IEEE Open Journal of the
Industrial Electronics Society

Received 9 October 2022; accepted 21 November 2022. Date of publication 25 November 2022;
date of current version 2 December 2022. The review of this paper was arranged by Associate Editor Francisco D. Freijedo.

Digital Object Identifier 10.1109/OJIES.2022.3224784

Fine-Tuned RNN-Based Detector for Electricity Theft Attacks in Smart Grid Generation Domain

MAYMOUNA EZ EDDIN[1], ABDULLATIF ALBASEER[1] (Member, IEEE), MOHAMED ABDALLAH[1] (Senior Member, IEEE), SERTAC BAYHAN[2] (Senior Member, IEEE), MARWA K. QARAQE[1] (Senior Member, IEEE), SAIF AL-KUWARI[1] (Senior Member, IEEE), AND HAITHAM ABU-RUB[3] (Fellow, IEEE)

[1]Division of Information and Computing Technology, College of science and engineering, Hamad Bin Khlifa University, 34110 Doha, Qatar
[2]Qatar Environment and Energy Research Institute (QEERI), Hamad Bin Khalifa University, 34110 Doha, Qatar
[3]Electrical and Computer Engineering Department, Texas A&M University at Qatar, 34110 Doha, Qatar

CORRESPONDING AUTHOR: ABDULLATIF ALBASEER (e-mail: aalbaseer@hbku.edu.qa)

This work was supported by NPRP Cluster project (NPRP-C) Twelve (12th) Cycle under Grant NPRP12C-33905-SP-67 from the Qatar National Research Fund (a member of Qatar Foundation).

ABSTRACT In this article, we investigate the problem of electricity theft attacks on smart meters when malicious customers (i.e., adversaries) claim injecting more generated energy into the grid to get more profits from utility companies. These attacks can be applied by accessing the smart meters monitoring renewable-based distributed generation (DG), and manipulating the reading. In this article, we propose approaches that rely on data sources with only a single generator (i.e., solar only) and multifuel type; and address the crucial effects of slight perturbations that the attacker can add, which can deceive the detector. In particular, this article introduces an efficient multitask deep-learning-based detector that offers a higher detection rate, copes with different fuel types, and uses only single data sources. The proposed detector incorporates months and days as two additional features to boost the performance and properly guide the model to successful detection. The proposed method is then extended to consider small perturbations that attackers may use to launch successful attacks. We conduct extensive simulations for two different detectors, one for solar DG and the other for multiple fuel types (i.e., solar and wind). Using a realistic dataset, the results reveal that the proposed recurrent neural network-based detectors identify adversaries at a higher rate than the existing solutions, even with minimal perturbations and different fuel types.

INDEX TERMS Cyberattacks on smart grid, electricity theft, generation domain, deep-learning (DL)-based detector, small perturbations.

I. INTRODUCTION

Traditional electricity grid (TEG) is based on one-way transmission within a hierarchical communication network. Electricity utilities have to realize the need to address the critical challenges faced, including the ever-increasing electricity demand, the low efficacy, rising electricity costs, and the bad environmental impact of existing grids [1]. With current power network requirements, the TEG may not be able to meet those needs, which necessitates the development of the smart grid (SG). An SG uses bidirectional power transmission and information flow, making it the next generation of the power grid.

The SG is conceptually divided into seven domains: generation, transmission, distribution, customers, operation, market, and service provider. The first four domains are involved in power flow, while the rest are related to control and communication in the SG system. These domains are enabled by new technologies, such as the Internet of Things (IoT), Supervisory Control and Data Acquisition (SCADA), and Advanced Metering Infrastructure (AMI), especially smart

meters. Smart meters are digital meters with microprocessors and onboard memory, which enable them to monitor and collect power usage on the consumer side. In particular, IoT enables bidirectional communication and data transformation between all smart devices in the network via the internet, including sensors, actuators, and smart meters, which allows energy monitoring and remote control of the SG system. Policymakers, developers, and researchers are motivated to use the SG system because of the increasing electricity demand, the aging of current electrical infrastructure, increasing energy charges, electricity reliability concerns, renewable power generation unit development, and electric vehicles, to name a few [2].

The SG brings valuable societal benefits, such as enhanced immeasurable utilization of current resources and ubiquitous control [3]. However, the emergence of smart technologies poses major cybersecurity risks due to the following:

1) legacy systems, such as industrial control systems (ICS) and SCADA, which are insecure [4];
2) the existing vulnerabilities in transmission control protocol/internet protocol [5];
3) novel attacks (e.g., false data injection and electricity theft) caused by new emerging smart technologies (i.e., smart meters) [6].

The first known cyberattack on a power grid was launched against the Ukrainian power infrastructure on December 23, 2015. The attack targeted three Ukrainian operators (providers to the Kyiv, Ivano-Frankivsk, and Chernivtsi regions). It used the BlackEnergy Trojan to infect ICSs, mainly SCADA, for power distribution. This attack caused power outages for nearly 230 000 customers that lasted several (one to six) hours; the damage to the grid took months to repair. A year after this initial successful hack, in December 2016, another one took place that disrupted power service for an hour in portions of Kyiv by deploying the Industroyer virus targeting ICSs [7]. A common attack vector against the SG includes denial of service (DoS), unauthorized access (UA), and false data injection (FDI) [8]. DoS attacks target the availability of the relevant systems, one common type of DoS attack is the jamming attack, where the attacker aims to increase the packet dropout rate of the channel [9]. At the same time, UA and FDI exploit vulnerabilities in the industry protocols to compromise the authenticity, confidentiality, and integrity of the data exchanged. FDI stealthy attack is a type of attack that considers bad data detection system functionality in the SG utility to increase the chances of the attack bypassing the detector [10]. Although these cyberattacks significantly impact smart grid functions, this article focuses on FDI attacks.

Meters and sensors lack tamper-resistance hardware, which increases the risk of the SG being compromised, as SGs may operate in hostile environments. For example, a malicious adversary might inject false measurements to disrupt SG operations by compromising the meters and sensors, disrupting the grid system state estimation and energy distribution. An example of a common threat is electricity theft (ET), which leads to major financial losses for electricity providers worldwide [11]. The SG has resulted in new forms of energy theft wherein malicious customers endeavor to execute cyberattacks rather than tap on the line or tamper with meters like in the case of the TEG. [12].

Malicious consumers or customers can generally cause ET attacks. The malicious consumer has no distributed generation (DG) unit and intends to manipulate the reported energy data to claim lower consumption, and consequently, reduce bills. In the case of a malicious customer attack, they manipulate the amount of energy they generate to be fed back to the grid. Some electric utility companies encourage customers to participate in energy generation and feed the generated energy back into the grid. Customers can generate their own energy by installing DG units, such as solar cells, photovoltaic (PV), and wind turbines. In this article, we focus on DG unit malicious customer attacks where the goal is to claim injecting more energy into the grid to earn more profits.

Electricity companies use feed-in tariffs (FITs) and net metering policies to encourage customers to use renewable energy sources. In FITs, customers who send all their generated energy to the grid receive a cashback from the utility company [13]. In the net metering policy, customers inject only the generated energy into the grid and receive a credit as a reduction on their next bill [14]. In an adversarial FITs scenario, the malicious customer attempts to manipulate the smart meter's reported energy data (ET attack), claiming a higher injected energy into the grid, and consequently, attaining more profit. Customers who have access to the firmware via the ANSI optical port of these smart meters can conceivably execute this attack by exploiting the weak authentication software installed in most of these meters [15]. The existing defense technique of the SCADA system against FDI is called bad data detection. It usually uses hypothesis testing by observing the largest normalized residual to detect the bad measurement data [16]. Liu et al. proved that attackers could initiate FDI attacks in electric grids against the existing state estimation and bad data detection techniques, assuming that attackers can compromise some meter devices and have some knowledge of electric grid connections and configurations [17]. Due to the simplicity of the bad data detection of SCADA, some ET cyberattacks may not be detected, causing the exact impact of FDI on the grid. This eventually leads to economic loss for the electricity utility. Thus, there is an emerging need to use more complex algorithms to efficiently detect such attacks. Machine learning (ML) technology has advanced rapidly in recent years, and detection-based defense against ET attacks is gradually shifting to adopt ML. Many ML techniques have been proposed in the literature [18], [19], [20], [21], [22] to detect ET attacks, either in the consumption domain, where the goal is to reduce the consumption bill or in the generation domain, where the goal is to inject more energy into the grid and gain more profit (i.e., FITs). Referring to the hourly ontario energy price (HOEP) [23] in 2020, and considering 4% additional energy, the adversary will earn $609.696 additional profit. However, most of these works only focused on

the consumption domain, necessitating more efforts for the generation domain. ML-based approaches can be divided into classification problems and anomaly detection. In the classification approach, the algorithm utilizes benign and malicious data in the training and testing stages. On the other hand, anomaly detection uses only benign data in the training stage; then, malicious data can be used in the testing stage.

The classification approach [24], [25], [26], [27], [28], [29] provides a high attack detection rate when a complex learning algorithm, such as deep learning (DL), is utilized to learn data patterns. However, a key bottleneck of this approach is the limited benign and malicious labeled dataset, which limits the ability to test how well the developed models generalize in larger or diverse malicious datasets. Moreover, this approach fails to detect unseen attacks, such as zero-day attacks. Similarly, building an ET detector (ETD) based on anomaly detection has received similar attention from the research community [11], [18], [30], [31], [32], [33], [34], [35]. Although anomaly detection methods can detect zero-day attacks, they usually provide a lower detection rate than the classification approach [36].

Despite the significant efforts focused on detecting ET attacks in the literature, most of these works focused mainly on the consumption domain, and attacks on the generation domain have received less attention. The authors in [19] recorded a high detection rate by integrating three data sources to detect a scenario where attackers claimed 20% additional power. However, they did not report the system's performance against small perturbation attacks. The data sources used in [19] are related to SCADA and generator capability, which are not instantly available to the electricity utility. Furthermore, all the work in the literature considered only a single generator type (i.e., solar) as in our previous work [37].

A. CONTRIBUTION

Motivated by the above remarks, this article investigates the aforementioned issues in more depth and proposes DL-based detectors to fill these gaps, considering model complexity, data availability integrated into the solution, and system performance against malicious behavior of small perturbations. We propose a DL-based detector that employs a single data source (i.e., energy generation profile per hour) for a single renewable energy fuel type (i.e., solar) and multiple renewable energy fuel types (i.e., solar and wind). To the best of our knowledge, this is the first work considering smaller perturbation impacts on the DG SG system and multiple fuel types. The contribution of our work can be summarized as follows.

1) We investigate the adoption of generation power profiles for solar and wind renewable energy, utilizing public datasets to detect ET in the generation domain. We prove that ET cyberattacks can be detected with a higher detection rate and accuracy using a single data source. This can be done by adopting month and day features to enhance the system performance and reduce model complexity.
2) We propose two ETDs, one unique for solar DG units and another unique hybrid for solar and wind DG units. Using the gated recurrent neural networks (GRU-RNN) model, we develop DL-based ETDs that capture the temporal features in the solar energy time-series dataset. We utilized the bidirectional long short term memory (BLSTM) model for the unique hybrid ETD. The proposed detection scheme achieves a detection rate of 96.68% and 88.58% for solar and hybrid ETD, respectively, when detecting malicious behavior of 20% claimed additional power.
3) We investigate how well ETDs behave against small perturbation attacks and propose training models using small perturbations to make them more robust. The proposed small perturbation-trained models achieve a detection rate of 91.61% in solar and 86.79% in hybrid environments, even if attackers only slightly change the reported energy by 8%.

