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

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

***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 between different types of anomalies, such as 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, resulting in 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 the 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.**

Manuscript received 18 July 2022; revised 17 December 2022 and 24 April 2023; accepted 5 May 2023. Date of publication 23 October 2023; date of current version 21 February 2024. This work was supported in part by the U.S. National Science Foundation under Grant DMS-1903226, Grant DMS-1925066, Grant DMS-2124493, Grant SaTC-2019311, and Grant ECCS-1946057; in part by the U.S. National Institute of Health under Grant R01GM122080; and in part by the U.S. Department of Energy under Grant DE-EE0009026. Paper no. TSG-00826-2022. *(Corresponding author: WenZhan Song.)*

Shushan Wu, Luyang Fang, T. N. Sriram, Wenxuan Zhong, and Ping Ma are with the Department of Statistics, University of Georgia, Athens, GA 30602 USA (e-mail: shushanwu@uga.edu; Luyang.Fang@uga.edu; tn@stat.uga.edu; wenxuan@uga.edu; pingma@uga.edu).

Jinan Zhang, Stephen J. Coshatt, Jin Ye, and WenZhan Song are with the College of Engineering, University of Georgia, Athens, GA 30602 USA (e-mail: jinan.zhang@uga.edu; stephen.coshatt@uga.edu; jin.ye@uga.edu; wsong@uga.edu).

Feraidoon Zahiri is with Robins Air Force Base, Warner Robins, GA, USA (e-mail: feraidoon.zahiri@us.af.mil).

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

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSG.2023.3325276.

Digital Object Identifier 10.1109/TSG.2023.3325276

I. Introduction

In SMART grids, power electronics are the fundamental building blocks. The expansion of 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., the converter, also threaten the safety and security of smart grids. A catastrophic failure of power electronic networks due to a malicious cyber-attack [1], [2], or an accidental physical fault would cause degradation of equipment and substantial economic loss. Furthermore, false identification of the root cause might lead to severe operational failure while performing mitigation strategies in the power electronic networks [3], [4]. Early detection and diagnosis of the anomalies are essential for the timely maintenance and recovery of power electronic networks [5], [6].

Most unsupervised anomaly detection and diagnosis algorithms are offline, which necessitates making decisions based on all the data across time [7], [8]. The approaches can present challenges when dealing with streaming data in real-world situations. Therefore, there is an urgent need to develop an online framework that can detect and diagnose anomalies in real-time. Such a framework would enhance the reliability and efficiency of smart grid operations, ensuring a prompt response to anomalies as they occur.

To overcome the aforementioned challenges, we propose an unsupervised, data-driven approach for anomaly detection and diagnosis called the Informative Leveraging for Anomaly Detection (ILAD) algorithm, which is combined with a Multivariate Functional Principal Component Analysis (MFPCA) clustering algorithm to distinguish between cyber-attacks and physical faults for anomaly diagnosis. We assume that the micro phasor measurement unit (μPMU) is installed at the point of coupling (PCC) of PV farms as shown in Figure 1. To detect the cyber-attacks and faults in PV farms, both voltage and current waveform data are measured at PCC first. Several features, including PMU data, Total Harmonic Distortion (THD), and mean current vector (MCV) are then extracted. Finally, the proposed method employs the extracted features to detect and identify cyber-attacks and faults. 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 TFD features using a Vector Autoregressive (VAR) model and calculate informative leverage scores [5], [9] for each time window. Both offline and online experiments show that the ILAD algorithm achieves high accuracy.

After performing anomaly detection, we assign the type of anomalies (cyber-attack or physical fault) to the identified anomalous windows based on our MFPCA clustering algorithm. To evaluate the performance of the proposed ILAD and MFPCA clustering algorithms, we conduct experiments using a PV farm as a case study and generate a variety of electric waveform data under both offline and online scenarios. In offline scenarios, the proposed offline ILAD (off-ILAD) successfully identifies the anomalies in 42 out of 43 cases, achieving an accuracy of about 0.94 for the anomaly diagnosis task, which is a competitive and comparable result to the classification task. For online anomaly detection, the proposed online ILAD (on-ILAD) algorithm achieves higher accuracy compared to other change point detection algorithms. For online 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 TFD features in power electronic networks.
2) We propose a novel 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 that triggers an alert if the informative leverage score rises significantly, instead of heuristically using a threshold [10], making it more robust to new anomalies.
4) We apply a novel MFPCA clustering algorithm to the power electronic network, which projects the TFD features onto lower-dimensional spaces spanned by eigenfunctions. Thus, the MFPCA clustering algorithm extracts features distinguishing cyber-attacks from physical faults.
5) Our algorithm can operate online to detect anomalies and diagnose their types based on TFD features in each time window as they occur.

The paper is organized as follows. In Section II, we review related literature. Section III presents the model of the power electronic network and attacks, which includes a typical power electronic network in a PV farm, a cyber-attack model, and a physical fault model. The problem setup is described in Section IV. In section V, we provide the necessary background before introducing our proposed algorithms. The proposed algorithms, including feature extraction, informative leverage score, and multivariate functional principal component analysis, are presented in Section VI. In Sections VII and VIII, we present the experimental results for offline and online scenarios, respectively. Finally, we conclude our work in Section IX.

