

ENHANCING THE CYBER RESILIENCE OF INFORMATION SYSTEMS THROUGH MACHINE LEARNING

Ndung'u H.,¹ Munyao V.,¹ Karani J.,² Wasike J.¹

¹Kirinyaga University, Kenya

²Karatina University, Kenya

Correspondence: Hndungukp15@gmail.com

ABSTRACT

Organizations today operate in an environment of relentless cyber threats that jeopardize continuity of critical services. Conventional security approaches, primarily focused on preventing intrusions, have proven insufficient. This shortfall has shifted attention to cyber resilience, defined as organizations' capacity to maintain operations and recover effectively after cyber disruptions. Among the technologies driving resilience, Machine Learning (ML) stands out for its ability to analyze vast datasets, identify anomalies, and respond effectively to evolving threats. Its ability to process vast and complex datasets enables detection of subtle, hidden patterns that conventional tools or human analysts might overlook. To examine ML's role in strengthening cyber resilience, this study undertook a systematic review of recent literature (2020–2026). The review encompasses 87 peer-reviewed studies from reputable sources, with the aim of mapping ML applications across different stages of resilience and identifying recurring benefits and challenges. The goal is to provide a clear understanding of how ML contributes to resilience, evaluate its performance in practice, and identify the key challenges that must be overcome to ensure its safe and robust adoption. Results show that ML significantly enhances resilience by enabling proactive threat detection and adaptive response mechanisms. However, challenges such as data quality, model robustness, and the need for human oversight limit its effectiveness. The evidence suggests that ML is not a standalone solution but rather an enabling technology whose success depends on thoughtful design, integration with human expertise, and mitigation of inherent limitations. This paper contributes a structured framework for understanding ML's role in cyber resilience and highlights critical issues that warrant further investigation, offering a foundation for both practitioners and researchers to advance resilient cybersecurity strategies.

Keywords: Machine Learning; Cyber Resilience; Cybersecurity; Intrusion Detection; Threat Detection; Incident Response; Anomaly Detection; Adaptive Cyber Defense; Information Systems; Adversarial Machine Learning.

INTRODUCTION

The increasing reliance of modern society on digital infrastructures—spanning finance, healthcare, and energy—has expanded the attack surface for cyber adversaries. Traditional security models, built on static rules and signature-based detection, struggle to counter today's sophisticated and adaptive threats. These include polymorphic ransomware capable of altering its appearance, long-term targeted campaigns backed by substantial resources, and exploits of previously unknown software vulnerabilities^[1]. In response, the focus of cybersecurity has shifted from prevention alone to cyber resilience based on systems that can anticipate attacks, withstand stress, recover rapidly from disruptions, and continuously learn to strengthen future defenses. The central objective is to ensure continuity of essential services even under active attack ^[2].

Within this paradigm, Machine Learning (ML) has emerged as a transformative technology. Its ability to process vast and complex datasets enables detection of subtle, hidden patterns that conventional tools or human analysts might overlook. This capability enables ML to power systems that can forecast emerging threats, identify anomalous behavior in real time, and automate elements of incident response to outpace human reaction times. (Radanliev et al., 2024) ^[3] observed that by shifting security operations from a reactive stance to a proactive and adaptive posture, ML is increasingly recognized not merely as a supporting tool but as a foundational component in building resilient and intelligent information systems.

Problem Statement

Although the promise of ML in the area of cybersecurity is evident, its implementation within a resilience approach poses significant challenges. Thus, scattered data, outdated technology not built for ML, and a lack of competent personnel in data science and security are common problems in organizations. Equally worrying is the fact that the ML models themselves may be exposed. Hammad et al., (2024)^[2] asserted that attackers may craft inputs in a way that deceives the models, corrupts the data with which they are trained, or exploit bias in the design, making the technology that is supposed to make systems more resilient ironically vulnerable to attack.

Given these challenges, there is an urgent need to move beyond isolated, small-scale experiments to a systematic approach that enables understanding of how diverse Machine Learning (ML) techniques can be applied securely and effectively across the full lifecycle of cyber incident management—spanning preparation, response, and recovery ^[4].

