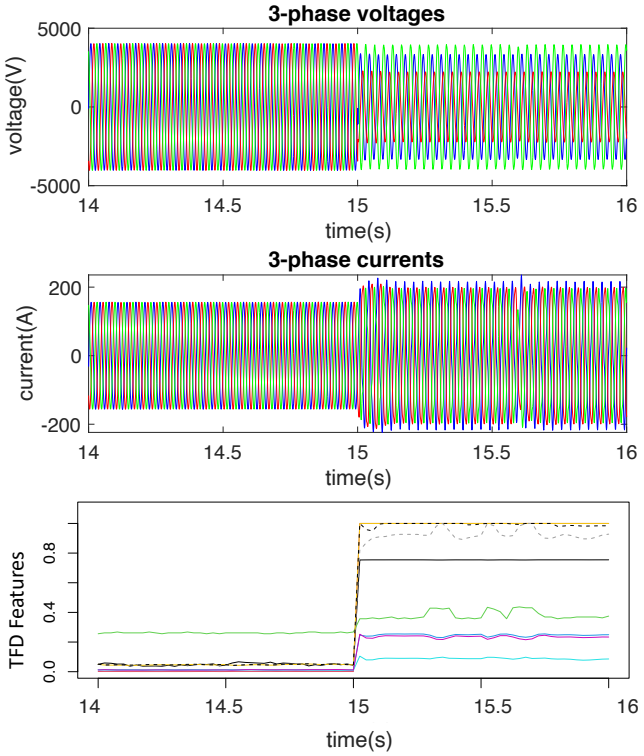

Fig. 3: An example of waveform data for three-phase voltages (bottom panel), three-phase currents (middle panel), and nine-dimensional TFD features (bottom panel), respectively, for one case. In this example, we show the data in a time range (14 16s). The anomaly happens at 15s.

vector when an anomaly happens. In statistics, leverage is a measure of how far away the value of the observation of TFD feature $\overline{\mathbf{X}}$ is from those of other observations. As shown in 3, the TFD features increase at 15 s, at which the anomaly happens. Thus, we formulate the problem as the identification of the time points with high leverage scores like the previous work [2], [13] did.

B. Anomaly Diagnosis Problem

We assume that there are two major anomaly types in the device-level power electronics converters (PEC), cyber-attack and physical faults. While these are two common types of anomalies, it is hard to distinguish between them. Wrong identification of the anomaly types might cause degradation of the devices and huge economic losses in the power electronic network. Thus, it is essential to identify the anomaly types of the anomalous time series after performing anomaly detection. To make sure our algorithm is still applicable to the online scenario, we slice the anomalous series $\left[\overline{\mathbf{X}}_i(t_{k+1}), ..., \overline{\mathbf{X}}_i(t_{k+T})\right]$ into pieces of anomalous windows. We are interested in predicting the cluster that each anomalous window belongs to with a label $z \in \{1, 2\}$, where 1 denotes cyber-attack, and 2 denotes physical fault. Note that this is an unsupervised problem where we do not have labels during the training phase, which is common in studies involving power electronic networks.

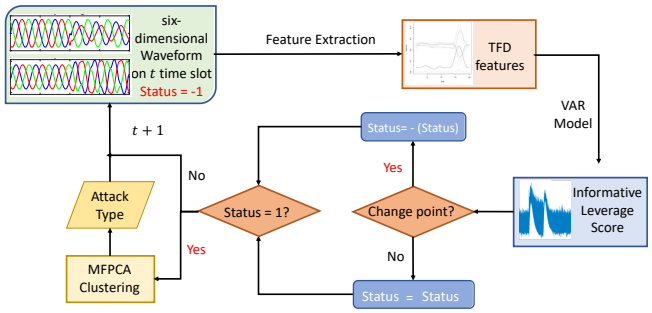

Fig. 4: The workflow chart of the online algorithm of anomaly detection and anomaly diagnosis.

IV. Algorithm Design

Our algorithm consists of three parts, as shown in Figure 4. First, through domain knowledge, we calculate the streaming TFD features for each time window. The extracted features contain information that not only helps distinguish normal data from anomalous data, but also enables us to distinguish between a cyber-attack and a physical fault. Second, we detect anomalies of the extracted features by the proposed ILAD algorithm. Since the informative leverage score of the extracted features will increase drastically if an anomaly starts or ends, we can easily detect the change points and raise flags when anomalies happen. The informative leverage score selects significant singular vectors for the leverage score calculation using a permutation test. The ILAD algorithm removes the noise and enlarges the difference between the anomalous period and normal period. The ILAD algorithm does not need labels in training and is effective in various emerging anomalies. Third, the anomaly diagnosis task would be triggered to assign labels (cyber-attack or physical fault) to the anomalous time windows after getting the anomalous data from the second step. This step also uses an unsupervised method, MFPCA, to cluster different anomaly types. Most classification methods need labels to train, while in power electronic networks, the true anomaly types are hard to obtain. Without needing the labels to train, our method extracts feature characterizing the difference between cyber-attacks and physical faults.

A. Feature extraction

Based on the raw waveform data, it is hard to distinguish the two anomaly types — cyber-attacks and physical faults. As shown in Figure 5, the plots of waveform data for two cases are on the left, one is under cyber-attack and the other has a physical fault. There is little difference between the two cases solely from the waveform data. This motivates us to use domain knowledge [18] to extract some higher-level time domain features and frequency domain features to help distinguish between the two anomaly types. We use the TFD features to identify the onset of anomalies and to distinguish between the two anomaly types via distinct patterns.

1) Frequency Domain Features: First, we obtain micro PMU (μPMU) features through fast Fourier transform (FFT) to convert a signal into individual spectral components and thereby extract frequency information about the signal. In

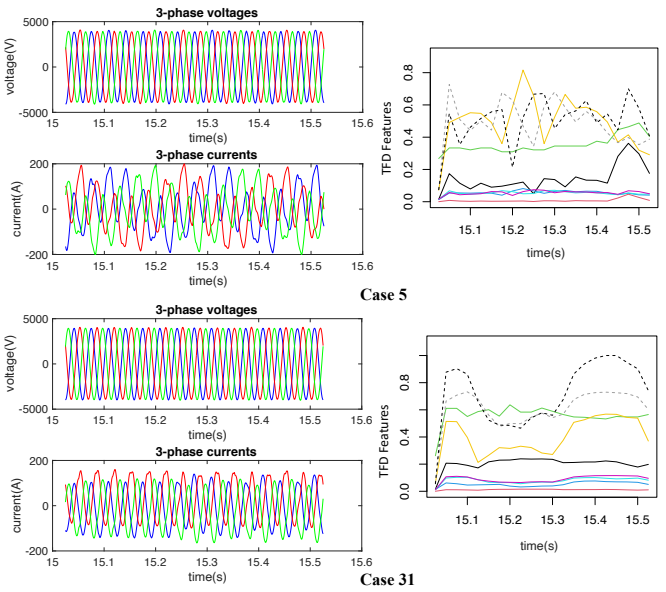

Fig. 5: An example of waveform data for two cases (5 and 31) and extracted TFD features. The first column shows the plot of waveform data, and the second column shows the plot of extracted TFD features. Sensor 5 encounters a cyber-attack while sensor 31 encounters a physical fault. In this example, we show the data in a time slot, from 15.025s to 15.525s.

addition, we use total harmonic distortion (THD) to capture the harmonic information of the distorted waveform. This yields a feature vector denoted by

$$\mathbf{F} = \left[M_{\{\cdot\}}, T_{\{\cdot\}}\right], \tag{4}$$

where $M_{\{\cdot\}}$, and $T_{\{\cdot\}}$ are six-dimensional vectors representing the magnitude (M) of the fundamental frequency and THD (T), respectively, for each phase of the waveform. The raw μPMU features sometimes lead to false positive results, especially when the magnitude is affected by a huge change in irradiance. Whereas the fundamental frequency, THD in a waveform are known to be lower than some boundary under normal conditions. Through expert knowledge, the maximum THD is set as $T_{max} = 5\%$. Then, THD for each phase is defined as:

$$\bar{T}_{\{\cdot\}} = \min\left\{\frac{T_{\{\cdot\}}}{T_{\max}}, 1\right\} \tag{5}$$

The six-dimensional total harmonic distortion feature is denoted by $\bar{T}$.