II. Related Works

To the best of our knowledge, several studies have explored anomaly detection and diagnosis by information embedded in electrical signals for cyber-threat in cyber-physical systems [11], presenting a great opportunity to advance cyberspace security and trustworthy research and design. Anomaly detection methods can be either supervised or unsupervised. Unsupervised methods. For the supervised ones, several supervised deep learning algorithms [3], [6], [12], [13], [14], [15], [16], [17], [18], [19] have been developed for anomaly detection. These methods use two different pathways to solve the anomaly detection problem. One pathway is to train Autoencoder models [3], [18] to reconstruct the distribution of normal data, and flags anomalies if the reconstruction error of testing data significantly exceeds that of the normal data. Such methods rely on labeled normal data to train and assume the same distribution for the training and testing data, which may limit their applicability. For example, when the system load increases, those methods will tend to cause 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 [6], [13], [14], [15], [17], [19] by Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) network. Since training deep learning models needs to balance the trade-off between training and generalization, inappropriate training will lead to bad performance in the testing data with different distribution from the training data [20]. In this case, novel anomalies with different patterns from the training data enter, and then the detector trained by deep learning models would misclassify the data, which leads to substantial loss. Another disadvantage of the supervised model is that the classification models need a large amount of labeled training data to increase the accuracy [21], which is challenging to obtain in smart grids.

Some anomaly detection methods are unsupervised, such as the isolation forest [8], [22], and do not need labels in the training process. However, the isolation forest algorithm cannot be updated in real time, which limits its usefulness for real-time anomaly detection. Several statistics-based methods [5], [23], [24] can be used for real-time anomaly detection. For example, the correlation-based [23] method can calculate an anomalous score based on the correlation between two PMU parameters, but it requires a pre-defined threshold to detect anomalies. Expert knowledge and experience are needed to determine a suitable threshold, and an inappropriate threshold can lead to false alarms and false negatives. The leverage score based on the VAR model, proposed in [5], achieves higher accuracy than other unsupervised methods.

TABLE I
SUMMARY OF ANOMALY DETECTION AND DIAGNOSIS METHODS

Method	Year	Online?	Unsupervised?	Objective	Gap
CRED [3]	2019	Yes	No	Detection	Need normal data to train
HTM [18]	2020	Yes	No	Detection	Need normal data to train
Statistical Correlation [23]	2019	Yes	Yes	Detection	Pre-defined threshold
Isolation Forest [8], [22]	2019, 2022	No	Yes	Detection	Non-online
SSL [12]	2019	Yes	No	Detection	Need labels to train
Leverage [5]	2019	Yes	Yes	Detection	Less Informative
CNN [15], [13], [6]	2020, 2021	Yes	No	Both	Need labels to train
LSTM [14], [19]	2020, 2022	Yes	No	Both	Need labels to train
PMUNET [17]	2021	Yes	No	Diagnosis	Need labels to train
SVM [16]	2021	Yes	No	Diagnosis	Need labels to train
Incremental Classifier [25]	2018	Yes	No	Diagnosis	Need labels to train
BMF [5]	2019	Yes	Yes	Diagnosis	No time-dependence
Proximity-based Clustering [26]	2019	No	Yes	Diagnosis	No time dependence

The leverage score is more effective, as shown in Figure 6 where the red lines indicate its performance. To overcome the limitations of the aforementioned methods, we propose an unsupervised real-time anomaly detection algorithm with higher accuracy than other unsupervised methods in the experiment results in Table III and V. Moreover, the proposed framework utilizes the change point detection algorithm, thus, it can circumvent the limitations of the pre-defined threshold.

Anomaly diagnosis is the task of distinguishing cyber-attacks from physical faults, which can be approached as either a classification or a clustering problem, depending on whether labels are used in training. Most available approaches focus on supervised learning methods [6], [13], [14], [15], [16], [17], [19], [25] that use a support-vector-based algorithm or deep learning framework for binary-classification to distinguish cyber-attacks from physical faults. Similar to the deficiencies in supervised methods of anomaly detection, supervised methods for anomaly diagnosis need large labeled training data to achieve high performance, which can be a significant limitation. Unsupervised methods [5], [26] are more potent in applications since they do not need label information during training. The proximity methods [26] treat each time point independently, and apply clustering methods such as K-means on the multivariate features of each time point. Binary Matrix Factorization (BMF) [5] can extract high-dimensional features from the entire time series, and t-SNE can reduce the dimension of the features. Then, proximity-based clustering methods can be easily applied to the whole dataset. However, the above two methods have limitations, as they do not capture the inter-dependence of different time series dimensions and the temporal dependency of single time series. To address these limitations, we propose to use MFPCA-based clustering. MFPCA approximates the data probability distribution function (p.d.f.) by the product of the p.d.f. of the principal components, which is the projection of the input features on the lower dimensional space spanned by eigenfunctions, calculated by the singular value decomposition of the covariance matrix of multivariate time series. Thus, MFPCA is able to capture the inter-dependence of the multi-dimensional time series and auto-correlation within a single dimension of the time series.

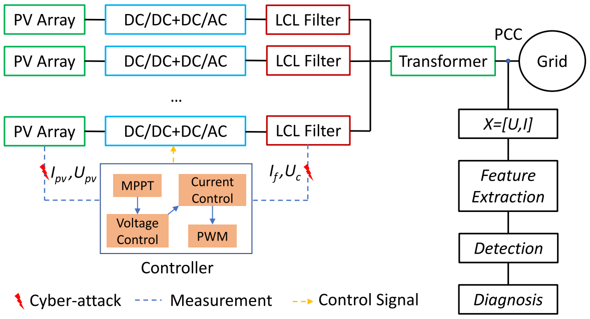

Fig. 1. Schematic diagram of the power electronic converter-enabled PV farm. I_{pv}, U_{pv}, I_f, and U_c are PV array current, PV array voltage, inductance current in the LCL, and capacitor voltage in the LCL, respectively.

For a comprehensive comparison of different methods, we summarize several aspects that are important to the real application of anomaly detection and diagnosis, whether it's online or not, unsupervised or not, for anomaly detection or diagnosis, and the gap existing in the methods to achieve the goal, accurate and real-time unsupervised anomaly detection and diagnosis. To make the comparison clear, we present the summary of all related works in Table I.

