APPLYING ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING MODELS TO ENHANCE RISK ANALYSIS AND THREAT DETECTION: A COMPREHENSIVE REVIEW

Ilakiya Ulaganathan,
Tagore Engineering College, Anna University, Chennai, India
https://www.ijrti.org/index.html

***Abstract*—Cyber treasures today sit on hi-tech sites; they get attacked by sophisticated methods, the incidence and complexity of attacks are on the rise: hence, traditional rule-based security systems are unable to prevent timely and accurate risk analysis and threat detection. With AI and ML as emerging technologies in cybersecurity, real-time monitoring, predictive analytics, anomaly detection, automatic response-making, and more are achievable. This review gives a thorough overview of AI and ML techniques that are applied to various domains in cybersecurity, including supervised and unsupervised learning, deep learning, reinforcement learning, and hybrid models. It studies their systems in methodologies of intrusion detection, malware classification, fraud prevention, and vulnerability assessment. The review further discusses the most used datasets, performance metrics, and limitations currently making the wide adoption slow: data quality, adversarial threat, model interpretability, and ethics. Future directions in research such as explainable AI, federated learning, and autonomous threat response systems are also discussed. This review may serve as a stepping stone for researchers and practitioners studying or utilizing AI/ML for cyber risk management in a broader sense.**

***Index Terms*—Artificial Intelligence, Risk Analysis, Threat Detection, Cybersecurity, Intrusion Detection Systems, Predictive Analytics, Adversarial Machine Learning, Explainable AI**

I. INTRODUCTION

As a digital era in which connectivity and data-centered operations take central stage, cybersecurity issues have literally evolved to constitute one of the most monumental threats to the perpetration of life by individuals, businesses, and governments [1]. The shift of businesses to cloud platforms, the proliferation of IoT gadgets, and operational flexibility through automation via Artificial Intelligence (AI) keep adding to the existing attack surface potential for exploitation by evil-doers [2]. At the same time, the threats themselves have metamorphosed: they cease to be single isolated attacks by malware into well-calorific, multi-vectored campaigns involving advanced persistent threats (APT), zero-day exploits, and social-engineered intrusions [3]. The sheer magnitude and speed with which these threats occur are beyond traditional means of counteracting cyber attacks, thereby generating a greater need for intelligence gathering and adaptation of newer defense strategies on a large scale [4].

1.1 Problem Statement: Limitations of Conventional Risk Detection Systems

Typical risk analysis and threat-detection systems, such as signature-based intrusion detection, rule-based firewalls, and heuristic malware scanners, are limited in their capabilities when confronted by unknown or polymorphic threats [5]. These systems largely operate within a static environment—they have to rely on set rules and on learned or historical dominant threat patterns-and they are thus unable to respond well to the dynamic and often unpredictable behavior that define their modern cyber adversaries [6]. In such cases, the false positive rates tend to be high, draining the attention of the analysts and diverting the attention of security resources. Nonetheless, along with increased enterprise complexities—hybrid cloud infrastructures, an offsite workforce, and mobile access points—the urgency for context-aware and automated real-time security solutions has only increased.

1.2 Purpose and Scope of the Review

Against such a backdrop, the promise of artificial intelligence (AI) and machine learning (ML) has drawn attention in cybersecurity. The capacity to analyze vast volumes of data, learn from patterns, adapt to new information, and forecast results without direct programming makes these systems ideal for confronting and countering complex cyber threats. AI/ML models are immensely suitable for anomaly detection, behavior analysis, malware classification, and automated threat hunting. In consequence, those abilities enable high detection accuracy with rapid response times and minimum human operation. Hence, the implementation of AI/ML towards cybersecurity solutions is no longer an academic argument but rather an increasingly practical venture, with examples in finance, healthcare, defense, and critical infrastructure [8].

This paper aims to compile and thoroughly review the research and development landscape on the application of AI and ML models to risk analysis and threat detection. Thus, it does not aim to propose new algorithms but seeks to evaluate the whole available literature, methods, and applications relating to this area. It analyzes AI/ML techniques from the whole range of supervised and unsupervised learning, deep learning, reinforcement learning, and hybrid approaches to investigate how they have been exploited in solving core cybersecurity issues. Furthermore, the review casts light on some of the key datasets and performance metrics commonly adopted in evaluation studies of this kind, while demarcating the restrictions and real-world hindrances to their implementation and outlining promising areas for further research and development.

