RESEARCH ARTICLE OPEN ACCESS

A Review: AI Enabled Threat Detection: Leveraging Artificial Intelligence for Advanced Security and Cyber Threat Mitigation

Ms. Sankesha. A. Rajguru*, Prof. Sayema. A. Khusro*
*(Cyber Security, P.E.S.College of Enginearing, and Chh.Sambhajinagar
Email: sankesha9899@gmail.com)
** (Computer Department, P.E.S.College of Enginearing, and Chh.Sambhajinagar, and Chh.Sambhajinagar
Email: sayema@pescoe.ac.in)

--************************----------------------------------

Abstract:

The increasing digitalization of modern infrastructures—ranging from Industry 5.0 and Internet of Things (IoT) ecosystems to 5G networks and autonomous systems—has led to a dramatic rise in sophisticated cyber threats that traditional security mechanisms struggle to detect. Artificial Intelligence (AI), particularly machine learning (ML) and deep learning (DL) techniques, has emerged as a powerful tool for enhancing threat detection, enabling the identification of network intrusions, malware, adversarial attacks, and zero-day vulnerabilities with greater accuracy and automation. This review provides a comprehensive analysis of recent AI-driven threat detection approaches, highlighting cutting-edge advancements such as transformer-based architectures, generative adversarial networks (GANs), federated learning, blockchain-enabled detection, and interpretable AI models. The papers examined span diverse application domains, demonstrating AI's adaptability to different cybersecurity challenges. Despite these advances, persistent issues remain, including imbalanced datasets, lack of explainability, adversarial manipulation of AI models, high computational overhead, and the difficulty of achieving real-time, scalable solutions. This review synthesizes current findings, identifies open research gaps, and outlines key directions for developing robust, trustworthy, and future-ready AI-enabled cybersecurity systems. The analysis underscores that while substantial progress has been made, continued interdisciplinary collaboration and innovation are essential for fully leveraging AI's potential in securing the rapidly evolving digital landscape.

Keywords — **Put your keywords here, keywords are separated by comma.**

--************************----------------------------------

I. INTRODUCTION

The rapid expansion of digital ecosystems, fueled by advancements in cloud computing, Industry 5.0, Internet of Things (IoT), and intelligent cyber-physical systems, has significantly increased exposure to sophisticated cyber threats. Modern attackers employ advanced strategies such as zero-day exploits, adversarial machine learning, ransomware automation, and persistent infiltration techniques that can easily bypass traditional signature-based security mechanisms. As a result, cybersecurity has transitioned from a reactive discipline to a domain requiring proactive, adaptive, and intelligent threat detection strategies. In this landscape, Artificial Intelligence (AI) has emerged as a transformative technology capable of analyzing large-scale, heterogeneous data and autonomously identifying hidden or evolving attack patterns.

AI-driven threat detection leverages machine learning (ML), deep learning (DL), and data-driven behavior analytics to detect anomalies, classify malicious activities, and predict potential intrusions with high precision. Unlike conventional systems that rely heavily on predefined rules or known attack signatures, AI-enabled solutions learn from data, identify patterns of normal and abnormal behavior, and adapt continuously to new cyberattack strategies. The integration of advanced models such as Generative Adversarial Networks (GANs), transformers, graph neural networks, federated learning, and reinforcement learning has further expanded the capabilities of modern intrusion detection systems (IDS) and malware classification frameworks.

Moreover, emerging domains including 5G communication networks, autonomous vehicles, smart healthcare, blockchain-integrated architectures, and microgrid infrastructures demand intelligent and scalable security solutions. The studies reviewed in this paper demonstrate AI's versatility across these sectors, highlighting its potential to enhance threat visibility, reduce false alarms, and detect complex multi-stage cyberattacks in real time. At the same time, critical challenges persist, including data imbalance, adversarial manipulation of AI models, a lack of explainability in decision-making, privacy preservation issues, and the computational cost of real-time deployments.

