Journal of Science Research International (JSRI)
ISSN: 2456-6365 (Online)

OPEN ACCESS

AI-BASED THREAT DETECTION IN CYBERSECURITY: OPPORTUNITIES AND CHALLENGES

Simran Shinde*, Pal Adnan Umar Saheb and Huzaif Bagwan

Department of Computer Science,

Pillai College of Arts, Commerce and Science (Empowered Autonomous), New Panvel

**Corresponding author E-mail: simranshinde@mes.ac.in*

Received: 14 January 2026	Revised: 22 February 2026	Accepted: 01 April 2026	Published: 16 April 2026

DOI: **https://doi.org/10.5281/zenodo.19613268**

Abstract:

The rapid proliferation of digital infrastructure has made cybersecurity one of the most critical domains in modern computing. As cyber threats grow in frequency, sophistication, and scale, traditional rule-based and signature-driven security systems are increasingly inadequate. This paper investigates the dual role of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) in modern cybersecurity: as a transformative force strengthening defensive capabilities, and as a source of novel risks introducing new attack vectors. The research examines AI's contribution to Intrusion Detection Systems (IDS), malware identification, network traffic analysis, and automated incident response, while exploring challenges including adversarial machine learning, false positive rates, data privacy concerns, and computational overhead. Through structured literature review, comparative analysis, and case-based evaluation, a comprehensive understanding of responsible AI deployment in cybersecurity is provided.

Keywords: *Artificial Intelligence, Cybersecurity, Machine Learning, Threat Detection, Intrusion Detection Systems, Network Security, Adversarial Machine Learning,*

1. Introduction

The digital revolution has transformed virtually every sector of modern society, expanding the attack surface available to malicious actors. Cybercrime is expected to inflict damages exceeding $10.5 trillion annually by 2025. Traditional signature-based Intrusion Detection Systems (IDS), firewall rules, and static access control policies are fundamentally reactive — effective only against known threats. As threat actors deploy zero-day exploits, polymorphic malware, and Advanced Persistent Threats (APTs), these conventional defenses become increasingly ineffective.

Artificial Intelligence offers a paradigm shift. AI systems learn from vast datasets of network behavior, log files, and historical attack signatures to identify subtle anomalies invisible to human analysts or rule-based engines. Unlike signature-based systems, ML models generalize from training data to detect novel attack patterns. Neural

networks, ensemble learning algorithms, and NLP models are deployed across the threat detection pipeline for tasks ranging from behavioral analysis and traffic classification to automated threat response and cyber threat intelligence generation.

1.1 Research Objectives

This paper aims to: (i) investigate how AI technologies improve detection, prevention, and response to cyber threats; (ii) evaluate risks and limitations of AI-driven security systems; (iii) compare AI-based solutions against traditional cybersecurity methodologies; and (iv) identify best practices for responsible AI integration in organizational cybersecurity frameworks.

2. Literature review

2.1 Machine learning in intrusion detection

Liao *et al.* (2013) established that signature-based systems exhibit severe limitations in detecting novel attacks. Tsai *et al.* (2009) demonstrated that Decision Trees and Random Forests outperformed other models in network intrusion classification on the KDD Cup 1999 dataset. Buczak and Guven (2016), surveying over 40 studies, identified Random Forests, Artificial Neural Networks, and Naive Bayes as most effective, noting that no single algorithm consistently outperformed others across all attack categories — reinforcing the case for ensemble and hybrid approaches.

2.2 Deep learning and adversarial ML

Kim *et al.* (2016) proposed an RNN-based IDS achieving 98.7% detection accuracy on sequential network traffic data. Nataraj *et al.* (2011) pioneered malware binary visualization using CNNs, achieving >95% classification accuracy — inherently resistant to simple obfuscation. Goodfellow *et al.* (2014) introduced adversarial examples, showing neural networks can be reliably deceived by imperceptible input perturbations. Papernot *et al.* (2016) extended this to IDS, demonstrating evasion rates exceeding 70% against neural network classifiers via gradient-based adversarial crafting.