III. POWER ELECTRONIC NETWORK AND ATTACK MODELS

A. A General Power Electronic Network Model

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

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

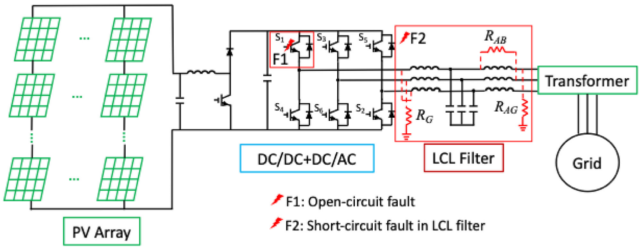

Fig. 2. Physical faults in PV converter and high-voltage line.

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, and

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

where R and L are the resistance and inductance.

B. Cyber-Attack Model

As discussed in many studies [27], [28], [29], 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

$$Y_A(t) = \alpha Y_o(t) + \beta \quad (3)$$

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 as shown in Figure 1, α is a multiplicative factor, and β is the false data injection.

C. Physical Fault

Besides cyber-attacks, physical faults also threaten PV farms. As shown in Figure 2, two types of physical faults, including open-circuit faults in the switch (F1), and short-circuit faults in the transmission line (F2), are modeled and simulated in the real-time testbed. In F1, the open-circuit fault occurs in a switch of the PV converter, which leads to the open transistor. Short-circuit fault causes a heavy current which creates overheating or destroys the equipment in the power grid. As shown in Figure 2, three-phase, two-phase, and single-phase short-circuit faults are modeled. R_{AB} is the fault resistance between Phase A and Phase B. R_{AG} is the fault resistance between Phase A and the ground. R_G is the ground resistance. The model for F2 results in incorrect connections between each phase on the high-voltage line. More details of the setting and types of physical faults we used in the article are presented in Table II.

IV. 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 [6]. We denote the nine-dimensional TFD features, a multivariate time series, by $\overline{\mathbf{X}}(t) = [\overline{X}^1(t), \ldots, \overline{X}^\ell(t), \ldots, \overline{X}^9(t)]$, 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 diagnose the anomalous portion of the series, $[\overline{\mathbf{X}}_i(t_{k+1}), \ldots, \overline{\mathbf{X}}_i(t_{k+T})]$. Given this anomalous series, the second goal is to assign an anomaly type (*cyber-attack* or *physical fault*) to each detected anomalous time period.

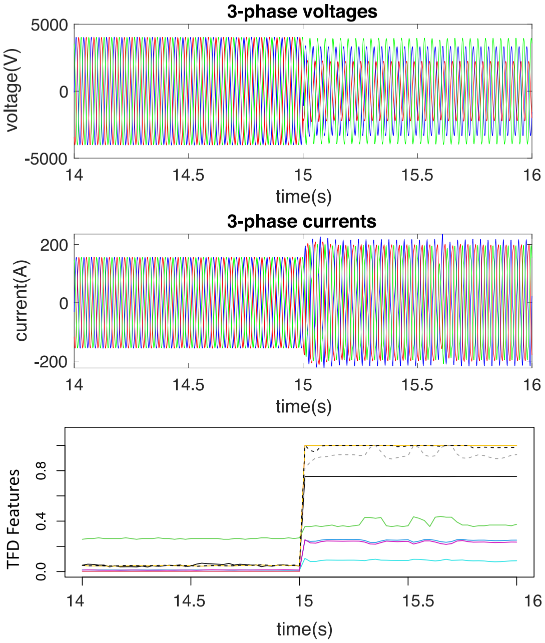

Fig. 3. An example of waveform data for three-phase voltages (bottom panel), three-phase currents (middle panel), and nine-dimensional TFD features (bottom panel), respectively, for one case. This example shows the data in a time range (14-16s). The anomaly happens at 15s.

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 $i = 1, \ldots, n$. 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 $t = t_{k+1}, \ldots, t_{k+T}$ 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 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

TABLE II
SETTING OF PHYSICAL FAULTS

Fault type	Location	Definition	Case No.
F1	IGBT	IGBT S_1	F_1
		IGBT S_2	F_2
		IGBT S_3	F_3
		IGBT S_5	F_4
F2	High voltage (HV) line phase A	$R_G \in \{1, 0.1\}$ Ohm	F_5,F_6
	HV line phase A	R_A = 0.2 Ohm & R_G = 0.1 Ohm	F_7
	HV line phase AB	R_{AB} = 2 Ohm	F_8
		R_{AB}= 2 Ohm & $R_G \in \{1, 0.1\}$ Ohm	F_9,F_{10}
		R_{AB}= 0.2 Ohm & R_G= 0.1 Ohm	F_{11}
	HV line phase AC	R_{AC} = 2 Ohm	F_{12}
		R_{AC}= 2 Ohm & $R_G \in \{1, 0.1\}$ Ohm	F_{13}, F_{14}
		R_{AC}= 0.2 Ohm & R_G= 0.1 Ohm	F_{15}
	HV line phase ABC	R_{ABC}= 2 Ohm & $R_G \in \{1, 0.1\}$ Ohm	F_{16}, F_{17}
		R_{ABC}= 0.2 Ohm & R_G= 0.1 Ohm	F_{18}

observations. As shown in Figure 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 [5], [9] did.

B. Anomaly Diagnosis Problem

There are two major anomaly types in the device-level power electronics converters (PEC), cyber-attacks and physical faults. While these are two common types of anomalies, it is hard to distinguish cyber-attacks from physical faults. 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 $[\overline{\mathbf{X}}_i(t_{k+1}), \ldots, \overline{\mathbf{X}}_i(t_{k+T})]$ 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.

