Communication in Physical Sciences 2023, 9(4): 754-772

A Review of Machine and Deep Learning Approaches for Enhancing Cybersecurity and Privacy in the Internet of Devices

Joy Nnenna Okolo

Received: 21 June 2023/Accepted: 24 August 2023/Published: 19 September 2023

Abstract: *This review on deep learning (DL) and machine learning (ML) approaches for network analysis of intrusion detection was described in this study. Each ML/DL method is clearly outlined in the paper. The study identifies the datasets which function as primary tools for tracking network traffic and abnormality detection because the study found that data holds a dominant role in ML/DL methods. The study goes into more detail on the problems of using ML/DL in cybersecurity and suggests potential fixes and directions for further research. Applications and services for the Internet of Devices (IoD) are extensively used in fields including eHealth, smart industry, smart cities, and driverless cars. The Internet of electronic devices is therefore extensively networked and capable of sending sensitive and private data without the need for human contact. For this reason, protecting data privacy is vital. A deep review of current machine learning (ML) and deep learning (DL)-based privacy solutions for the Internet of Devices is presented in this study. In conclusion, we pinpoint a few feasible solutions for various risks and threats.*

Keywords*: Cybersecurity**, privacy protection, ML, DL, Internet, systems and devices*.

Joy Nnenna Okolo
Department of Computer and Information Sciences, Western Illinois University, Macomb, Illinois.United States.
Email: okolojoy2704@gmail.com

1.0 Introduction

The application of machine learning (ML), deep learning (DL), and data mining (DM) techniques in cybersecurity is examined in this research, with an emphasis on data breach detection and prevention. An in-depth review is provided based on current practices indicative of each method, while emphasizing the intricacy and usefulness of these various algorithms. A framework for comparing the methods is proposed, as well as recommendations and best practices for the various types of cyber threats types.

Cybersecurity refers to the suite of technologies and protocols that are implemented to protect a digital asset, such as a network, system, application, or data, from unauthorized access, malevolent assault, tampering or destruction (Buczak & Guven, 2015). The safety of digital assets typically involves a protective infrastructure with a combinatorial host and network architecture (Olawale et al., 2020). The overall base systems typically consist of an assortment of firewalls, intrusion detection systems (IDS), and anti-virus software (Buczak & Guven, 2015). Between the different families of hardware connected to a digital network, intrusion detection systems (IDS) serve a purpose beyond merely detecting how many devices are connected to a network; they will alert when there is unauthorized use of a device, data tampering, data duplication or data destruction (Salloum et al., 2021; Mukkamala et al., 2005). Threats can usually come from an internal actor inside an organization, or an external actor outside an organization (Areghan, 2023). As the sophistication and volume of cyber-attacks continue to grow, applying them in systems that use AI and ML, is increasingly required to carry out modern threat detection and mitigation (Yavanoglu & Aydos, 2017). When implementing a cyber security approach, it is important to consider the most recognized security standards to ensure that you are applying the secure, effective and scalable cyber defense solution

(like in the case of Sdn), and there is no question that these standards support cyber security research including, intrusion detection systems (Areghan, 2023). From this perspective, they are required. Other research projects may be able to verify the machine/deep learning methods and possibly the underlying datasets. However, as we noted, there has been limited cybersecurity research (including security ontologies) limited to (ML/DL)/DM methodologies and security datasets. This study surveys ML/DL)/DM techniques with a path towards cybersecurity, while considering methodologies and descriptions. While numerous reviews have been published (including Salloum et al., 2021); and there are various references to publications which use these methods (including Yin et al., 2017; Xiao et al., 2018).

Unlike prior reviews, this paper analyzes studies that are identified and selected based on effective selection criteria. The literature was assembled through targeted searches in Google Scholar, applying keywords entitled: machine learning, deep learning, cybersecurity, and data mining. Priority was given in a given order to highly cited studies that applied known methodologies. However, a small selection of lesser cited studies that are innovative were seen as appropriate to include to ensure entrance into the analysis of less-established techniques. Overall, the selection strategy was to provide at least one and ideally multiple representative works for each of the categories of ML, DL, and DM.

Globally, tens of billions of smart gadgets are connected and share information with little assistance from humans (Rodríguez et al., 2023). The volume of created data has grown rapidly due in part to the success of Internet of devices services and apps, as well as the notable increase in the number of IoT devices. According to a forecast by the International Data Corporation, between 2020 and 2023, the volume of data would rise from 4 to 140 zettabytes (Rodríguez et al., 2023). Concerns about user privacy are growing as a result of the substantial volume of personal data that IoD collects and shares. Numerous recent papers list the various security and privacy risks related to this.

By the end of 2022, approximately 0.015 trillion smart gadgets will be linked to the computer network, according to Gartner (2022). The vast volume of unprotected data that is gathered and kept online is a liability in addition to the devices' potential vulnerability (Rodríguez et al., 2023). The inability of users to manage the data that their gadgets share has made privacy one of the main issues. IoD solutions need to ensure that user data is highly protected. A company may suffer a significant financial loss if its data is compromised or leaked, giving competitors access to private information).

2.0 Literature Review

Intelligent algorithms and software offer substantial solutions to intricate engineering and scientific problems, and this has triggered the emergence of machine learning (ML) (Jordan & Mitchell, 2015). HoweveDavidr, ML has advanced significantly in the past 2 decades, and is now easier to use and more approachable by people without solid technical knowledge (Fraley & Cannady, 2017). The technology that started as an experimental idea, which was limited to opaque black-box systems, has now become a practical technology that is already used in many industries (Jordan & Mitchell, 2015).

Jordan and Mitchell (2015) highlight the role of machine learning in the substantial expansion of a range of software functions, namely computer vision (Alazab & Tang, 2019), natural language processing (Alazab & Tang, 2019; Li, 2018), speech recognition, and robotic automation (Li, 2018). Google, Facebook, Amazon, and other tech powerhouses today use ML to improve user interaction, create individual recommendations, and improve promotion tactics. To a developer, training an ML model

can be more effective than conventional programming processes, where systems are commonly created by imitating an expected output instead of coding procedures (Fraley & Cannady, 2017). Machine learning is creating a significant impact that goes well beyond technology companies and is revolutionizing data-driven domains such as cybersecurity (Jordan & Mitchell, 2015). It similarly promises to be transformative in other fields, including cosmology, biology, and the social sciences (Alhashmi et al., 2019; Darwish et al., 2020; Salloum et al., 2017; 2021) and is starting to open new possibilities and understandings (Fraley & Cannady, 2017).

Machine learning (ML) proposes novel approaches to the interpretation and handling of experimental results, providing new insights into the data analysis (Fraley & Cannady, 2017). On a theoretical level, the analysis of ML algorithms allows an understanding of the principles of big data. The tools may also be incorporated in domain applications to enhance performance standards and to tune system behaviour.

