energies

MDPI

Review

Review of the Data-Driven Methods for Electricity Fraud Detection in Smart Metering Systems

Mahmoud M. Badr [1,2], Mohamed I. Ibrahem [2,3], Hisham A. Kholidy [1], Mostafa M. Fouda [4,5] and Muhammad Ismail [6,*]

[1] Department of Network and Computer Security, College of Engineering, SUNY Polytechnic Institute, Utica, NY 13502, USA; badrm@sunypoly.edu (M.M.B.); kholidh@sunypoly.edu (H.A.K.)
[2] Department of Electrical Engineering, Faculty of Engineering at Shoubra, Benha University, Cairo 11672, Egypt; mibrahem@gmu.edu
[3] Department of Cyber Security Engineering, George Mason University, Fairfax, VA 22030, USA
[4] Department of Electrical and Computer Engineering, College of Science and Engineering, Idaho State University, Pocatello, ID 83209, USA; mfouda@ieee.org
[5] Center for Advanced Energy Studies (CAES), Idaho Falls, ID 83401, USA
[6] Department of Computer Science, Tennessee Technological University, Cookeville, TN 38501, USA
* Correspondence: mismail@tntech.edu

Abstract: In smart grids, homes are equipped with smart meters (SMs) to monitor electricity consumption and report fine-grained readings to electric utility companies for billing and energy management. However, malicious consumers tamper with their SMs to report low readings to reduce their bills. This problem, known as electricity fraud, causes tremendous financial losses to electric utility companies worldwide and threatens the power grid's stability. To detect electricity fraud, several methods have been proposed in the literature. Among the existing methods, the data-driven methods achieve state-of-art performance. Therefore, in this paper, we study the main existing data-driven electricity fraud detection methods, with emphasis on their pros and cons. We study supervised methods, including wide and deep neural networks and multi-data-source deep learning models, and unsupervised methods, including clustering. Then, we investigate how to preserve the consumers' privacy, using encryption and federated learning, while enabling electricity fraud detection because it has been shown that fine-grained readings can reveal sensitive information about the consumers' activities. After that, we investigate how to design robust electricity fraud detectors against adversarial attacks using ensemble learning and model distillation because they enable malicious consumers to evade detection while stealing electricity. Finally, we provide a comprehensive comparison of the existing works, followed by our recommendations for future research directions to enhance electricity fraud detection.

Keywords: smart grid; false data injection; electricity fraud; deep learning; clustering; privacy preservation; adversarial attacks; ensemble learning

Citation: Badr, M.M.; Ibrahem, M.I.; Kholidy, H.A.; Fouda, M.M.; Ismail, M. Review of the Data-Driven Methods for Electricity Fraud Detection in Smart Metering Systems. *Energies* **2023**, *16*, 2852. https://doi.org/10.3390/en16062852

Academic Editor: Tek Tjing Lie

Received: 31 December 2022
Revised: 10 February 2023
Accepted: 17 March 2023
Published: 19 March 2023

1. Introduction

The smart grid is the current upgrade of the traditional power grid. A typical smart grid architecture is shown in Figure 1. The figure shows that the smart grid architecture comprises five main components, including an electricity generation system, transmission and distribution systems to deliver electricity to consumers, an advanced metering infrastructure (AMI) network, and a system operator [1,2]. The AMI allows two-way transmission of data between the consumers and the electric utility company system operator [3–6]. In the AMI, homes, buildings, and factories are equipped with smart meters (SMs) to monitor the consumers electricity consumption and report fine-grained readings, either through wireless communication networks [7–18] or wired communication alternatives, e.g., power line communication (PLC) [19,20], to the electric utility company system operator for billing

and energy management [21–31]. Also, in the smart grid, renewable energy resources such as solar panels and wind turbines are used for generating environmentally-friendly electricity [32–34]. Despite the benefits brought by the smart grid, it suffers from security and privacy issues [35–50].

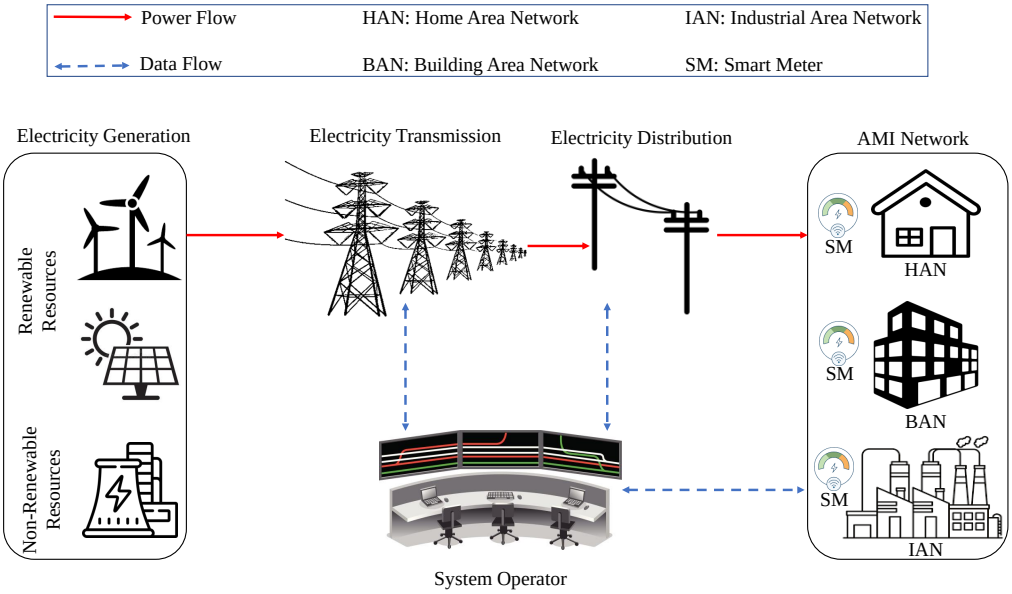

Figure 1. Smart grid architecture.

In this paper, we investigate the electricity fraud problem. Malicious smart grid consumers can compromise their SMs to report low readings to the electric utility company to reduce their bills. SMs can be hacked as follows. Given that passwords used to secure the ANSI optical ports of SMs are not strong enough [32,34], malicious consumers can get access to their SMs by launching a brute force attack against the ANSI optical port using tools, such as Terminator [32,51–53]. After that, the malicious consumer can write a malicious script and install it to get control of the SM [32]. Furthermore, it has been shown recently how malicious consumers can exploit the vulnerabilities of the AMI wireless networks to commit electricity fraud [54]. The electricity fraud problem has devastating consequences. On the one hand, electric utility companies worldwide lose billions of dollars every year. As an example, the annual loss in India due to electricity fraud is $17 billion [2]. In a similar case, the annual losses in Brazil and China are about 16% and 6% of their gross electricity generation, respectively [55,56]. This is not only the case in developing countries, but developed countries also have the problem of electricity fraud. For instance, the annual losses in the United States, United Kingdom, and Canada are $6 billion, $173 million, and $100 million, respectively [2,56,57]. On the other hand, electricity fraud affects the power grid stability and may result in complete blackouts [1,58,59].

To detect electricity fraud, various methods have been proposed in the literature. These methods can be classified as hardware-based methods and data-driven methods. The hardware-based methods require special hardware, i.e., additional devices, for detecting electricity fraud. There are many limitations to the hardware-based methods [60]. First, how expensive it is to deploy special devices. Second, how vulnerable they are. Last, is the issue of maintenance, where it is difficult and burdensome to keep the devices' functionality. One common issue is the failure of batteries in smart devices and the necessity to replace them. On the other hand, there are data-driven electricity fraud detection methods that work on data gathered from the SMs, sensors, and infrastructure [60]. Moreover, by looking into the literature on electricity fraud detection, we find that machine learning (ML)-based detectors offer superiority over the other methods, such as state-estimation-based and game-theory-based detectors [61]. The ML-based detectors employ either unsupervised techniques, e.g., clustering [62], or supervised techniques, including shallow architectures, such as decision

tree (DT) [63], random forest (RF), extreme gradient boosted trees (XGBoost) [64], and support vector machine (SVM) [58], and deep architectures, such as feed-forward neural networks (FFNNs) [2], convolutional neural networks (CNNs) [1], and recurrent neural networks (RNNs) [61]. However, it has been proved in the literature that deep learning (DL) architectures outperform shallow learning architectures in the area of electricity fraud detection [65].

In this paper, we review the main existing data-driven electricity fraud-detecting methods. In addition, we discuss how to perform electricity fraud detection without violating the privacy of smart grid consumers. Also, we clarify that it is not enough to design an accurate electricity fraud detector without considering robustness against adversarial attacks because malicious consumers can simply evade the detectors and continue stealing electricity. Given the severity of electricity fraud, we recommend a set of future research directions to benefit the research community in enhancing electricity fraud detection. Our major contributions in this paper are highlighted as follows.

- We address the electricity fraud problem in smart grid metering systems and review the main existing data-driven approaches for electricity fraud detection with emphasis on the pros and cons of each approach.
- We investigate the new trends and challenges in electricity fraud detection, including efficient and privacy-preserving electricity fraud detection and robust electricity fraud detectors against adversarial attacks.
- We provide a comprehensive comparison of the existing works in terms of the type of metering system, the dataset used, data analysis, data-driven approach, privacy preservation, robustness against adversarial attacks, and special hardware requirement.
- We recommend a set of future research directions for interested researchers in investigating electricity fraud detection.