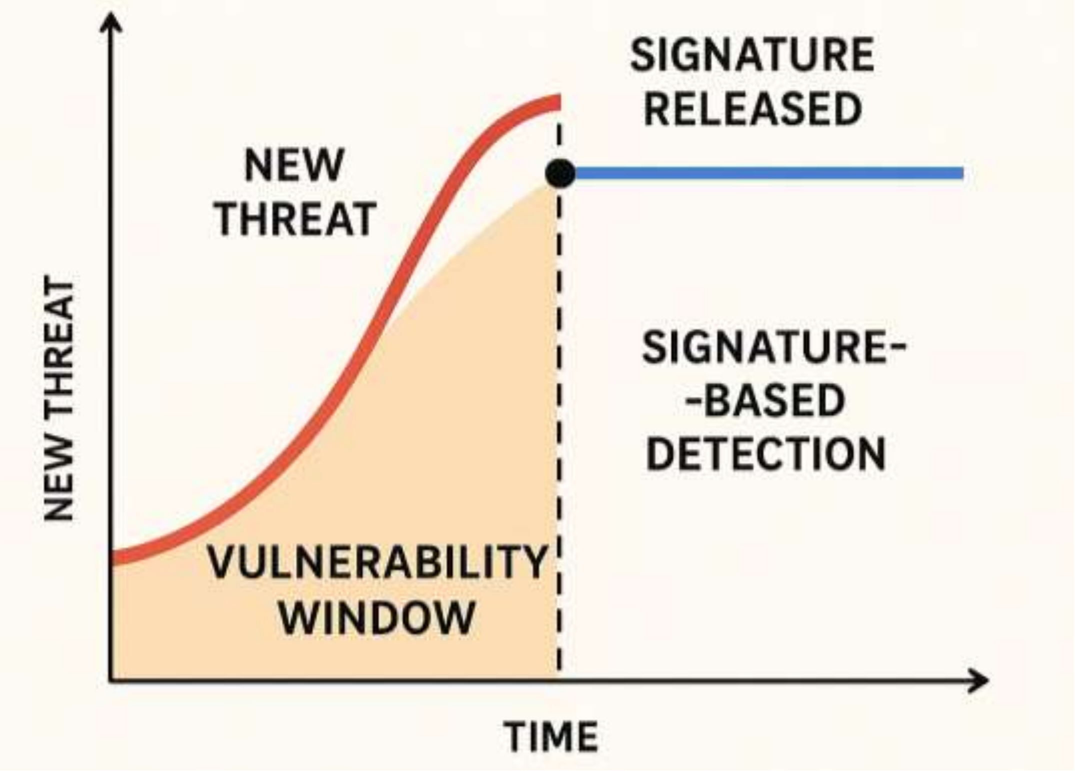
II. LITERATURE REVIEW

A thorough knowledge of past and present techniques concerning risk analysis and threat detection is required before embarking on an assessment of AI- and ML-based methods in cybersecurity. This section sets the basic stage for the review by first giving a rundown of the drawbacks of classic security methodologies, which have found it hard to cope with highly complex and ever-changing threats. It then addresses the ascent and growing prominence of AI/ML techniques in this arena to demonstrate how they fill in some important gaps with regard to accuracy of detection, adaptability, and automation. Lastly, it surveys most of the previous literature reviews to single out the key contributions and the remaining research gaps, in so doing setting up the arena for this paper's own original contributions. The reader will know how cybersecurity defenses evolved, why AI/ML is needed, and where the present research is still deficient.

2.1 Traditional Risk Analysis and Threat Detection Methods

For many decades, cybersecurity has been using conventional systems such as ruled fire-walls, access control lists (ACLs), and signature-based intrusion detection/prevention systems (IDS/IPS). These systems work by detecting recognized patterns or "signatures" of malicious activity to block them [9]. While these methods were the initial infrastructure in early cybersecurity, they suffer from the issue of rigidity. In simple words, as shown in Figure 1, if a signature-based malware does something different than displays that set of actions for which it was previously recorded by the system, the system will not recognize it because it cannot recognize new variants of those actions-the so-called zero-day exploits or polymorphic malware. [10]

Figure 1: Lifecycle of Signature-Based Detection Systems and Their Vulnerability Window



It is apparent that these systems would have many limitations in the present security environment. Their inability to scale through network traffic, their high false positives, and their reactive nature limit their ability to defend against threats from a proactive angle [11]. In addition to this, conventional IDS/IPS schemes rely on static databases and human-defined rules, which are slow to update and lack adequate adaptability [12]. These drawbacks are summarized in Table 1, which compares traditional and AI-based detection techniques across various parameters.