There are two typical physical faults, short circuit fault, and open circuit fault. Since the difference between the magnitudes of the three-phase waveforms R_m increases drastically when a short circuit fault happens, we could first extract features that help distinguish short circuit fault from other types of attack. R_m is defined as:

$$\begin{aligned} R_{m,I} &= \sqrt{\Delta M_{I_1}^2 + \Delta M_{I_2}^2 + \Delta M_{I_3}^2} \\ R_{m,U} &= \sqrt{\Delta M_{U_1}^2 + \Delta M_{U_2}^2 + \Delta M_{U_3}^2} \\ R_m &= (R_{m,I} + R_{m,U})/2 \end{aligned} \tag{6}$$

where $\Delta M_{I_1} = M_{I_a} - M_{I_b}$, $\Delta M_{I_2} = M_{I_b} - M_{I_c}$, $\Delta M_{I_3} = M_{I_a} - M_{I_c}$, and $R_{m,U}$ is defined similarly. After normalization

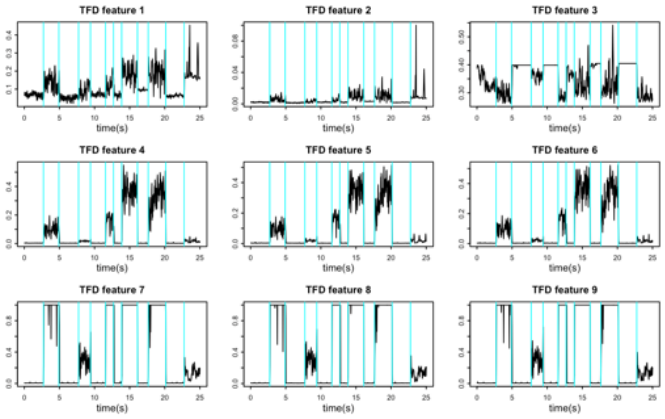

Fig. 6: An example of 9-dimensional TFD features for one case. Blue lines indicate the anomaly start/end time. There are 6 anomaly periods in total.

and scaling, the magnitude based features become:

$$\begin{aligned} \bar{R}_{m1} &= \min\left\{\frac{R_m}{R_{m1,\max}}, 1\right\} \\ \bar{R}_{m2} &= \min\left\{\frac{\ln\left(R_m + e\right) - 1}{R_{m2,\max}}, 1\right\}, \end{aligned} \tag{7}$$

where $R_{m1,max}$ is the maximum of R_m, and $R_{m2,max}$ is the maximum of $\ln(R_m + e) - 1$.

2) Time Domain Features: Except for the frequency domain feature, the transformation of the time domain features, three-phase currents, helps distinguish open-circuit fault from other attacks. We use a variant of the mean current vector (MCV) by current Concordia transformation:

$$\begin{aligned} I_\alpha &= \sqrt{\frac{2}{3}} I_a - \sqrt{\frac{1}{6}} I_b - \sqrt{\frac{1}{6}} I_c \\ I_\beta &= \sqrt{\frac{1}{2}} I_b - \sqrt{\frac{1}{2}} I_c. \end{aligned}$$

The cyber-attack and short circuit fault cases show a circle pattern centered in the origin in the plot of the vector (I_α, I_β) for all the time points. Whereas the open circuit fault's pattern is distorted due to the poor circuit contacts. Thus, to reflect the degree of distortion of points (I_α, I_β) at a time point t_k, we define the MCV point at time t_k as:

$$P_{mcv}(t_k) = \left(\frac{1}{N_k}\sum_{i=t_k-N_k+1}^{t_k} I_\alpha(i), \frac{1}{N_k}\sum_{i=t_k-N_k+1}^{t_k} I_\beta(i)\right) \tag{8}$$

Next, the time-domain feature $\bar{P}_{mcv}$ is defined based on the randomness of the distribution of P_{mcv} in a time window $[t_{k_1}, t_{k_2}]$. The more concentrated the distribution of P_{mcv} is, the more likely the anomaly type is the open circuit fault.

In all, we combine both the time and frequency domain features, and use the following set of features to do anomaly detection and anomaly diagnosis:

$$\overline{\mathbf{X}} = \left[\bar{R}_{m1}, \bar{R}_{m2}, \bar{P}_{mcv}, \bar{T}\right] \tag{9}$$

We refer to the above 9-dimensional feature as the TFD features. We use this feature to carry out anomaly detection and anomaly diagnosis.

B. Informative Leveraging for Anomaly detection

After extracting TFD features that could signal anomalous patterns of power electronic networks, we further model the 9-dimensional TFD features $\overline{\mathbf{X}}(t)$ by a VAR model, and determine the highly influential time points based on the leverage score of the VAR model. The original leverage score calculation method cannot eliminate the random noise, resulting in an insignificant difference between the normal and anomalous periods. This insignificant difference would result in false detection of the starts and ends of the anomalies. To overcome this issue, we propose an informative leverage score to remove the random noise from the small singular values.

After extracting the bump pattern through the informative leverage score, we use a sequential change point method [19] to identify the starts and ends of anomalies automatically. Our method can also be generalized to an online scenario to detect the starts and ends of the anomalies using the informative leverage scores.

1) Leverage Score for Offline Anomaly Detection: For each case, we assume that the 9-dimensional TFD feature at a time point depends on the past p observations of the series. Thus, we use the following p-th order VAR model representation to characterize the temporal dependence structure of the time series $\overline{\mathbf{X}}(t)$:

$$\overline{\mathbf{X}}(t_k) = \overline{\mathbf{X}}(t_{k-1})\mathbf{A}_1 + \overline{\mathbf{X}}(t_{k-2})\mathbf{A}_2 + \cdots + \overline{\mathbf{X}}(t_{k-p})\mathbf{A}_p + \boldsymbol{\epsilon}(t) \tag{10}$$

where $\{\mathbf{A}_i\}_{i=1}^p$ are 9×9 unknown parameters matrices and $\boldsymbol{\epsilon}(t)$ is the vector of error terms that are independently and identically distributed with mean zero and constant variance.

The VAR(p) model in (10) can also be expressed in the form of a linear model:

$$\overline{\mathbf{Y}} = \overline{\mathbf{D}}^p \boldsymbol{A} + \boldsymbol{\epsilon}, \tag{11}$$

where $\overline{\mathbf{Y}} = \left[\overline{\mathbf{X}}^T(t_k), \overline{\mathbf{X}}^T(t_{k+1}), ..., \overline{\mathbf{X}}^T(t_{k+T})\right]^T$, $\overline{\mathbf{D}}^p$ is the lag matrix of time series $\overline{\mathbf{X}}(t)$ defined as:

$$\begin{bmatrix} \overline{\mathbf{X}}(t_{k-1}) & \overline{\mathbf{X}}(t_{k-2}) & \cdots & \overline{\mathbf{X}}(t_{k-p}) \\ \overline{\mathbf{X}}(t_k) & \overline{\mathbf{X}}(t_{k-1}) & \cdots & \overline{\mathbf{X}}(t_{k-p+1}) \\ \vdots & \vdots & \ddots & \vdots \\ \overline{\mathbf{X}}(t_{k+T-1}) & \overline{\mathbf{X}}(t_{k+T-2}) & \cdots & \overline{\mathbf{X}}(t_{k+T-p}) \end{bmatrix}, \tag{12}$$

and $\boldsymbol{A} = [\mathbf{A}_1^T, \cdots, \mathbf{A}_p^T]^T$ is the parameter matrix to be estimated, and $\boldsymbol{\epsilon}$ is the random noise.