The rest of this paper is organized as follows. In Section 2, we study the existing data-driven methods for electricity fraud detection, how to preserve the consumers' privacy while enabling electricity fraud detection, and how to secure the electricity fraud detectors against adversarial attacks. Section 3 compares the existing works. In Section 4, we list some future research directions for interested scholars. Finally, Section 5 concludes our paper.

2. Data-Driven Electricity Fraud Detection Methods

2.1. Wide and Deep Convolutional Neural Networks

2.1.1. Existing Research Issues

Electricity fraud detection has recently become easier with the introduction of smart grids. Although it has made it easier, there are still multiple obstacles to overcome. To look for anomalies in electricity usage, the algorithm has to sift through tons and tons of data. Parts of the data can be noisy and inaccurate. The algorithm has to determine which data points qualify and which need to be discarded. The data is also in one-dimensional time-series format and with the extensive amount, it is difficult to read and look for patterns. Electricity fraud data is also different in the sense that traditional machine learning and neural network methods do not work on it. SVM and AI neural networks are not able to be applied to the dataset due to their complex nature [60].

There are several limitation factors in the research and the methods as well. Data collection and monitoring require specific devices that are able to include artificial intelligence features to have the correct data points. This leads to a further limitation as not all traditional machine learning methods work correctly on the dataset. In smart grids, hardware devices such as SMs and sensors need to be installed in all areas to get accurate readings. Smart devices, just like any other cyber-physical system, are susceptible to cyber-attacks and data leaks. There are also physical limitations such as weather conditions and security issues depending on locations. These devices also require maintenance such as protection against vandalism and regular updates such as replacing batteries. The limitations make the task harder to solve with a traditional solution.

2.1.2. Electricity Fraud Problem Analysis

To demonstrate the electricity fraud problem, Zheng et al. have used the dataset provided by the State Grid Corporation of China (SGCC) [60]. The dataset is from a reported duration of 1035 days, which totals around three years. The number of consumers from whom the data was collected is 42,372 individuals.

In [60], the authors depict the energy consumption within one month (4-week duration) for two randomly selected consumers from the SGCC dataset, a regular consumer and an electricity thief. For each consumer, they first showed the consumption by the dates, and then showed the consumption by the weeks. The representation of consumption by dates does not give any distinction between regular consumers and thieves. On the other hand, the representation of consumption by dates enables us to distinguish between regular consumers and thieves as follows. For regular consumers, there is periodicity in energy consumption. The periodicity demonstrates that peak energy usage is typically on the third day of each week and the low point of consumption is typically on the fifth day of each week. This was the case for the whole three years of the SGCC dataset. This is vital to understand electricity fraud because it reflects what a normal periodicity of usage from regular consumers would look like. For electricity thieves, it has been shown in [60] that electricity fraud has an effect on periodicity. In particular, the electricity consumption of the investigated electricity thief throughout the whole month was not constant in terms of patterns, unlike how it was for the investigated regular consumer, where the periodicity of consumption matched a pattern for the whole month. This inequality in periodicity is one form of detecting electricity fraud anomalies. In other words, electricity fraud shows lesser periodic data when compared to normal consumption from legitimate consumers.

Although the presence/absence of periodicity can enable a distinction between legitimate consumers and thieves, analyzing the periodicity in energy consumption is not easy for many reasons. These reasons include the enormous size and noisy nature of the electricity consumption data. In addition to that, conventional machine learning models, such as SVM, cannot capture the data periodicity due to their limited generalization capabilities. This makes it challenging to figure out if electricity fraud is occurring. A solution to this is to use the wide and deep convolutional neural networks framework proposed by Zheng et al. in [60].

2.1.3. Proposed Solution

In the proposed wide and deep convolutional neural networks framework, there are two components; wide component and deep CNN component. The Wide component is represented in the framework through a layer of a fully connected neural network. The wide component extracts global knowledge from the one-dimensional electricity consumption date. On the other hand, the deep CNN component is meant to identify the periodicity/non-periodicity of the electricity consumption data. When taking a glance at the CNN detection method, it is hard to figure out the periodicity/non-periodicity with one-dimensional data due to the consumption of electricity fluctuating every day in a relatively independent way. A solution to this is to transform the one-dimensional electricity consumption data into two-dimensional data and feed it to the deep CNN component. The deep CNN component analyzes the anomalies by looking at several weeks, instead of several days.

2.2. Novel Combined Data-Driven Approach

In [66], Zheng et al. proposed a novel combined data-driven approach for electricity fraud detection. In particular, they combine two novel data mining techniques to detect electricity fraud. Many of the existing electricity fraud detection techniques require the usage of labeled datasets or additional system information, which causes issues in detection accuracy. The combination of two techniques used in this paper, maximum information coefficient (MIC) and clustering technique by fast search and find of density peaks (CFSFDP), allows for better electricity fraud detection. MIC finds the correlation between non-technical loss (NTL) and certain electricity behavior of the consumer whereas

CFSFDP finds abnormal consumers amongst thousands of load profiles by using different arbitrary shapes. Both methods utilize ML to automate the process of analyzing the data and detecting anomalies to detect electricity fraud.

For both MIC and CFSFDP methods to be applied properly to successfully detect electricity fraud, observer meters are required to be installed for every area that contains a group of consumers. For every area, an observer meter measures the sum of all consumers' electricity consumption within that area. Observer meters are more secure than regular SMs, making it harder for malicious users to tamper with the meter data. The data in these observer meters is required by both the MIC and the CFSFDP methods to be implemented properly.

2.2.1. Maximum Information Coefficient (MIC)

This method quantifies the association between tampered load profiles and the NTLs. NTLs are things such as electricity theft, unbilled accounts, billing errors, or systematic errors. To calculate the number of NTLs, the following equation is used:

$$e_t = E_t - \sum_{i \in A} \tilde{x}_{i,t}, \tag{1}$$

where e_t is the amount of NTLs at time t, E_t is the observer meter recorded data at time t, and $\tilde{x}_{i,t}$ is the reported electricity consumption data by the SM of consumer i at time t. Thus, the amount of NTLs at any time is calculated as the difference between the observer meter data and the sum of all consumer's SMs data. Once this is calculated, it is compared to the actual SM data using MIC to detect any electricity fraud. The calculations are done automatically, and the comparisons are done by an ML algorithm to analyze anomalies to detect electricity fraud. This method is set up by installing observer meters in addition to the SMs to get accurate information about electricity fraud. However, if the observer meters are tampered with by false data injection (FDI), the information will be highly inaccurate, and this method fails to work.

2.2.2. Clustering Technique by Fast Search and Find of Density Peaks (CFSFDP)

This method clusters the data by finding density peaks and comparing them to other clusters of data nearby. CFSFDP is able to detect outliers in energy fraud based on data points gathered through SM data. It detects anomalies and finds outliers in load profiles in energy fraud that cannot be detected using MIC.

This method tackles the issue of FDI by using algorithms to deter FDI and correct the data. There are six FDI types that this method tackles and they are mathematically defined in Table 1 [66]. In the table, x_t is the original electricity consumption data at time t, and $\tilde{x}_t$ is the tampered data. In FD11, the reported data is generated by multiplying the original consumption data by a constant small fraction all the time. FD12 shows that consumption data above a threshold are clipped and in FD13, a constant value is subtracted from all the reported data so that the reported data cannot be less than zero. FD14 uses a random period defined each day during which the original consumption data are replaced by zeros before being reported. In FD15, all reports are modified by scaling down the original consumption data by different percentages. Lastly, in FD16, synthetic reports are created by multiplying the average consumption of the previous month by a random percentage defined in each of the reports.

2.2.3. Combining MIC and CFSFDP

Finally, to improve both methods, Zheng et al. [66] proposed to combine both MIC and CFSFDP to resolve the issues around electricity fraud. Figure 2 shows the framework of how to combine MIC and CFSFFDP in order to detect electricity fraud and how the results of the proposed methods are used in the next steps of the detection technique [66]. For an area with k consumers and z-day recorded data series, a time series of NTL is first calculated using SMs data and observer meter data according to Equation (1). The next step

is the normalization of each load profile, $\tilde{\mathbf{x}}_p$, by dividing it with $max_t\ \tilde{\mathbf{x}}_p$, and then the SM dataset is reconstructed into a normalized load profile dataset of $k \times z$ vectors. A correlation calculation is then performed with the normalized load profiles and NTL using the MIC. Normalized load profiles are also used as input to the CFSFDP method for calculating the degree of abnormality $\beta_{i,j}$. The MIC and degree of abnormality $\beta_{i,j}$ are then used to calculate suspicion $Rank_1$ and suspicion $Rank_2$, respectively. The two ranks are combined using one of two ways, either $(\frac{Rank_1+Rank_2}{2})$ or $(\sqrt{Rank_1 \times Rank_2})$. After calculating the combined rank for each consumer, if it turns out to be too high for a certain consumer, then the consumer is flagged for potential electricity fraud. This combination of both the MIC and CFSFDP methods complements each other as CFSFDP allows for the detection of abnormal profiles and MIC allows for the correction of those profiles to the NTLs.

Table 1. Six types of FDI.