This huge bedrock of ML methods, including logistic regression, linear regression, Naive Bayes, decision trees, random forests, support vector machines, gradient boosting, and deep learning, among others, plays different roles based on the task to be solved. The variety of these algorithms may be in their differences; however, they are united by the goal to study and process large-scale sets of data effectively, resulting in adaptive and strategic answers. Even greater optimization is likely in the future due to the improvements in the algorithms (Fraley & Cannady, 2017). Cybersecurity is another area where machine learning is crucial to handle the vast amounts of information that digital systems produce (Fraley & Cannady, 2017). Since networks are prone to external attacks, the success of cyberattacks is usually dependent on the sophistication of tools that are utilized to scan and exploit targets (Jordan & Mitchell, 2015). Even adversaries are starting to use ML to conduct more precise and more damaging attacks..

However, in the recent past, the research community started to focus on the relationship between artificial intelligence and cybersecurity, and a few surveys discussed the security implications of ML (Salloum et al., 2021). Remarkably, Salloum and colleagues (2021) reviewed the recent applications of machine learning in Internet of Things (IoT) and intrusion detection, noting the increasing applicability of machine learning in securing computer networks against advancing cyber threats.

The following sections include important literature reviews that concentrated on machine learning (ML) and deep learning (DL) methodologies in network-based intrusion detection systems, where detailed descriptions of each method's contribution were provided (Xin et al., 2018). Xin et al. performed a comprehensive review of security issues and defensive strategies during the training, testing, and inference of machine learning models from a data-centric view. In addition to this, Salloum et al. (2021) also reviewed the scope of artificial intelligence security risks and particularly highlighted vulnerabilities in supervised learning and reinforcement learning systems. Papernot et al. (2016), on the other hand, provide updates on defensive strategies and security concerns.

2.1 Cybersecurity

Cybersecurity refers to the measures used to safeguard electronic data from theft or damage. It is customary to avoid abusing these tools and information. Software, hardware, and Internet data are all included in cybersecurity, which is useful for safeguarding everything from private information to complex government systems. Businesses must keep up efficient risk management structures that focus on information security vulnerabilities and address corrective, detection, monitoring, and assessment issues in order to address the main cyber security problems (Jahwar and Ameen,

2021 and Maseer et al., 2021). Cybersecurity is the key component of technological operations and processes that guard networks, devices, and programs against access, harm, and illegality (Olawale et al., 2020).

Personal data is gathered, processed, and stored on computers by any sector, like as corporations, governments, financial organizations, and others, and then transmitted through other computers or networks. Consequently, the frequency of cyberattacks is increasing. Fig. 1 comes close to it as well. The concepts are described here. Network, information, and application security are the three primary components of cybersecurity (Sarker, 2021).

Fig. 1: Security domains, including cybersecurity (Torres et al., 2019)

Application security is the process of securing an application by taking various actions. The application security vulnerabilities are normally dealt with through continuous monitoring and identification, resolution, and finally prevention of a possible threatening situation. Information security is an extensive system of measures aimed at maintaining the availability, integrity, and anonymity of corporate information on different platforms and in different formats.

Network security is particularly interested in the protection of computer systems against unauthorised access or malicious programs, regardless of whether these risks are intentional or accidental (Torres et al., 2019; Sarker, 2021). Conversely, operations security (OPSEC) focuses on the detection and safeguarding of unclassified, nonetheless, sensitive information; this is the information that enemies or business rivals can use to gain credible knowledge.On the other hand, internet security can be identified as the safeguards that are put in place in order to keep the privacy of the activities that take place online. This involves implementing policies and technical defences on browsers, operating-systems, networks and applications to counter cyberattacks. Information and Communication Technology (ICT) security, which may be understood in alignment with the CIA triad (Confidentiality, Integrity, Availability), and governed by the ISO standards, can be viewed as the principles of cybersecurity. Nevertheless, the end user is one of the most important and insecure elements. The human factor remains a crucial element, and about 50 percent of cybercrimes happen because of the user negligence or ignorance of digital risks, while human activity is responsible for about 90% of cyberattacks (Torres et al., 2019; Sarker, 2021). The aim of the cyberattacks is to gain access to or disable the target device. A range of attacks on the target can be used to achieve the goal.

Daily, a large number of cyberattacks take place and are evolving (Torres et al., 2019). Below is a quick discussion of them. (i) Malware. One kind of malicious software designed to damage a single device or network is called malware. Both more contemporary hostile software like spyware and ransomware, as well as more conventional malicious software like worms, malware, and trojans, fall under this category (Torres et al., 2019). Malware infects a device or network when a user reads an email file, clicks on a risky link, or downloads dangerous apps. The most important thing to keep in mind is that malware

spreads and reproduces through interactions with other devices or systems.

Among the causes are information collection, blocking connectivity to the internet, and installing new malicious software (Jahwar and Ameen, 2021). Jahwar and Ameen (2021) define phishing as the process of delivering spam messages that look like they are from a reliable source, typically by email. The objective is either to steal personal data, such as credit card numbers and login credentials, or to infect the victim's machine with malware. One type of cybercrime that is becoming more and more common is phishing (Torres et al., 2019). (iii) Spam: This type of email is unwanted. 12 In addition to wasting readers' time, spam emails may include automatically launching Java applets when the message is seen (Jahwar and Ameen, 2021).

2.2 Related work

In the past decade, IoD, ML, and privacy have come into contact. Several researchers have undertaken tutorials and surveys (mentioned in Tables 1 and 2; Rodríguez et al., 2022) to offer a roadmap for future solutions and a practical guide to manage cybersecurity concerns, given the importance of ML and cybersecurity in IoD contexts. But the majority of polls that are currently available focus on cybersecurity in IoD contexts while ignoring privacy. They tend to prioritize other cybersecurity risks, such as software-based or network-based attacks. However, the scant privacy-focused polls that have been released present a limited perspective because they mainly address particular privacy-preserving measures.

Table 1 presents a comprehensive overview of existing survey papers related to machine learning (ML) and deep learning (DL) applications in cybersecurity and privacy protection, particularly within the context of the Internet of Devices (IoD). This table, adapted from Rodríguez et al. (2022), categorizes these surveys based on their primary focus (e.g., cybersecurity ML, privacy-preserving DL), the area of application (e.g., IoD, 5G, mobile networking, cloud computing), and the specific scenario addressed. The inclusion of this table highlights the current landscape of research and identifies gaps that this review aims to address, specifically the limited focus on privacy in IoD contexts despite the growing concerns.

