A Deep Learning-based Solution for Securing the Power Grid against Load Altering Threats by IoT-enabled Devices

Hamidreza Jahangir, Subhash Lakshminarayana, Carsten Maple, and Gregory Epiphaniou

Abstract—**The growing integration of high-wattage Internet-of-Things (IoT)-enabled electrical appliances at the consumer end has created a new attack surface that an adversary can exploit to disrupt power grid operations. Specifically, dynamic load-altering attacks (D-LAAs), accomplished by an abrupt or strategic manipulation of a large number of consumer appliances in a botnet-type attack, have been recognized as major threats that can potentially destabilize power grid control loops. This paper introduces a novel approach based a multi-output network (two-dimensional convolutional neural networks classifier and reconstruction decoder) — called "2DR-CNN"— to detect and localize D-LAAs with high resolution. To achieve this, we leverage the frequency and phase angle data of the generator buses monitored by phasor measurement units (PMUs) installed in the power grid. To verify the effectiveness of the proposed method, simulations are conducted on IEEE 14- and 39-bus systems. The performance of the 2DR-CNN method is compared against several benchmark machine learning-based approaches. The results confirm that the proposed method outperforms other techniques in detection and localizing D-LAAs with high resolution in a number of practical scenarios, including PMU measurement noises and missing measurements.**

Index Terms—**cyber physical systems, dynamic load altering attack, Internet of Things, cyber security, deep learning.**

I. Introduction

THE proliferation of Internet-of-Things (IoT) enabled high-wattage electrical appliances (e.g., air conditioners, plug-in-electric vehicles) has created a new attack surface to target power grids. Although IoT intends to increase operating efficiency and provide convenience to end-users, they are an effective entry point for adversaries to intrude into the power grid and thus creates cyber security risks [1]. In particular, the adversary can disrupt power grid operation by manipulating the electrical demand via a large-scale Botnet-type attack against IoT-based electrical appliances [2]. Thus, it is of critical importance to devise defense strategies that can secure the power grid against these so-called *load-altering attacks* (LAAs).

H. Jahangir (Corresponding author) and S. Lakshminarayana are with the School of Engineering, University of Warwick, UK. C. Maple and G. Epiphaniou are with the Warwick Manufacturing Group, University of Warwick, UK.

E-mails: (Hamidreza.Jahangir, Subhash.Lakshminarayana, CM, Gregory.Epiphaniou)@warwick.ac.uk.

This work has been supported by the PETRAS National Centre of Excellence for IoT Systems Cybersecurity, which has been funded by the UK EPSRC under grant number EP/S035362/1.

The stability of the power grid is determined, in large part, by the balance between generation and demand. When an adversary abruptly changes a large amount of demand by manipulating high-wattage IoT-enabled devices, it can upset this balance. Existing works on this topic can be broadly divided into (i) attack impact analysis, and (ii) attack detection and localization.

Attack Impact Analysis: Prior research has shown that LAAs can lead to severe consequences such as causing unsafe frequency excursions, and/or line outages [2]. Power grid emergency actions, such as, under-frequency load shedding, can limit the damage to some extent. However, LAAs can still cause a partition of the bulk power system and controlled load shedding [3]. The so-called *dynamic* LAAs (D-LAAs), in which the attacker manipulates the system load by monitoring the grid's frequency dynamics, can potentially destabilize the power grid's frequency control loop [4]. An analytical framework to analyze static and D-LAAs using the theory of second-order dynamical systems was proposed in [5] to identify the victim nodes from which an attacker can launch the most impactful attacks. Recent works [6], [7] have also investigated how different loading and renewable energy penetration conditions impact the grid's vulnerability to D-LAAs. In particular, low-inertia conditions due to high renewable energy penetration can exacerbate this problem [7].

Detection and localization: The focus of this work is on developing a fast and robust tool to detect and localize D-LAAs with high precision, which is an essential step toward limiting the damage due to D-LAAs. To this end, existing works have proposed utilizing the phase angle/frequency dynamics monitored by phasor measurement units (PMUs) [8]–[10]. Similar approaches using PMU measurement data are also adopted in the broader context of detecting cyber-physical attacks against wide-area monitoring systems [11], [12], localizing the source of faults/oscillations [13]–[15] and disturbance type classification [16]. These approaches can be broadly divided into methods that explicitly use the underlying power system model in their formulation [8], [10], [14] and purely data-driven methods [13], [15], [16].

Approaches that use the power grid model in their formulation are specific to a linear model and cannot be extended to a non-linear model in a straightforward manner [8], [11], [14]. Moreover, due to the computational complexity associated with the algorithms, the localization can be only performed up to the zone of operation, but not identify the exact bus from which the attack/fault originates [8], [12], [14]. Alternatively, an unscented Kalman filter approach was proposed in [9]. However, this method does not scale well as the number of attack nodes increase. Reference [10] proposed a D-LAA localization method using physics-informed machine learning approaches, which can be applied to non-linear models. While the proposed algorithms achieve good performance in specific scenarios, the implementation difficulties associated with combining the power grid's operation constraints with machine learning techniques' training loss limit the application of these algorithms (e.g., the physics-informed neural network algorithm does not perform well in systems that exhibit slow oscillatory dynamics).

Regarding the application of pure data-driven methods, the authors of reference [17] have developed a data-driven hierarchical monitoring framework that can detect and localize cyber threats in distribution power systems with sparse monitoring sensors. This study is shown to provide accurate localization at the distribution system level by monitoring the voltage and current of the selected buses using waveform measurement devices. However, at the power grid transmission level, the frequency and phase angle profiles are the two most essential factors in stability studies [18]. The authors of [13] proposed a data-driven strategy to detect and localize forced oscillations using robust principal component analysis. This technique produces sufficient outcomes, but is contingent on the operators' selection of internal parameters. This constrains the approach's applicability in online applications such as localizing D-LAAs in which the operator cannot have prior knowledge of the actual parameters used by the attacker.

Finally, one-dimensional convolution neural networks (1D-CNNs) have been used in the context of fault localization/classification [15], [16]. However, as we show in this work, a multi-output network based on two-dimensional convolutional neural networks classifier and reconstruction decoder — here after called "2DR-CNN" — achieves superior feature extraction and classification performance, especially in the proposed context in which we have two different physical signals (i.e., phase angle and frequency data) for localizing the attack. In addition, the 2DR-CNN is resistant to PMU noise and data quality concerns such as missing/outlier data points because of the parallel integration of reconstruction decoder and the classification process.

To summarize, we note that in the specific case of LAAs, we require a high-resolution attack localization method that can be generalized to linear/non-linear grid models, and that is robust to PMU noise/data quality issues.

To address the aforementioned shortcomings, this paper proposes a deep learning-based solution named as 2DR-CNN. CNN is a robust tool for high-quality feature extraction in high-dimensional cases (e.g., image processing [19]). In the context of this work, we have two features in this study: phase angle and frequency of the generator buses. By implementing 2D kernels, we simultaneously consider both these signals in the feature extraction task. This 2D feature extraction helps us engage the interaction between the phase angle and frequency of the generator buses, which are interdependent parameters in power grid swing equations [20]. To implement a 2D kernel (moving in two directions), the input data set is transformed into an image structure, i.e., two-dimensional data with channels. (Further details are provided in Section III-A.) Furthermore, we only utilize the PMU measurements from the generator buses to infer the attack parameters. This is particularly relevant in our scenario, since the measurements corresponding to the load buses cannot be trusted, e.g., a sophisticated attacker may launch a coordinated false data injection attack on the load-bus measurements to hide their attack. Thus, the proposed method is intrinsically secure against such attacks. In addition to the above mentioned details, being a purely data-driven method, the proposed approach is not restricted to the linear system model and can be generalized to non-linear models. The main contributions of this paper can be summarized as follows:

- Designing a data-driven framework that relies solely on generator bus profiles (phase angle and frequency) to precisely detect and locate D-LAAs with high resolution – If we install PMUs on load buses, they may be vulnerable to fault data injection attacks, which will also raise the cost of monitoring.
- Assessing a narrow observation window —– just 2 seconds of the explored profiles to be in accordance with the power grid relays' response time under IEEE Standard 1547 [21] – to detect and localize the D-LAAs with the help of implementing a 2D feature extraction environment (2DR-CNN) that concurrently reads both the phase angle and the frequency of generator buses in the feature extraction task.
- Securing the proposed method in real-world scenarios involving noisy PMU data as well as missing and outlier points by considering robust feature extraction layers — designing a reconstruction decoder network equipped with sparsity regularization parallel with the classification layer that prevents the loss of spatiotemporal features in backward gradient descent via a bypass way.