Types	Mathematical Representation
FDI1	$\tilde{x}_t \leftarrow \alpha x_t$ where $0.2 < \alpha < 0.8$ is randomly generated
FDI2	$\tilde{x}_t \leftarrow \begin{cases} x_t, & \text{if } x_t \leq \gamma \\ \gamma, & \text{if } x_t > \gamma \end{cases}$ where γ is a randomly defined cut-off point, and $\gamma < \max \mathbf{x}$
FDI3	$\tilde{x}_t \leftarrow \max\{x_t - \gamma, 0\}$ where γ is a randomly defined cut-off point, and $\gamma < \max \mathbf{x}$
FDI4	$\tilde{x}_t \leftarrow f(t) \cdot x_t$ where $f(t) = \begin{cases} 0, & \text{if } t_1 < t < t_2 \\ 1, & \text{otherwise} \end{cases}$ $t_1 - t_2$ is a randomly defined time period longer than 4 h
FDI5	$\tilde{x}_t \leftarrow \alpha_t x_t$ where $0.2 < \alpha_t < 0.8$ is randomly generated
FDI6	$\tilde{x}_t \leftarrow \alpha_t \overline{\mathbf{x}}$ where $0.2 < \alpha_t < 0.8$ is randomly generated, $\overline{\mathbf{x}}$ is the average consumption of the load profile

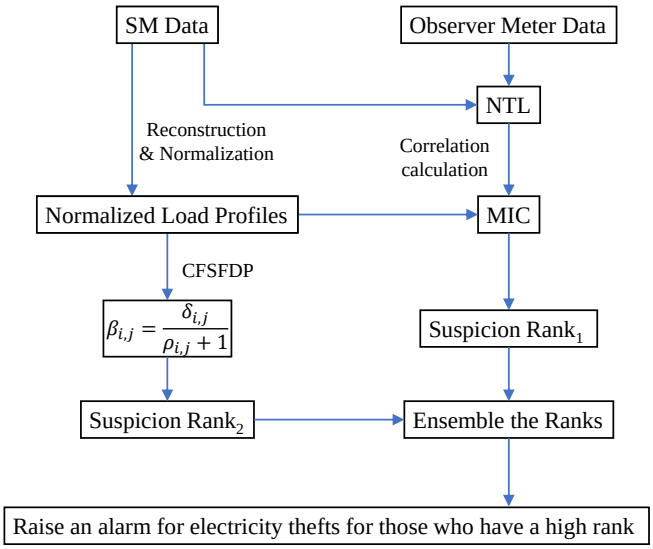

Figure 2. Combined framework.

2.3. Detecting Electricity Fraud in Net-Metering Systems

2.3.1. Existing Research Issues

In smart grids, there are three types of metering systems, including consumption-metering, feed-in-tariff, and net-metering systems [32–34,67,68]. In the consumption-

metering systems, homes are not equipped with renewable energy sources and satisfy their electricity consumption needs from the power grid. In these systems, homes are only equipped with SMs reporting fine-grained electricity consumption readings to the electric utility company. On the other hand, in feed-in-tariff and net-metering systems, homes are equipped with renewable energy resources so that they can generate electricity to satisfy their consumption needs and sell the excess generated electricity to electric utility companies. Apart from [34,68], the existing works only investigated electricity fraud in either consumption-metering [60] or feed-in-tariff systems [32,33]. However, the electricity fraud problem is different and more challenging in the net-metering systems, where homes are equipped with net meters that report net readings accounting for the difference between the consumed energy by homes and generated energy from renewable resources. This is because the net readings depend on both the consumers' consumption patterns and renewable energy resources generation patterns. Moreover, there is no publicly available dataset for the net-metering systems to be used for investigating the electricity fraud problem in these systems.

2.3.2. Data Analysis

To investigate the electricity fraud problem in net-metering systems, Badr et al. [34] have prepared a dataset depending on the Ausgrid dataset [69] and the SOLCAST website [70]. The Ausgrid dataset is a real dataset released by Ausgrid, the largest distributor of electricity on Australia's east coast. This dataset contains real electricity consumption and generation readings recorded every 30 min by a group of customers from Sydney and New South Wales, whose homes are equipped with rooftop solar panels. The recorded readings are from 1 July 2010 to 30 June 2013. SOLCAST is a website providing weather information for any place in the world during a specified time range. By processing the Ausgrid dataset, Badr et al. [34] obtained net readings of 31 customers. By exploiting the SOLCAST website and the available location information about the participating customers from the Ausgrid dataset, Badr et al. [34] obtained weather information, including the solar irradiance and temperature, in the same locations of the customers and in the same time range of the Ausgrid dataset. Finally, for each customer, the prepared dataset contains the customer net readings, the corresponding solar irradiance and temperature values, the day of the week, the season of the year, and C_{Max}, which is the maximum generation capacity from the customer installed solar panels.

Furthermore, because the Ausgrid dataset contains only true readings from benign customers, i.e., it does not contain any electricity fraud examples, Badr et al. [34] proposed a set of realistic attacks that emulate the electricity fraud behavior of malicious customers. they used these attacks to synthesize electricity fraud examples. To understand the electricity fraud problem in net-metering systems, Badr et al. [34] analyzed the prepared dataset. In particular, they found time correlations among the net readings reported by benign customers. Moreover, they found correlations between the reported net readings reported by benign customers and the corresponding solar irradiance and temperature values.

2.3.3. Proposed Solution

Building upon the above data analysis, Badr et al. [34] designed a multi-stage, multi-data-source DL model for detecting electricity fraud in net-metering system. The idea behind their design is that although a malicious customer is capable of reporting false net readings to deceive the electric utility company to achieve higher profit, he/she has no control over the data collected from trustworthy sources, including the solar irradiance, temperature, day, week, and C_{Max}. Badr et al. [34], designed their detector in three stages. The first stage is a hybrid CNN and GRU model that takes the net readings of a particular customer on a certain day. The second stage is a stack of GRU layers whose input is a concatenation of the output from the first stage and the corresponding solar irradiance and temperature values. The final stage is a stack of dense layers whose input is a concatenation of the output from the second stage and the corresponding day, season, and C_{Max}.

The results in [34] demonstrate that depending on some auxiliary data from trustworthy sources in addition to the reported net readings from the customers enhances electricity fraud detection.

2.4. Privacy-Preserving Electricity Fraud Detection

Most of the existing data-driven approaches depend on the reported fine-grained electricity consumption readings for detecting electricity fraud. However, the fine-grained readings may reveal sensitive information about the smart grid consumers, including the appliances being used and whether they are on travel [1,2]. This sensitive information threatens the consumers' privacy and may be misused for criminal activities, such as burglary. To enable electricity fraud detection while preserving the consumers' privacy, several solutions have been proposed in the literature [1,2,71–75].

In [71,72], Salinas et al. proposed a set of distributed peer-to-peer (P2P) algorithms to preserve the consumers privacy. These algorithms are used for exchanging messages among the SMs for solving a linear system of equations (LSE) for calculating SMs honesty coefficients. Instead of depending on the fine-grained readings, the calculated coefficients are used by the electric utility company for detecting electricity fraud. However, the schemes in [71,72] suffer from the following limitations [1,2]. First, [71,72] assume that the SMs apply the schemes honestly but they fail if the SMs manipulate the messages sent to their peers. Second, these schemes require the availability of power line losses beforehand, which is practically hard. Third, these schemes only considered one type of electricity fraud cyber-attacks but attackers can steal electricity in different ways as illustrated in Table 1.

In [73], Salinas et al. proposed a privacy-preserving electricity fraud detection scheme based on the Kalman filter. In particular, a P2P state estimation approach based on the Kalman filter is executed for detecting electricity fraud. In this scheme, the SMs collaborate to calculate the line currents and voltages for each SM. Then, instead of reporting their fine-grained readings, the SMs report the calculated currents and voltages to the electric utility company to be compared with predefined thresholds for detecting electricity fraud. However, this scheme suffers from the following limitations [1,2,51]. First, it depends on state estimation for detecting electricity fraud, which is less accurate than using ML [58]. Second, like [71,72], the scheme in [73] assumes that the SMs apply the state estimation protocol honestly, which means that it fails if the SMs manipulate the messages sent to their peers.

In [75], Yao et al. proposed a privacy-preserving electricity fraud scheme based on encrypting the fine-grained reading before sending them to the electric utility company. In particular, the SMs report their encrypted readings to two entities. The first entity called the server gateway, is assumed to be fully trusted. Therefore, it is allowed to decrypt the reported fine-grained readings and run a CNN-based electricity fraud detector that reports the results to the electric utility company. The second entity, called gateway, is not trusted and is allowed only to aggregate the individual encrypted readings for a group of consumers in a certain area and report the plain-text aggregated reading to the electric utility company for energy management without being able to access the individual plain-text readings. However, in reality, the entity, which is assumed trusted, could misuse the data itself [1,2].

To resolve the research issues identified in [71–75], Nabil et al. [1] proposed a privacy-preserving electricity fraud detection scheme based on secure multi-party computation (SMC). To preserve the consumers' privacy, the proposed scheme allows them to report masked fine-grained readings to the electric utility company. This scheme depends on secret sharing for allowing electric utility companies to perform billing and load monitoring from the received masked readings. Moreover, this scheme depends on an interactive SMC protocol between the electric utility company and each consumer SMs using arithmetic and binary circuits for detecting electricity fraud. By executing this protocol, a CNN-based electricity fraud detector is evaluated on the fine-grained readings reported by each consumer per day for detecting electricity fraud in a privacy-preserving manner. However, this

scheme suffers from the following limitations [2]. First, it suffers from large computation and communication overheads, which is not suitable for SMs because these devices have limited resources. Second, the CNN-based detector involves nonlinearities that have to be approximated by linear functions for allowing privacy-preserving evaluation. These approximations affect the accuracy of the detector. Third, the prediction results of the CNN-based detector are known to both the electric utility company and consumers, which allows malicious consumers to conceal any signs of electricity fraud before on-site inspections.

