A Review of Artificial Intelligence Techniques for Cybersecurity Threat Detection

Shilpa Nawale
Dept. of Computer Science School of Information Technology Indira University, Pune

Sarika Shingate
Dept. of Computer Science School of Information Technology Indira University, Pune

DOI: 10.29322/IJSRP.16.02.2026.p17037
https://dx.doi.org/10.29322/IJSRP.16.02.2026.p17004

Paper Received Date: 6th January 2026
Paper Acceptance Date: 7th February 2026
Paper Publication Date: 12th February 2026

ABSTRACT

The effectiveness of earlier signature-based and rule-driven methods of detection was declining due to the increasing number and complexity of threats. Artificial intelligence (AI) is now a vital tool in cybersecurity, which enables the analysis of patterns of behaviour and the detection of unusual activity within complicated networks. This paper offers an in-depth examination and assessment of AI-driven threat detection methods. It studies supervised machine learning frameworks, graph-based algorithms, deep learning architectures, and unsupervised anomaly detection techniques. The evaluated methods demonstrate better accuracy over standard approaches, less false positives, and greater detection capabilities for unknown threats. The research paper also identifies significant obstacles, such as the need for application awareness modelling, competitive avoidance, and restriction on prompt deployment. Future research possibilities to develop scalable, reliable, and understandable artificial intelligence threat detection tools are examined in the paper's conclusion.

I. INTRODUCTION

The modern digital age is evolving rapidly because to services in the cloud, web-based apps, and interconnected devices. This development has led to a spike in both the severity and frequency of cyberattacks. Traditional threat detection methods, like rule-based or signature-based systems, are restricted to detecting known attacks. They are incapable to identify innovative, undetected or sophisticated threats like multi-step attacks, developing malware, or zero-day vulnerabilities. These outdated approaches are consequently unable to protect modern businesses.

Because it can gain insight from data and make wise decisions, artificial intelligence (AI) has become known as a potent cybersecurity tool. Vast amounts of network traffic, information, and user activity can be evaluated by AI-based systems to detect unusual patterns which could point to a threat. [1], [2]

Although unsupervised models may recognize unusual behavior even in lack of labeled data, machine learning may group attacks based on past instances. Deep learning models, which include CNNs, RNNs, and transformers, are especially useful for threat detection since they can automatically identify patterns from complex security data.

Security teams may detect threats quicker, more precisely, and with fewer false alarms when AI is built into the intrusion detection systems (IDS) and Security Operations Centers (SOC). In order to locate hidden attack paths, advanced methods like neural networks with graphs (GNNs) can understand the relationships between people, devices, and processes. AI may be utilized as well to forecast possible dangers and provide responses. [5], [7]

The objective is to aid in the development of AI-based systems which are more capable, reliable, and secure to be able to protect against modern cyber threats. Current threat detection methods frequently ignore new and complex cyberthreats and only detect established attacks. Thus, enhanced systems that can quickly and accurately detect unknown attacks must be developed. The present research analyzes various AI-based threat detection techniques and compares their benefits with their drawbacks. False alerts, data problems with security, and immediate detection issues are also addressed. The present study is not intended to implement an innovative approach, instead choosing to examine and review current strategies.

