data

MDPI

Review

The Comparison of Cybersecurity Datasets

Ahmed Alshaibi *, Mustafa Al-Ani, Abeer Al-Azzawi, Anton Konev and Alexander Shelupanov

Department of Complex Information Security of Computer Systems, Faculty of Security, Tomsk State University of Control Systems And Radioelectronics, 634000 Tomsk, Russia; al-ani@fb.tusur.ru (M.A.-A.); al-azzawi@fb.tusur.ru (A.A.-A.); kaa@fb.tusur.ru (A.K.); saa@tusur.ru (A.S.)
* Correspondence: alshaibi@fb.tusur.ru

Abstract: Almost all industrial internet of things (IIoT) attacks happen at the data transmission layer according to a majority of the sources. In IIoT, different machine learning (ML) and deep learning (DL) techniques are used for building the intrusion detection system (IDS) and models to detect the attacks in any layer of its architecture. In this regard, minimizing the attacks could be the major objective of cybersecurity, while knowing that they cannot be fully avoided. The number of people resisting the attacks and protection system is less than those who prepare the attacks. Well-reasoned and learning-backed problems must be addressed by the cyber machine, using appropriate methods alongside quality datasets. The purpose of this paper is to describe the development of the cybersecurity datasets used to train the algorithms which are used for building IDS detection models, as well as analyzing and summarizing the different and famous internet of things (IoT) attacks. This is carried out by assessing the outlines of various studies presented in the literature and the many problems with IoT threat detection. Hybrid frameworks have shown good performance and high detection rates compared to standalone machine learning methods in a few experiments. It is the researchers' recommendation to employ hybrid frameworks to identify IoT attacks for the foreseeable future.

Keywords: cybersecurity; network security; datasets; machine learning; cyberattacks; IoT

Citation: Alshaibi, A.; Al-Ani, M.; Al-Azzawi, A.; Konev, A.; Shelupanov, A. The Comparison of Cybersecurity Datasets. *Data* **2022**, *7*, 22. https://doi.org/10.3390/data7020022

Academic Editors: Aleksandr Ometov and Joaquín Torres-Sospedra

Received: 27 December 2021
Accepted: 25 January 2022
Published: 29 January 2022

1. Introduction

The internet of things (IoT) was introduced to work as a feedback system in terms of adaptability, real-time functionality, intelligence, and predictability. It could be networked or distributed and requires enhanced design tools that enable a design methodology [1]. Furthermore, machine learning (ML) is a vital part of cyber–physical systems (CPSs) and helps the user to convert the data rapidly and precisely into meaningful decisions, actions, and forms. ML techniques are helpful in the analysis of data for descriptive purposes. These techniques will help the user in prediction matters, as they describe what will happen in future. Finally, and most importantly, they are also practiced in prescriptive matters that include providing decision support and automation [2].

In this regard, modern computing technology is used to integrate a large number of cyber and physical components in CPSs. Due to the IoT, a safe and energy-efficient data flow between the physical and digital worlds is maintained [3]. Several applications, including smart health, smart automobiles, and cities, as well as mobile and military systems, make use of cognitive radio frequency CPS. CPS was created to improve human lives while also eliminating routine labor, serving as an essential breakthrough. However, because of its importance, it has been the target of recent high-profile data breaches [4]. Critical infrastructure systems are vulnerable to both physical and cyberattacks because of their intricate interdependencies. Physical attacks can corrupt or harm the information system, while cyberattacks can cause physical problems. Despite their scalability, the physical space of the system in which they show themselves is affected by cyber risks that materialize by nature [5].

Safety and security concerns prevent CPS from being used to solve real-world problems, despite all of the hype surrounding them. It is necessary to operate in real time. Due to this expectation, systems might run into network challenges, such as delays, which is why CPSs have stringent requirements. Furthermore, as compared to standard information technology (IT) systems, the damage that a failure brings to human life and infrastructure is far more severe in this case [6]. In the fields of computer science and IT, cybersecurity has emerged as a significant research area. Even though it initially focused on protecting information systems against attacks from malware, adware, spyware, and ransomware, cybersecurity has since expanded to include intrusion detection systems (IDSs) and firewalls, as well as other technologies. This is partially due to the increase in the interconnectivity of the CPS sensor, actuator, and controller, which has increased the attack surface, making these systems vulnerable to hostile operations. The application of ML for cybersecurity began with IDS, which helped to detect and study malware and other anomalies in information and communication systems. Upon success, ML was utilized to secure IoT systems [7]. However, that was not the only issue that had to be dealt with.

Personal health records must be transferred from the personal health record system to the concerned doctor(s) or medical devices to enable a healthcare CPS on the internet to function properly. By encrypting a personal health record during transmission and restricting access to the locations where it is stored, it is possible to maintain the privacy of the information (databases log files and backups). This is done so because when an unauthorized entity attempts to gain access to a patient's medical records, a confidentiality breach is detected. In this regard, it is important that confidentiality be maintained so as to safeguard user privacy in CPS [8].

The loss of integrity of data or devices in the system can affect security as well. Such attacks could affect the physical or cyber aspects of the process. This includes the assaults carried out via malware, software, or network elements, for example, by using fake sensor data. Since CPS uses granular and other different sensors, the user's privacy may be compromised. Privacy assaults are mainly passive and need private data access or public data conclusions about specific information. In this regard, transparency is key, as it is the ability to prevent the unlawful disclosure of information [9].

Integrity refers to the preservation of data without modification until done by a person who has been granted permission to do so. When an adversary mistakenly, or with the aim of inflicting harm, edits or deletes critical data, the integrity of the system is compromised. As a result, receivers obtain erroneous information and mistakenly believe it to be correct. By preventing, detecting, and/or blocking deception attacks on the information given/received by the sensors and actuators or controllers, integrity can be ensured in the system. During the calculation and communication processes, it is vital to ensure that the data, transactions, and communications are legitimate in order to avoid fraud. Given this requirement, it is critical to validate the identities of both persons participating in the authentication process [9,10].

In the light of these findings, the paper covers privacy issues and cyberattacks on the network of industrial internet of things (IIoT). The authors aim to identify challenges and ideas for future research for the enhancement of IoT security. The study concentrates on different areas in the detection of IoT attacks. Its aim is to describe in detail the development of the cybersecurity datasets used to train the algorithms that are used for building IDS detection models as well as analyzing and summarizing different and famous IoT attacks. Their main security concern is intrusion detection. ML and big data analytics are often used in these systems to assure security. However, when it comes to real-world application, these algorithms can fall short.

2. Motivation

When it comes to the application of machine learning algorithms, cybersecurity is an important topic. Mathematical models alone are not sufficient to combat today's cybersecurity risks. The appropriate software must be installed in the manufacturing company's

network in order to secure all the systems that hackers could use to gain access to critical information. These systems are referred to as 'endpoints' since they are where the data are accessible to the employer. We need to install a ML model in the network of the client healthcare organization and allow it to analyze the network's actions in real time in order to detect cybersecurity risks [11].

ML learns to recognize the characteristics of typical network activity and uses this knowledge as a basis for estimating the likelihood of suspicious activity. It shows a red flag if a user's behavior deviates significantly from the usual. In industrial organizations, Darktrace is an example of a machine learning vendor that can be used for this purpose. It is common practice to deploy machine learning-based security solutions for IT systems. These particular traits, such as skewed datasets, might make it difficult for trained ML algorithms to detect an attack. The paper will go over what it takes to produce high-quality datasets in the following paragraphs [12].

ML enables cybersecurity systems to quickly identify patterns and gain the information necessary to develop countermeasures in order to prevent similar assaults from occurring in the future. For the uninitiated, ML aids cybersecurity groups in both the effective prevention of threats and the rapid response to active attacks. In this regard, and more significantly, the introduction of deep learning (DL) and reinforcement learning (RL) has greatly aided the deployment of ML algorithms to solve real-world problems that were previously intractable by shallow algorithms and the more familiar supervised and unsupervised algorithms. High-dimensional data created by CPS, as well as the continuous growth of data, are factors that encourage the use of DL in CPS. To cater accordingly to the latest developments, deep reinforcement learning (DRL)—which combines DL and RL—was introduced. This development has resulted in a tremendous revolution in CPS research and continues to demonstrate great potential for providing solutions to the current and future challenges in cybersecurity [13].

Despite this massive shift, the fact remains that data are essential for ML to thrive in cybersecurity applications. It is worth noting that ML can also be defined as the process of creating and analyzing new patterns. In this regard, having a big amount of data from a variety of different scenarios is necessary to achieve this goal. Quantity is not everything in this scenario, as the data must also be accurate and up to date. Massive volumes of data are being generated by the IoT and other technologies. Even more massive amounts of information have eluded traditional data management systems although they can be managed with the help of big data solutions. ML techniques are, therefore, critical in the development of cybersecurity systems that can manage huge amounts of data more efficiently. To sum up, this paper focuses on the function of ML in cybersecurity. The usage of data in ML will also be covered. After that, the researchers will look at the future of CPS in conjunction with the current methods.