To resolve the research issues identified in [1], Ibrahem et al. [2] proposed an efficient privacy-preserving electricity fraud detection based on lightweight functional encryption (FE). In particular, FE allows the consumers to send encrypted fine-grained readings to the electric utility company to preserve their privacy while allowing the electric utility company to perform billing, monitoring, and electricity fraud detection without accessing the individual plain-text readings. Using FE, each consumer SM is assigned a unique secret key for encrypting the readings and the electric utility company is given a decryption key for privacy-preserving electricity fraud detection. The electric utility company has an FFNN-based detector with its first layer encrypted using FE. Then, using the decryption key, the electric utility company implements an inner product between the encrypted readings and the encrypted model to detect electricity fraud.

Federated Learning-Based Electricity Fraud Detection

Unlike the previous works [1,2,71–75], Mi et al. [76] proposed a privacy-preserving electricity fraud detection framework based on federated learning (FL). FL allows multiple data owners to train a global ML model in a privacy-preserving way [77–81]. Instead of sharing their data with a central server for training a global model, the data owners train local ML models on their private data and only share the parameters of their models with the server to be aggregated for building the global model [77,78]. In [76], for detecting electricity fraud two servers and a number of detection stations are required. In particular, the consumers use differential privacy (DP) to preserve the privacy of their readings before sending them to the detection stations. The detection stations train local ML models on the received data. Then, the servers and detection stations collaboratively build a global electricity fraud detection model using FL. However, using DP for preserving privacy comes at the cost of the electricity fraud detection accuracy, i.e., there is a trade-off between privacy and accuracy [82,83].

2.5. Robust Electricity Fraud Detection against Adversarial Attacks

2.5.1. Existing Research Issues

Although most of the existing ML-based detectors achieve acceptable performance in detecting electricity fraud cyber-attacks, they are not reliable due to their vulnerability to adversarial attacks targeting ML. It has been shown in [84–87] that the existing ML-based electricity fraud detectors are vulnerable to poisoning and evasion attacks. In poisoning attacks, the attackers attack the ML models during the training phase. In particular, if an attacker has access to the training dataset, he/she can either modify the existing samples or insert newly crafted samples. In [84], Takiddin et al. have launched a poisoning attack by assuming attackers have the ability to do label flipping, i.e., mislabel some malicious samples as legitimate and some benign samples as electricity fraud. Takiddin et al. have run some experiments with different percentages of the adversarial samples, i.e., mislabeled samples, and the results demonstrated the effectiveness of the attack. Moreover, the results showed that the higher the percentage of adversarial samples, the more the performance of the detector deteriorates. Takiddin et al. [84] proved that poisoning attacks can lead to up to a 17% reduction in electricity fraud detection performance of the existing ML-based detectors.

On the other hand, in evasion attacks, the attackers attack the ML models during the run-time phase. In particular, given a properly trained ML model, an attacker tries to push the ML model for providing wrong outputs. In [86], Li et al. have launched evasion attacks

against the existing ML-based electricity fraud detectors. They used popular algorithms, including the fast gradient sign method (FGSM) [88], the fast gradient value (FGV) [89], and DeepFool [90], to create adversarial samples that can bypass the existing detectors. Using the previous algorithms, they can calculate slight perturbations to be added to electricity fraud samples so that they remain malicious while seen by the detectors as benign. Moreover, Li et al. [85,86] have proposed a new algorithm, called SearchFromFree, that is capable of generating adversarial samples evading the existing detectors while maximizing the attacker's achievable profit. In [87], Takiddin et al., have shown the seriousness of evasion attacks on the performance of the existing ML-based detectors by proposing a more powerful evasion attack that does not only depend on the attacker's SM readings but also on its neighboring readings.

Moreover, Badr et al. [91] have shown that the existing global electricity fraud detectors are prone to a new kind of evasion attack, which can be launched by using a generative adversarial network (GAN). The idea of this attack is to exploit the variance in the electricity consumption levels of the different consumers. In particular, the global electricity fraud detection approach employs one detector for detecting electricity fraud from all consumers. Some consumers are characterized by low electricity consumption levels, while other consumers are characterized by high electricity consumption levels. For a malicious high-consumption consumer to commit electricity fraud without being detected, he/she can train a GAN to generate fake low-consumption readings and report them instead of his/her real consumption readings as shown in Figure 3 [91]. Badr et al. [91] have proved the seriousness of this attack by training a GAN to generate fake electricity consumption samples and using the generated samples for evading various global detectors of different architectures with high success rates.

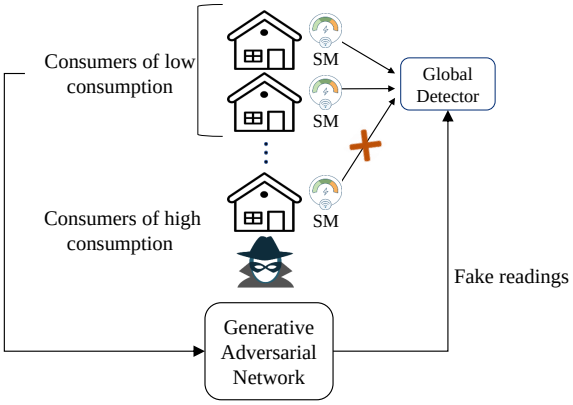

Figure 3. Evasion attack based on GAN.

2.5.2. Proposed Solutions

Although adversarial attacks against ML models cannot be completely avoided, there are some defense strategies that can be used to alleviate them or make it harder for the attacker to find adversarial samples. Three popular defense strategies are adversarial training [92], model distillation [93], and ensemble learning [94]. To defend against the existing adversarial attacks and produce robust electricity fraud detectors, several solutions have been proposed in the literature [84,86,87,91,95]. To defend against poisoning attacks, Takiddin et al. [84] proposed a robust detector based on ensemble learning. The proposed detector is a combination of a deep auto-encoder with attention (AEA), GRU, and feed-forward neural networks (FFNNs). Takiddin et al. provided two variations of the proposed detector. The first one is based on ensemble averaging, where three different models, including AEA, GRU, and FFNN, are trained, and then the detector's final decision is based on the average of the outputs of the three. The second detector is based on a sequential ensemble, where there is a sequence of three models so that the output of each model is processed by the following model and the detector's final decision is taken from the last model in the sequence. The results in [84] indicate that the second detector is more robust

than the first detector and provides at least a 10% increase in robustness against poisoning attacks compared to the existing electricity fraud detectors.

To defend against evasion attacks, Li et al. [86] proposed a robust electricity fraud detector based on the model distillation method. Their proposed detector makes it hard for the attacker to find an adversarial sample with high profit. In other words, their detector forces the attacker to significantly minimize his/her achievable profit to be able to evade detection. The idea of the proposed detector is shown in Figure 4 [86]. The detector is built in two steps. In the first step, a training dataset $\{X, Y\}$ is used to train an ML model M_1. Then, M_1 is used for giving new labels $M_1(X)$ for the training samples X, where the new label for each sample is the vector of output probabilities from the softmax layer of M_1. In the second step, the new dataset $\{X, M_1(X)\}$ is used for training a distilled model M_2 that has a similar structure to M_1. Papernot et al. [93] have proved that the distilled model is less sensitive to the changes in the input sample, and thus more robust against evasion attacks. Also, to defend against evasion attacks, Takiddin et al. [87] proposed a robust detector that is based on sequential ensemble learning. The proposed detector involves an attentive auto-encoder, convolutional-recurrent, and FFNNs. The detector is also an anomaly detector that is trained only on benign samples of true readings aiming at identifying both traditional electricity fraud cyber-attacks and evasion attacks. The results in [87] indicate that the proposed detector is far more robust than the existing electricity fraud detectors.

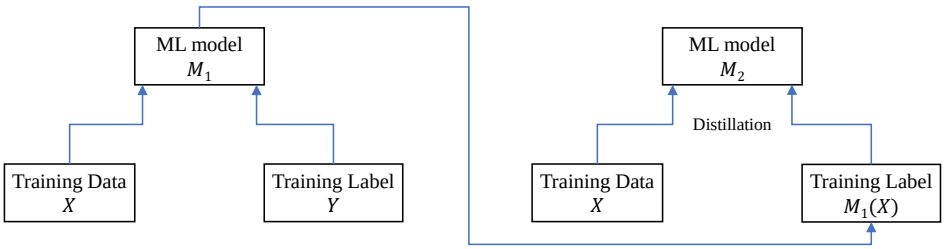

Figure 4. Model distillation.

To defend against the GAN-based evasion attacks, Badr et al. [91] proposed a clustering-based electricity fraud detection approach. Instead of building one global detector to detect electricity fraud from all consumers, multiple detectors can be employed. In particular, an electric utility company can cluster its consumers based on some trustworthy factors affecting their electricity consumption level, including but not limited to geographical location, house size, and contracted power. After that, it builds a specific electricity fraud detector for each cluster. As a result, launching the GAN-based evasion attack against the cluster-specific detector is not gainful because all the consumers of the same cluster have similar electricity consumption levels. On the other hand, faking the readings of the low-consumption consumers in the other clusters is not successful because the fake-generated samples can easily be detected by the cluster-specific detectors as indicated by the results in [91].