By the linear model representation in equation (11), the leverage score of the q-th data point can be interpreted as the amount of leverage or influence the q-th observed value exerts on the q-th fitted value [20]. The leverage score in the linear model has been generalized to the VAR model in time series [13]. In the VAR model, the time points with drastic fluctuation tend to have higher leverage scores, and we call them influential data points. In this way, we can convert the problem of detecting anomalies into the problem of identifying time points with high leverage scores.

The time points associated with the drastic fluctuation indicate the starts or ends of anomalies. Figure 6 shows an example of the TFD features for one case. We see anomalies happen when the TFD features drastically change, which motivates us to use the leverage scores to detect anomalies. For each case, the leverage score of the q-th observation can be expressed as

$$l_{qq} = \overline{\mathbf{d}}_{(q)}^{p\,T} (\overline{\mathbf{D}}^{p\,T} \overline{\mathbf{D}}^p)^{-1} \overline{\mathbf{d}}_{(q)}^p, \tag{13}$$

where $\overline{\mathbf{d}}_{(q)}^{p\,T}$ is the q-th row of $\overline{\mathbf{D}}^p$, and we call $\overline{\mathbf{D}}^{p\,T}\overline{\mathbf{D}}^p$ the lag-covariance matrix of the TFD features $\overline{\mathbf{X}}(t)$.

In real applications, some scenarios involve an online setting where we need to determine whether a streaming data window exhibits an anomaly based only on the information from the data collected before the current time point. For this, a generalization of the idea in [13] yields a streaming leverage score that only utilizes the history and the current information to approximate the leverage score.

2) Streaming Leverage score for Online Anomaly Detection: When the anomaly detection problem is extended to a real-time task, some additional difficulties arise. The main challenge is that one usually needs to make an immediate decision when a change has occurred as soon as a new data point streams in. However, the calculation of the lag covariance matrix needs the input of the whole time series. To overcome this, a natural and effective way is to use a pilot sample to approximate the true lag covariance matrix. Here, we use the method introduced by [13] to calculate the streaming leverage score, which guarantees the accuracy of the estimation while reducing the computational cost. We use the pilot sample of size r to approximate the lag-covariance matrix $\overline{\mathbf{D}}^{p\,T}\overline{\mathbf{D}}^p$. The streaming leverage score of the q-th observation, $\tilde{l}_{qq}$, is defined as:

$$\tilde{l}_{qq} = \overline{\mathbf{d}}_{(q)}^{p\,T} (\boldsymbol{\Gamma}_r^p)^{-1} \overline{\mathbf{d}}_{(q)}^p, \tag{14}$$

where $\boldsymbol{\Gamma}_r^p$ represents the approximation to the lag-covariance matrix based on the pilot sample with size r, and we call it the sketched lag-covariance matrix.

We show a simplified version of the streaming leverage score. We denote the singular value decomposition (SVD) of the sketched lag-covariance matrix $\boldsymbol{\Gamma}_r^p$ by $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$, where $\boldsymbol{\Sigma}$ is the diagonal matrix of singular values, $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices such that $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}$. Let

$$\tilde{l}_{qq} = \sum_{j=1}^{r-p} \left(\overline{\mathbf{d}}_{(q)}^{p\,T} \boldsymbol{v}^{(j)}\right)^2 / \boldsymbol{\sigma}_j^2, \tag{15}$$

where $\boldsymbol{v}^{(j)}$ is the j-th column of $\mathbf{V}$, $\boldsymbol{\sigma}_j$ is the j-th singular value, and $r - p$ is the total number of singular values of the sketched lag-covariance matrix $\boldsymbol{\Gamma}_r^p$. The singular values of the lag-covariance matrix are also referred to as spectrum in this article.

Here, the information of the lag-covariance matrix is projected onto orthogonal directions of singular vectors $\boldsymbol{v}^{(j)}$, and each singular value is the variance of the data in the corresponding singular vector space. The Principal Component Analysis (PCA) method selects the most significant singular vectors representing the whole data distribution and removes the ones with smaller singular values. In our case, each pair of nearly equal eigenvalues and associated PCs of the lag-covariance matrix characterizes an oscillatory mode, e.g.,

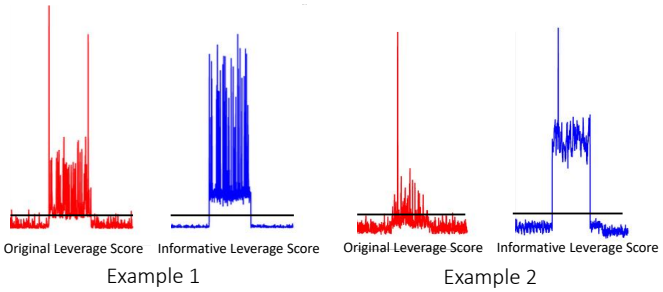

Fig. 7: An example of the original leverage score and informative leverage score. By removing the noise and obtaining an informative leverage score, the gap between the normal and the anomalous rises. This leads to higher accuracy while detecting the starts and ends of anomalies.

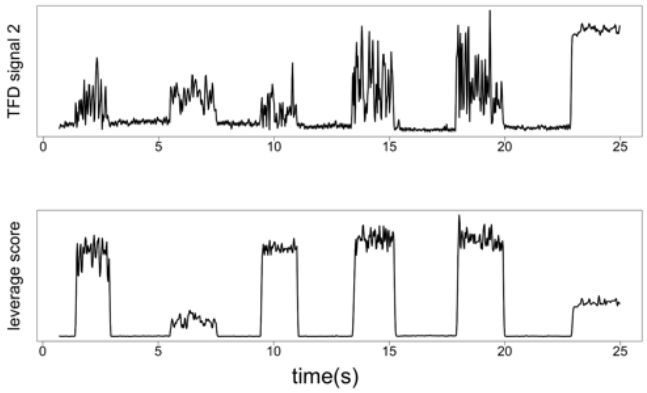

Fig. 8: An example of one dimension of the TFD features (top panel) and the informative leverage score of the total TFD features (bottom panel).

trend, periodicity, and noise. However, not every PC can help distinguish between normal and anomalous data. For example, the first PC characterizing the trend is not informative to anomaly detection, and anomalies often appear in other oscillatory modes.