A critical gap persists in the form of limited, evidence-based guidance on how organizations should select, implement, and govern ML systems to both strengthen resilience and mitigate associated risks ^[5]. This study seeks to address that gap by conducting a comprehensive analysis of recent studies with a view to providing a clear understanding of how ML contributes to resilience, evaluating its performance in practice, and identifying the key challenges that must be overcome to ensure its safe and robust adoption ^[6].

Objective

This study sought to;

1. analyze the specific contributions of ML to each stage of the cyber resilience lifecycle.
2. evaluate performance of different ML approaches using synthesized evidence from recent literature.
3. Identify and critically discuss the technical, operational, and ethical challenges in deploying ML for resilience
4. Propose an integrated framework and future research directions to realize ML's full potential in building resilient information systems.

Scope of the Study

The study is a systematic review of recent literature, industry reports, and real-world case studies published from 2020 to 2026. The review analyzes the use of ML in major areas of cyber resilience, i.e., detection of intrusions in the network and system, user behavior analysis to identify threats, prediction of attacks, and automated actions to deal with incidents and recovery. The analysis covers applications in various key sectors, including Industrial Control Systems for infrastructure such as power grids, medical infrastructure and medical devices, financial systems, and international supply chains ^[7]. Notably, the review is not limited to

evaluating technical performance, i.e., the accuracy with which a model would identify a threat, but also explores practical issues such as the challenge of adopting ML to older systems, the necessity of models that a human can comprehend and feel confident about, and organizational issues that predetermine the success or failure of these technologies. This methodology will help produce not only well-grounded academic results but also those that are actually applicable to professionals to create more resilient systems [6].

ML as a Foundational Pillar for Cyber Resilience

Existing literature does not perceive machine learning as a new tool of cybersecurity but rather as a primary element in developing resilient systems. This change fits into a contemporary concept of cyber resilience, which goes beyond merely not being hacked or being hacked. The four major capacities of resilience are described by modern frameworks as the 4Rs: Robustness, Rapid Recovery, Resourcefulness, and Redundancy. ML offers tangible, scientific methods of meeting all these objectives. Robustness- the inherent strength of any system is enhanced when ML actively hunts for any latent vulnerability within software or network environments before the attacker can detect it. ML helps teams to find actual incidents and react instantly, hence reducing recovery time significantly; rapid recovery is accelerated when the first sorting and analysis of the security alerts are automated. Resourcefulness, or solving new problems in times of crisis, is increased when ML assists in prioritization of actions and use of limited personnel in the optimal way during an attack [2].

Lastly, ML gives the chance to be smarter about Redundancy. It allows simulating various failure scenarios and providing intelligent backup plans instead of simply copying the systems at a high cost, keeping the services online and providing better services [10]. Notably, the four principles of resilience cannot occur without the role of the brain, which is played by ML in the current complicated digital contexts, as opined by [8].

How ML Enhances Resilience: Key Mechanisms

Research points to four main ways ML strengthens cyber resilience. **Anticipating threats before they hit.** ML models have the potential of transforming security to proactive. ML is capable of detecting patterns by studying the historical data on attacks, global threat reports,

and even chatter on hacker forums that indicate an impending attack. This enables defenders to build defenses and address weaknesses in advance, which is the "anticipate" aspect of resilience [3].

With Anomaly Detection, the Unknown can be detected. In case of new attacks that have previously unknown signatures, ML applies unsupervised learning. Methods such as Autoencoders get to know what normal network or user behavior is like [9]. They, in turn, raise red flags over a major anomaly to this trend, which may imply a new threat or a rogue insider attack. This plays a vital role in the detection of attacks that cannot be detected by traditional tools.

Response with Reinforcement Learning Automation. Reinforcement Learning (RL) can be used to enhance the response stage. In this case, an AI agent works in a free, simulated environment, and it is taught by trial and error which actions (such as isolating a device or blocking an IP address) result in the most desirable behavior to contain the threat [1]. As time passes, it learns the best response policies, which enable a rapid, automated response that decreases the damage window.

Lessons on Incidents to Improve Recovery. ML assists in streamlining recoveries after an attack and can prioritize the systems restoration to limit disruption of the business [10]. What is more important, ML examines the incident in full to study the lessons: What went wrong in the defense? What are the differences in the approaches? This forms a feedback loop where the whole security system would become stronger and more adapted to completing the cycle of recovery to better preparedness.

