International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 6, Nov-Dec-2026, ISSN (Online): 2395-566X

AI and Ml in Cybersecurity: Techniques, Applications, Challenges and Future Directions

Shreehari Shankarrao Lunge[1] Prof.Divya Ashok Rahane[2]
Dept. Of Artificial Intelligence And Data Science Sahyadri Valley College Of Engineering And Technology Rajuri, Pune, India

Abstract – Cyber threats are growing in volume, speed and sophistication, and defenses based only on fixed signatures and hand-written rules struggle to keep pace. Artificial Intelligence (AI) and Machine Learning (ML) offer data-driven alternatives that can learn normal behavior, classify malicious activity and prioritize alerts. This paper presents a literature-based study of AI and ML in cybersecurity drawn from five recent peer-reviewed works. It reviews supervised, unsupervised, reinforcement and deep learning methods and discusses their use in intrusion detection, malware detection, anomaly detection, threat intelligence and security automation. A reference architecture for an ML-based security pipeline is proposed, the measures used to evaluate such systems are summarized, and reported results of representative approaches are compared and read critically. The study also examines adversarial attacks, data quality and privacy, bias, resource cost, skills shortage and explainability as the main obstacles, and outlines future directions such as hybrid rule-and-ML defense, explainable AI, federated learning and automated retraining. The findings indicate that AI and ML strengthen detection and response when combined with human oversight, but they must be deployed responsibly.

Keywords – Artificial intelligence, machine learning, cybersecurity, intrusion detection, malware detection, anomaly detection, adversarial machine learning.

I. INTRODUCTION

Cybersecurity is the practice of protecting computers, networks, applications and data from unauthorized access, disruption and misuse. As organizations depend more on digital services, the number, speed and sophistication of attacks have grown, and security teams must separate real threats from very large volumes of legitimate activity [1], [4]. Traditional defenses rely on fixed signatures, predefined rules and manual analysis. These methods remain useful against known threats, but they struggle with new, polymorphic or zero-day attacks and become slower and less scalable as data volumes rise [3], [5].

Machine Learning (ML), a subset of Artificial Intelligence (AI), allows computers to learn patterns from data without being explicitly programmed for every case [1]. In security this means models that learn normal behavior, classify files and traffic as benign or malicious, rank alerts and support incident response. Interest in these methods is high: a survey reported in [4] found that 45% of the organizations studied had already adopted AI and ML in their security systems, 35% planned to do so, and 20% felt the time was not yet right. At the same time, AI is not a complete solution. It can be manipulated by adversaries, depends on good data, and raises privacy and fairness questions [3].

This paper brings together the findings of five recent studies [1]–[5] to give a structured view of AI and ML in cybersecurity. Its contributions are: (i) a comparison of the selected studies and the gaps they identify, (ii) a summary of learning methods and their security applications, (iii) a reference architecture for an ML-based security pipeline, (iv) a discussion of evaluation measures with a worked reading of reported results, and (v) a consolidated list of challenges and future directions. No new experiments were carried out, and all numerical results are quoted as reported by the original authors. Section II reviews related work, Section III introduces the fundamentals, Section IV describes applications, Section V presents the architecture, Section VI covers evaluation, Sections VII and VIII discuss challenges and ethics, Section IX outlines future directions and Section X concludes.

II. LITERATURE REVIEW

1. Overview Of Selected Studies

The five studies were chosen because they are recent (2023–2025), peer reviewed and complementary: four surveys or reviews and one hands-on implementation. Table I summarizes them. Karki et al. [1] survey ML and AI in cybersecurity and propose a data-flow framework in which collected data is cleaned, split into labeled and unlabeled sets, processed by supervised and unsupervised learners, then weighed and visualized for the organization. El Gharbaoui et al. [2] follow a systematic procedure: a Scopus search returned 1,453 records that were screened to 436 and finally 25 eligible papers, and the synthesis answers questions on applications, threat-detection benefits and integration obstacles. Hashmi et al. [3] review AI in information security, compare detection studies on benchmark datasets and give particular weight to ethics, privacy by design and adversarial resilience. Mohamed [4] covers eleven application areas, from intrusion detection to antivirus software and security education, and reports survey figures on adoption and barriers. Nimmagadda and Mehr [5] describe an implementation that combines supervised and

© 2026 IJSRET

unsupervised models with a Cowrie honeypot, a Flask API and an automated retraining plan.