3) Informative Leverage Scores for Anomaly Detection: The aforementioned challenges motivate us to propose an informative leveraging for anomaly detection (ILAD) algorithm to select more informative PCs to differentiate between the normal and anomalous periods. Instead of directly using the original leverage scores, we perform a test to see if each singular vector is informative by examining the amount of noise it contains. If a singular vector contains excessive random noise, we exclude it while calculating the leverage score. Mimicking the idea of a permutation test, we randomize different rows of the lag-covariance matrix $\overline{\mathbf{D}}_i^{p\,T}\overline{\mathbf{D}}_i^p$ for each feature in the offline setting and the sketched lag-covariance matrix $\mathbf{\Gamma}_r^p$ in the online setting, and perform an SVD again. The result of the SVD in online and offline settings is usually denoted by $\tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^T$. We repeat this procedure many times, and each time compare the actual values to the randomized ones. If the true singular value is outside the 95% confidence interval, then we declare that the singular value and the associated singular vector are informative. Through the permutation test, we get a set $\mathcal{I}$ of informative singular vectors. Then, we let

$$\tilde{l}_{qq}^k = \sum_{j=\in\mathcal{I}} \left(\overline{\mathbf{d}}_{(q)}^{p\,T}\tilde{\boldsymbol{v}}^{(j)}\right)^2 / \tilde{\boldsymbol{\sigma}}_j^2, \tag{16}$$

where $\tilde{\boldsymbol{v}}^{(j)}$ is the j-th column of $\tilde{\mathbf{V}}$, $\tilde{\boldsymbol{\sigma}}_j$ is the j-th entry of $\tilde{\mathbf{\Sigma}}$, and k is the cardinality of the set of the informative singular vectors $\mathcal{I}$. We illustrate the advantages of filtering informative singular vectors via a comparison of original leverage scores and the proposed informative leverage scores for two cases in Figure 7; these are calculated in an offline manner. We can clearly see that the patterns of the two leverage scores are entirely different. The red lines shown in Figure 7 are the original leverage scores, and the blue lines shown in Figure 7 are the informative leverage scores. Comparing the two sets of leverage scores, we see that the informative leverage scores seem to remove the noise, especially before the anomaly starts and after the anomaly ends , making the refined algorithm perform better for anomaly detection. In addition, switching from original leverage scores to informative leverage scores, the gap between the score of the normal to that of the anomalous rises significantly. Thus, the performance of the anomaly detection improves by removing the information in the direction of the least important singular vectors.

To illustrate that the proposed ILAD algorithm still works in online settings, we show that informative leverage scores reflect drastic changes caused by the starts and ends of anomalies in an online manner. Figure 8 shows an example of streaming data with five cyber-attacks and one physical fault. The top panel shows one dimension of the TFD features, and the bottom panel shows the informative leverage scores calculated by the total TFD features. We can see that the time points with high leverage scores and those signaling presence of anomalies always coincide, which confirms our belief that influential points with high leverage scores are where anomalies happen. Due to the drastic change in the informative leverage scores as soon as there is an anomaly, we subsequently use a sequential change point detection algorithm [19] to identify the starts and ends of anomalies. Most available anomaly detection methods use a pre-specified threshold to raise a flag. The threshold based methods are ad-hoc and need a fine-tuning step to set an appropriate value. Instead, the sequential change point method is a data-driven approach, making decisions based on past information. Thus, the anomaly detector prevents information leakage from future observations, and identifies anomalies adaptive to the data.

C. Multivariate Functional Principal Component Analysis Clustering for Anomaly Diagnosis

Most approaches for anomaly diagnosis [11], [21] use a supervised classification model, where information from labels is used for prediction. However, for anomalies in power electronic networks, the labels for the anomaly types are hard to obtain. Thus, accurate unsupervised methods are urgently needed for anomaly diagnosis in power electronic networks. Currently, existing unsupervised anomaly diagnosis methods distinguish between anomaly types using distance based methods, such as K-means and hierarchical clustering [22]. These methods ignore the dependency between different data features and are sensitive to outliers. In addition, distance based and dissimilarity based methods do not assume models,

and therefore, we cannot find the probability that a new data point belongs to a certain cluster. In order to model the dependence and assign a probability of cluster membership to each data point, we use the MFPCA to approximate the data distribution and maximize the likelihood of the mixture model. MFPCA can embed the multivariate time series into a low-dimensional space spanned by eigenfunctions based on Karhunen-Loeve expansion [23]. Through such projection, the density of the multivariate time series can be approximated by the product of the densities of the principal component scores.

Assume that the data is generated from multiple clusters, then the multivariate time series follows a mixture model, whose likelihood can be maximized by the iterative Expectation–maximization (EM) algorithm [24], [25]. To find the optimal representation of the time series in a functional space, we further assume that $\overline{\mathbf{X}}(t)$ is an L_2-continuous stochastic process, that is,

$$\begin{aligned} &\forall t \in [t_1, t_2], \quad \lim_{h\to 0} \mathbb{E}\left[\|\overline{\mathbf{X}}(t+h) - \overline{\mathbf{X}}(t)\|^2\right] \\ &= \lim_{h\to 0} \int_{t_1}^{t_2} \sum_{\ell=1}^{9} \mathbb{E}\left[\left(\overline{X}^{\ell}(t+h) - \overline{X}^{\ell}(t)\right)^2\right] dt = 0 \end{aligned} \tag{17}$$

Note that most real data satisfy this assumption, and so does the TFD feature, which is normalized in $[0, 1]$. We also denote the mean of the ℓ-th variate as $\mu^{\ell} = \left\{\mu^{\ell}(t) = \mathbb{E}\left[\overline{X}^{\ell}(t)\right]\right\}_{t\in[0,T]}$, and let $\mu(t) = \mathbb{E}[\overline{\mathbf{X}}(t)] = \left(\mu^1, \ldots \mu^{\ell}, \ldots, \mu^9\right)^T$. We further define the covariance function of $\overline{\mathbf{X}}(t)$ as:

$$V(s,t) = \mathbb{E}[(\overline{\mathbf{X}}(s) - \mu(s)) \otimes (\overline{\mathbf{X}}(t) - \mu(t))], \tag{18}$$

where $s, t \in [t_1, t_2]$, and $\otimes$ is the tensor product on $\mathbb{R}^p$. Then, the eigenfunctions $\left\{\boldsymbol{f}_m = \left(f_m^1, \ldots, f_m^{\ell}, \ldots, f_m^9\right)^T\right\}_{m\geqslant 1}$ are defined as:

$$\int_{t_1}^{t_2} V(\cdot, t)\boldsymbol{f_m}(t)dt = \lambda_m \boldsymbol{f}_m, \tag{19}$$

which satisfy $\int_{t_1}^{t_2} \sum_{\ell=1}^{9} f_m^{\ell}(t)' f_{m'}^{\ell}(t)dt = 1$ if $m = m'$ and 0 otherwise, and $\{\lambda_m\}_{m\geqslant 1}$ are associated eignvalues. Consequently, the principal component $\{\mathrm{C}_m\}_{m\geqslant 1}$ are the projections of $\mathbf{F}$ on the space spanned by the eigenfunctions $\{\boldsymbol{f}_m\}_{m\geqslant 1}$ of the covariance function:

$$\mathrm{C}_m = \int_{t_1}^{t_2} \sum_{\ell=1}^{9} \left(\overline{X}^{\ell}(t) - \mu^{\ell}(t)\right) f_m^{\ell}(t)dt, \tag{20}$$

where the principal components $\{\mathrm{C}_m\}_{m\geqslant 1}$ are zero-mean uncorrelated random variables with variance $\{\lambda_m\}_{m\geqslant 1}$, respectively. After removing the mean effect of $\overline{\mathbf{X}}(t)$, we truncate the first q' terms of the Karhunen-Loeve expansion of $\overline{\mathbf{X}}(t)$ and write it as:

$$\overline{\mathbf{X}}(t) = \sum_{m=1}^{q'} \mathrm{C}_m \boldsymbol{f}_m(t), \quad t \in [t_1, t_2]. \tag{21}$$