V. PRELIMINARIES

Before proceeding into the details of our proposed algorithm, we introduce the background knowledge of the paper including the VAR model, calculation of leverage score in the VAR model, and details about the embedding method for time series, MFPCA.

A. Vector Autoregressive Model

A classical p-th order VAR model representation characterizes the temporal dependence structure of the time series $\overline{\mathbf{X}}(t)$:

$$\overline{\mathbf{X}}(t_k) = \overline{\mathbf{X}}(t_{k-1})\mathbf{A}_1 + \overline{\mathbf{X}}(t_{k-2})\mathbf{A}_2 + \cdots + \overline{\mathbf{X}}\big(t_{k-p}\big)\mathbf{A}_p + \boldsymbol{\epsilon}(t) \tag{4}$$

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 (4) can also be expressed in the form of a linear model:

$$\overline{\mathbf{Y}} = \overline{\mathbf{D}}^{p}\boldsymbol{A} + \boldsymbol{\epsilon}, \tag{5}$$

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

$$\begin{bmatrix} \overline{\mathbf{X}}(t_{k-1}) & \overline{\mathbf{X}}(t_{k-2}) & \cdots & \overline{\mathbf{X}}\big(t_{k-p}\big) \\ \overline{\mathbf{X}}(t_k) & \overline{\mathbf{X}}(t_{k-1}) & \cdots & \overline{\mathbf{X}}\big(t_{k-p+1}\big) \\ \vdots & \vdots & \ddots & \vdots \\ \overline{\mathbf{X}}(t_{k+T-1}) & \overline{\mathbf{X}}(t_{k+T-2}) & \cdots & \overline{\mathbf{X}}\big(t_{k+T-p}\big) \end{bmatrix}, \tag{6}$$

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

B. Leverage Score in VAR

By the linear model representation in equation (5), 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. The leverage score in the linear model has been generalized to the VAR model in time series [9]. 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. Since the TFD features drastically change when an anomaly happens, we are motivated to use the leverage scores to detect anomalies. For each case, the leverage score of the q-th observation can be expressed as

$$l_{qq} = \overline{\mathbf{d}}_{(q)}^{p\;T}\left(\overline{\mathbf{D}}^{pT}\overline{\mathbf{D}}^{p}\right)^{-1}\overline{\mathbf{d}}_{(q)}^{p}, \tag{7}$$

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

C. Multivariate Functional Principal Component Analysis

MFPCA can embed the multivariate time series into a low-dimensional space spanned by eigenfunctions based on Karhunen-Loeve expansion [30]. Compared with other deep learning classification models [5], [14], [17], [31], the advantage of the proposed MFPCA clustering method is that it is interpretable and does not use labels to train. Compared with other unsupervised clustering approaches [3], [5], [26], the advantage of MFPCA is that it can model both the interdependence of different dimensions of time series and the auto-correlation of a single dimension of time series.

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,

$$\forall t \in [t_1, t_2], \quad \lim_{h\to 0} \mathbb{E}\Big[\|\overline{\mathbf{X}}(t+h) - \overline{\mathbf{X}}(t)\|^2\Big]$$
$$= \lim_{h\to 0} \int_{t_1}^{t_2} \sum_{\ell=1}^{9} \mathbb{E}\Big[\Big(\overline{X}^{\ell}(t+h) - \overline{X}^{\ell}(t)\Big)^2\Big] dt = 0. \qquad (8)$$

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 $\mu^{\ell} = \{\mu^{\ell}(t) = \mathbb{E}[\overline{X}^{\ell}(t)]\}_{t\in[0,T]}$, and let $\mu(t) = \mathbb{E}[\overline{\mathbf{X}}(t)] = (\mu^1, \ldots \mu^{\ell}, \ldots, \mu^9)^T$. We further define the covariance function of $\overline{\mathbf{X}}(t)$ as:

$$V(s,t) = \mathbb{E}\Big[\Big(\overline{\mathbf{X}}(s) - \mu(s)\Big) \otimes \Big(\overline{\mathbf{X}}(t) - \mu(t)\Big)\Big], \qquad (9)$$

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

$$\int_{t_1}^{t_2} V(\cdot, t)\boldsymbol{f_m}(t)dt = \lambda_m \boldsymbol{f}_m, \qquad (10)$$

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 $\{\lambda_m\}_{m\geq 1}$ are associated eigenvalues. Consequently, the principal component $\{\mathrm{C}_m\}_{m\geq 1}$ are the projections of $\overline{\mathbf{X}}$ on the space spanned by the eigenfunctions $\{\boldsymbol{f}_m\}_{m\geq 1}$ of the covariance function:

$$\mathrm{C}_m = \int_{t_1}^{t_2} \sum_{\ell=1}^{9} \Big(\overline{X}^{\ell}(t) - \mu^{\ell}(t)\Big) f_m^{\ell}(t)dt, \qquad (11)$$

where the principal components $\{\mathrm{C}_m\}_{m\geq 1}$ are zero-mean uncorrelated random variables with variance $\{\lambda_m\}_{m\geq 1}$, 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:

$$\overline{\mathbf{X}}(t) = \sum_{m=1}^{q'} \mathrm{C}_m \boldsymbol{f}_m(t), \qquad t \in [t_1, t_2]. \qquad (12)$$

The truncation leads to a dimension-reduced subspace.

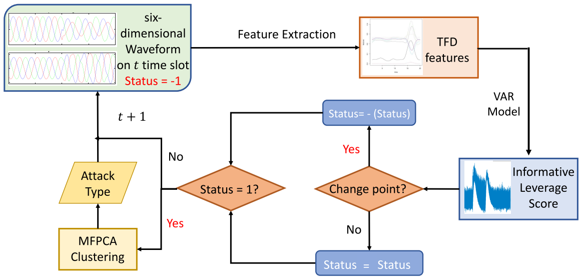

Fig. 4. The workflow chart of the online algorithm of anomaly detection and anomaly diagnosis.

