International Journal of Advanced Engineering and Technology
www.allengineeringjournal.com
ISSN: 2456-7655
Received: 12-03-2026, Accepted: 12-04-2026, Published: 27-04-2026
Volume 10, Issue 2, 2026, Page No. 79-81

AI-based threat detection in cybersecurity: Opportunities and challenges

Simran Shinde, Pal Adnan Umar Saheb, Huzaif Bagwan
Department of Computer Science, Pillai College of Arts Commerce & Science (Empowered Autonomous), Panvel, Navi Mumbai, Maharashtra, India

Abstract
The rapid proliferation of digital infrastructure has made cybersecurity critical to modern computing. Traditional rule-based systems are inadequate against evolving threats. This paper investigates AI's dual role in cybersecurity: strengthening defenses through ML/DL-based Intrusion Detection Systems (IDS), malware identification, network traffic analysis, and automated incident response; while also introducing adversarial risks. Through literature review and comparative analysis, the paper provides a comprehensive view of responsible AI deployment for resilient cybersecurity.

Keywords: Artificial Intelligence, cybersecurity, machine learning, threat detection, IDS, network security, adversarial ML

Introduction

The digital revolution has expanded attack surfaces exponentially. Cybercrime damages are projected to exceed $10.5 trillion annually by 2025. Traditional signature-based IDS, firewall rules, and static access control policies are fundamentally reactive — effective only against known threats. As adversaries deploy zero-day exploits, polymorphic malware, and APTs, these defenses fail during critical vulnerability windows.

AI and Machine Learning offer a paradigm shift: learning from vast datasets to detect anomalies invisible to human analysts or rule-based engines. Neural networks, ensemble algorithms, and NLP models are deployed across the threat detection pipeline for behavioral analysis, traffic classification, automated threat response, and cyber threat intelligence generation.

Research Objectives

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

Literature Review

Research in AI-driven cybersecurity has accelerated over the past decade, spanning ML algorithms, intrusion detection, malware analysis, and automated response systems.

1. ML in Intrusion Detection

Liao *et al*. (2013) [1] established that signature-based IDS exhibit severe limitations detecting novel attacks. Tsai *et al*. (2009) [2] evaluated Decision Trees, SVMs, and k-NN on KDD Cup 1999, finding Decision Trees and Random Forests most accurate. Buczak and Guven (2016) [3] surveyed 40+ studies, identifying Random Forests, ANNs, and Naive Bayes as most effective, noting no single algorithm consistently outperforms across all attack categories — reinforcing the need for ensemble approaches.

2. Deep Learning & SOAR

Kim *et al*. (2016) [4] achieved 98.7% detection accuracy using RNN-based IDS on sequential network traffic. Nataraj *et al*. (2011) [6] pioneered malware binary visualization — converting executables into grayscale images for CNN classification with >95% accuracy, resistant to obfuscation. Gartner research confirms SOAR platforms leveraging AI reduce Mean Time to Respond (MTTR) by up to 90%, with Settanni *et al*. (2017) [9] demonstrating 80% of routine security operations can be automated.

3. Adversarial ML

Goodfellow *et al*. (2014) [7] demonstrated adversarial examples can fool neural networks via imperceptible perturbations. Papernot *et al*. (2016) [8] showed gradient-based attacks achieve >70% evasion rates against neural network classifiers — underscoring the need for robust adversarial training and defense-in-depth strategies.

Methodology & Core Techniques

This research employs qualitative and comparative methodology through systematic literature review, case study analysis, and performance benchmarking across four dimensions: technical efficacy, operational impact, risk profile, and ethical implications.

1. Supervised Learning

Random Forest (RF), an ensemble of decision trees, achieves 95.01% accuracy and 99.23% precision in IoT Mirai botnet detection, and 97.8% in general network anomaly detection. SVMs find optimal hyperplanes in high-dimensional feature space with strong generalization. XGBoost adds built-in feature importance metrics for interpretability.