The truncation leads to a dimension reduced subspace. We further assume the density of each principal component C_m is univariate Gaussian distribution. Since the structure of the distribution of the multivariate time series can be retained in the spectrum of the covariance of the data, one natural density surrogate of TFD feature $\overline{\mathbf{X}}(t)$ is the density of the first q' principal components:

$$f_{\overline{\mathbf{X}}(t)}^{(q')}(\overline{x}) = \prod_{j=1}^{q'} f_{\mathrm{C}_m}\left(\mathrm{c}_m(\overline{x}); \lambda_m\right), \tag{22}$$

where $\mathrm{c}_m(\overline{x})$ is the principal component score of data $\overline{x}$, and f_{C_m} is the density of the m-th principle component C_m. Assume the data generation procedure follows a mixture model, the probability of generating data from g-th cluster π_g satisfies $\sum_{g=1}^{K} \pi_g = 1$. We denote the indicator of the cluster g as Z^g, which takes the value 1 when the data belongs to g-th cluster and 0 otherwise. Then, we approximate the density of $\overline{\mathbf{X}}_{|Z^g=1}(t)$ by product of the densities of random variables $\left\{\mathrm{C}_{m|Z^g=1}\right\}_{m=1,\ldots,q'}$ with zero mean and variance $\{\lambda_{m,g}\}_{m=1,\ldots,q'}$. Thus, the density of $\overline{\mathbf{X}}(t)$ can be represented by:

$$f_{\overline{\mathbf{X}}(t)}^{(q')}(\overline{x}; \theta) = \sum_{g=1}^{K} \pi_g \prod_{j=1}^{q'_g} f_{\mathrm{C}_{m|Z^g=1}}\left(\mathrm{c}_{m,g}(\overline{x}); \lambda_{m,g}\right), \tag{23}$$

where $\mathrm{c}_{m,g}(\overline{x})$ is the pricipal component score of $\overline{x}$ belonging to g-th cluster, and q'_g is the number of principal components for g-th cluster, and $\theta = \left\{\left(\pi_g, \lambda_{1,g}, \ldots, \lambda_{q'_g,g}\right)_{1\leqslant g\leqslant K}\right\}$ are unknown parameters to be estimated. We can represent the likelihood of the observed data $\overline{x} = \{\overline{x}_i\}$ by:

$$l^{(q')}(\theta; \overline{x}) = \prod_{i=1}^{n} \sum_{g=1}^{K} \pi_g \prod_{m=1}^{q'_g} \frac{1}{\sqrt{2\pi\lambda_{m,g}}} \exp\left(-\frac{1}{2} \frac{\mathrm{c}_{m,g}^2(\overline{x}_i)}{\lambda_{m,g}}\right), \tag{24}$$

where $c_{m,g}(\overline{x}_i)$ is the m-th principal component score of i-th observation $\overline{x}_i$ belonging to the g-th group. We use the iterative EM algorithm to maximize the above likelihood function with respect to θ. By finding the optimal representation of the data $\overline{x}$, we can estimate the most probable clustering assignment for each observation $\overline{x}_i$.

To make this algorithm applicable to anomaly diagnosis in power electronic networks, we use the sliding window approach to slice the long time series into small fragments since the duration of the attack may be short and it could shift between different anomaly types. Thus, we assign clustering labels to each sliding window. In our context, there are only two anomaly types to be distinguished. Thus, we set the number of clusters as two. Another implementation issue of the MFPCA clustering algorithm is how to decide the number of principal components for approximating the likelihood function. We use the Cattell scree test [26] to select q'_g for each g-th group.

V. Offline Testing Results

A. Experiment setup

The model and data used in this study are based on a *testbed* model co-developed by the Intelligent Power Electronics Electric Machine Lab and the Sensorweb Research Lab at the University of Georgia (UGA) for generating electric waveform

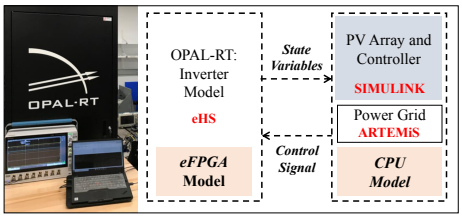

Fig. 9: Real-time testbed.

data. In this study, we refer to the data from this *testbed* as the UGA dataset. The PV farm consisting of seven converters and an IEEE 37-node distribution grid is simulated in OPAL-RT as shown in Figure 9. To simulate the dynamics of the PV farms, PV converters are modeled in Embedded Field Programmable Gate Array (eFPGA). The IEEE 37-node distribution grid is simulated in Advanced Real-Time Electro-Magnetic Solvers (ARTMEiS) to realize the real-time simulation. In the real-time testbed, a number of cases are simulated. The offline dataset consists of 43 abnormal cases. As we show in Table I, among all 43 anomalous cases, there are 25 cyber-attack cases, of which 14 are single-DIA cyber-attack cases, 10 are coordinated-DIA cyber-attack cases, and 1 is a replay attack, and 18 are physical fault cases, of which 14 are short circuit fault cases and 4 are open circuit fault cases. The data is the six-dimensional raw waveform data composed of three-phase currents and three-phase voltages. Each case has a total of 800,000 time points with a sampling frequency of 20,000 Hz. As a pre-processing step, we first down-sample the raw time series every ten points to prevent the high computational cost. Then, we extract TFD features from the raw waveform data. For the down-sampled six-dimensional waveform of length 1000, we could extract nine-dimensional TFD features of length 20. After feature extraction, we get a multivariate time series with dimension $(1600, 9)$.

TABLE I: Number of Each Type of Anomaly in Dataset

Anomaly SubType	Number of Cases	Cyber or Physical?
Single DIA	14	Cyber-Attack
Coordinated DIA	10	Cyber-Attack
Replay attack	1	Cyber-Attack
Short-circuit fault	14	Physical Fault
Open-circuit fault	4	Physical Fault

B. Offline test results

1) Offline Anomaly Detection: For offline anomaly detection, our task is to identify the starts and ends of the anomalies. The input for our algorithm is the 9-dimensional TFD features with 1600 time points. The true anomalies start at 15 seconds and end at 25 seconds. If the delay of the detector's responses to the true starts or ends is no later than 5 seconds, we say the detection is successful.

Before implementing the ILAD algorithm, we first fit the VAR(p) model to the TFD features, then we calculate the informative leverage scores for all time points and estimate the breakpoints of the scores, which are our estimated starts and ends of anomalies. It should be noted that the choice of the hyper-parameter p in the VAR(p) model is data-driven. Since the initial part of the streaming data is mostly normal, we take this part as the pilot sample to determine the order p of the time-dependence structure. Specifically, we aim to find the VAR(p) model which best represents the underlying dependence structure of the normal patterns of the TFD features. Considering both the prediction loss and the model complexity, we choose p with the smallest BIC value in the range of $p \in [1, 15]$. We also build the model under different pilot sample sizes (from 35 to 65) to test if our model is sensitive to the pilot sample size. We find that the optimal choice of the order p remains the same. Thus, we set the pilot sample size as 50.

TABLE II: Experiment Results of Offline Anomaly Detection

Approach	Start	End
off-ILAD	**42/43**	**32/43**
Leverage	40/43	21/43
Hotelling T^2	33/43	3/43
MCUSUM	17/43	17/43

To show the benefits of the proposed informative leverage score, we compared it with the original leverage score in terms of the accuracy of identifying the starts and ends of the anomalies. We also compared two unsupervised score based algorithms, Hotelling T^2 [27], [28] and Multivariate CUSUM [29], [30], for detecting the starts and ends of anomalies. We deployed these two methods since they are designed to deal with the multivariate time series data. The same sequential change point detection algorithm is applied in the proposed ILAD algorithm to ensure fairness. Results are shown in Table II. The performance of the proposed algorithm denoted by "off-ILAD" is better than that of the original "Leverage" approach and is superior to the other score based methods. Note that "off-ILAD" identifies 42 starts and 32 ends of anomalies out of the 43 cases. The reason why the accuracy of "off-ILAD" in detecting the ends of the anomalies is lower than detecting the starts is that, even though some physical faults are withdrawn, the system cannot return to its normal state. This is why detecting of the ends of anomalies fails in some cases.