ML in Action: Sector-Specific Applications

Use of ML mechanisms varies significantly across different critical industries. **Securing Critical Infrastructure.** In the case of systems such as power grids or water treatment plants, ML models are optimally created to identify attacks with a goal of causing physical damage. They can detect False Data Injection, where attackers provide operators with artificial sensor data so that they can cause disastrous failures. ML assists in differentiating between genuine

sensor data and the interfered data, which keeps the integrity of these essential systems intact [11].

Protecting the Healthcare Systems. Healthcare should strike a balance between high security and stringent laws of patient privacy [7]. In this case, Federated Learning is a prospective solution. It enables training of ML models with data directly obtained by the medical device (such as a heart monitor) without ever leaving the hospital, and can be used to detect damaged equipment without violating privacy.

Fighting Financial Fraud. In finance, resilience refers to preventing real-time fraud. This is centered on ML, which examines millions of transactions with an aim of identifying patterns associated with credit card fraud or money laundering. Such systems keep on learning every new fraud type, and banks can modify their system in real time to keep the criminals at bay [8].

Communicating and Building Supply Chain Resilience. Contemporary supply chains are big information networks. ML can make them more resilient because it predicts problems such as a port closure or bankruptcy of a supplier by using shipping trackers, news, and weather reports data. It may then suggest alternative pathways or suppliers, as a means to allow the supply chain to change and keep moving on [10].

Current Challenges and Open Questions

Despite its promise, existing literature highlights several persistent hurdles that limit the effectiveness of Machine Learning (ML) in cyber resilience.

a) Technical and Adversarial Weaknesses

ML models remain vulnerable to deception. Attackers can subtly manipulate malicious data—for example, altering a malware file—so that an ML classifier misidentifies it as benign, allowing it to bypass defenses. This dynamic creates a perpetual arms race between attackers and defenders. Compounding the issue, high-quality training data for rare but severe attacks is scarce, leaving ML systems ill-prepared to counter the most consequential threats [5].

b) Real-World and Human-Centric Constraints

Integrating ML into legacy systems is often complex and costly. Furthermore, many advanced ML models function as “black boxes,” producing decisions without clear explanations. Thus, even when effective, this lack of transparency undermines trust among security analysts, raising regulatory concerns, and can lead organizations to disable such systems altogether [10].

c) Measurement and Research Gaps

A critical gap lies in the absence of robust methods for measuring resilience itself. Most studies rely on narrow metrics such as detection accuracy for specific tasks, which fail to capture system-wide outcomes. What is missing are comprehensive measures that reflect practical impacts—such as reductions in downtime during incidents or mitigation of data loss [3]. Developing such broad, outcome-oriented metrics is essential for demonstrating the tangible value of ML in building resilient organizations.

METHODOLOGY

Research Approach and Design

The study aimed to create an evidence-based framework constructed on a large portion of existing research. Since the machine learning and cyber resilience sphere is complicated and constantly evolving, implementing new experiments would be unrealistic. Rather, a strict Systematic Literature Review (SLR) was selected. This approach entailed locating, reviewing, and analyzing all the available literature on the subject to make general inferences. It is a clearly structured process that reduces prejudice and produces a trustworthy map of existing knowledge. This is the best way to answer "how and why" questions and also to be able to construct new ideas based on the available evidence.

Systematic Search and Selection Process

To ensure a comprehensive review, the search for literature followed a strict, three-step protocol.

Step 1: *All-inclusive search of the database.* We used four large academic databases of high-quality research in the field of technology and security, including IEEE Xplore, ACM Digital Library, ScienceDirect, and SpringerLink. We applied a particular search query that included a combination of keywords such as machine learning, cyber resilience, intrusion detection, and incident response^[8]. To have the latest and most significant work, the search was restricted to the period between January 2020 and March 2026. The search yielded 2347 articles.

Step 2: *Intense Relevance and Quality Rigorous Filtration.* To reduce this big list, we used two-stage filtering in which we had explicit rules:

- Stage 1 Title and Abstract Screen. Abstracts for all the 2,347 articles were read, and only those that were (a) empirical research or elaborate theoretical models, (b) oriented towards system resilience (not just detection), and (c) on information systems or critical infrastructure were chosen. We eliminated opinion articles, guides, and articles that talked about hardware alone. This filter left 215 prospective papers.
- Stage 2 Full-Text Quality Appraisal. We read the text of the 215 papers and evaluated them based on methodological clarity and adequacy. We filtered out the papers that had ambiguous methods or that did not focus on ML. The last group of articles analyzed included 87 high-quality and highly relevant studies.