B. ARTICLE ORGANIZATION

The rest of this article is organized as follows. Section II briefly summarizes the existing literature. In Section III, we present our methodology, including data generation, feature selection, data cleansing, and ETD model training and inference. In Section IV, we present the experimental setup and the utilized evaluation metrics. The development of the detection model is introduced in Section V. The results are discussed in Section V, in which we evaluate the performance of the proposed models. Finally, Section VII concludes this article.

II. RELATED WORK

Researchers introduced different detection techniques for ET attacks, focusing on either the consumption or generation domains. Recent research has primarily used ML and DL algorithms to construct an ETD. ETD approaches based on ML may be divided into two categories: Anomaly detection based on supervised learning (i.e., classification problem) [38], [39], [40], [41], [42] and anomaly detection based on unsupervised learning [36]. Anomaly detection based on supervised learning employs both benign (truthful) and malicious (manipulated) electricity data samples to train and validate the detector. While anomaly detection is based on unsupervised learning, anomaly detection systems use only benign data to learn the true client consumption behavior. Malicious data are discovered during testing based on deviation from the learned honest pattern.

A. ELECTRICITY THEFT IN CONSUMPTION DOMAIN

Data-driven methods for detecting ET in the consumption domain became popular due to large volumes of data collected from smart meters installed at customers' premises. A few existing studies are based on data-driven ML algorithms, which classify consumers as honest or bad based on their load profile, [19] where the characteristics of customers' benign and malicious energy usage were used to create supervised classifiers. In [24], the authors employed

a gradient-boosting theft detector with feature engineering-based preprocessing. The authors in [27] developed the andhra pradesh central power distribution corporation limited (AP-SPDCL) naive bayes (NB) classifier for detecting unexpected customer consumption trends in power distribution networks. Random forests [28] and AdaBoost [29] were utilized as ETD. On the other hand, the authors in [43] used an extreme gradient boosting classifier. The researchers chose DL algorithms for ET detection to capture the energy datasets' intricate patterns. In [22], a DL-based detector was added, primarily to detect malicious customers on the consumption side, where synthetics attacks were generated, presuming suspicious consumers may lower consumption by a fixed or random amount. The authors in [44] created a sequential ensemble algorithm based on a deep autoencoder with attention (AEA) for detecting various cyberattacks. The research published in [45] used an RNN classifier based on GRU that captures the temporal correlation in the customer's load profile. DL-based classifiers including feed-forward neural networks [46], RNNs [45], and vector embedding [47] were utilized as ETD.

Several studies investigated the anomaly detection approach because the classification-supervised approach does not account for zero-day attacks. When only benign data are used, for instance, outlier detection [31] was used to design an anomaly detection based ETD. Xiao et al. [32] developed an ETD based on the random matrix theory.

Anomalies in meter readings were also detected using the integrated convolutional neural networks technique as in [18] where Xue et al. created a powerful learning machine, which localizes the FDI-attacked buses in a power system. To deal with the identified false data, a recovery procedure was also implemented.

B. ELECTRICITY THEFT IN GENERATION DOMAIN

Some research focused on classification-supervised ML data-driven systems to detect energy generation theft from PV panels. The authors in [19] concentrated on creating a supervised learning classifier based on deep (stacked) autoencoders with an LSTM-based sequence-to-sequence (seq2seq) structure. Furthermore, the authors in [19] created a malicious database synthetically to train the detector. The main advantage of this work is the high detection rate. However, this rate only depends on the proposed attack functions used in the study. Thus, zero-day attacks may not be detected. As a result, the authors in [20] developed an unsupervised detector based on anomaly detection that is trained solely on benign data collected by the operator during regular system operation. Similarly, in [26], the authors utilized a detector based on the least-square error and a sliding temporal window. The main advantage of the unsupervised detector is that it can detect zero-day attacks. However, it has a low detection rate compared to supervised learning models. Furthermore, the authors in [15] suggested optimal cyberattack functions on the DG units (assuming that attackers know the detection method). Moreover, the authors in [15] built a detector using autoregressive integrated moving average (ARIMA) models, kullback-leibler divergence (KLD), and principle component analysis (PCA). Table 1 summarizes related work advantages and drawbacks.

Although considerable studies have been devoted to addressing the ET in the SG, there are some common limitations. We present our reflection on the designs of the proposed ETDs based on the summary shown in Table 2. We argue the following.

1) Most of the existing research addresses ET detection in the energy consumption domain rather than generation.
2) Because there is an insufficient dataset with benign and malicious ground-truth samples, the researcher's approaches are hampered. More specifically, due to the scarcity of real datasets containing both benign and malicious data, researchers had to artificially introduce some malicious data, which could lead to bias.
3) Some research efforts focused on an anomaly detection approach to detect ET due to dataset restrictions. While this approach does not usually yield a high detection rate, it can detect zero-day attacks.
4) To improve the detection rate, supervised ML techniques were used. However, the scarcity of datasets, including malicious records, limits the applicability of this approach.
5) Some studies employ shallow ML algorithms, which are incapable of accurately capturing multiple consumption patterns seen in the complex structure of power metering data.
6) Despite prior research on ET attacks on the SG, no studies have been conducted to investigate the effects of small malicious perturbations on ET.

III. METHODOLOGY

Currently, SCADA's bad data detection system in AMI is the only defender against ETA [43]. According to HOEP [23], a malicious customer who could add 20% energy could steal approximately 18290$ in the year 2020, with an hourly weighted average of 1.74$ per kW·h. Researchers use DL models to learn the complex nature of energy generation data and better detect the ETA. However, this approach cannot effectively detect the variations of ET attacks, such as smaller perturbations.

Hence, this article designs an efficient and robust DL-based detector for ETA against malicious behavior variation, particularly smaller perturbations. We utilize renewable energy generator profiles, used in [19], to create a synthetic benign and malicious dataset. Moreover, we utilize wind turbines' energy generation dataset considering ETA detection in a mixed DG unit environment, including solar panels and wind turbine DG units. In particular, we train an RNN-based model capable of learning temporal features in time-series data. We also include small perturbations of malicious behavior within the data to make the system more robust against such attack types. This will support SCADA's bad data detection function to detect ET attacks in solar and wind DG units, as it is more complex.

TABLE 1. Existing ETD-Based ML Approaches in Generation Domain

Approach	Advantages	Drawbacks
LSTM-based detector [19]	High detection rate	Limited attack functions Undetected zero-day attack
Datadriven based on regression tree [20]	Detected zero-day attack	Low accuracy High false alarm
Leastsquare error with sliding window [26]	Detected zero-day attack	Low accuracy High false alarm

TABLE 2. Literature ETD Summary Based on SG Domain and Adapted ML Approach

SG domain	ML Approach	References
Consumption domain	Supervised ML	[22], [24], [25], [26], [27], [28], [29], [43], [44], [45], [45], [46], [47], [48], [49], [50], [51], [52]
	Anomaly detection	[11], [18], [30], [31], [32], [33], [34], [34], [35]
Generation domain	Supervised ML	[19]
	Anomaly detection	[15], [20], [26], [53]

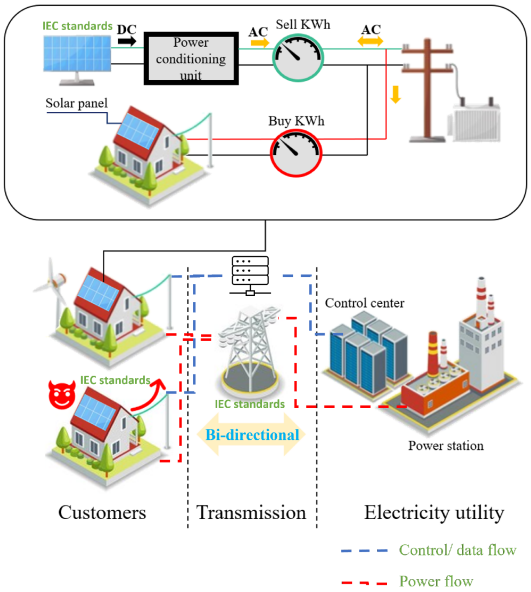

FIGURE 1. DG unit in a smart grid with an FIT payment system where a malicious customer manipulates power generation reports to increase their profit.

A. SYSTEM OVERVIEW

In DG units, most energy utilities' cashback is based on the generated energy, similar to FITs. When malicious customers have access to smart meters installed in their homes, they can manipulate the generation report to maximize their profit. Fig. 1 depicts this scenario. We assume the following.

1) Malicious customers can access the smart meter port.
2) Each generation unit has its own energy report.
3) Electricity utility has information about the fuel type used in the generator.
4) Different percentages can be applied to create an attack during the day.
5) Malicious customers apply one type of attack per day and do not mix between them.

B. DATA GENERATION

In this article, we use two investigation schemes: one with a single fuel type (i.e., solar only) and another with multiple fuel types (i.e., solar and wind). The data generation steps are depicted in Fig. 2.

1) RAW DATASET DESCRIPTION

This article uses a real smart meter dataset from Ontario, Canada [54]. Specifically, a public report by the Independent Electricity System Operator (IESO) considers hourly energy measurements in Ontario, Canada, as a 5-min average per hour [54]. The generator energy reports create realistic load profiles for residential households. The measurements for each generator unit are reported daily under the delivery date feature. The IESO provides features for each generator, including the generator type, fuel type, measurements, and generated energy per hour (24 features). There are several generator types for each fuel type: wind, solar, gas, hydro, biofuel, and nuclear. For each DG, the measurement feature has four categories: capability, output, available capacity, and forecasting. It is worth noting that the measurement feature depends on the fuel type. Each measurement category has an energy report entry, where capability represents the maximum energy per hour that the generator can produce if it depends on biofuel, gas, hydro, and nuclear fuel type. Similarly, the available capacity measurement category represents the maximum energy per hour a generator can produce, depending on wind or solar fuel type. In addition, forecast measurement indicates the maximum energy wind or solar generators can produce depending on the weather that day. For all fuel types, output measurement reports the actual unit-generated energy per hour that day. In this article, we use the raw datasets of the

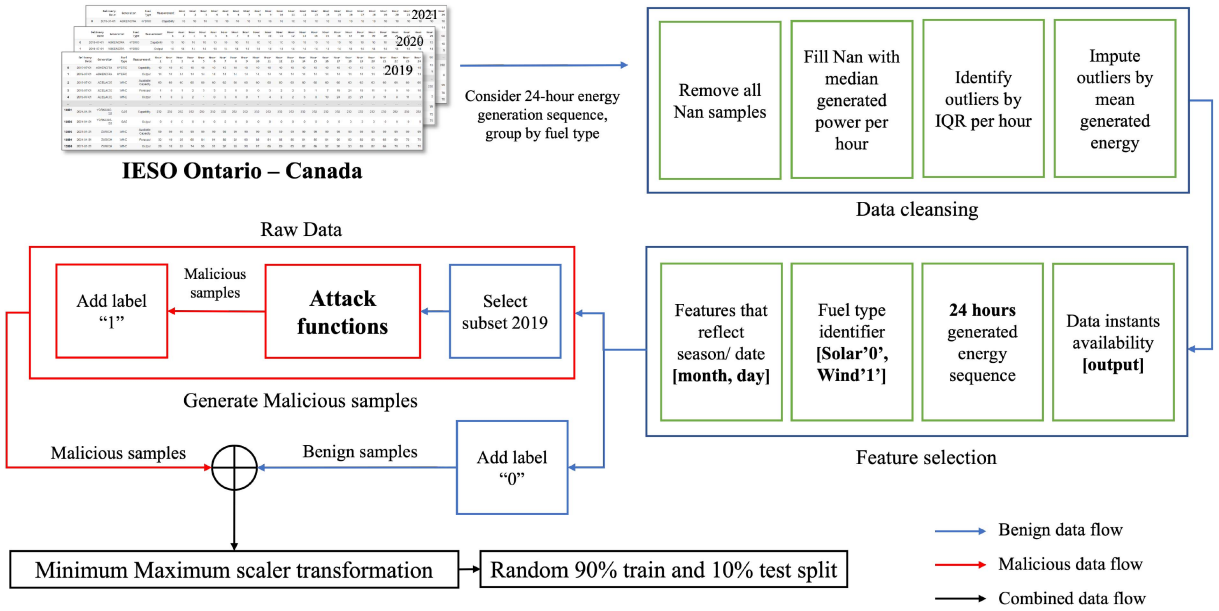

FIGURE 2. Data generation flowchart.

