Grenze International Journal of Engineering and Technology, January Issue

Cyber Threat Analysis using Machine Learning for Proactive Mitigation: A Comprehensive Review

Akshay R. Jain[1] and Atul Agrawal[2]
[1-2]Department of Computer Science and Engineering, Oriental University, Indore (M.P.), India
Email: jainakshay781@gmail.com, atul.273@gmail.com

***Abstract*— With the rapid increase of digital connectivity we have seen some extraordinary opportunities as well as an ominous escalation of cyber threats. Conventional rule-based approaches typically fail to identify the dynamic sophisticated cyber-attacks. As a result, machine learning (ML) has become a powerful tool in forward security for pro-active defense against cyber threats to detect anomaly, anticipation of an attack, and tampering by evolving threats. This review study provides an overview of the field of cyber-threat analysis using machine learning, and an in-depth analysis of a variety of supervised, unsupervised, and deep learning models employed in cybersecurity. It classifies some important cyber threats, analyzes popular datasets like CICIDS2017 and UNSW-NB15, and evaluates the ML models effectiveness using accuracy, F1-score, and detection rate performance measures. Additionally, the paper discusses emerging challenges, including but not limited to, imbalanced data, model interpretability, and adversarial evasion, as well as potential future directions, such as federated learning and explainable AI in this field. The paper tries to connect the divide between the cybersecurity practitioners and the ML researchers by giving an organized survey of the state of the art and suggesting the directions in which the path to developing intelligent and resilient security systems.**

***Index Terms*— Cyber threat detection, machine learning in cybersecurity, proactive threat mitigation, intrusion detection systems, deep learning models, adversarial attacks and defense.**

I. INTRODUCTION

The past decades have seen the digital age revolutionize the modern daily lifestyle, allowing fluid interactivity, direct communication, and data format services. Unfortunately, this evolution has also brought an increase in cyber threats, from common viruses and network attacks to more advanced threats like Advanced Persistent Threats (APTs), zero-day and ransomware. The frequency, complexity, and impact of these cyberattacks have grown rapidly, putting individuals, organizations, and nations at significant risk. According to recent studies, global cybercrime damages are projected to reach $10.5 trillion annually by 2025 [1].

Traditional cybersecurity approaches, signature-base detection and static rule engines, become more and more inadequate to handle these new attack vectors. These techniques commonly have high false positive rates and are unable to recognise zero-day exploitation and polymorphic attacks. In contrast, machine learning (ML) demonstrates great potential to characterize intricate patterns, respond to new threats and automatically detect threats in nearly real time [2].

Grenze ID: 01.GIJET.12.1.520

Methods such as supervised/unsupervised anomaly, along with complex models (i.e., deep networks) have proven to be useful in different cybersecurity tasks (e.g., intrusion detection, malware classification, botnet identification or phishing URL detection). Such smart models can be trained on historical data and generalize across unseen situations to take preventive measures instead of reactive approach. Unlike previous reviews, this paper uniquely correlates threat categories with machine learning paradigms, offering a structured strategy for proactive cyber threat detection.

This review paper aims to:

- Provide an overview of different cyber threat categories and their characteristics.
- Examine the application of ML methods in detecting and mitigating such threats.
- Analyze widely used datasets and benchmarking approaches.
- Identify prevailing challenges in implementing ML-based cybersecurity systems.
- Propose future research directions such as federated learning, explainable AI, and real-time adaptive models.

By offering a comprehensive analysis of current trends and techniques, this study bridges the gap between theory and practice in machine learning-based cyber threat mitigation, contributing toward the development of intelligent and secure digital infrastructures.

II. Cyber Threat Landscape

The changing cyber threat environment reflects the increased complexity, size, and frequency of attacks on digital systems. The days when cyber agents resorted to unsophisticated hacking are becoming less and less frequent, as attackers now take advantage of the latest exploits (such as zero-day exploits), the creation of new malware variants, and click-fraud attacks (often initiated by phishing) that easily charge for fake advert clicks (social engineering). Cybercrimes have expanded beyond computer systems to vital areas like health, finance, energy, industrial control systems, which makes cybersecurity an open and the most important issue on a global scale.