3. Related Work

There is a great deal of potential for ML to help with cybersecurity, as highlighted in the beginning. In order to solve a variety of computer security issues, many ML techniques have been effectively used. To identify and categorize brute force attacks, ML tasks might be employed. Some applications of ML, along with the datasets used in dealing with cyberattacks, are briefly discussed below.

Network intrusion detection (NID) systems are designed to detect malicious network activity that compromises a network's confidentiality, availability, and integrity. A system for classifying distributed denial of service (DDoS), detecting it at the same time, was developed by [10].

In [14], the authors presented a Fog layer-based DDoS attack detection approach to identify malicious nodes in the IoT network. The suggested solution employs clustering and an entropy-based strategy and is implemented on the OMNeT++ simulator.

A significant cyber threat is the use of social spamming, which relies heavily on bulk messaging, fake accounts, and the dissemination of harmful links. According to [15],

spammers use social media to conduct phishing attacks, spread malware, and promote affiliate websites. A social honeypot, which can be used to identify spammers on social networks such as Twitter and Facebook, was created to help safeguard social systems against such assaults in the future. High accuracy and low false positive rates are achieved with the use of support vector machines (SVMs). Several ML algorithms were used to investigate the intrusion detection challenge in smart grids. They have looked at some possible ways to fix a dataset that is not evenly split. About 12.5% of the attack data in the ADFA-LD dataset that was utilized by them can be found in this regard. For IIoT applications, it is crucial to keep in mind that this ratio does not hold up, according to the results which reveal that using resampling in combination with different ML classification algorithms improves classification accuracy by more than 10% compared to the state of the art, making the findings more realistic [16].

Ullah created an IDS using a mix of J48 and naive Bayes (NB) techniques. For this, the Mississippi State University's gas pipeline infrastructure was used to create the dataset. The response injection, code injection, reconnaissance, and command injection were among the attacks included in their dataset. An attribute filter was first used with the J48 classifier, but it did not work well. Furthermore, an anomaly-based IDS was built using the NB classifier. Their study's assault traffic ratio was 21.87%, significantly greater than in actual life [17].

As code injection attacks, they are two of the most common forms of attacks (command injection and data injection) that generate large amounts of traffic. Attempts were made to adjust the pressure values of the pipeline using seven different variants of data injection attacks and to use the instructions that operate the gas pipeline using four different kinds of command injection attacks. Their accuracy and recall measurements were used to offer a fair evaluation despite the dataset's imbalance (17% attack traffic) [18].

Ambusaidi used the K-means approach for IDS, that is, an unsubstantiated clustering algorithm. To imitate an ICS, an open-source virtual PLC (OpenPLC platform) is employed and coupled with AES-256 encryption. The attacks on their system included denial of service (DoS), interception (eavesdrop), and code injection. However, the proportion of attack data utilized for training was not disclosed [19]. Muhammad et al. presented one-class support vector machine (OCSVM) as an appropriate anomaly-based IDS. Since the dataset is imbalanced, they claimed that OCSVM is a suitable option. The packet size and data rate are the two characteristics of an electric grid's traffic that were employed by the authors. During a normal operation, a supervisory control and data acquisition (SCADA) system was used to train the model. There were no malicious attack data in the data used to train the model [20].

On the other hand, parallel computing is required for the existing stochastic gradient descent (SGD) algorithm for fog-to-things computing. As a result, the SGD will be overburdened by the huge volume of data released via IoT. Given this, the paper suggested a distributed DL-driven IDS utilizing the NSL-KDD dataset, where the features were extracted using the stacked autoencoder (SAE) and classified using softmax regression (SMR). It was demonstrated in their research that the SAE performed better as a DL in terms of accuracy (99.27%) than conventional shallow models [21].

In [21], an optimized extraction of features in the multivariate correlation analysis (MCA) was done by utilizing a triangle-area-based approach MCA. Considering the data that reached the network's destination, certain features were implemented to minimize the overhead. Additionally, the 'triangle area map' feature was used to identify geometrical connections between two unique variables in order to improve the zero-day attack detection accuracy. The author determined the differences between a pre-built normal profile and the observed traffic using the earth mover's distance (EMD). The traffic on the network was then translated into pictures using MCA and inspected for anomalies using the ISCX and KDDCup99 datasets. Their findings were 90.12% (ISCX) and 99.95% (KDD) accurate when they used sample-wise correlation.

An IoT botnet attack is another type of DoS attack. The authors built an IDS, which is a combination of an artificial neural network (ANN), NB, and decision tree (DT), to avoid botnet attacks against the message queuing telemetry transport (MQTT) and domain name system (DNS). ANN, NB, and DT were chosen, as they could classify these vectors proficiently due to the close correntropy values between the malicious and benign vectors. This ensemble outperformed every single algorithm with regard to the performance parameters of the false-positive and detection rates. The end result was that the accuracies recorded for the UNSW and NIMS datasets were 99.54% and 98.29%, respectively [22].

Man-in-the-middle (MitM) attacks, which are similar to DoS attacks, are commonly occurring attacks in an IoT network as well. Several technical solutions have been presented to tackle this issue. Since the traditional feed forward neural networks are not capable of capturing the sequence or time-series data, the authors used the long short-term memory (LSTM) and recurrent neural network (RNN). They were used to avert the instances of impersonation in a smart healthcare context by using a combination of supervised and unsupervised ML algorithms, extracting features with ANN and SVM, and then classifying them with ANN. Firstly, they employed stack autoencoder (SAE) to extract features in their selection and a deep-feature extraction approach (D-FES), followed by ANN and SVM for feature selection. In the end, classification was conducted using ANN. Through the use of the AWID dataset, the researchers were able to attain an accuracy of 99.92% [23].

In [24], an unsupervised DL technique was employed on the basis of SMR and SAE, called self-taught learning (STL). The NSL-KDD dataset was used to compare three different types of classification: the normal and anomaly (2-class), normal and 4 different attack categories (5-class) and normal and 22 different attacks (23-class) classifications. It was found that the 2-class classification outperformed SMR, as it developed a multi-class machine learning-based categorization characterized by mutual information (MI). Moreover, the mutual information feature selection (MIFS) along with linear correlation coefficient (LLC) was utilized for the selection of the linearly dependent variable.

Training an ML model relies heavily on data. One can, for example, utilize a patient's previous medical history to predict the outcome of a new patient. However, the disadvantage here is that patients are wary of disclosing their personal information because of obvious privacy issues. In the research published by [19,25], these issues have been addressed. To categorize medical information using non-linear kernel SVM while protecting the privacy of both the customer and the service provider, Ref. [26] developed an entirely new framework called eDiag. Before this study, researchers employed methods which they found to be unsuitable for use in online medical pre-diagnosis.

Concerned with safeguarding the user data and model outcomes, Ref. [19] categorized privacy difficulties as model privacy issues and learning privacy issues, respectively. According to [25], the first step of detection happened on a mobile terminal, and the information was then sent to a cloud server for additional analysis with the help of DL. Using the Sino Weibo dataset, the authors were able to attain a 91% accuracy rate using the convolutional neural network (CNN) as a categorization algorithm.

4. Role of ML in CPS

Expeditious decisions and actions can now be made, thanks to ML. There are a range of uses for the data in ML approaches, including descriptive and diagnostic analyses, and prediction and prescriptive reasons and purposes (such as decision assistance and decision automation) [24]. For the first time, machine learning may be used to teach an artificial intelligence system without requiring it to be explicitly designed. Being a subclass of artificial intelligence itself, ML allows a system to learn from the training data. It can also assist in creating predictions when new data are available (which may or may not be correct). Technology's rapid advancements have resulted in a massive volume of data being created, often referred to as 'big data' [26].

The ML technique can enhance cyber security to a great extent. A few examples are as follows:

- Cybersecurity companies can employ various tools of data science to process and analyze big data that are historic or acting as a threat to intelligence data over the recent years [27].
- Cybersecurity companies make use of ML algorithms to deal with the problems related to classification, clustering, dimensionality reduction and regression [28].
- The use of ML is significant in the implementation and evaluation of various systems, such as the implementation of authentication systems, evaluating the protocol implementation, assessing the security of human interaction proofs, smart meter data profiling, etc. [27].

Cybersecurity makes great use of ML techniques in several ways. Thanks to ML, modern-day cybersecurity has the ability to perform malware detection, intrusion detection and data leakage, which are cumbersome to be solved only by mathematical models.