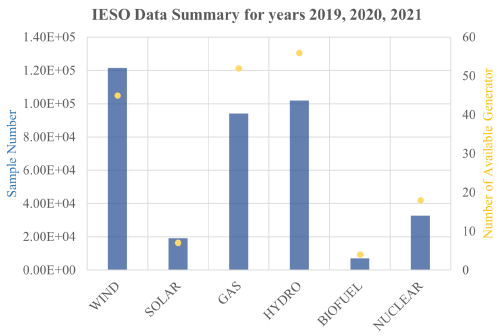

FIGURE 3. Raw dataset summary for the years 2019, 2020, and 2021 combined based on fuel type, with each type figure displaying sample number as a bar chart and number of different generators as a scatter chart.

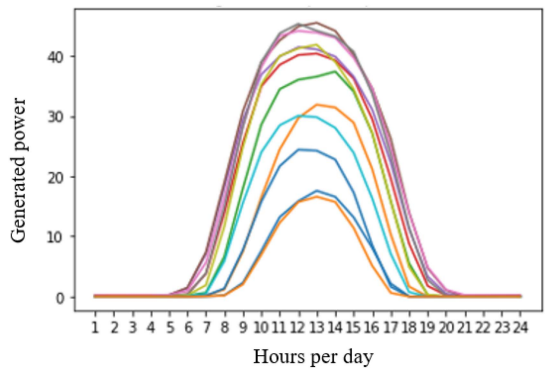

FIGURE 4. Hourly solar generated energy averaged by month.

past three years (i.e., 2019, 2020, and 2021). Fig. 3 depicts the data summary based on the fuel type.

2) FEATURE SELECTION

To cope with data availability challenges, we only consider the output measurement from the existing readings, representing the actual generated energy. We aim to design a detector that works for wind and solar DG unit environments. Therefore, we select a subset of the data where the fuel type is based on wind and solar. Then, we encode the fuel type feature because each fuel has a different data pattern. As solar and wind renewable energies generation capability depends on the weather, as proved in [54] for solar energy, we consider seasons representative features. In practice, month and day can represent the weather yet are instantly available in the raw dataset. Fig. 4 shows the solar energy generation variation per month as an example. Without loss of generality, we consider both generator types (i.e., solar and wind) and create a feature space containing 27 features, including the 24 actual generated energy sequences (i.e., the hourly based reported injected energy during the day), and the three added features (i.e., the month, day, and fuel type).

3) DATA CLEANSING

The data are grouped by fuel type (solar or wind), and the following preprocessing steps are applied. First, all faulty samples are removed from the data. This includes data samples with empty entries (NAN) or data samples showing zero generation. Next, the missing data (NAN) are imputed by considering the median value across the day. Finally, outliers in the benign data, identified using the interquartile range method, are imputed with the mean across each feature.

TABLE 3. On Renewable-Based DG Units, Cyberattack Functions for Electricity Theft, [19]

Attack type	Mathematical modeling
Partial increment attack: fixed α	$A_1(B_{t,d}) = (1+\alpha)B_{t,d}$
Partial increment ttack: Random α	$A_2(B_{t,d}) = (1+\alpha)B_{t,d}, \alpha \in \text{random}(l,h)$
Minimum generation attack	$A_3(\min(B_{t,d})) = (1+\alpha)\max(B_{t,d})$
Peak generation attack	$A_4(B_{t,d}) = \max(B_{t,d}, B_{t+1,d})$

* $B_{t,d}$ is benign data matrix that includes the hourly generated energy per day d and hour t. α is a fractional number representing claimed additional energy. l and h are the boundaries of the minimum and maximum α.

4) MALICIOUS DATASET

There is no malicious and benign ground-truth data in the power system field. Therefore, a group of cyberattack functions is applied to a subset of the benign dataset to construct the malicious dataset (i.e., 2019). Several articles investigated possible cyberattack functions [19], [20]. In [19], the authors derived mathematical models from mimicking ET attack, assuming the adversary accessed the smart meter. In such a scenario, the adversary aims to increase its profit by increasing the reported generated energy. Table 3 summarizes the mathematical formulation of the attack functions [19]. For 2019, four attack samples are generated for each benign sample in the dataset.

There are four attacks type, partial increment attack (fixed α), partial increment attack (random α), minimum generation attack, and peak generation attack, where α is a fractional number that represents the malicious behavior that is the additional energy. A partial increment attack (fixed α) is an attack where the malicious customer adds a fixed percentage α of the generated energy to the daily reported energy, aiming to increase the injected reported energy. A partial increment attack (random α) is similar to attack type 1. However, the malicious customer adds a random percentage α of the generated energy to the daily reported energy. A minimum generation attack is when the malicious customer replaces the minimum generation reported energy during the day with partial α of the maximum generated energy. A peak generation attack is an attack where the malicious customer track the maximum generated energy during the day, then replaces all the following reported energy with the maximum generation. The severity of the attacks depends on α where there is a tradeoff between increasing the reported energy to get more profit and deviating severely from the normal pattern and being detectable by the bad data detection function on the utility company side.

Training a balanced dataset is mandatory to avoid model biases; thus, we apply the attack function only on a subset of the dataset where we capture the attack effect in different seasons. At this stage, the preprocessed samples are considered benign and labeled as 0 (negative class), and the attack functions' output samples are considered malicious and labeled as 1 (positive class). Finally, we apply the Min–Max scalar to normalize the dataset and perform a 90% random split for training and 10% testing samples, as we need more training samples. Algorithm 1 illustrates the steps we followed to generate a synthetic benign/malicious dataset. It is worth noting that the main complexity of Algorithm 1 lies in the training part, where the complexity of generating the malicious samples is $O(4N)$ with four types of attacks and N samples. On the other hand, feeding the inputs into the trained model is straightforward for testing and practical deployment, making it a better fit for resource-constrained devices.

C. ELECTRICITY THEFT DETECTOR

In this section, two detectors are designed where the first is designed to detect ET in solar distributed generation unit (DGU), and the other is utilized in solar and wind DGU. The proposed detectors are expected to identify complex patterns in data. Thus, we employ four structures for the detector based on DL models that can capture temporal features in the time-series dataset, namely, RNN, BLSTM, GRU, and bidirectional GRU (BGRU). The listed DL models are known to perform well in time-series data problems [55], [56].

1) TRAINING STAGE

Even though the feed-forward neural networks require low computational cost, it does not learn temporal features [19]. Hence, RNN models are the best candidate to handle the time-series data effectively (e.g., energy generation profiles), improving the detector's performance. However, the RNN suffers from a vanishing gradient problem, which occurs when learning from long sequences. Hence, LSTM and GRU are improved versions of the RNN, which use gates and memory blocks to solve gradient vanishing problems, making them more capable models.

Initially, we adopt a simple model structure consisting of input, hidden, and output layers. The input layer consists of N neurons equal to the number of passed sequences or features. In our study, the first layer consists of 26 neurons, the generated energy reported per hour (24 features), the day of the month, and the month for a single fuel type investigation. Similarly, we only add the fuel type feature in the hybrid environment investigation and have 27 neurons in the input

Algorithm 1: Synthesis Dataset Generation.

Input: Prepossessed energy generation profile as Benign Dataset B, Subset B as benign sample $B_{t,d}$, additional energy percent as α, lower random variable boundary as l, upper random variable boundary as h
Output: Benign/Malicious datasets as X for features, Y for labels
Attack Synthesis dataset Generation($B_{t,d}[:, 0:23], \alpha, l, h$):
for each $b \in \mathcal{B}_{t,d}$ **do**
  $A_1(B_{t,d}) = (1+\alpha)B_{t,d}$
  $A_2(B_{t,d}) = (1+\alpha)B_{t,d}, \alpha \in random(l,h)$
  $A_3(\min(B_{t,d})) = (1+\alpha)\max(B_{t,d})$
  $A_4(B_{t,d}) = \max(B_{t,d}, B_{t+1,d})$
end for
$Attack = concatenate([A_1, A_2, A_3, A_4])$ (Synthesis Malicious dataset including only 24 hours energy generation features)
$M = B_{t,d}[:, :-2]$ (additional proposed feature, month, and day of the week)
$Attackfeature = concatenate([M, M, M, M])$
$Attack = concatenate([Attack, Attack_{feature}], axis = 1)$ (Synthesis Malicious dataset including all features)
[Labeling]
$Benignlabel = zeros(length(B))$
$Attacklabel = ones(length(Attack))$
$X = concatenate(B, Attack)$
$Y = concatenate(Benignlabel, Attacklabel)$

layer. The hidden layers are based on the basic GRU or LSTM layer based on the selected model for the experiment. Finally, the output layer has a single neuron to decide the sample class. Generally, we perform the same experimental setups for each model, and then, select the best-performing model.

2) HYPERPARAMETER

The hyperparameters define the characteristic of deep learning models in the training and validation stage (e.g., learning rate, number of layers, hidden size, activation functions, and loss functions). It is worth noting that these hyperparameters can be automated using tuning algorithms such as random search or grid search. We use the state-of-the-art ET detector model as in [19], and consider similar parameters as our starting point; then, we fine tune the hyperparameter through a trial-and-error process. In each trial, we try different combinations of the grid values till we cover all possible combinations listed in Table 4, then we consider the hyperparameters set that gives the best results. It is worth noting that we follow this process for all possible models (e.g., RNN, BRNN, LSTM, BLSTM, GRU, and BGRU). For all four experimental settings, we investigate in our study the solar DG units environment with 20% ET attack variation, solar and wind DG units environment with 20% ET attack variation, solar DG units environment with 8% ET attack variation, and solar and wind DG units environment with 8% ET attack variation. Some hyperparameters are fixed for all models in all investigations. Those are the hidden size of 128, the optimizer Adam, and cross-entropy as loss function. The final hyperparameters for each model are reported in the results section.