VI. 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 the 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 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 [6] to identify the onset of anomalies and to distinguish between the two anomaly types via distinct patterns.

1) Frequency Domain Features: First, we obtain μPMU features through fast Fourier transform (FFT) to project a signal into the frequency domain. Since the signal is distorted when an anomaly happens, we use THD to capture the harmonic information of the distorted waveform. This yields

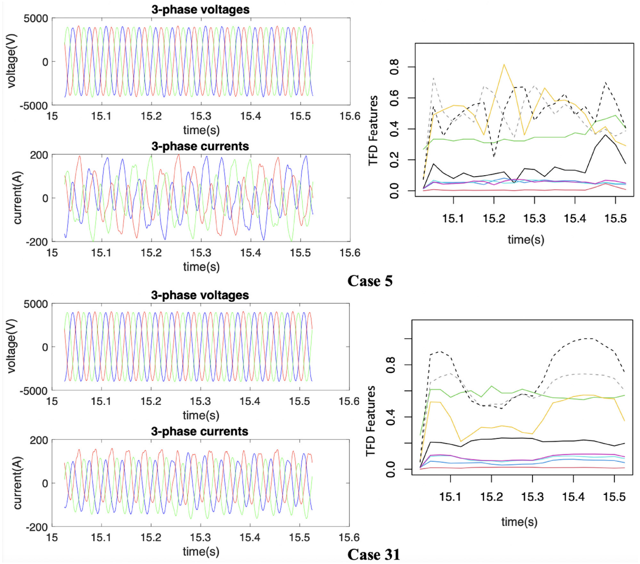

Fig. 5. An example of waveform data for two cases (5 and 31) and extracted TFD features. The first column shows the plots of waveform data, and the second column shows the plots of extracted TFD features. Case 5 encounters a cyber-attack while case 31 encounters a physical fault. In this example, we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

a feature vector denoted by

$$\mathbf{F} = \left[M_{\{\cdot\}}, T_{\{\cdot\}}\right], \tag{13}$$

where $M_{\{.\}}$, 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. Whereas the THD in a waveform is known to be lower than some boundaries under normal conditions. Through expert knowledge, the maximum THD is set as $T_{max} = 5\%$. Then, THD for each phase is normalized as follows:

$$\bar{T}_{\{\cdot\}} = \min\left\{\frac{T_{\{\cdot\}}}{T_{\max}}, 1\right\} \tag{14}$$

The raw μPMU features sometimes lead to false positive results, especially when the magnitude is affected by a huge change in irradiance. Thus, we extract the difference between the magnitudes of the three-phase waveforms R_m to distinguish physical faults from cyber-attacks:

$$\begin{aligned} R_{m,I} &= \sqrt{\Delta M_{I_1}^2 + \Delta M_{I_2}^2 + \Delta M_{I_3}^2} \\ R_{m,U} &= \sqrt{\Delta M_{U_1}^2 + \Delta M_{U_2}^2 + \Delta M_{U_3}^2} \\ R_m &= \left(R_{m,I} + R_{m,U}\right)/2 \end{aligned} \tag{15}$$

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 and scaling, the magnitude-based features become:

$$\begin{aligned} \bar{R}_{m1} &= \min\left\{\frac{R_m}{R_{m1,\max}}, 1\right\} \\ \bar{R}_{m2} &= \min\left\{\frac{\ln(R_m + e) - 1}{R_{m2,\max}}, 1\right\}, \end{aligned} \tag{16}$$

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 features, the transformation of the time domain features, three-phase currents, helps distinguish physical faults from cyber-attacks. We use a variant of the MCV by current Concordia transformation, which is used in anomaly detection for power electronic networks [6]:

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

The degree of distortion of points (I_α, I_β) at a time point t_k indicates physical faults. Thus, we define the MCV point at time t_k as:

$$P_{mcv}(t_k) = \left(\frac{1}{N_k}\sum_{i=t_k-N_k+1}^{t_k} I_\alpha(i), \frac{1}{N_k}\sum_{i=t_k-N_k+1}^{t_k} I_\beta(i)\right) \tag{17}$$

According to the domain knowledge from [6], since the poor circuit contacts would affect the MCV locations, thus, Concordia transformation of MCV has clear patterns when an open circuit fault happens. Thus, $\bar{P}_{mcv}$ is defined based on the maximum number of points of all regions in the panel (I_α, I_β). The $\bar{P}_{mcv}$ feature represents the concentration of MCV points, which is helpful when distinguishing open-circuit faults from other threats.

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:

$$\overline{\mathbf{X}} = \left[\bar{R}_{m1}, \bar{R}_{m2}, \bar{P}_{mcv}, \bar{T}\right] \tag{18}$$

We refer to the above 9-dimensional feature as the TFD features, where $\bar{R}_{m1}$, $\bar{R}_{m2}$, and $\bar{P}_{mcv}$ are all scaler features, and $\bar{T}$ are six-dimensional 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 [32] 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. For this, a generalization of the idea in [9] yields a streaming leverage score that only utilizes the history and the current information to approximate the leverage score.

1) 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 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 [9] 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:

$$\tilde{l}_{qq} = \overline{\mathbf{d}}^{p\ T}_{(q)}\left(\boldsymbol{\Gamma}^p_r\right)^{-1}\overline{\mathbf{d}}^{p}_{(q)}, \tag{19}$$

where $\boldsymbol{\Gamma}^p_r$ 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 $\boldsymbol{\Gamma}^p_r$ 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

$$\tilde{l}_{qq} = \sum_{j=1}^{r-p}\left(\overline{\mathbf{d}}^{p\ T}_{(q)}\boldsymbol{v}^{(j)}\right)^2/\boldsymbol{\sigma}^2_j, \tag{20}$$

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 $\boldsymbol{\Gamma}^p_r$. 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 projected data in the corresponding singular vector space. In our case, each pair of eigenvalue and associated Principal Component (PC) of the lag-covariance matrix characterizes an oscillatory mode, e.g., 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.