Major categories of cyber threats include:

- Malware Attacks: Malicious software like ransomware, trojans, and worms that disrupt operations, steal data, or demand ransom payments.
- Phishing and Social Engineering: Techniques that manipulate users into divulging sensitive information, often leading to identity theft or system compromise.
- Denial-of-Service (DoS) and Distributed DoS (DDoS) Attacks: Overwhelm a system or network to render it unusable, often used to sabotage services or distract from other attacks.
- Insider Threats: Breaches initiated by individuals within an organization, either maliciously or unintentionally, which are difficult to detect and mitigate.
- Advanced Persistent Threats (APTs): Coordinated, long-term cyberattacks often attributed to nation-state actors, aimed at espionage or critical disruption.

The use of IoT devices, cloud platforms, and telecommuting job sites has increased the attack surface, making it easy for systems to get compromised. Over 90% of data breach are reported to be caused by human mistakes or vulnerabilities due to phishing and credential which have been compromised [3]. In addition, adversaries have started to use AI to generate realistic phishing emails and to distribute malicious software more efficiently, making it harder for existing defense mechanisms to keep up [4].

To tackle these upcoming threats, it is crucial to change from being reactive towards proactive threat analysis. A machine learning approach is well suited to this task by allowing systems to learn from the past attack patterns and detect the anomalies, which may indicate the initial stages of an attack [5].

III. Machine Learning Techniques in Cybersecurity

ML is becoming one of the most popular applications in modern cybersecurity systems, because it is capable of learning automatically from data, finding hidden patterns and generalizing to detect unknown threats. Unlike rule-based systems that depend on some predefined signatures, ML models can help in pro- active detection by identifying anomalies or learning the patterns of threat behaviour. ML based techniques in cyber domain include intrusion detection, malware classification, spam filtering, botnet detection and so on [2].

A. Supervised Learning

Supervised methods are commonly employed for the cyber threat detection, particularly when annotated databases exist. These classifiers are trained to divide inputs into several known categories such as benign traffic or malicious traffic. Popular algorithms such as SVM, DT, RF, k-NN [1], etc. For example, CICIDS2017 and

NSL-KDD were datasets in which Random Forest has known to boast high accuracy in network flow classification [3].

B. Unsupervised Learning

Unsupervised learning methods do not need labeled data and could be applied for anomaly detection under the problem of identifying novel threats. Commonly used algorithms are k-Means clustering, DBSCAN and Principal Component Analysis (PCA) [5]. Auto-encoders, as a type of deep learning unsupervised model, have become popular to detect anomalies of network traffic [7]. They learn to reconstruct input features and alert to deviations as potential intrusions. Using unsupervised approach however, false positive rates might be higher and interpretation from experts is needed [8].

C. Deep Learning

Deep learning can achieve better performance in both accuracy and flexibility due to its ability to learn hierarchical feature representations. Convolutional Neural Networks (CNNs) are employed to perform static malware image classification; Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) are used for sequential log analysis and threat prediction [9]. Deep Autoencoders and hybrid CNN-LSTM architectures have been used successfully to detect APTs and varieties of Malware with high accuracy [10].

D. Hybrid and Ensemble Models

In order to improve the accuracy and stability, hybrid models and ensemble approaches can leverage the advantages of multiple learning methods. CNNs are able to learn spatial features from data, whereas LSTMs can learn temporal dependencies and they work together to provide an efficient intrusion detection system [12]. Ensemble methods such as majority voting, bagging [9] (e.g., Random Forest), and boosting [10] (e.g., XGBoost) are effective at reducing overfitting leading to better detection rates [13]. These techniques are especially attractive for noisy or mixed data sources.

E. Reinforcement Learning (RL)

RL is a kind of dynamic process in which agents learn the optimal security policies by interacting with their environment. RL-based models have been created to model intrusion response plans and firewalls rule adaptability [14].

