Journal of Soft Computing and Computer Applications

Volume 1 | Issue 1 Article 1000

2024

Surveying Machine Learning in Cyberattack Datasets: A Comprehensive Analysis

Azhar F. Al-zubidi
Department of Computer Science, University of Technology -- Iraq, Baghdad 10066, Iraq, cs.21.07@grad.uotechnology.edu.iq

Alaa Kadhim Farhan
Department of Computer Science, University of Technology -- Iraq, Baghdad 10066, Iraq, alaa.k.farhan@uotechnology.edu.iq

El-Sayed M. El-Kenawy
Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura, 35111, Egypt, skenawy@ieee.org

Recommended Citation

Al-zubidi, Azhar F.; Farhan, Alaa Kadhim; and El-Kenawy, El-Sayed M. (2024) "Surveying Machine Learning in Cyberattack Datasets: A Comprehensive Analysis," *Journal of Soft Computing and Computer Applications*: Vol. 1: Iss. 1, Article 1000.
DOI: https://doi.org/10.70403/3008-1084.1000

JOURNAL OF SOFT COMPUTING AND COMPUTER APPLICATIONS 2024;1:1000

REVIEW

Surveying Machine Learning in Cyberattack Datasets: A Comprehensive Analysis

Azhar F. Al-zubidi [a,*], Alaa Kadhim Farhan [a], El-Sayed M. El-Kenawy [b]

[a] Department of Computer Science, University of Technology – Iraq, Baghdad 10066, Iraq
[b] Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura, 35111, Egypt

ABSTRACT

Cyberattacks have become one of the most significant security threats that have emerged in the last couple of years. It is imperative to comprehend such attacks; thus, analyzing various kinds of cyberattack datasets assists in constructing the precise intrusion detection models. This paper tries to analyze many of the available cyberattack datasets and compare them with many of the fields that are used to detect and predict cyberattack, like the Internet of Things (IoT) traffic-based, network traffic-based, cyber-physical system, and web traffic-based. In the present paper, an overview of each of them is provided, as well as the course of machine learning that has employed these datasets. From this survey, the researchers and the cybersecurity professional can derive a convenient classification of these datasets and their usages based on reviewing recent papers in this field. Furthermore, the types of machine learning involved in such systems as well as the intrusion detection and anomaly detection systems used to learn the models are presented in this paper. These techniques include deep learning models, random forests, support vector machines, and other commonly applied methods. Each technique has its advantages and limitations in the context of cyberattack prediction and detection. The paper is also consider factors like the specific technique and tools used, the type of attacks taken and the accuracy rate achieved. Of a total of 85 papers, 34 were selected for review in this paper. This survey is an essential tool for improving knowledge about the state of cyber detection and prediction techniques today.

Keywords: Cyberattack datasets, Machine learning, Network traffic-based datasets, Internet of things traffic-based datasets, Cyber-physical system traffic-based datasets, Web traffic-based datasets

1. Introduction

Cybersecurity is a rapidly growing field dedicated to protecting critical infrastructure, sensitive information, and user privacy, and this rapid growth is attributed to the advanced

Received 12 March 2024; accepted 12 May 2024.
Available online 27 June 2024

* Corresponding author.
E-mail addresses: cs.21.07@grad.uotechnology.edu.iq (A. F. Al-zubidi), alaa.k.farhan@uotechnology.edu.iq (A. Kadhim Farhan), skenawy@ieee.org (E. M. El-Kenawy).

https://doi.org/10.70403/3008-1084.1000

digital technologies used by hackers. As the threats of cyberattacks have become more crucial, cybersecurity is now a discipline that is required. Its goal is not limited to prevention of attacks but also embraces guarding devices, networks, and valuable data. Cybersecurity specialists strive to stay one step ahead of the threats and to safeguard the digital world against new and evolving threats. This is especially so given the nature and complexity of contemporary attacks that are not only highly intelligent but also adaptable, leading to calls for reconsideration of traditional network security approaches [1].

There have been various approaches in the use of the machine learning and deep learning models in creating good models that can identify cyber threats. However, the challenges of the application of these methods are the quality and the availability of data sets that can provide various types and behavior [2]. The datasets are used for training, testing, and evaluating the detection models and to perform comparative analysis and benchmarking. Hence, we need to review the available cyberattack datasets and evaluate their properties, advantages, and disadvantages [3].