Table I: Summary of the Selected Studies

Ref.	Type and scope	Main emphasis
[1]	Survey of ML and AI in cybersecurity with a proposed framework	Data pipeline, data security, third-party risk
[2]	Systematic review of AI and ML in network security (25 eligible papers)	Applications, threat detection, obstacles, research gaps
[3]	Survey of AI in information security	Detection tasks, datasets, ethics, adversarial risk
[4]	State-of-the-art survey across eleven security areas	Adoption, techniques, barriers
[5]	Implementation of an ML-based IDS with a honeypot	Lab architecture, Random Forest, API, retraining

2. Rule-Based Versus Learning-Based Defense

Hashmi et al. [3] distinguish rule-based systems, which enforce explicit conditions such as blocking an IP address after five failed logins within a minute, from learning-based systems that infer patterns from historical data without hand-written rules. Rule-based methods are consistent and easy to interpret but must be updated by people, while learning-based methods adapt to new patterns but need data and careful validation. Sommer and Paxson [6] caution that applying ML to intrusion detection is harder than in many other fields, because errors are costly, traffic is highly varied and results must be evaluated with care. Buczak and Guven [7] surveyed the ML and data-mining methods used for intrusion detection and give broader technical background. Several of the reviewed studies point toward hybrid designs that keep traditional controls and add ML on top [3], [4].

3. Research Gaps

Three gaps recur. First, El Gharbaoui et al. [2] call for more empirical research on how well AI and ML security solutions work in real deployments. Second, Karki et al. [1] note the absence of standardized security procedures for ML in cybersecurity. Third, Hashmi et al. [3] stress that high accuracy on a benchmark dataset does not guarantee protection in practice. This paper responds in a limited way by consolidating the evidence and by showing evaluation practices that give a more honest picture of performance.

III. FUNDAMENTALS OF AI AND ML IN CYBERSECURITY

1. Learning Paradigms

1) Supervised Learning: Models are trained on labeled examples to map inputs to outputs [1], [3]. In security, labels such as benign or attack are used to train classifiers such as decision trees, random forests, support vector machines, logistic regression and naive Bayes [3], [5]. Supervised models are accurate against known threats but need labeled data, which is costly to build and may carry bias [4].

2) Unsupervised Learning: These methods find structure in unlabeled data, for example by clustering similar traffic or flagging points that deviate from normal behavior [1], [3]. K-means clustering, isolation forests and autoencoders are typical examples [1], [5]. They can reveal previously unknown attacks, but their output is harder to interpret and can produce many false alarms [4].

3) Reinforcement Learning: An agent learns by trial and error from rewards and penalties [3]. In security it has been proposed for optimizing the behavior of network defenses and for choosing the best order in which to apply patches [4]. Its main costs are computing resources and expertise [4].

4) Deep Learning: Convolutional and recurrent networks, including LSTM models, learn features automatically from large datasets and are used for malware and intrusion detection [3], [4]. Generative models such as GANs and variational autoencoders can create synthetic training data [4]. Deep models need much data and computing power and are hard to interpret [4].