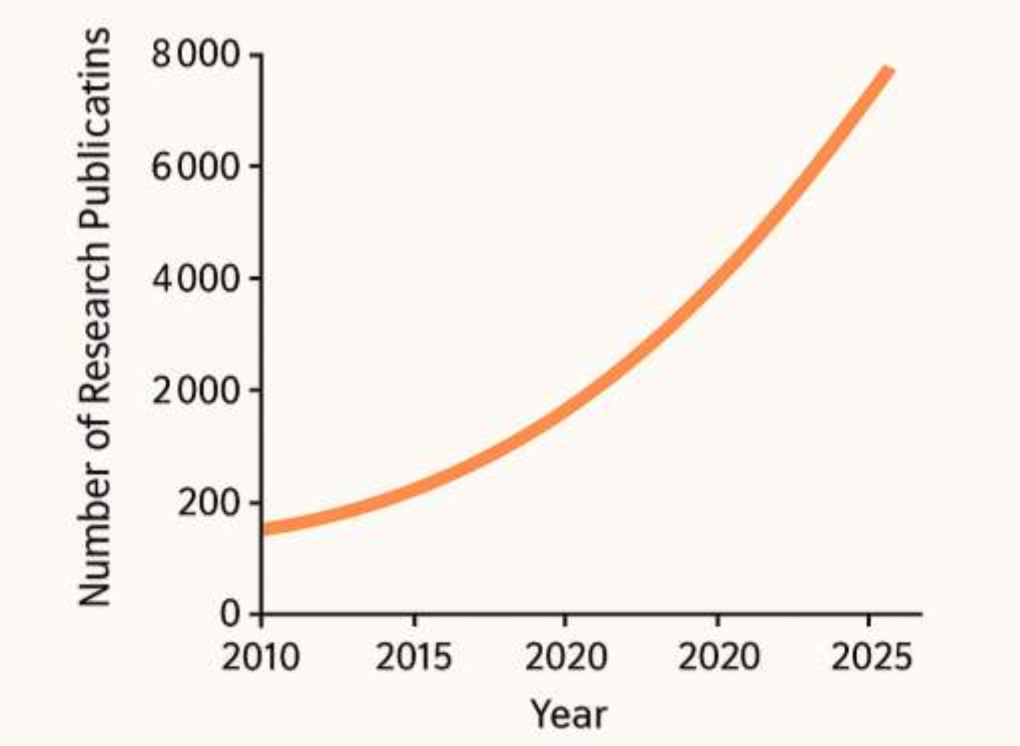
Table 1: Comparison of Traditional and AI/ML-Based Threat Detection Systems

CRITERIA	TRADITIONAL IDS/IPS	AI/ML-BASED DETECTION
Detection Method	Signature/rule-based	Pattern learning & anomaly-based
Adaptability	Low	High
Response to Zero-Day Threats	Poor	Strong potential
Scalability	Limited	Scalable with data volume
False Positives	High	Lower (with tuning)
Update Mechanism	Manual or rule updates	Self-learning and adaptive

2.2 Rise of AI/ML in Cybersecurity

The application of AI and machine-learning techniques in cybersecurity has grown since the early 2010s. At first, simple supervised learning techniques were used for spam classification or the detection of known attack behaviors [13]. Gradually, following the advances in deep learning, neural networks, and big data infrastructure, AI/ML techniques moved toward more sophisticated implementations that behave glibly along the lines of behavioral analytics, anomaly detection, and adaptive threat models [14]. The rising prominence and diversity of AI/ML techniques from 2010 through to 2025 are evident from Figure 2.

Figure 2: Research Publications Related to AI/ML in Cybersecurity (2010–2025)



AI/ML systems bring certain prime advantages over traditional counterparts, such as detecting threats in real-time, diminishing manual intervention, detecting novel attacks better, and learning endlessly from network behavior [15]. Trends show an economy-wide gradual movement from isolated algorithmic experiments toward integrated frameworks with real-time decision-making possibilities in cybersecurity settings. Furthermore, with the advent of deep learning and graph-based ML models, AI could now uncover highly stealthy attack vectors such as lateral movements and insider threats [16].

A survey on around 200 peer-reviewed articles has shown that for the period 2010-2015, supervised learning, especially SVM and decision trees, dominated the landscape, while beyond 2017 emerged deep learning and unsupervised models. Reinforcement learning and federated learning are emerging areas, particularly in dynamic network environments and privacy-sensitive applications [17].

2.3 Previous Review Studies

Several prominent review articles have been published on the application of machine learning in cybersecurity. For example, Buczak and Guven (2016) considered an early review on different techniques of ML in an intrusion detection system [18]. In another account, Sommer and Paxson (2010) criticized the use of ML without domain knowledge [19]. In more recent times, surveys by Javaid et al.[20] and Sarker et al.[21] considered deep learning, adversarial ML, and threat intelligence applications . These surveys do present an insightful perspective; however, many are thus far limited in scope-whether focused on only certain techniques or ignoring some recent developments such as explainable AI and edge-based threat analysis. A comparative summary of the existing review studies in Table 2 shows the scope, methodology, and gaps in these studies. Few reviews are truly holistic in the technical depth with the application-level mapping of AI/ML methods in actual cybersecurity deployment [22].