Data Analysis and Synthesis

From the 87 articles, we retrieved important data from each of the 87 articles using a standardized form focused on the types of machine learning applied, the role of the resilience process the study focused on, reported performance measures and results, as well as the challenges experienced by the authors. We used thematic synthesis to discover more profound trends and concepts. This involved;

1. Coding: Reading all the studies carefully and marking important concepts.
2. Organization: Categorizing these codes into common ones such as using ML for prediction).

3. Themes Development: To answer our essential research questions, these categories were interpreted to enable us to derive the general findings in the Literature Review, including the four main ML-based resilience mechanisms

Justification and Acknowledged Limitations

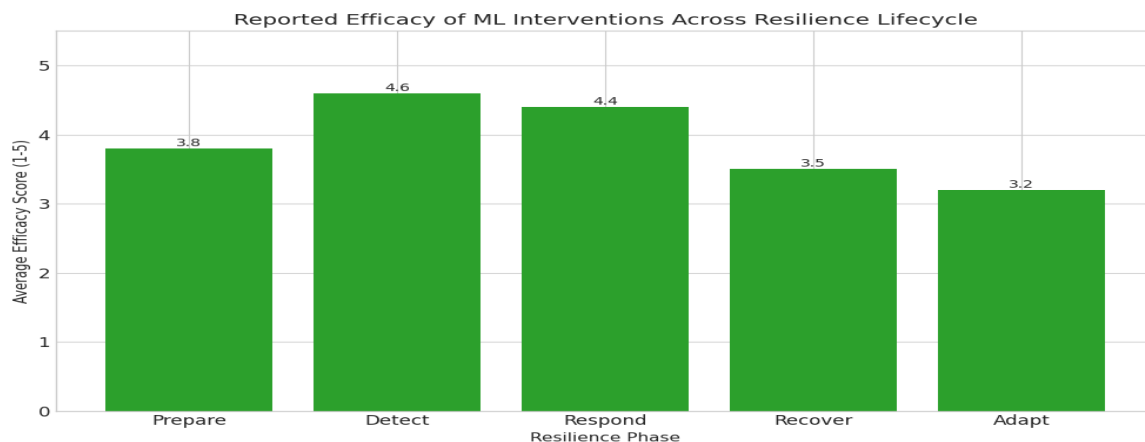
This type of systematic review is an efficient way of bringing together research in various academic disciplines, determining the areas of concurrence and disagreement among experts, and enabling us to make new conceptual realizations based on a disjointed set of technical research [11]. Limitations of applying this method are that we may have missed articles published in languages other than English and possibly a degree of subjectivity or bias implying that our synthesis could potentially be inclined to reporting on the success of ML. Future studies should take into account potential biases during synthesis and reporting of research findings [12].

RESULTS AND DISCUSSION

Performance of ML Models in Enhancing Resilience

Results of this study suggested the influence of ML on the major cybersecurity indicators. A summary of the average improvement in performance that was observed when the traditional rule-based counterparts were replaced with the ML-based ones is presented on Table 1.