In summary, this study presents a novel data-driven solution using a two-dimensional CNN structure that is capable of detecting and localizing LAAs by analyzing only two seconds of the power grid dynamic profiles, including the frequency and phase angle of the generator buses. In addition, the proposed method is provided with a reconstruction network that enables it to deliver robust performance in noisy environments and in the presence of real-world obstacles like missing and outlier points in the observed profiles. Finally, we note that this study introduces a novel security solution for the power grids based on monitoring the physical parameters of the power grid, such as, frequency and phase angles of the power grid dynamics, and adds as an additional layer of security (in addition to the cryptographic security and those that monitor the network traffic data), which enhances the power grid's protection against malicious behavior such as LAAs.

The focus of this work is on monitoring the power grid from the perspective of a transmission system operator (TSO). In particular, we focus on transmission network oscillations, with the goal of localising the "substation(s)" from which the LAAs originate. We assume the power grid that the TSO has an accurate knowledge of the grid topology, which is a reasonable assumption for transmission networks and the deployed PMUs can monitor the dynamics of the power grid in real-time [22].

The remaining parts of this paper are organized as follows: a short description of the LAAs problem is provided in section II. Section III introduces our solution to address the detection and localization of the D-LAAs with 2DR-CNN. In section IV, the simulation framework and results are presented. Finally, Section V concludes the findings of this study.

II. Problem Statement

A. Power Grid Model Under LAAs

We consider a generic power grid consisting of N buses connected by M transmission lines. We denote the set of buses by $\mathcal{N}$, which can be divided as $\mathcal{N} = \mathcal{N}_G \cup \mathcal{N}_L$, where $\mathcal{N}_G$ and $\mathcal{N}_L$ denote the set of generator and load buses respectively. Let $N_G = |\mathcal{N}_G|$ and $N_L = |\mathcal{N}_L|$. The power grid dynamics under LAAs are given by the following set of differential equations [4]:

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \dot{\boldsymbol{\delta}} \\ \dot{\boldsymbol{\omega}} \\ \dot{\boldsymbol{\theta}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{p}^{LS} + \boldsymbol{\epsilon}^L \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{K}^I + \mathbf{B}^{GG} & \mathbf{K}^P + \mathbf{D}^G & \mathbf{B}^{GL} \\ \mathbf{B}^{LG} & -\mathbf{K}^L & \mathbf{B}^{LL} \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta} \\ \boldsymbol{\omega} \\ \boldsymbol{\theta} \end{bmatrix}, \tag{1}$$

where $\boldsymbol{\delta} \in \mathbb{R}^{N_G}$ and $\boldsymbol{\theta} \in \mathbb{R}^{N_L}$ are the phase angle of the generator and the load buses respectively, and $\boldsymbol{\omega} \in \mathbb{R}^{N_G}$ are the generator bus frequencies. The matrices $\mathbf{M}, \mathbf{D}^G, \mathbf{K}^I, \mathbf{K}^P \in \mathbb{R}^{N_G \times N_G}$ are diagonal matrices whose diagonal entries are the generator inertia, damping, proportional, and integral coefficients, respectively. Matrices $\mathbf{B}^{GG} \in \mathbb{R}^{N_G \times N_G}, \mathbf{B}^{LL} \in \mathbb{R}^{N_L \times N_L}, \mathbf{B}^{GL} \in \mathbb{R}^{N_G \times N_L}$ are sub-matrices of the admittance matrix, derived as $\mathbf{B}_{bus} = \begin{bmatrix} \mathbf{B}^{GG} & \mathbf{B}^{GL} \\ \mathbf{B}^{LG} & \mathbf{B}^{LL} \end{bmatrix}$. The vectors $\mathbf{p}^{LS}, \boldsymbol{\epsilon}^L \in \mathbb{R}^{N_L}$ along with the matrix $\mathbf{K}^L$ model the system load under LAAs (more details are presented in the following).

B. Description of LAAs

We assume that the total system load consists of two components, i.e., $\mathbf{p}^L = \mathbf{p}^{LS} + \mathbf{p}^{LV}$, where $\mathbf{p}^{LS}$ is the secure part of the system load (i.e., it includes non-smart and/or protected loads) and $\mathbf{p}^{LV}$ is the vulnerable portion of the load. Under LAAs, the net load of the system is given by

$$\mathbf{p}^L = \mathbf{p}^{LS} + \boldsymbol{\epsilon}^L - \mathbf{K}^L \boldsymbol{\omega}, \tag{2}$$

where $\boldsymbol{\epsilon}^L \in \mathbb{R}^L$ is the *static* LAA component and $-\mathbf{K}^L\boldsymbol{\omega}$ is the *dynamic* LAA (D-LAAs). The static LAA is a one-time load perturbation introduced by the attacker. On the other hand, the D-LAA is a time-varying load perturbation that follows the frequency fluctuations of the system. Note that to execute a D-LAA the attacker is required to monitor the frequency fluctuations of the system (i.e., $\boldsymbol{\omega}$), and adjust their load perturbations accordingly [4]. $\mathbf{K}^L \in \mathbb{R}^{N_L} \times \mathbb{R}^{N_G}$ denotes a matrix consisting of attack controller gain values (feedback coefficients multiplying the observed frequencies $\boldsymbol{\omega}$) which has the effect of increasing the frequency deviations from the setpoint. Note that the total load altered by the attacker is bounded by $\mathbf{p}^{LV}$ as follows [4]:

$$\mathbf{K}^L \omega \leq (P^{LV} - \epsilon^L)/2. \tag{3}$$

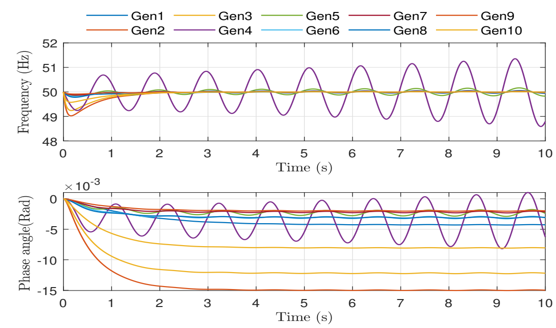

Fig. 1: Frequency and phase angle profiles of Generator buses in IEEE 39-bus system following the initiation of a D-LAA with $K_L(19, 4) = 24.02$ pu and $\epsilon_{19} = 0.8$ pu.

To illustrate the effect that D-LAAs have on the power grid operations, Fig.1 depicts the oscillation of the generator buses following the initiation of a D-LAA in the IEEE 39-bus system (attack parameters specified in the figure description). It can be observed that the frequency oscillations grow in an unbounded manner, eventually leading to system instability.

C. Power Grid Monitoring and Attack Localization Problem

We assume that the system operator has installed PMUs at generator buses, enabling them to monitor the phase angles $\{\delta_i^{(\tau)}\}_{i \in \mathcal{N}_G, \tau=1,\ldots,T}$ and frequency fluctuations $\{\dot{\delta}_i^{(\tau)}\}_{i \in \mathcal{N}_G, \tau=1,\ldots,T}$ over time (i.e., a sampled version of the signals shown in Fig.1). Herein, $x^{(\tau)}$ is the value of the signal x at time slot τ, the slots being sampled at a time interval of T_s, and T being the total number of time slots. IEEE/IEC standards specify that the sampling frequency of the PMU for a 50 Hz system can be between 10 and 100 frames per second. That is, T_s is between $10 - 100$ ms [23].

The objective of this work is to detect and localize the LAAs by monitoring the signals $\{\delta_i^{(\tau)}\}_{i \in \mathcal{N}_G, \tau=1,\ldots,T}$ and $\{\omega_i^{(\tau)}\}_{i \in \mathcal{N}_G, \tau=1,\ldots,T}$. In particular, we aim to detect the locations corresponding to non-zero values of the elements of $\boldsymbol{\epsilon}$ and $\mathbf{K}^L$ for destabilizing attacks (see Section III-B for more details).

III. Design of Attack Detection and localization Framework

This section discusses the proposed D-LAA detection and localization approach in detail. The proposed deep learning-based method comprises two phases – (i) offline training phase and (ii) online inference phase. In the offline training phase, we provide the CNN with extensive data samples under various normal operation/attack scenarios to train the classifier. In the online phase, the trained classifier is deployed at a power grid control center takes PMU signals as inputs and performs attack detection/localization.