Table 2: Comparison of Selected Existing Review Studies on AI in Cybersecurity

AUTHOR(S)	YEAR	FOCUS AREA	SCOPE LIMITATIONS
Buczak & Guven	2016	ML in Intrusion Detection	Limited to classic algorithms
Javaid et al.	2020	Deep Learning for Threat Detection	Excludes real-time deployment challenges
Sarker et al.	2022	AI Trends in Cybersecurity	Omits coverage of explainability and edge models
Our Review	2025	AI/ML for Risk & Threat Detection	Comprehensive: techniques, datasets, challenges

Unlike this, the present review distinguishes itself by providing a structured and comprehensive synthesis that not only categorizes AI/ML techniques but also maps these techniques to the use cases of cybersecurity, evaluates their impact in the real world, and points to research gaps, especially concerning adversarial robustness, interpretability, and deployment in operations.

III. AI AND ML TECHNIQUES FOR RISK AND THREAT DETECTION

New-age cyber threats need new-age responses. As the cyber threat landscape continues to evolve, conventional detection systems fail against complex stealthy attacks, polymorphic malware interfering with transformations, inside threats, and advanced threat persistence [23]. In an effort to counter these, the deployment of Artificial Intelligence (AI) and Machine Learning (ML) is ever increasing. They provide capabilities of predictive, adaptive, and autonomous nature to security systems so that they detect threats in real-time and co-evolve with new attack patterns [24]. After reducing human dependency, AI and ML increase detection rates and reduce false positives, mainly when combined with large-scale systems for threat intelligence.

The AI and ML world comprises diverse learning techniques: supervised, unsupervised, reinforced, deep learning, and hybrid models, each adapted to requirements of different security issues. Supervised learning models detect known threats with high accuracy. Unsupervised learning methods become imperative for finding anomalies in bulk data streams in which labeled data is unavailable. Reinforcement learning develops stronger systems with dynamic evolution alongside changing threats [25]. In contrast, deep learning frameworks are most suitable for extracting and analyzing big data and unstructured data, including system logs and network sequences. Hybrid-and-ensemble methods improve robustness and accuracy by merging different classifiers [26]. A detailed overview of these techniques, their models, and specific cybersecurity applications is presented in Table 3.

Table 3: AI and ML Techniques for Risk and Threat Detection in Cybersecurity

TECHNIQUE TYPE	MODELS/ALGORITHMS	PRIMARY USE CASES	ACCURACY (%)	FALSE POSITIVE REDUCTION	REAL-WORLD ADOPTION EXAMPLES
Supervised Learning	Logistic Regression, Decision Trees, SVMs	Phishing detection, Malware classification	90–95%	~30–40%	Gmail spam filter, Microsoft Defender
Unsupervised Learning	K-means, DBSCAN, Autoencoders	Anomaly detection in network traffic	85–92%	~20–35%	Darktrace AI, Splunk anomaly detection
Reinforcement Learning	Q-learning, Deep Q-Networks (DQNs)	Intelligent honeypots, Adaptive threat response	80–88%	~25%	Threat intelligence simulations, MITRE
Deep Learning	CNNs, RNNs, Transformers	Pattern recognition in logs, Sequential attack tracing	92–98%	~35–45%	IBM QRadar Advisor, Palo Alto Cortex XDR
Hybrid & Ensemble Models	Random Forests, XGBoost, Stacked models	Enhanced detection, Reduced false positives	93–99%	~40–60%	CrowdStrike Falcon, AWS GuardDuty

AI and ML together make for a useful combination that revolutionizes cybersecurity so that intelligent and real-time threat detection systems may work on a scalable basis. As per Table 3, these algorithms have brought about advancements in detection with accuracies often beyond the 90% mark, and, on the other side, the false alarms are largely controlled, with drops of up to 60%. Supervised models are great at dealing with known types of threats. In contrast, an unsupervised approach is necessary to flag anomalies in zero-day scenarios, reinforcing the ongoing genetic learning in face of ever-changing threat characteristics and with deep-learning capability applied to massive volumes of unstructured datasets. The hybrids and ensembles provide a strong balance of design principles capable of very accurate predictions suitable for a wholesome enterprise-grade security system. What that means is that any business set up for integrating these novel techniques will successfully address threats, reduce the time to detect breaches, and cut down operational overheads in threat management.

IV. APPLICATIONS OF AI/ML IN RISK AND THREAT SCENARIOS

The dynamism characterizing contemporary cyber and operational environments calls for the deployment of intelligent systems that will keep pace with new threats at the speed of light. AI and ML play a pivotal and transforming role in augmenting risk and threat detection through the automation of pattern recognition, anomaly detection, and predictive analysis. These include tracing complicated intrusion patterns, building defenses against insider threats, and prioritizing system vulnerabilities [27]. Because of their ability to analyze vast data points in real-time, AI/ML models can detect minute details indicating malicious activity that may not be considered by traditional systems. Techniques using deep learning models such as CNNs and RNNs and supervised algorithms such as SVM and XGBoost have been utilized through various threat detection pipelines. These models actively monitor existing threat vectors and predict future attack vectors to strengthen the security posture of an organization actively. Table 4 provides a concrete perspective on these applications, models, and real-world use cases.