2) Informative Leverage Scores for Anomaly Detection: The aforementioned challenges motivate us to propose an informative leveraging for anomaly detection 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 $\overline{\mathbf{D}}^{p^T}_i\overline{\mathbf{D}}^{p}_i$ for each feature in the offline setting and the sketched lag-covariance matrix $\boldsymbol{\Gamma}^p_r$ in the online setting, and perform an SVD again. The result of the SVD in online and offline settings is usually denoted by $\tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^T$. 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

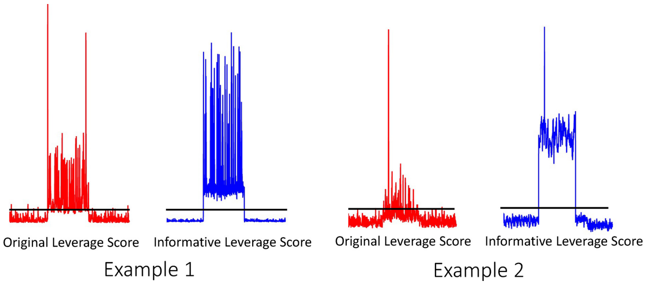

Fig. 6. An example of the original leverage score and informative leverage score. By removing the noise and obtaining an informative leverage score, the gap between the normal and the anomalous rises. This leads to higher accuracy while detecting the starts and ends of anomalies.

$$\tilde{l}^k_{qq} = \sum_{j=\in\mathcal{I}}\left(\overline{\mathbf{d}}^{p\ T}_{(q)}\tilde{\boldsymbol{v}}^{(j)}\right)^2/\tilde{\boldsymbol{\sigma}}^2_j, \tag{21}$$

where $\tilde{\boldsymbol{v}}^{(j)}$ is the j-th column of $\tilde{\mathbf{V}}$, $\tilde{\boldsymbol{\sigma}}_j$ is the j-th entry of $\tilde{\boldsymbol{\Sigma}}$, 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 6; these are calculated in an offline manner. The red lines shown in Figure 6 are the original leverage scores, and the blue lines shown in Figure 6 are the informative leverage scores. The informative leverage scores are able to remove the noise, and the gap between the score of the normal to that of the anomalous rises significantly. Thus, the performance of anomaly detection improves by removing the information from 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 examples of streaming data with five cyber-attacks and one physical fault. The coincidence between time points with high leverage scores and those indicating the presence of anomalies confirms our belief that influential points with high leverage scores are where anomalies occur. 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 [32] 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 data-driven, 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 [14], [31] use a supervised classification model, where information from labels is used for training and 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 proximity-based methods, such as K-means and hierarchical clustering [5], [26]. These methods ignore the dependency between different data features and are sensitive to outliers. In addition, these methods do not assume models. 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. Through projections by MFPCA in Section V, 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 [33]. After we apply MFPCA and embed the time series into a dimension-reduced subspace, we further assume each principal component C_m follows 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:

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

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 $\{\mathrm{C}_{m|Z^g=1}\}_{m=1,\ldots,q'}$ with zero mean and variance $\{\lambda_{m,g}\}_{m=1,\ldots,q'}$. Thus, the density of $\overline{\mathbf{X}}(t)$ can be represented by:

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

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

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

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 $\overline{x}_i$.

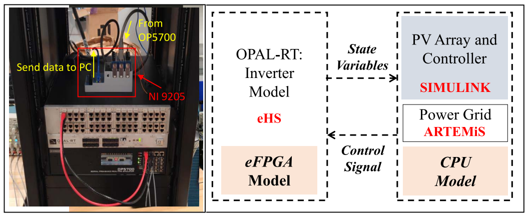

Fig. 7. Real-time testbed using OPAL-RT and NI device.

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. 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 [33] to select q'_g for g-th cluster.

VII. 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 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 7. 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. 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, 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 $(1600, 9)$.

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 detect the change points 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.

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 [34] and Multivariate CUSUM [35], for detecting the starts and ends of anomalies. We deployed these two methods since they are designed to deal with multivariate time series data. The same sequential change point detection algorithm is applied to the proposed ILAD algorithm to ensure fairness. Results are shown in Table III. 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 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 are significantly different, it is necessary to distinguish cyber-attack from physical faults 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 compare the benchmark deep embedding methods, t-SNE and UMAP, to embed the data onto a two-dimensional space and cluster the data by K-means. In addition, we compare the MFPCA method with the combination of two dimension reduction approaches, Principal Component Analysis (PCA), and t-SNE. The results are shown in Table IV. 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 four methods considered here. This is because the MFPCA, unlike other unsupervised dimension reduction approaches, could model both the inter-dependence of different dimensions of time series and the auto-correlation of a single dimension of time series. The Accuracy measure of the MFPCA algorithm is 99.12% and the F1 score is 100.00%, which are relatively higher numbers and even comparable to some of the classification algorithms [6]. 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.