TABLE 4. ETD Models Grid Search Hyperparameters Tuning

Models	Parameter	Grid search
LSTM, BLSTM	Batch size	4, 8, 16, 32, 64, 128
GRU, BGRU	learning rater	0.0001, 0.0005, 0.0007, 0.001
RNN,	#Layers	2, 4, 6, 8
BRNN	#epoch	250, 500, 1000, 1500, 2000, 2500

D. TRAINING AGAINST SMALL PERTURBATION

The adversary may reduce the added power percentage to make it even harder for the electricity utility companies to detect malicious behavior.

To generate smaller perturbation attack samples, we set α to 0.08 instead of 0.2 as in [19], then test the trained model with these samples to study its robustness. Furthermore, we build a smaller perturbation malicious dataset as described in Section III-B, train the model with this new dataset, and study the model robustness. The performance is compared to the previous training approaches.

E. DETECTION SCHEME WORKFLOW

The flowchart in Fig. 5 depicts the whole system flow. The ETD workflow shows three main stages referring to the ML project pipelines: data processing, model training, and model inferences. In the first stage, the data generation algorithm uses realistic power generation profiles assumed as benign data, applies attack functions to a portion of the data, and yields a benign/ malicious dataset suitable for classification problems. To prepare our data, we extract all proposed features at the top of 24 h of reported energy as the month, day of the week, and fuel type. Considering our investigation settings, we select a subset dataset to contain solar fuel type for a single fuel type investigation. While for a hybrid fuel type environment, both the solar and wind fuel types are included in the dataset. To prepare the dataset for the next stage, ML model training, we apply the Min–Max scalar and split the dataset into 10% disjoint training and testing datasets; due to the small data size, we considered a larger split for the training 90% to avoid the overfitting problem. Second, we split our dataset into batches, defining the batch size parameter (bs). We select one DL model from the list before (i.e., RNN, BRNN, LSTM, BLSTM, GRU, and BGRU). For each selected model, we initialize the same hyperparameters, such as learning rate (LR), number of hidden layers (L), and number of training epochs (ep), which are fine tuned in the next stages to enhance the model performance. Optimization parameters are fixed, where we use the Adam optimizer and cross-entropy loss function in all experiments. Finally, we perform detection

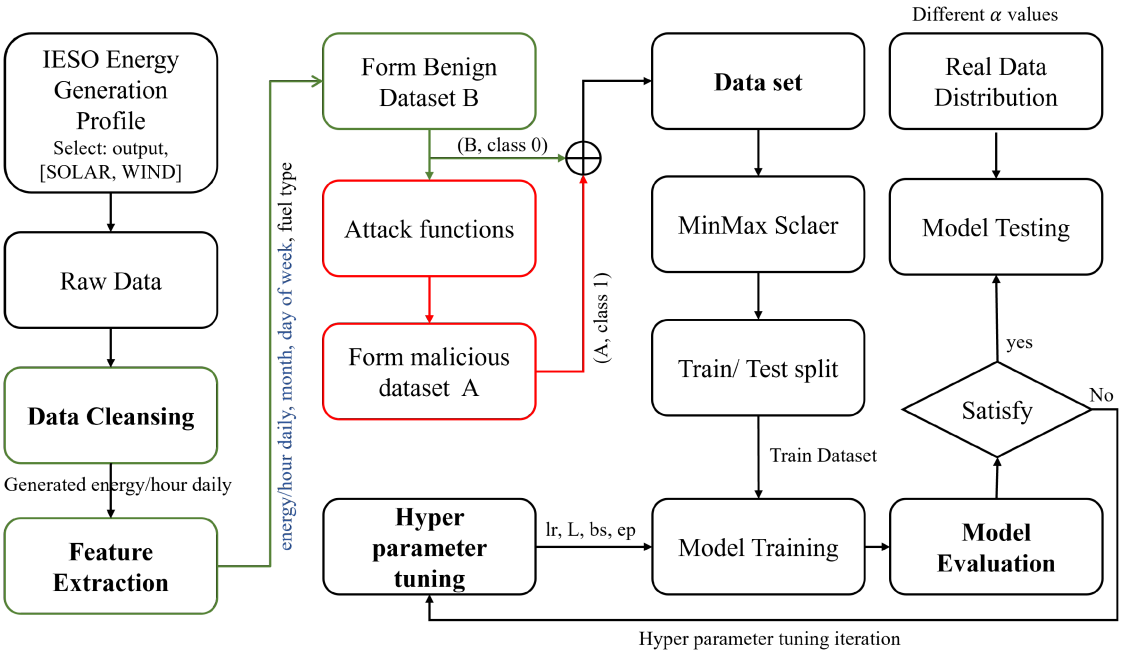

FIGURE 5. Overall workflow meant for ETA-based DL detector design.

evaluation and iterate to do hyperparameter tuning. Once we get satisfactory performance, we select the best-performing model and continue with further experiments to evaluate the model under different scenarios.

IV. EXPERIMENTAL SETUP AND EVALUATION METRICS

This section evaluates the proposed models' performances based on field evaluation matrices, as detailed in the following sections.

A. EXPERIMENTAL STEPS

Our experiments can be divided into two categories based on adversary behavior. In the first, we consider the reasonable adversary behavior training (summarized in the first two points), and in the second, we consider the smaller perturbation training, as discussed in Section III-D

For training the model with reasonable adversary behavior, we prepare the data as stated in Section III-B by setting α to 20%, then we add or subtract 5% to set the randomness boundary. We train different models with the same data settings, then test their performance from the same data distribution. Next, we create testing dataset, each is designed to mimic different variation of ETA, where we set α to smaller perturbation, i.e., $0.02 \leq \alpha \leq 0.16$. Thus, we can measure the model's effectiveness based on the adversary's behavior. We follow the same approach for training the model with smaller perturbations. We propose smaller perturbation training, where we initially set α to 8% to prepare the dataset for this investigation. Then, we test the model with different variations of ETA, i.e., $0.02 \leq \alpha \leq 0.16$, to have a fair comparison with the reasonable alpha trained model.

B. EVALUATION METRICS

Quantifying the model performance is mandatory for evaluation and comparison. Since we deal with a classification problem, we use common evaluation metrics such as F1-score and accuracy. We further use the metrics in [19] and [20] to benchmark their model performance. We use detection rate (DR), false alarm (FA), and the highest difference (HD) as evaluation metrics. In particular, these metrics have been used in [19] and [20] where DR measures the ratio of correctly detected attack samples, FA estimates the ratio of miss labeled benign samples, and HD represents the difference between the DR and FA. Considering attacks as a positive class, we can define the evaluation criteria as follow [20]:

$$\text{DR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1}$$

$$\text{FA} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{2}$$

$$\text{HD} = \text{DR} - \text{FA} \tag{3}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

$$F1 = 2 \times \frac{\text{Precision.DR}}{\text{Precision} + \text{DR}} \tag{5}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{6}$$

Furthermore, we test the robustness of the proposed approaches against small perturbation attacks. Finally, we compare the performance of the proposed DL-based detector with the state-of-the-art detectors in [19].

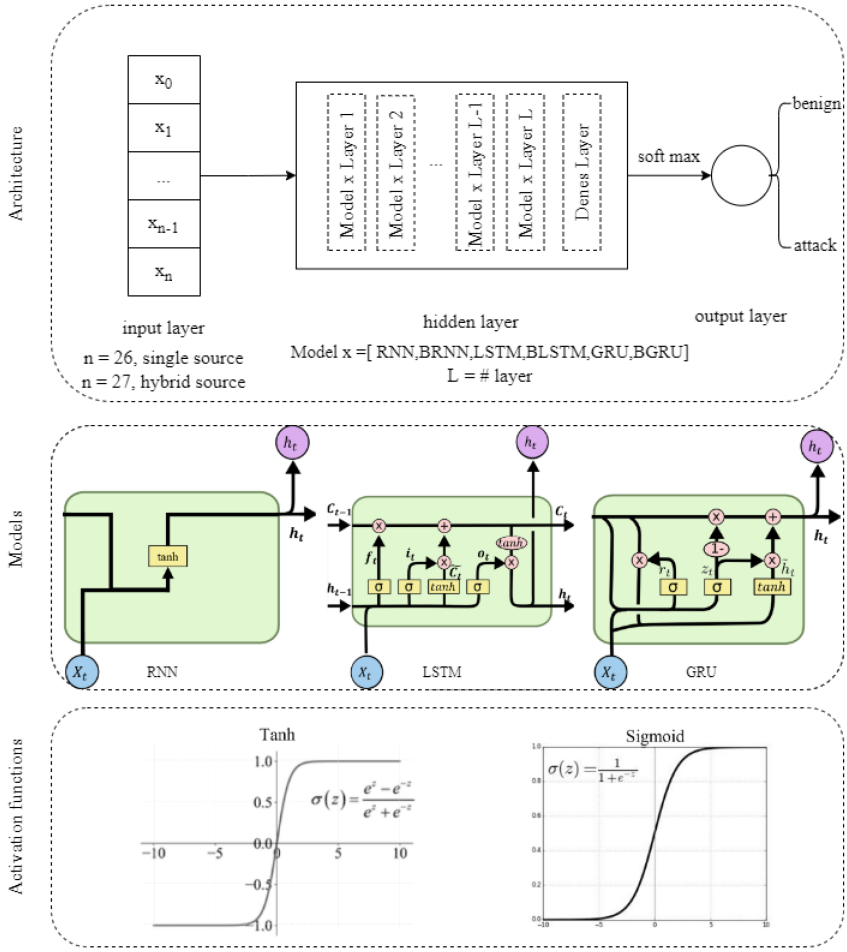

FIGURE 6. ET attack detector based on the DL model architecture.

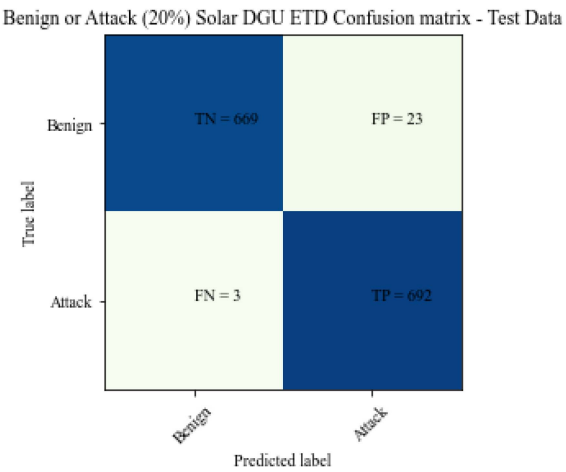

FIGURE 7. Confusion matrix of GRU-based ETD to solve the solar DGU ET issue.

V. ET DETECTION MODEL DEVELOPMENT

To implement ETD for multiple fuel types, we start by implementing the system based on a single fuel type: solar. Then, we follow the same approach after adding wind energy profiles. We investigate the detector's performance with a single source of training data, energy generation profiles [12]. We aim to determine which DL models presented with different training settings offer the best detection performance.