Given these advancements and challenges, this review provides a comprehensive examination of recent AI-enabled threat detection techniques, focusing on their methodologies, applications, strengths, and limitations. The objective is to analyze how state-of-the-art AI models contribute to improving cybersecurity resilience while identifying existing research gaps that must be addressed to build trustworthy and future-ready defense systems.

With the rapid digitalization of industries, IoT ecosystems, 5G networks, autonomous systems, and cloud infrastructures, cyber threats have become increasingly sophisticated and difficult to detect using conventional security mechanisms. Traditional signature-based and rule-based intrusion detection systems fail to identify new, evolving, and stealthy cyberattacks such as zero-day exploits, advanced persistent threats (APTs), insider threats, and adversarial attacks on AI models. Although Artificial Intelligence (AI) provides promising capabilities for automated threat detection, its deployment introduces new challenges, including model explainability, robustness, data imbalance, resource consumption, and vulnerability to adversarial manipulation. Therefore, there is a critical need to systematically examine AI-enabled threat detection approaches, understand their strengths and limitations across various application domains, and identify gaps that hinder their practical deployment in real-world cybersecurity environments.

II. Literature Review

A. Artificial Intelligence (AI) has become a critical enabler of next-generation cybersecurity systems due to its ability to learn complex patterns, adapt to evolving threats, and analyze massive volumes of network and system data. Traditional intrusion detection systems rely heavily on signature-based rules, which are ineffective against zero-day attacks, obfuscated malware, and advanced persistent threats (APTs). AI-driven approaches, especially machine learning (ML) and deep learning (DL), are capable of identifying subtle deviations in behavior, enabling the early detection of sophisticated cyber threats. Recent studies have expanded the role of AI beyond anomaly detection to include adversarial defense, blockchain-enabled security, and intelligent malware analysis.

B. Network Threat Detection Using AI

Wang et al. [1] propose an AI-powered network threat detection system capable of analyzing traffic data to detect anomalies with high accuracy. Their system leverages ML models and feature engineering techniques to identify malicious patterns in network flows. While the system achieves strong classification accuracy, its primary limitation is the lack of real-time evaluation, which is critical for detecting high-speed attacks in enterprise networks.

C. Federated Learning, Blockchain, and Distributed Security

Federated learning and blockchain have emerged as promising techniques for secure and decentralized cybersecurity solutions. Aliyu et al. [3] combine blockchain with a federated forest to detect adversarial examples in in-vehicle network intrusion detection systems. Their method enhances security and robustness by ensuring tamper-proof data exchange across distributed nodes. Nonetheless, blockchain's computational overhead remains a concern for real-time implementations.

Taher et al. [9] focus on securing DC microgrids from false data injection attacks (FDIA) using nonlinear autoregressive observer models. Their system provides real-time detection and mitigation capabilities, which are vital for critical infrastructure protection. However, scalability for large interconnected grids remains an open challenge.

D. Explainable AI (XAI) for Intrusion Detection

Explainability has become a key requirement for the deployment of AI in mission-critical cybersecurity environments. Neupane et al. [4] provide a comprehensive survey of explainable intrusion detection systems (X-IDS), highlighting the strengths and limitations of popular XAI methods such as SHAP, LIME, and rule-based explanations. Their study emphasizes that despite the growing number of AI-based IDS solutions, the lack of interpretability hinders trust and operational adoption. This gap underscores the need for explainable deep learning models capable of providing human-understandable insights without sacrificing detection accuracy.

E. GANs and Deep Learning Approaches

Generative models, particularly GANs, have shown promise in improving intrusion detection systems by generating realistic synthetic samples. Park et al. [5] demonstrate how GAN-generated data can be used to enhance anomaly detection in IoT environments. Their enhanced IDS architecture improves performance on under-represented attack classes and reduces false positives. However, challenges remain regarding GAN training instability and overfitting.

Torres et al. [6] explore malware detection using feature engineering and behavioral analysis. Their approach highlights the importance of dynamic analysis in identifying unknown malware families. Although the method achieves high accuracy, its reliance on handcrafted features limits adaptability to new malware variants.