2.6. Blockchain-Based Electricity Fraud Detection

In [96], Casado-Vara et al. investigated how blockchain could improve electricity fraud detection. Blockchain is the underlying technology behind cryptocurrencies. It has gained popularity since the appearance of Bitcoin and is currently used in various applications beyond cryptocurrencies, including smart transportation systems [97,98], smart healthcare systems [99,100], and smart grids [96]. Blockchain removed the need for a trusted third party by replacing the central architecture model with a decentralized architecture avoiding the single point of failure and transparency issues. Blockchain is a distributed ledger that is shared among multiple network entities to store transactions in a secure and immutable way. Given the blockchain advantages, Casado-Vara et al. [96] proposed a blockchain-based electricity fraud detection system. In their system, a wireless sensor network (WSN) of

nodes is used to monitor the power distribution grid, and the WSN nodes form a private blockchain. The consumers' SMs act as blockchain users that can only transmit their reading transactions to the blockchain nodes. On the other hand, the WSN nodes act as blockchain miners who can record data on the ledger and also send their sensed data to the blockchain. Through the difference between the WSN nodes transmitted data and the SMs transmitted readings, NTLs can be calculated and localized. Then, a clustering algorithm is used for detecting electricity fraud.

3. Comparison of the Existing Works

In this section, we discuss the limitations of the existing works in the literature and provide a comparison of them in terms of the type of metering system, the dataset used, data analysis, data-driven approach, privacy preservation, robustness against adversarial attacks, and special hardware requirement. Table 2 summarizes our comparison. First, we can observe from the table that most of the existing works focus on detecting electricity fraud in consumption-metering systems, few works [32,33] investigated the problem in fit-in-tariff systems, and [34] is the only existing work investigating the problem in net-metering systems. Second, we can observe that although different datasets have been used in the literature, the Irish dataset [101] is the most used one. Third, we can observe that few works [34,60,66] did data analysis in order to design a suitable detector. Fourth, we can observe that unlike most of the existing works, the works [1,2,76] investigated practical privacy-preserving electricity fraud detection. However, privacy in [1] comes at the cost of high computation and communication overheads and privacy in [76] comes at the cost of reduced model accuracy. Fifth, unlike most of the existing works that focus on designing accurate detectors without considering the vulnerability to adversarial attacks, the works [84,86,87,91] investigated robust electricity fraud detectors. Finally, unlike most of the existing works that do not need special hardware for detecting electricity fraud, the works [66], [76] and [96] require observer meters, detection stations, and WSN nodes, respectively.

Table 2. Comparison of the existing works.

	Work	**[60]**	**[66]**	**[32]**	**[33]**	**[34]**	**[1]**	**[2]**	**[84]**	**[86]**	**[87]**	**[91]**	**[76]**	**[96]**
Features	Metering System	C	C	F	F	N	C	C	C	C	C	C	C	C
	Dataset	S	I	A	*	A	I	I	I	I	I	I	I	*
	Data-Driven Approach	*I*	*II*	*III*	*IV*	*IV*	CNN	FFNN	*V*	*VI*	*V*	GAN & Clustering	*VII*	Clustering
	Data Analysis	$\surd$	$\surd$	$\times$	$\times$	$\surd$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
	Privacy Preservation	$\times$	$\times$	$\times$	$\times$	$\times$	$\surd^{-}$	$\surd^{++}$	$\times$	$\times$	$\times$	$\times$	$\surd^{+}$	$\times$
	Robustness against Adversarial Attacks	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\surd^{P}$	$\surd^{E}$	$\surd^{E}$	$\surd^{E}$	$\times$	$\times$
	Special Hardware Requirement	$\times$	$\surd^{O}$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\surd^{D}$	$\surd^{W}$

Note: C $\longrightarrow$ Consumption, F $\longrightarrow$ Fit-in-tariff, and N $\longrightarrow$ Net-metering; S $\longrightarrow$ SGCC [102], I $\longrightarrow$ Irish [101], A $\longrightarrow$ Ausgrid [69], and * $\longrightarrow$ Synthetic dataset; *I* $\longrightarrow$ Wide and deep CNN, *II* $\longrightarrow$ Combined MIC and CFSFDP, *III* $\longrightarrow$ Auto-regressive integrated moving average (ARIMA) and the Kullback-Leibler divergence (KLD), *IV* $\longrightarrow$ Multi-data-source hybrid DL model, *V* $\longrightarrow$ Sequential ensemble model, *VI* $\longrightarrow$ Distilled FFNN/CNN/RNN, and *VII* $\longrightarrow$ Temporal convolutional network (TCN); $\surd$ $\longrightarrow$ Feature is achieved and $\times$ $\longrightarrow$ Feature is not achieved or considered; $-$ $\longrightarrow$ With high computation and communication overheads, $++$ $\longrightarrow$ With low computation and communication overheads, and $+$ $\longrightarrow$ With reduction in the model accuracy; *P* $\longrightarrow$ Robust against poisoning attacks and *E* $\longrightarrow$ Robust against evasion attacks; *O* $\longrightarrow$ Requires observer meters, *D* $\longrightarrow$ Requires detection stations, and *W* $\longrightarrow$ Requires WSN nodes.

4. Recommendations for Future Directions

As we have seen electricity fraud is a big problem that causes huge financial losses and threatens the power grid stability. Given that, work in electricity fraud detection should

continue until reaching accurate, practical, lightweight, privacy-preserving, and robust electricity fraud detection methods that remain effective against zero-day attacks [103]. Therefore, we recommend the following research directions for interested scholars.

1. *Lightweight Privacy-Preserving Detectors.* SMs are usually cost-effective devices and do not have too much computational power and communication resources. Thus, privacy-preserving electricity fraud detection methods should be continuously upgraded for improved efficiency.
2. *Integrating Relevant Data.* We have seen some reach efforts to integrate data from relevant sources to enhance electricity fraud detection [33,104–107]. Investigating more relevant data sources beyond the control of malicious consumers should continue. Moreover, researchers are encouraged to seek integration between hardware-based and data-driven-based methods for accurate detection.
3. *Security against Adversarial Attacks.* We have seen some reach efforts to secure electricity fraud detectors against adversarial attacks. However, the existing works only consider one attack in designing their detector. Different defense strategies are required for thwarting different attacks. Therefore, seeking an electricity fraud detector robust against multiple attacks is still required.
4. *Continuous Learning Detectors.* Along with implementing methods to help detect electricity fraud, these methods should be continuously monitored and upgraded. This is because they will eventually be discovered by malicious consumers and they will try new ways to circumvent them. Smart grid security personnel should monitor these methods for any discrepancies and errors that could come up. A discovered discrepancy or problem should be flagged and patched to maintain availability and reliability.

In light of the existing limitations, we recommend an electricity fraud detection model with the following properties. It will employ more advanced DL architecture and combine data from different sources to provide high detection accuracy. It will adopt continuous learning to be ready for the zero-day attacks the detector is not trained on. It will employ a lightweight cryptosystem to efficiently preserve the consumers' privacy. It will combine ensemble learning with model distillation to be robust against various adversarial attacks.

5. Conclusions

To combat the rise of electricity fraud, many different approaches have been tried. Among the existing approaches, the data-driven approaches have proved to provide the state-of-art-performance. Therefore, in this paper, we studied the main existing works and analyzed the advantages and disadvantages of each work. Then, we have compared the existing privacy-preserving electricity fraud methods in terms of computation and communication overheads and degradation in the detection accuracy. Also, we have discussed the vulnerability of the electricity fraud detectors to adversarial attacks, including poisoning and evasion attacks. Then, we discussed some defense strategies to make the detectors robust against adversarial attacks. Moreover, we have provided a comprehensive comparison between the existing electricity fraud detection works in terms of the type of metering system, the dataset used, data analysis, data-driven approach, privacy preservation, robustness against adversarial attacks, and special hardware requirement. In the end, we have recommended future research directions, including lightweight privacy-preserving detectors, Integrating relevant data for accurate detection, security against various attacks simultaneously, and continuous learning detectors.