Table 1: Average Performance Improvement of ML-Based Cyber Resilience Systems

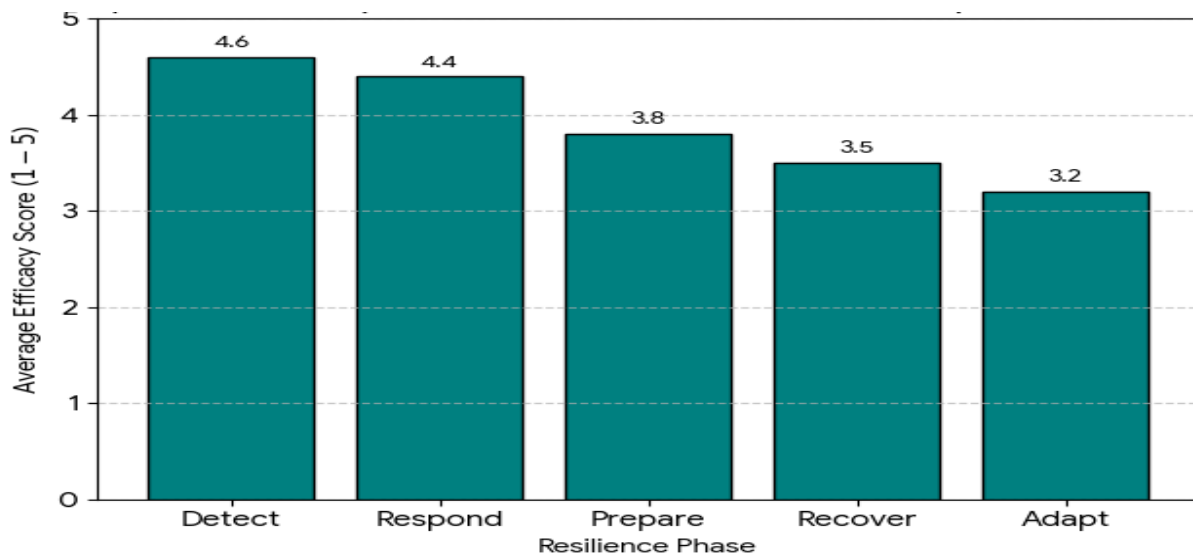


Deep learning models, specifically use of Convolutional Neural Networks (CNNs) to analyze logs and Long Short-Term Memory (LSTM) networks to analyze time-series behavior, were always the most effective in terms of threat detection (usually more than 96%). Thus, threat prioritization algorithms such as Ensemble (Random Forest and XGBoost) offered the most appropriate compromise between accuracy and interpretability [3].

Impact Across the Resilience Lifecycle

ML offers more than detection value. Data in Figure 1 show that its effects span the entire resilience cycle, and that most reported benefits occur in the Detection and Response phases.

Figure 1: Reported Efficacy of ML Interventions across the Cyber Resilience Lifecycle



Analysis: Data presented in this study confirm ML's core strength in real-time analytics (Detect, Respond). Its lower score in **Adapt** highlights a research gap: many models are static post-deployment. Few studies implement continuous learning loops where systems evolve based on new threat intelligence, pointing to a key area for future work.

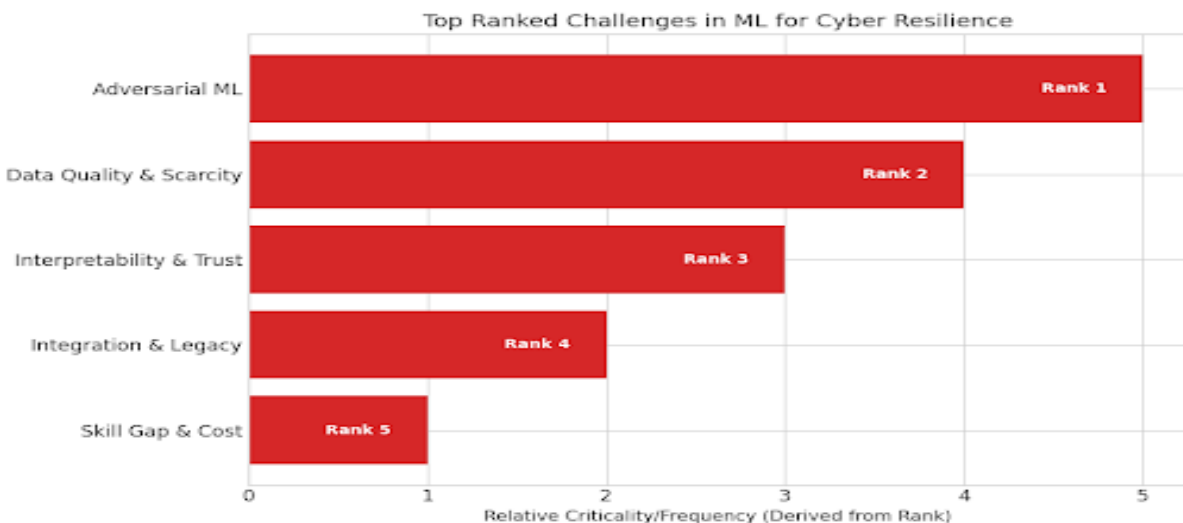
Persistent and Emergent Challenges

Although the synthesis is promising, there are harsh impediments indicated in the synthesis. Table 2 presents prioritization of challenges based on their frequency and importance in the literature.