One of the methods for detecting threats to cybersecurity is installing the ML model in the network of a healthcare company and allowing it to analyze various activities on and off the network in real time. In comparison to regular network activity, machine learning is useful in creating a parallel between a normal and suspicious activity as well as in determining the probability of its occurrence. If the activity is far from the normal activity of the system, it is flagged as a fraudulent activity. One example of machine learning vendors is Darktrace, which is used for detecting anomalies in healthcare companies [8].

Simply put, ML can be of aid to cybersecurity organizations to take prompt and effective measures against attacks and threats in real time [13]. It is important to note that data are important for determining how successful machine learning is in cyber security. A broad range of data is needed to create various potential outcomes. Both the quantity and quality of data are important. The massive amount of data produced by IoT along with other applications cannot be managed in conventional ways, but a big data framework is what works best in such scenarios.

5. Cyberattacks and IIoT

Manufacturing is known to be the most attacked industry. Industrial control systems have raised many questions over some manufacturers who have just started networking with some corporate systems. It has been seen that many internet manufacturing systems are highly exposed to cyberattacks due to the network connections established between the industrial control systems and internet providers. It has serious effects on the companies, such as threats, physical damage, delayed production, and manufacturing disruptions [28]. Many ICS attacks do not see the need to waste time on software vulnerabilities due to the lack of basic security controls (authentication and encryption) and ICS network design. The attackers gain unregulated access to every controller due to such weak security, and they change their configuration, logic, and state to cause disruption once the network is ruptured [5].

The investigation seems to be the first step against ICS that enables the attackers to survey the digital environment very easily. In order to step into the foothold of the targeted network, they apply various technical tactics. All the possible configurations and vulnerabilities will be looked at carefully by the attacker the moment they decide to launch the malware. The effects of an attack can make changes to some operations and adjustments to the present controls and configurations [29]. Given this context, minimizing the attacks could be the major objective of cybersecurity, while knowing that attacks cannot be fully avoided. The number of people resisting the attacks and protection system is likely less than those who prepare attacks. Well-reasoned and learning-backed problems must be addressed by the cyber machine. The existing tools and architecture should be integrated with it as well. In order to detect fraud, a classification technique could be applied for detection and verification. Moreover, for face detection, the dimensionality reduction can be used [29].

One of the commonly used infrastructures in cyber matters by governments and different industries is the CPS, which helps them to solve the challenges of everyday cyber

life. One of the primary objectives of computational technology is ensuring rapid and correct decision making in an environment with big data systems. In order to transform the manufacturing industry to the next level, the integration of the IoT with CPSs could be helpful. Moreover, we face various issues in our life, such as the absence of smart analytical tools that directly affect industries. In this regard, the tools and equipment mentioned above are very reliable for the betterment of CPS [30]. To handle such advanced devices, skilled human resources will not be sufficient alone. There is a greater chance for security breaches to happen in IoTs that are usually interconnected. So, in CPSs, it is important for the cybersecurity and ML to be integrated, as this will help in meeting the compulsory requirements [29].

ICSs used to be isolated from the outside world to protect them from malicious attacks. Due to the increased connectivity with business networks and the use of internet communications to conveniently convey information, these systems are now vulnerable to cyberattacks. Advanced security is now paramount due to the sensitive nature of these industrial applications. The IIoT applications, in particular, require an IDS to create a safe environment because infiltration is a primary security concern [31].

5.1. IIoT End Point Security Challenges

- Since assaults are deployed into the wild before the antivirus signatures for the attacks are known, antivirus systems are unable to eliminate common malware with any reliability. If the malware manages to infiltrate a vulnerable system between the moment of its launch and the time that the antivirus signatures are applied, the system becomes compromised, even if an antivirus system is installed [31].
- Due to the time that it takes for a vendor to create security updates and end users to install them, the exploitation of the known vulnerabilities is not always prevented by security updates. During this time period, systems are particularly vulnerable. Furthermore, security updates are occasionally incorrect, and when incorrect, they are ineffective in addressing the known vulnerability that is the source of their purpose for being released [1,32].
- IDSs and security monitoring systems are detective in nature, not preventive. In cybersecurity, these systems document and monitor the system continuously to detect, as reliably as possible, any abnormal activity and respond to it. IDSs and monitoring systems are important, but they do not deal with the attacks. There is still a lack of consistent success with intrusion detection and monitoring systems. This is due to the time-consuming nature of intrusion detection and incident response [1,19,32].
- The October 2016 Dyn attack [31]—the world's most significant attack of its kind—established a new trend in how cyberattacks operate. The assault launched by 100,000 malicious endpoints overwhelmed Dyn's internet DNS infrastructure DDoS. Aside from this, the machine-to-machine (M2M) connectivity and real-time analytics are sources of innovation in IoT [33]. However, they remain a source of security vulnerability, as M2M communication still has security issues as well as resource efficiency and scalability issues. The following are the common and major problems with IoT, where block chain technology plays its role [34]. The blockchain has garnered significant attention, as more people have become aware of its potential benefits in various domains. It has a substantial impact on an organization's business model, by reducing expenses, increasing efficiency, and adding additional costs and dangers. The term 'blockchain' refers to distributed ledger technology. Users add transactions by establishing a block with an associated cryptographic hash, timestamp, and transaction data [35].

Each block is sent to each blockchain participant for verification, using the proof-of-work consensus mechanism. Additionally, the block is linked to other blocks kept in the distributed ledger and transparently exchanged with participants via their computers [23]. In this regard, a blockchain is a distributed ledger technology based on peer-to-peer networks, cryptographic technologies, and distributed systems. There are two types of

blockchain models: public permissionless and private permissioned. In a private permissioned blockchain, the transactions are validated and decentralized. A private permissioned blockchain is restricted to the predetermined trusted users with verified identities, while a central authority controls it. As a result, transaction validation is centrally managed on a private permissioned blockchain [36]. One cannot ignore the importance of blockchain in an ML algorithm for intrusion detection, as it can help counteract the attackers at different layers, such as the application layer, perception layer, and processing layer.

5.2. Types of Cyberattacks on IIoT

- When it comes to security risks, DoS is the most straightforward to implement. Due to the growing number of IoT devices with insufficient security, DoS attacks are becoming increasingly popular among attackers [31]. One of the primary objectives of a DoS attack is to overwhelm the network with invalid requests, causing the bandwidth to be wasted. As a result, the legitimate users are unable to access the services. DDoS is an attack in which multiple sources attack a single target at the same time, making it difficult to identify and avoid. Although DDoS attacks occur in a variety of shapes and sizes, their ultimate purpose is the same [37].
- MiTM attacks are among the earliest types of cyberattacks to be discovered. Spoofing and impersonation are two types of this. It is possible for the MiTM attacker to be interacting with node X while pretending to be destination B. Additionally, a secure sockets layer (SSL) stripping allows an attacker to establish a connection with the server through hypertext transfer protocol secure (HTTPS) while connecting with the victim over hypertext transfer protocol (HTTP) [2,31].
- Malware is an abbreviation for malicious software. The number of IoT devices has expanded in recent years, as has the number of IoT software patches, which an attacker can employ to install malware and perform other criminal operations. It comprises viruses, spyware, worms, Trojan horses, rootkits, and other forms of deceptive advertising. Examples include smart home devices, healthcare equipment, and automobile sensors. These attackers are typically state sponsored, well funded, and well trained, which makes them particularly dangerous [38].
- Malicious attacks infiltrate a network and spread malware in the network from infected devices to other devices. A botnet is a hostile attack in which a group of infected devices connects to the internet and engages in illegal, criminal activities together [39].
- Password attacks enable access to a third person's passwords through malicious entities. These include two methods: one is the dictionary method, and the other is the brute force method. The dictionary method is used to decrypt an encrypted password. In contrast, under brute force, multiple possible usernames and passwords are used [39].
- Distributed attacks are where only a single or specific server is not attacked; the surrounding network's infrastructure is also affected. Through a vulnerable entry point, an attacker gains access to a website, which is termed as a backdoor attack [40].
- DDoS attacks prevent other users from accessing network resources, such as servers, by flooding a network with overburdened and overloaded requests [39].
- Spam attacks use messaging systems. In this case, spam messages are sent to a large number of people. These messages contain scams and are a sort of phishing scheme to target consumers [39].

5.3. Some Examples of a False Data Injection Attack (FDIA)

- The data from a patient's equipment, such as blood pressure, pulse, heart rate, and body temperature, are critical to the success of a difficult surgical procedure, according to surgeons. Hackers may tamper with this information in order to cause death. For high-value targets, such as national leaders, influential figures, politicians, activists, and researchers, falsified data can be used to kill them. Given this context, the oc-

currence of FDIAs cannot be ruled out when discussing internet-based healthcare (e-Healthcare) or remote surgery [37].