Time-series-like datasets need a model that learns from time features. Thus, we use RNN and its variations models and select the best-performing model based on DR and FA. We randomly initiate hyperparameters for all models, perform five tuning cycles based on their performance, and then, report the best trial. Finally, we select the bestperforming model and fine tune its hyperparameters further.

As shown in Fig. 6, we follow the same approach for single and multifuel detectors, fix the hidden size to 128, and split the random seed of train and test datasets. Then, we investigate by varying the batch size, the number of layers, the learning rate, and the number of training epochs. We consider 20% adversary behavior in this stage.

We further test the best-performing models' generalization. We try different random state seeds to split our train/test datasets to generalize our model performance. We run the detection scheme with different random state seeds five times

TABLE 5. ETD Models Hyperparameters

ETD	Model	bs	lr	L	ep
Solar reasonable α	GRU	32	0.0007	4	500
Hybrid reasonable α	BLSTM	64	0.0005	6	1000
Solar smaller α	GRU	16	0.0007	4	2000
Hybrid smaller α	BLSTM	64	0.0005	6	2500

for all conducted scenarios while keeping the same adversary behavior settings.

1) SINGLE FUEL TYPE SCENARIO

We start by selecting the solar energy profile only and dropping the fuel type feature. In this stage, we train and test the model with a reasonable $\alpha = 0.2$. Then, we apply the 10% train and test the dataset split to increase training samples. Table 6 presents different model performance. To prove the concept of our added features (i.e., month and day), we start our investigation with a classical ML algorithm. According to the literature, decision trees outperform other algorithms [25]. We first passed 24 h energy generation profile as feature space to the classifier, then added our proposed two features, month and day. As reported in Table 6, the added features improve the classifier's DR and FA. As the solar energy generation pattern is simple and relatively simple, GRU-based models perform better than advanced ones, such as LSTM-based models. The degraded RNN model performance is expected due to the gradient vanishing issue. The GRU model outperforms other DL models, with DR of 96.68% and FA of 0.43%. The confusion matrix of GRU against 20% malicious behavior attack is shown in Fig. 7. Similarly, we investigate a single fuel type source using the GRU model following the experimental settings as in Section IV, and the hyperparameters as in Table 5.

As GRU-based ETD provided the best detection rate, so we report its generalized performance. Table 7 presents our experimental results. The model detection rate varies between 94.05% and 96.68%, with an average DR of 94.84% and an interval of confidence (IoC) of $\pm 0.05611\%$. Thus, we are 95% confident that the average system performance varies between [94.78,94.90], which is satisfactory. We also highlight that the best electricity theft detector would maximize the detection rate and minimize the false alarm. It is worth mentioning that all experiments are replicated with different seeds to generalize the results and select the best-performing models.

2) HYBRID FUEL TYPE SCENARIO

In the Hybrid ETD scenario, we design the model to detect ET in solar and wind DGU. We prepare our data as in Section III-B while considering the whole dataset, including wind and solar energy profiles, and we keep the fuel type feature. We then apply the same process we follow in model selection for solar DGU ETD. As shown in Table 8, generally, the models' performance is degraded compared to the case where we used only solar DGU. At the same time, adding wind energy generation profiles increased the dataset complexity.

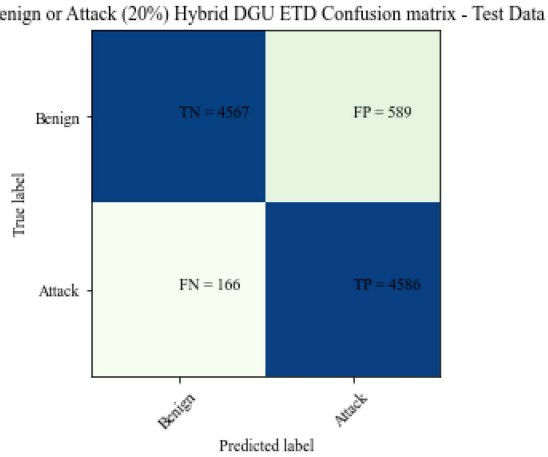

FIGURE 8. Confusion matrix of BLSTM-based ETD to solve the hybrid DGU ET issue.

This stems from the fact that we cannot define a common pattern in wind energy profiles, while solar energy generation profiles have a unique pattern that is easier to learn. Although GRU and BGRU have slightly better DR than the BLSTM model, BLSTM has much better FA. The confusion matrix of BLSTM against 20% malicious behavior attack is shown in Fig. 8. Using the settings as in Section IV, we continue our experiments and investigation for a unique hybrid detector using the BLSTM model with DR of 88.58% and FA of 3.49%. (please refer to Table 5 for hyperparameters)

As the BLSTM model outperforms other models for hybrid ETD, we report its generalized performance. Table 9 presents the experimental results. The model detection rate varies between 86.04% and 88.58%, with an average DR of 87.20% and IoC of $\pm 0.02\%$. Thus, we are 95% confident that the average system performance varies between [87.18.,87.22]. As the best ETD should maximize the detection rate and minimize the false alarm, the grid search algorithm is applied to select the best-performing hyperparameters, including the seed of model initialization.

VI. ET DETECTION PERFORMANCE EVALUATION

Using the best performing models from Section V, trained with $\alpha = 0.2$, we perform extensive testing to evaluate the designed ET detectors' robustness against smaller perturbations. We prepare the testing dataset with smaller perturbation (e.g., $\alpha = [0.16, 0.12, 0.08, 0.04, 0.02]$), and report the ETD performance for each investigation (i.e., solar, solar and wind DGU ETDs) as in Section VI-A. Furthermore, we carry out our proposed smaller perturbation training, where we train GRU and BLSTM-based model with dataset prepared according to Section III-B; however, we set the adversary behavior variable to $\alpha = 0.08$. Then, we follow the same testing approach used with reasonable α trained models to check ETD robustness against smaller perturbations. We report small perturbation-trained model performance in Section VI-B.

TABLE 6. Solar DGU ETD Different Model Performance, Where We Set $\alpha = 0.2$ and Test With Same Data Distribution

Model	DR	FA	HD	Precision	F1-score	Accuracy
Decision Tree(24)	0.9827	0.086	0.9740	0.9913	0.9869	0.9870
Decision Tree(26)	0.9841	0.0072	0.9769	0.9927	0.9884	0.9885
SVM (26)	0.5566	0.0057	0.5508	0.9896	0.7124	0.7768
RNN	0.4489	0.0087	0.4403	0.9811	0.6160	0.7195
BLSTM	0.9133	0.0245	0.8888	0.9738	0.9463	0.9445
GRU	0.9668	0.0043	0.9624	0.9955	0.9809	0.9813
BGRU	0.9480	0.0144	0.9336	0.9850	0.9661	0.9668

TABLE 7. Best Performing Model Generalization Evaluation With Five Different Random State Seeds, With $\alpha = 0.2$

Trial number	Random seed	DR	FA	HD	Precision	F1-score	Accuracy
1	10	0.9405	0.0229	0.9176	0.9759	0.9579	0.9589
2	60	0.9565	0.0115	0.9450	0.9880	0.9720	0.9726
3	7	0.9351	0.0295	0.9056	0.9707	0.9526	0.9524
4	150	0.9433	0.0176	0.9257	0.9823	0.9624	0.9625
5	42	0.9668	0.0043	0.9624	0.9955	0.9809	0.9813
Average		0.9484	0.0172	0.9313	0.9825	0.9651	0.9655
IoC: 95%		0.0006	0.0005	0.0011	0.0005	0.0005	0.0005

TABLE 8. Hybrid DGU ETD Different Model Performance, Where We Set $\alpha = 0.2$ and Test With Same Data Distribution

Model	DR	FA	HD	Precision	F1-score	Accuracy
RNN	0.8516	0.5688	0.2828	0.6190	0.7169	0.6500
LSTM	0.8530	0.0524	0.8006	0.9464	0.8973	0.8984
BLSTM	0.8858	0.0349	0.8508	0.9649	0.9237	0.9238
GRU	0.8991	0.3314	0.5677	0.7464	0.8157	0.7886
BGRU	0.9048	0.1107	0.7941	0.8987	0.9017	0.8974

A. TRAINED MODEL PERFORMANCE

The model is trained on the generated data as described in Section III-B while setting the minimum perturbation that the adversary can add to 0.2 (i.e., $\alpha = 0.2$). We apply the same setting for solar DGU ETD and hybrid DGU ETD, then tests the robustness of each as follows.

1) SINGLE FUEL TYPE SCENARIO

The GRU-based detector shows a better detection rate of 96.68% when malicious customers claim 20% additional power, given that the model is trained on 20% malicious behavior. However, the GRU-based ETD acts differently with smaller perturbations. Table 10 shows the model performance in the inference stage while we vary the severity of malicious behavior samples in the testing dataset. The model's performance drops when a smaller perturbation is present in the testing set. As model detection rate goes from 96.6% to 66.04%, when varying α between 20% and 2%. We use smaller perturbation training to enhance the system robustness against such malicious behavior. We train the model for 500 epochs, with a learning rate of 0.0007, while passing a batch size of 32 to our four layers GRU model. We fixed the hidden size to 128 and used the Adam optimizer and cross-entropy loss function.

2) HYBRID FUEL TYPE SCENARIO

The Hybrid ETD problem is more complex than Solar ETD, but BLSTM reached a DR of 88.60% to detect 20% claimed additional energy. Given that the model trained on 20% malicious behavior, the model performance dropped further with smaller perturbations. As shown in Table 11, the model's performance dropped when smaller perturbations were present in the testing set. As the model detection rate goes from 88.60% to 83.48%, the model's accuracy generally drops from 92.39% to 78.60%. We further investigate the impact of injecting smaller perturbations on the system during the training phase. We train the model for 2000 epochs, with a learning rate of 0.0005, while passing a batch size of 64 to our six layers BLSTM model. We also fixed the hidden size to 128 and used the Adam optimizer and cross-entropy loss function.

B. SMALL PERTURBATION PERFORMANCE

It is worth emphasizing that the trained model with reasonable alpha (i.e., $\alpha \geq 0.20$) cannot efficiently detect the smaller perturbation on the malicious data. As a result, we investigate the effect of smaller perturbations during training on the detector's behavior. This method entails training GRU and BLSTM models on the generated data, as described in Section III-B while setting the malicious behavior (i.e., the perturbations) $\alpha = 0.08$. We apply the same settings for solar DGU ETD and hybrid DGU ETD, then test the robustness of each as follows.