F. AI for Healthcare, IoT, and Emerging Cyber-Physical Systems

Industry 5.0 and smart healthcare environments demand robust cybersecurity solutions due to increasing data volume, connected devices, and life-critical applications. Wazid et al. [7] develop an ensemble-based IDS combining random forest, SVM, and boosting algorithms. Their solution achieves strong performance in healthcare IoT networks; however, it requires high computational resources.

Allafi and Alzahrani [8] propose an artificial orca optimization algorithm integrated with ensemble learning for IoT threat detection. Their bio-inspired technique enhances accuracy and convergence speed, although it lacks cross-domain validation.

In enterprise networks, insider threats remain one of the most challenging attack vectors. Al-Shehari et al. [10] utilize a density-based local outlier factor (LOF) algorithm to detect anomalous insider activities in imbalanced datasets. The approach effectively identifies subtle malicious behaviors but struggles with high-dimensional data.

Finally, Soliman et al. [11] introduce RANK, an end-to-end AI-assisted architecture designed to detect persistent attacks in enterprise infrastructures. Their system integrates analytics, detection, and response mechanisms. While comprehensive, the model requires large high-quality datasets, which are often unavailable due to privacy concerns.

III. PAGE STYLE

Ref No.	Authors & Year	Focus Area	Techniques / Models Used	Dataset / Environmen t	Key Contributions	Limitations
[1]	Wang et al., 2022	AI-powered network threat detection	ML-based classification, feature analysis	Network traffic datasets	Developed scalable AI-powered NIDS with high	Limited evaluation on real-time attacks

					detection accuracy	
[2]	Gao et al., 2022	Trojan detection in DNN models	Multi-domain Trojan detection, adversarial analysis	Deep neural networks	Effective detection of Trojan attacks across domains	High computational complexity
[3]	Aliyu et al., 2022	Adversarial example detection using blockchain federated forest	Statistical detection, federated learning, blockchain	In-vehicle network IDS	Secure, decentralized IDS resistant to adversarial attacks	High overhead in blockchain operations
[4]	Neupane et al., 2022	Explainable Intrusion Detection Systems (X-IDS)	XAI methods (SHAP, LIME, rule extraction)	Multiple IDS datasets	Comprehensive survey on X-IDS challenges & opportunities	Lack of unified benchmarks for XAI
[5]	Park et al., 2023	AI-based NIDS using GANs	Generative Adversarial Networks + Deep Learning	IoT network datasets	Improved anomaly detection using GAN-generated samples	Model instability due to GAN training
[6]	Torres et al., 2023	Malware detection	Feature engineering, behavior analysis	Dynamic malware samples	High accuracy via engineered behavioral features	Limited generalization to new malware families
[7]	Wazid et al., 2024	IDS for Industry 5.0 healthcare	Ensemble learning (RF, XGBoost, SVM)	Healthcare IoT datasets	Robust IDS for Industry 5.0 with improved classification	Higher training time due to ensemble models
[8]	Allafi & Alzahrani, 2024	IoT security enhancement	Artificial Orca Algorithm + Ensemble Learning	IoT sensor data	Optimized detection accuracy using bio-inspired AI	Focused mainly on IoT; lacks cross-domain validation
[9]	Taher et al., 2024	FDIA (False Data Injection Attack)	Nonlinear autoregressive observer models	DC microgrid system	Real-time FDIA detection and mitigation	Computationally heavy for large-scale microgrids

		detection in microgrids				
[10]	Al-Shehari et al., 2024	Insider threat detection in imbalanced data	Density-Based Local Outlier Factor (LOF)	Enterprise security logs	Enhanced insider threat detection in imbalanced datasets	Limited performance on high-dimensional data
[11]	Soliman et al., 2024	Persistent attack detection	RANK: AI-assisted end-to-end architecture	Enterprise networks	End-to-end persistent attack detection pipeline	Requires large high-quality dataset for training