Table II: Common Techniques and Their Limits

Technique	Typical use	Main limit
Decision tree, random forest	Intrusion and malware classification [3], [5]	Needs labeled data; may miss new threats [4]
Logistic regression, SVM	Baseline classifiers [3], [5]	Depends on feature quality
K-means, isolation forest	Grouping traffic; finding unknown threats [1], [5]	Limited interpretability [4]
CNN, LSTM, autoencoder	Malware, sequences, anomalies [3], [5]	Large data and compute [4]
Reinforcement learning	Adaptive response, patch order [4]	Trial-and-error, compute [4]
NLP	Advisories, threat hunting, training [4]	Accuracy and scalability [4]

2. Typical Ml Workflow In Security

A typical workflow has seven stages: data collection, preprocessing, feature extraction, model training, evaluation,

© 2026 IJSRET

deployment and continuous monitoring. In [5], network traffic is captured with tcpdump into .pcap files, converted to .csv with Tshark, labeled 0 for normal and 1 for attack, cleaned of missing values, encoded for categorical fields and then used to train and test models. Karki et al. [1] describe preprocessing as cleaning, transformation, integration, reduction and discretization, and note that it improves both model performance and the accuracy of threat detection.

IV. APPLICATIONS OF AI AND ML IN CYBERSECURITY

1. Intrusion Detection

Intrusion detection and response aims to identify and react to threats before they cause damage. AI-based systems apply anomaly detection, supervised classifiers and clustering to logs, traffic and user activity [4]. Studies compiled in [3] range from classical classifiers on the KDD99 dataset to deep models: a CNN with attention feeding a Bi-LSTM on NSL-KDD reported an F1 score of 0.91, and a transformer-based system evaluated on CICIDS2017 and CIC-DDoS2019 reported 0.99. In [2], ML-based detection of DDoS attacks in programmable P4 networks is highlighted as a real-time application, together with its heavy computing demand.

2.Malware Detection

Signature-based antivirus recognizes only known samples. ML and deep learning classify files using static and dynamic features, n-grams or image representations of binaries [3], [4]. Studies in [3] include a pseudo-label deep neural network for Android malware categories that reached an F1 score of 98% on the CICMalDroid2020 dataset, and a hybrid deep learning system called ScaleMalNet that reached 99%. Mohamed [4] adds that commercial antivirus products now use ML, deep learning and behavioral detection to find zero-day threats.

3. Anomaly Detection

Anomaly detection finds behavior that departs from what is normal. In [3], a stacked Bi-LSTM model reached 94% accuracy on OpenStack cloud data, a set of ML and deep models reached 99% on IoT traffic from the DS2OS dataset, and an ensemble method reached 99.7% on KDDcup99. Hashmi et al. [3] warn that high accuracy on one dataset does not show that a system will cope with unexpected threats, and they recommend that ML flags anomalies while human analysts validate them.

4. Threat Intelligence And Security Automation

AI helps collect, analyze and share threat intelligence, including threat hunting with natural language processing, deep learning and graph analysis, and threat prediction from historical data [4]. In security automation, rule-based systems, expert systems and decision trees automate tasks such as incident response, vulnerability management and compliance monitoring [4]. Karki et al. [1] add a weighing table that scores results on performance, cost, reliability, scalability and security before dashboards and reports are produced.

5.ulnerability Management And Security Education

ML can scan systems, prioritize vulnerabilities by risk and, with NLP, extract affected products and severity from security advisories; reinforcement learning has been suggested for patch ordering [4]. AI chatbots and personalized training content support security awareness, for example by helping users recognize phishing e-mails [4].

6. Honeypot-Assisted Detection

Nimmagadda and Mehr [5] deploy the Cowrie low-interaction honeypot on a victim virtual machine. It emulates SSH and Telnet services and records login attempts, commands and credentials from attackers, producing behavioral data that conventional datasets lack. These logs enrich the training data and support retraining, so the detector can keep up with changing attacker tactics.

7. Reference Architecture And Methodology

Combining the framework of Karki et al. [1] with the laboratory setup of Nimmagadda and Mehr [5], Fig. 1 shows a reference architecture for an ML-based security pipeline. It has five processing layers and a feedback loop.