- Modern healthcare facilities can generate a lot of medical imaging data. For example, the dental scan helps dentists locate any atypical wisdom teeth. If the hacker alters the image, the dentist and patient will be taken unawares. False or distorted images may also endanger the patient's life, especially when it comes to detecting malignant tumors [37].
- Drones or unmanned aerial vehicles (UAVs) are often utilized in military activities. If the FDIA hacks these drones, the drone user party will acquire false intelligence, causing catastrophic damage. Sensors are used extensively in various drone applications to collect data. Bad sensor data can lead to bad intelligence and bad military decisions [37].
- As a result of a recent cyberattack on the Australian parliament interest, there has been a renewed interest in cybersecurity, particularly with regard to the FDIA. On a national and international scale, a successful FDIA could have significant consequences. Things could get even worse in terms of international relations [41].
- User-to-root (U2R) assaults are the second most common type of IoT attack. In a U2R attack, the attacker employs illegal techniques and methods (for example, sniffing passwords and malware injection) to acquire access to devices or obtain access from a normal user account on the victim's computer [40].
- Remote-to-local (R2L) assaults are the third most common type of IoT attack. These attacks are exploitations in which the attacker discovers a security vulnerability in a network and exploits it in order to gain access to it under the guise of a legitimate local user [40].

5.4. Privacy Threats

Users of the IoT are vulnerable to privacy attacks, such as sniffing, de-anonymization, and inference attacks, in addition to security risks. Moreover, the confidentiality of data is jeopardized, regardless of whether the data are at rest or in transit. With regard to privacy concerns, blockchain technology plays an important role in addressing them. It plays its role in privacy matters in the following ways:

- Militarized intrusion techniques (MITs) can be divided into two categories: active MIT attacks (AMAs) and passive MIT attacks (PMAs). The PMA is a passive listener that monitors the data transit between two devices. Despite the fact that the PMA infringes privacy, the data are not altered. An attacker who gains access to a device can silently observe the device for months before launching an assault on the device. With the increasing number of cameras in IoT devices, such as toys, smartphones, and wristwatches, the impact of PMA is becoming increasingly significant [42].
- Passive data privacy attacks (PDPAs), on the other hand, are classified as active data privacy attacks (ADPAs). Data privacy is the root cause of identity theft and reidentification. In this regard, anonymization, location detection, and data aggregation are used in re-identification attacks. They seek to gather information from a variety of sources in order to identify their targets. Malware can be used to impersonate a user. ADPA includes data tampering, while PDPA includes data leakage and re-identification [42].

6. Cybersecurity Datasets

The development of a dataset used to train the IDS detection models is a very important task. Over the years, different datasets have been created [10], and the ones that are most commonly used in the field of IDS are briefly explained below.

6.1. CIC-IDS2017

The Canadian Institute of Cybersecurity and the University of New Brunswick collaborated on the CICIDS dataset in which the Bprofile system was used. The network traffic that

was generated used HTTP, HTTPS, file transfer protocol (FTP), secure shell protocol (SSH), and email protocols to infer abstract behaviors for 25 users. The CI-CIDS-2017 dataset was used, as it has more complicated features and a lot of traffic and attributes that can be used to find anomalies. The entropy was 0.716 between malicious and benign traffic and 0.523 across attack types. In comparison to benign to malicious traffic, the attack types were comparatively less balanced [20]. Figure 1 shows the statistics of attacks and normal behavior in the CICIDS 2017 dataset.

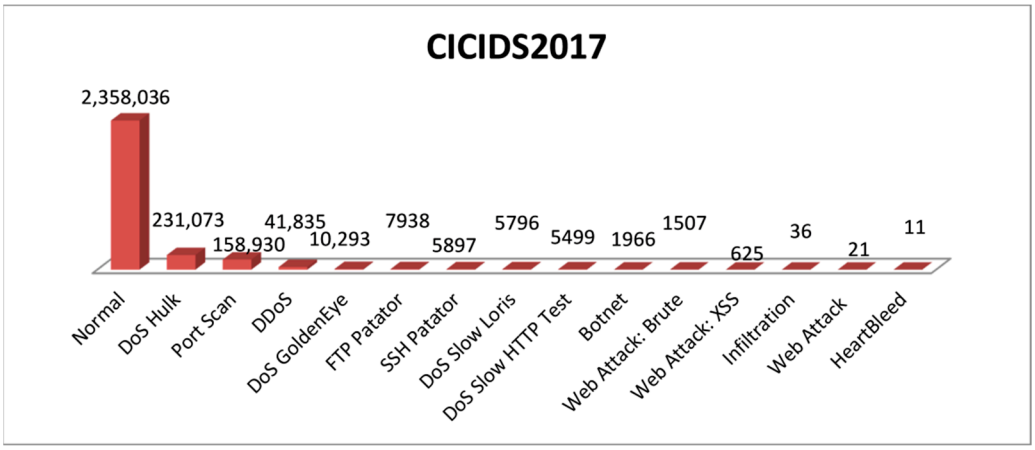

Figure 1. Statistics of attacks and normal behavior in the CICIDS 2017 dataset.

6.2. UNSW-NB15

The UNSW NB15 dataset was created in the lab of the Australian Centre for Cyber Security (ACCS), where the IXIA PerfectStorm tool was used to create this dataset. The dataset includes CSV (comma-separated values) source files. An IXIA traffic generator with transmission control protocol (TCP) connections to three servers was used to create the dataset. A router was used to link two TCP dump servers and three clients. This resulted in CSV files coming from the TCP dump. A router with three clients was also linked to the third server. A firewall separated the two routers to which the first two and third servers were connected. However, the realness of the data in the UNSW NB15 dataset posed problem here.

The reason for this is that there are a lot of attacks that are listed as 'generic', but the precise type of attack is not always clear. The entropy was 0.548 when comparing anomaly and normal types. The entropy of solely attack types was 0.514, indicating that attack types, in general, are somewhat more imbalanced than other traffic types combined [41]. Figure 2 shows the statistics of attacks and normal behavior in the UNSW NB15 train dataset. In addition to this, Figure 3 shows the statistics of attacks and normal behavior in the UNSW NB15 test dataset.

6.3. DS2OS

The communication among the different IoT nodes were included in this dataset as well. The distributed smart space orchestration system (DS2OS), which connects all of these nodes, is a common middleware. Here, T = the user-service communication was recorded and saved in a CSV format. In total, 357,952 samples and 13 features were included in the dataset, while temperature, window, and light controllers were included in the IoT-based system. An example of this is the DS2OS [43]. Figure 4 shows the statistics of attacks and normal behavior in the DS2OS dataset.

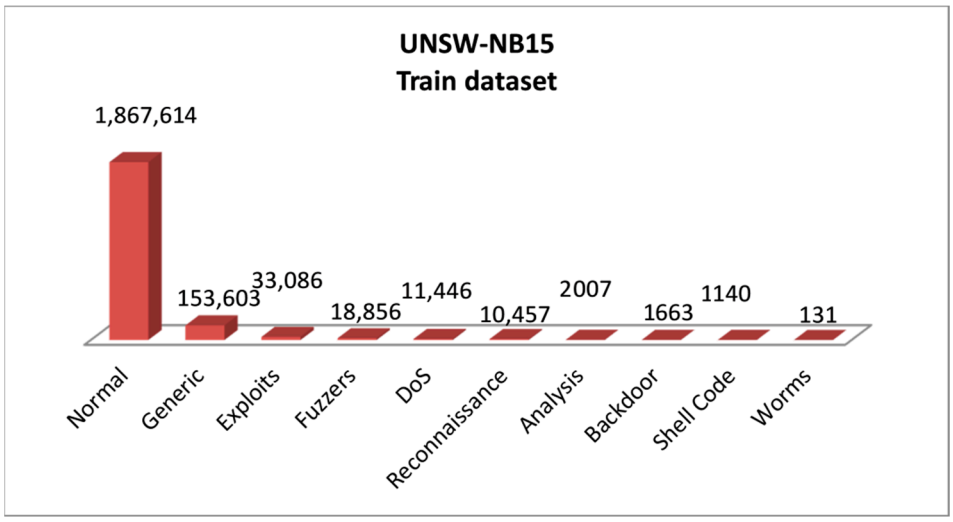

Figure 2. Statistics of attacks and normal behavior in the UNSW-NB15 train dataset.