Machine learning and deep learning methods have proven to be effective approaches for combating cyber threats. The kind of algorithms that can be implemented in this context are capable of processing large amount of data, identify patterns, and make informed decisions. These applications, for example, can vary from intruding detection to anomaly recognition. The validity of these models depends on many characteristics like, quality of data set which is one of the most essential factors while training a good and efficient machine learning model for recognizing different cyberattacks for defense purpose. They act as the basis for the reliable models by giving the various and balanced inputs essential elements in training accurate models. Challenges arise due to imbalanced class distributions, noisy data, and the need for features engineering. So, researchers must address these issues to build effective and accurate detection of agathism and privacy concerns in cyberattack datasets [4]. Through these techniques they are able to sift through large volumes of data, look for patterns, and make informed decisions. From intrusion detection to anomaly recognition, their application is diverse and impactful when working with real-world cyberattack datasets, and privacy and ethical considerations become crucial because they contain information about victims, attackers, and affected organizations. These concerns arise when the data includes personally identifiable information, confidential business details, or sensitive government information. Accordingly, researchers and practitioners must handle such data carefully to prevent accidental exposure and uphold privacy standards [5, 6].

The contribution of this paper is to survey the existing literature and open data sources on cybersecurity and cyber risk, emphasizing on the datasets that have been used to enhance academic knowledge and progress the current state-of-the-art in cybersecurity detection. In addition, it takes relevant information about these datasets and advocates for open data and the standardization of cyber risk data for academic comparability and replication.

The paper organization will proceed as follows: Section 2 presents a comprehensive overview of available cyberattack datasets for studying cyber threats. Section 3 explains machine learning techniques and deep learning methods used in intrusion detection. Section 4 presents the cyberattack datasets reviewed, analyzing specific datasets, their applications, and limitations. Finally, the conclusion of key findings and importance of standardized data in cybersecurity research are presented in Section 5.

2. Cyberattack datasets

In this section, a review of publicly available cyberattack datasets is provided and organized into a classification diagram based on their sources, features, formats, and attack

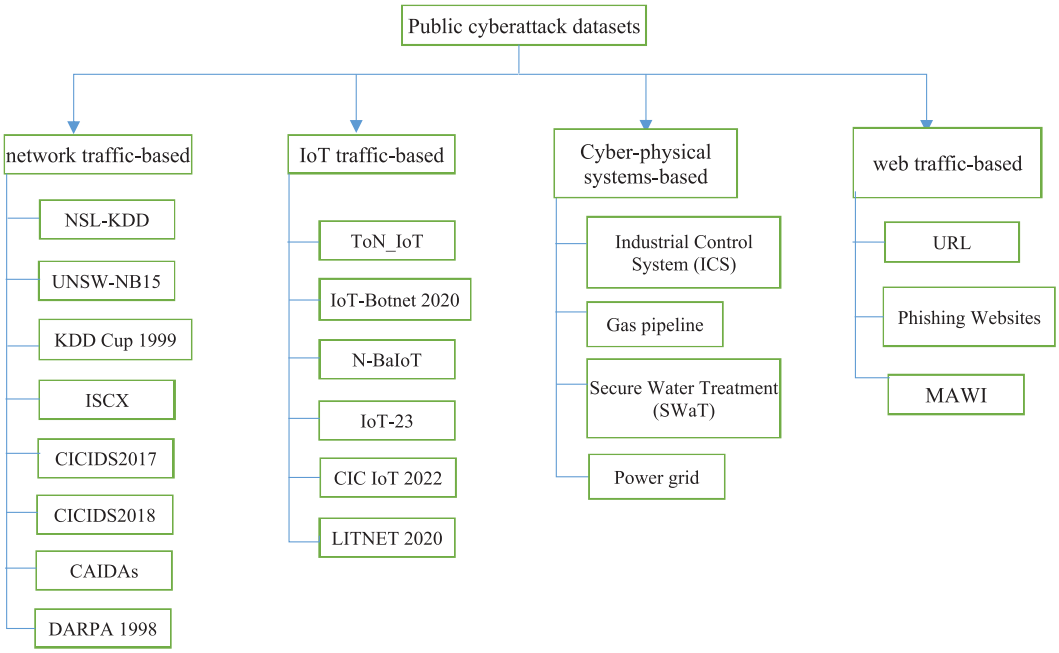

Fig. 1. Classification of cyber security datasets [1, 2].

types. Moreover, the diagram serves as a way to sort the currently available databases as well as examine each of them regarding the presence of certain features, size, number of records, and attack types. Some criteria for choosing the best and the most usable data sets are described beside in general, more exhaustive comparison of their strengths and weaknesses for various goals. Cyberattack datasets are classified into four main categories: network-based datasets, Internet of Things (IoT)-based datasets, cyber-physical system-based datasets, and web traffic-based datasets. Each category has several subcategories that could describe the precise types of cyberattack data [7]. Fig. 1 shows the classification of cyberattack datasets.