Author Contributions: Conceptualization, M.M.B., M.I.I., H.A.K., M.M.F. and M.I.; methodology, M.M.B., M.M.F. and M.I.I.; investigation, M.M.B., M.I.I., H.A.K., M.M.F. and M.I.; writing—original draft preparation, M.M.B., M.I.I. and M.M.F.; writing—review and editing, H.A.K., M.M.F. and M.I.; supervision, H.A.K., M.M.F. and M.I.; resources, M.M.B., M.I.I., H.A.K., M.M.F. and M.I.; data curation, M.M.B., M.I.I., H.A.K., M.M.F. and M.I.; visualization, M.M.B. and M.I.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Nabil, M.; Ismail, M.; Mahmoud, M.M.E.A.; Alasmary, W.; Serpedin, E. PPETD: Privacy-Preserving Electricity Theft Detection Scheme With Load Monitoring and Billing for AMI Networks. *IEEE Access* **2019**, *7*, 96334–96348. [CrossRef]
2. Ibrahem, M.I.; Nabil, M.; Fouda, M.M.; Mahmoud, M.M.E.A.; Alasmary, W.; Alsolami, F. Efficient Privacy-Preserving Electricity Theft Detection With Dynamic Billing and Load Monitoring for AMI Networks. *IEEE Internet Things J.* **2021**, *8*, 1243–1258. [CrossRef]
3. Fadlullah, Z.M.; Fouda, M.M.; Kato, N.; Takeuchi, A.; Iwasaki, N.; Nozaki, Y. Toward intelligent machine-to-machine communications in smart grid. *IEEE Commun. Mag.* **2011**, *49*, 60–65. [CrossRef]
4. Ibrahem, M.I.; Badr, M.M.; Fouda, M.M.; Mahmoud, M.; Alasmary, W.; Fadlullah, Z.M. PMBFE: Efficient and Privacy-Preserving Monitoring and Billing Using Functional Encryption for AMI Networks. In Proceedings of the 2020 International Symposium on Networks, Computers and Communications (ISNCC), Montreal, QC, Canada, 20–22 October 2020; pp. 1–7. [CrossRef]
5. Ibrahem, M.I.; Badr, M.M.; Mahmoud, M.; Fouda, M.M.; Alasmary, W. Countering Presence Privacy Attack in Efficient AMI Networks Using Interactive Deep-Learning. In Proceedings of the 2021 International Symposium on Networks, Computers and Communications (ISNCC), Dubai, United Arab Emirates, 31 October–2 November 2021; pp. 1–7. [CrossRef]
6. Ibrahem, M.I.; Abdelfattah, S.; Mahmoud, M.; Alasmary, W. Detecting Electricity Theft Cyber-attacks in CAT AMI System Using Machine Learning. In Proceedings of the 2021 International Symposium on Networks, Computers and Communications (ISNCC), Dubai, United Arab Emirates, 31 October–2 November 2021; pp. 1–6. [CrossRef]
7. Badr, M.M.; Fouda, M.M.; Eldien, A.S.T. A novel vision to mitigate pilot contamination in massive MIMO-based 5G networks. In Proceedings of the 2016 11th International Conference on Computer Engineering & Systems (ICCES), Cairo, Egypt, 20–21 December 2016; pp. 366–371. [CrossRef]
8. Badr, M.M.; Fouda, M.M.; Tag Eldien, A.S. A spatiotemporal scenario to mitigate pilot contamination in 5G massive MIMO systems. In Proceedings of the 2017 12th International Conference on Computer Engineering and Systems (ICCES), Cairo, Egypt, 19–20 December 2017; pp. 95–100. [CrossRef]
9. Badr, M.; Fouda, M.; El-dien, A. Enhanced FFR Scenario for Pilot Contamination Mitigation in 5G Systems with Massive Mimo. *Benha J. Appl. Sci.* **2017**, *2*, 99–104. [CrossRef]
10. Abdulaal, M.J.; Ibrahem, M.I.; Mahmoud, M.M.E.A.; Khalid, J.; Aljohani, A.J.; Milyani, A.H.; Abusorrah, A.M. Real-Time Detection of False Readings in Smart Grid AMI Using Deep and Ensemble Learning. *IEEE Access* **2022**, *10*, 47541–47556. [CrossRef]
11. Ibrahem, M.I.; Mahmoud, M.; Fouda, M.M.; Alsolami, F.; Alasmary, W.; Shen, X. Privacy Preserving and Efficient Data Collection Scheme for AMI Networks Using Deep Learning. *IEEE Internet Things J.* **2021**, *8*, 17131–17146. [CrossRef]
12. Alsharif, A.; Nabil, M.; Mahmoud, M.; Abdallah, M. Privacy-preserving collection of power consumption data for enhanced AMI networks. In Proceedings of the 2018 25th International Conference on Telecommunications (ICT), Saint-Malo, France, 26–28 June 2018; pp. 196–201.
13. Sherifl, A.; Alsharif, A.; Mahmoud, M.; Abdallah, M.; Song, M. Efficient privacy-preserving aggregation scheme for data sets. In Proceedings of the 2018 25th International Conference on Telecommunications (ICT), Saint-Malo, France, 26–28 June 2018; pp. 191–195.
14. Kholidy, H.A. Multi-layer attack graph analysis in the 5G edge network using a dynamic hexagonal fuzzy method. *Sensors* **2021**, *22*, 9. [CrossRef] [PubMed]
15. Kholidy, H.A.; Karam, A.; Sidoran, J.L.; Rahman, M.A. 5G Core Security in Edge Networks: A Vulnerability Assessment Approach. In Proceedings of the 2021 IEEE Symposium on Computers and Communications (ISCC), Athens, Greece, 5–8 September 2021; pp. 1–6.
16. Kholidy, H.A. A Triangular Fuzzy based Multicriteria Decision Making Approach for Assessing Security Risks in 5G Networks. *arXiv* **2021**, arXiv:2112.13072.
17. Abuzamak, M.; Kholidy, H. UAV Based 5G Network: A Practical Survey Study. *arXiv* **2022**, arXiv:2212.13329.
18. Kholidy, H.A.; Karam, A.; Sidoran, J.; Rahman, M.A.; Mahmoud, M.; Badr, M.; Mahmud, M.; Sayed, A.F. Toward Zero Trust Security IN 5G Open Architecture Network Slices. In Proceedings of the MILCOM 2022—2022 IEEE Military Communications Conference (MILCOM), Rockville, MD, USA, 28 November–2 December 2022; pp. 577–582. [CrossRef]
19. Yigit, M.; Gungor, V.C.; Tuna, G.; Rangoussi, M.; Fadel, E. Power line communication technologies for smart grid applications: A review of advances and challenges. *Comput. Netw.* **2014**, *70*, 366–383. [CrossRef]
20. Galli, S.; Scaglione, A.; Wang, Z. For the grid and through the grid: The role of power line communications in the smart grid. *Proc. IEEE* **2011**, *99*, 998–1027. [CrossRef]
21. Aladdin, S.; El-Tantawy, S.; Fouda, M.M.; Tag Eldien, A.S. MARLA-SG: Multi-Agent Reinforcement Learning Algorithm for Efficient Demand Response in Smart Grid. *IEEE Access* **2020**, *8*, 210626–210639. [CrossRef]

22. Fouda, M.M.; Fadlullah, Z.M.; Kato, N.; Takeuchi, A.; Nozaki, Y. A novel demand control policy for improving quality of power usage in smart grid. In Proceedings of the 2012 IEEE Global Communications Conference (GLOBECOM), Anaheim, CA, USA, 3–7 December 2012; pp. 5154–5159. [CrossRef]
23. Rajagopalan, A.; Swaminathan, D.; Alharbi, M.; Sengan, S.; Montoya, O.D.; El-Shafai, W.; Fouda, M.M.; Aly, M.H. Modernized Planning of Smart Grid Based on Distributed Power Generations and Energy Storage Systems Using Soft Computing Methods. *Energies* **2022**, *15*, 8889. [CrossRef]
24. Abdulkader, R.; Ghanimi, H.M.A.; Dadheech, P.; Alharbi, M.; El-Shafai, W.; Fouda, M.M.; Aly, M.H.; Swaminathan, D.; Sengan, S. Soft Computing in Smart Grid with Decentralized Generation and Renewable Energy Storage System Planning. *Energies* **2023**, *16*, 2655. [CrossRef]
25. Ibrahem, M.I.; Mahmoud, M.M.E.A.; Alsolami, F.; Alasmary, W.; AL-Ghamdi, A.S.A.M.; Shen, X. Electricity-Theft Detection for Change-and-Transmit Advanced Metering Infrastructure. *IEEE Internet Things J.* **2022**, *9*, 25565–25580. [CrossRef]
26. Alsharif, A.; Nabil, M.; Mahmoud, M.M.; Abdallah, M. EPDA: Efficient and privacy-preserving data collection and access control scheme for multi-recipient AMI networks. *IEEE Access* **2019**, *7*, 27829–27845. [CrossRef]
27. Alsharif, A.; Shafee, A.; Nabil, M.; Mahmoud, M.; Alasmary, W. A multi-authority attribute-based signcryption scheme with efficient revocation for smart grid downlink communication. In Proceedings of the 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Atlanta, GA, USA, 14–17 July 2019; pp. 1025–1032.
28. Alsharif, A.; Nabil, M.; Sherif, A.; Mahmoud, M.; Song, M. MDMS: Efficient and privacy-preserving multidimension and multisubset data collection for AMI networks. *IEEE Internet Things J.* **2019**, *6*, 10363–10374. [CrossRef]
29. Hamed, M.; ElHalawany, B.; Fouda, M.; Tag Eldien, A. Performance analysis of applying load balancing strategies on different SDN environments. *Benha J. Appl. Sci.* **2017**, *2*, 91–97. [CrossRef]
30. Hamed, M.I.; ElHalawany, B.M.; Fouda, M.M.; Eldien, A.S.T. A novel approach for resource utilization and management in SDN. In Proceedings of the 2017 13th International Computer Engineering Conference (ICENCO), Cairo, Egypt, 27–28 December 2017; pp. 337–342. [CrossRef]
31. Hamed, M.I.; ElHalawany, B.M.; Fouda, M.M.; Tag Eldien, A.S. A new approach for server-based load balancing using software-defined networking. In Proceedings of the 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), Cairo, Egypt, 5–7 December 2017; pp. 30–35. [CrossRef]
32. Krishna, V.B.; Gunter, C.A.; Sanders, W.H. Evaluating Detectors on Optimal Attack Vectors That Enable Electricity Theft and DER Fraud. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 790–805. [CrossRef]
33. Ismail, M.; Shaaban, M.F.; Naidu, M.; Serpedin, E. Deep Learning Detection of Electricity Theft Cyber-Attacks in Renewable Distributed Generation. *IEEE Trans. Smart Grid* **2020**, *11*, 3428–3437. [CrossRef]
34. Badr, M.M.; Ibrahem, M.I.; Mahmoud, M.; Fouda, M.M.; Alsolami, F.; Alasmary, W. Detection of False-Reading Attacks in Smart Grid Net-Metering System. *IEEE Internet Things J.* **2022**, *9*, 1386–1401. [CrossRef]
35. Abdalzaher, M.S.; Fouda, M.M.; Ibrahem, M.I. Data Privacy Preservation and Security in Smart Metering Systems. *Energies* **2022**, *15*, 7419. [CrossRef]
36. Hegazy, H.I.; Tag Eldien, A.S.; Tantawy, M.M.; Fouda, M.M.; TagElDien, H.A. Real-Time Locational Detection of Stealthy False Data Injection Attack in Smart Grid: Using Multivariate-Based Multi-Label Classification Approach. *Energies* **2022**, *15*, 5312. [CrossRef]
37. Fouda, M.M.; Fadlullah, Z.M.; Kato, N.; Lu, R.; Shen, X.S. A Lightweight Message Authentication Scheme for Smart Grid Communications. *IEEE Trans. Smart Grid* **2011**, *2*, 675–685. [CrossRef]
38. Fadlullah, Z.M.; Fouda, M.M.; Kato, N.; Shen, X.; Nozaki, Y. An early warning system against malicious activities for smart grid communications. *IEEE Netw.* **2011**, *25*, 50–55. [CrossRef]
39. Fouda, M.M.; Fadlullah, Z.M.; Kato, N.; Lu, R.; Shen, X. Towards a light-weight message authentication mechanism tailored for Smart Grid communications. In Proceedings of the 2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Shanghai, China, 10–15 April 2011; pp. 1018–1023. [CrossRef]
40. Fouda, M.M.; Fadlullah, Z.M.; Kato, N. Assessing attack threat against ZigBee-based home area network for Smart Grid communications. In Proceedings of the 2010 International Conference on Computer Engineering & Systems, Cairo, Egypt, 30 November–2 December 2010; pp. 245–250. [CrossRef]
41. Hegazy, H.I.; Eldien, A.S.T.; Tantawy, M.M.; Fouda, M.M.; TagElDien, H.A. Online Location-based Detection of False Data Injection Attacks in Smart Grid Using Deep Learning. In Proceedings of the 2022 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS), Bali, Indonesia, 24–26 November 2022; pp. 153–159. [CrossRef]
42. Abdelfattah, S.; Baza, M.; Badr, M.M.; Mahmoud, M.M.E.A.; Srivastava, G.; Alsolami, F.; Ali, A.M. Efficient Search Over Encrypted Medical Data With Known-Plaintext/Background Models and Unlinkability. *IEEE Access* **2021**, *9*, 151129–151141. [CrossRef]
43. Alotaibi, M.; Ibrahem, M.I.; Alasmary, W.; Al-Abri, D.; Mahmoud, M. UBLS: User-Based Location Selection Scheme for Preserving Location Privacy. In Proceedings of the 2021 IEEE International Conference on Communications Workshops (ICC Workshops), Montreal, QC, Canada, 14–23 June 2021; pp. 1–6. [CrossRef]
44. Habbak, H.; Mahmoud, M.; Metwally, K.; Fouda, M.M.; Ibrahem, M.I. Load Forecasting Techniques and Their Applications in Smart Grids. *Energies* **2023**, *16*, 1480. [CrossRef]