2.3 SOAR and automated response

Gartner research highlights that AI-powered SOAR platforms reduce Mean Time to Respond (MTTR) by up to 90%. Settanni *et al.* (2017) demonstrated that automated playbook execution handles up to 80% of routine security operations without human intervention, allowing analysts to focus on higher-order threat hunting and strategic defense.

3. Methodology and AI techniques

This research employs a qualitative and comparative methodology grounded in systematic literature review, case study analysis, and performance benchmarking. The analysis framework encompasses: technical efficacy (detection accuracy and response speed), operational impact (automation and scalability), risk profile (adversarial robustness), and ethical implications (transparency and accountability).

3.1 ML algorithms

Random Forest (RF) constructs multiple decision trees and outputs the mode of their predictions. In IoT security, RF achieved 95.01% accuracy with 99.23% precision detecting Mirai botnet malware; for general network anomaly detection, hyperparameter-optimized RF reached 97.8% accuracy with F1-scores of 98%. SVMs find optimal hyperplanes separating classes in high-dimensional feature space, demonstrating strong performance in

binary classification with good generalization. XGBoost offers exceptional predictive performance with built-in feature importance metrics aiding interpretability.

3.2 Anomaly detection & behavioral analysis

Anomaly detection identifies deviations from baselines without requiring labeled attack samples. Gaussian Mixture Models, Isolation Forests, and unsupervised Autoencoders establish probabilistic models of normal behavior — particularly valuable for zero-day attacks and insider threats. User and Entity Behavior Analytics (UEBA) builds dynamic behavioral profiles for users and devices; deviations (e.g., accessing sensitive systems at unusual hours) trigger risk scores for escalation. This approach is highly effective for detecting APT campaigns.

3.3 Deep learning architectures

CNNs extract spatial features from malware binary images and network flow matrices. LSTMs analyze log files, packet sequences, and user activity timelines where temporal dependencies carry predictive significance. Transformer-based models, adapted from NLP, perform cybersecurity tasks including threat intelligence extraction and vulnerability report classification. Graph Neural Networks (GNNs) applied to communication graphs detect lateral movement and coordinated multi-host attacks invisible when analyzing individual flows.

4. AI-based threat detection systems

4.1 Intrusion Detection Systems (IDS)

AI-enhanced IDS represent the most widely deployed ML application in cybersecurity. AI-based IDS learn behavioral models from network data to detect both known and previously unseen attacks. Hybrid architectures combining signature detection with ML-based anomaly detection represent current best practice — achieving higher detection rates and lower false positive rates than either approach alone. Ensemble IDS using stacked classifiers achieve detection accuracies exceeding 99% on NSL-KDD and CICIDS2017 benchmarks.

4.2 Malware detection

Malware detection has moved beyond hash-based blacklisting to dynamic behavioral analysis and deep learning classification. Static analysis uses ML to extract features from executable file structures and opcode sequences. Dynamic analysis executes suspicious files in sandboxed environments and applies ML to behavioral traces including API call sequences, network connections, and registry changes. CNNs applied to binary visualization achieved >95% classification accuracy across 25 malware families. GANs are being explored for malware variant generation to augment training datasets.

4.3 Network traffic analysis and automated response

AI-based network traffic analysis enables continuous monitoring at machine speed. Flow-based analysis using packet size distribution, inter-arrival times, and protocol flags allows ML classifiers to distinguish normal traffic from DDoS, port scanning, data exfiltration, and botnet C2 communications. Upon threat confirmation, SOAR platforms execute playbooks orchestrating responses across multiple security tools: isolating endpoints, blocking malicious IPs, revoking credentials, and capturing forensic images. Reinforcement learning enables adaptive playbook optimization, refining response strategies from past incidents.