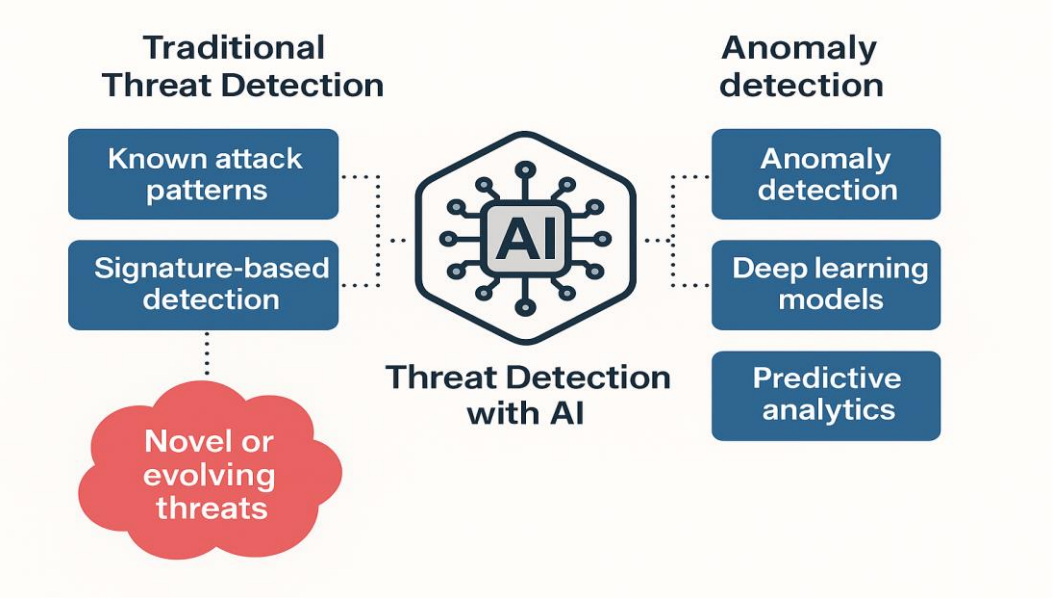

Figure 1. demonstrates the difference between AI-based and conventional threat detection.

II. LITERATURE REVIEW

A. Conventional Methods of Threat Detection

Early cybersecurity systems identified vulnerabilities through simple, fixed approaches. Incoming files or traffic on the network were compared to established patterns of attack utilizing detection using signatures. The method was effective at recognizing known and dated threats, but it was incapable to detect new or altered attacks like polymorphic malware or zero-day exploits.

Human-written rules have been used by systems based on rules to define unusual or suspicious actions. These systems were unable to keep up with constantly changing cyberattacks and required frequent update manually. An attack remained undiscovered if it did not conform to an established norm.

By emphasizing odd activity, statistical anomaly detection made an attempt to find threats. Despite its ability to identify unforeseen threats, it typically generated a large number of false alarms due to significant variations in normal user and system behaviour.

In general, conventional detection techniques are hampered by:

- Lack of ability to adapt
- High maintenance.
- failure to recognize new or complicated attacks.
- High rates of false positives.

As a consequence, these innovations are inappropriate for contemporary, dynamic cyber circumstances.

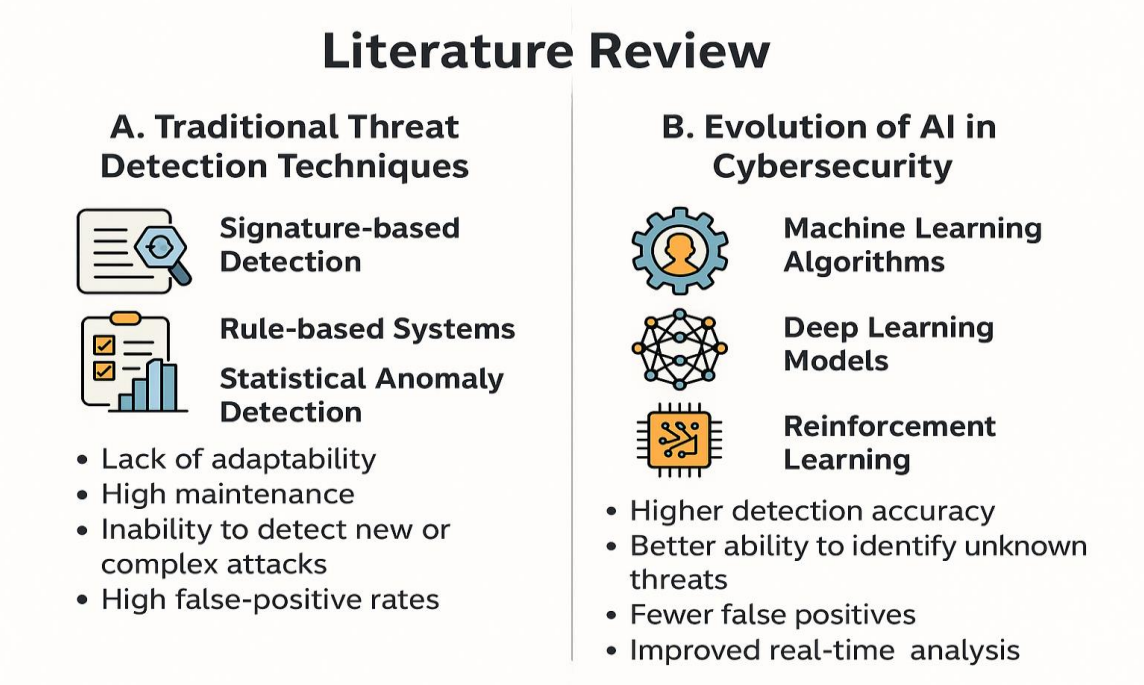

Figure 2 summarizes the examination of the literature.