A- **Network traffic-based**: This type of dataset is very important in identifying and preventing an attack. These data sets can capture information about traffic flow in the network, such as packet source, destinations, protocols, and data volume. This type of dataset comprises National Science Laboratory (NSL)-Knowledge Discovery and Data Mining (KDD) (NSL-KDD), University of New South Wales Network-Based 15 (UNSW-NB15), KDD Cup 1999 (KDD CUP 1999), Information Security Center of Excellence (ISCX) datasets, the Canadian Institute for Cybersecurity Intrusion Detection (CICIDS) 2017 and 2018 dataset, Defense Advanced Research Project Agency 1998 (DARPA98) and Cooperative Association for Internet Data Analysis (CAIDAs).

B- **IoT Traffic-based**: This dataset contains a wide set of useful data to investigate cyberattacks on IoT devices. The Traffic of Networks in IoT (ToN-IoT) dataset focuses on analyzing malware traffic spread in IoT networks, and the data collection period is for 30 days across different devices. The IoT Botnet (IoT-Botnet 2020) dataset has samples from 61 different IoT malware families. The Network Behavior of IoT (N-BaIoT) dataset captures both benign and anomalous IoT traffic, while the IoT-23 dataset offers different network traffic of controlling the IoT devices. Lastly, the comprehensive Canadian Institute for Cybersecurity IoT 2022 (CIC IoT dataset

2022) covers various devices, protocols, and attacks, aiding research and enhancing cybersecurity measures. The LITNET-2020 dataset refers to annotated real-world network flow datasets for network intrusion detection obtained from a real-world academic network. It contains examples of both normal and under-attack network traffic.

C- **Cyber-Physical Systems (CPS)**: This type of dataset encompasses data from engineering systems that integrate physical and computational components; CPS is prevalent in industries like manufacturing, transportation, energy, and healthcare, connecting physical devices with sensors, actuators, and computing elements. This type of dataset includes data such as the Industrial Control System (ICS) dataset, the Gas Pipeline dataset, Special Weapons and Tactics (SWaT), and the power grid.

D- **Web traffic-based**: This type of dataset captures network traffic dataset, including packets and flows exchanged between devices on a network. These datasets serve various purposes, from studying network behavior to developing and testing techniques for detecting and mitigating cyberattacks, such as Uniform Resource Locator (URL), Phishing Websites, Measurement and Analysis on the WIDE Internet (MAWI) Working Group that supports research and analysis in network security and performance [7].

3. Machine learning techniques

Cybersecurity refers to various methods and tools used to protect data, information, networks, and programs from different types of attacks, such as data tampering, theft, unauthorized access, and destruction over the Internet or network. Cybersecurity components mainly focus on host protection and network security systems. It is applied to many domains, such as cloud computing, wireless sensor networks, and IoT. However, despite the existence of various security measures such as antivirus, firewall, and Intrusion Detection System (IDS). This paper reviews 34 papers selected from 85 papers on machine learning techniques for cyberattack detection using different cyberattack datasets [8], as shown in Table 1.

Previous works have some advantages and disadvantages. For instance, the authors in [9] utilized Bi-directional Long Short-Term Memory (BLSTM) layers for effective feature learning from raw data, classifying network intrusions using a softmax layer. Dropout and batch normalization prevent overfitting, but the complexity of deep learning architecture and computational intensity during training are potential challenges. The authors in [10] offered versatile models such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) for identifying malicious activities in Industrial IoT (IIoT), learning features from raw data using various neural network layers. However, scalability challenges and computational efficiency may arise with large datasets. The authors in [11] employed an ensemble-based semi-supervised learning approach that effectively handles unlabeled data through co-training and self-training algorithms. However, scalability challenges with large datasets and computational efficiency may arise. The authors in [12] introduced an improved BackPropagation (BP) neural network using the Levenberg-Marquardt algorithm, effectively detecting and classifying network intrusions. While it learns features from raw data using BP neural networks, potential overfitting and sensitivity to outliers are concerns. The authors in [13] provided a comprehensive overview of available datasets for cyber security intrusion detection, comparing their characteristics and identifying challenges. However, it did not propose novel detection techniques directly impacting practical systems. The authors in [16] presented an unsupervised intelligent system for detecting IoT botnet attacks without prior

Table 1. Techniques comparison.