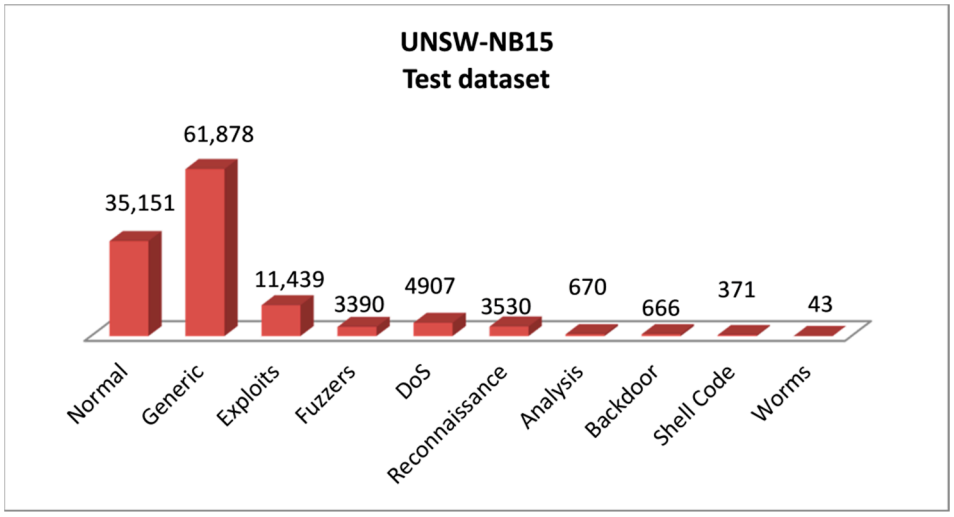

Figure 3. Statistics of attacks and normal behavior in the UNSW-NB15 test dataset.

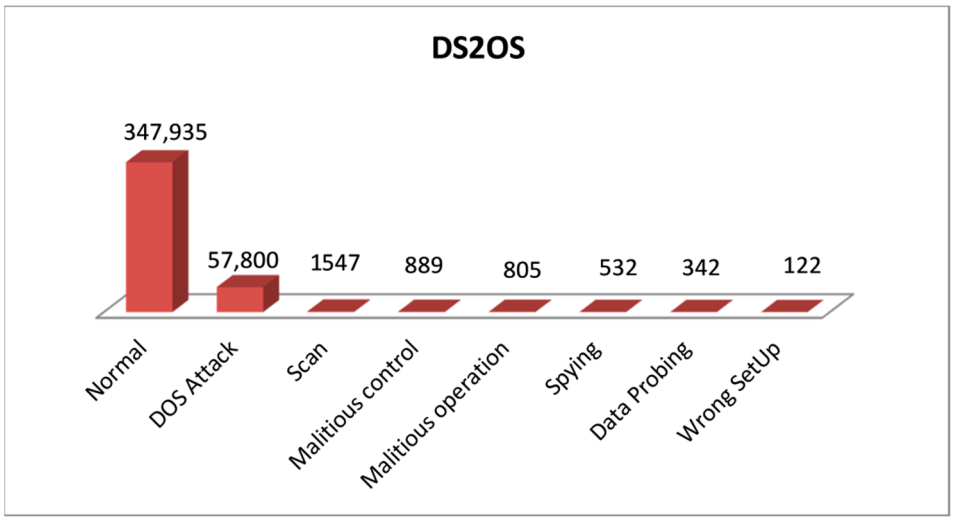

Figure 4. Statistics of attacks and normal behavior in the DS2OS dataset.

6.4. BoT-IoT

A network of multiple bots designed to undertake malicious activities on the target network is known as Botnet. It is managed by a botmaster via a command-and-control protocol (a single unit manager for malicious attacks). They are controlled remotely and help in removing any malicious operations. They are available in different sizes from small botnets to big botnets. Small botnets range up to a few hundred bots, while big botnets can have up to 50,000 bots. A Botnet virus can be effective and functional for years. Hackers distribute botnet software and work in stealth, leaving no trace of their presence [44].

6.5. KDD Cup 1999

One of the most extensively used datasets for detecting network intrusion is the KDD Cup 1999. This was a version of the DARPA Intrusion Detection Evaluation Program from 1998. Raw TCP data were gathered in packet traces from an air force local area network (LAN) by Massachusetts Institute of Technology (MIT) Lincoln Labs for about nine weeks. DoS attacks include syn-floods, unauthorized access to a remote machine R2L. This dataset was large with approximately 41 features. Essential TCP attributes, such as the source and destination bytes, were also available. However, a duplication between the training and testing data occurred, and several important elements, such as IP addresses, were found to be lacking. Moreover, the data gathered were rejected on the fact that they were based on synthetic generation and were obtained over two decades ago, which is out of date [43,45]. In this regard, Figure 5 shows the statistics of attacks and normal behavior in the KDD Cup 1999 train dataset, while Figure 6 shows the same for the KDD Cup 1999 test dataset.

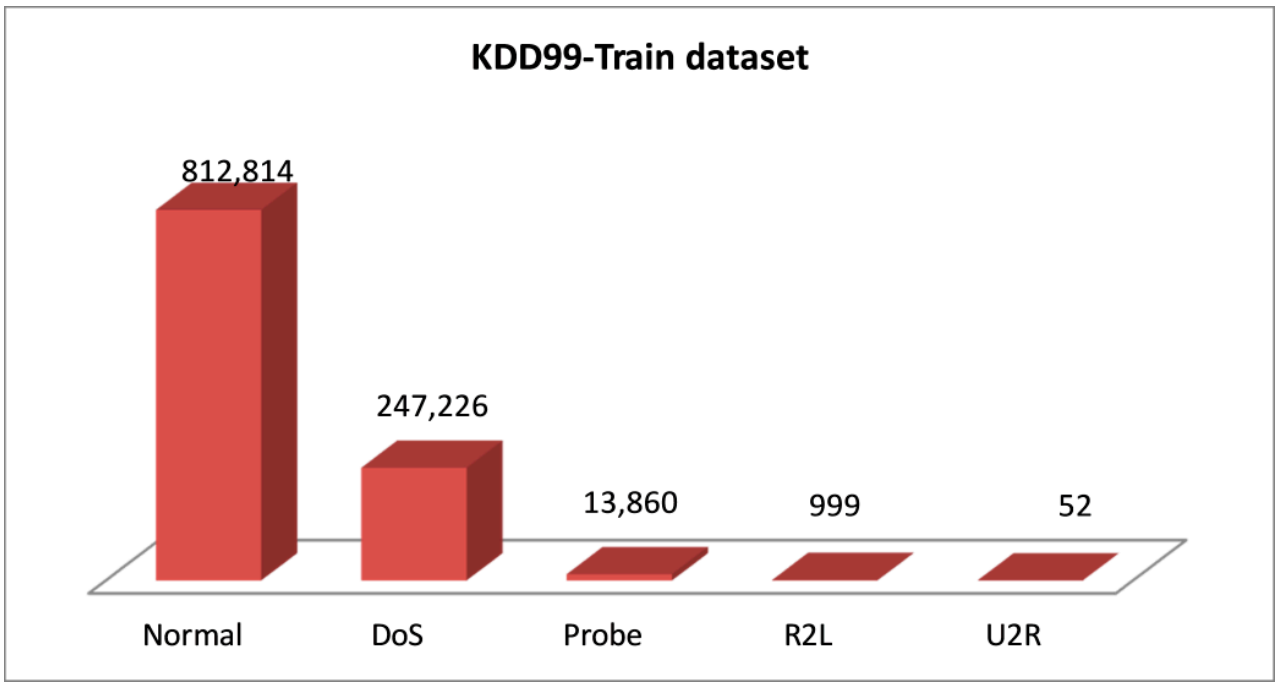

Figure 5. Statistics of attacks and normal behavior in the KDD Cup 1999 train dataset.

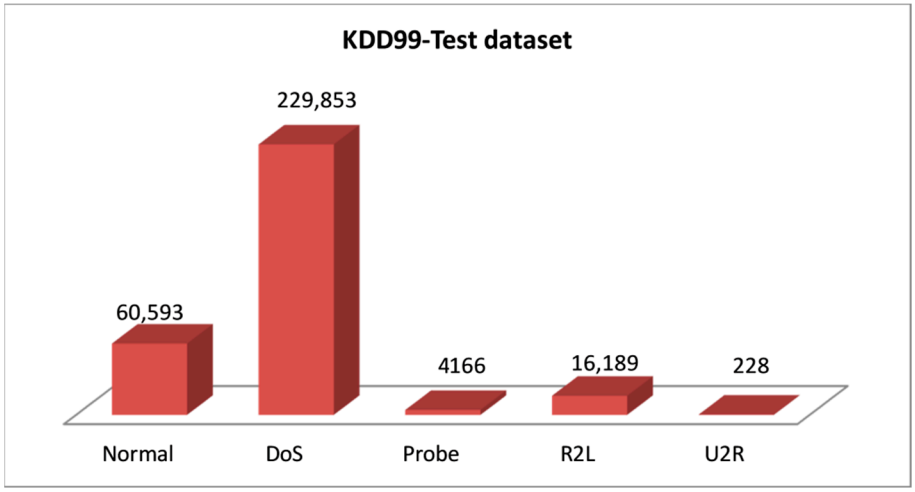

Figure 6. Statistics of attacks and normal behavior in the KDD Cup 1999 test dataset.