IV. Dataset Overview

The performance of ML based cyber threat detection systems relies heavily on the quality and representativeness of the datasets on which such systems are trained and evaluated. Public cybersecurity datasets provide a way to benchmark algorithms under realistic attack and background conditions. These databases differ in scale, feature richness, attack heterogeneity, and suitability for contemporary threats. There are a number of standard datasets used for the evaluation of Intrusion Detection Systems and other threat analysis models:

NSL-KDD

As a supplement to the KDD Cup 1999 dataset, the NSL-KDD dataset accounts for redundancy and imbalance. It includes labeled records indicating normal traffic, as well as different attack patterns (DoS, R2L, U2R and Probe). But it misses the modern real-world threats and traffic diversity [15].

CICIDS2017

CICIDS2017 is an extensive dataset that consists of benign and modern attacks traffic a such as Brute Force, Botnet, DDoS, and Web, compiled by Canadian Institute for Cybersecurity. It has features based on flow characteristics and imitates real network traffic [3].

UNSW-NB15

Modelling itself on current generation network traffic, UNSW-NB15 combines normal and artificial attack characteristics. The dataset contains features that are based on the Argus and Bro tools to catch advanced attacks including Analysis, Fuzzers and Exploits [16].

TON_IoT

TON_IoT is a purpose-designed dataset for the Internet of Things (IoT) ecosystem that includes telemetry, network-pack, and system logs data, and offers rich context to detect cyber threats in IoT environments [17, 18].
Bot-IoT

One more IoT dataset, called Bot-IoT, contains labeled traffic for DDoS, DoS, and data exfiltration attacks in a simulated IoT environment by collecting packet-level and flow-level data [18].

TABLE I: COMPARATIVE ANALYSIS OF PUBLICLY AVAILABLE DATASET

Datasets	Year	Institution	Data Type	Attack Types Covered	Features	Remarks
NSL-KDD	2009	University of New Brunswick	Labeled connection records	DoS, Probe, R2L, U2R	41	Outdated attack scenarios; good for baseline
CICIDS2017	2017	Canadian Institute for Cybersecurity	Network flow logs	DDoS, Botnet, Brute Force, Web attacks	80+	Realistic traffic; up-to-date attack types
UNSW-NB15	2015	Australian Centre for Cyber Security	Packet-based and flow data	Exploits, Fuzzers, Reconnaissance	49	Rich features; modern attacks simulated
TON_IoT	2020	University of New South Wales	IoT telemetry and logs	Ransomware, DoS, Injection, MITM	Multiple	Multimodal; suited for IoT/edge environments
Bot-IoT	2018	University of New South Wales	Packet and flow data	DDoS, DoS, Reconnaissance, Data theft	Multiple	Imbalanced data; good for IoT-based threat study

V. LITERATURE REVIEW

Previous surveys have documented the progress of ML and DL models in cybersecurity. However, these often lack a structured linkage between threat types, datasets, and detection methods. This paper aims to fill that gap by providing a functional classification of ML techniques based on cyber threat profiles and dataset properties, which is not sufficiently addressed in existing works.

[4] The paper presents a decade-long review of AI application in cybersecurity with emphasis on ML, DL, and hybrid approaches. It evaluates their efficacy in detecting as well as in automating threats. VIII Next Directions Some of the key future directions are: real-time detection, edge integration and explainable AI. This work further lays the groundwork for academic and industry exploitation.

[5] The work reviews the literature on network-based intrusion detection data sets and proposes fifteen evaluation criteria such as label noise/trustworthiness, type of traffic, and realism. It criticizes old datasets such as NSL-KDD and recommends CICIDS2017 or UNSW-NB15. The work is an important source for datasets used in the research of IDS.

[6] This survey provides ML methods for detecting anomalies in network traffic, including supervised and unsupervised methods. It reviews essential performance metrics and it identifies also the imbalance of the data and the high dimensionality as the main issues. Real time detection and adaptive learning are emphasized.