B. AI's Progress in Cybersecurity

The identification of threats has shifted as an outcome of the use of artificial intelligence (AI), moving from reactive to predictive and flexible methods. To classify malicious behavior, early research included standard machine learning techniques like as decision tree models, SVMs, Naïve Bayes, and clustering. Though these algorithms improved detection accuracy, they weren't very effective with noisy or imbalanced data and mostly relied on human feature selection.

More potent AI techniques appeared as a result of mathematical innovations and the easy availability of huge cybersecurity sets (including KDD, NSL-KDD, CICIDS, and UNSW-NB15). Machine learning techniques, such as CNNs, RNNs, LSTMs, and transformers, made it possible for machines to identify significant patterns from intricate network data. These techniques worked well for identifying highly sophisticated malware variants, botnets, and DDoS attacks.

Systems that learn the most effective tactics for defense based on constant input were introduced by Reinforcement Learning (RL).

By understanding the links between devices, individuals, and network nodes, Graph Neural Networks (GNNs) improve detection by helping in recognizing of lateral network movements. [8]

Evidence shows that systems powered by AI offer:

- Enhanced accuracy of detection
- Improved ability for identifying unidentified risks
- Decreased false positives
- Improved evaluation in real time
- Unfortunately, problems like conflicting attacks, explanation issues, and installation problems continue to be addressed.

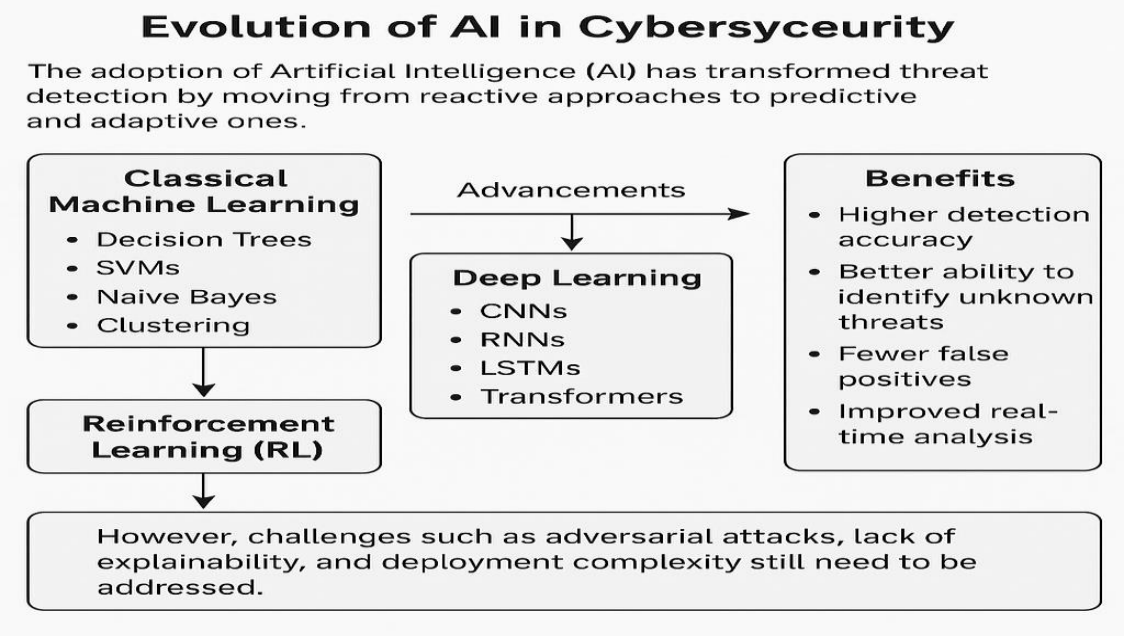

Figure 3. illustrates the growing role of artificial intelligence in security.

III. TECHNIQUES

This section presents a structured summary of popular AI techniques to detect dangers to cybersecurity, including graph-based models, deep learning, supervised training, and unsupervised detectable anomalies. The models provided here are conceptually evaluated and compared based on findings given in past research and standard studies, instead of having experimentally applied in this study.

1. Machine models for supervised learning

Each network flow is classified as either benign or hazardous when the models in question are trained on certain data sets.

Techniques Used:

- Logistic regression: presents a simple linear selection boundaries and a binary classification as the starting point.
- A random forest approach: Utilizing several decision trees for handling difficult problems in order to improve durability via shared learning.