A. Motivation for Using 2D Feature Extraction

As noted in Section I, power grid monitoring using PMUs provides the operator with two signals to exploit – the frequency and the phase angle of the generator buses. As we note in Fig.1, the frequency/phase angle dynamics under D-LAAs can rapidly grow in magnitude, leading to unsafe frequency excursions within a short time interval. Thus, attack detection/localization must be timely. This, in turn, requires precise feature extraction to accurately and quickly attribute the observed oscillations to the correct victim buses that are under the attacker's control. Thus, we reshape the input data into a 2D structure to concurrently integrate these two parameters in the feature extraction task. Our kernel travels in two directions as a result of this method, enabling us to effectively interpret the input data.

B. Data Generation to Train the Supervised Classifier

The training data for the 2DR-CNN corresponds to a sampled version of the frequency/phase angle data as shown in Fig.1. Online monitoring of generator buses using PMUs has been proven as an efficient way for assessing the transient stability of the power grid in studies on power system stability [18]. Accordingly, the data to train the machine learning (ML) models are arranged in matrices of size $N_G \times T \times 2$ (recall T corresponds to the number of time samples considered in the training data). For instance, considering a two-second observation window in Fig.1 with a PMU sampling frequency of 50 samples per second would result in $T = 100$. The two layers in the training data (corresponding to the third dimension) are the phase angle and frequency data, respectively.

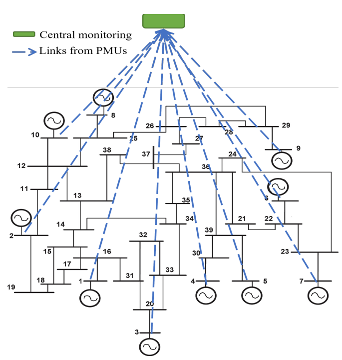

Fig. 2: Central monitoring of IEEE 39-bus system.

To train a supervised classifier for detecting and localizing D-LAAs, we must ensure that our training dataset completely covers all potential victim buses and attack magnitudes. To this end, we inject a variety of D-LAAs from diverse load buses and attack controller gains to generate the training datset. The samples that include LAAs are labelled with the location of the bus from which the attacker's inputs have been injected. We consider a number of different attacks, including single-point and multi-point attacks (i.e., simultaneous attacks at multiple nodes) to cover the different types of attacks that are likely to occur. Let N_{tr} denote the total number of distinct attacks generated in this process. Thus, the training data consists of N_{tr} matrices of size $N_G \times T \times 2$.

Differentiating D-LAAs from Normal System Operation: D-LAAs will have an effect on the stability of the power grid. In other words, for destabilizing D-LAAs, at least one of the eigenvalue defined in the state-space representation of the power grid will lie on the right-half plane [24]. Additionally, the attacker can also transfer the eigenvalues to a "zone of vulnerability", from which intrinsic disturbances associated with power grid operations can result in grid instability. We refer to such attacks as *semi-destabilizing* attacks, which we formalize in the following. First, we define the power grid eigenvalues in terms of damping ratio ζ and natural oscillation frequency ω_n: $a = -\zeta\omega_n, b = \omega_n\sqrt{1-\zeta^2}$ [25]. According to the North American Electric Reliability Corporation, the vulnerable plane is defined as: $\zeta \leq 3\%, 2.5 \leq \omega_n \leq 12.6rad/s$ [26]. Thus, any attacks that transfer the eigenvalues into this region can be defined as semi-destabilizing attacks. To illustrate the notion, Fig.3 illustrates the eigenvalue configuration for a semi-unstable D-LAA with D-LAA ($K_L(24,4) = 55.62$ pu), in which two eigenvalues are moved near to the right-half plane. To differentiate D-LAAs from normal power grid fluctuations, we examine the real part of the system's eigenvalues. If at least one of the eigenvalues lies in the right-half plane, or in the zone of vulnerability, then the corresponding data in the training set is labelled as "attack" data. Otherwise, the system is deemed to be under normal operation (or a benign attack that is unlikely to destabilize the system).

In order to provide a clear overview of the generated profiles, a concise description of the overall data generation procedure for this study is provided below[1] (in accordance with the framework presented in reference [27]):

- *Step 1:* Obtaining power grid topological data from the MATPOWER simulator and power grid dynamic parameters from Appendix section for the understudied IEEE case to initialize the general structure of the Equation (1).
- *Step 2:* Getting a random value for the attack controller gain matrix $\mathbf{K}^L \in \mathbb{R}^{N_L} \times \mathbb{R}^{N_G}$ while considering the constraints in Equation (3).
- *Step 3:* Checking the eigenvalues of the power grid according to the detailed formulation presented in reference [5].
- *Step 4:* Evaluating the impact of the launched attack: (i) attack: if an eigenvalue is in the unstable plane ($a = -\zeta\omega_n > 0$) or semi-unstable plane $\zeta \leq 3\%, 2.5 \leq \omega_n \leq 12.6rad/s$. (ii) normal fluctuation: if neither the semi-unstable nor the unstable plane contain any eigenvalues.
- *Step 5:* Storing the frequency and phase angle profiles of the generator buses for two seconds (100 samples, 50 samples per second) in both attack and normal fluctuation scenarios.
- *Step 6:* Storing the location of the launched attack according to the $\mathbf{K}^L \in \mathbb{R}^{N_L} \times \mathbb{R}^{N_G}$ values for the attack profiles – location is determined based on the position of the load bus in the attack controller gain matrix ($\mathbb{R}^{N_L}$).
- *Step 7:* Repeating the mentioned steps to cover a variety of possible scenarios (we considered 3000 samples for

[1]The generated data set for the numerical results of this study are available at https://github.com/omiddeeplearning/IEEE-IoTJ-LAAs.git.

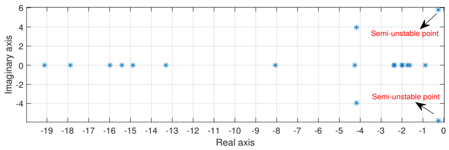

Fig. 3: Eigenvalues of IEEE 39-bus system for $K_L(24, 4) = 55.62$ pu.

each of the under-studied IEEE cases, which present accurate results in our tests).

C. Designing 2DR-CNN for Detection and localization of D-LAAs

The overall structure of the proposed method is outlined in **Algorithm 1**. Following the algorithm description, we note that the proposed method consists of two classification networks: the first one detects attacks, while the second one localizes the primary source of attack with high resolution (that is, identifies the bus from which the attack was launched). As discussed in Sections III-A and III-B, our detection and localization tasks require high-dimensional input data. To deal with this vast volume of data, we utilize a 2DR-CNN, a deep network tailored for both noise-free and noisy data environments. Since CNNs are designed primarily for image processing applications, we view the input data as an image with dimensions $N_G \times T \times 2$, which is effectively a picture with dimensions $N_G \times T$ and 2 channels. To facilitate the handling of our input data with the CNN structure, the initial convolution layer of the 2DR-CNN network converts the input data into a square image. To accomplish this goal, we use 512 filters with $1 \times T/N_G$ size by $1 \times T/N_G$ stride value to convert our original rectangular image of dimensions $N_G \times T \times 2$ into a square image of dimensions $N_G \times N_G \times 512$. Following this, different layers are applied to the input data during feature extraction in both the detection and localization networks. The detailed architecture of the 2DR-CNN developed to detect and localize tasks in the IEEE 14- and 39-bus systems are shown in Table I. As illustrated, the dimension of the output layer for the detection and localization networks is set to 2 (binary classification: attack/non-attack) and $|\mathcal{N}_L|$ (multi-classification: location of the attacked bus, where, for the IEEE 14-bus case $|\mathcal{N}_L| = 9$ and for the IEEE 39-bus case $|\mathcal{N}_L| = 29$), respectively. As outlined in Table I, the reconstruction network is introduced immediately after the flattening layer, which is comprised of three dense layers. In order to prevent overfitting concerns in the reconstruction network, the first dense layer — the downsampling layer — is equipped with a sparsity regularization constraint (10^{-5}). In this configuration, we have a network with two outputs: detection/localization and reconstruction. The loss function is therefore defined as follows:

$$Loss = Loss_{Det/Loc} + \lambda \ Loss_{Rec} \tag{4}$$

where λ is defined as 0.0005. Adding the reconstruction network is primarily intended to prevent the loss of spatiotemporal features due to pooling layers (before the flattening layer, shown in Fig. 4) during feedforward flow of data. The reconstruction network parallel to the classification network is used to recover the removed features from the original input data during backward flow of data (gradient descent). In this structure, it is important to consider the low coefficient for reconstruction loss in order to ensure the stability of the overall structure (avoiding convergence issues and misleading gradients due to two sources of errors – classification and reconstruction networks). The reconstruction network is only responsible for fine-tuning the training task by providing a by-pass channel to recover the eliminated features by the polling layers during the training procedure of the main network's parameters (W_{Conv}), as shown below:

$$\frac{\partial Loss}{\partial W_{Conv}} = \frac{\partial Loss}{\partial Loss_{Det/Loc}} \times \frac{\partial Loss_{Det/Loc}}{\partial W_{Conv}} + \lambda \ \frac{\partial Loss}{\partial Loss_{Rec}} \times \frac{\partial Loss_{Rec}}{\partial W_{Conv}} \tag{5}$$