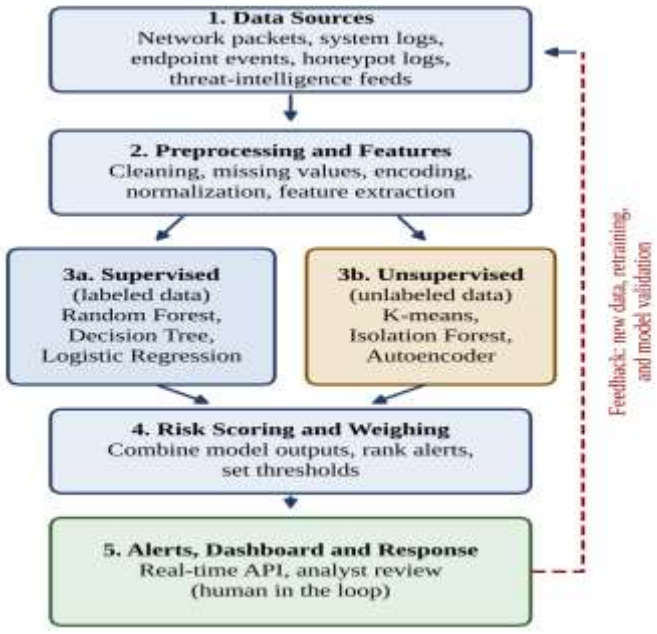

Fig. 1. Reference architecture for an ML-based security pipeline, adapted from the frameworks in [1] and [5].

© 2026 IJSRET

8. Data Collection Layer
Sources include network packets, system and application logs, endpoint events, honeypot logs and threat-intelligence feeds. Karki et al. [1] stress that only the data required for analysis should be stored and that it must be protected, since it often contains sensitive information.

B. Preprocessing and Feature Layer
Data is cleaned, missing values are handled, categorical fields are encoded and features are extracted. Data with trusted labels flows to the supervised path, while unlabeled data flows to the unsupervised path [1].

9. Detection And Decision Layers
Supervised models such as Random Forest, decision trees and logistic regression classify known attack types, while unsupervised models such as K-means, isolation forests and autoencoders look for unusual behavior [1], [5]. A weighing or risk-scoring step combines the outputs and ranks alerts before they reach analysts through dashboards, for example Kibana or Power BI [1], [5]. In [5], a Flask REST API accepts packet features in JSON and returns a benign or attack label for real-time use.

10. Feedback Layer
New traffic and honeypot data are fed back so that models can be retrained on a schedule, for example with cron jobs, validated before replacing the old model, and kept under version control [5].

11. Laboratory Setup
The implementation in [5] uses three VirtualBox machines on an isolated internal network: a Kali Linux attacker running Nmap and Metasploit, an Ubuntu victim running the Cowrie honeypot and tcpdump, and an optional analyzer machine for model training and dashboards. Isolation allows attacks to be simulated without risk to production systems.

V. EVALUATION METRICS AND PERFORMANCE

1. Confusion-Matrix Measures
Detection quality is usually described with the confusion matrix, which counts true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). Common measures are:
Precision = TP / (TP + FP) (1)
Recall = TP / (TP + FN) (2)
F1 = 2 × Precision × Recall / (Precision + Recall) (3)
FPR = FP / (FP + TN) (4)
Precision shows how many alerts are real, recall shows how many attacks are caught, F1 balances the two, and the false-positive rate (FPR) shows how often normal activity is wrongly flagged. ROC-AUC summarizes the trade-off between recall and FPR across thresholds. Accuracy alone can mislead when classes are imbalanced, a difficulty raised for imbalanced datasets in [4].

2. Operational Measures
Operational measures include detection latency, alert volume, analyst workload and retraining frequency. False alarms waste resources and can disrupt normal operations [4], whereas real-time classification with near-instant predictions is described as a practical strength in [5].

3.Reported Performance
Fig. 2 plots results that Hashmi et al. [3] compile for selected approaches. Values range from 91% to 99.7%, but the studies use different datasets, tasks and metrics (accuracy or F1), so they should not be ranked against one another. They show that modern ML and deep models can reach high scores on benchmark data.