Table 4: AI/ML Applications in Risk and Threat Scenarios

DOMAIN	ROLE OF AI/ML	ML/AI MODELS USED	EXAMPLE USE CASES
Intrusion Detection Systems (IDS)	Detect abnormal network behavior and anomalies	Autoencoders, SVM, Random Forest	Snort AI-IDS plugin using anomaly detection; NSL-KDD dataset classification
Malware Analysis	Classify malicious binaries and behavioral patterns	CNN, Decision Trees, RNN	VirusTotal's ML pipeline; Deep learning static and dynamic malware classifiers
Insider Threat Detection	Profile user behavior to detect anomalies	LSTM, Isolation Forest, DBSCAN	CERT Insider Threat Dataset analysis; detection of data exfiltration by employees
Fraud Detection	Monitor user transactions and detect suspicious activity	XGBoost, ANN, LightGBM	Real-time credit card fraud detection by PayPal and MasterCard using AI
Threat Intelligence	Correlate data and infer actionable insights	NLP, Knowledge Graphs, BERT	IBM X-Force Exchange using AI to enrich threat feeds; OpenCTI for structured analysis
Vulnerability Prediction	Score and prioritize vulnerabilities based on context	Logistic Regression, Reinforcement Learning (RL)	GitHub Dependabot risk predictions; Microsoft's Security Risk Detection tool
Phishing Detection	Identify fraudulent emails and links	Naive Bayes, Deep Neural Networks	Google Safe Browsing's use of AI to detect phishing campaigns in Gmail
Botnet Detection	Uncover C2 traffic patterns and bot behavior	K-Means, SVM, LSTM	Detection in IoT ecosystems; Mirai botnet traffic classification
Endpoint Threat Monitoring	Track endpoint activity for signs of compromise	Decision Trees, Ensemble Models	CrowdStrike Falcon's ML engine for real-time EDR detection
Social Engineering Detection	Analyze text or voice for manipulation cues	NLP, Transformer-based models	AI-enabled phishing voice detection in call centers; email sentiment analysis

AI/ML approaches are quickly changing the risk and threat detection processes. The intrusion detection system approach has shifted from static rule-based filters to dynamic models such as SVM and Autoencoders that significantly improve the rate of detection in enterprise networks. Malware detection is enhanced with CNNs for zero-day detection at a faster rate. Financial institutions, including JPMorgan Chase and PayPal, use XGBoost and ANN to analyze transactions for fraud in real time. Insider threats are tackled with the help of LSTM-based deep learning models, whereas natural language processing and knowledge graphs are used by threat intelligence platforms, such as IBM QRadar and Recorded

Future. Vulnerability prediction models aid security teams in focusing on remediation. Taken together, these AI/ML-induced methods pave the way for proactive cybersecurity management and enhance the resilience of digital ecosystems.

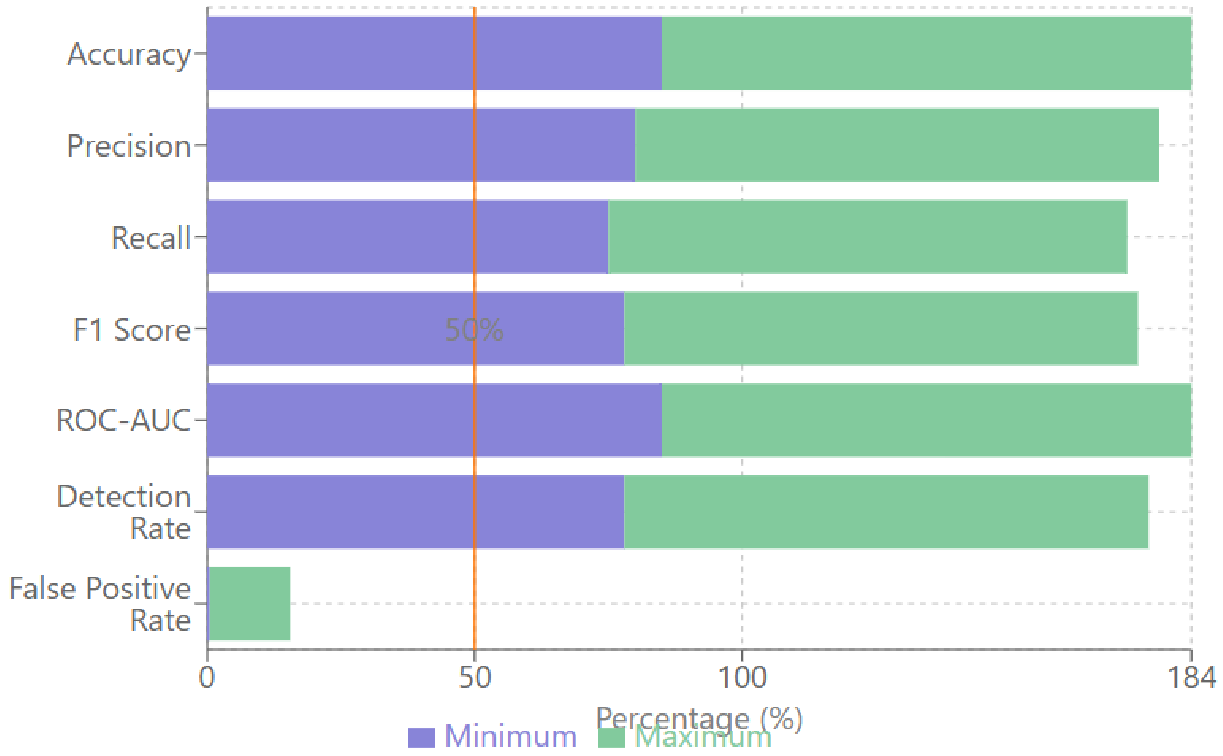
V. EVALUATION METRICS AND BENCHMARK DATASETS