5. Comparative analysis

Table 1: Traditional vs. AI-Based Cybersecurity Systems

Feature	Traditional Systems	AI-Based Systems
Core Approach	Rule-based, signature-driven	Behavioral learning, pattern recognition
Threat Coverage	Known threats only	Known and unknown (zero-day) threats
Adaptability	Low — manual rule updates required	High — continuous model retraining
False Positive Rate	High — broad rule sensitivity	Lower with well-trained models
Zero-Day Detection	Poor — no matching signatures	Strong — anomaly-based generalization
Automation Level	Low — primarily manual analysis	High — up to 80% task automation
Transparency	High — rules are human-readable	Moderate — black-box concerns
Scalability	Limited by rule complexity	Scales with data and compute resources

Table 2: AI Algorithm Performance in Cybersecurity Applications

Algorithm	Accuracy	Key Strengths	Primary Limitations
Random Forest	95%–99.6%	High precision, handles large datasets	Computationally intensive
Decision Tree	Up to 99.39%	Highly interpretable, fast training	Prone to overfitting on noisy data
Deep Neural Network	AUC ~98%+	Complex pattern recognition, auto feature learning	High resource demand, opaque
SVM	92%–98%	Effective in high dimensions, good generalization	Slow on very large datasets
LSTM/RNN	96%–99%	Excellent for sequential/temporal data	Long training times, vanishing gradient
Autoencoder	90%–97%	Effective unsupervised anomaly detection	Tuning reconstruction threshold is challenging

Traditional systems operate with high specificity but low sensitivity for novel threats. AI-based systems trade some computational efficiency for dramatically improved coverage of unknown attack vectors. Ensemble methods and deep learning architectures consistently achieve the highest detection accuracy, albeit at greater computational cost and reduced interpretability.

6. Opportunities of AI in cybersecurity

- Real-time monitoring: AI processes terabytes of daily log data in real time, correlating events across thousands of endpoints. This reduces average dwell time from 200+ days (traditional detection) to near real-time threat surfacing.
- Predictive threat detection: ML models analyzing historical attack patterns, threat intelligence feeds, and vulnerability databases forecast likely attack vectors, enabling proactive defense before exploitation occurs.
- Automated incident response: AI-powered SOAR platforms execute containment and remediation within milliseconds of threat confirmation, reducing MTTR by over 90% and ensuring consistent policy application.

- Reduced analyst burnout: SOC analysts receive upwards of 10,000 alerts per day. AI-based triage and prioritization systems filter, correlate, and score alerts, allowing analysts to focus on genuine threats and improving both job satisfaction and response quality.

7. Challenges and limitations

- Adversarial machine learning: Adversarially crafted malware samples or network packets modified to evade AI classifiers while retaining malicious functionality achieve evasion rates exceeding 70%. Data poisoning attacks inject maliciously crafted samples into training pipelines to corrupt learned representations.
- High false positive rates: AI-based systems continue to struggle with false positives in heterogeneous network environments. These impose direct operational costs, consume analyst time, and erode trust in automated systems — requiring careful calibration of classification thresholds.
- Data privacy and regulatory compliance: AI security systems require large volumes of network traffic and user activity data, creating challenges under GDPR and similar regulations. Organizations must implement robust data governance frameworks balancing security with privacy obligations.
- Lack of explainability: Deep neural networks and gradient boosting ensembles are inherently opaque (the "black box problem"). In high-stakes security environments, accountability demands explainability. XAI methods such as SHAP values, LIME, and attention mechanisms partially address this, but fully satisfactory explainability remains an open research problem.
- Computational requirements: Real-time deep learning inference requires GPU/AI chip acceleration, representing significant capital expenditure. For SMEs, cloud-based security services and managed AI security platforms provide a partial solution.