45. Badr, M.M. Security and Privacy Preservation for Smart Grid AMI Using Machine Learning and Cryptography. Ph.D. Thesis, Tennessee Technological University, Cookeville, TN, USA, 2022.
46. Baza, M.I.; Fouda, M.M.; Tag Eldien, A.S.; Mansour, H.A. An efficient distributed approach for key management in microgrids. In Proceedings of the 2015 11th International Computer Engineering Conference (ICENCO), Cairo, Egypt, 29–30 December 2015; pp. 19–24. [CrossRef]
47. Kholidy, H.A. Autonomous mitigation of cyber risks in the Cyber–Physical Systems. *Future Gener. Comput. Syst.* **2021**, *115*, 171–187. [CrossRef]
48. Kholidy, H.A. Detecting impersonation attacks in cloud computing environments using a centric user profiling approach. *Future Gener. Comput. Syst.* **2021**, *117*, 299–320. [CrossRef]
49. Haque, N.I.; Ashiqur Rahman, M.; Chen, D.; Kholidy, H. BIoTA: Control-Aware Attack Analytics for Building Internet of Things. In Proceedings of the 2021 18th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), Rome, Italy, 6–9 July 2021.
50. Kholidy, H.A. Correlation-based sequence alignment models for detecting masquerades in cloud computing. *IET Inf. Secur.* **2020**, *14*, 39–50. [CrossRef]
51. Ibrahem, M.I. Privacy-Preserving and Efficient Electricity Theft Detection and Data Collection for AMI Using Machine Learning. Ph.D. Thesis, Tennessee Technological University, Cookeville, TN, USA, 2021.
52. Abdalzaher, M.S.; Fouda, M.M.; Emran, A.; Fadlullah, Z.M.; Ibrahem, M.I. A Survey on Key Management and Authentication Approaches in Smart Metering Systems. *Energies* **2023**, *16*, 2355. [CrossRef]
53. Alsharif, A.; Nabil, M.; Tonyali, S.; Mohammed, H.; Mahmoud, M.; Akkaya, K. EPIC: Efficient privacy-preserving scheme with EtoE data integrity and authenticity for AMI networks. *IEEE Internet Things J.* **2018**, *6*, 3309–3321. [CrossRef]
54. Zanetti, M.; Jamhour, E.; Pellenz, M.; Penna, M.; Zambenedetti, V.; Chueiri, I. A Tunable Fraud Detection System for Advanced Metering Infrastructure Using Short-Lived Patterns. *IEEE Trans. Smart Grid* **2019**, *10*, 830–840. [CrossRef]
55. Antmann, P. *Reducing Technical and Non-Technical Losses in the Power Sector*; World Bank: Washington, DC, USA, 2009.
56. Javaid, N. A PLSTM, AlexNet and ESNN Based Ensemble Learning Model for Detecting Electricity Theft in Smart Grids. *IEEE Access* **2021**, *9*, 162935–162950. [CrossRef]
57. Takiddin, A.; Ismail, M.; Nabil, M.; Mahmoud, M.M.E.A.; Serpedin, E. Detecting Electricity Theft Cyber-Attacks in AMI Networks Using Deep Vector Embeddings. *IEEE Syst. J.* **2021**, *15*, 4189–4198. [CrossRef]
58. Jokar, P.; Arianpoo, N.; Leung, V.C.M. Electricity Theft Detection in AMI Using Customers' Consumption Patterns. *IEEE Trans. Smart Grid* **2016**, *7*, 216–226. [CrossRef]
59. Abdulaal, M.J.; Ibrahem, M.I.; Mahmoud, M.; Bello, S.A.; Aljohani, A.J.; Milyani, A.H.; Abusorrah, A.M. DRFD: Deep Learning-Based Real-time and Fast Detection of False Readings in AMI. In Proceedings of the SoutheastCon 2022, Mobile, AL, USA, 26 March–3 April 2022; pp. 682–689. [CrossRef]
60. Zheng, Z.; Yang, Y.; Niu, X.; Dai, H.N.; Zhou, Y. Wide and Deep Convolutional Neural Networks for Electricity-Theft Detection to Secure Smart Grids. *IEEE Trans. Ind. Inform.* **2018**, *14*, 1606–1615. [CrossRef]
61. Nabil, M.; Ismail, M.; Mahmoud, M.; Shahin, M.; Qaraqe, K.; Serpedin, E. Deep Recurrent Electricity Theft Detection in AMI Networks with Random Tuning of Hyper-parameters. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018; pp. 740–745.
62. Peng, Y.; Yang, Y.; Xu, Y.; Xue, Y.; Song, R.; Kang, J.; Zhao, H. Electricity theft detection in AMI based on clustering and local outlier factor. *IEEE Access* **2021**, *9*, 107250–107259. [CrossRef]
63. Tehrani, S.O.; Moghaddam, M.H.Y.; Asadi, M. Decision Tree based Electricity Theft Detection in Smart Grid. In Proceedings of the 2020 4th International Conference on Smart City, Internet of Things and Applications (SCIOT), Mashhad, Iran, 16–17 September 2020; pp. 46–51. [CrossRef]
64. Buzau, M.M.; Tejedor-Aguilera, J.; Cruz-Romero, P.; Gómez-Expósito, A. Detection of Non-Technical Losses Using Smart Meter Data and Supervised Learning. *IEEE Trans. Smart Grid* **2019**, *10*, 2661–2670. [CrossRef]
65. Bhat, R.R.; Trevizan, R.D.; Sengupta, R.; Li, X.; Bretas, A. Identifying Nontechnical Power Loss via Spatial and Temporal Deep Learning. In Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), Anaheim, CA, USA, 18–20 December 2016; pp. 272–279.
66. Zheng, K.; Chen, Q.; Wang, Y.; Kang, C.; Xia, Q. A Novel Combined Data-Driven Approach for Electricity Theft Detection. *IEEE Trans. Ind. Inform.* **2019**, *15*, 1809–1819. [CrossRef]
67. Badr, M.M.; Ibrahem, M.I.; Baza, M.; Mahmoud, M.; Alasmary, W. Detecting Electricity Fraud in the Net-Metering System Using Deep Learning. In Proceedings of the 2021 International Symposium on Networks, Computers and Communications (ISNCC), Dubai, United Arab Emirates, 31 October–2 November 2021.
68. Badr, M.M.; Ibrahem, M.I.; Mahmoud, M.; Fouda, M.M.; Alasmary, W. Detection of false-reading attacks in the AMI net-metering system. *arXiv* **2020**, arXiv:2012.01983.
69. Ausgrid's Solar Home Electricity Data. Available online: https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Solar-home-electricity-data (accessed on 1 September 2020).
70. SOLCAST. Available online: https://solcast.com/historical-and-tmy/ (accessed on 1 September 2020).