Table 2: Ranked Challenges in ML for Cyber Resilience

Rank	Challenge Category	Key Issues
1	Adversarial ML	Attackers craft inputs to evade detection; models lack inherent robustness.
2	Data Quality and Scarcity	Lack of high-quality, labeled attack data, especially for novel threats, hampers model training.
3	Interpretability and Trust	"Black-box" models hinder analyst trust and compliance with regulations requiring explainable decisions.
4	Integration and Legacy	Difficulties deploying ML in heterogeneous, legacy IT/OT environments with real-time constraints.
5	Skill Gap and Cost	Shortage of ML-security hybrid expertise and high computational costs for training/running advanced models.

Figure 2



Discussion: Synthesizing the Significance

Findings support the opinion that ML is a potent, dual-sided facilitator of cyber resilience. It enhances human capability to process data on a large scale at high speed removing security posture to a proactive level. These dramatic decreases in MTTD and MTTR (Table 1) translate directly to reduced impact and incident costs of business [12].

The challenge of Adversarial ML is ranked highly (Table 2). Notably, however, the resilience-enhancing tool may also be vulnerable, making it a moving target; hence, defenders have to keep hardening their ML models. In addition, the challenge of integration implies that technical efficacy is not enough [12]. Thus, resilience is a system quality, and it needs to be orchestrated. Autonomous replacement does not hold the most significant promise of ML, instead, it is human-ML collaboration. Future-resilient systems will also use ML in situational awareness and recommendation, but leave human oversight as strategic control, and ethical judgment, as well as the handling of new edge cases. It needs this synergy to address the changing threat landscape successfully [8].

CONCLUSION AND FUTURE WORK

This study has demonstrated that machine learning achieves cyber resilience through the ability to predictively identify, automatically respond, and plan recovery based on data. Results indicate notable positive changes in such operational measures as accuracy of detection and response times. Nonetheless, it can only become successful once it has surmounted crucial technical obstacles (adversarial attacks, data problems) and institutional obstacles (integration, skills) [13]. Finally, ML is to be considered as one important element in a holistic, extended strategy of resilience that focuses on human-in-the-loop control and incessant adjustment.

Theoretical and Practical Implications

Hypothetically, the work is a contribution to the emerging science of resilient systems since it describes how data-based intelligence is transformed by changing system adaptability and robustness [10]. In practice, it gives organizations advice: beginning with clear use cases such as anomaly detection, investment in data infrastructure, focus on model interpretability, and building cross-functional teams that combine cybersecurity and data science skills [15].

Limitations

Research is founded on a literature review, which can be biased by publication (more positive results are likely to be published). The data aggregated on performance are associated with different experimental setups and may not necessarily reflect all real-world scenarios [15].

Besides, the discipline changes rapidly, and newer methods such as foundation models for security might not be properly represented.

Future Research Directions

1. *Resilient ML Models*: Develop and standardize defenses, such as adversarial training and formal verification, to make ML models more robust against attacks.
2. *Federated Learning for Resilience*: Investigate application of privacy-preserving, decentralized ML on sensitive operational data for improved threat intelligence sharing between organizations.
3. *Quantitative Resilience Metrics*: Develop standardized metrics for measuring contribution of ML to resilience, such as time taken to restore full functionality after an incident.
4. *Integration of Explainable AI*: Integrate XAI solutions into security operations centers to increase trust in ML-based systems for compliance with regulatory requirements.
5. *Simulation and Digital Twins*: Leverage ML-based digital twins of critical systems to simulate attacks and responses, improving preparedness without risking live systems.