  Support Vector Machine (SVM): It offers the most effective classification boundary for large volumes of threat data.
- Gradient Boosting (XGBoost): By gradually building models which correct previous errors, this method ensures reliability for complex risk patterns. These models have been useful in classifying internet traffic and recognizing common attacks such as brute force attacks, distributed denial of service (DDoS) attacks, and port scans. [5], [6]

2. Models for Unsupervised Anomaly Detection

These do not require labelled attack data, compared to supervised models. They get familiarized with the network's usual conduct and recognize any deviations that may present dangers. They are consequently useful for recognizing unexpected threat patterns, uncommon incidents, and zero-day threats.

Models Useful:

- Isolation Forest: Suitable for massive network logs, it efficiently isolates outliers through random partitioning.
- One-Class SVM: Recognizes spots that are outside of the typical pedestrian area after establishing the borders around it.
- Autoencoders: Neural networks that replicate common traffic patterns are known as autoencoders. Inappropriate or malicious behaviour is detected by high reconstruction errors.

  These models were important in spotting minute inefficiencies and threats that had not been detected previously.

3. Types of Deep Learning

Complex, high-dimensional features were directly obtained from raw or prepared information using deep learning models. Both the spatial and temporal characteristics of network traffic are taken into account by these models. [5], [7]

Models Employed:

Convolutional neural networks (CNNs): Find patterns of structure in packets and extract beneficial characteristics from traffic matrices.

Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM): Detect multi-step or forming assaults through examining sequential actions in time-series logs. [5], [6]

Transformer models can be helpful to recognize complex and carefully executed attacks as they employ self-attention to understand dependencies that last.

Deep learning reduced the requirement for manual feature engineering and significantly enhanced detection accuracy.

4. Graph-Based Models

Many gadgets and intricate routes of communication are frequently employed during modern cyberattacks. By displaying network pieces as node and interactions as edges, network-based models have the ability to capture these connections.

Models Developed:

Graph Neural Networks (GNNs): To understand how assaults propagate across a network, model interactions among hosts, IPs, ports, and activities.

Graph Convolutional Networks (GCN): Integrate data from nodes nearby to find abnormal connections or hidden routes of attack.

Graph Attention Networks (GAT): Improve the detection of hidden lateral movement by employing attention mechanisms to identify relevant network connections. Insider threats and complicated attack chains were very recognized by graph-based algorithms. [8]

5. Optimization of Hyperparameters

To improve performance and lower false positives, every model was carefully adjusted. The following methods were applied:

- Grid Search: An in-depth search using predefined parameter combinations.
- Random Search: To quickly cover an additional search space, parameters were regularly selected.
- Bayesian Optimization: A smart method of searching based on past results to swiftly arrive at optimal parameters.

Each model was ensured to operate at its most effective configuration due to this optimization process.

Summary of Model Architecture

Category	Models Used	Purpose/Function	Strengths	Limitations
Supervised Machine Learning Models	Logistic Regression, Random Forest, SVM, XGBoost	Classify network traffic as benign or malicious using labeled data	High accuracy for known attacks; interpretable (LR, RF)	Limited ability to detect unknown or zero-day attacks
Unsupervised Anomaly Detection Models	Isolation Forest, One-Class SVM Autoencoders	Identify unusual or abnormal network behaviors without labeled data	Detect zero-day threats: useful when labada is limited	May generate false positives; sensitive to noise
Deep Learning Models	CNN, RNN/LSTM Transformers	Learn complex spatial and temporal patterns from network traffic	Excellent for detecting complex attack paths and	High model complexity, computationally expensive
Optimization Techniques	Grid Search Random Search Bayesian Optimization	Improve model accuracy and reduce false positives	Systematic an efficient tuning	

IV. EVALUATION METRICS

In comparison to the results of each model, the evaluation metrics employed include:

- **Accuracy:**
  Calculates how many total predictions were correct.
- **Precision:**
  Among all the predictions classified as attacks how many were attacks.

- **Recall:**
  Number of actual attacks successfully identified by the model among all actual attacks.
- **F1-S:**
  It provides a balanced accuracy that takes into account both its precision and recall.
- **ROC-AUC:**
  Indicates the quality of separation between the attack traffic and the normal traffic.
- **False Positive Rate (FPR):**
  Frequency at which normal traffic was incorrectly identified as an attack.
  All of these factors were used in combination in order to assess the detection accuracy and the efficiency of the threat detection models in real-time processes.