2) Offline Anomaly Diagnosis: Among all the anomalies, two major anomaly types are to be categorized. Since the repair involved after attacks of different types of anomalies is significantly different, it is necessary to distinguish cyber-attack from physical fault accurately.

The extracted TFD features for each case are long and periodic, therefore, we slice the long time series into several time slots (each slot has 20 time points). Thus, we have 80 time slots from one case. Furthermore, we filter the data in the anomalous duration detected by our proposed ILAD algorithm. Thus, we obtain 893 windows in total. We apply the MFPCA clustering to diagnose the 893 observations of multivariate time series. Our method embeds the data onto a low-dimensional space spanned by eigenfunctions. Thus, we

TABLE III: Experimental Results of Offline Anomaly Diagnosis

Approach	Accuracy	F1	TPR	TNR
MFPCA	0.9440	0.9490	1.0000	0.9029
t-SNE	0.6002	0.6557	0.5650	0.4630
UMAP	0.6663	0.6469	0.5709	0.6190

compare the benchmark deep embedding methods t-SNE, and UMAP to embed the data onto a two-dimensional space and apply the K-means clustering algorithm. The results are shown in Table III. We measure the performance of clustering through Accuracy, F1 score, TPR (True Positive Rate), and TNR (True Negative Rate). In terms of all four measures, the proposed MFPCA algorithm is the best among the three methods considered here. The Accuracy measure of the MFPCA algorithm is 94.40% and the F1 score is 94.90%, which are relatively higher numbers and even comparable to some of the classification algorithms [18]. Our MFPCA clustering algorithm successfully identifies all the cyber-attacks. However, some physical faults are wrongly identified as cyber-attacks because some are hard to distinguish from cyber-attacks.

VI. Online Testing Results

A. Online Experiment Setup

To validate the proposed method, we develop a real-time detection and diagnosis testbed using NI device. As shown in Figure 10, the NI 9205 is connected to the OPAL-RT. The real-time data obtained by NI 9205 is sent to the PC through Ethernet. To perform a comprehensive real-time data analysis,

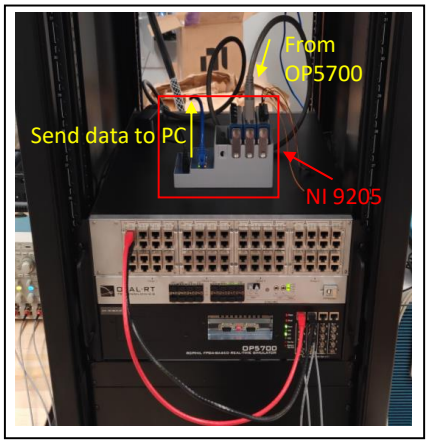

Fig. 10: Real-time testbed using OPAL-RT and NI device.

we also distinguish between the two types of physical faults by considering the short circuit fault and the open circuit fault. We obtained streaming data consisting of different anomaly types under two scenarios: (1) Scenario one consists of a set of streaming data with five cyber-attacks, and one physical attack due to a short circuit fault; (2) Scenario two consists of another set of streaming data with five cyber-attacks, and one physical attack due to an open circuit fault. An illustration of the real-time experiment setup of the two scenarios is in Figure 11. The y-axis represents the state of the streaming data, whether it is normal, under cyber-attack (anomaly type 1), or under physical fault (anomaly type 2). There are 6 starts and 5 ends of anomalies to be detected.

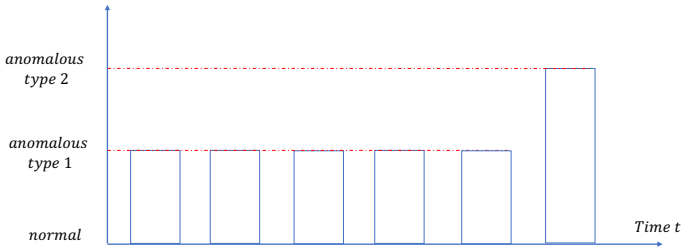

Fig. 11: The figure shows the set-up of the real-time data. The first five anomalies are cyber-attacks, and the last is the physical fault.

B. Online test results

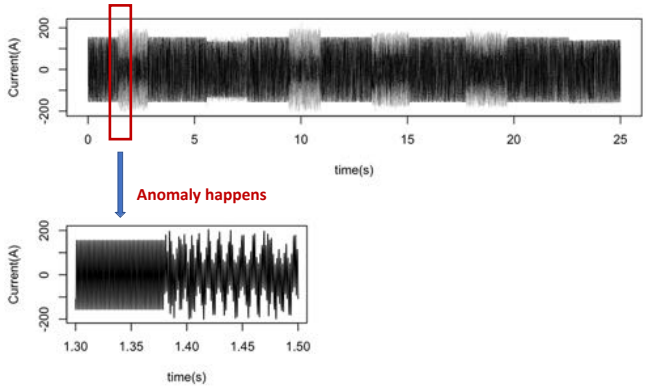

Fig. 12: Top panel: One phase of the current for the second scenario; Bottom panel: Zoom-in of the anomalous duration under cyber-attack (The anomaly happens between 1.35s and 1.40s).

1) Online Anomaly Detection: The proposed online-ILAD algorithm is implemented on the above online datasets to test its performance. Under each scenario, we continuously collect waveform data and detect the anomaly as the new data streams. The raw streaming waveform data contain around 500,000 time points($\approx 25s$). Figure 12 shows one phase of the current under the second scenario. Our goal is to detect the starts and ends of all attacks. We first down-sample the long time series every 10 time points to prevent high computational cost, and then extract the nine-dimensional TFD features. Our following analysis is based on the TFD features. We use a similar procedure in the offline setting to choose the best VAR(p) model and apply the online-ILAD algorithm to the streaming data. As in the offline experiment, the pilot sample size for selecting the best VAR model is 50. Varying different order values p, we choose the best hyper-parameter for the VAR(p) model with the smallest BIC value. According to Figure 13, the VAR(3) model is chosen for scenario one, and VAR(5) is for scenario two. Figure 14 shows the calculated online informative leverage scores for both scenarios. The top panel is the result of scenario one and the bottom panel is the result of scenario two. The blue vertical lines are where the anomalies happen. The red vertical lines are the detected starts and ends of anomalies. We can see that the time points with high leverage scores are consistent with

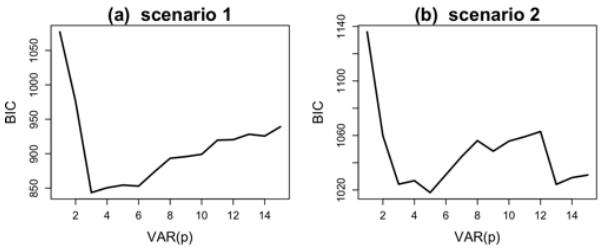

Fig. 13: (a): For scenario 1, BIC score under different p; (b):For scenario 2, BIC score under different p

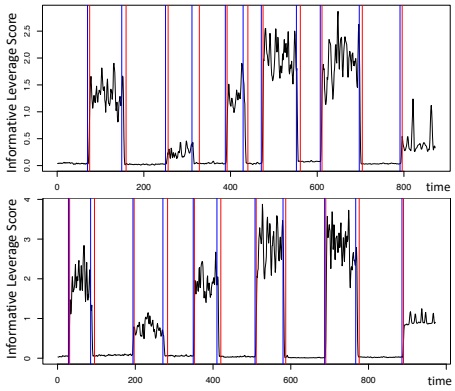

Fig. 14: The results of online Informative Leveraging for anomaly detection. The solid black line is the informative leverage score. The blue vertical lines are where the anomalies happen. The red vertical lines are the detected starts and ends of anomalies. The upper one shows the results of scenario one, and the lower one shows the results of scenario two.

the anomalies on waveform data. We then used the change point detection algorithm to sequentially detect change points of the informative leverage score. Table IV and Table V show the results of online anomaly detection for scenario one and scenario two, respectively. Our proposed online ILAD algorithm is denoted by "on-ILAD". We compare the proposed methods with other score based anomaly detection methods and identify the anomalies by the same sequential change point algorithm. The performance of the anomaly detection task in the two scenarios is good, with the 100% accuracy. Thus, our method is superior in performance to other competing methods. It should be noted that the anomalous data returns to the normal state after the attack ends. Thus, our proposed method successfully detects all the ends of anomalies and validates the efficiency of the proposed algorithm.