TABLE III
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

TABLE IV
EXPERIMENT RESULTS OF OFFLINE ANOMALY DIAGNOSIS

Approach	Accuracy	F1	TPR	TNR
MFPCA	**0.9912**	**1.0000**	**0.9825**	**0.9912**
t-SNE	0.6002	0.6557	0.5650	0.4630
UMAP	0.6663	0.6469	0.5709	0.6190
PCA+t-SNE	0.5604	0.5529	0.5379	0.5833

VIII. ONLINE TESTING RESULTS

A. Online Experiment Setup

To validate the proposed method, we develop a real-time detection and diagnosis testbed using the NI device. As shown in Figure 7, 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, 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. There are

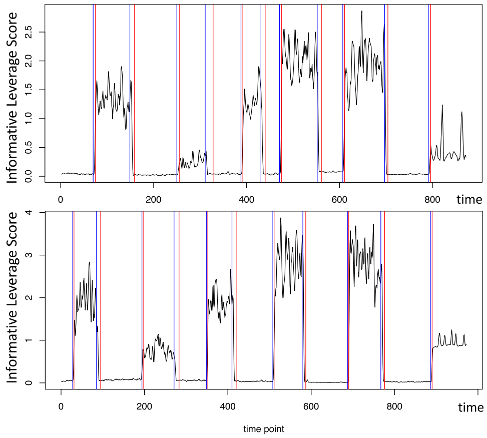

Fig. 8. The results of online Informative Leveraging for anomaly detection. The solid black line is the informative leverage score. The blue vertical lines are where the anomalies happen. The red vertical lines are the detected starts and ends of anomalies. The upper one shows the results of scenario one, and the lower one shows the results of scenario two.

6 starts and 5 ends of anomalies to be detected under both scenarios.

B. Online Test Results

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($\approx 25s$). 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. Figure 8 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 indicate where the anomalies happen. The red vertical lines indicate the detected starts and ends of anomalies. We can see that the time points with high leverage scores are consistent with the anomalies on waveform data. We then use the change point detection algorithm to sequentially detect change points of the informative leverage score. Table V shows the results of online anomaly detection for scenario one and scenario two. 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 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 V
COMPARISON OF PREDICTION RESULTS FOR SCENARIOS 1 AND 2

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

TABLE VI
EXPERIMENTAL RESULTS OF REAL-TIME ANOMALY DIAGNOSIS FOR SCENARIO 1

Approach	Accuracy	F1	TPR	TNR
MFPCA	**0.9524**	**0.9697**	**1.0000**	**0.9412**
t-SNE	0.7619	0.8571	0.8824	0.2500
UMAP	0.8095	0.8947	1.0000	0.0000
PCA+t-SNE	0.8095	0.8824	0.8824	0.5000

TABLE VII
EXPERIMENTAL RESULTS OF REAL-TIME ANOMALY DIAGNOSIS FOR SCENARIO 2

Approach	Accuracy	F1	TPR	TNR
MFPCA	**1.0000**	**1.0000**	**1.0000**	**1.0000**
t-SNE	0.8095	0.8667	0.7647	1.0000
UMAP	0.9047	0.9444	1.0000	0.5000
PCA+t-SNE	0.8095	0.8947	1.0000	0.0000

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 [6], 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 95.24% of all the time slots for scenario one. For scenario two, our method identifies all the open circuit faults and cyber-attacks. Compared to the open circuit fault, it is harder to distinguish the short circuit fault from the cyber-attack.

IX. CONCLUSION

This paper presents a novel framework for solving anomaly detection and diagnosis problems in power electronic networks. To detect anomalies, we use a novel ILAD 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 is not threshold-based and 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 MFPCA clustering method which does not need labels to train. Based on the model trained by offline cases, for each time window, we tested the data in an online manner to decide the clustering labels. 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 anomaly diagnosis algorithm 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.

REFERENCES

[1] R. Alguliyev, Y. Imamverdiyev, and L. Sukhostat, "Cyber-physical systems and their security issues," *Comput. Ind.*, vol. 100, pp. 212–223, Sep. 2018.

[2] A. Singh and A. Jain "Study of cyber attacks on cyber-physical system," in *Proc. 3rd Int. Conf. Internet Things Connect. Technol. (ICIoTCT)*, 2018, pp. 26–27.

[3] C. Zhang et al., "A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 01, pp. 1409–1416, 2019.

[4] H. T. Reda, A. Anwar, and A. Mahmood, "Comprehensive survey and taxonomies of false data injection attacks in smart grids: Attack models, targets, and impacts," *Renew. Sustain. Energy Rev.*, vol. 163, Jul. 2022, Art. no. 112423.

[5] F. Li et al., "Detection and identification of cyber and physical attacks on distribution power grids with PVs: An online high-dimensional data-driven approach," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 10, no. 1, pp. 1282–1291, Feb. 2022.

[6] L. Guo, J. Zhang, J. Ye, S. J. Coshatt, and W. Song, "Data-driven cyber-attack detection for pv farms via time-frequency domain features," *IEEE Trans. Smart Grid*, vol. 13, no. 2, pp. 1582–1597, Mar. 2022.

[7] G. Fenza, M. Gallo, and V. Loia, "Drift-aware methodology for anomaly detection in smart grid," *IEEE Access*, vol. 7, pp. 9645–9657, Jan. 2019.

[8] S. Ahmed, Y. Lee, S.-H. Hyun, and I. Koo, "Unsupervised machine learning-based detection of covert data integrity assault in smart grid networks utilizing isolation forest," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 10, pp. 2765–2777, Oct. 2019.

[9] R. Xie, Z. Wang, S. Bai, P. Ma, and W. Zhong, "Online decentralized leverage score sampling for streaming multidimensional time series," in *Proc. 22nd Int. Conf. Artif. Intell. Stat.*, 2019, pp. 2301–2311.

[10] Q. Yang et al., "Incipient residual-based anomaly detection in power electronic devices," *IEEE Trans. Power Electron.*, vol. 37, no. 6, pp. 7315–7332, Jun. 2022.

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

[12] Q. Cui and Y. Weng, "Enhance high impedance fault detection and location accuracy via μ-PMUs," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 797–809, Jan. 2019.