Table 1: The findings of previous surveys on ML and DL for privacy protection (Rodríguez et al., 2022)

Publication	Summary	Scope	Area	Scenario
Al-Garadi et al. (2020)	IoD security survey of ML and DL techniques.	Cybersecurity ML	IoD	
Hussain et al. (2020)	Survey of IoT network security solutions based on ML and DL.	Cybersecurity ML	IoD	
Waheed et al. (2020)	IoT security and privacy survey utilizing ML and BC.	Cybersecurity ML	IoD	
Khan et al. (2020)	analysis of the technology that went into creating the 5G privacy and security model.	Cybersecurity ML	5G	
Dixit &Silakari (2021)	review of deep learning techniques for cybersecurity applications.	Cybersecurity DL	IoD	

Rodriguez et al. (2021)	analysis of cybersecurity solutions for mobile networks based on DL.	Cybersecurity DL	Mobile networking
Gosselin et al. (2022)	An Evaluation of Security and Privacy in Federated Learning.	Cybersecurity FL	IoD
Rigaki& Garcia (2021)	A Survey of Security and Privacy in Federated Learning.	Privacy attacks ML	IoD
Tanuwidjaja et al. (2019)	Survey of privacy-preserving DL techniques.	Privacy-preserving DL	Mobile networking
Boulemtafes et al. (2020)	survey of DL privacy-preserving methods.	Privacy-preserving DL	Mobile networking
Liu et al. (2021)	Survey that reviews interactions between privacy and machine learning.	Privacy ML	Mobile networking
Zheng et al. (2019)	Privacy-preserving ML review for Cloud Computing and IoD.	Privacy-preserving ML	Cloud Computing
Seliem et al. (2018)	Survey of privacy threats in IoD environments.	Privacy-preserving	IoD
Amiri-Zarandi et al. (2020)	investigation of ML-based IoD privacy protection solutions.	Privacy-preserving ML	IoD
Kounoudes et al. (2020)	Survey of user-centric privacy protection approaches in IoD.	User-centric privacy-preserving	IoD
Zhu et al. (2021)	In IoD aggregation scenarios, examine privacy-preserving machine learning training strategies.	Privacy-preserving ML Training	IoD
El Ouadrhiri& Abdelhadi (2022)	A Survey on Differential Privacy for Federated and Deep Learning.	Differential Privacy FL	IoD

Table 1 reveals a growing interest in the application of ML and DL for cybersecurity and privacy within various technological domains. A significant number of the surveyed works, such as Al-Garadi et al. (2020), Hussain et al. (2020), and Waheed et al. (2020), primarily focus on general cybersecurity aspects in IoD environments, often leveraging ML and DL techniques for intrusion detection and threat mitigation. This indicates a strong foundation in using these advanced computational methods to bolster the security posture of IoD systems.

However, a closer examination of the "Scope" column highlights a crucial distinction: while many studies address "Cybersecurity ML" or "Cybersecurity DL," a smaller subset explicitly targets "Privacy-preserving DL," "Privacy ML," or "Privacy-preserving ML Training." This imbalance suggests that privacy, while acknowledged as a critical concern, may not always be the primary focus or is often

addressed implicitly within broader cybersecurity frameworks. For instance, Tanuwidjaja et al. (2019) and Boulemtafes et al. (2020) specifically delve into privacy-preserving DL techniques, and Zheng et al. (2019) and Amiri-Zarandi et al. (2020) investigate privacy-preserving ML solutions in Cloud Computing and IoD, respectively. This specialized focus on privacy-preserving methods is particularly relevant given the sensitive nature of data collected and transmitted by IoD devices.

The "Area" column further clarifies the context of these surveys. The prevalence of "IoD" across many entries underscores the relevance of this review to the Internet of Devices. Other areas like "5G," "Mobile networking," and "Cloud Computing" are also represented, indicating that privacy and security challenges often extend across interconnected technological ecosystems. The inclusion of Federated Learning (FL) in studies like Gosselin et al. (2022) and Rigaki& Garcia (2021) is noteworthy, as FL offers a promising paradigm for privacy-preserving machine learning by enabling collaborative model training without direct data sharing. This aligns with the increasing demand for privacy-by-design solutions in distributed IoD environments.

In conclusion, Table 1 demonstrates a robust body of research on ML/DL for cybersecurity in IoD. However, it also underscores the need for more explicit and dedicated investigations into privacy-preserving techniques. While general cybersecurity measures are essential, the unique challenges of data privacy in IoD, especially concerning sensitive personal information, necessitate a more targeted approach, which some of the more recent privacy-focused surveys have begun to address. This review aims to build upon these insights by offering a more detailed examination of specific ML-based privacy solutions for IoD.

Table 2 provides a detailed overview of specific research efforts focused on implementing Machine Learning (ML)-based privacy solutions within centralized learning Internet of Devices (IoD) setups. Adapted from Rodríguez et al. (2022), this table delineates key aspects of each study, including the type of attack addressed, the privacy protection model employed, the ML model utilized, the dataset used for experimentation, reported performance results (accuracy or AUC), and the specific scenario of application. This granular analysis is crucial for understanding the practical implementation and effectiveness of ML techniques in safeguarding data privacy in centralized IoD architectures.

Table 2: ML-based privacy in Centralized Learning IoD setups

Reference	Privacy Protection Model	ML Model	Dataset	Results (Accuracy or AUC)	Scenario
Rahulamathavan et al. (2014)	Encryption (Classification)	SVM	WBCPIDIRISJAFFE	98.24%, 86.98%, 87.33%, 89.67%	Cloud Computing
Wang et al. (2017)	Encryption (Classification)	ML-ELM	MNIST	79.83% (AES), 90.44% (DES)	Cloud Computing

Zhu et al. (2017)	Encryption (Classification)	SVM	PID	94%	IoD eHealth
Jiang et al. (2019)	Encryption (Training)	Homomorphic surf and fast image matching	DR1RetiDBMessidor	AUC: [86%, 89%]	IoD eHealth

****Attack for all = data leakage (DL)**

Table 2 offers valuable insights into the concrete applications of ML for privacy preservation in centralized IoD environments, primarily addressing "Data leakage" as the significant attack vector. A common theme across these studies is the reliance on "Encryption" as the primary privacy protection model. This underscores the fundamental role of cryptographic techniques in securing sensitive data during various stages of ML processing, including classification and training.

For instance, Rahulamathavan et al. (2014) and Wang et al. (2017) both employed encryption for classification tasks, albeit in a "Cloud Computing" scenario rather than direct IoD. Rahulamathavan et al. (2014) achieved high accuracies (ranging from 86.98% to 98.24%) using Support Vector Machines (SVM) on diverse datasets like WBC, PID, IRIS, and JAFFE. Similarly, Wang et al. (2017) utilized ML-ELM with encryption schemes (AES and DES) on the MNIST dataset, demonstrating varying but respectable accuracies of 79.83% and 90.44%. These results indicate the feasibility of combining ML with encryption to maintain data confidentiality without significantly compromising classification performance.

More directly relevant to IoD, Zhu et al. (2017) applied SVM with encryption for classification in an "IoD eHealth" scenario, achieving a 94% accuracy on the PID dataset. This highlights the applicability of these methods in critical domains where sensitive health data is handled. Jiang et al. (2019) further extended this by focusing on "Encryption (Training)" within an "IoD eHealth" context, using homomorphic techniques for image matching on medical datasets (DR1, RetiDB, Messidor) and reporting impressive AUC values between 86% and 89%. The use of homomorphic encryption is particularly significant as it allows computations on encrypted data, thus preserving privacy throughout the entire training process—a crucial advancement for ML in privacy-sensitive IoD applications.

The datasets used in these studies, such as PID for medical applications and MNIST for image recognition, reflect the diverse data types commonly encountered in IoD. The reported accuracy and AUC metrics consistently demonstrate that ML models can achieve high performance even when privacy-preserving techniques are integrated. Finally, Table 2 illustrates that encryption-based privacy protection models, combined with various ML algorithms, are effective in mitigating data leakage in centralized IoD and related cloud computing setups. The successful application in "IoD eHealth" scenarios is particularly promising, showcasing the potential for ML to enable secure and privacy-preserving data analysis in sensitive domains. These findings lay a strong foundation for further research into more advanced and scalable privacy-preserving ML techniques for the rapidly expanding Internet of Devices.