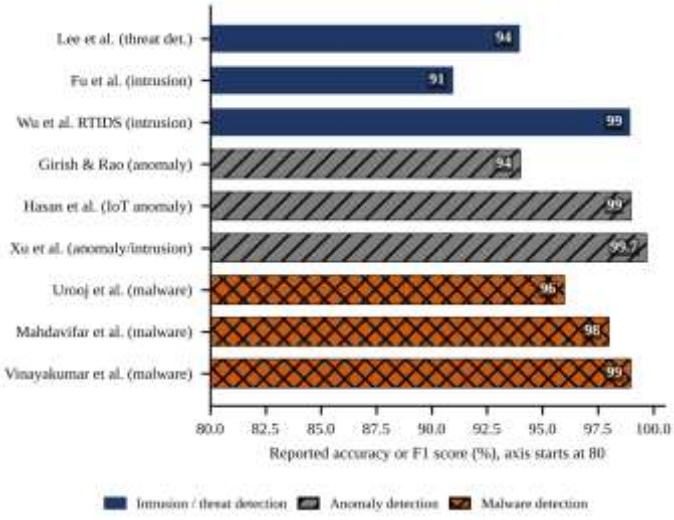

Fig. 2. Reported accuracy or F1 score (%) of selected approaches compiled in [3]. Datasets and metrics differ between studies.

4. A Closer Look At A Laboratory Result
Nimmagadda and Mehr [5] report that a Random Forest classifier exceeded 95% accuracy and describe false positives as minimal. The confusion matrix they show contains about 19 true negatives, 11 false positives, no false negatives and about 270 true positives (the largest cell is printed as 2.7e+02). From these values the overall accuracy is (19 + 270) / 300, or about 96%, which agrees with the reported figure, and recall for attacks is 100%. However, the false-positive rate is 11 / 30, or about 37%, because only 19 of the 30 benign samples were classified correctly. This calculation was made by the authors of this paper from the published figure and is approximate. It illustrates why

© 2026 IJSRET

accuracy alone is not enough and why test data should be balanced and realistic [3], [4].

5. Dataset Considerations

The studies compiled in [3] rely on benchmark datasets such as NSL-KDD, CICIDS2017, KDD99, UNSW-NB15 and CICMalDroid2020. Hashmi et al. [3] note that models must still be validated on data from other environments, retrained as threats change and supervised by people, because a model that is accurate on its training data may not generalize to new conditions.

VI. CHALLENGES AND LIMITATIONS

- **Adversarial Attacks:** Attackers can deliberately change inputs to mislead a model into wrong predictions [3]. Building resilient systems requires adversarial training, new architectures and constant adaptation, and the large datasets this needs raise privacy concerns [3].
- **Data Security, Privacy and Quality:** ML systems depend on large datasets that often contain personal information, so strong encryption, access control and compliance with privacy law are needed [1]. Poor data quality, accuracy and integrity weaken every technique reviewed in [4], and in federated learning the participating devices must be confirmed to be free of malicious intent [1].
- **Bias and Fairness:** Models can reproduce biases present in training data and produce unfair results, so they must be monitored, audited and tested before deployment [3].
- **Resource Intensiveness:** Training and running advanced models, especially deep and generative ones, requires high computing power and energy. This raises cost and environmental concerns and is hardest for small organizations [3]. El Gharbaoui et al. [2] note that such demands can hinder real-time use in large networks.
- **Skills and Understanding:** Mohamed [4] reports the barriers in Table III. A shortage of skilled professionals is also noted in [2] and [3].

Table III: Barriers to Adoption Reported in [4]

Barrier	Organizations reporting
Lack of understanding of the technology	36.9%
Shortage of skilled personnel	34.0%
High costs	29.1%

- **Interpretability and False Alarms:** Many models are hard to explain, and imbalanced data can bias them toward the majority class, giving poor detection of rare attacks or many false alarms that waste analyst time [4].
- **Standards and Third-Party Dependency:** No agreed security standards yet exist for ML in cybersecurity [1]. Outside libraries, datasets and sources add vulnerabilities, so third-party resources should be screened and software dependencies regularly updated and tested [1].