2. Anomaly Detection & Behavioral Analysis

Gaussian Mixture Models, Isolation Forests, Autoencoders, and Variational Autoencoders detect zero-day attacks without labeled samples by modeling baseline traffic. Autoencoders flag high reconstruction error as anomalous — effective for novel exfiltration and C2 channels. UEBA platforms build dynamic behavioral profiles per user/device; deviations trigger risk scores for APT lateral movement detection.

3. Deep Learning Architectures

CNNs extract spatial features from malware binary images and network flow matrices. LSTMs analyze sequential log

files, packet sequences, and user activity timelines for temporal dependencies. Transformer models, adapted from NLP, perform threat intelligence extraction from unstructured text and vulnerability report classification.

AI-Based Threat Detection Systems

1. Intrusion Detection Systems (IDS)

AI-enhanced IDS outperform traditional tools like Snort/Suricata. Network-based IDS (NIDS) analyze packet headers, payloads, and flow statistics. Host-based IDS (HIDS) monitor system calls, file access, and process behaviors. Hybrid IDS combining signature detection with ML-based anomaly detection represent current best practice, achieving >99% detection accuracy on NSL-KDD and CICIDS2017 benchmarks

2. Malware Detection

Beyond hash-based blacklisting: static analysis extracts features from executables; dynamic analysis in sandboxed environments tracks API calls, network connections, and registry changes. CNNs on binary visualization achieve >95% accuracy across 25 malware families. RNNs on API call sequences detect polymorphic variants. GANs generate adversarial samples to augment training data.

3. Network Traffic Analysis

Flow-based ML classifiers distinguish normal traffic from DDoS, port scanning, exfiltration, and botnet C2 using packet size distributions, inter-arrival times, and protocol flags. Graph Neural Networks (GNNs) model network communications as dynamic graphs, detecting lateral movement and coordinated multi-host campaigns — especially valuable for slow-moving APTs.

4. Automated Threat Response (SOAR)

AI-powered SOAR platforms execute playbooks upon threat confirmation: isolating endpoints, blocking malicious IPs, revoking credentials, capturing forensic images, and notifying response teams — all within milliseconds. Reinforcement learning algorithms enable adaptive playbook optimization, learning optimal strategies from past incidents.

Comparative Analysis

Table 1: Traditional vs. AI-Based Cybersecurity Systems

Feature	Traditional Systems	AI-Based Systems
Threat Coverage	Known threats only	Known + zero-day threats
Adaptability	Manual rule updates	Continuous model retraining
Zero-Day Detection	Poor — no signatures	Strong — anomaly generalization
Automation Level	Low — manual analysis	High — up to 80% automated
False Positive Rate	High — broad sensitivity	Lower with well-trained models
Transparency	High — readable rules	Moderate — black-box concerns
Scalability	Limited by rule complexity	Scales with data & compute

Table 2: AI Algorithm Performance in Cybersecurity

Algorithm	Accuracy	Strengths	Limitations
Random Forest	95%–99.6%	High precision, handles large datasets	Computationally intensive
Decision Tree	Up to 99.39%	Interpretable, fast training	Prone to overfitting
Deep Neural Net	AUC ~98%+	Complex pattern recognition, auto feature learning	Opaque, high resource demand
SVM	92%–98%	Effective in high dimensions	Slow on very large datasets
LSTM/RNN	96%–99%	Excellent for sequential/temporal data	Long training, vanishing gradients
Autoencoder	90%–97%	Unsupervised anomaly detection	Threshold tuning challenging

Opportunities of AI in Cybersecurity

Real-Time Monitoring: AI processes terabytes of daily log data in real time, correlating events across thousands of endpoints to surface threats within seconds — reducing dwell time from an average of 200+ days to near-instantaneous detection.

Predictive Threat Detection: By analyzing historical attack patterns, threat feeds, and vulnerability databases, AI forecasts likely attack vectors before exploitation — shifting security posture from reactive to proactive.

Automated Incident Response: SOAR platforms encode expert knowledge into automated playbooks, reducing MTTR by over 90% while ensuring consistent policy application without human fatigue variability.

Reduction in Alert Fatigue: SOC analysts face 10,000+ alerts/day. AI-based triage filters and prioritizes alerts, allowing analysts to focus on genuine threats and reducing burnout — a chronic problem in modern security operations.