3.0 Machine Learning (ML) in Cybersecurity

Artificial intelligence (AI) and Machine learning (ML) is an effective data analysis methodology that allows the automatic generation of predictive models (Ademilua & Areghan, 2022). It is central to the process of improving the performance of numerous

computational tasks through optimizing algorithms and being involved in continuous learning (Jahwar & Ameen, 2021). Abdulqader et al. (2020) remark that machine learning methods can be categorized into two supervised and unsupervised learning, among others. The key objective of ML is designing systems that have the ability to learn relying on the data and that can conduct operations without the immediate intervention of the individual (Jahwar & Ameen, 2021). The most prevalent strategies that will be applied in the scenario of supervised learning are classification and regression. They can be applied bookishly in cybersecurity to identify numerous kinds of cyberattacks, such as scanning, spoofing, or denial-of-service attacks by naming them with the aid of classification algorithms (Sarker et al., 2019). The importance of machine learning in cybersecurity is hard to overestimate because this tool enables the enhancement of threat detection, prediction, and response. Its applications have been used in diverse areas like malware detection, Supervisory Control and Data Acquisition (SCADA) security, Industrial Control Systems (ICS) and Vehicular Ad-hoc Networks (VANETs) intrusion detection.

3.2 Classification of Machine Learning Algorithms to Cybersecurity

3.2. 1 Classical Machine Learning Techniques

Machine learning (ML) stands out as one of the rapidly expanding branches of computer science, which already provides various real-life applications (Yavanoglu & Aydos, 2017). The purpose of model construction is the analysis of input data using statistical (output prediction within set limits) methods: supervised learning, unsupervised learning, and reinforcement learning are the three most important machine learning algorithms (Yavanoglu & Aydos, 2017). Supervised learning is the most utilized methodology since most of the real-world applications rely on its core techniques, classification, and regression. Different algorithms have been presented and are utilized by the researchers on numerous machine learning problems (Ben Salem et al., 2008). Some of the famous algorithms employed in different research are the K-Means clustering (Xie et al., 2004), K-Nearest Neighbours (KNN) (Yusof et al., 2016), Random Forest (Hasan et al., 2016), Decision Trees (Salloum et al., 2021), Naive Bayes (Bhamare et al., 2016), Support Vector Machines (SVM) (Bhamare et al., 2016)

3.2.2 Deep Learning Techniques

The application of deep learning technology in intrusion detection has gained substantial popularity in recent years. According to Salloum et al. (2021) deep learning methods perform better than traditional approaches in intrusion detection tasks. The research of Javaid et al. (2016) demonstrates how a deep neural network (DNN) model can detect flow-based anomalies. Deep learning provides an effective solution for anomaly identification within software-defined networks (SDNs) as demonstrated by their research results Tang et al. (2016) developed their intrusion detection system using self-taught learning (STL) based deep learning techniques on the NSL-KDD benchmark dataset. Their model demonstrated superior detection performance when compared to previous models. These studies explain the utilization of deep learning for feature extraction as well as dimension reduction. Deep learning models require pre-training as their essential component. Researchers use classical supervised learning structures in their deep learning models for classification problems yet these models have rarely been applied to multiclass classification problems in existing literature. Furthermore, a minimal and effective architecture was advocated by Salloum et al. (2021) through the use of Recurrent Neural Networks (RNs) for intrusion detection. An RNN with three layers was proposed, which conveys 41 features as input and output four different types of

intrusion in a misuse-based intrus detection system (IDS)..

RNNs are not fully capable of representing high-dimensional data due to the incomplete connection between layer nodes. This limits their ability to support deep learning. Furthermore, previous studies have neglected to consider the evaluation of such models in the binary classification task. The shortcomings were resolved by Yin und al. (2017,18) through the introduction of an intrusion detection system (IDS) Recurrent Neural Networks (RN-IDSO), which utilized deep learning techniques. They have examined the model's performance in solving binary and multiclass classification problems. Important hyperparameters such as the number of neurons and learning rate were found to have a significant impact on the effectiveness of the system. The performance of RNN-IDS demonstrated that it offers superior classification accuracy and surpasses traditional machine learning classifiers in intrusion detection. The method does not only propose a novel way of developing IDS but also increases the accuracy of the detection of various classes of attacks.

4.0 Cybersecurity Datasets

Currently, numerous studies generate data for both community repositories and individual research (Yavanoglu and Aydo, 2017). Using artificial intelligence and machine learning research, the current security-related datasets are described in this part.

4.1 KDD Cup 1999 Dataset

The tcpdump and BSM list files are part of the data obtained from MIT Lincoln Labs, in addition to KDD Cup 1999. This dataset was created using the information collected during the DARPA'98 study (Yavanoglu and Aydo, 2017). Additionally, This dataset is utilized as reference for the IDS evaluation (Fraley and Cannady, 2017). The KDD'99 is a widely used data set to evaluate anomaly detection methods and its effectiveness (Yavanoglu and Aydo, 2018). A number of studies are presently utilizing the KDD dataset (Neethu, 2013; Kozik et al. 2015).

4.2 ISOT (Information Security and Object Technology)

The 1,675,424 traffic flows in the ISOT dataset are made up of both ordinary network traffic and publicly accessible botnet traffic. The Storm and Waledac botnets, which are a component of the French Honeynet Project, are the source of the malicious traffic in the ISOT dataset (Yavanoglu and Aydos, 2017).

4.3 HTTP CSIC 2010 Dataset

A set of thousands upon thousands of web requests for testing web attack detection and defense systems were created by the Information Security Institute of CSIC using automated tools (Salloum et al, 2010). In the dataset, there are 25,000 malicious queries and 6,000 valid requests. Normal and abnormal are the two categories of HTTP queries in the dataset, as explained by Yavanoglu et al. (2017). It is useful as a source for web-based intrusion detection in research papers such as Salloum et al. (2021) and Saad & Co (2011).

4.4 CTU-13(Czech Technical University) Dataset

CTU-13 dataset was made by the Czech Technical University for 13 different malware capture scenarios simulating a real-world network environment. This dataset exists for delivering real-world heterogeneous botnet traffic data. According to Yavanoglu and Aydos (2017), normal traffic is generated from the legitimate hosts verified by the dataset, while botnet traffic is generated from compromised machines. Even though the CTU-13 dataset has the major advantage of a highly detailed annotation of activities performed in controlled condition (Yusof et al., 2016; Salloum et al., 2021),

4.5 UNSW-NB15 Dataset

Using the IXIA Perfect Storm software within the Cyber Range Lab at the ACCS, the UNSW-NB15 dataset was manufactured (Yavanoglu and Aydos, 2017). The dataset contains one

hour of anonymised traffic data, which includes a 2007 DDoS attack (Salloum et al., 2021). There are nine types of attacks in this dataset, including fuzzers, backdoors, and analysis, as well as exploits, DoS, and reconnaissance, generic attacks, worms, and shellcodes (Yavanoglu and Aydos, 2017).