Ref.	Year	Datasets	Machine learning techniques	Type of attacks	Accuracy range
[8]	2018	NSL-KDD, CICIDS2017	B-LSTM RNN	DoS, Probe, R2L, U2R	NSL-KDD: 99.11% CICIDS2017: 99.32%
[9]	2018	UNSW-NB15, CICIDS2017	CNN RNN LSTM	Malicious activities in the IIoT	UNSW-NB15: 98.67%, CICIDS2017: 98.89%
[10]	2019	NSL-KDD, CICIDS2017	A semi-supervised learning approach based on an ensemble of co-training and self-training algorithms	DoS, Probe, R2L, U2R	NSL-KDD: 99.32%, CICIDS2017: 99.48%
[11]	2019	KDD CUP 99	Improved BP neural network	DOS, R2L, U2L, and Probing	93.31%
[12]	2019	KDD Cup 99, NSL-KDD	Machine learning datasets for cybersecurity applications	Various types of attacks depending	99.92%
[13]	2019	NSL-KDD, CICIDS2017	Intelligent IDS based on deep learning using CNN	DoS, Probe, R2L, U2R	NSL-KDD: 99.11% CICIDS2017: 99.32%
[14]	2019	Twitter dataset	A study of the challenges of applying machine learning algorithms for cybersecurity threat classification models	Threats in social media communication data related to ISIS propaganda	91.3%
[15]	2020	UNSW-NB15, CICIDS2017	Grey wolf optimization and one class SVM-based unsupervised intelligent system	IoT botnet attacks	UNSW-NB15: 98.76%, CICIDS2017: 98.92%
[16, 17]	2020	NSL-KDD, CICIDS2017	Vector convolutional deep learning approach for anomaly detection framework	Anomalies in IoT traffic	NSL-KDD: 99.23% CICIDS2017: 99.41%
[18]	2020	UNSW-NB15	Fuzzy logic and DNN-based HIDS	DoS, sinkhole, and eavesdropping	98.6%
[19, 20]	2020	NSL-KDD	FNN with Software-Defined Network-IoT (SDN-IoT)	MITM, DDoS, side-channel, and malicious code	83%
[21]	2020	DARBA99	GCN for alert correlation (Alert-GCN)	DoS, R2L, U2R, Probing	92.42%
[22]	2020	Malimg, Microsoft malware	A deep learning framework for malware based on graph convolutional networks: A MalNet	Ramnit, Lollipop, Kelihos_ver3, Vundo, Simda,	Malimg: 98.67% Microsoft malware: 99.41%
[23]	2020	DARPA 2000	An approach for reconstructing attack scenarios using attack graph and alert data mining	DDoS, Port Scan, Web Attack	96.67%
[24]	2020	CTU-13	A simulated cloud environment-based experimental study using a proposed MSPC-based (IDS)	IRC, Spam traffic, Click Fraud, Port	98.6%
[25]	2020	IoT-23	A dataset generation scheme based on network traffic analysis	Brute force, scan, DoS, MITM, etc.	97.9%

(*continued on next page*)

Table 1. *(continued).*

Ref.	Year	Datasets	Machine learning techniques	Type of attacks	Accuracy range
[26]	2020	IoT Network Intrusion Detection (IoTID20)	Anomaly-based intrusion detection approach for IoT networks	(DoS), (MITM), scanning attacks.	99.97%
[27]	2020	LITNET	Ensemble of: DNN LSTM DSAE	Port scan, brute force, DoS	99.7%
[28]	2020	KDD Cup 99	Proposed Hybrid Machine Learning Technique (HMLT)	DoS, probe, R2L, and U2R	99.6%
[29]	2020	LITNET-2020 d	Deep neural network classifier	Port scan, brute force, DoS.	99.7%
[30]	2020	UNSW-NB15	detect cyberattacks from the NetFlow data by using CNN	Analysis, backdoor, DoS, exploitation,	99.8%
[31]	2021	KDD Cup 99, NSL-KDD	Cost-Sensitive Stacked Auto-Encoders (CSSAE)	DoS, R2L, Probe, and U2R	KDD Cup 99: 99.53% NSL-KDD: 98%
[32]	2021	SWaT, WADI	Graph Convolutional Networks (GCN) for alert correlation approach	SWaT, WADI	SWaT: 0.99% WADI: 0.98%
[33]	2021	DARPA SC2	Deep learning for frame error prediction using a DARPA	DARPA	97.83%
[34]	2021	NSL-KDD, CICIDS2017	A deep learning approach CNN for the network data security detection system	DOS, DDOS, malware	93,4%
[35]	2021	Information Security and Object Technology-Cloud Intrusion Detection System (ISOT-CID)	An experimental study using three machine learning algorithms (k-means, k-medoids, and DBSCAN)	SQL injection, brute force attack, DOS, DDOS, malware	99.9%
[36]	2021	UNSW-NB15	proposed Multi-Stage Classification Approach (MSCA)	Analysis, backdoor, DoS, exploits, fuzzers	99.7%
[37]	2021	ISCX-IDS2012	Deep Convolutional Neural Networks (DCNN), Deep Q-Networks (DQN)	Brute force, DoS, infiltration	98.9%
[38]	2022	NSL-KDD	Relaxation-based anomaly detection using the ensemble Kalman filter	DDoS	90,32%
[39]	2022	UNSW-NB15	Use a combination of AdaBoosting and Bagging methods with four different classifiers: Naïve Bayesian (NB) SVM Random Forest (RF) K-Nearest Neighbor (KNN).	Analysis, backdoor, DoS, exploitation, fuzzing, generic, reconnaissance, shellcode, and worm	85.49%
[40]	2022	UNBS-NB 15 KDD 99 CICIDS2018	An IDS based on Gradient Boost Decision Trees with Optimization (OGBDT-IDS)	Backdoor, DoS, exploitation, fuzzing, generic, recon.	99.8%
[41, 42]	2023	NSL-KDD, IoT-Botnet 2020	Use Feed Forward Neural Network (FNN) and LSTM	DoS, probe, R2L, U2R), (analysis, backdoor, exploits)	99.95%