To provide a better visual representation of the designed 2DR-CNN network, Fig. 4 illustrates the structure of the developed networks (detection/localization classifiers and reconstruction network) for the IEEE 39-bus system. The overall structure of the proposed networks (i.e., number of CNN and pooling layers, size of filters (kernels), stride values, etc.) is specifically designed to deliver robust and stable feature extraction procedures based on the size of the input data (monitoring profiles from PMU units) and target classes (number of load buses that might be used for LAA launch) of the understudied cases. For instance, the IEEE 39-bus system, which has more load buses (candidate targets for our localization work) than the IEEE 14-bus system, includes more hidden layers and larger stride values in the initial layers due to their larger input data size. According to the numerical tests, the structures proposed had the most efficient performance compared to other possible structures.

Algorithm 1 Detection and localization Algorithm.

Input: Phase angle data $\{\delta_i^{(\tau)}\}_{i \in \mathcal{N}_G, \tau=1,\dots,T}$ and frequency data $\{\omega_i^{(\tau)}\}_{i \in \mathcal{N}_G, \tau=1,\dots,T}$, where, for the IEEE 14-bus system, $\mathcal{N}_G = \{1, 2, 3, 6, 8\}, |\mathcal{N}_G| = 5$ for IEEE 39-bus case, $\mathcal{N}_G = \{30, \dots, 39\}, |\mathcal{N}_G| = 10$.
Output: Detect the D-LAAs and localize the attacked bus ($i \in \mathcal{N}_L$), where, for IEEE 14-bus case $|\mathcal{N}_L| = 9$ and for IEEE 39-bus case $|\mathcal{N}_L| = 29$.

```
Monitor phase angle {δ_i^(τ)}_{i∈N_G,τ=1,...,T} and frequency {ω_i^(τ)}_{i∈N_G,τ=1,...,T} data.
Discriminate between the regular fluctuations of the power grid and D-LAAs using the first binary classifier.
Output of first classifier: D-LAAs attacks or regular fluctuations.
if first classifier detects D-LAAs attacks then
    Localize the victim buses using the second multi-class classifier.
    Output of second classifier: location of the attacked bus ( i ∈ N_L)
else
    first classifier confirms regular fluctuations and goes to monitor the next time step
end if
```

TABLE I: Detailed structure of 2DR-CNN for IEEE 14- and 39-bus systems.

Operation Layer		Number of Filters		Size of Each Filter		Stride Value		Output Data Detection/Localization	Output Data Reconstruction	Output Data Detection/Localization	Output Data Reconstruction
		IEEE 14-bus	IEEE 39-bus	IEEE 14-bus	IEEE 39-bus	IEEE 14-bus	IEEE 39-bus	IEEE 14-bus		IEEE 39-bus	
Input Data		–	–	–	–	–	–	$5 \times 100 \times 2$		$10 \times 100 \times 2$	
Convolution Layer	Convolution	512	512	1×20	1×10	1×20	1×10	$5 \times 5 \times 512$		$10 \times 100 \times 512$	
	ReLU	–	–	–	–	–	–	$5 \times 5 \times 512$		$10 \times 10 \times 512$	
Pooling Layer	Max pooling	1	1	2×2	2×2	1×1	2×2	$4 \times 4 \times 512$		$5 \times 5 \times 512$	
Dropout Layer	Dropout (0.5)	1	1	–	–	–	–	$4 \times 4 \times 512$		$5 \times 5 \times 512$	
Convolutional Layer	Convolutional	256	256	2×2	2×2	1×1	1×1	$3 \times 3 \times 256$		$4 \times 4 \times 256$	
	ReLU	–	–	–	–	–	–	$3 \times 3 \times 256$		$4 \times 4 \times 256$	
Pooling Layer	Max pooling	1	1	2×2	2×2	1×1	2×2	$2 \times 2 \times 256$		$2 \times 2 \times 256$	
Dropout Layer	Dropout (0.5)	1	1	–	–	–	–	$2 \times 2 \times 256$		$2 \times 2 \times 256$	
Flatten Layer	Flat	1	1	–	–	–	–	1024		1024	
Dense Layer	Fully connected	128	128	–	–	–	–	128	256	128	512
	ReLU	–	–	–	–	–	–	128	256	128	512
Dense Layer	Fully connected	1	1	–	–	–	–	2(detection)/9(localization)	512	2(detection)/29(localization)	1024
	Softmax/ReLU	–	–	–	–	–	–	2(detection)/9(localization)	512	2(detection)/29(localization)	1024
Dense Layer	Fully connected	1	1	–	–	–	–	–	1000	–	2000
	Reshape	–	–	–	–	–	–	–	$5 \times 100 \times 2$	–	$10 \times 100 \times 2$

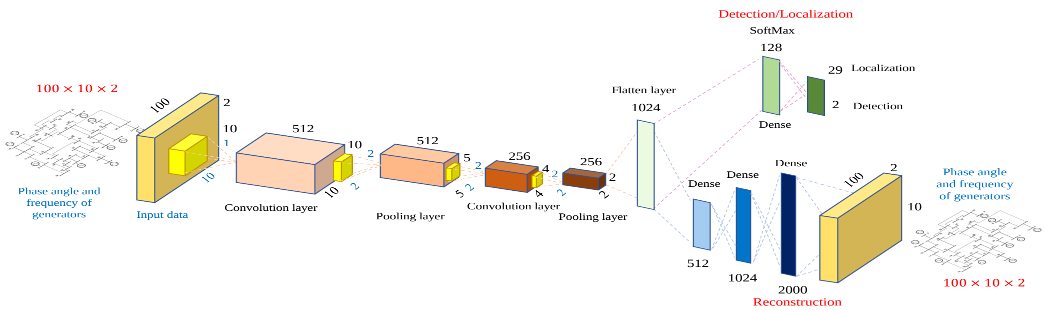

Fig. 4: Overall structure of 2DR-CNN for IEEE 39-bus system.

IV. Simulations

A. Simulation Settings

Two test cases are included to verify the effectiveness of the proposed method, including IEEE 14- and 39-bus systems. The MATPOWER simulator contains the topological data for the IEEE cases. The Appendix includes further information on the dynamic simulation parameters. To evaluate the suggested method's efficiency against various D-LAAs, the simulation results include single-point (static and dynamic attacks launched from the same location) and multi-point (static and dynamic attacks launched from multiple locations) D-LAAs. Finally, to evaluate the suggested method's performance in real-world application scenarios, a section on applying appropriate PMU noise and data-quality issues is included at the end of this section. In all the detection/localization tasks, we consider data from a two-second interval following the attack sampled at 50-times per second. Thus, T is set at 100.

To demonstrate the superiority of the proposed 2DR-CNN method, it is compared to several benchmark techniques deployed in supervised classification tasks, including 1D-CNN , auto-encoder-based deep neural network (AE-DNN), DNN support vector machines (SVM)Ṫhese benchmark algorithms are also implemented in a cascaded manner, similar to 2DR-CNN, with the first classifier detecting attacks and the second classifier localizing them. To run the simulation results, a wide library of $N_{\text{tr}} = 3000$ samples for each of the under-studied scenarios is established. The training, validation, and test data sets are partitioned by 80%, 10%, and 10%, respectively.

The proposed method is implemented in Python 3.9 using the TensorFlow framework. The categorical cross-entropy loss function is considered, and the Adam optimizer method is applied to perform the training process. Fig. 5 illustrates the overall training loss and accuracy for validation and test data sets. The simulations are conducted on a Windows PC with 11th Gen Intel(R) Core(TM) i7-1185G7 @ 3.00GHz processor, RAM: 16 GB. Offline training procedures for different algorithms vary significantly, and more intricate algorithms such as 2DR-CNN and 1D-CNN require more time than SVM and DNN. The crucial element, which is the online execution time, is less than $89ms$ for the proposed method (2DR-CNN) (the average value for all three cases, including IEEE 14-, 39- and 57-bus systems), making it suitable for online applications. Table II provides a thorough comparison of the response times of all adopted benchmark techniques.