4.6 Privacy Challenges and Threats in IoD

This segment investigates privacy issues relating to IoD. We present, therefore, an overview of IoD systems to identify the privacy issues. The IoT architecture contains four basic layers comprising the Perception Layer, together with the Network Layer and Middleware Layer and Application Layer (Rodríguez et al., 2023; Juuti et al., 2019). The Perception Layer contains sensors together with actuators and embedded devices which perform data collection from physical environments. Sensors detect parameters such as motion, acceleration, temperature, humidity, and air quality, whereas actuators regulate how physical objects behave—for instance, adjusting a vehicle's acceleration in response to real-time data.

A vast volume of information is produced by devices in the Perception Layer and transmitted for further processing and secure routing to the Network Layer. Through the IoD architecture, the data is processed and transmitted by the Network Layer. In order to handle vendor-specific services and application requirements, a middleware layer is occasionally introduced as a link between the network and software layers. The IoD top layer is called the layer of apps. It uses data that has been processed from various IoD devices. It performs functions unique to a given application.

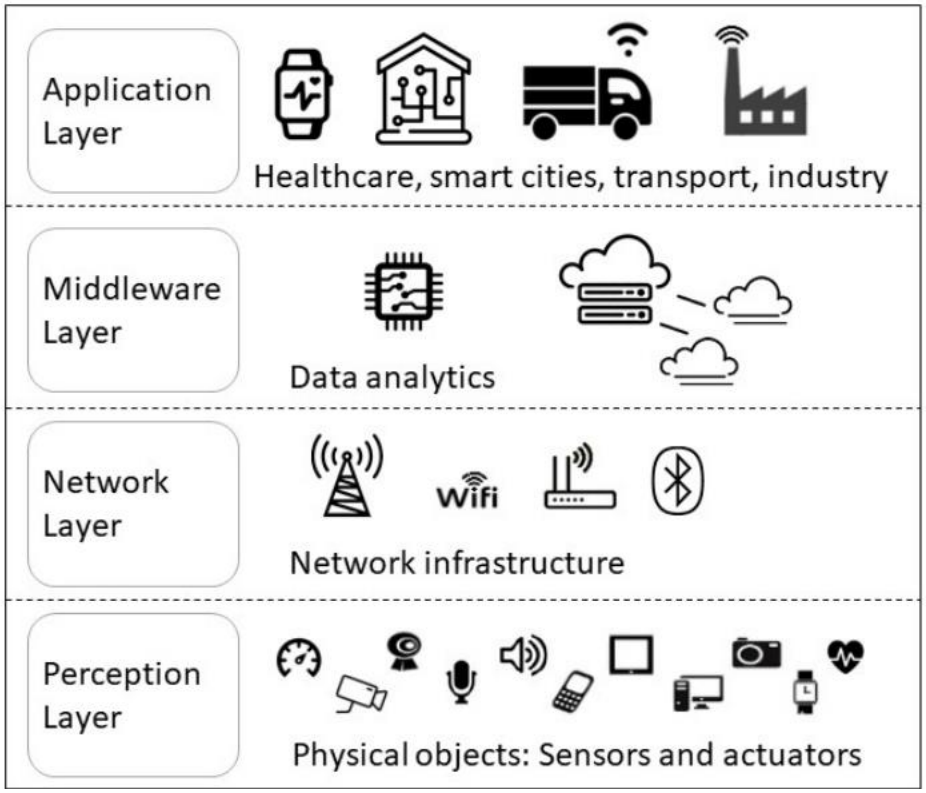

Fig. 2: Framework of the Internet of Devices (Rodríguez et al., 2023).

4.7 ML-Powered IoD Privacy Devices

ML and cybercrime have been interacting for a while. The new IoT applications and surroundings are not sufficiently protected by traditional privacy protection methods. Figure 3 illustrates the three primary layers that make up fog-IoT frameworks. A group of end devices that gather and analyze data make up

one end of the Internet of Things layer. They send information to the Fog layer, which processes and stores data and establishes a connection with the cloud, the next tier of the hierarchy. The middle layer of the construction is made up of the fog layer. IoT devices and fog nodes can rely on the cloud, the top layer, for dependable services and support, particularly for jobs requiring a lot of resources (Adetunji & Areghan, 2022).

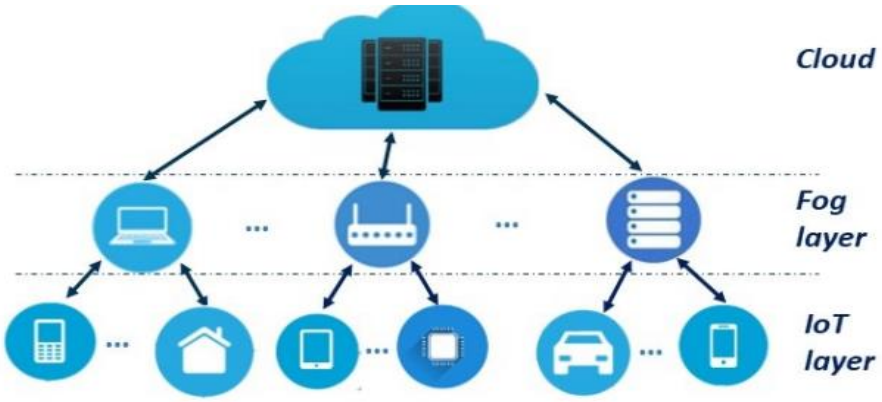

Fig.3: Design of an IoD system (Rodríguez et al., 2023)

ML and DL techniques are currently being examined for privacy due to their success in other cybersecurity domains. Nevertheless, there are two sides to ML and DL approaches. On the one hand, their detection skills are superior. The adoption of these technologies for Internet of Devices (IoD) services and applications remains limited, thus creating privacy risks because sensitive user data might get exploited. This section examines the current literature about how machine learning (ML) approaches work to boost privacy safeguards within IoT systems. The analyzed studies follow two main protection methods, which include encryption alongside differential privacy while also using one of three learning designs, such as centralized distributed or federated systems. We examine the threat or attack that each remedy averts. We further investigate the architecture, paying particular focus on the machine learning approach or methods, how they are executed, the testing data sets, and the outcomes. The full study is included in Tables 2-3, while the overview is given in Figure 3 (Rodríguez et al., 2023).

Table 3 presents a concise summary of various research efforts that have applied machine learning (ML) and deep learning (DL) techniques to address cybersecurity challenges, with a particular focus on intrusion detection systems (IDS). This table, derived from Jahwar and Ameen (2021), details the specific ML/DL techniques employed, the datasets used for experimentation, the primary purpose of the developed system, and the achieved accuracy. This compilation serves to illustrate the diversity of approaches and the performance benchmarks attained in recent studies aimed at enhancing cybersecurity through intelligent algorithms.

Table 3 provides a comprehensive overview of various ML and DL techniques applied to intrusion detection, demonstrating the diverse methodologies and their efficacy. The "Techniques" column highlights a range of algorithms, from traditional machine learning methods like K-Nearest Neighbours (KNN), Decision Trees (DT), Support Vector Machines (SVM), and Random Forest (RF) to advanced deep learning architectures such as Deep Belief Networks (DBN), Deep Neural Networks (DNN), Recurrent Neural Networks (RNN), and Multi-Layer Perceptrons (MLP). This diversity indicates that researchers are

exploring a wide spectrum of computational intelligence to address the complex problem of cybersecurity. Some studies also incorporate feature reduction or selection techniques like Principal Component Analysis (PCA) or rule-based methods (e.g., Signature Apriori algorithm) alongside classifiers to enhance performance.