Future scope

- Autonomous cyber defense: Reinforcement learning provides the framework for AI systems that independently investigate, contain, and remediate incidents. DARPA's Cyber Grand Challenge has demonstrated the feasibility of autonomous vulnerability discovery and patching, though governance challenges must be resolved before production deployment.
- Federated learning: Organizations train local models and share only model updates — not raw data — with a central coordinator. This enables cross-industry threat intelligence sharing (banks, hospitals, utilities, government agencies) while preserving the confidentiality of each organization's network telemetry.
- Quantum computing integration: Quantum computers threaten widely-used public-key cryptography (Shor's algorithm), necessitating post-quantum cryptographic migrations. Conversely, quantum ML algorithms may dramatically accelerate AI security model training and inference, enabling real-time analysis of previously intractable data volumes.
- Large Language Models (LLMs): LLMs such as GPT-4 can automate analysis of threat intelligence reports, generate incident summaries, assist in policy development, and support security awareness training. However, they also introduce new attack surfaces including prompt injection and automated generation of sophisticated phishing content.

Conclusion

This paper presents a comprehensive analysis of AI-based threat detection in cybersecurity. The evidence demonstrates that AI offers transformative capabilities addressing the fundamental limitations of traditional security systems. ML algorithms, particularly Random Forests and deep neural networks, achieve detection accuracies of 95–99% across diverse threat categories. AI-powered SOAR platforms automate up to 80% of routine security operations and reduce MTTR by over 90%. Predictive analytics enables a shift from reactive to proactive defense, potentially eliminating the long dwell times that allow threat actors to persist undetected.

At the same time, adversarial ML attacks can undermine AI detector reliability. High false positive rates impose operational costs. Model opacity creates accountability challenges. Data privacy regulations constrain collection practices. Computational demands create barriers for smaller organizations. The path forward requires integrating AI's analytical capabilities with human expertise, robust governance frameworks, and continuous adversarial hardening. Organizations should implement Explainable AI (XAI) frameworks, maintain human-in-the-loop validation for critical decisions, and invest in adversarial training. As the threat landscape evolves at unprecedented speed, AI is not merely an enhancement but a fundamental necessity for maintaining viable defensive posture. Organizations and policymakers who invest in understanding, deploying, and governing AI-based cybersecurity systems today will be best positioned to navigate the threats of tomorrow.

References

1. Liao, H. J., Lin, C. H. R., Lin, Y. C., & Tung, K. Y. (2013). Intrusion detection system: A comprehensive review. *Journal of Network and Computer Applications, 36*(1), 16–24.
2. Tsai, C. F., Hsu, Y. F., Lin, C. Y., & Lin, W. Y. (2009). Intrusion detection by machine learning: A review. *Expert Systems with Applications, 36*(10), 11994–12000.
3. 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.
4. Kim, J., *et al.* (2016). Long short-term memory recurrent neural network classifier for intrusion detection. *IEEE PlatCon*, 1–5.
5. Tang, T. A., *et al.* (2016). Deep learning approach for network intrusion detection in SDN. *IEEE WINCOM*, 258–263.
6. Nataraj, L., *et al.* (2011). Malware images: Visualization and automatic classification. *ACM VizSec*, 1–7.
7. Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. *arXiv*. https://arxiv.org/abs/1412.6572
8. Papernot, N., *et al.* (2016). The limitations of deep learning in adversarial settings. *IEEE European Symposium on Security and Privacy (EuroS&P)*, 372–387.
9. Settanni, G., *et al.* (2017). Acquiring cyber threat intelligence through security information correlation. *IEEE CYBCONF*, 1–7.
10. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys, 41*(3), 1–58.
11. Mirsky, Y., *et al.* (2018). Kitsune: An ensemble of autoencoders for online network intrusion detection. *Network and Distributed System Security Symposium (NDSS)*.
12. Apruzzese, G., *et al.* (2018). On the effectiveness of machine and deep learning for cyber security. *Proceedings of the 2018 10th International Conference on Cyber Conflict (CyCon)*, 371–390.