TABLE II: Online response time of different approaches

Method	Online response time (msec)
2DR-CNN	89
1D-CNN	69
MLP	66
SVM	61

In this study, the accuracy for both detection and localization tasks is calculated as follows [28]:

$$accuracy = \frac{1}{T_n} \sum_{i=1}^{T_n} \frac{C^i}{Q^i} \tag{6}$$

where, C^i and Q^i represent the number of successfully categorized samples and query samples in the i_{th} test series, respectively, and T_n is the total number of test series.

B. Simulation Results and Discussion

Detection of D-LAAs: As indicated in **Algorithm 1**, the first classifier distinguishes D-LAAs from regular power grid

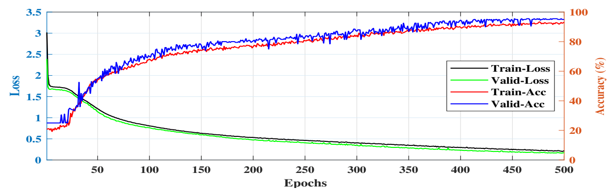

Fig. 5: Training plot of 2DR-CNN.

TABLE III: Detection results of D-LAAs with different approaches.

Methods	IEEE 14-Bus	IEEE39-Bus
2DR-CNN	99.66%	99.27%
1D-CNN	97.49%	96.54%
AE-DNN	96.36%	96.11%
DNN	96.44%	95.37%
SVM	95.13%	91.34%%

oscillations. Table III summarizes the accuracy of the detection compared to the other benchmarks. The detection phase, particularly for the IEEE 14-bus system, is not difficult for any of the benchmark techniques, and the accuracy ranges from 95.13% to 99.66% for SVM and 2DR-CNN algorithms, respectively. However, for the larger of the two systems, i.e., the IEEE 39-bus system, the accuracy of SVM decreases to 91.34%, and other benchmarks also see a minor decrease (in the region of 1%). In comparison, 2DR-CNN once again achieves greater than 99% accuracy. Overall, we observe that most of the benchmark techniques have acceptable performance in differentiating D-LAAs from normal power grid operations, with the 2DR-CNN performing the best.

Localization of single-point and multi-point D-LAAs: As stated in **Algorithm 1**, once the first classifier detects an attack, the second classifier begins high-resolution localization of the source of the attack. Table IV summarizes the performance of the 2DR-CNN and other benchmark methods for single- and multi-point attacks. As expected, due to the increased number of classes in the localization task (i.e., multi-task classification as compared to binary classification in the detection task), the accuracy of classification in this step drops slightly, in both single- and multi-point attack scenarios. However, we observe that 2DR-CNN surpasses the other benchmark techniques in both IEEE 14- and IEEE 39-bus scenarios. In particular, the 2DR-CNN step achieves 99.02% and 98.07% accuracy for single- and multi-point attacks on the IEEE 14-bus system, respectively; and 95.33% and 93.57% accuracy for the IEEE 39-bus system. Additionally, the results demonstrate that although other benchmark approaches perform satisfactorily in both single and multi-point assaults on the IEEE 14-bus system, SVM attains poor performance in both single- and multi-point attacks in the IEEE 39-bus system, with an accuracy of 70.15% and 69.93% percent, respectively. Other benchmarks such as 1D-CNN, AE-DNN, and DNN, obtain an accuracy of less than 90%. When we delve deeper, we observe that the distinction between single- and multi- point attacks on IEEE 39-bus is more pronounced, as the performance gap between the algorithms is widened for multi-point attacks in comparison to the single-point attacks. To support this claim, Fig.6 depicts the frequency profiles of generator buses during single- and multi-point attacks. Due to the fact that multi-

TABLE IV: localization results of D-LAAs with different approaches.

Methods	IEEE 14-Bus		IEEE 39-Bus	
	Single-point attacks	Multi-point attacks	Single-point attacks	Multi-point attacks
2DR-CNN	99.02%	98.07%	95.53%	93.57%
1D-CNN	97.17%	96.55%	89.76%	87.59%
AE-DNN	97.69%	96.63%	88.09%	86.81%
DNN	97.23%	96.29%	85.21%	81.23%
SVM	91.14%	89.36%%	70.15%	69.93%

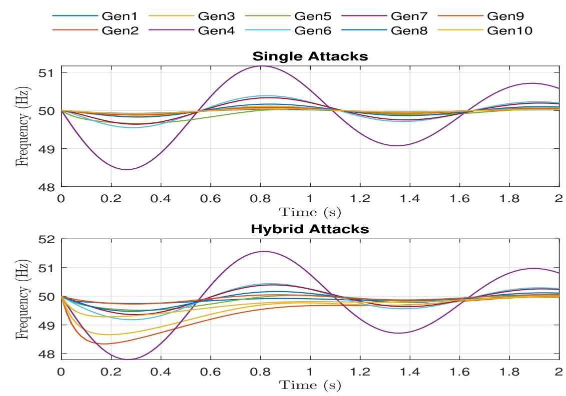

Fig. 6: Frequency profiles IEEE 39-bus after attacks: $K_L(24, 4) = 55.62pu$ for single-point attack ϵ_{19} =1MW and for multi-point attack $\epsilon_5 = 3$MW, $\epsilon_6 = 2$MW and $\epsilon_{19} = 1$MW.

point attacks include at least two sources of static attacks in addition to the D-LAAs, the frequency profiles of multi-point attacks exhibit more volatility, which makes it harder for the algorithms to perform the classification task.

The comparison of 2DR-CNN and 1D-CNN demonstrates the paper's central idea, which is to convert the input data into an image-like structure and implement 2DR-CNN for feature extraction. In the 2DR-CNN, the interconnection between the frequency and phase angle profiles is considered directly during the feature extraction task (by moving the kernel in two directions). This results in a difference of 6% between the accuracies of 2DR-CNN and 1D-CNN in localizing DLAAs for the IEEE 39-bus system. Following that, we observe AE-DNN with AE at the initial layer outperforms DNN by approximately 3% and 5% in single and multi-point attacks, respectively. However, we do not observe a major difference in the performance between AE-DNN and DNN for the IEEE 14-bus system, since AE serves as a denoising element for the AE-DNN and there is no substantial fluctuation in the input data profiles in small cases. To illustrate the performance of 2DR-CNN in detail, Fig. 7 and Fig. 8 exhibit the localization outcomes as confusion matrices for single- and multi-point attacks, respectively. As previously noted, the localization task is more challenging with multi-point attacks. Comparing the localization performances of attacks launched from buses 8 and 16 in single- and multi-point attacks elucidates this point.

Localization of D-LAAs in noisy environments: To assess the proposed method's robustness in the presence of noisy PMU measurements, this section examines the localization performance for IEEE 14- and 39-bus systems under three different noise scenarios. As suggested in reference [16], Gaussian noise can be used to model the PMU noise. Table V

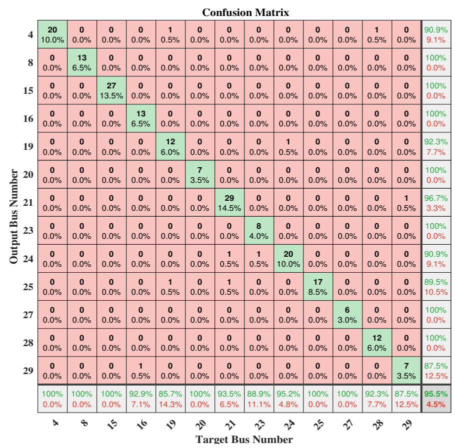

Fig. 7: Confusion matrix of 2DR-CNN single-point attacks.