Table 3: A summary of cybersecurity-related machine learning and deep learning tasks (Jahwar and Ameen, 2021)

Authors	Techniques	Datasets	Purpose	Accuracy (%)
Jahwar and Ameen (2021)	PSA and KNN	NSL-KDD	Intrusion Detection System	94.00
Atefi et al. (2019)	KNN and DNN	CICIDS-2017	Anomaly Analysis for the Classification Purpose of Intrusion Detection Systems	KNN = 92.93 DNN = 88.24
Ahmed et al. (2019)	DT	KDDCUP99	Enhance a hybrid Intrusion Detection System	99.80
Kwon et al. (2019)	DBN	NSL-KDD	Anomaly Detection System	97.5%
Almiani et al. (2020)	RNN	NSL-KDD	Anomaly intrusion detection system and attack classification	98.27
Waskle et al. (2020)	PCA and RF	KDDCUP99	Intrusion Detection System	96.78
Kumari, A., and Mehta (2020)	J48 DT and SVM	KDDCUP99	Hybrid Intrusion Detection System	99.60
Chen et al. (2020)	Signature Apriori algorithm	CICIDS2017	Novel Network Intrusion Detection System	99.56
Thirumalairaj and Jeyakarthi (2020)	MLP and PID	CICIDS2017	Intrusion Detection System	98.96
Krishna et al. (2020)	MLP	KDDCUP99	Intrusion Detection and Prevention System	91.40

The "Datasets" column reveals that NSL-KDD, KDDCUP99, and CICIDS-2017 are the most frequently used benchmark datasets for evaluating intrusion detection systems. KDDCUP99, while historically significant, is known for its limitations such as redundant records, which NSL-KDD aims to address. CICIDS-2017, being a more recent dataset, offers a broader range of up-to-date attack scenarios and realistic network traffic, making it a valuable resource for contemporary intrusion detection research. The choice of dataset significantly influences the generalizability and real-world applicability of the developed IDS.The "Purpose" column consistently centers on "Intrusion Detection System" or variations thereof, such as "Anomaly Detection System" and "Intrusion

Detection and Prevention System." This reinforces the critical role of ML/DL in identifying and often preventing malicious activities in networks. The aims range from general intrusion detection to specific anomaly analysis and the enhancement of hybrid IDS, indicating a focus on both known and unknown threats.

Notably, the "Accuracy" column presents high performance metrics across almost all studies, with many achieving accuracies well above 90%, and several even reaching close to 100% (e.g., Ahmed et al. (2019) at 99.8% and Kumari and Mehta (2020) at 99.6% on KDDCUP99, and Chen et al. (2020) at 99.56% on CICIDS2017). While these high accuracy rates are promising, it's important to consider other metrics like false positive rates and recall in a real-world cybersecurity context, as a high accuracy might sometimes be misleading if the system fails to detect certain types of critical attacks or generates too many false alarms. However, these reported accuracies do demonstrate the strong potential of ML/DL for robust intrusion detection. Finally, Table 3 vividly illustrates the robust and evolving landscape of ML and DL applications in cybersecurity. The continuous development and refinement of these intelligent techniques, coupled with the use of increasingly realistic datasets, are contributing significantly to the creation of more effective and proactive intrusion detection and prevention systems, which are vital for securing the Internet of Devices and other critical digital infrastructures.

The majority of the ML and DL jobs in the cyber area are covered in this review. We are aware that the majority of them are rather accurate. But according to the published research in Table 1, a number of investigations have revealed the enhanced accuracy of intrusion detection by using machine learning (ML) and deep learning (DL) methods on different datasets (Almian et al., 2021; Kumari and Mehta, 2020; Jahwar and Ameen, 2021). Almian et al. (2021) improved a hybrid intrusion detection system with the KDDCUP99 dataset and presented a higher accuracy. They suggested that the new method, which uses decision tree methods, i.e., the C4.5 algorithm, can improve the accuracy of detection and reduce the false positive rate. Likewise, Thirumalairaj and Jeyakarthic (2020) proposed a hybrid intrusion detection system introduced with two ML algorithms, i.e., J48 Decision Tree (DT) and Support Vector Machine (SVM), to select the relevant features using the KDD CUP dataset. To detect intrusion messages faster, Krishna et al. (2020) proposed a method to quickly find the related attack signature based on known signatures. Jahwar and Ameen (2021) presented two learning methods, namely, Perimeter Intrusion Detection (PID) with Multi-Layer Perceptron (MLP) and PID with MLP and quantum classifier algorithm. Quantum classification is used in these methods to further improve intrusion detection performance, and indeed, these methods have been shown to improve the major performance measures of accuracy, efficiency, and consistency. In addition, the model suggested by (Kumari and Mehta, 2020) uses a multilayer RNN.

It is intended to be used in fog computing security in a fashion similar to Internet-based security. It has been demonstrated that the proposed model used a balanced version of the difficult dataset, NSL-KDD. Nevertheless, compared to the high-accuracy algorithms previously published, some of the techniques covered in this paper are less accurate. As an illustration, consider the Intrusion Detection System suggested by Jahwar and Ameen (2021). The KDDCUP99 dataset was used to train the Multi-Layer Perceptron (MLP), a deep learning model, in their instance. The system makes detection and avoidance of intrusions more dependable and efficient by integrating detection and response systems into a single framework. In general, we discovered that the

majority of ML and DL algorithms have improved accuracy after reviewing them.

5.0 Conclusion and Future Direction

The issue of cybersecurity has become one of the most important national and global-level issues since the number and complexity of cyberattacks have grown significantly. This paper provides a full survey of the problem of cybersecurity as it is solved using different machine learning (ML) methods. Various researchers have relied on an array of data to investigate the effectiveness of ML and artificial intelligence (AI) in securing computer systems. The paper has a comprehensive literature review about ML, deep learning (DL), and data mining (DM) approaches that have been applied in the cybersecurity field. It classifies the popular datasets in use, their benefits and shortcomings, and shows exemplary research works on how the ML/DL and DM approaches are applied in cyber threat detection. Besides, the paper focuses on recent developments of ML and DL-based intrusion detection in the ast three years. Since the heart of the matter in training and testing of intrusion detection systems lies with the datasets, the accessibility of the extensive and representative datasets is a significant concern. In the absence of such datasets, the ML and DL models cannot work effectively. The applicability of ML and DL to detect malware and intrusions is also discussed in the work and the strengths and weaknesses of these approaches are provided. In the future, it is suggested that well-organized datasets that can be accessed publicly will be created to allow the further advancement of this area. For upcoming researchers and technologists in this field, this study offers a knowledge base. The study examines each work's design, data set, published findings, implementation, and ML approach or methods. We attempted to compare the performance of the various proposals. Before classifying their datasets, works that address direct privacy data leakage attacks encrypt them.