Challenges and Limitations

Adversarial ML: Carefully crafted malware samples and network packets can bypass AI classifiers while retaining malicious functionality. Data poisoning attacks inject maliciously labeled data into training pipelines; clean-label variants are especially insidious as they carry correct labels and are invisible to standard quality checks.

False Positives: Despite improvements, false positives persist in heterogeneous environments, consuming analyst time and eroding trust in automated systems. Calibrating detection thresholds without sacrificing sensitivity remains a fundamental design tension.

Data Privacy & Compliance: GDPR and similar regulations constrain data collection scope and retention duration, limiting AI model effectiveness. Organizations must implement data governance frameworks balancing security objectives with privacy obligations.

Lack of Explainability: Deep learning models are inherently opaque. In high-stakes decisions — isolating a critical server, blocking a network segment — security teams need interpretable reasoning. XAI methods (SHAP, LIME, attention mechanisms) partially address this, but fully satisfactory explainability remains an open problem.

Computational Requirements: Deep learning inference requires GPUs/AI chips. Training on updated threat data demands high-performance infrastructure. For SMEs, these costs are prohibitive, though cloud-based managed security platforms partially lower the barrier.

Future Scope

Autonomous Cyber Defense: Reinforcement learning agents that independently investigate, contain, and remediate incidents. DARPA's Cyber Grand Challenge demonstrated feasibility of autonomous vulnerability discovery and patching. Full production deployment requires resolving significant technical and governance challenges.

Federated Learning: Train robust models across organizations without centralizing sensitive data — each participant shares model updates, not raw telemetry. Enables cross-industry threat intelligence sharing (banks, hospitals, utilities) while preserving confidentiality.

Quantum Computing & LLMs: Quantum algorithms may accelerate ML training/inference dramatically; organizations should prepare post-quantum cryptographic migrations now. LLMs (GPT-4, Claude) automate threat report analysis, incident summaries, and policy generation, but also introduce new attack surfaces including prompt injection and automated phishing generation.

Conclusion

AI offers transformative capabilities addressing fundamental limitations of signature-based security. Random Forests and deep neural networks achieve 95–99% detection accuracy. SOAR platforms automate 80% of routine operations and reduce MTTR by 90%. Predictive analytics eliminates long dwell times that allow threat persistence.

However, adversarial ML attacks, high false positive rates, model opacity, privacy constraints, and computational demands require disciplined management. The path forward integrates AI's analytical power with human expertise, robust governance, and continuous adversarial hardening.

Organizations should implement XAI frameworks, maintain human-in-the-loop validation for critical decisions, and invest in adversarial training. As the threat landscape evolves, AI is not merely an enhancement but a fundamental necessity for viable defense in the digital age.

References

1. Liao *et al*. Intrusion detection system: A comprehensive review. JNCA,2013:36(1):16–24.
2. Tsai *et al*. Intrusion detection by machine learning: A review. ESWA,2009:36(10):11994–12000.
3. Buczak & Guven. Survey of data mining/ML methods for cyber security IDS. IEEE CST,2016:18(2):1153–1176.
4. Kim *et al*. LSTM RNN classifier for intrusion detection. PlatCon, IEEE, 2016.
5. Tang *et al*. Deep learning for network intrusion detection in SDN. WINCOM, IEEE, 2016.
6. Nataraj *et al*. Malware images: Visualization and automatic classification. VizSec, ACM, 2011.
7. Goodfellow *et al*. Explaining and harnessing adversarial examples. arXiv, 2014, 1412.6572.
8. Papernot *et al*. Limitations of deep learning in adversarial settings. EuroS&P, IEEE, 2016.
9. Settanni *et al*. Acquiring cyber threat intelligence through security correlation. CYBCONF, 2017.
10. Chandola *et al*. Anomaly detection: A survey. ACM CSUR,2009:41(3):1–58.
11. Mirsky *et al*. Kitsune: Ensemble of autoencoders for online network intrusion detection. NDSS, 2018.
12. Apruzzese *et al*. On the effectiveness of machine and deep learning for cyber security. CyCon, 2018.