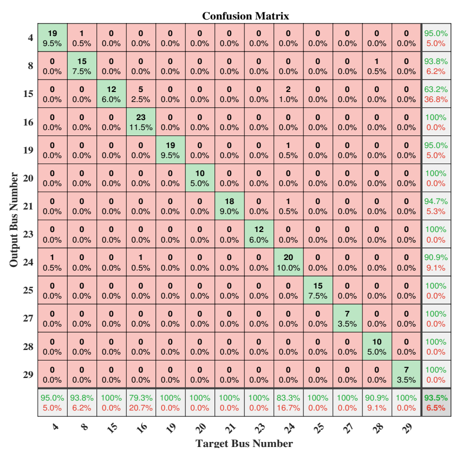

Fig. 8: Confusion matrix of 2DR-CNN multi-point attacks.

compares the performance of the 2DR-CNN to that of other benchmark methods. The numerical results demonstrate the proposed method's robustness with noisy data. For instance, for multi-point attacks in the IEEE 39-bus system, 2DR-CNN provides 92.51%, 91.21%, and 90.04% accuracies for 26, 20, and 16.5 dB SNR ($20\log(signal/noise)$ [29]) values, respectively. An inspection of Table V demonstrates that the proposed method achieves 95.04% and 92.52% accuracy in the localization of single- and multi-point attacks on the IEEE 39-bus case with the 26 dB SNR value, respectively, while other benchmarks achieve less than 87.92% and 86.56% accuracy. Moreover, we observe that AE-DNN delivers a similar performance to that of 1D-CNN. This is because of the additional AE in this structure, which aids in noisy measurements (about 2% gain against DNN). The 2DR-CNN has robust feature extraction that outperforms the AE-DNN's additional AE layer, and as a result, the AE layer is not considered for CNN-based networks. As expected, raising the noise level (i.e., decreasing the SNR values) degrades the accuracy of 2DR-CNN and other benchmarks, although 2DR-CNN is significantly more robust than other methods – implementing reconstruction network along with the 2D feature extraction –, achieving approximately 90% accuracy in localization of various D-LAAs at a 16.5 dB SNR rate.

TABLE V: localization results of D-LAAs with different approaches by considering the PMU Noise.

Methods	IEEE 14-Bus		IEEE 39-Bus	
26 DB SNR (noise variance 5%)				
	Single-point attacks	Multi-point attacks	Single-point attacks	Multi-point attacks
2DR-CNN	97.01%	96.14%	95.04%	92.51%
1D-CNN	96.74%	95.11%	87.93%	86.56%
AE-DNN	96.75%	94.81%	88.09%	85.98%
DNN	96.41%	94.66%	87.33%	81.13%
SVM	91.07%	89.31%	62.91%	59.78%
20 DB SNR (noise variance 10%)				
	Single-point attacks	Multi-point attacks	Single-point attacks	Multi-point attacks
2DR-CNN	96.07%	93.13%	92.28%	91.21%
1D-CNN	96.08%	90.23%	86.15%	84.95%
AE-DNN	96.73%	91.38%	84.49%	82.78%
DNN	95.13%	90.61%	82.33%	79.43%
SVM	90.39%	85.43%	60.44%	58.12%
16.5 DB SNR (noise variance 15%)				
	Single-point attacks	Multi-point attacks	Single-point attacks	Multi-point attacks
2DR-CNN	93.41%	91.92%	91.14%	90.04%
1D-CNN	89.13%	85.98%	83.59%	85.98%
AE-DNN	89.77%	84.49%	82.74%	84.46%
DNN	88.71%	81.31%	80.66%	81.31%
SVM	87.84%	83.97%	54.79%	53.97%

Verifying the robustness of the proposed method with PMU data-quality issues: Finally, we evaluate the proposed method's effectiveness towards other practical challenges, such as missing and outlier points in the PMU data. We consider two possible scenarios: (i) 5% of randomly-selected data points are missing (over the 2-second time horizon). (ii) 5%-10% of the randomly-selected data points contain outliers. As shown in Fig. 9, the missing data points are set to zero (i.e., they are replaced with the reference power system frequency of 50 Hz). To create outlier data points, we multiply the true value of the signal at a particular time instant with a random number generated in the range $[0.5, 1.5]$. Then, we replace the corresponding data point with the modified value. It should be noted, the outlier data points deviate significantly from their neighbouring data samples in comparison with the missing points (shown in Fig. 9). Therefore, the training process of machine learning algorithms may be more biased and influenced by outliers in input data than by missing point scenarios, which may result in longer training times, less accurate models, and ultimately worse results. Robust methods should focus on the general trend of the monitoring profiles rather than chasing down outliers. As indicated in Table VI, the 2DR-CNN surpasses the other benchmarks in terms of lost and outlier points, achieving accuracies of 89.21% and 85.91%, respectively (outlier scenarios result in lower accuracy). The primary reason for the 2DR-CNN's robustness in this challenging scenario lies in its strategy of reconstructing data via a decoder network along with the classification task. In addition, as illustrated in Fig.9, 2DR-CNN employs a rectangular kernel that moves in two directions across the data, distinguishing it from the point-by-point reading of the data employed in techniques such as AE-DNN, DNN, and SVM. Following 2DR-CNN, we see that 1D-CNN exceeds other benchmarks; nevertheless, 1D-CNN's performance declines dramatically in these cases (for instance, below 80% for the outlier point scenario), which is the fundamental motivation for employing 2D kernels and reconstruction network in this study.

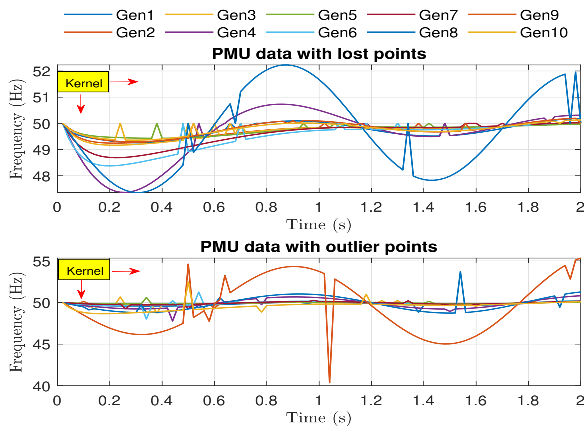

Fig. 9: Frequency profiles for IEEE 39-bus system with lost points in PMU data (multi-point attack: $K_L(28,5) = 59.45pu$, $\epsilon_{28} = 0.79$MW, $\epsilon_{24} = 5.98$MW, $\epsilon_{29} = 0.15$MW) and outlier points in PMU data (multi-point attack: $K_L(27,5)$=95.12pu, $\epsilon_{27} = 1.35$MW, $\epsilon_{23} = 8.62$MW, $\epsilon_{24} = 1.35MW$).

TABLE VI: Localization of multi-point D-LAAs on IEEE 39-bus system with lost/outlier points in PMU data.

Scenarios	2DR-CNN	1D-CNN	AE-DNN	DNN	SVM
Lost points in PMU data	89.21%	83.58%	77.24%	74.63%	43.19%
Outlier points in PMU data	85.91%	78.63%	69.26%	68.09%	35.47%

C. Scalability of the suggested framework

In this study, the monitoring framework is built on CNN structure, which has been shown to be successful in different big-data classification tasks like face recognition [19]. However, to verify the effectiveness of the proposed method in a larger case study, a challenging case, such as the IEEE 57-bus system, which is composed of 7 generator buses and 50 load buses is also investigated (see Section III-B). To design a monitoring framework for the IEEE 57-bus case, we followed the general concept shown in Fig. 4. The input data for this case is defined as $7 \times 100 \times 2$ and 50 classes are defined for the output of the localization task (see Section III-B). Our proposed method (2DR-CNN) delivers 98.51% accuracy for the detection task, whereas the second-best method (1D-CNN) achieves 95.18% and after that AE-DNN, DNN and SVM methods reach 93.82%, 93.02% and 90.17%, respectively. Table VII summarizes the results of the localization task in different scenarios for the IEEE 57-bus case. As shown in Table VII, difference between 2DR-CNN and other benchmark methods is more pronounced in this case than in the IEEE 14- and 39-bus cases. While the proposed method delivers a robust performance in noisy environments with more than 90% accuracy, the second-best method delivers an accuracy of 81.08% for multi-point attacks in noisy environments. In a similar manner, other methods lose significant performance and produce results below 80% (AE-DNN, DNN) and 70% (SVM). The localization results of the various methods with lost and outlier points in PMU data are shown in Table VIII that confirms the robustness of the 2DR-CNN method in comparison to other benchmarks by presenting significantly more accurate results (at least 7%) than the next-best method (1D-CNN).

TABLE VII: Localization results of D-LAAs on the IEEE 57-bus system with different approaches in different scenarios (without noise and average outcome of the noisy environments (26, 20 and 16.5 DB SNR).