6.0 References

Abdelghafar, S., Darwish, A., &Hassanien, A. E. (2020). Intelligent health monitoring systems for space missions based on data mining techniques. In *Machine learning and data mining in aerospace technology* (pp. 65–78). Springer.

Ademilua, D. A., and Areghan, E. (2022). AI-Driven Cloud Security Frameworks: Techniques, Challenges and Lessons from Case Studies. *Communication in Physical Sciences*, 8(4): 674-688

Ahmed, M. R. A. G., & Ali, F. M. A. (2019). Enhancing hybrid intrusion detection and prevention system for flooding attacks using decision tree. In *2019 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)* (pp. 1–4). IEEE.

Alazab, M., & Tang, M. (2019). *Deep learning applications for cyber security*. Springer.

Alhashmi, S. F. S., Salloum, S. A., & Abdallah, S. (2019). Critical success factors for implementing artificial intelligence (AI) projects in Dubai government United Arab Emirates (UAE) health sector: Applying the extended technology acceptance model (TAM). In *International Conference on Advanced Intelligent Systems and Informatics* (pp. 393–405). Springer.

Almiani, M., AbuGhazleh, A., Al-Rahayfeh, A., Atiewi, S., &Razaque, A. (2020). Deep recurrent neural network for IoT intrusion detection system. *Simulation Modelling Practice and Theory*, *101*, 102031.

Alomari, K. M., AlHamad, A. Q., &Salloum, S. (2019). Prediction of the digital game rating systems based on the ESRB. *Opción*, *35*(19), 1368–1393.

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., &Mané, D. (2016). *Concrete problems in AI safety*. arXiv preprint arXiv:1606.06565.

Areghan, E. (2023). From Data Breaches to Deep Fakes: A comprehensive Review of Evolving Cyber Threats and Online Risk

Management. *Communication in Physical Sciences*, *9*(4), 738–753.

Atefi, K., Hashim, H., &Kassim, M. (2019). Anomaly analysis for the classification purpose of intrusion detection system with K-nearest neighbors and deep neural network. In *2019 IEEE 7th Conference on Systems, Process and Control (ICSPC)* (pp. 269–274). IEEE.

Benaddi, H., Ibrahimi, K., & Benslimane, A. (2018). Improving the intrusion detection system for nslkdd dataset based on pca-fuzzy clustering-knn. In *2018 6th International Conference on Wireless Networks and Mobile Communications (WINCOM)* (pp. 1–6). IEEE.

Ben Salem, M., Hershkop, S., &Stolfo, S. J. (2008). A survey of insider attack detection research. In *Insider attack and cyber security* (pp. 69–90). Springer.

Bhamare, D., Salman, T., Samaka, M., Erbad, A., & Jain, R. (2016). Feasibility of supervised machine learning for cloud security. In *2016 International Conference on Information Science and Security (ICISS)* (pp. 1–5).

Bonaci, T., Herron, J., Yusuf, T., Yan, J., Kohno, T., &Chizeck, H. J. (2015). *To make a robot secure: An experimental analysis of cyber security threats against teleoperated surgical robots*. arXiv preprint arXiv:1504.04339.

Buczak, A. L., &Guven, E. (2015). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, *18*(2), 1153–1176.

Chen, L., Kuang, X., Xu, A., Suo, S., & Yang, Y. (2020). A Novel Network Intrusion Detection System Based on CNN. In *2020 Eighth International Conference on Advanced Cloud and Big Data (CBD)* (pp. 243–247). IEEE.

Chowdhury, S., et al. (2017). Botnet detection using graph-based feature clustering. *Journal of Big Data*, *4*(1), 14.

da Costa, K. A. P., Papa, J. P., Lisboa, C. O., Munoz, R., & de Albuquerque, V. H. C. (2019). Internet of Things: A survey on machine learning-based intrusion detection approaches. *Computer Networks*, *151*, 147–157.

Darwish, A., Ezzat, D., &Hassanien, A. E. (2020). An optimized model based on convolutional neural networks and orthogonal learning particle swarm optimization algorithm for plant diseases diagnosis. *Swarm and Evolutionary Computation*, *52*, 100616.

Dua, S., & Du, X. (2016). *Data mining and machine learning in cybersecurity*. Auerbach Publications.

Elsayad, D., Ali, A., Shedeed, H. A., &Tolba, M. F. (2020). PAGeneRN: Parallel architecture for gene regulatory network. In *Data analytics in medicine: Concepts, methodologies, tools, and applications* (pp. 1052–1075). IGI Global.

Fraley, J. B., & Cannady, J. (2017). The promise of machine learning in cybersecurity. *SoutheastCon 2017*, 1–6.

Gallagher, B., &Eliassi-Rad, T. (2009). *Classification of http attacks: A study on the ECML/PKDD 2007 discovery challenge*. Lawrence Livermore National Lab. (LLNL), Livermore, CA (United States).

Hacioglu, U., &Sevgilioglu, G. (2019). The evolving role of automated systems and its cyber-security issue for global business operations in Industry 4.0. *International Journal of Business Ecosystem & Strategy*, *1*(1), 1–11.

Haddadi, F., Le Cong, D., Porter, L., &Zincir-Heywood, A. N. (2015). On the effectiveness of different botnet detection approaches. In *International Conference on Information Security Practice and Experience* (pp. 121–135). Springer.

Hasan, M. A. M., Nasser, M., Ahmad, S., &Molla, K. I. (2016). Feature selection for intrusion detection using random forest.

Journal of Information Security, *7*(03), 129.

Hoque, N., Bhattacharyya, D. K., &Kalita, J. K. (2016). A novel measure for low-rate and high-rate DDoS attack detection using multivariate data analysis. In *2016 8th International Conference on Communication Systems and Networks (COMSNETS)* (pp. 1–2). IEEE.

Jahwar, A. F., & Ameen, S. Y. (2021). A Review on Cybersecurity based on Machine Learning and Deep Learning Algorithms. *Journal of Soft Computing and Data Mining*, *2*(2), 14–25.

Javaid, A., Niyaz, Q., Sun, W., &Alam, M. (2016). A deep learning approach for network intrusion detection system. In *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)* (pp. 21–26).

Jones, C. L., Bridges, R. A., Huffer, K. M. T., & Goodall, J. R. (2015). Towards a relation extraction framework for cyber-security concepts. In *Proceedings of the 10th Annual Cyber and Information Security Research Conference* (p. 11).

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260.

Kato, K., &Klyuev, V. (2014). An intelligent DDoS attack detection system using packet analysis and support vector machine. *IJICR*, 478–485.

Kozik, R., Choraś, M., Renk, R., &Hołubowicz, W. (2015). A proposal of algorithm for web applications cyber attack detection. In *IFIP International Conference on Computer Information Systems and Industrial Management* (pp. 680–687). Springer.

Krishna, A., Lal, A., Mathewkutty, A. J., Jacob, D. S., & Hari, M. (2020). Intrusion Detection and Prevention System Using Deep Learning. In *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 273–278). IEEE.

Kumari, A., & Mehta, V. (2020). Hybrid Intrusion Detection System using J48 Decision Tree and SVM. *International Journal of Computer Applications*, *174*(17), 1-5.