knowledge. The one-class Support Vector Machine (SVM) effectively learns a boundary around normal data, but potential false positives due to outlier-based detection and interpretability challenges exist. The authors in [17] exceled at identifying IoT traffic anomalies using vector convolutional deep learning, but computational resources and data privacy concerns in fog environments are considerations. The authors in [19] proposed two innovative intrusion detection approaches. First, the Hybrid Intrusion Detection System (HIDS) fuzzy logic and Deep Neural Network (DNN), enhance adaptability and handle uncertainty efficiently. However, its computational demands and complexity in tuning both components pose challenges. Second, the Fuzzy Neural Network (FNN) model efficiently detects network intrusions using software-defined networks, leveraging fuzzy rules for feature learning. Yet, interpretability remains a concern due to fuzziness. Additionally, the authors in [22] introduced a Graph Convolutional Networks (GCN) model that discovers attack scenarios from intrusion alerts using graph-structured data. While it correlates alerts effectively, its applicability is limited to specific attack representations. The authors in [41] developed a relaxation-based anomaly detection method for cyber-physical systems, utilizing ensemble Kalman filters. Although accurate, it may not suit real-time applications due to computational intensity. Threshold selection based on standard deviation impacts its precision. The authors in [35] introduced a deep learning approach combining CNN and LSTM networks to handle complex network data security threats. It effectively learns features from raw data using convolutional and recurrent layers, while chi-square-based feature selection enhances model accuracy. However, substantial computational resources are required for training deep neural networks, and a large dataset is necessary for effective learning. [34] specialized in predicting frame errors in wireless communication networks with high accuracy, leveraging Gated Recurrent Unit (GRU) networks for feature learning. Correlation-based feature selection optimizes error prediction. However, its applicability may be limited to frame error scenarios. The authors in [42] utilized an optimized gradient boost decision tree model, enhanced by the African buffalo optimization method, for efficient cyber security intrusion detection. It effectively learns features from raw data, improving classification accuracy. However, the complexity arising from integrating advanced algorithms and potential computational intensity may be limitations. The authors in [23] employed graph convolutional networks for malware detection from graph-structured data. It effectively learns features using graph convolutional layers, while dropout and batch normalization prevent overfitting. The authors in [34] introduced an attack scenario reconstruction approach that discovers attack paths from intrusion alerts using attack graphs and alert data mining techniques. While it effectively ranks attack scenarios based on likelihood and severity, its specificity to the used attack graph and alert data may limit generalization to all attack types. The authors in [35] presented a deep learning approach for intelligent intrusion detection using CNN and LSTM networks. It captures temporal dependencies and classifies different network intrusions using a softmax layer. However, computational resources are necessary for deep learning, and model performance may vary based on data quality and network characteristics. Finally, [24] uses social media communication data to evaluate machine learning algorithms for cybersecurity. It addresses data quality, labeling, imbalance, and privacy challenges.

4. Cyberattack datasets reviewed

Researchers widely use these cyberattack datasets to evaluate and compare different attack detection techniques. However, some are more popular and frequently used than

Table 2. Cyberattack datasets used in the reviewed papers.

Dataset	Reference	Number of uses
CICIDS2017	[8–10, 12, 13, 15, 16, 34]	8
NSL-KDD	[8, 10, 12, 15, 31, 34, 44]	7
UNSW-NB15	[9, 12, 15, 18, 37, 41, 42]	5
KDD CUP 99	[11, 12, 28, 31, 42]	5
DARPA	[21, 23, 33]	3
IoT-23	[25, 27]	2
LITNET-2020	[27, 29]	2
ISCX-IDS-2012	[12, 38]	2
BoT-IoT	[43]	1
SWaT and WADI	[32]	1
CICIDS2018	[41]	1
URL-2016	[36]	1