[7] The authors design a CNN-LSTM deep learning model for DDoS detection in SDN networks. The abnormal traffic sequences are detected by the model accurately. The performance of the estimator is demonstrated with experiments in dynamic conditions. It articulates the promise of deep learning in network security.

[8] A deep learning-based IDS is presented using stacked autoencoders trained on NSL-KDD. The model improves accuracy over classical ML methods, especially on minority classes. It demonstrates unsupervised feature learning's utility in security applications.

[9] This review analyses DL architectures (CNN, RNN, GAN) applied to cybersecurity tasks. Use cases include malware analysis, phishing detection, and threat intelligence. It emphasizes model explainability and adversarial resilience as future priorities.

[10] The paper presents a hybrid DL approach combining autoencoders for feature extraction with SoftMax classifiers. Tested on NSL-KDD, it achieves low false positives and high accuracy. The model demonstrates the value of unsupervised pretraining for IDS.

[11] A DL-based anomaly detection model using LSTM and GRU is proposed for ICS security. Evaluated on real-world ICS datasets, it outperforms traditional classifiers. The study addresses critical infrastructure vulnerabilities.

[12] The Res-TranBiLSTM model integrates ResNet, Transformer, and BiLSTM to detect IoT-based intrusions. It captures both spatial and temporal features, outperforming baseline models. Results validate its strength in handling complex IoT data.

[14] A dynamic ML-based framework is introduced for mitigating DDoS in cloud systems. It adapts to attack behaviour in real time, improving detection rates and service uptime. The system emphasizes responsiveness and scalability.

[15] In this paper, we criticize the redundancy and attack homogeneity of the KDD 99 dataset. This is what led to the creation of NSL-KDD. This paper will form the groundwork of dataset design for IDS research.

[17] TON_IoT processes telemetry, system logs, and network data to realistic anomaly detection. It offers multi-model ML training for internet of things security. The farewell dataset can be used to benchmark edge-to-cloud threat analysis models.
[18] This work utilizes ML to categorize IoT devices using network traffic patterns. It enhances network visibility as well as security policy implementation. The technique allows passive profiling in intelligent spaces.
[19] Various ML classifiers are evaluated for the detection of cyber-attacks in industrial systems. The detection accuracy of Random Forest and SVM is high. A validation has been done by real world SCADA data.
[20] This survey discusses ML for malware detection, including static and dynamic techniques. It talks about feature engineering, classifier war and adversarial perils. These omissions include polymorphic malware recognition and normalization of the data.
[21] Deep IDS is suggested with CNN, LSTM and mixed models with CICIDS2017. The model is of high accuracy and can support real-time inference. It demonstrates the scalability of DL for cybersecurity.
[23] The authors present a survey of methods for interpreting predictions of black-box ML models. Methods (SHAP, LIME, counterfactuals) are compared. We pay particular attention to their use in security-critical domains.
[25] Federated Learning is proposed for decentralized model training without the need for data transmission. Applications range from mobile apps to health care to cyber security. Communication efficiency and trust establishment are challenging.

VI. Research Challenges

Although ML shows great promise in the identification of cyber threats and the mitigation of cyber threats in a preventive manner, there are challenges that prevent it from being used in real-world settings and for general purposes. These issues are due to the evolving nature of cyberattacks, data issues and model interpretability. The main research challenges are as follows:

A. Imbalanced and Limited Datasets

Literally all publicly available datasets (e.g., NSL-KDD, CICIDS2017) have class imbalance issue to some extent, with normal traffic instances far more than that of the malicious traffic. Such imbalance tends to bias ML models towards prediction of the majority class, thus making them less effective in identifying rare but important types of attacks like zero-day attacks or insider threat [19].

B. Evolving Threat Landscape

Cyber-attacks are constantly evolving with new forms and attack tactics, and the static model is also outdated. Oftentimes, ML models that are trained on historical data do not generalize to new or adversarial threats, re-enforcing the need of adaptive learning [20].

C. Feature Engineering and Data Quality

The high-quality and relevant feature is the key to the accurate threat. Nevertheless, real-time feature extraction from encrypted traffic, free text logs, or diverse IoT data is very challenging. Additional differences in data collection also impede cross-system generalization [21].