TABLE IV: Comparison of prediction results for Scenario 1

Approach	Start	End
on-ILAD	6/6	5/5
Leverage	5/6	2/5
Hotelling T^2	5/6	4/5
MCUSUM	3/6	2/5

TABLE V: Comparison of prediction results for Scenario 2

Approach	Start	End
on-ILAD	6/6	5/5
Leverage	4/6	3/5
Hotelling T^2	2/6	1/5
MCUSUM	5/6	4/5

TABLE VI: Experimental Results of Real-time Anomaly Diagnosis for Scenario 1

Approach	Accuracy	F1	TPR	TNR
MFPCA	0.8571	0.9032	1.0000	0.8235
t-SNE	0.7143	0.2500	0.2500	0.2500
UMAP	0.8095	0.8947	1.0000	0.0000

TABLE VII: Experimental Results of Real-time Anomaly Diagnosis for Scenario 2

Approach	Accuracy	F1	TPR	TNR
MFPCA	0.9524	0.9697	0.9412	1.0000
t-SNE	0.8095	0.8667	0.7647	1.0000
UMAP	0.9047	0.9444	1.0000	0.5000

2) Online Anomaly Diagnosis: As in the offline experiment, we slice the TFD feature in the anomalous period into small time slots and predict the TFD feature label in each time slot based on the mixture model we trained in the offline experiment. For each incoming time slot, we estimate its principal components in each cluster, and compare the likelihood of the window belonging to each cluster. Finally, we assign the clustering label to the one with a higher likelihood. The online testing result of the MFPCA clustering algorithm is shown in Table VI and Table VII. In the online testing, the performance of our clustering algorithm is still comparable to the classification method mentioned in [18], and our method is superior in performance to other deep embedding based clustering methods in terms of the binary classification metrics we use. For scenario one, our MFPCA clustering method identifies all the cyber-attacks successfully. Besides, our method successfully identifies 82.35% of all the time slots with short circuit faults. For scenario two, our method identifies all the open circuit faults, and 94.12% of the time slots with cyber-attacks. Compared to the open circuit fault, it is harder to distinguish the short circuit fault from the cyber-attack.

VII. Conclusion

This paper presents a novel framework for solving the anomaly detection and diagnosis problems in power electronic networks. To detect anomalies, we use a novel informative leveraging for anomaly detection (ILAD) algorithm that does not need any data to train the algorithm. Compared to other deep learning algorithms that need labels of the normal data or labels of both the normal and anomalous data, the proposed algorithm is unsupervised and does not need labels to train. Compared to other unsupervised score based anomaly detection methods, the proposed method has higher accuracy. Furthermore, it is shown that our offline ILAD algorithm can be generalized to the online ILAD by sketching the lag-covariance matrix.

Most available work uses supervised classification models for the anomaly diagnosis task. However, the labels for anomaly types in the power electronic networks are not easily accessible in real applications. Therefore, we use an unsupervised Multivariate Functional Principal Component Analysis (MFPCA) clustering method to train the algorithm without labels. Based on the model trained by offline cases, for each time window, we tested the data in an online manner to decide on the cluster association. To the best of our knowledge, this

is the first article to use unsupervised anomaly detection and diagnosis algorithm for the power electronic network.

It should be mentioned that more work needs to be done in the future to make our method discover novel anomaly types. Our clustering model cannot discover new clusters in an online scenario as more data streams in. To make the algorithm identify new clusters, we may need to borrow ideas from dynamic linear models to generalize the MFPCA clustering algorithm to a dynamic version.

APPENDIX

APPENDIX A
PSEUDO CODE OF THE ONLINE INFORMATIVE LEVERAGING FOR ANOMALY DETECTION ALGORITHM

Algorithm 1: Online Informative Leveraging Anomaly Detection and Diagnosis Algorithm

Result: The detected anomaly start time t^{str}, end time t^{end} and the anomaly type g.

Input: Streaming Input of the waveform, that is, a window of a streaming input $[X_i(t-h),..,X_i(t)]$; and time lag between two data windows l; extract a short period of starting data' feature (details of feature extraction refers to IV-A) as initials $[\overline{\mathbf{X}}_i(t_0),...\overline{\mathbf{X}}_i(t_1)]$; the anomalous state $S=-1$, which means that the status is normal;

for *time interval* $[t-h,t]$ **do**

- 1) Extract the TFD features $[\overline{\mathbf{X}}_i(t-h^*),...,\overline{\mathbf{X}}_i(t)]$ from the input data window $[X_i(t-h),..,X_i(t)]$ (details of extracting TFD features can be found in section IV-A);
- 2) Based on the TFD features, get the streaming informative leverage scores $\tilde{l}_{qq}^k$ of each data point q in time window $[t-h^*,...,t]$ (details of calculating informative leverage scores can be found in section IV-B);
- 3) Based on the calculated informative leverage score, apply the sequential change point detection algorithm (See details in B);
- 4) **if** *the informative leverage score is flagged as a change point* **then**
  - The anomalous state changes, and $S=-S$;
  - **if** *the anomalous state* $S=1$*, which means the status is abnormal* **then**
    - Do anomaly diagnosis by MFPCA clustering to identify if it's a cyber-attack or physical fault. (details of the MFPCA clustering algorithm can be found in section IV-C);
  - **end**
  - **if** *the anomalous state* $S=-1$*, which means the status is normal* **then**
    - The anomaly ends;
  - **end**
- **end**

end

APPENDIX B
CHANGE POINT DETECTION ALGORITHM FOR ANOMALY DETECTION

Here, we provide details on the change point detection algorithm we used in the article ([31]). To test whether a change point occurred at some time point p, we first divide the observations into two samples, $x_1,...,x_p$, and $x_{p+1},...,x_n$, and apply a likelihood ratio test to determine whether the data before the time point p has the same mean and variance as the data after time point p. The null hypothesis of the likelihood ratio test is:

$$H_0 : x_i \sim \mathrm{Exp}\,(\lambda_0) \quad \forall i \tag{25}$$

The alternative hypothesis of the likelihood ratio test is:

$$H_1 : x_i \sim \begin{cases} \mathrm{Exp}\,(\lambda_0) & \text{if } i \leqslant p \\ \mathrm{Exp}\,(\lambda_1) & \text{if } i > p \end{cases} \tag{26}$$

where Exp denotes an exponential distribution and λ_0 and λ_1 are unknown parameters. Then, the statistic corresponding to the generalized likelihood ratio test is given by:

$$M_{p,n} = -2\left(n\log\frac{n}{S_{0,n}} - p\log\frac{p}{S_{0,p}} - (n-p)\log\frac{n-p}{S_{p.n}}\right) \tag{27}$$

where $S_{r,s} = \sum_{i=r+1}^{s}\left(x_i - \bar{x}_{r,s}\right)^2/(s-r)$, and $\bar{x}_{r,s}$ is the mean value of data from time point r to s. This statistic is used for testing whether the time point p is the change point in a sequence of length n. Since the value p is unknown, we use the statistic defined below to identify whether the sequence contains a change point:

$$\begin{aligned} M_n &= \max_p M_{p,n}, \quad 2 \leqslant p \leqslant\ t-2 \\ p^* &= \arg\max_p M_{p,n}, \quad 2 \leqslant p \leqslant\ t-2 \end{aligned} \tag{28}$$

If the test statistics $M_n > h_n$ for some threshold h_n, then the point p^* is the detected change point.