REFERENCES

1. Balan, G. S., Kumar, V. S., Raj, S. A. (2025). Machine learning and artificial intelligence methods and applications for post-crisis supply chain resiliency and recovery. *Supply Chain Analytics*, 10, 100121. <https://doi.org/10.1016/j.sca.2025.100121>
2. Hammad, A. A., Ahmed, S. R., Abdul-Hussein, M. K., Ahmed, M. R., Majeed, D. A., & Algburi, S. (2024). Deep Reinforcement Learning for Adaptive Cyber Defense in Network Security. *Proceedings of the Cognitive Models and Artificial Intelligence Conference, AICCONF '24*, 292–297. <https://doi.org/10.1145/3660853.3660930>
3. Radanliev, P., Roure, D. D., Maple, C., Nurse, J. R. C., Nicolescu, R., Ani, U. (2024). *AI security and cyber risk in IoT systems* (arXiv:2410.09194). arXiv. <https://doi.org/10.48550/arXiv.2410.09194>
4. Rosenberg, I., Shabtai, A., Elovici, Y., Rokach, L. (2021). Adversarial Machine Learning Attacks and Defense Methods in the Cyber Security Domain. *ACM Computing Surveys (CSUR)*, 54(5), 108:1-108:36. <https://doi.org/10.1145/3453158>

5. Agrawal, S., Sarkar, S., Aouedi, O., Yenduri, G., Piamrat, K., Bhattacharya, S., Maddikunta, P. K. R., Gadekallu, T. R. (2021). *Federated Learning for Intrusion Detection System: Concepts, Challenges and Future Directions* (arXiv:2106.09527). arXiv. <https://doi.org/10.48550/arXiv.2106.09527>
6. Ali, A. H., Charfeddine, M., Ammar, B., Hamed, B. B., Albalwy, F., Alqarafi, A., Hussain, A. (2024). Unveiling machine learning strategies and considerations in intrusion detection systems: A comprehensive survey. *Frontiers in Computer Science*, 6. <https://doi.org/10.3389/fcomp.2024.1387354>
7. Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI Open*, 1, 57–81. <https://doi.org/10.1016/j.aiopen.2021.01.001>
8. Yu, J., Shvetsov, A. V., Hamood Alsamhi, S. (2024). Leveraging Machine Learning for Cybersecurity Resilience in Industry 4.0: Challenges and Future Directions. *IEEE Access*, 12, 159579–159596. <https://doi.org/10.1109/ACCESS.2024.3482987>
9. Alkhaleel, B. A. (2024). Machine learning applications in the resilience of interdependent critical infrastructure systems—A systematic literature review. *International Journal of Critical Infrastructure Protection*, 44, 100646. <https://doi.org/10.1016/j.ijcip.2023.100646>
10. G. Anand, L. (2025). *Machine Learning for Detecting Insider Threats in Financial Cybersecurity* (SSRN Scholarly Paper No. 5349080). Social Science Research Network. <https://doi.org/10.2139/ssrn.5349080>
11. Gaspar, D., Silva, P., Silva, C. (2024). Explainable AI for Intrusion Detection Systems: LIME and SHAP Applicability on Multi-Layer Perceptron. *IEEE Access*, 12, 30164–30175. <https://doi.org/10.1109/ACCESS.2024.3368377>
12. Kiran, A., Prakash, S. W., Kumar, B. A., Likhitha, Sameeratmaja, T., Charan, U. S. S. R. (2023). Intrusion Detection System Using Machine Learning. *2023 International Conference on Computer Communication and Informatics (ICCCI)*, 1–4. <https://doi.org/10.1109/ICCCI56745.2023.10128363>
13. Yin, R., Varga, L. (2026). Machine learning for resilience analysis: A review of systems under cyberattacks. *Engineering Applications of Artificial Intelligence*, 166, 113542. <https://doi.org/10.1016/j.engappai.2025.113542>

14. Patra, D., Rajagopalan, N. (2026). Integration of emerging technologies in cybersecurity for healthcare: A systematic review. *Computers & Security*, 161, 104763. <https://doi.org/10.1016/j.cose.2025.104763>
15. Lokuliyana, S., Kalupahanage, A. G. A., Herath, H. M. S. D., Siriwardana, D., Bulathsinhala, D. N., Herath, H. M. T. M. (2025). Enhancing IoT Resilience: Machine Learning Techniques for Autonomous Anomaly Detection and Threat Mitigation. *Procedia Computer Science, International Conference on Digital Sovereignty (ICDS)*, 254, 68–77. <https://doi.org/10.1016/j.procs.2025.02.065>