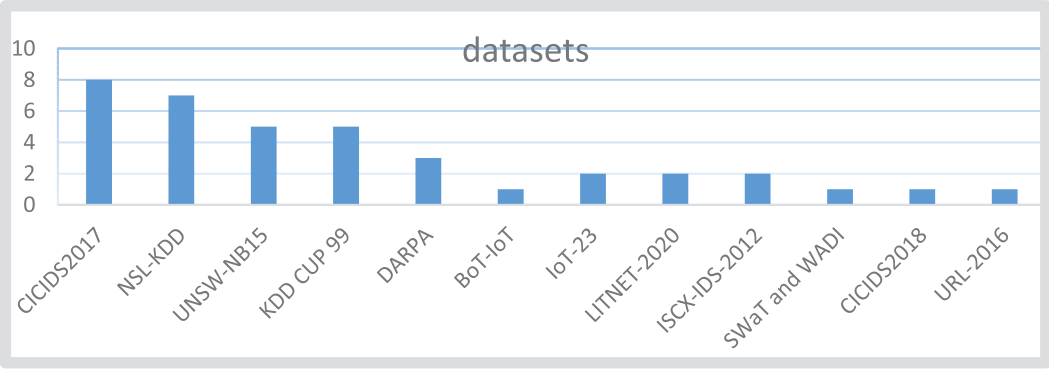

Fig. 2. Cyber security dataset redundancy.

others, which may indicate their relevance, quality, or availability, as shown in Table 2 and Fig. 2.

5. Conclusion

This paper presents an analysis of the redundancy or usefulness of different cyberattack datasets and machine learning models for cyberattack detection. It has shown that some datasets and models are more popular and frequently used than others, and possible reasons for this have been discussed. This analysis can help researchers to understand the current trends and challenges in cyberattack detection datasets and to choose the appropriate datasets and models for their experiments. Nevertheless, the analysis also has some limitations and directions for future work. These include employing alternative metrics to assess the quality or impact of the datasets and models, conducting a more comprehensive survey and evaluation of both existing and new datasets and models, and performing more in-depth analysis and comparison of the advantages and disadvantages of different datasets and models. It is anticipated that this work will inspire and motivate further research in this important, evolving field.

References

1. J. Sharma, C. Giri, O. C. Granmo, and M. Goodwin, "Multi-layer intrusion detection system with ExtraTrees feature selection, extreme learning machine ensemble, and softmax aggregation," *EURASIP Journal on Information Security*, vol. 2019, no. 1, Art. no. 15, 2019, doi: 10.1186/s13635-019-0098-y.
2. K. Barik, "Cybersecurity deep: approaches, attacks dataset, and comparative study," *Applied Artificial Intelligence*, vol. 36, no. 1, Jan. 2022, doi: 10.1080/08839514.2021.1977229.
3. A. Prasad and S. Chandra, "Machine learning to combat cyberattack: a survey of datasets and challenges," *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, vol. 20, no. 4, pp. 577–588, doi: 10.1177/15485129211094881
4. Y. H. Ali *et al.*, "Optimization system based on convolutional neural network and Internet of Medical Things for early diagnosis of lung cancer," *Bioengineering*, vol. 10, no. 3, Art. no. 320, 2023, doi: 10.3390/bioengineering10030029.
5. W. Lee, S. J. Stolfo, and K. W. Mok, "A data mining framework for building intrusion detection models," *Proceedings of the 1999 IEEE Symp. on Security and Privacy (Cat. No.99CB36344)*, Oakland, CA, USA, pp. 120–132, 1999, doi: 10.1109/SECPRI.1999.766909.
6. A. Al-Abassi, H. Karimipour, H. HaddadPajouh, A. Dehghantanha, and R. M. Parizi, "Industrial big data analytics: challenges and opportunities," In: K. K. Choo, A. Dehghantanha (eds) Handbook of Big Data Privacy. Springer, Cham. https://doi.org/10.1007/978-3-030-38557-6_3.
7. S. M. Naser and Y. H. Ali, "Study of cyber security effects on wireless sensor networks," *Iraqi Journal of Computers, Communication, Control & Systems Engineering*, vol. 21, no. 4, pp. 74–81, 2021.
8. B. Roy and H. Cheung, "A deep learning approach for intrusion detection in internet of things using bi-directional long short-term memory recurrent neural network," in *Proc. of the 28th Int. Telecommunication Networks and Applications Conf.*, Sydney, NSW, Australia, pp. 1–6, 2018, doi: 10.1109/ATNAC.2018.8615294.