For sequential change point detection, we process the observations sequentially. The statistics M_n is calculated using the observations between the current data point and the past observations. If $M_n > h_n$ then the change point is marked, otherwise M_{n+1} is computed. This sequential change point detection algorithm can help identify the duration of anomalies when the informative leverage scores change drastically.

REFERENCES

1 Sabbaghpur Arani, M. and Hejazi, M. A., "The comprehensive study of electrical faults in pv arrays," *Journal of Electrical and Computer Engineering*, vol. 2016, 2016.

2 Li, F., Xie, R., Yang, B., Guo, L., Ma, P., Shi, J., Ye, J., and Song, W., "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*, 2019.

3 Tan, S., Guerrero, J. M., Xie, P., Han, R., and Vasquez, J. C., "Brief survey on attack detection methods for cyber-physical systems," *IEEE Systems Journal*, vol. 14, no. 4, pp. 5329–5339, 2020.

4 Alguliyev, R., Imamverdiyev, Y., and Sukhostat, L., "Cyber-physical systems and their security issues," *Computers in Industry*, vol. 100, pp. 212–223, 2018.

5 Singh, A. and Jain, A., "Study of cyber attacks on cyber-physical system," in *Proceedings of 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT)*, 2018, pp. 26–27.

6 Anwar, A., Mahmood, A. N., and Shah, Z., "A data-driven approach to distinguish cyber-attacks from physical faults in a smart grid," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 2015, pp. 1811–1814.

7 Zhang, C., Song, D., Chen, Y., Feng, X., Lumezanu, C., Cheng, W., Ni, J., Zong, B., Chen, H., and Chawla, N. V., "A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 1409–1416.

8 Saraswat, D., Bhattacharya, P., Zuhair, M., Verma, A., and Kumar, A., "Ansmart: A svm-based anomaly detection scheme via system profiling in smart grids," in *2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)*. IEEE, 2021, pp. 417–422.

9 Ahmed, A., Sajan, K. S., Srivastava, A., and Wu, Y., "Anomaly detection, localization and classification using drifting synchrophasor data streams," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3570–3580, 2021.

10 Yaacob, A. H., Tan, I. K., Chien, S. F., and Tan, H. K., "Arima based network anomaly detection," in *2010 Second International Conference on Communication Software and Networks*. IEEE, 2010, pp. 205–209.

11 Li, F., Li, Q., Zhang, J., Kou, J., Ye, J., Song, W., and Mantooth, H. A., "Detection and diagnosis of data integrity attacks in solar farms based on multilayer long short-term memory network," *IEEE Transactions on Power Electronics*, vol. 36, no. 3, pp. 2495–2498, 2020.

12 Li, Q., Li, F., Zhang, J., Ye, J., Song, W., and Mantooth, A., "Data-driven cyberattack detection for photovoltaic (pv) systems through analyzing micro-pmu data," in *2020 IEEE Energy Conversion Congress and Exposition (ECCE)*. IEEE, 2020, pp. 431–436.

13 Xie, R., Wang, Z., Bai, S., Ma, P., and Zhong, W., "Online decentralized leverage score sampling for streaming multidimensional time series," in *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, vol. 89. PMLR, 16–18 Apr 2019, pp. 2301–2311.

14 Yang, Q., Gultekin, M. A., Seferian, V., Pattipati, K., Bazzi, A. M., Palmieri, F. A., Rajamani, R., Joshi, S. N., Farooq, M., and Ukegawa, H., "Incipient residual-based anomaly detection in power electronic devices," *IEEE Transactions on Power Electronics*, vol. 37, no. 6, pp. 7315–7332, 2022.

15 Gajanur, N. R., Greidanus, M. D., Mazumder, S. K., and Abbaszada, M. A., "Impact and mitigation of high-frequency side-channel noise intrusion on the low-frequency performance of an inverter," *IEEE Transactions on Power Electronics*, pp. 1–1, 2022.

16 Barua, A. and Al Faruque, M. A., "Hall spoofing: A non-invasive dos attack on grid-tied solar inverter," in *29th {USENIX} Security Symposium ({USENIX} Security 20)*, 2020, pp. 1273–1290.

17 Kune, D. F., Backes, J., Clark, S. S., Kramer, D., Reynolds, M., Fu, K., Kim, Y., and Xu, W., "Ghost talk: Mitigating emi signal injection attacks against analog sensors," in *2013 IEEE Symposium on Security and Privacy*. IEEE, 2013, pp. 145–159.

18 Guo, L., Zhang, J., Ye, J., Coshatt, S. J., and Song, W., "Data-driven cyber-attack detection for pv farms via time-frequency domain features," *IEEE Transactions on Smart Grid*, 2021.

19 Ross, G. J., Tasoulis, D. K., and Adams, N. M., "Sequential monitoring of a bernoulli sequence when the pre-change parameter is unknown," *Computational Statistics*, vol. 28, no. 2, pp. 463–479, 2013.

20 Ma, P., Mahoney, M., and Yu, B., "A statistical perspective on algorithmic leveraging," in *International Conference on Machine Learning*. PMLR, 2014, pp. 91–99.

21 Justin, D., Concepcion, R. S., Calinao, H. A., Lauguico, S. C., Dadios, E. P., and Vicerra, R. R. P., "Application of ensemble learning with mean shift clustering for output profile classification and anomaly detection in energy production of grid-tied photovoltaic system," in *2020 12th International Conference on Information Technology and Electrical Engineering (ICITEE)*. IEEE, 2020, pp. 286–291.

22 Dey, M., Rana, S. P., Simmons, C. V., and Dudley, S., "Solar farm voltage anomaly detection using high-resolution μpmu data-driven unsupervised machine learning," *Applied Energy*, vol. 303, p. 117656, 2021.

23 Huang, S., Quek, S., and Phoon, K., "Convergence study of the truncated karhunen–loeve expansion for simulation of stochastic processes," *International journal for numerical methods in engineering*, vol. 52, no. 9, pp. 1029–1043, 2001.

24 Jansen, R., "Maximum likelihood in a generalized linear finite mixture model by using the em algorithm," *Biometrics*, pp. 227–231, 1993.

25

26

27 Doğu, E. and Kocakoç, İ. D., "A multivariate change point detection procedure for monitoring mean and covariance simultaneously," *Communications in Statistics-Simulation and Computation*, vol. 42, no. 6, pp. 1235–1255, 2013.

28 Montgomery, D. C., *Introduction to statistical quality control*. John Wiley & Sons, 2020.

29 Woodall, W. H. and Ncube, M. M., "Multivariate cusum quality-control procedures," *Technometrics*, vol. 27, no. 3, pp. 285–292, 1985.

30 Mason, R. L. and Young, J. C., *Multivariate statistical process control with industrial applications*. SIAM, 2002.

31 Ross, G. J., "Sequential change detection in the presence of unknown parameters," *Statistics and Computing*, vol. 24, no. 6, pp. 1017–1030, 2014.