Ref. No.	**Methodology Used**	**AI Technique / Model**	**Reported Accuracy / Performance**
[1]	AI-based network traffic analysis	ML classifiers with feature engineering	~94–97% detection accuracy
[2]	Multi-domain Trojan detection	Deep Neural Networks (DNN)	>96% detection accuracy
[3]	Blockchain-enabled federated IDS	Federated Forest + statistical detection	~93–95% accuracy with high robustness
[4]	Explainable intrusion detection (survey)	XAI methods (SHAP, LIME, rules)	Accuracy varies; focus on explainability over performance
[5]	GAN-enhanced IDS	GAN + Deep Learning	~95–98% detection accuracy
[6]	Behavior-based malware analysis	ML with engineered behavioral features	~96–97% accuracy
[7]	Ensemble IDS for Industry 5.0 healthcare	RF + SVM + Boosting	~97–99% accuracy
[8]	Bio-inspired optimization-based IDS	Artificial Orca Algorithm + Ensemble ML	~96–98% accuracy
[9]	Control-theoretic detection	NARX observer-based AI model	>95% detection rate
[10]	Statistical anomaly detection	Density-Based Local Outlier Factor (LOF)	~92–94% accuracy

[11]	End-to-end AI-assisted architecture	Deep Learning + Analytics (RANK)	~96–98% detection accuracy

Gap Analysis

Despite notable advancements in AI-enabled threat detection, the literature reveals several critical gaps that limit the effectiveness, scalability, and real-world deployment of existing cybersecurity solutions. This gap analysis synthesizes findings from recent studies and highlights unresolved challenges that require further investigation.

1. Explainability vs. Detection Accuracy Gap

Most AI-based intrusion detection and malware classification systems prioritize high detection accuracy using complex deep learning architectures such as GANs, transformers, and ensemble models. However, these models often lack interpretability, making it difficult for security analysts to understand or trust their decisions. Although explainable AI (XAI) approaches have been proposed, they are typically applied as post-hoc solutions and are not tightly integrated into detection pipelines. This creates a gap between high-performance AI models and the need for transparent, auditable security systems—especially in critical domains such as healthcare, autonomous vehicles, and industrial control systems.

2. Real-Time Processing and Scalability Gap

Many reviewed studies demonstrate strong performance on offline datasets but fail to address real-time deployment challenges. High computational complexity, large feature spaces, and deep neural architectures often hinder scalability in high-speed networks and large-scale IoT environments. The lack of lightweight and low-latency AI models suitable for edge and fog computing highlights a significant gap between experimental results and operational feasibility.

3. Dataset Quality and Generalization Gap

A persistent limitation across studies is the reliance on benchmark datasets that are outdated, imbalanced, or domain-specific. Models trained on such datasets often exhibit reduced generalization when exposed to real-world traffic patterns or emerging attack types. Furthermore, limited availability of labeled data for rare or zero-day attacks restricts supervised learning performance, indicating a need for more robust unsupervised, self-supervised, and semi-supervised learning approaches.

4. Adversarial Robustness Gap

While AI enhances threat detection, it also introduces new attack surfaces. Several studies acknowledge vulnerabilities to adversarial attacks, data poisoning, and model evasion techniques, yet few propose comprehensive defense mechanisms. Existing adversarial detection methods are often evaluated in isolation and lack adaptability against evolving attack strategies, revealing a gap in resilient and self-defensive AI models.

5. Privacy-Preserving Security Gap

Although federated learning and blockchain-based approaches have been introduced to address privacy concerns, they remain limited by communication overhead, synchronization latency, and susceptibility to poisoning attacks. Most implementations focus on architectural feasibility rather than end-to-end privacy guarantees, highlighting a gap in scalable, secure, and privacy-preserving AI frameworks suitable for real-world cybersecurity applications.

IV. CONCLUSIONS

This review has presented a comprehensive analysis of recent advancements in AI-enabled threat detection, emphasizing the transformative role of artificial intelligence in modern cybersecurity frameworks. The surveyed studies demonstrate that machine learning, deep learning, and hybrid AI models significantly enhance the detection of complex cyber threats, including zero-day attacks, adversarial intrusions, insider threats, and advanced persistent threats across diverse application domains such as Industry 5.0, Internet of Things (IoT), healthcare systems, microgrids, and enterprise networks. Advanced techniques including generative adversarial networks, transformer-based architectures, federated learning, and blockchain-integrated security solutions have shown notable improvements in detection accuracy, adaptability, and robustness compared to traditional security mechanisms.