1) SINGLE FUEL TYPE SCENARIO

The GRU-trained model with $\alpha = 0.2$ has an unstable detection rate against smaller perturbations. On the other hand, the model trained with small perturbations shows more robustness against the variation of such malicious behavior. The confusion matrix of GRU against 8% malicious behavior attack is shown in Fig. 9. As shown in Table 12, the model is trained using $\alpha = 0.08$, and then, tested with different variations of the malicious behavior represented as an additional percentage, α. We can see that the model performance is stable against smaller perturbations, outperforms the reasonable α trained model, and is still effective for reasonable α values

TABLE 9. Best Performing Model Generalization Evaluation With Five Different Random State Seeds, With $\alpha = 0.2$

Trial number	Random seed	DR	FA	HD	Precision	F1-score	Accuracy
1	10	0.8746	0.0397	0.8349	0.9597	0.9152	0.9157
2	60	0.8672	0.0543	0.8129	0.9455	0.9046	0.9048
3	7	0.8604	0.0594	0.8011	0.9403	0.8986	0.8989
4	150	0.8720	0.0479	0.8241	0.9501	0.9094	0.9112
5	42	0.8858	0.0349	0.8508	0.9649	0.9237	0.9238
Average		0.8720	0.0472	0.8247	0.9521	0.9103	0.9109
IoC: 95%		0.0002	0.0002	0.0005	0.0002	0.0120	0.0119

TABLE 10. Solar Model Trained With $\alpha = 0.2$ Performance Under Small Perturbation

Testing settingα	l	h	DR	FA	HD	Precision	F1-score	Accuracy
0.2	**0.15**	**0.25**	0.9667	0.0057	0.9610	0.9941	0.9802	0.9805
0.16	0.11	0.12	0.9219	0.0129	0.9090	0.9860	0.9529	0.9545
0.12	0.07	0.17	0.9277	0.0215	0.9061	0.9772	0.9518	0.9531
0.08	0.06	0.1	0.7153	0.0230	0.6922	0.9686	0.8229	0.8464
0.04	0.02	0.09	0.6864	0.0388	0.6475	0.9462	0.7956	0.8241
0.02	0.015	0.07	0.6604	0.0446	0.6158	0.9364	0.7746	0.8082

Training parameters are in bold.

TABLE 11. Hybrid Model Trained With $\alpha = 0.2$ Performance Under Small Perturbation

Testing settingα	l	h	DR	FA	HD	Precision	F1-score	Accuracy
0.2	**0.15**	**0.25**	0.8860	0.0349	0.8510	0.9649	0.9240	0.9239
0.16	0.11	0.12	0.8662	0.0509	0.8152	0.9486	0.9055	0.9059
0.12	0.07	0.17	0.8625	0.0703	0.7922	0.9301	0.8950	0.8947
0.08	0.03	0.13	0.8392	0.1511	0.6881	0.8577	0.8483	0.8439
0.04	0.02	0.09	0.8365	0.2321	0.6044	0.7963	0.8159	0.8036
0.02	0.015	0.07	0.8348	0.2668	0.5679	0.7724	0.8024	0.7860

Training parameters are in bold.

TABLE 12. Solar DGU ETD Model Trained With $\alpha = 0.08$ Performance Under Small and Reasonable Perturbation

Testing settingα	l	h	DR	FA	HD	Precision	F1-score	Accuracy
0.2	0.15	0.25	0.9161	0.0604	0.8557	0.9378	0.9269	0.9279
0.16	0.11	0.12	0.9421	0.0302	0.9119	0.9687	0.9553	0.9560
0.12	0.07	0.17	0.9393	0.0705	0.8688	0.9289	0.9346	0.9343
0.08	**0.06**	**0.1**	0.9421	0.0230	0.9191	0.9760	0.9588	0.9596
0.04	0.02	0.09	0.9364	0.0604	0.8759	0.9391	0.9377	0.9379
0.02	0.015	0.07	0.8973	0.0705	0.8268	0.9268	0.9119	0.9135

Training parameters are in bold.

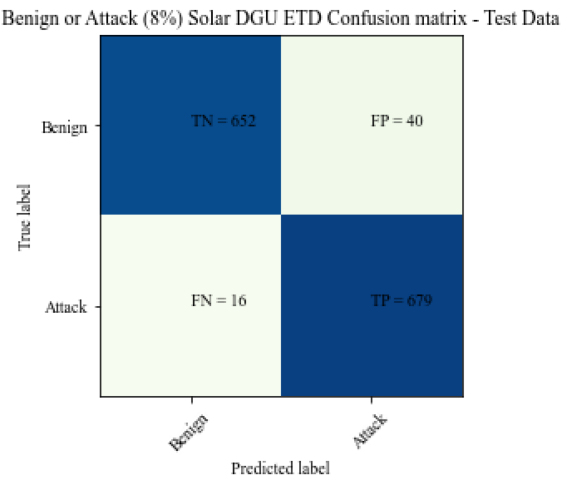

FIGURE 9. Confusion matrix of GRU-based ETD to solve the solar DGU ET issue, smaller perturbation trained model.

attack. Generally, this model gives more stable results than the $\alpha = 0.2$ model for smaller perturbations. That is because the model accuracy did not drop below 90%, and varied between 91.35% and 95.96%, which supports our hypothesis that the model should be trained against smaller perturbations. It is important to highlight that we train the model for 1000 epochs and use a batch size of 16, for our four layers GRU model, with the same learning rate and other hyperparameters. Fig. 10 visualizes our findings and proves our hypothesis, where the proposed training settings model, smaller perturbation ET attack training, shows enhanced and more robust performance than the typical training setting model, solving the solar DGU ET issue. The smaller perturbation train detector, in blue, shows a stable trend compared to the 20% malicious behavior trained detector, in green, which shows a decaying trend.

2) HYBRID FUEL TYPE SCENARIO

The optimized BLSTM model shows a considerable DR drop when tested by smaller perturbation. On the other hand,

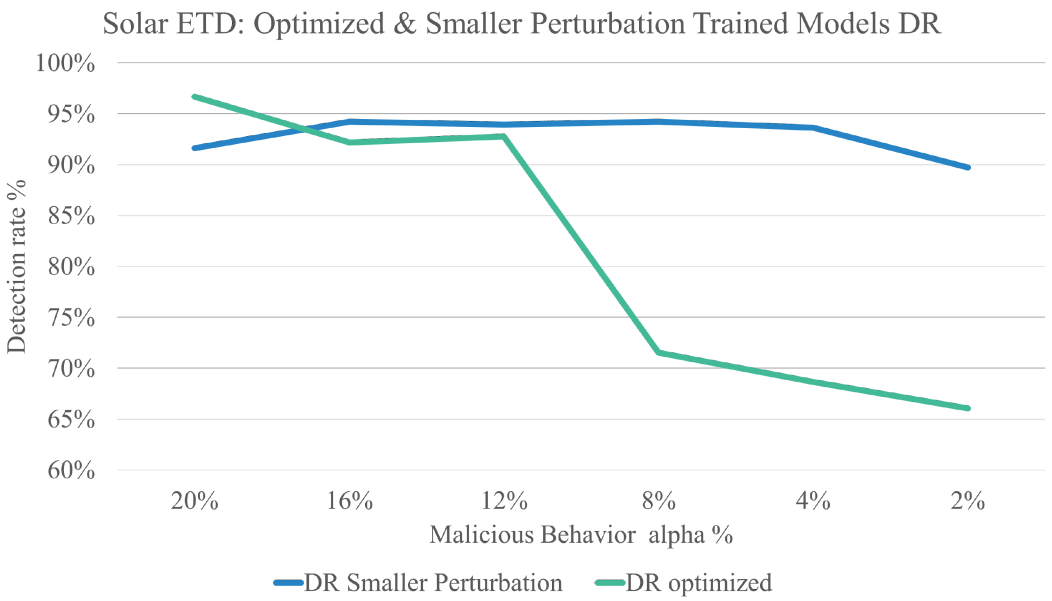

FIGURE 10. Detection rate compression between optimized malicious behavior trained model, and proposed small perturbation trained model versus different percentage, solving the solar DGU ET issue.

TABLE 13. Hybrid DGU ETD With $\alpha = 0.08$ Performance Under Small and Reasonable Perturbation

Testing settingα	l	h	DR	FA	HD	Precision	F1-score	Accuracy
0.2	0.15	0.25	0.8679	0.1907	0.6773	0.8316	0.8494	0.8398
0.16	0.11	0.12	0.8660	0.1014	0.7646	0.9026	0.8839	0.8816
0.12	0.07	0.17	0.8656	0.0814	0.7842	0.9202	0.8921	0.8910
0.08	**0.06**	**0.1**	0.8567	0.0476	0.8091	0.9513	0.9015	0.9026
0.04	0.02	0.09	0.8533	0.1366	0.7187	0.8717	0.8634	0.8592
0.02	0.015	0.07	0.8559	0.1978	0.6581	0.8244	0.8399	0.8301

Training parameters are in bold.

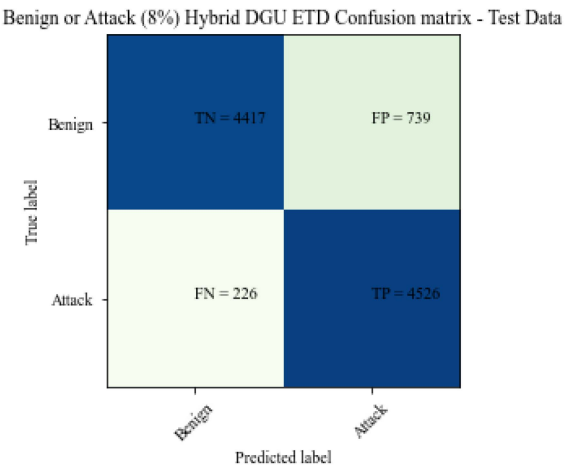

FIGURE 11. Confusion matrix of BLSTM-based ETD to solve the hybrid DGU ET issue, smaller perturbation trained model.

the trained BLSTM with $\alpha = 0.08$ shows different behavior against such attacks. The confusion matrix of BLSTM against 8% malicious behavior attack is shown in Fig. 11. Table 13 presents the testing results of the small perturbation $\alpha = 0.08$ trained model against the different values for α. In contrast to larger perturbation attacks, the model's performance is stable for reasonable α values. Generally, BLSTM trained with $\alpha = 0.08$ improved the detection rate against smaller perturbations compared to the model trained with reasonable malicious behavior $\alpha = 0.2$. Fig. 12 visualizes our findings and proves our hypothesis, where the proposed training setting shows enhanced and more robust performance than the typical training setting model, solving the solar and wind DGU ET issue. The smaller perturbation train detector, in blue, shows a stable trend compared to the 20% malicious behavior trained detector, in green, which shows a decaying trend.