Methods	Without noise		Average of noisy environments	
	Single-point attacks	Multi-point attacks	Single-point attacks	Multi-point attacks
2DR-CNN	93.71%	91.78%	91.59%	90.03%
1D-CNN	88.44%	86.09%	84.68%	81.08%
AE-DNN	82.11%	79.31%	79.31%	73.55%
DNN	80.42%	76.59%	77.12%	70.41%
SVM	77.23%	72.58%	69.28%	63.51%

TABLE VIII: Localization of multi-point D-LAAs on IEEE 57-bus system with lost/outlier points in PMU data.

Scenarios	2DR-CNN	1D-CNN	AE-DNN	DNN	SVM
Lost points in PMU data	87.32%	80.50%	71.85%	67.33%	39.17%
Outlier points in PMU data	83.29%	76.27%	63.57%	62.97%	30.10%

D. Efforts to address the new attack surfaces:

As compared to unsupervised and semi-supervised methods, supervised methods are able to achieve high accuracy, especially when it comes to localization, which involves multi-class classification [30], [31]. As a result, the proposed supervised localization solution is able to achieve accurate detection and localization outcomes in challenging scenarios (semi-destabilizing LAAs, see Section III-B), which is difficult with unsupervised methods. However, as with other supervised data-driven intrusion detection methods, a large number of labelled samples is needed for the training task [32]. To address this issue, in this study, we enrich our training data library by considering various types of static and dynamic LAAs (more than 3000 samples), details are given in SectionIII-B. In terms of generating a significant amount of power grid dynamic profiles (i.e., frequency and phase angle) while the system is under attack, two points should be considered: (i) the proposed algorithm relies on "physical measurement data" and there are well-known power grid simulators (e.g., MATPOWER, PowerWorld, etc.) that accurately model the system (widely used by the power system industry). (ii) the LAAs that this work concentrates on are attacks that disrupt the "physical controllers" of the power grid, which may also be easily simulated using the given simulators. As a result, while the system is under attack, we can generate enough training data utilising power grid simulators during the "offline training phase" (this approach has been widely applied to detecting cyber-physical attacks in power grids, such as false data injection attacks [30].) We note that due to the offline character of the training, the vast size of the training data will not be an impediment to the proposed method's deployment in real-world applications. In our numerical results, 10% of the generated data that was not observed by our supervised technique is assigned to the test procedure, and the precise detection and localization of load-altering threats are reported (more than 85% accuracy even in highly noisy environments). This verifies the extensiveness of our training data and the viability of the proposed technique in unexpected scenarios. Despite all of the above considerations, as with any other supervised intrusion detection techniques, there is a risk associated with the new attack surfaces [33]. Nevertheless, this risk can be mitigated by updating the training data with new samples, which is a common approach in supervised deep learning-based solutions [34], [35].

E. Approach Feasibility Discussion

The focus of this work is on monitoring the power grid from the perspective of a transmission system operator (TSO). In particular, we focus on transmission network oscillations, with the goal of localising the "substation(s)" from which the LAAs originate. We assume the power grid that the TSO has a thorough knowledge of the grid topology, which is a valid assumption for transmission networks, and the deployed PMUs can track the dynamics of the power grid in real-time [22]. The synchrophasor technology has been widely used to improve the observability of power systems over the past decade. The proposed method uses 50 samples per second. With the aid of global positioning system time stamps (GPS), the acquired data (such as voltage, current, frequency, and phase angle) can be synchronized. A phasor data concentrator (PDC) in a control station receives the measurements recorded by each PMU using the IEEE C37.118 protocol. Several PMU applications (such as wide-area monitoring, disturbance detection, and voltage stability) have been developed in order to enhance the stability of the power grid [36]. In light of the aforementioned details, the proposed method is capable of being implemented in real-world scenarios as an effective monitoring solution against different kinds of LAAs in power grids.

V. Conclusions

This paper proposes a multi-output data-driven technique based on the 2DR-CNN structure for the detection and localization of D-LAAs. To accomplish this, a reconstruction decoder is applied along with the classification networks, in which the first classifier detects D-LAAs by distinguishing them from natural power grid oscillations, and the second classifier precisely localizes the location of D-LAAs. To demonstrate the suggested method's effectiveness, a range of attacks, including single- and multi-point attacks, are examined in two case studies involving IEEE 14- and 39-bus systems. The proposed method achieves an accuracy of approximately 95% by using 2D feature extraction and reconstruction decoder network equipped with sparsity regularization constraints, thus outperforming other benchmarks. Additionally, to investigate the proposed method's robustness in practical situations, the findings address diverse scenarios such as noise and lost/outlier points in the PMU data. By incorporating a robust feature extraction approach with its 2D kernel, the suggested method exhibits robust performance in the aforementioned conditions. As demonstrated by the numerical findings, the proposed method has the potential to be considered as an extra security surface for increasing the protection of modern power grids. Future work for this study will focus on designing a robust network by incorporating additional scenarios, such as the delay in receiving profiles from PMUs, and implementing a hybrid method with the support of advanced semi-supervised techniques in the localization task to mitigate the risks of new attack vectors.

Appendix: Simulation Parameters

Dynamic parameters for IEEE 14-bus system:

$$M_1 - M_5 = [0.125; 0.034; 0.016; 0.010; 0.015];$$
$$D_1 - D_5 = [0.125; 0.068; 0.032; 0.068; 0.072];$$
$$K_1^P - K_5^P = [0.02; 0.09; 0.03; 0.03; 0.08];$$
$$K_1^I - K_5^I = [0.35; 0.40; 0.35; 0.35; 0.40];$$
$$D_i = 0.01, \forall i \in \mathcal{N}_L;$$

Dynamic parameters for IEEE 39-bus system:

$$M_1 = 2.3186; M_2 : M_8 = 2.6419; M_9 : M_{10} = 2.4862.$$
$$K_1^P - K_{10}^P = [1; 0.45; 0.45; 0.1; 0.5; 0.4; 0.3; 0.2; 0.4; 0.5];$$
$$K_i^I = 0.6, \forall i \in \mathcal{N}_G; D_i = 2, \forall i \in \mathcal{N}_G; D_i = 0.01, \forall i \in \mathcal{N}_L;$$

Dynamic parameters for IEEE 57-bus system:

$$M_1 - M_7 = [2.6309; 1.200; 5.078; 1.200; 2.6309; 1.200; 2.6309];$$
$$D_1 - D_7 = [2; 0; 2; 0; 2; 0; 2];$$
$$K_1^P - K_7^P = [25; 35; 10; 20; 30; 10; 30];$$
$$K_1^P - K_7^P = [25; 20; 20; 20; 30; 15; 30];$$
$$D_i = 0.2, \forall i \in \mathcal{N}_L;$$

References

[1] H.-N. Dai, Z. Zheng, and Y. Zhang, "Blockchain for internet of things: A survey," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8076–8094, 2019.

[2] S. Soltan, P. Mittal, and H. V. Poor, "BlackIoT: IoT botnet of high wattage devices can disrupt the power grid," in *Proc. USENIX Security*, 2018, pp. 15–32.

[3] B. Huang, A. A. Cardenas, and R. Baldick, "Not everything is dark and gloomy: Power grid protections against iot demand attacks," in *Proc. USENIX Security Symposium*, Aug. 2019, pp. 1115–1132.

[4] S. Amini, F. Pasqualetti, and H. Mohsenian-Rad, "Dynamic load altering attacks against power system stability: Attack models and protection schemes," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 2862–2872, July 2018.

[5] S. Lakshminarayana, S. Adhikari, and C. Maple, "Analysis of IoT-based load altering attacks against power grids using the theory of second-order dynamical systems," *IEEE Transactions on Smart Grid*, vol. 12, no. 5, pp. 4415–4425, 2021.

[6] J. Ospina *et al.*, "On the feasibility of load-changing attacks in power systems during the covid-19 pandemic," *IEEE Access*, vol. 9, pp. 2545–2563, 2021.

[7] S. Lakshminarayana, J. Ospina, and C. Konstantinou, "Load-altering attacks against power grids under COVID-19 low-inertia conditions," *IEEE Open Access Journal of Power and Energy*, pp. 1–1, 2022. [Online]. Available: 10.1109/OAJPE.2022.3155973

[8] S. Amini, F. Pasqualetti, M. Abbaszadeh, and H. Mohsenian-Rad, "Hierarchical location identification of destabilizing faults and attacks in power systems: A frequency-domain approach," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2036–2045, 2019.

[9] M. Izbicki, S. Amini, C. R. Shelton, and H. Mohsenian-Rad, "Identification of destabilizing attacks in power systems," in *Proc. American Control Conference (ACC)*, 2017, pp. 3424–3429.