6.6. NSL-KDD

The normal and anomalous classes have a 0.719 entropy difference. The entropy was 0.214 across all attack categories, implying that there were more variances across attack classes, compared to those between anomalous and typical traffic. The lower the entropy, the more unbalanced the data. The dataset was gathered in collaboration with the Canadian Institute for Cybersecurity and the University of New Brunswick. The issues with the duplicate data in the training and testing datasets from the KDD Cup 1999 were resolved through the NSL-KDD 2009 dataset. NSL-KDD, on the other hand, eliminated certain redundant and more frequent records from the KDD Cup 1999 dataset, which can still be useful for the training set.

As a result, given that the raw TCP dump data should still be retained, this could lead to even more biases. In this regard, the NSL-KDD dataset has an underlying issue: it comprises data from a network that dates back to the DARPA dataset from 1998. Regular traffic accounts for 51.88% of the data, while anomalous traffic accounts for 48.12%, resulting in a nearly equal balance. Normal and anomalous traffic have a closer proximity to a balanced dataset with an entropy of 0.999 [45]. Figure 7 shows the statistics of attacks and normal behavior in the NSL-KDD training dataset, while Figure 8 shows the same for the NSL-KDD test dataset.

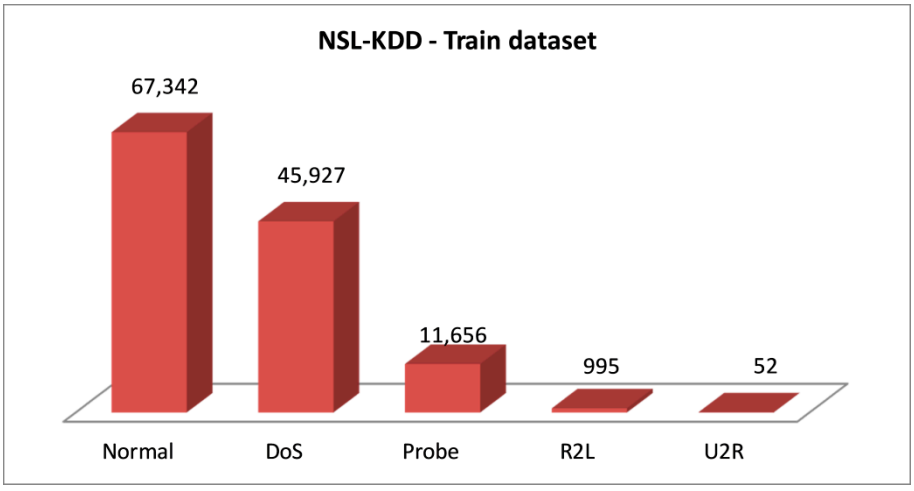

Figure 7. Statistics of attacks and normal behavior in the NSL-KDD train dataset.

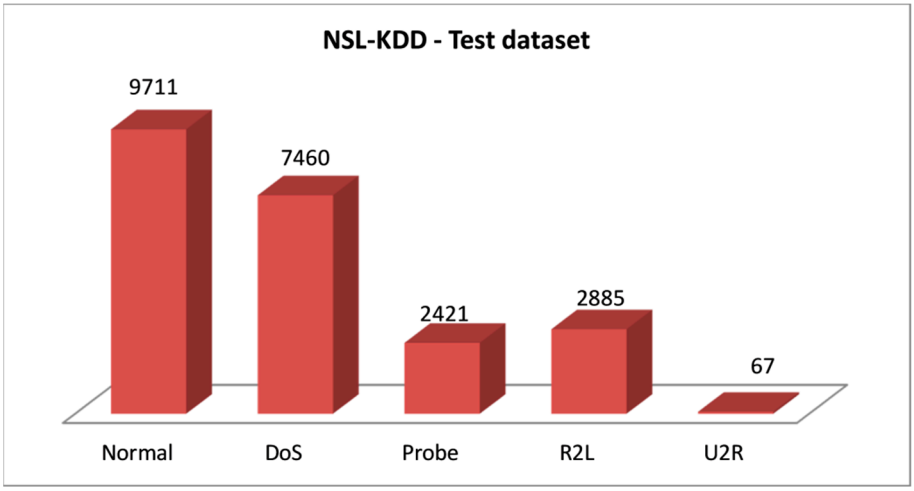

Figure 8. Statistics of attacks and normal behavior in the NSL-KDD test dataset.

7. Discussion

7.1. Limitations

ML plays a vital role along with quality of the datasets. Several detection systems are compared in Table 1, in this regard. It can be observed from the table that each study utilized different datasets for different attacks by using different ML methods. Some of them used the same attacks and the same datasets with different ML methods but attained different accuracy rates.

Table 1. Study outlines presented in the literature.

Reference	ML Technique	IoT Attacks	Dataset	Accuracy
[46]	OS-ELM	Dataset Multiple	NSL-KDD	97.3
[47]	NN	DOS, U2R and R2L.	NSL-KDD	82.3
[35]	DT and NB.	Probing, U2R and R2L.	NSL-KDD	85.8
[23]	TAB	DoS Flooding	KDD99	99.95
[35]	DT	DOS, Reconnaissance U2R, R2L., Backdoor	KDD99	98
[26]	Ensemble Learning	Malware	AndroZoo, Drebin	94
[48]	DT.	DOS	RPL-NIDDS17	98.1
[47]	DT	DOS, Reconnaissance U2R, R2L.	UNSW-NB15	97.8
[21]	NN.	Probing, U2R and R2L	NSL-KDD	99.2
[47]	DT	DoS Reconnaissance, U2R, R2L.	NSL-KDD	98
[49]	NN.	DOS, reconnaissance and DDOS	BoT-IoT	98.26
[19]	LSSVM	Anomaly	KDD99	99.7
[27]	DFEL	Dataset Multiple	UNSW-NB15,	98.5
[21]	LSTM	DoS Flooding	ISCX2012,	99.9
[24]	Adaboost	Botnet Flooding	UNSW-NB15	99.5

A majority of the studies used KDD99, NSL-KDD, and real private datasets. The data volume had to be sufficient to test the algorithm. A majority of the studies used DL methods in which ANN performed well with sufficient data volume, limiting an attack in the process.

NSL-KDD is widely used by researchers with different ML methods. This dataset gave better accuracy when it was used with a DL method, given its large volume. It was widely used for Probing, U2R, DoS, and R2L attacks. KDD99 has the same features as NSL-KDD, which comprises 41 features and 1 class attribute that falls under 4 types of attacks: probe attacks, U2R attacks, R2L attacks, and DoS attacks. NSL-KDD is mostly used, compared to KDD99, but the latter's results are better than other public datasets. However, ML gave better accuracy in the studies, despite being criticized for being outdated.

The results for the NN method, used with KDD99, could be different if it was used for attacks other than probing, U2R, DoS, and R2L attacks. The ISCX2012 dataset was used by [43], although it did not give satisfactory results, whereas the UNSW-NB15 dataset seems the second most widely used dataset after NSL-KDD, owing to its satisfactory accuracy rates. UNSW-NB15 also seems to be the most effective dataset after KDD99, whereas DL methods such as NN seem the most useful ML method to develop IDS models to deal with large datasets such as KDD99 in a real scenario, such as an IoT system. In this regard, a massive amount of data produced by IoT devices helps methods such as NN to learn better and give better accuracy, compared to shallow algorithms.

7.2. Future Research

However, the use of ML in cybersecurity is becoming a matter of great concern. The use of ML algorithms for security was recently shown to be vulnerable to adversarial attacks [50]. They use data from sensors in networked systems to form conclusions or predictions. Thus, tampering with the data or input is a common cyber assault on CPS and other systems. As a result, the model generates incorrect results.

In the case of DNNs, in particular, which have become more popular in the field of CPSs, this is particularly true for CPS. Anxiety has arisen around the thought of repurposing defense-oriented strategies in order to attack a system. This is because adversaries can now use ML algorithms to, instead, attack various systems and carry out adversarial attacks [37]. To their surprise, researchers have discovered that using defense systems to their advantage can produce attacks that are more effective, faster, and less expensive than traditional methods because they take advantage of the resources already available. Given this context, researchers must overcome the challenge of compromising the ML algorithms used to increase cybersecurity in networked systems if they are to succeed [38].