C. BENCHMARKING AND DISCUSSION

We benchmark our results with the state-of-the-art model in [19], where they trained and tested their model on benign and malicious datasets from different sources. The authors in [19] used only the solar energy generation profiles, achieved a DR of 94.6% and a high false alarm of 26%, to detect 20% malicious behavior. Although their best performing model (i.e., the three data source-based models) achieved a higher detection rate of 99.3%, its FA reaches 22% and requires a lot of data resources. As shown in Table 14, our proposed approaches achieve a 96.67% DR, which

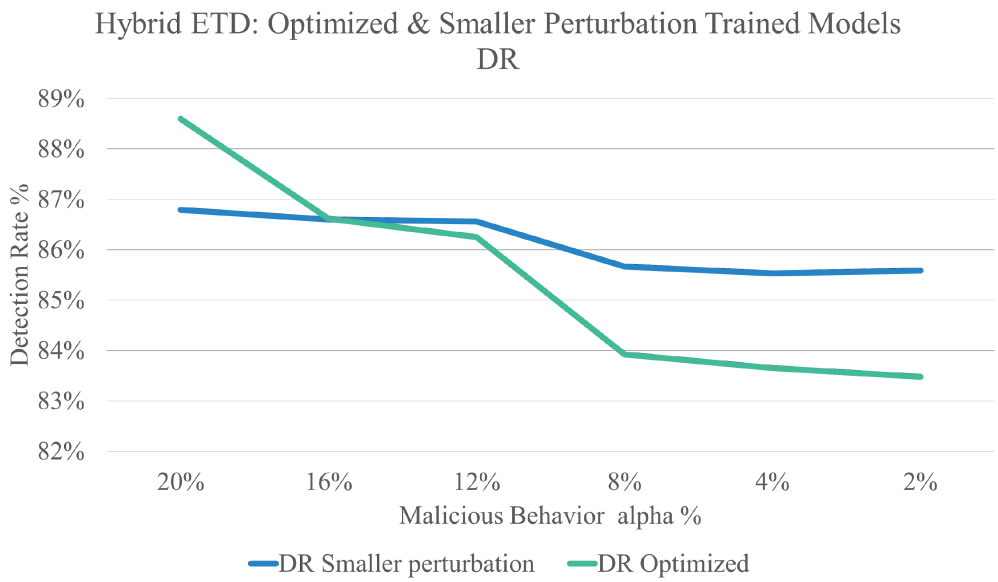

FIGURE 12. Detection rate compression between optimized malicious behavior trained model, and proposed small perturbation trained model versus different percentage, solving solar and wind DGU ET issue.

TABLE 14. Model Performance With Balanced Data Compared With State of Art Study, Testing $\alpha = 20\%$ [19]

Model	DR	FA	HD	F1-score
C-GRU-RNN	0.946	0.26	0.974	0962
Our detector (α=0.2)	0.9667	0.0057	0.961	0.9802
Our detector (α=0.08)	0.9161	0.0604	0.8557	0.9269

outperforms the state-of-the-art one-source model. In [18], a regression classification approach achieved an average accuracy of 95.3%, using $\alpha = 0.1$. The same approach was used in [20] and achieved a 91.5% accuracy. Both articles had different settings than ours and [19], but they both addressed the same problem. Overall, the proposed detectors show outstanding performance in identifying the ETA, with a high detection rate for single or multiple fuel types.

VII. CONCLUSION

This article investigated FIT generation ET attack detection in renewable energy-based DG units. We proposed efficient detectors with high detection rates, and we used a realistic energy profile dataset that includes the reported injected energy for three years. The main limitation is the lack of a benchmark dataset, where there are no benign and attack samples for ET attacks in the generation domain. We have solved this limitation by generating synthetic attack samples. We used a series of cyberattack functions to compromise smart meter data, thus attacking the integrity of the injected power readings from the DG devices. Two features were selected to enhance the system detection and address the data variation during the seasons. We use the GRU-RNN and BLSTM-RNN models as they are more suited for pattern learning when the data have a temporal relationship. We showed that 20% malicious behavior attacks are easy to detect, and a model trained with such a percentage cannot detect smaller perturbation attacks. Therefore, we proposed small perturbation malicious behavior attack training. We started our investigation with solar fuel type data and designed a GRU-RNN DL-based ETD. The results showed that the model offers outstanding performance with the highest detection rate, using the energy generation profile of one data source. Then, we designed a unique model that is used to solve wind and solar DG units' electricity theft issues by utilizing the BLSTM model. Our results from both investigations proved that smaller perturbation-trained models have more robust performance than the models trained with reasonable malicious behavior regarding ETA variation. The main limitation of the proposed scheme is that the performance of the detector decreases against zero-day attacks or unseen attacks, which is a common limitation among supervised learning approaches deployed in ETA detection. A research direction with more generator types and the ensemble method will be interesting for future extensions.

ACKNOWLEDGMENT

The findings herein reflect the work, and are solely the responsibility, of the authors. Open Access funding provided by the Qatar National Library.

REFERENCES

[1] M. Trabelsi, H. Vahedi, and H. Abu-Rub, "Review on single-dc-source multilevel inverters: Topologies, challenges, industrial applications, and recommendations," *IEEE Open J. Ind. Electron. Soc.*, vol. 2, pp. 112–127, Jan. 2021.

[2] A. H. Al-Badi, R. Ahshan, N. Hosseinzadeh, R. Ghorbani, and E. Hossain, "Survey of smart grid concepts and technological demonstrations worldwide emphasizing on the Oman perspective," *Appl. Syst. Innov.*, vol. 3, no. 5, p. 5, 2020, doi: 10.3390/asi3010005.

[3] S. Tan, D. De, W.-Z. Song, J. Yang, and S. K. Das, "Survey of security advances in smart grid: A data driven approach," *IEEE Commun. Surv. Tuts.*, vol. 19, no. 1, pp. 397–422, Oct. 2016.

[4] D. Pliatsios, P. Sarigiannidis, T. Lagkas, and A. G. Sarigiannidis, "A survey on SCADA systems: Secure protocols, incidents, threats and tactics," *IEEE Commun. Surv. Tut.*, vol. 22, no. 3, pp. 1942–1976, Jul.–Sep. 2020.

[5] M. De Vivo, G. O. de Vivo, R. Koeneke, and G. Isern, "Internet vulnerabilities related to TCP/IP and T/TCP," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 29, no. 1, pp. 81–85, 1999.

[6] P. I. R. Grammatikis, P. G. Sarigiannidis, and I. D. Moscholios, "Securing the Internet of Things: Challenges, threats and solutions," *Internet Things*, vol. 5, pp. 41–70, 2019.

[7] D. Serpanos and T. Komninos, "The cyberwarfare in Ukraine," *Computer*, vol. 55, no. 7, pp. 88–91, 2022.

[8] Z. El Mrabet, N. Kaabouch, H. El Ghazi, and H. El Ghazi, "Cyber-security in smart grid: Survey and challenges," *Comput. Elect. Eng.*, vol. 67, pp. 469–482, 2018.

[9] M. Wang and B. Xu, "Observer-based guaranteed cost control of cyber-physical systems under dos jamming attacks," *Eur. J. Control*, vol. 48, pp. 21–29, 2019.

[10] J. Shi, S. Liu, B. Chen, and L. Yu, "Distributed data-driven intrusion detection for sparse stealthy FDI attacks in smart grids," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 68, no. 3, pp. 993–997, Mar. 2021.

[11] P. Jokar, N. Arianpoo, and V. C. Leung, "Electricity theft detection in ami using customers' consumption patterns," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 216–226, Jan. 2016.

[12] Y. Tang, C.-W. Ten, and K. P. Schneider, "Inference of tampered smart meters with validations from feeder-level power injections," in *Proc. IEEE Innovative Smart Grid Technologies-Asia*, 2019, pp. 2783–2788.

[13] K. Cory, T. Couture, and C. Kreycik, "Feed-in tariff policy: Design, implementation, and RPS policy interactions," National Renewable Energy Lab., Golden, CO, USA, Tech. Rep., 2009.

[14] S. Seme, K. Sredensek, and Z. Praunseis, "Smart grids and net metering for photovoltaic systems," in *Proc. IEEE Int. Conf. Modern Elect. Energy Syst.*, 2017, pp. 188–191.

[15] V. B. Krishna, C. A. Gunter, and W. H. Sanders, "Evaluating detectors on optimal attack vectors that enable electricity theft and DER fraud," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 4, pp. 790–805, Aug. 2018.

[16] B. Li, R. Lu, and G. Xiao, *Detection of False Data Injection Attacks in Smart Grid Cyber-Physical Systems*. Berlin, Germany: Springer, 2020.

[17] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Inf. Syst. Secur.*, vol. 14, no. 1, pp. 1–33, 2011.

[18] D. Xue, X. Jing, and H. Liu, "Detection of false data injection attacks in smart grid utilizing elm-based OCON framework," *IEEE Access*, vol. 7, pp. 31762–31773, 2019.

[19] M. Ismail, M. F. Shaaban, M. Naidu, and E. Serpedin, "Deep learning detection of electricity theft cyber-attacks in renewable distributed generation," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3428–3437, Jul. 2020.

[20] M. Shaaban, U. Tariq, M. Ismail, N. A. Almadani, and M. Mokhtar, "Data-driven detection of electricity theft cyberattacks in PV generation," *IEEE Syst. J.*, vol. 16, no. 2, pp. 3349–3359, Jun. 2022.

[21] N. Bhusal, M. Gautam, and M. Benidris, "Detection of cyber attacks on voltage regulation in distribution systems using machine learning," *IEEE Access*, vol. 9, pp. 40402–40416, 2021.

[22] M. Ismail, M. Shahin, M. F. Shaaban, E. Serpedin, and K. Qaraqe, "Efficient detection of electricity theft cyber attacks in AMI networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2018, pp. 1–6.

[23] [Online]. Available: https://www.ieso.ca/en/Power-Data/Data-Directory

[24] R. Punmiya and S. Choe, "Energy theft detection using gradient boosting theft detector with feature engineering-based preprocessing," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2326–2329, Jan. 2019.

[25] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, "Decision tree and SVM-based data analytics for theft detection in smart grid," *IEEE Trans. Ind. Inform.*, vol. 12, no. 3, pp. 1005–1016, Jun. 2016.

[26] Y. He, G. J. Mendis, and J. Wei, "Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism," *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2505–2516, Sep. 2017.

[27] T. S. Murthy, N. Gopalan, and V. Ramachandran, "A naive bayes classifier for detecting unusual customer consumption profiles in power distribution systems-APSPDCL," in *Proc. IEEE 3rd Int. Conf. Inventive Syst. Control*, 2019, pp. 673–678.

[28] S. Li, Y. Han, X. Yao, S. Yingchen, J. Wang, and Q. Zhao, "Electricity theft detection in power grids with deep learning and random forests," *J. Elect. Comput. Eng.*, vol. 2019, 2019.

[29] R. Wu, L. Wang, and T. Hu, "Adaboost-SVM for electrical theft detection and GRNN for stealing time periods identification," in *Proc. IEEE 44th Annu. Conf. IEEE Ind. Electron. Soc.*, 2018, pp. 3073–3078.

[30] S. K. Singh, R. Bose, and A. Joshi, "Pca based electricity theft detection in advanced metering infrastructure," in *Proc. IEEE 7th Int. Conf. Power Syst.*, 2017, pp. 441–445.

[31] J. Yeckle and B. Tang, "Detection of electricity theft in customer consumption using outlier detection algorithms," in *Proc. IEEE 1st Int. Conf. Data Intell. Secur.*, 2018, pp. 135–140.

[32] F. Xiao and Q. Ai, "Electricity theft detection in smart grid using random matrix theory," *IET Gener., Transmiss. Distrib.*, vol. 12, no. 2, pp. 371–378, 2018.

[33] Y. Liu and S. Hu, "Cyberthreat analysis and detection for energy theft in social networking of smart homes," *IEEE Trans. Comput. Social Syst.*, vol. 2, no. 4, pp. 148–158, Dec. 2015.

[34] V. Badrinath Krishna, R. K. Iyer, and W. H. Sanders, "ARIMA-based modeling and validation of consumption readings in power grids," in *Proc. Int. Conf. Crit. Inf. Infrastructures Secur.*, 2015, pp. 199–210.