[10] S. Lakshminarayana, S. Sthapit, H. Jahangir, C. Maple, and H. V. Poor, "Data-driven detection and identification of iot-enabled load-altering attacks in power grids," *IET Smart Grid*, 2022.

[11] K. Pan and P. Mohajerin, "From static to dynamic anomaly detection with application to power system cyber security," *IEEE Transactions on Power Systems*, vol. 35, no. 2, pp. 1584–1596, 2019.

[12] V. K. Singh and M. Govindarasu, "A cyber-physical anomaly detection for wide-area protection using machine learning," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3514–3526, 2021.

[13] T. Huang, N. M. Freris, P. Kumar, and L. Xie, "A synchrophasor data-driven method for forced oscillation localization under resonance conditions," *IEEE Transactions on Power Systems*, vol. 35, no. 5, pp. 3927–3939, 2020.

[14] T. R. Nudell, S. Nabavi, and A. Chakrabortty, “A real-time attack localization algorithm for large power system networks using graph-theoretic techniques,” *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2551–2559, 2015.

[15] W. Li, D. Deka, M. Chertkov, and M. Wang, “Real-time faulted line localization and pmu placement in power systems through convolutional neural networks,” *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4640–4651, 2019.

[16] Z. Li, H. Liu, J. Zhao, T. Bi, and Q. Yang, “A power system disturbance classification method robust to pmu data quality issues,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 1, pp. 130–142, 2021.

[17] 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 Transactions on Smart Grid*, vol. 13, no. 3, pp. 2369–2380, 2022.

[18] Y. Wu, M. Musavi, and P. Lerley, “Synchrophasor-based monitoring of critical generator buses for transient stability,” *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 287–295, 2015.

[19] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[20] J. Sun, G. Qi, N. Mazur, and Z. Zhu, “Structural scheduling of transient control under energy storage systems by sparse-promoting reinforcement learning,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 2, pp. 744–756, 2021.

[21] “Ieee standard for interconnection and interoperability of distributed energy resources with associated electric power systems interfaces,” *IEEE Std 1547-2018 (Revision of IEEE Std 1547-2003)*, pp. 1–138, 2018.

[22] O. Samuelsson, M. Hemmingsson, A. H. Nielsen, K. O. H. Pedersen, and J. Rasmussen, “Monitoring of power system events at transmission and distribution level,” *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 1007–1008, 2006.

[23] “IEEE/IEC international standard - measuring relays and protection equipment - part 118-1: Synchrophasor for power systems - measurements,” *IEC/IEEE 60255-118-1:2018*, pp. 1–78, 2018.

[24] S. Acharya, Y. Dvorkin, and R. Karri, “Public plug-in electric vehicles+ grid data: Is a new cyberattack vector viable?” *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5099–5113, 2020.

[25] K. P, B. N, and L. M. G, *Power system stability and control*. McGraw-hill New York, USA, 1994, vol. 7.

[26] M. A. Tabrizi, N. Prakash, M. Sahni, H. Khalilinia, P. Saraf, and S. Kolluri, “Power system damping analysis on large power system networks: An entergy case study,” in *2017 IEEE Power & Energy Society General Meeting*. IEEE, 2017, pp. 1–5.

[27] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, “Edge-iiotset: A new comprehensive realistic cyber security dataset of iot and iiot applications for centralized and federated learning,” *IEEE Access*, vol. 10, pp. 40 281–40 306, 2022.

[28] H. Elayan, M. Aloqaily, and M. Guizani, “Sustainability of healthcare data analysis iot-based systems using deep federated learning,” *IEEE Internet of Things Journal*, vol. 9, no. 10, pp. 7338–7346, 2022.

[29] M. Brown, M. Biswal, S. Brahma, S. J. Ranade, and H. Cao, “Characterizing and quantifying noise in pmu data,” in *2016 IEEE Power and Energy Society General Meeting (PESGM)*. IEEE, 2016, pp. 1–5.

[30] S. Wang, S. Bi, and Y.-J. A. Zhang, “Locational detection of the false data injection attack in a smart grid: A multilabel classification approach,” *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8218–8227, 2020.

[31] W. Zhou, H. Li, and M. A. Anastasio, “Approximating the ideal observer for joint signal detection and localization tasks by use of supervised learning methods,” *IEEE transactions on medical imaging*, vol. 39, no. 12, pp. 3992–4000, 2020.

[32] L. Erhan, M. Ndubuaku, M. Di Mauro, W. Song, M. Chen, G. Fortino, O. Bagdasar, and A. Liotta, “Smart anomaly detection in sensor systems: A multi-perspective review,” *Information Fusion*, vol. 67, pp. 64–79, 2021.

[33] F. Li, Q. Li, J. Zhang, J. Kou, J. Ye, W. Song, and H. A. Mantooth, “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.

[34] J. Du, Y. Zhou, P. Liu, C.-M. Vong, and T. Wang, “Parameter-free loss for class-imbalanced deep learning in image classification,” *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[35] Z. Xu, D. Shen, Y. Kou, and T. Nie, “A synthetic minority oversampling technique based on gaussian mixture model filtering for imbalanced data classification,” *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[36] C.-C. Sun, A. Hahn, and C.-C. Liu, “Cyber security of a power grid: State-of-the-art,” *International Journal of Electrical Power & Energy Systems*, vol. 99, pp. 45–56, 2018.

Hamidreza Jahangir is a Postdoctoral Research Fellow at the School of Engineering, University of Warwick, UK. He received his Ph.D. and M.Sc from the Faculty of Electrical Engineering at the K. N. Toosi University, Tehran, Iran in 2021 and 2015, B.Sc from the Faculty of Electrical Engineering at Iran University of Science and Technology, Tehran, Iran in 2013. His research focuses on advanced Artificial Intelligence applications in smart energy management systems. His works mainly revolve around developing deep learning-based approaches to estimate the electric vehicle charging demand, generation of renewable energy resources, load demand, electricity price, modelling cyber-physical systems and designing smart decision-making frameworks.

Subhash Lakshminarayana is an Associate Professor at the School of Engineering, University of Warwick, UK. He received his Ph.D. from the Alcatel Lucent Chair on Flexible Radio and the Department of Telecommunications at École supérieure d’électricité, France in 2013, M.S. degree in Electrical and Computer Engineering from The Ohio State University in 2009 and B.S. from Bangalore University, India in 2007. His research interests include cyber-physical system security (power grids and urban transportation) and wireless communications. His works have been selected among the Best conference papers on integration of renewable & intermittent resources at the IEEE PESGM - 2015 conference, and the “Best 50 papers” of IEEE Globecom 2014 conference. Currently, his research is funded by Innovate UK, EPSRCPETRAS National Centre of Excellence for Cybersecurity of IoT Systems in UK, and the EUTOPIA European Alliance.

Carsten Maple is the Principal Investigator of the NCSC–EPSRC Academic Centre of Excellence in Cyber Security Research, University of Warwick, where he is a Professor of Cyber Systems Engineering with Warwick Manufacturing Group. He is also a Fellow of the Alan Turing Institute, the National Institute for Data Science, and AI in the U.K., where he is a Principal Investigator on a $5 million project developing trustworthy national identity to enable financial inclusion. He is a CoInvestigator of PETRAS, the National Centre of Excellence for IoT Systems Cyber Security and works with numerous banking organizations advising on security, privacy, and use of artificial intelligence. He has an international research reputation and extensive experience of institutional strategy development and interacting with external agencies. He has published over 250 peer-reviewed papers and has coauthored the U.K. Security Breach Investigations Report 2010, supported by the Serious Organized Crime Agency and the Police Central e-crime Unit.

Gregory Epiphaniou received the B.Sc. degree in industrial design from the University of Western Macedonia, Kozani, Greece, in 1993, and the Ph.D. degree from the University of Bedfordshire, Luton, U.K., in 2011. He is currently holds a position as an Associate Professor of Security Engineering with the University of Warwick, Coventry, U.K. His role involves bid support, applied research, and publications. He led and contributed to several research projects funded by EPSRC, IUK, and local authorities totalling over £4M. He currently holds a subject matter expert panel position in the Chartered Institute for Securities and Investments. Part of his current research activities is formalized around cyber effects modeling, wireless communications with the main focus on crypto-key generation, exploiting the time-domain physical attributes of V-V channels, and cyber resilience. Dr. Epiphaniou acts as a technical committee member for several scientific conferences in Information and network security and served as a Key Member in the development of WS5 for the formation of the U.K. Cybersecurity Council