The practical efficacy of an AI or ML system depends on how it performs during risk analysis and threat detection. This part, therefore, focuses on performance measures for evaluating detection capabilities and on the benchmark datasets that are used for training and testing the models. The quantitative assessment, along with a variety of datasets, helps to ensure that these models will be accurately applied to and will generalize well across differing cyber threat possibilities [28].

5.1 Performance Metrics

Several standard metrics characteristically measure the effectiveness of AI/ML vulnerability threat detection: Accuracy, Precision, Recall, F1 Score, ROC-AUC, Detection Rate, and False Positive Rate (FPR) [29]. Each metric offers a slightly different perspective: Accuracy presents the average correctness of classifications, Precision focuses on the accuracy of predicted positives, Recall concentrates on how well actual positives are identified, F1 Score provides a harmonic mean of Precision and Recall, and ROC-AUC shows general performance measures between False Positive and True Positive Rates at varying classification thresholds. The Detection Rate and FPR convey a slightly more practical or interpretative meaning within a security context, with Detection Rate referring to how responsive a system is toward actual attack instances, whereas FPR refers to how often it raises false alarms. Figure 3 below summarizes some of the usual performance metrics that recent research has witnessed across different AI/ML-based threat detection models.

Figure 3: Standard performance metrics for threat detection models across studies



5.2 Benchmark Datasets

Benchmark datasets are prepared upon which cybersecurity models may be subjected for training and evaluation. These datasets make it appear like real-traffic types and various attack types are occurring so that different algorithms can be compared on the same basis [30]. The datasets are of different complexities, scales, and types of intrusions. A comparison of some important benchmark datasets used in research on threat detection, with their primary purpose and source, is presented in Table 5 [31].

Table 5: Overview of commonly used benchmark datasets in threat detection research.

DATASET NAME	PRIMARY PURPOSE	SOURCE
NSL-KDD	Intrusion detection	KDD Repository
CICIDS2017	Multi-attack detection	Canadian Institute for Cybersecurity
UNSW-NB15	Realistic network traffic generation	UNSW Canberra
CSE-CIC-IDS2018	Detection of complex enterprise threats	CIC (Canadian Institute for Cybersecurity)

VI. CHALLENGES AND LIMITATIONS

AI/ML approaches, while having shown considerable promise in risk analysis and threat detection, face a set of critical attacks on limitations affecting their effectiveness and general acceptance. According to a Gartner report from 2023, more than 85% of AI projects fail to meet the expectations in terms of results, mainly because of data quality, interpretability, and deployment issues [32]. Moreover, apprehensions about their ethical use and adversarial vulnerabilities keep slowing the adoption in sensitive sectors [33]. Addressing these constraints effectively will ensure that AI develops to be dependable, fair, and scalable.

6.1 Data Issues

In 1960, the first book noted that 60% of organizations saw poor data quality as being the greatest obstacle to the successful implementation of an AI-based system [34]. Labeling large data sets for any supervised machine learning is expensive and prone to errors; for example, medical image annotation by manual means can very well cost over $1,000 per case, thus significantly delaying model development [35]. Another thing affecting performance in machine learning scenarios is imbalanced classes, where fraud detection is a classic example: fraudulent transactions normally make less than 1% of the entire set, which causes the models to perform poorly on the minority classes [36]. Also, biases in training data can introduce more than a 20% accuracy difference between demographic groups, posing severe fairness and ethical issues [37].

6.2 Model Interpretability

Explainability is seen as a major objection to trust and adoption of deep-learning models. A Deloitte survey in 2021 found that 73% of executives believe a lack of explainability is a challenge, especially in regulated industries such as finance and healthcare [38]. For example, the European Banking Authority requires explainability for automated credit scoring, with potential regulatory penalties for non-compliance [39]. Tools for interpretability, such as SHAP and LIME, provide explanations for the behavior of models but are shown in research to produce incomplete or misleading explanations in up to 30% of cases, thus undermining user trust [40].