1. Ethical And Governance Considerations

Hashmi et al. [3] argue that ethical AI goes beyond legal compliance and rests on fairness, transparency, accountability, sustainability and privacy. Privacy by Design asks that privacy be the default setting, be built into the system and be supported by end-to-end security and transparency. UNESCO's recommendation on the ethics of AI and Ethical Impact Assessments are offered as tools for judging effects on individuals and society before deployment [3]. Intrusion detection also collects large amounts of sensitive data, which raises data-protection concerns [4]. For practitioners this means documenting data use, limiting retention, auditing models and keeping a human in the loop for decisions with real consequences [3].

2.Future Directions

- **Hybrid Defense:** Future systems should combine the interpretability of rule-based methods with the adaptability of ML, and should be tuned to the needs of each sector [3].
- **Explainable and Privacy-Preserving AI:** Explainable AI can make model decisions transparent and raise user trust [2], [3]. Federated learning, which keeps data on local devices, is one route to better privacy and is discussed together with explainability in work summarized in [2].
- **Automated Retraining and Live Log Parsing:** Automated pipelines can retrain models on new honeypot data, validate them and keep version control, and parsing honeypot logs in real time would close the feedback loop [5].
- **Deep Learning for Sequences:** LSTM networks for time-series traffic and CNNs for localized traffic patterns are planned extensions in [5].
- **Generative AI and Standard Benchmarks:** Generative AI and large language models are increasingly used in security tasks but are resource intensive [3]. Further empirical studies [2] and standardized procedures [1] are needed so that results can be compared fairly.

VII. CONCLUSION

AI and ML have changed how cyber threats are detected and handled. The studies reviewed here show strong results in intrusion, malware and anomaly detection, useful support for

© 2026 IJSRET

threat intelligence and automation, and growing adoption in industry. They also show that accuracy figures from benchmark data can hide weaknesses, as the closer reading of one laboratory result in this paper illustrates. Adversarial attacks, data privacy, bias, cost, skills and explainability remain open problems. The most promising path is a hybrid, human-supervised approach that keeps traditional controls, retrains models as threats change, evaluates them with balanced and realistic data, and follows ethical and privacy principles. Future work should include empirical testing of the proposed architecture on current datasets.

1. Acknowledgment
The authors thank the Department of Artificial Intelligence and Data Science, Sahyadri Valley College of Engineering and Technology, Rajuri, Pune, for the facilities and support provided, and the authors of the studies cited in this paper.

REFERENCES

1. S. Karki, A. B. M. M. Hasan and C. Sanin, “Use of ML and AI in cybersecurity – A survey,” Procedia Computer Science, vol. 246, pp. 1260–1270, 2024. doi: 10.1016/j.procs.2024.09.552.
2. O. El Gharbaoui, I. Kiyadi and H. El Boukhari, “Evaluating AI and ML in network security: A comprehensive literature review,” Procedia Computer Science, vol. 251, pp. 727–733, 2024. doi: 10.1016/j.procs.2024.11.176.
3. E. Hashmi, M. M. Yamin and S. Yildirim Yayilgan, “Securing tomorrow: a comprehensive survey on the synergy of artificial intelligence and information security,” AI and Ethics, vol. 5, pp. 1911–1929, 2025. doi: 10.1007/s43681-024-00529-z.
4. N. Mohamed, “Current trends in AI and ML for cybersecurity: A state-of-the-art survey,” Cogent Engineering, vol. 10, no. 2, Art. 2272358, 2023. doi: 10.1080/23311916.2023.2272358.
5. A. Nimmagadda and S. Yavary Mehr, “AI-powered intrusion detection system with honeypot integration,” International Journal of Intelligent Information Systems, vol. 14, no. 4, pp. 70–76, 2025. doi: 10.11648/j.ijiis.20251404.11.
6. R. Sommer and V. Paxson, “Outside the closed world: On using machine learning for network intrusion detection,” in Proc. IEEE Symposium on Security and Privacy, pp. 305–316, 2010.
7. A. L. Buczak and E. Guven, “A survey of data mining and machine learning methods for cyber security intrusion detection,” IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153–1176, 2016. A

© 2026 IJSRET