9. M. AL-Hawawreh, N. Moustafa, and E. Sitnikova, "Identification of malicious activities in industrial internet of things based on deep learning models," *Journal of Information Security and Applications*, vol. 41, pp. 1–11, Jun. 2018, doi: 10.1016/j.jisa.2018.05.002.
10. S. R. Khonde and V. Ulagamuthalvi, "Ensemble-based semi-supervised learning approach for a distributed intrusion detection system," *Journal of Cyber Security and Technology*, vol. 3, no. 3, pp. 163–188, Jul.–Dec. 2019, doi: 10.1080/23742917.2019.1623475.
11. A. Yang, Y. Zhuansun, C. Liu, J. Li, and C. Zhang, "Design of intrusion detection system for internet of things based on improved BP neural network," *IEEE Access*, vol. 7, pp. 106043–106052, 2019, doi: 10.1109/ACCESS.2019.2929919.
12. C. Nilă, V. V. Patriciu, and I. Bica, "Machine learning datasets for cyber security applications," *Security & Future*, vol. 3, no. 3, pp. 109–112, 2019.
13. R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," *IEEE Access*, vol. 7, pp. 41525–41550, 2019, doi: 10.1109/ACCESS.2019.2895334.
14. A. Q. Lima and B. Keegan, "Challenges of using machine learning algorithms for cybersecurity: a study of threat classification models applied to social media communication data," *Cyber Influence and Cognitive Threats*, pp. 33–52, 2020, doi: 10.1016/B978-0-12-819204-7.00003-8.
15. A. Al Shorman, H. Faris, and I. Aljarah, "Unsupervised intelligent system based on one class support vector machine and grey wolf optimization for IoT botnet detection," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 7, pp. 2809–2825, Jul. 2020, doi: 10.1007/s12652-019-01387-y.
16. N. G. Bhuvaneswari Amma and S. Selvakumar, "Anomaly detection framework for Internet of things traffic using vector convolutional deep learning approach in fog environment," *Future Generation Computer System*, vol. 113, pp. 255–265, Sep. 2020, 10.1016/j.future.2020.07.020.
17. E. M. Alsaedi and A. K. Farhan, "Retrieving encrypted images using convolution neural network and fully homomorphic encryption," *Baghdad Science Journal*, vol. 20, no. 1, pp. 206–220, Mar. 2023, doi: 10.21123/bsj.2022.6550.
18. D. S. Smys, "Hybrid intrusion detection system for internet of things (IoT)," *Journal of IoT in Social, Mobile, Analytics, and Cloud*, vol. 2, no. 4, pp. 190–199, Dec. 2020, doi: 10.36548/jismac.2020.4.002.
19. Y. H. Ali *et al.*, "Multi-layered non-local bayes model for lung cancer early diagnosis prediction with the internet of medical things," *Bioengineering*, vol. 10, no. 2, Art. no. 138, 2013, doi: 10.3390/bioengineering10020138.
20. F. Farhin, I. Sultana, N. Islam, M. S. Kaiser, M. S. Rahman, and M. Mahmud, "Attack detection in internet of things using software defined network and fuzzy neural network," in *Proc. 2020 Joint 9th Int. Conf. on Informatics, Electronics & Vision (ICIEV) and 2020 4th Int. Conf. on Imaging, Vision & Pattern Recognition (icIVPR)*, Kitakyushu, Japan, pp. 1–6, 2020, doi: 10.1109/ICIEVicIVPR48672.2020.9306666.