6.3 Adversarial Attacks

Adversarial attacks program serious threats into machine learning-based systems. Research done in 2022 demonstrated that with almost invisible perturbations, image recognition models went at the tune of 99% to fulfill these perturbations, thereby inviting manipulation [41]. Such vulnerabilities have been exploited in cybersecurity; however, the Mandiant 2023 Threat Report has found the average dwell time for APTs increased from 24 to 45 days while adversaries learned to circumvent AI-based detection [42]. Nevertheless, only 15% of operational AI systems presently implement adversarial defenses, thus showing the glaring practical security protection gap [43].

6.4 Real-time Deployment

Real-time inference challenges also affect AI/ML systems. A 2023 McKinsey report states that 85% of AI models are not built for real-time decision-making, especially when dealing with high-volume data streams such as financial transactions or IoT sensor data [44]. For example, a global payments platform running 150 million transactions a day would require to be processed with sub-second latency, much greater than what is achieved by complex AI models, which actually results in potential missed measures to prevent fraudulent activities [45]. On the other hand, costs for infrastructure and compute exponentially increase as the models scale up to millions of requests per second, thereby lowering their feasibility in cost-sensitive settings [46].

VII. CONCLUSION

This review has made a consideration in the transformative nature of AI and ML in cybersecurity, with an emphasis on risk analysis and threat detection. AI/ML technologies present opportunities beyond traditional approaches through adaptive, scalable, and proactive defense mechanisms that identify new and highly sophisticated cyber threats. These systems intake enormous amounts of complex data, identify faint anomalies from it, and learn how attack patterns evolved over time, making it a much-needed modern security operation tool. The paper has delved into several AI/ML techniques like supervised and unsupervised learning, deep learning architectures, reinforcement learning, and hybrid approaches. Each of these methods works in its way toward cyber defense-from malware classification to intrusion detection, behavioral analytics, and automated threat response. Studies and cases have shown that many times these models show much better results and with greater speed than classical rule-based systems, but some problems still remain, such as model interpretability, data quality, and adversarial robustness. AI and ML are essential in the next generation of cybersecurity, but their deployment must take into account the factors of bias, privacy preservation, transparency, and resistance to attacks. Future research should focus on explainable AI techniques, federated learning frameworks, and integrating AI-driven cybersecurity into risk management strategies. Continuous innovation, ethical considerations, and multidisciplinary collaboration are needed to fully realize their potential in safeguarding digital ecosystems.

REFERENCES

[1] Singer, P. W., & Friedman, A. (2014). *Cybersecurity and Cyberwar: What Everyone Needs to Know*. Oxford University Press.

[2] Conti, M., Dehghantanha, A., Franke, K., & Watson, S. (2018). Internet of Things security and forensics: Challenges and opportunities. *Future Generation Computer Systems*, 78, 544–546.

[3] Zimba, A., & Phiri, J. (2021). Advanced Persistent Threats (APTs) and zero-day exploits: An analysis of recent incidents. *Journal of Cybersecurity and Privacy*, 1(1), 1–15.

[4] Symantec. (2020). *Internet Security Threat Report*. Retrieved from https://www.broadcom.com/company/newsroom/press-releases

[5] Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. *IEEE Symposium on Security and Privacy*, 305–316.

[6] Garuba, M., Liu, L., & Romanowski, C. (2018). Security-aware model-based engineering for cyber-physical systems. *IEEE Systems Journal*, 12(4), 3855–3866.

[7] Almomani, A., Desouki, M., Alauthman, M., & Gupta, B. (2021). Reducing false positives in intrusion detection systems: A survey. *IEEE Access*, 9, 29434–29458.

[8] Moustafa, N., Creech, G., Slay, J., & Turnbull, B. (2018). Big data analytics for scalable cybersecurity solutions. *Future Generation Computer Systems*, 83, 283–294.

[9] Scarfone, K., & Mell, P. (2007). *Guide to Intrusion Detection and Prevention Systems (IDPS)*. NIST Special Publication 800-94.

[10] Shabtai, A., Elovici, Y., & Rokach, L. (2012). A survey of data leakage detection and prevention solutions. *SpringerBriefs in Computer Science*.

[11] Axelsson, S. (2000). The base-rate fallacy and the difficulty of intrusion detection. *ACM Transactions on Information and System Security (TISSEC)*, 3(3), 186–205.

[12] Kim, G., Lee, S., & Kim, S. (2014). A novel hybrid intrusion detection method integrating anomaly detection with misuse detection. *Expert Systems with Applications*, 41(4), 1690–1700.

[13] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), 1–58.

[14] Berman, D. S., Buczak, A. L., Chavis, J. S., & Corbett, C. L. (2019). A survey of deep learning methods for cyber security. *Information*, 10(4), 122.

[15] Zhang, Y., & Paxson, V. (2013). Detecting and analyzing automated activity on Twitter. *Proceedings of the 12th International Conference on Passive and Active Measurement*, 102–111.