Kwon, D., Kim, H., Kim, J., Suh, S. C., Kim, I., & Kim, K. J. (2019). A survey of deep learning-based network anomaly detection. *Cluster Computing*, *22*(1), 949–961.

Li, J. (2018). Cyber security meets artificial intelligence: A survey. *Frontiers of Information Technology & Electronic Engineering*, *19*(12), 1462–1474.

Lippmann, R., Haines, J. W., Fried, D. J., Korba, J., & Das, K. (2000). The 1999 DARPA off-line intrusion detection evaluation. *Computer Networks*, *34*(4), 579–595.

Liu, Q., Li, P., Zhao, W., Cai, W., Yu, S., & Leung, V. C. M. (2018). A survey on security threats and defensive techniques of machine learning: A data driven view. *IEEE Access*, *6*, 12103–12117.

McNeil, N., Bridges, R. A., Iannacone, M. D., Czejdo, B., Perez, N., & Goodall, J. R. (2013). Pace: Pattern accurate computationally efficient bootstrapping for timely discovery of cybersecurity concepts. In *2013 12th International Conference on Machine Learning and Applications* (Vol. 2, pp. 60–65). IEEE.

Mukkamala, S., Sung, A., & Abraham, A. (2005). Cyber security challenges: Designing efficient intrusion detection systems and antivirus tools. In V. R. Vemuri (Ed.), *Enhancing computer security with smart technology 2006* (pp. 125–163).

Neethu, B. (2013). Adaptive intrusion detection using machine learning. *International Journal of Computer Science & Network Security*, *13*(3), 118.

Nguyen, H. T., Torrano-Gimenez, C., Alvarez, G., Petrović, S., & Franke, K. (2011). Application of the generic feature selection measure in detection of web attacks. In *Computational intelligence in security for information systems* (pp. 25–32). Springer.

Olawale, A., Ajoke, O., &Adeusi, C. (2020). Quality Assessment and Monitoring of Networks Using Passive Technique. *Review of Computer Engineering Research*, *7*(2), 54–61. https://doi.org/10.18488/journal.76.2020.72.54.61

Pacheco, A. G. C., Ali, A.-R., &Trappenberg, T. (2019). *Skin cancer detection based on deep learning and entropy to detect outlier samples*. arXiv preprint arXiv:1909.04525.

Papernot, N., McDaniel, P., Sinha, A., & Wellman, M. (2016). *Towards the science of security and privacy in machine learning*. arXiv preprint arXiv:1611.03814.

Rodríguez, E., Otero, B., & Canal, R. (2023). A Survey of Machine and Deep Learning Methods for Privacy Protection in the Internet of Things. *Sensors*, *23*(3), 1252. https://doi.org/10.3390/s23031252

Saad, S., et al. (2011). Detecting P2P botnets through network behavior analysis and machine learning. In *2011 Ninth Annual International Conference on Privacy, Security and Trust* (pp. 174–180). IEEE.

Salloum, S. A., et al. (2017). Mining social media text: Extracting knowledge from facebook. *International Journal of Computer Digital Systems*, *6*(2), 73–81.

Salloum, S. A., et al. (2017). Mining text in news channels: A case study from facebook. *International Journal of Information Technology and Language Studies*, *1*(1), 1–9.

Salloum, S. A., Al-Emran, M., &Shaalan, K. (2017). A survey of text mining in social media: Facebook and twitter perspectives. *Advanced Science, Technology and Engineering Systems Journal*, *2*(1), 127–133.

Salloum, S. A., Al-Emran, M., Abdallah, S., &Shaalan, K. (2017). Analyzing the Arab Gulf newspapers using text mining techniques. In *International Conference on Advanced Intelligent Systems and Informatics* (pp. 396–405). Springer.

Salloum, S. A., AlHamad, A. Q., Al-Emran, M., &Shaalan, K. (2018). *A survey of Arabic text mining* (Vol. 740). Springer.

Salloum, S. A., Mhamdi, C., Al-Emran, M., &Shaalan, K. (2017). Analysis and classification of Arabic newspapers' facebook pages using text mining techniques. *International Journal of Information Technology and Language Studies*, *1*(2), 8–17.

Salloum, S. A., et al. (2020). Machine Learning and Deep Learning Techniques for Cybersecurity: A Review. In *The International Conference on Artificial Intelligence and Computer Vision*. Springer. https://doi.org/10.1007/978-3-030-44289-7_5

Tang, T. A., Mhamdi, L., McLernon, D., Zaidi, S. A. R., &Ghogho, M. (2016). Deep learning approach for network intrusion detection in software defined networking. In *2016 International Conference on Wireless Networks and Mobile Communications (WINCOM)* (pp. 258–263). IEEE.

Thirumalairaj, A., &Jeyakarthic, M. (2020). Perimeter Intrusion Detection with Multi-Layer Perception using Quantum Classifier. In *2020 Fourth International Conference on Inventive Systems and Control (ICISC)* (pp. 348–352). IEEE.

Torrano-Gimenez, C., Perez-Villegas, A., & Alvarez, G. (2009). A self-learning anomaly-based web application firewall. In *Computational Intelligence in Security for Information Systems* (pp. 85–92). Springer.

Torrano-Gimenez, C., Pérez-Villegas, A., Álvarez, G., Fernández-Medina, E., Malek, M., & Hernando, J. (2009). An anomaly-

based web application firewall. In *SECRYPT* (pp. 23–28).

Torrano-Giménez, C., Perez-Villegas, A., & Alvarez Maranón, G. (2010). *An anomaly-based approach for intrusion detection in web traffic*.

Waskle, S., Parashar, L., & Singh, U. (2020). Intrusion detection system using PCA with random forest approach. In *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 803–808). IEEE.

Xiao, L., Wan, X., Lu, X., Zhang, Y., & Wu, D. (2018). *IoT security techniques based on machine learning*. arXiv preprint arXiv:1801.06275.

Xie, M., Hu, J., & Slay, J. (2014). Evaluating host-based anomaly detection systems: Application of the one-class SVM algorithm to ADFA-LD. In *2014 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)* (pp. 978–982). IEEE.

Xin, Y., et al. (2018). Machine learning and deep learning methods for cybersecurity. *IEEE Access*, *6*, 35365–35381.

Yavanoglu, O., &Aydos, M. (2017). A review on cyber security datasets for machine learning algorithms. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 2186–2193). IEEE.

Yin, C., Zhu, Y., Fei, J., & He, X. (2017). A deep learning approach for intrusion detection using recurrent neural networks. *IEEE Access*, *5*, 21954–21961.

Yusof, A. R., Udzir, N. I., &Selamat, A. (2016). An evaluation on KNN-SVM algorithm for detection and prediction of DDoS attack. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems* (pp. 95–102). Springer.

Zhang, Q., Man, D., & Yang, W. (2009). Using HMM for intent recognition in cyber security situation awareness. In *2009 Second International Symposium on Knowledge Acquisition and Modeling* (Vol. 2, pp. 166–169). IEEE.

Compliance with Ethical Standards

Declaration

Ethical Approval

Not Applicable

Availability of Data

Data shall be made available upon request.

Competing interests

The author declared no competing interests

Funding

The authors declare that they have no known competing financial interests

Author's Contribution

The work was designed and written by the author.