21. H. Karimipour and H. Leung, "Relaxation-based anomaly detection in cyber-physical systems using ensemble Kalman filter," in *IET Cyber-Physical System: Theory and Application*, vol. 5, no. 1, pp. 49–58, Mar. 2020, doi: 10.1049/iet-cps.2019.0031.
22. X. Pei, L. Yu, and S. Tian, "AMalNet: a deep learning framework based on graph convolutional networks for malware detection," *Computers and Security*, vol. 93, Art. no. 101792, 2020, doi: 10.1016/j.cose.2020.101792.
23. H. Hu, J. Zhang, Y. Liu, X. Xu, and J. Tan, "Attack scenario reconstruction approach using attack graph and alert data mining," *Journal of Information Security and Applications*, vol. 54, Art. no. 102522, 2020, doi: 10.1016/j.jisa.2020.102522.
24. A. Aldribi, I. Traoré, B. Moa, and O. Nwamuo, "Hypervisor-based cloud intrusion detection through online multivariate statistical change tracking," *Computers & Security*, vol. 88, Art. no. 101646, 2020, doi: 10.1016/j.cose.2019.101646.
25. I. Ullah and Q. H. Mahmoud, "A scheme for generating a dataset for anomalous activity detection in IoT networks," in *Proc. 33rd Canadian Conf. on Artificial Intelligence (Candian AI 2020)*, Ottawa, ON, Canada, pp. 508–520, May 13–15, 2020.
26. P. Maniriho, E. Niyigaba, Z. Bizimana, V. Twiringiyimana, L. J. Mahoro, and T. Ahmad, "Anomaly-based intrusion detection approach for IoT networks using machine learning," in *Proc. 2020 Int. Conf. on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*, Surabaya, Indonesia, pp. 303–308, 2020, doi: 10.1109/CENIM51130.2020.9297958.
27. V. Dutta, M. Choraś, M. Pawlicki, and R. Kozik, "A deep learning ensemble for network anomaly and cyber-attack detection," *Sensors*, vol. 20, no. 16, Art. no. 4583, 2020, doi: 10.3390/s20164583.
28. H. M. Tahir *et al.*, "Hybrid machine learning technique for intrusion detection system," in *Proc. of the 5th Int. Conf. on Computing and Informatics (ICOCI 2015)*, Istanbul, Turkey, Aug. 11–13, pp. 464–472, 2015.
29. R. Damasevicius *et al.*, "LITNET-2020: an annotated real-world network flow dataset for network intrusion detection," *Electronics*, vol. 9, no. 5, Art. no. 800, 2020, doi: 10.3390/electronics9050800.
30. C.-T. Yang, J.-C. Liu, E. Kristiani, M.-L. Liu, I. You, and G. Pau, "NetFlow monitoring and cyberattack detection using deep learning with ceph," *IEEE Access*, vol. 8, pp. 7842–7850, 2020, doi: 10.1109/ACCESS.2019.2963716.
31. A. Telikani and A. H. Gandomi, "Cost-sensitive stacked auto-encoders for intrusion detection in the internet of things," *Internet of Things*, vol. 14, Art. no. 100122, 2021, doi: 10.1016/j.iot.2019.100122.
32. Q. Cheng, C. Wu, and S. Zhou, "Discovering attack scenarios via intrusion alert correlation using graph convolutional networks," *IEEE Communications Letters*, vol. 25, no. 5, pp. 1564–1567, May 2021, doi: 10.1109/LCOMM.2020.3048995.
33. A. S. M. M. Jameel, A. P. Mohamed, X. Zhang, and A. E. Gamal, "Deep learning for frame error prediction using a DARPA spectrum collaboration challenge (SC2) dataset," *IEEE Networking Letters*, vol. 3, no. 3, pp. 133–137, Sept. 2021, doi: 10.1109/LNET.2021.3096813.
34. N. R. Sai, G. S. C. Kumar, M. A. Safali, and B. S. Chandana, "Detection system for the network data security with a profound deep learning approach," in *Proc. 2021 6th Int. Conf. on Communication and Electronics Systems (ICCES)*, Coimbatre, India, pp. 1026–1031, 2021, doi: 10.1109/ICCES51350.2021.9488967.
35. A. Alshammari and A. Aldribi, "Apply machine learning techniques to detect malicious network traffic in cloud computing," *Journal of Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00475-1.
36. R. Qaddoura, A. M. Al-Zoubi, H. Faris, and I. Almomani, "A multi-layer classification approach for intrusion detection in IoT networks based on deep learning," *Sensors*, vol. 21, no. 9, Art. no. 2987, 2021, doi: 10.3390/s21092987.
37. J. Yang, G. Liang, B. Li, G. Wen, and T. Gao, "A deep-learning-and reinforcement-learning-based system for encrypted network malicious traffic detection," *Electron. Letters*, vol. 57, no. 9, pp. 363–365, 2021, doi: 10.1049/ell2.12125.
38. J. Yuan, G. Chen, S. Tian, and X. Pei, "Malicious URL detection based on a parallel neural joint model," in *IEEE Access*, vol. 9, pp. 9464–9472, 2021, doi: 10.1109/ACCESS.2021.3049625.
39. D. N. Mhawi and S. H. Hashim, "Proposed hybrid ensemble learning algorithms for an efficient intrusion detection system," *Iraqi Journal of Computers, Communications, Control and Systems Engineering*, vol. 22, no. 2, pp. 73–84, 2022, doi: doi.org/10.33103/uot.ijccce.22.2.7.
40. S. Mishra, "An optimized gradient boost decision tree using enhanced african buffalo optimization method for cyber security intrusion detection," *Applied Sciences*, vol. 12, no. 24, Art. no. 12591, 2022, doi: 10.3390/app122412591.
41. E. M. Alsaedi and A. K. Farhan, "RCAE_BFV: retrieve encrypted images using convolution autoencoder and BFV," *Iraqi Journal of Computers, Communications, Control and Systems Engineering*, vol. 22, no. 3, pp. 48–61, 2022, doi: 10.33103/uot.ijccce.22.3.5.

42. I. N. Mahmood and H. S. Abdullah, “Lung cancer prediction and risk factors identification using artificial neural network,” *Iraqi Journal of Computers, Communications, Control and Systems Engineering*, vol. 22, no. 1, pp. 55–62, 2022, doi: 10.33103/uot.ijccce.22.1.6.
43. A. F. Al-zubidi, A. K. Farhan, and S. M. Towfek, “Predicting DoS and DDoS attacks in network security scenarios using a hybrid deep learning model,” *Journal of Intelligent Systems*, vol. 33, no. 1, 2014, doi: 10.1515/jisys-2023-0195.
44. E. M. Alsaedi, A. K. Farhan, M. W. Falah, and B. K. Oleiwi, “Classification of encrypted data using deep learning and legendre polynomials,” in *Proc. of the Int. Conf. on Innovations in Computing Research (ICR’22)*, pp. 331–345, 2022.