[35] D. Yao, M. Wen, X. Liang, Z. Fu, K. Zhang, and B. Yang, "Energy theft detection with energy privacy preservation in the smart grid," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7659–7669, Oct. 2019.

[36] A. Takiddin, M. Ismail, U. Zafar, and E. Serpedin, "Deep autoencoder-based anomaly detection of electricity theft cyberattacks in smart grids," *IEEE Syst. J.*, vol. 16, no. 3, pp. 4106–4117, Sep. 2022.

[37] M. Ezeddin, A. Albaseer, M. Abdallah, S. Bayhan, M. Qaraqe, and S. Al-Kuwari, "Efficient deep learning based detector for electricity theft generation system attacks in smart grid," in *Proc. 3rd Int. Conf. Smart Grid Renewable Energy*, 2022, pp. 1–6.

[38] X. Yuan, P. He, Q. Zhu, and X. Li, "Adversarial examples: Attacks and defenses for deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2805–2824, Sep. 2019.

[39] D. Wang et al., "Daedalus: Breaking nonmaximum suppression in object detection via adversarial examples," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 7427–7440, Aug. 2022.

[40] G. Zhang, C. Yan, X. Ji, T. Zhang, T. Zhang, and W. Xu, "Dolphinattack: Inaudible voice commands," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2017, pp. 103–117.

[41] J. Li, R. Ji, H. Liu, X. Hong, Y. Gao, and Q. Tian, "Universal perturbation attack against image retrieval," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 4899–4908.

[42] T. Zheng, C. Chen, and K. Ren, "Distributionally adversarial attack," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 2253–2260.

[43] Z. Yan and H. Wen, "Electricity theft detection base on extreme gradient boosting in AMI," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 2504909-1–2504909-9, Jan. 2021.

[44] A. Takiddin, M. Ismail, U. Zafar, and E. Serpedin, "Robust electricity theft detection against data poisoning attacks in smart grids," *IEEE Trans. Smart Grid*, vol. 12, no. 3, pp. 2675–2684, May 2021.

[45] M. Nabil, M. Mahmoud, M. Ismail, and E. Serpedin, "Deep recurrent electricity theft detection in AMI networks with evolutionary hyper-parameter tuning," in *Proc. IEEE Int. Conf. Internet Things, IEEE Green Comput. Commun., IEEE Cyber, Phys. Soc. Comput., IEEE Smart Data*, 2019, pp. 1002–1008.

[46] M. Nabil, M. Ismail, M. Mahmoud, M. Shahin, K. Qaraqe, and E. Serpedin, "Deep learning-based detection of electricity theft cyber-attacks in smart grid ami networks," in *Deep Learning Applications for Cyber Security*. Berlin, Germany: Springer, 2019, pp. 73–102.

[47] A. Takiddin, M. Ismail, M. Nabil, M. M. Mahmoud, and E. Serpedin, "Detecting electricity theft cyber-attacks in AMI networks using deep vector embeddings," *IEEE Syst. J.*, vol. 15, no. 3, pp. 4189–4198, Sep. 2021.

[48] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and M. Mohamad, "Nontechnical loss detection for metered customers in power utility using support vector machines," *IEEE Trans. Power Del.*, vol. 25, no. 2, pp. 1162–1171, Apr. 2010.

[49] E. W. S. Angelos, O. R. Saavedra, O. A. C. Cortés, and A. N. De Souza, "Detection and identification of abnormalities in customer consumptions in power distribution systems," *IEEE Trans. Power Del.*, vol. 26, no. 4, pp. 2436–2442, Oct. 2011.
[50] C. C. O. Ramos, A. N. de Sousa, J. P. Papa, and A. X. Falcao, "A new approach for nontechnical losses detection based on optimum-path forest," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 181–189, Feb. 2011.
[51] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and F. Nagi, "Improving SVM-based nontechnical loss detection in power utility using the fuzzy inference system," *IEEE Trans. Power Del.*, vol. 26, no. 2, pp. 1284–1285, Apr. 2011.
[52] A. Nizar, Z. Dong, and Y. Wang, "Power utility nontechnical loss analysis with extreme learning machine method," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 946–955, Aug. 2008.
[53] X. Yuan, M.-g. Shi, and Z. Sun, "Research of electricity stealing identification method for distributed PV based on the least squares approach," in *Proc. IEEE 5th Int. Conf. Electric Utility Deregulation Restructuring Power Technol.*, 2015, pp. 2471–2474.
[54] M. M. Othman, H. M. Ahmed, M. H. Ahmed, and M. M. Salama, "A techno-economic approach for increasing the connectivity of photovoltaic distributed generators," *IEEE Trans. Sustain. Energy*, vol. 11, no. 3, pp. 1848–1857, Jul. 2020.
[55] J. Zheng, C. Xu, Z. Zhang, and X. Li, "Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network," in *Proc. IEEE 51st Annu. Conf. Inf. Sci. Syst.*, 2017, pp. 1–6.
[56] Y. Wang, W. Shi, Q. Jin, and J. Ma, "An accurate false data detection in smart grid based on residual recurrent neural network and adaptive threshold," in *Proc. IEEE Int. Conf. Energy Internet*, 2019, pp. 499–504.

MAYMOUNA EZ EDDIN received the B.Sc. degree in electrical engineering from Qatar University, Doha, Qatar, in 2020, and the M.Sc. degree in data science and engineering from Hamad Bin Khalifa University, Doha, in 2022. She is currently working toward the Ph.D. degree in electrical and computer engineering with Texas A&M University, College Station, TX, USA.

She worked as a Research Assistant with Qatar University from 2020 to 2022, with the Qatar University Machine Learning Group. She is currently working as an Associate Research Assistant with Texas A&M University at Qatar, Doha. Her research interests include the application of deep learning and machine learning in smart grid security, and healthcare.

ABDULLATIF ALBASEER (MEMBER, IEEE) received the M.Sc. degree in computer networks from King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia, in 2017, and the Ph.D. degree in computer science and engineering from Hamad Bin Khalifa University, Doha, Qatar, in 2022.

He is a Postdoctoral Research Fellow with the Smart Cities and IoT Lab, Hamad Bin Khalifa University. He has authored and co-authored more than 15 conference and journal papers in IEEE International Conference on Communications, Globecom, and IEEE transactions and invented five patents in the area of federated learning and wireless network edge. His current research interests include federated learning over wireless network edge, IoT, smart cities, and cybersecurity in smart grid.

MOHAMED ABDALLAH (SENIOR MEMBER, IEEE) received the B.Sc. degree in electrical and computer engineering from Cairo University, Giza, Egypt, in 1996, and the M.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Maryland at College Park, College Park, MD, USA, in 2001 and 2006, respectively.

From 2006 to 2016, he held academic and research positions with Cairo University and Texas A & M University in Qatar, Doha, Qatar. He is currently a Founding Faculty Member with the rank of Associate Professor with the College of Science and Engineering, Hamad Bin Khalifa University, Doha. He has authored and co-authored more than 150 journals and conferences and four book chapters and coinvented four patents. His current research interests include wireless networks, wireless security, smart grids, optical wireless communication, and blockchain applications for emerging networks.

Dr. Abdallah was the recipient of the Research Fellow Excellence Award at Texas A& M University in Qatar in 2016, the Best Paper Award in multiple IEEE conferences, including IEEE BlackSeaCom 2019 and the IEEE First Workshop on Smart Grid and Renewable Energy in 2015, and the Nortel Networks Industrial Fellowship for five consecutive years, 1999–2003. His professional activities include an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS and theIEEE OPEN ACCESS JOURNAL OF COMMUNICATIONS, the Track Co-Chair of the IEEE VTC Fall 2019 Conference, the Technical Program Chair of the 10th International Conference on Cognitive Radio-Oriented Wireless Networks, and a Technical Program Committee Member of several major IEEE conferences.

SERTAC BAYHAN (SENIOR MEMBER, IEEE) received the B.S. degree, M.S. degree and Ph.D. degrees in electrical engineering as valedictorian and engineering degrees from Gazi University, Ankara, Turkey, in 2006, 2008 and 2012, respectively.

In 2008, he joined the Electronics and Automation Engineering Department, Gazi University, as a Lecturer, where he was promoted to an Assistant Professor and Associate Professor in 2013 and 2017, respectively. From 2014 to 2018, he also worked with Texas A&M University at Qatar, as an Associate Research Scientist. He is currently working with Qatar Environment and Energy Research Institute (QEERI), Doha, Qatar, as a Scientist. He has authored 150 high-impact journal and conference papers. He is the coauthor of two books and three book chapters. His research interests include the areas of advanced control of photovoltaic systems, microgrids, and smart grid applications.

Prof. Bayhan has led several international projects with collaborators from all over the world. Because of the visibility of his research, he has been elected as a Chair of IES Power Electronics Technical Committee. He is currently an Associate Editor for the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, IEEE JOURNAL OF EMERGING AND SELECTED TOPICS IN INDUSTRIAL ELECTRONICS, and a Guest Editor for the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS.

MARWA K. QARAQE (SENIOR MEMBER, IEEE) received the bachelor's degree with Summa Cum Laude in electrical engineering from Texas A&M University at Qatar, Doha, Qatar, in 2010, the Master of Science and Ph.D. degrees in electrical engineering from Texas A&M University, College Station, TX, USA, in 2012 and 2016, respectively.

She is currently an Associate Professor with Hamad Bin Khalifa University, Doha. Her research interests include wireless communication, signal processing, and machine learning, and their application in multidisciplinary fields, including but not limited to security, Internet of Things, and health. His main research interests include physical layer security, federated learning over wireless networks, and machine learning for wireless communication, security, and health.

SAIF AL-KUWARI (SENIOR MEMBER, IEEE) received the Bachelor of Engineering degree in computers and networks from the University of Essex, Essex, U.K., in 2006 and the two Ph.D. degrees in computer science from the University of Bath, Bath, U.K. and Royal Holloway, University of London, Egham, U.K., both in 2012.

He is currently an Assistant Professor with the College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar. His research interests include applied cryptography, quantum computing, computational forensics, and their connections with machine learning.

Dr. Al-Kuwari is a Fellow of the Institution of Engineering and Technology and British Computer Society, and a Senior Member of the Association for Computing Machinery.

HAITHAM ABU-RUB (FELLOW, IEEE) received the M.Sc. degree in electrical engineering from the Gdynia Marine Academy, Gdynia, Poland, in 1990, and the Ph.D. degree in electrical engineering from the Gdansk University of Technology, Gdansk, Poland, in 1995.

Since 2006, he has been with the Texas A&M University at Qatar, Doha, Qatar, where he is currently a Professor and the Managing Director of the Smart Grid Center Extension. He is currently leading many projects on photovoltaic and hybrid renewable power generation systems with different types of converters and electric drives. He has authored more than 300 journal and conference papers and has earned and supervised many research projects. He is the coauthor of four books, two of which are published by Wiley. He is also an author or coauthor of five book chapters. His research interests include energy conversion systems, including electric drives, power electronic converters, renewable energy, and smart grid.

Dr. Abu-Rub was the recipient of many prestigious international awards, such as the American Fulbright Scholarship, the German Alexander von Humboldt Fellowship, the German DAAD Scholarship, and the British Royal Society Scholarship. He is also an Editor for several IEEE journals.