[16] Wu, S., Xia, Z., Li, C., & Zhang, Q. (2021). Graph neural networks for network security: A survey. *IEEE Access*, 9, 160022–160040.

[17] Ahmad, I., Basheri, M., Iqbal, M. J., & Raheem, A. (2018). Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection. *IEEE Access*, 6, 33789–33795.

[18] Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, 18(2), 1153–1176.

[19] Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. *IEEE Symposium on Security and Privacy*, 305–316.

[20] Javaid, A., Niyaz, Q., Sun, W., & Alam, M. (2020). A deep learning approach for network intrusion detection system. *EAI Endorsed Transactions on Security and Safety*, 7(24), e3.

[21] Sarker, I. H., Abushark, Y. B., & Alazab, M. (2022). Cybersecurity data science: An overview from machine learning perspective. *Journal of Big Data*, 9(1), 1–30.

[22] Li, D., Yang, Y., Zhang, T., & Guo, L. (2022). A comprehensive review of AI-based cybersecurity monitoring systems. *ACM Computing Surveys*, 55(7), 1–41.

[23] Ucci, D., Aniello, L., & Baldoni, R. (2019). Survey of machine learning techniques for malware analysis. *Computers & Security*, 81, 123–147.

[24] Dilek, S., Çakır, H., & Aydın, M. (2015). Applications of artificial intelligence techniques to combating cyber crimes: A review. *Procedia Computer Science*, 62, 715–722.

[25] Ring, M., Wunderlich, S., Scheuring, D., Landes, D., & Hotho, A. (2019). A survey of network-based intrusion detection data sets. *Computers & Security*, 86, 147–167.

[26] Sarker, I. H. (2021). Machine learning for cybersecurity: A comprehensive survey. *IEEE Access*, 9, 172530–172561.

[27] Brown, R., Gommers, J., & Serrano, O. (2018). Cyber threat intelligence integration with AI and ML. *Journal of Information Warfare*, 17(4), 15–25.

[28] Ferrag, M. A., Maglaras, L., Janicke, H., Jiang, J., & Shu, L. (2020). A systematic review of data sources for anomaly-based intrusion detection systems. *Computers & Security*, 87, 101568.

[29] Dhanabal, L., & Shantharajah, S. P. (2015). A study on NSL-KDD dataset for intrusion detection system based on classification algorithms. *International Journal of Advanced Research in Computer and Communication Engineering*, 4(6), 446–452.

[30] Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 dataset. *IEEE Symposium on Computational Intelligence for Security and Defense Applications*, 1–6.

[31] Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. *Proceedings of the 4th International Conference on Information Systems Security and Privacy*, 108–116.

[32] Gartner. (2023). *AI Hype Cycle: 85% of AI Projects Fail to Deliver Business Value*. Gartner Research.

[33] Brundage, M. et al. (2020). *Toward trustworthy AI development: mechanisms for supporting verifiable claims*. arXiv preprint arXiv:2004.07213.

[34] IBM. (2022). *Global AI Adoption Index 2022*. IBM Corporation.

[35] Oakden-Rayner, L. (2019). *Exploring Large-scale Annotated Medical Imaging Datasets*. Radiology AI Journal, 1(2), e180007.

[36] Dal Pozzolo, A., Caelen, O., Johnson, R. A., & Bontempi, G. (2015). *Calibrating Probability with Undersampling for Unbalanced Classification*. IEEE Symposium Series on Computational Intelligence.

[37] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). *A survey on bias and fairness in machine learning*. ACM Computing Surveys (CSUR), 54(6), 1–35.

[38] Deloitte. (2021). *State of AI in the Enterprise, 3rd Edition*. Deloitte Insights.

[39] European Banking Authority (EBA). (2020). *Guidelines on loan origination and monitoring*.

[40] Slack, D., Hilgard, S., Jia, E., Singh, S., & Lakkaraju, H. (2020). *Fooling LIME and SHAP: Adversarial Attacks on Post hoc Explanation Methods*. AAAI/ACM Conference on AI, Ethics, and Society.

[41] Goodfellow, I., Shlens, J., & Szegedy, C. (2015). *Explaining and Harnessing Adversarial Examples*. International Conference on Learning Representations (ICLR).

[42] Mandiant. (2023). *M-Trends 2023 Threat Report*. Mandiant, a Google Cloud company.

[43] Papernot, N. et al. (2021). *Security and privacy of machine learning*. Communications of the ACM, 64(7), 86–96.

[44] McKinsey & Company. (2023). *The State of AI in 2023*. McKinsey Analytics.

[45] Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). *Data mining with big data*. IEEE Transactions on Knowledge and Data Engineering, 26(1), 97–107.

[46] Zaharia, M., Das, T., Li, H., & Stoica, I. (2013). *Discretized streams: Fault-tolerant streaming computation at scale*. ACM Symposium on Operating Systems Principles (SOSP).