[13] W. Qiu, Q. Tang, K. Zhu, W. Wang, Y. Liu, and W. Yao, "Detection of synchrophasor false data injection attack using feature interactive network," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 659–670, Jan. 2021.

[14] F. Li et al., "Detection and diagnosis of data integrity attacks in solar farms based on multilayer long short-term memory network," *IEEE Trans. Power Electron.*, vol. 36, no. 3, pp. 2495–2498, Mar. 2021.

[15] Q. Li, F. Li, J. Zhang, J. Ye, W. Song, and A. Mantooth, "Data-driven cyberattack detection for photovoltaic (PV) systems through analyzing micro-PMU data," in *Proc. IEEE Energy Convers. Congr. Expo. (ECCE)*, 2020, pp. 431–436.

[16] D. Saraswat, P. Bhattacharya, M. Zuhair, A. Verma, and A. Kumar, "Ansmart: A svm-based anomaly detection scheme via system profiling in smart grids," in *Proc. 2nd Int. Conf. Intell. Eng. Manage. (ICIEM)*, 2021, pp. 417–422.

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

[18] A. Barua, D. Muthirayan, P. P. Khargonekar, and M. A. Al Faruque, "Hierarchical temporal memory based machine learning for real-time, unsupervised anomaly detection in smart grid: WiP abstract," in *Proc. ACM/IEEE 11th Int. Conf. Cyber-Phys. Syst. (ICCPS)*, 2020, pp. 188–189.

[19] Q. Li, J. Zhang, J. Zhao, J. Ye, W. Song, and F. Li, "Adaptive hierarchical cyber attack detection and localization in active distribution systems," *IEEE Trans. Smart Grid*, vol. 13, no. 3, pp. 2369–2380, May 2022.

[20] J. E. Zhang, D. Wu, and B. Boulet, "Time series anomaly detection for smart grids: A survey," in *Proc. IEEE Electr. Power Energy Conf. (EPEC)*, 2021, pp. 125–130.

[21] L. Cui, Y. Qu, L. Gao, G. Xie, and S. Yu, "Detecting false data attacks using machine learning techniques in smart grid: A survey," *J. Netw. Comput. Appl.*, vol. 170, Nov. 2020, Art. no. 102808. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1084804520302769

[22] A. M. Saber, A. Youssef, D. Svetinovic, H. H. Zeineldin, and E. F. El-Saadany, "Anomaly-based detection of cyberattacks on line current differential relays," *IEEE Trans. Smart Grid*, vol. 13, no. 6, pp. 4787–4800, Nov. 2022.

[23] H. Karimipour, S. Geris, A. Dehghantanha, and H. Leung, "Intelligent anomaly detection for large-scale smart grids," in *Proc. IEEE Can. Conf. Electr. Comput. Eng. (CCECE)*, 2019, pp. 1–4.

[24] J. Yu et al., "Congo: Scalable online anomaly detection and localization in power electronics networks," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 13862–13875, Aug. 2022.

[25] D. Jung, K. Y. Ng, E. Frisk, and M. Krysander, "Combining model-based diagnosis and data-driven anomaly classifiers for fault isolation," *Control Eng. Pract.*, vol. 80, pp. 146–156, Nov. 2018.

[26] R. Punmiya, O. Zyabkina, S. Choe, and J. Meyer, "Anomaly detection in power quality measurements using proximity-based unsupervised machine learning techniques," in *Proc. Electr. Power Quality Supply Rel. Conf. (PQ) Symp. Electr. Eng. Mechatron. (SEEM)*, 2019, pp. 1–6.

[27] A. Barua and M. A. Al Faruque, "Hall spoofing: A non-invasive DoS attack on grid-tied solar inverter," in *Proc. 29th {USENIX} Secur. Symp. ({USENIX} Secur. 20)*, 2020, pp. 1273–1290.

[28] N. R. Gajanur, M. D. R. Greidanus, S. K. Mazumder, and M. A. Abbaszada, "Impact and mitigation of high-frequency side-channel noise intrusion on the low-frequency performance of an inverter," *IEEE Trans. Power Electron.*, vol. 37, no. 10, pp. 11481–11485, Oct. 2022

[29] J. Zhang, L. Guo, and J. Ye, "Cyber-attack detection for photovoltaic farms based on power-electronics-enabled harmonic state space modeling," *IEEE Trans. Smart Grid*, vol. 13, no. 5, pp. 3929–3942, Sep. 2022.

[30] C. Happ and S. Greven, "Multivariate functional principal component analysis for data observed on different (dimensional) domains," *J. Amer. Stat. Assoc.*, vol. 113, no. 522, pp. 649–659, 2018.

[31] D. Justin, R. S. Concepcion, H. A. Calinao, S. C. Lauguico, E. P. Dadios, and R. R. P. Vicerra, "Application of ensemble learning with mean shift clustering for output profile classification and anomaly detection in energy production of grid-tied photovoltaic system," in *Proc. 12th Int. Conf. Inf. Technol. Electr. Eng. (ICITEE)*, 2020, pp. 286–291.

[32] J. K. Lindeløv et al., "mcp: An R package for regression with multiple change points," *OSF Preprints*, vol. 10, 2020.

[33] A. Schmutz, J. Jacques, C. Bouveyron, L. Cheze, and P. Martin, "Clustering multivariate functional data in group-specific functional subspaces," *Comput. Stat.*, vol. 35, no. 3, pp. 1101–1131, 2020.

[34] D. C. Montgomery, *Introduction to Statistical Quality Control*. Hoboken, NJ, USA: Wiley, 2020.

[35] A. Haq, M. B. C. Khoo, M. Ha Lee, and S. A. Abbasi, "Enhanced adaptive multivariate EWMA and CUSUM charts for process mean," *J. Stat. Comput. Simul.*, vol. 91, no. 12, pp. 2361–2382, Mar. 2021.