IMPORTANT FINDINGS

- The algorithms are more effective than traditional algorithms. The traditional approach is unable to handle the latest, dynamic, and polymorphic threats.
- Feature gathering helps in improved detections through Deep Learning models. The AI models ensure better danger detection through patterns that keep changing. [5], [7]
- CNNs can detect prominent packet-level and flow-level patterns.
- LSTMs are used to detect sequential events of an attack such as multi-step attacks or login attempts. [5], [6]
- Graph Neural Networks (GNNs) are known for their capability to handle complex relationships in graphs. [8]
- They are highly effective in tracing hidden attack paths and lateral movement between computers.
- They have great importance for monitoring at the SOC level and for other activities of threat hunting.
- Unsupervised learning models have shown efficacy in identifying zero-day or “unknown” threats.
- Auto encoders are anomaly detectors that rely on analysis of reconstruction errors.
  Isolation Forest identifies unusual anomalies without the need for labelled attacked data

V. CHALLENGES

- **Vulnerability to Adversarial Attacks** AI models, particularly deep learning systems, are susceptible to adversarial manipulation. Attackers can subtly modify packet features, inject noise, or craft malicious traffic that mimics normal behavior. These small perturbations can mislead ML/DL models into misclassification, allowing intrusions to bypass detection systems.
- **Limited Explainability and Transparency** Deep learning models (CNNs, LSTMs, Transformers, GNNs) often operate as “black boxes,” making it difficult for cybersecurity analysts to understand why a particular alert was generated. This lack of interpretability reduces trust, complicates incident investigation, and challenges regulatory compliance where explanation of decisions is required. [5], [7]
- **Real-Time Constraints and High Computational Demand** Deploying AI systems in real-world networks requires low-latency processing to detect threats instantly. However, many deep models are computationally intensive and slow when handling high-volume traffic streams. Achieving real-time inference requires model compression, hardware acceleration, or system-level optimization.
- **Data Imbalance and Scarcity of Malicious Samples** Cybersecurity datasets typically contain far fewer malicious events compared to large volumes of normal traffic. This imbalance causes models to become biased toward benign classifications, leading to poor detection of rare yet critical attack classes. Techniques like oversampling, synthetic data generation, or class-weight adjustments are required to stabilize training. [1]-[4]

VI. CONCLUSION

- AI-based threat detection methods perform better than traditional signature and rule-based systems.
- Supervised models are effective for identifying well-known and previously observed attacks.
- Unsupervised and deep learning models can detect new, unknown, or evolving threats without relying on labeled data. [5], [7]
- Graph Neural Networks enhance detection by understanding relationships between devices and identifying multi-step attack paths. [8]
- Key challenges remain, including:
  - vulnerability to adversarial attacks,
  - limited explainability of deep models,
  - data imbalance in intrusion datasets, [1]–[4]
  - and the need for faster real-time processing.

- Overall, AI is a strong foundation for next-generation cybersecurity and will continue to support more adaptive and resilient intrusion detection systems.

VII. FUTURE SCOPE

Future research in AI-driven threat detection can focus on several promising directions:

- **Explainable AI (XAI) for Cybersecurity** Developing transparent and interpretable deep learning and GNN-based models to improve analyst trust and support real-world SOC operations.
- **Federated and Privacy-Preserving Learning** Enabling collaborative threat detection across organizations without sharing sensitive network data, enhancing both privacy and model generalization.
- **Real-time and Edge-based Intrusion Detection** Creating lightweight, low-latency AI models suitable for IoT, mobile devices, and 5G edge nodes where rapid response is critical.
- **Hybrid AI–Human Decision Systems** Combining machine intelligence with human expertise to support threat triage, alert prioritization, and adaptive response strategies.

REFERENCES

1. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization,” Proc. ICISSP, 2018.
2. M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, “A Detailed Analysis of the KDD CUP 99 Data Set,” IEEE Symposium on Computational Intelligence for Security and Defense Applications, 2009.
3. N. Moustafa and J. Slay, “UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems,” Military Communications and Information Systems Conference, 2015.
4. I. Sharafaldin, A. Lashkari, and A. Ghorbani, “CICIDS2017 Dataset,” Canadian Institute for Cybersecurity, 2017.
5. A. Javaid, Q. Niyaz, W. Sun, and M. Alam, “A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks,” IEEE MILCOM, 2016.
6. T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," arXiv:1301.3781, 2013. (For sequential models / embeddings) [5], [6]
7. A. Vaswani et al., “Attention Is All You Need,” Advances in Neural Information Processing Systems (NeurIPS), 2017.
8. T. Kipf and M. Welling, “Semi-Supervised Classification with Graph Convolutional Networks,” International Conference on Learning Representations (ICLR), 2017.