71. Salinas, S.; Li, M.; Li, P. Privacy-preserving energy theft detection in smart grids. In Proceedings of the 2012 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), Seoul, Republic of Korea, 18–21 June 2012; pp. 605–613. [CrossRef]
72. Salinas, S.; Li, M.; Li, P. Privacy-preserving energy theft detection in smart grids: A P2P computing approach. *IEEE J. Sel. Areas Commun.* **2013**, *31*, 257–267. [CrossRef]
73. Salinas, S.A.; Li, P. Privacy-preserving energy theft detection in microgrids: A state estimation approach. *IEEE Trans. Power Syst.* **2015**, *31*, 883–894. [CrossRef]
74. Richardson, C.; Race, N.; Smith, P. A privacy preserving approach to energy theft detection in smart grids. In Proceedings of the 2016 IEEE International Smart Cities Conference (ISC2), Trento, Italy, 12–15 September 2016; pp. 1–4.
75. Yao, D.; Wen, M.; Liang, X.; Fu, Z.; Zhang, K.; Yang, B. Energy theft detection with energy privacy preservation in the smart grid. *IEEE Internet Things J.* **2019**, *6*, 7659–7669. [CrossRef]
76. Wen, M.; Xie, R.; Lu, K.; Wang, L.; Zhang, K. Feddetect: A novel privacy-preserving federated learning framework for energy theft detection in smart grid. *IEEE Internet Things J.* **2021**, *9*, 6069–6080. [CrossRef]
77. Badr, M.M.; Ibrahem, M.I.; Mahmoud, M.; Alasmary, W.; Fouda, M.M.; Almotairi, K.H.; Fadlullah, Z.M. Privacy-Preserving Federated-Learning-Based Net-Energy Forecasting. In Proceedings of the SoutheastCon 2022, Mobile, AL, USA, 26 March–3 April 2022; pp. 133–139. [CrossRef]
78. Badr, M.M.; Mahmoud, M.; Fang, Y.; Abdulaal, M.; Aljohani, A.J.; Alasmary, W.; Ibrahem, M.I. Privacy-Preserving and Communication-Efficient Energy Prediction Scheme Based on Federated Learning for Smart Grids. *IEEE Internet Things J.* **2023**. [CrossRef]
79. Kamaludeen, H.A.K.R. An Innovative Hashgraph-based Federated Learning Approach for Multi Domain 5G Network Protection. In Proceedings of the IEEE Future Networks (5G World Forum), Montreal, QC, Canada, 10–14 October 2022.
80. Kholidy, H.A.; Hariri, S. Toward An Experimental Federated 6G Testbed: A Federated Leaning Approach. In Proceedings of the 2022 IEEE/ACS 19th International Conference on Computer Systems and Applications (AICCSA), Abu Dhabi, United Arab Emirates, 5–8 December 2022.
81. Ibrahem, M.I.; Mahmoud, M.; Fouda, M.M.; ElHalawany, B.M.; Alasmary, W. Privacy-preserving and Efficient Decentralized Federated Learning-based Energy Theft Detector. In Proceedings of the GLOBECOM 2022—2022 IEEE Global Communications Conference, Rio de Janeiro, Brazil, 4–8 December 2022; pp. 287–292. [CrossRef]
82. Jayaraman, B.; Evans, D. Evaluating differentially private machine learning in practice. In Proceedings of the 28th USENIX Security Symposium (USENIX Security 19), Santa Clara, CA, USA, 14–16 August 2019; pp. 1895–1912.
83. Hitaj, B.; Ateniese, G.; Perez-Cruz, F. Deep models under the GAN: Information leakage from collaborative deep learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, Dallas, TX, USA, 30 October–3 November 2017; pp. 603–618.
84. Takiddin, A.; Ismail, M.; Zafar, U.; Serpedin, E. Robust Electricity Theft Detection Against Data Poisoning Attacks in Smart Grids. *IEEE Trans. Smart Grid* **2021**, *12*, 2675–2684. [CrossRef]
85. Li, J.; Yang, Y.; Sun, J.S. SearchFromFree: Adversarial Measurements for Machine Learning-based Energy Theft Detection. In Proceedings of the 2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), Tempe, AZ, USA, 11–13 November 2020; pp. 1–6. [CrossRef]
86. Li, J.; Yang, Y.; Sun, J.S. Exploiting vulnerabilities of deep learning-based energy theft detection in AMI through adversarial attacks. *arXiv* **2020**, arXiv:2010.09212.
87. Takiddin, A.; Ismail, M.; Serpedin, E. Robust Data-Driven Detection of Electricity Theft Adversarial Evasion Attacks in Smart Grids. *IEEE Trans. Smart Grid* **2023**, *14*, 663–676. [CrossRef]
88. Goodfellow, I.J.; Shlens, J.; Szegedy, C. Explaining and harnessing adversarial examples. *arXiv* **2014**, arXiv:1412.6572.
89. Rozsa, A.; Rudd, E.M.; Boult, T.E. Adversarial diversity and hard positive generation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 25–32.
90. Moosavi-Dezfooli, S.M.; Fawzi, A.; Frossard, P. Deepfool: A simple and accurate method to fool deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2574–2582.
91. Badr, M.M.; Mahmoud, M.; Abdulaal, M.; Aljohani, A.J.; Alsolami, F.; Balamsh, A. A Novel Evasion Attack Against Global Electricity Theft Detectors and a Countermeasure. *IEEE Internet Things J.* **2023**. [CrossRef]
92. Szegedy, C.; Zaremba, W.; Sutskever, I.; Bruna, J.; Erhan, D.; Goodfellow, I.; Fergus, R. Intriguing properties of neural networks. *arXiv* **2013**, arXiv:1312.6199.
93. Papernot, N.; McDaniel, P.; Wu, X.; Jha, S.; Swami, A. Distillation as a defense to adversarial perturbations against deep neural networks. In Proceedings of the IEEE Symposium on Security and Privacy (SP), San Jose, CA, USA, 22–26 May 2016; pp. 582–597.
94. Strauss, T.; Hanselmann, M.; Junginger, A.; Ulmer, H. Ensemble methods as a defense to adversarial perturbations against deep neural networks. *arXiv* **2017**, arXiv:1709.03423.
95. Takiddin, A.; Ismail, M.; Serpedin, E. Robust Detection of Electricity Theft Against Evasion Attacks in Smart Grids. In Proceedings of the ICC 2021—IEEE International Conference on Communications, Montreal, QC, Canada, 14–23 June 2021; pp. 1–6. [CrossRef]

96. Casado-Vara, R.; Prieto, J.; Corchado, J.M. How blockchain could improve fraud detection in power distribution grid. In Proceedings of the 13th International Conference on Soft Computing Models in Industrial and Environmental Applications, San Sebastian, Spain, 6–8 June 2018; pp. 67–76.
97. Badr, M.M.; Amiri, W.A.; Fouda, M.M.; Mahmoud, M.M.E.A.; Aljohani, A.J.; Alasmary, W. Smart Parking System with Privacy Preservation and Reputation Management Using Blockchain. *IEEE Access* **2020**, *8*, 150823–150843. [CrossRef]
98. Badr, M.M.; Baza, M.; Abdelfattah, S.; Mahmoud, M.; Alasmary, W. Blockchain-Based Ride-Sharing System with Accurate Matching and Privacy-Preservation. In Proceedings of the 2021 International Symposium on Networks, Computers and Communications (ISNCC), Dubai, United Arab Emirates, 31 October–2 November 2021; pp. 1–8. [CrossRef]
99. Alansari, S.A.; Badr, M.M.; Mahmoud, M.M.E.A.; Alasmary, W.; Alsolami, F.; Ali, A.M. Efficient and Privacy-Preserving Infection Control System for COVID-19-Like Pandemics Using Blockchain. *IEEE Internet Things J.* **2022**, *9*, 2744–2760. [CrossRef]
100. Alansari, S.A.; Badr, M.M.; Mahmoud, M.; Alasmary, W. Efficient and Privacy-Preserving Contact Tracing System for COVID-19 using Blockchain. In Proceedings of the 2021 IEEE International Conference on Communications Workshops (ICC Workshops), Montreal, QC, Canada, 14–23 June 2021; pp. 1–6. [CrossRef]
101. Irish Social Science Data Archive. Available online: https://www.ucd.ie/issda/data/commissionforenergyregulationcer/ (accessed on 1 December 2022).
102. State Grid Corporation of China. Available online: http://www.sgcc.com.cn/ (accessed on 1 September 2020).
103. Takiddin, A.; Ismail, M.; Zafar, U.; Serpedin, E. Deep autoencoder-based anomaly detection of electricity theft cyberattacks in smart grids. *IEEE Syst. J.* **2022**, *16*, 4106–4117. [CrossRef]
104. Takiddin, A.; Rath, S.; Ismail, M.; Sahoo, S. Data-Driven Detection of Stealth Cyber-Attacks in DC Microgrids. *IEEE Syst. J.* **2022**, *16*, 6097–6106. [CrossRef]
105. Takiddin, A.; Atat, R.; Ismail, M.; Boyaci, O.; Davis, K.R.; Serpedin, E. Generalized Graph Neural Network-Based Detection of False Data Injection Attacks in Smart Grids. *IEEE Trans. Emerg. Top. Comput. Intell.* **2023**. CrossRef]
106. Boyaci, O.; Umunnakwe, A.; Sahu, A.; Narimani, M.R.; Ismail, M.; Davis, K.R.; Serpedin, E. Graph neural networks based detection of stealth false data injection attacks in smart grids. *IEEE Syst. J.* **2021**, *16*, 2946–2957. [CrossRef]
107. Boyaci, O.; Narimani, M.R.; Davis, K.R.; Ismail, M.; Overbye, T.J.; Serpedin, E. Joint detection and localization of stealth false data injection attacks in smart grids using graph neural networks. *IEEE Trans. Smart Grid* **2021**, *13*, 807–819. [CrossRef]