Hackers are now capable of launching sophisticated attacks with potentially life-threatening consequences. Today, the concern is often how to deal with inaccurate material that is given through even authorized channels, such as social media. Owing to the increasing digital connectivity, hackers now have the ability to manipulate election results, demand ransom for private and sensitive data, and destroy critical national infrastructure, such as smart grids, all at the same time.

When a network is infiltrated, the attacker manipulates the data, which is similar to FDIA. According to [39], the contextual anomaly is described as a network abnormality. They are only available in DARPA/KDD Cup 1999 datasets, which are often criticized for being outdated. The concept of FDIA is virtually absent from the publicly available network traffic analysis data. To detect FDIA and other cyber threats accurately, quality datasets are required.

As a future work, first and foremost, we require high-quality data to implement ML algorithms on the real-time network traffic data. Second, we must study the possibility of merging ML algorithms for intrusion detection with blockchain technology in order to address IIoT privacy concerns.

8. Conclusions

Each layer of the IoT has its own set of threats because of its distinct properties. The majority of attacks on the IoT occur at the data transmission layer, according to the study. At any level of the IoT architecture, an attack on the IoT can be detected. In the IoT, data gathering, storage, analysis, and distribution are all essential components. A comprehensive approach is necessary, which includes adhering to best practices and conducting regular testing to ensure that a system is free of vulnerabilities. The system should be able to detect and respond to new threats (such as zero-day attacks) as they emerge because harmful activities are constantly evolving. In this regard, ML and DL can be useful for traffic analysis. It is possible to maintain a log of logs and communicate in an IoT context while utilizing the blockchain. Since it is impossible to tamper with this data, it can be relied upon as evidence in a court of law. For the most part, research on the IoT

security and privacy has focused on either securing or protecting data. The researchers feel that when it comes to system security, it is essential to consider both security and privacy. Furthermore, ensuring the privacy of user data is of paramount importance and can only be done in a way that is consistent from start to finish. The datasets used to train a model are inconsistent in the current systems. In order to achieve the results that they desire, any attacker can alter these datasets. An IoT security and privacy can be achieved by combining ML algorithms with blockchain technology, although this research is still in its infancy.

The data from IoT device testbeds were used in the bulk of the research, although publicly available data were used in others. Among these, the data from the NSL-KDD, UNSW-NB15, and KDDCUP99 sources are the most frequently used by researchers. Concerns about the quality of publicly available data may lead to inaccurate detection results. However, using data from real-world IoT traffic improves the performance of ML algorithms. When it comes to IoT threat detection models, NNs are especially common because of their prevalence in ML approaches that are commonly used in this field. NNs, on the other hand, require substantially more time and money to run on normal CPUs. IoT attack detection models have rarely employed ensemble learning (EL) techniques. Using SVMs to analyze very large datasets is not suggested because the training procedure takes a very long time. Fortunately, the research has shown alternatives by showing that distributed attack detection is more effective than centralized methods [21].

Many problems with IoT threat detection, as well as an outline of the work that has to be done to overcome these obstacles, can also be found in this study. Academics have shown that incorporating publicly available datasets, such as NSLKDD, UNSW-NB15, and KDDCUP99, into IoT threat detection models offers a substantial challenge. When IoT attack models are combined with public datasets with quality concerns, the results can be subpar. The researchers recommend that some data preprocessing and data cleaning procedures be used to improve the quality of publicly available datasets. IoT attack models are also another challenge to conquer. There are few studies that evaluate different ML methods when it comes to detecting IoT hazards. Thus, further research is needed to obtain more generalizable results. EL algorithms and other classifiers can also be used to detect IoT threats, in addition to the ones listed above. Hybrid frameworks have shown good performance and high detection rates compared to standalone machine learning methods in a few experiments. It is the researchers' recommendation to employ hybrid frameworks to identify IoT attacks for the foreseeable future. Since distributed models are more accurate than central models at detecting attacks, the researchers propose using them instead of central models for this purpose.

Funding: This research was funded by the Ministry of Science and Higher Education of Russia, Government Order for 2020–2022, project no. FEWM-2020-0037 (TUSUR).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Rayes, A.; Salam, S. Internet of Things (IOT) Overview. In *Internet of Things from Hype to Reality*; Springer: Cham, Switzerland, 2016; pp. 1–34.
2. Zhang, M.; Selic, B.; Ali, S.; Yue, T.; Okariz, O.; Norgren, R. Understanding Uncertainty in Cyber-Physical Systems: A Conceptual Model. In Proceedings of the European Conference on Modelling Foundations and Applications, Vienna, Austria, 6–7 July 2016; pp. 247–264.
3. Lee, E. The Past, Present and Future of Cyber-Physical Systems: A Focus on Models. *Sensors* **2015**, *15*, 4837–4869. [CrossRef] [PubMed]
4. Golani, N.; Rajasekaran, R. IoT Challenges: Security. In *Internet of Things (IoT)*; CRC Press: Boca Raton, FL, USA, 2017; pp. 211–234.

5. Gupta, Y.; Shorey, R.; Kulkarni, D.; Tew, J. The Applicability of Blockchain in the Internet of Things. In Proceedings of the 2018 10th International Conference on Communication Systems & Networks (COMSNETS), Bengaluru, India, 3–7 January 2018; pp. 561–564.
6. Kang, J.; Yu, R.; Huang, X.; Maharjan, S.; Zhang, Y.; Hossain, E. Enabling Localized Peer-to-Peer Electricity Trading among Plug-in Hybrid Electric Vehicles Using Consortium Blockchains. *IEEE Trans. Ind. Inform.* **2017**, *13*, 3154–3164. [CrossRef]
7. Rohr, J.; Wright, A. Blockchains, Private Ordering, and the Future of Governance. In *Regulating Blockchain*; Oxford University Press: Oxford, UK, 2019; pp. 43–57.
8. Zhu, H.; Liu, X.; Lu, R.; Li, H. Efficient and Privacy-Preserving Online Medical Prediagnosis Framework Using Nonlinear SVM. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 838–850. [CrossRef] [PubMed]
9. Cinque, M.; Cotroneo, D.; Di Martino, C.; Russo, S.; Testa, A. AVR-Inject: A Tool for Injecting Faults in Wireless Sensor Nodes. In Proceedings of the 2009 IEEE International Symposium on Parallel & Distributed Processing, Rome, Italy, 23–29 May 2009; pp. 1–8.
10. Sedjelmaci, H.; Feham, M. Novel Hybrid Intrusion Detection System for Clustered Wireless Sensor Network. *Int. J. Netw. Secur. Its Appl.* **2011**, *3*, 1–14. [CrossRef]
11. Paul, T.; Rakshit, S. Big Data Analytics for Marketing Intelligence. In *Big Data Analytics*; Auerbach Publications: Boca Raton, FL, USA, 2021; pp. 215–230.
12. Gupta, B.B.; Sahoo, S.R. Machine-Learning and Deep-Learning-Based Security Solutions for Detecting Various Attacks on Osns. In *Online Social Networks Security*; Routledge: London, UK, 2021; pp. 57–69.
13. Thiyagarajan, P. A Review on Cyber Security Mechanisms Using Machine and Deep Learning Algorithms. In *Handbook of Research on Machine and Deep Learning Applications for Cyber Security*; IGI Global: Hershey, PA, USA, 2020; pp. 23–41.
14. Gaurav, A.; Gupta, B.B.; Hsu, C.-H.; Yamaguchi, S.; Chui, K.T. Fog Layer-Based DDoS Attack Detection Approach for Internet-of-Things (IoTs) Devices. In Proceedings of the 2021 IEEE International Conference on Consumer Electronics (ICCE) 2021, Las Vegas, NV, USA, 10–12 January 2021; pp. 1–5.
15. Promper, C.; Engel, D.; Green, R.C. Anomaly Detection in Smart Grids with Imbalanced Data Methods. In Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI) 2017, Honolulu, HI, USA, 27 November–1 December 2017.
16. Shekarforoush, S.H.; Green, R.; Dyer, R. Classifying Commit Messages: A Case Study in Resampling Techniques. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN) 2017, Anchorage, AK, USA, 14–19 May 2017; pp. 1273–1280.
17. Ullah, I.; Mahmoud, Q.H. A Hybrid Model for Anomaly-Based Intrusion Detection in SCADA Networks. In Proceedings of the 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 11–14 December 2017; pp. 2160–2167.
18. Beaver, J.M.; Borges-Hink, R.C.; Buckner, M.A. An Evaluation of Machine Learning Methods to Detect Malicious SCADA Communications. In Proceedings of the 2013 12th International Conference on Machine Learning and Applications 2013, Miami, FL, USA, 4–7 December 2013; pp. 54–59.
19. Ambusaidi, M.A.; He, X.; Nanda, P.; Tan, Z. Building an Intrusion Detection System Using a Filter-Based Feature Selection Algorithm. *IEEE Trans. Comput.* **2016**, *65*, 2986–2998. [CrossRef]
20. Aminanto, M.E.; Choi, R.; Tanuwidjaja, H.C.; Yoo, P.D.; Kim, K. Deep Abstraction and Weighted Feature Selection for Wi-Fi Impersonation Detection. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 621–636. [CrossRef]
21. Diro, A.; Chilamkurti, N. Leveraging LSTM Networks for Attack Detection in Fog-to-Things Communications. *IEEE Commun. Mag.* **2018**, *56*, 124–130. [CrossRef]
22. Kolias, C.; Kambourakis, G.; Stavrou, A.; Gritzalis, S. Intrusion Detection in 802.11 Networks: Empirical Evaluation of Threats and a Public Dataset. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 184–208. [CrossRef]
23. Tan, Z.; Jamdagni, A.; He, X.; Nanda, P.; Liu, R.P. A System for Denial-of-Service Attack Detection Based on Multivariate Correlation Analysis. *IEEE Trans. Parallel Distrib. Syst.* **2014**, *25*, 447–456.
24. Moustafa, N.; Turnbull, B.; Choo, K.-K.R. An Ensemble Intrusion Detection Technique Based on Proposed Statistical Flow Features for Protecting Network Traffic of Internet of Things. *IEEE Internet Things J.* **2019**, *6*, 4815–4830. [CrossRef]
25. Jia, Q.; Guo, L.; Jin, Z.; Fang, Y. Preserving Model Privacy for Machine Learning in Distributed Systems. *IEEE Trans. Parallel Distrib. Syst.* **2018**, *29*, 1808–1822. [CrossRef]
26. Feng, P.; Ma, J.; Sun, C.; Xu, X.; Ma, Y. A Novel Dynamic Android Malware Detection System with Ensemble Learning. *IEEE Access* **2018**, *6*, 30996–31011. [CrossRef]
27. Zhou, Y.; Han, M.; Liu, L.; He, J.S.; Wang, Y. Deep Learning Approach for Cyberattack Detection. In Proceedings of the IEEE INFOCOM 2018—IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Honolulu, HI, USA, 15–19 April 2018; pp. 262–267.
28. Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput. Sci.* **2021**, *2*, 160. [CrossRef] [PubMed]
29. Gao, X.-Z.; Kumar Sangaiah, A.; Sugumaran, V. Cloud Based Cyber-Physical Systems in the Design of next-Generation Digital Systems. *Intell. Autom. Soft Comput.* **2017**, *23*, 475–476. [CrossRef]
30. Ahmad Yousef, K.; AlMajali, A.; Ghalyon, S.; Dweik, W.; Mohd, B. Analyzing Cyber-Physical Threats on Robotic Platforms. *Sensors* **2018**, *18*, 1643. [CrossRef]
31. Pfeiffer, A.; Gyulai, D.; Kádár, B.; Monostori, L. Manufacturing Lead Time Estimation with the Combination of Simulation and Statistical Learning Methods. *Procedia CIRP* **2016**, *41*, 75–80. [CrossRef]