D. Lack of Real-Time Detection Capabilities

Plenty of ML scenarios need batch-based training and offline inference which prevent from being applied to real-time or near real-time analysis and response. It is still an active research challenge on how to achieve high detection accuracy without increasing the processing latency in intrusion detection systems [20].

E. Adversarial ML and Evasion Attacks

A successful attack on ML systems is conducted through adversaries generating adversarial inputs such that the output of the ML model would result in misclassification. This raises the important issue of whether or not ML-based cybersecurity solutions are robust under attack [21].

F. Model Interpretability and Explainability

Deep learning models like CNNs or LSTMs operate as "black boxes" and it is hard for cybersecurity experts to interpret and trust their decisions. The lack of interpretability has also hampered the widespread use of ML models in institutions that regard security critical sectors, such as finance or national infrastructure [22].

G. Data Privacy and Ethical Concerns

Developing ML models upon sensitive network logs or user behaviour data raises issues around user privacy and regulation (e.g. GDPR). Privacy-preserving strategies, including federated learning, are beginning to be deployed in practice [23].

VII. Future Research Directions

As cyber threats become increasingly sophisticated, the integration of machine learning into cybersecurity must evolve to meet emerging challenges. The following research directions offer promising avenues for enhancing proactive threat mitigation:

- Federated Learning for Privacy-Preserving Security: Centralized training violates sensitive data. With federated learning, it is possible to train models collaboratively among the different edge devices without leaking raw data, thus preserving privacy while improving the generalization of the model in heterogeneous environments [24].
- Explainable AI (XAI) for Trustworthy Cyber Defense: If ML models are to be more widely adopted in these important settings, then interpretability is key. XAI frameworks like SHAP or LIME can be incorporated to facilitate better understanding and trust to model decisions by analysts [25].
- Adaptive and Continual Learning Systems: If we design models that can learn over time as new threats emerge and that can adapt in real time, we might be able to address these evolving cyber threats. Adaptive Systems based on online learning and reinforcement learning are promising candidates.
- Multi-modal Threat Detection: The fusion of various data modalities network traffic, logs, and user behaviour—may enhance the accuracy of detection. Hybrid-based models that incorporate both structured and unstructured data will be more robust to advanced attack vectors.
- Lightweight ML Models for Edge Devices: With increasing IoT ecosystems, there is a strong need to design efficient ML algorithms that are deployable in low-resource environments (edge devices). Methods including model pruning, quantization edge cloud collaborative models are also being investigated.
- Robustness Against Adversarial Attacks: Research in adversarial ML should further investigate how to develop models resistant to poisoning and evasion attacks through robust training methods and adversarial detection mechanisms.

VIII. Conclusion

In this work, we provide a comprehensive survey focusing on the impact of ML in cyber threat detection and mitigation. ML models, including classical supervised and unsupervised up to complex deep learning structure and hybrid ensemble, show promising results in detecting complex and dynamic attack patterns with high accuracy and efficiency.

These models are being evaluated on publicly available benchmark datasets such as CICIDS2017, NSL-KDD, and TON_IoT, however, the challenges related to evaluation of DGA detection models are the imbalance in data, interpretability of model, and effectiveness of transferability to new attack vectors. Shifting the pendulum from reactive to proactive ML-based cyber solutions necessitates more than strong detection algorithms; real-time responsiveness, scalability, and explanation for actions are all equally essential. This paper has pointed out a few crucial research directions such as adversarial robustness, privacy-preserving model training, and ethical issues. Future work should strive to design adaptive, explainable, lightweight ML framework for efficient operation in dynamic and heterogeneous environments like IoT and edge computing landscape. By linking gap between theoretical research and practical deployment, ML holds the promise of redefining cybersecurity, enabling intelligent systems capable of anticipating and countering cyber threats before they inflict damage. A unique aspect of this review is its attempt to strategically align threat categories with optimal ML solutions. This mapping serves as a practical guide for researchers and practitioners in selecting the right learning approach for specific cyber risks.