Despite these advancements, the review identifies several critical challenges that limit real-world deployment of AI-driven cybersecurity systems. Key issues include the lack of explainability in deep learning models, vulnerability to adversarial manipulation, dependence on imbalanced or outdated datasets, high computational overhead, and limited scalability in real-time environments. While explainable AI and privacy-preserving learning frameworks offer promising directions, their integration into operational threat detection systems remains at an early stage.

REFERENCES

[1] B.-X. Wang, J.-L. Chen, and C.-L. Yu, ‘‘An AI-powered network threat detection system,’’ IEEE Access, vol. 10, pp. 54029–54037, 2022.

[2] Y. Gao, Y. Kim, B. G. Doan, Z. Zhang, G. Zhang, S. Nepal, D. C. Ranasinghe, and H. Kim, ‘‘Design and evaluation of a multi-domain trojan detection method on deep neural networks,’’ IEEE Trans. Depend. Secure Comput., vol. 19, no. 4, pp. 2349–2364, Jul. 2022.

[3] I. Aliyu, S. Van Engelenburg, M. B. Mu’Azu, J. Kim, and C. G. Lim, ‘‘Statistical detection of adversarial examples in blockchain-based federated forest in-vehicle network intrusion detection systems,’’ IEEE Access, vol. 10, pp. 109366–109384, 2022.

[4] S. Neupane, J. Ables, W. Anderson, S. Mittal, S. Rahimi, I. Banicescu, and M. Seale, ‘‘Explainable intrusion detection systems (X-IDS): A survey of current methods, challenges, and opportunities,’’ IEEE Access, vol. 10, pp. 112392–112415, 2022.

[5] C. Park, J. Lee, Y. Kim, J.-G. Park, H. Kim, and D. Hong, ‘‘An enhanced AI-based network intrusion detection system using generative adversarial networks,’’ IEEE Internet Things J., vol. 10, no. 3, pp. 2330–2345, Feb. 2023.

[6] M. Torres, R. Álvarez, and M. Cazorla, ‘‘A malware detection approach based on feature engineering and behavior analysis,’’ IEEE Access, vol. 11, pp. 105355–105367, 2023.

[7] M. Wazid, J. Singh, A. K. Das, and J. J. P. C. Rodrigues, ‘‘An ensemble-based machine learning-envisioned intrusion detection in industry 5.0-driven healthcare applications,’’ IEEE Trans. Consum. Electron., vol. 70, no. 1, pp. 1903–1912, Feb. 2024.

[8] R. Allafi and I. R. Alzahrani, ‘‘Enhancing cybersecurity in the Internet of Things environment using artificial orca algorithm and ensemble learning model,’’ IEEE Access, vol. 12, pp. 63282–63291, 2024.

[9] M. A. Taher, M. Behnamfar, A. I. Sarwat, and M. Tariq, ‘‘False data injection attack detection and mitigation using nonlinear autoregressive exogenous input-based observers in distributed control for DC microgrid,’’ IEEE Open J. Ind. Electron. Soc., vol. 5, pp. 441–457, 2024.

[10] T. A. Al-Shehari, D. Rosaci, M. Al-Razgan, T. Alfakih, M. Kadrie, H. Afzal, and R. Nawaz, ‘‘Enhancing insider threat detection in imbalanced cybersecurity settings using the density-based local outlier factor algorithm,’’ IEEE Access, vol. 12, pp. 34820–34834, 2024.

[11] H. M. Soliman, D. Sovilj, G. Salmon, M. Rao, and N. Mayya, ‘‘RANK: AI-assisted end-to-end architecture for detecting persistent attacks in enterprise networks,’’ IEEE Trans. Depend. Secure Comput., vol. 21, no. 4, pp. 3834–3850, Jul. 2024.