32. Chowdhury, A.; Karmakar, G.; Kamruzzaman, J. Survey of Recent Cyber Security Attacks on Robotic Systems and Their Mitigation Approaches. In *Cyber Law, Privacy, and Security*; IGI Global: Hershey, PA, USA, 2019; pp. 1426–1441.
33. Golomb, T.; Mirsky, Y.; Elovici, Y. Ciota: Collaborative Anomaly Detection via Blockchain. In Proceedings of the 2018 Workshop on Decentralized IoT Security and Standards, San Diego, CA, USA, 18 February 2018.
34. Dina, A.S.; Manivannan, D. Intrusion Detection Based on Machine Learning Techniques in Computer Networks. *Internet Things* **2021**, *16*, 100462. [CrossRef]
35. Illy, P.; Kaddoum, G.; Miranda Moreira, C.; Kaur, K.; Garg, S. Securing Fog-to-Things Environment Using Intrusion Detection System Based on Ensemble Learning. In Proceedings of the 2019 IEEE Wireless Communications and Networking Conference (WCNC), Marrakesh, Morocco, 15–18 April 2019; pp. 1–7.
36. Pajouh, H.H.; Javidan, R.; Khayami, R.; Dehghantanha, A.; Choo, K.-K.R. A Two-Layer Dimension Reduction and Two-Tier Classification Model for Anomaly-Based Intrusion Detection in IOT Backbone Networks. *IEEE Trans. Emerg. Top. Comput.* **2019**, *7*, 314–323. [CrossRef]
37. Barreno, M.; Nelson, B.; Sears, R.; Joseph, A.D.; Tygar, J.D. Can Machine Learning Be Secure? In Proceedings of the 2006 ACM Symposium on Information, computer and communications security—ASIACCS '06 2006, Taipei, Taiwan, 21–24 March 2006; pp. 16–25.
38. Ning, Z.; Dong, P.; Wang, X.; Rodrigues, J.J.; Xia, F. Deep Reinforcement Learning for Vehicular Edge Computing. *ACM Trans. Intell. Syst. Technol.* **2019**, *10*, 1–24. [CrossRef]
39. Goswami, G.; Agarwal, A.; Ratha, N.; Singh, R.; Vatsa, M. Detecting and Mitigating Adversarial Perturbations for Robust Face Recognition. *Int. J. Comput. Vis.* **2019**, *127*, 719–742. [CrossRef]
40. Chaâri, R.; Ellouze, F.; Koubâa, A.; Qureshi, B.; Pereira, N.; Youssef, H.; Tovar, E. Cyber-Physical Systems Clouds: A Survey. *Comput. Netw.* **2016**, *108*, 260–278. [CrossRef]
41. Yulianto, A.; Sukarno, P.; Suwastika, N.A. Improving AdaBoost-Based Intrusion Detection System (IDS) Performance on CIC Ids 2017 Dataset. *J. Phys. Conf. Ser.* **2019**, *1192*, 012018. [CrossRef]
42. Ahmad, U.; Song, H.; Bilal, A.; Saleem, S.; Ullah, A. Securing Insulin Pump System Using Deep Learning and Gesture Recognition. In Proceedings of the 2018 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications. 12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE), New York, NY, USA, 1–3 August 2018; pp. 1716–1719.
43. Moustafa, N.; Slay, J. The Significant Features of the UNSW-NB15 and the KDD99 Data Sets for Network Intrusion Detection Systems. In Proceedings of the 2015 4th International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS), Kyoto, Japan, 5 November 2015; pp. 1–8.
44. Koroniotis, N.; Moustafa, N.; Sitnikova, E.; Turnbull, B. Towards the Development of Realistic Botnet Dataset in the Internet of Things for Network Forensic Analytics: Bot-IOT Dataset. *Futur. Gener. Comput. Syst.* **2019**, *100*, 779–796. [CrossRef]
45. Kocher, G.; Kumar, G. Machine Learning and Deep Learning Methods for Intrusion Detection Systems: Recent Developments and Challenges. *Soft Comput.* **2021**, *25*, 9731–9763. [CrossRef]
46. Prabavathy, S.; Sundarakantham, K.; Shalinie, S.M. Design of Cognitive Fog Computing for Intrusion Detection in Internet of Things. *J. Commun. Netw.* **2018**, *20*, 291–298. [CrossRef]
47. Liang, C.; Shanmugam, B.; Azam, S.; Jonkman, M.; Boer, F.D.; Narayansamy, G. Intrusion Detection System for Internet of Things Based on a Machine Learning Approach. In Proceedings of the 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), Vellore, India, 30–31 March 2019; pp. 1–6.
48. Fenanir, S.; Semchedine, F.; Baadache, A. A Machine Learning-Based Lightweight Intrusion Detection System for the Internet of Things. *Rev. Intell. Artif.* **2019**, *33*, 203–211. [CrossRef]
49. Verma, A.; Ranga, V. Machine Learning Based Intrusion Detection Systems for IOT Applications. *Wirel. Pers. Commun.* **2019**, *111*, 2287–2310. [CrossRef]
50. Ge, M.; Fu, X.; Syed, N.; Baig, Z.; Teo, G.; Robles-Kelly, A. Deep Learning-Based Intrusion Detection for IOT Networks. In Proceedings of the 2019 IEEE 24th Pacific Rim International Symposium on Dependable Computing (PRDC), Kyoto, Japan, 1–3 December 2019; pp. 256–25609.