References

[1] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: methods, systems and tools," IEEE Communications Surveys & Tutorials, vol. 16, no. 1, pp. 303-336, 2014.

[2] L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153-1176, 2016.

[3] Sharafaldin, I., Habibi Lashkari, A. and Ghorbani, A. A. (2018). Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. In Proceedings of the 4th International Conference on Information Systems Security and Privacy - ICISSP; ISBN 978-989-758-282-0; pages 108-116. DOI: 10.5220/0006639801080116.
[4] Ofusori, L., Bokaba, T., & Mhlongo, S. (2024). Artificial Intelligence in Cybersecurity: A Comprehensive Review and Future Direction. Applied Artificial Intelligence, 38(1).
[5] M. Ring, S. Wunderlich, D. Scheuring, D. Landes, and A. Hotho, "A survey of network-based intrusion detection data sets," Computers & Security, vol. 86, pp. 147–167, 2019.
[6] S. M. R. Islam et al., "A survey on anomaly detection in network traffic using machine learning," Computer & Security, vol. 87, 2019.
[7] X. Yuan, C. Li, and X. Li, "DeepDefense: Identifying DDoS attack via deep learning," in IEEE ICIS, 2017.
[8] Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in IEEE NetSoft, 2016.
[9] Alazab et al., "Deep learning applications for cybersecurity: A review," IEEE Access, vol. 9, pp. 24365–24365, 2021.
[10] M. Shone et al., "A deep learning approach to network intrusion detection," IEEE Trans. Emerging Topics in Computational Intelligence, vol. 2, no. 1, pp. 41–50, 2018.
[11] S. Bedi and J. Kim, "Anomaly detection in industrial control systems using deep learning," ICT Express, vol. 7, no. 4, pp. 413–419, 2021.
[12] S. Wang, W. Xu, and Y. Liu, "Res-TranBiLSTM: An intelligent approach for intrusion detection in IoT," Computer Networks, vol. 235, p. 109982, 2023.
[13] T. G. Dietterich, "Ensemble methods in machine learning," in MCS, Springer, 2000, pp. 1–15.
[14] B. Gupta, R. C. Joshi, and M. Misra, "Dynamic defense strategy against DDoS attacks in cloud computing," Journal of Network and Computer Applications, vol. 30, no. 4, pp. 1143–1154, 2021.
[15] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," in IEEE Symposium on Computational Intelligence for Security and Defense Applications, 2009.
[16] N. Moustafa and J. Slay, "UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems," in Military Communications and Information Systems Conference (MilCIS), 2015.
[17] N. Moustafa, "TON_IoT Datasets: The Ultimate Realistic IoT Datasets for AI-based Intrusion Detection Systems," arXiv preprint arXiv:2003.08544, 2020.
[18] M. H. Sivanathan et al., "Classifying IoT Devices in Smart Environments Using Network Traffic Characteristics," IEEE Transactions on Mobile Computing, vol. 18, no. 8, pp. 1745–1759, 2019.
[19] W. Wang et al., "Cyber-attacks detection in industrial systems using machine learning algorithms," Reliability Engineering & System Safety, vol. 202, p. 107061, 2020.
[20] Alazab et al., "Machine learning-based malware detection: A review," Future Generation Computer Systems, vol. 115, pp. 211–221, 2021.
[21] R. Vinayakumar et al., "Deep Learning Approach for Intelligent Intrusion Detection System," IEEE Access, vol. 7, pp. 41525–41550, 2019.
[22] N. Papernot et al., "The limitations of deep learning in adversarial settings," in IEEE EuroS&P, 2016.
[23] R. Guidotti et al., "A Survey of Methods for Explaining Black Box Models," ACM Computing Surveys, vol. 51, no. 5, pp. 1–42, 2018.
[24] K. Shokri and V. Shmatikov, "Privacy-preserving deep learning," in Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS), 2015.
[25] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 10, no. 2, pp. 1–19, 2019.