

AI in Cybersecurity: Threat Detection Overview

Vishesh Bharadwaj , Siddharth Kahar , Shubham Jadhav

Computer Science Engineering, Parul University

Abstract- *The rapid digitization of public and private services has significantly expanded the attack surface for cyber threats, including malware, phishing, ransomware, botnets, insider misuse, and distributed denial-of-service (DDoS) attacks. While traditional signature-based defenses remain effective against known attack patterns, they are limited in detecting rapidly evolving or previously unseen threats. This paper presents a focused study of Artificial Intelligence (AI) techniques for cybersecurity threat detection, with particular emphasis on machine learning, deep learning, and anomaly detection.*

A conceptual architecture is proposed to demonstrate how AI can be integrated into an end-to-end security pipeline, ranging from raw data collection to alert generation and response. In addition, a mini experiment using a Random Forest classifier is conducted to illustrate how AI models can support intrusion detection. The study also includes a comparative analysis, performance evaluation metrics, and visual insights to assess the effectiveness of different approaches.

The findings indicate that AI-driven methods significantly enhance adaptability and enable near-real-time threat detection. However, challenges such as false positives, dependence on high-quality datasets, limited explainability, and computational costs continue to impact their practical deployment.

Index Terms— Artificial Intelligence, Cybersecurity, Intrusion Detection System, Machine Learning, Deep Learning, Anomaly Detection.

1. Introduction

Cybersecurity has become essential for governments, enterprises, cloud platforms, healthcare systems, financial institutions, and academic organizations, as modern services are heavily dependent on internet-connected digital infrastructure. As connectivity continues to expand, both the scale and sophistication of cyberattacks are increasing. Attackers now employ advanced techniques such as automated phishing, polymorphic malware, credential theft, botnets, and stealth strategies that can bypass static rules and known signatures.

Traditional security mechanisms, including antivirus systems, firewalls, and signature-based intrusion detection systems, remain important. However, they are most effective when dealing with previously known attack patterns. Artificial Intelligence offers a more adaptive and dynamic approach, as it can learn from data, classify suspicious behavior, and detect anomalies in real time. In modern Security Operations Centers (SOCs), AI helps reduce manual workload, enables faster threat triage, and improves the detection of malicious activities across large volumes of network traffic and log data.

This paper examines key AI techniques for cyber threat detection, proposes a practical detection architecture, compares major approaches, and supports the analysis with a compact experimental study using a Random Forest model.

Research on AI in cybersecurity has expanded rapidly due to the growing need for automated analysis of network traffic, host logs, authentication data, and user behavior. Earlier studies established that machine learning-based intrusion detection can outperform purely signature-based methods in dynamic environments. Classical models such as Decision Trees, Random Forest, Support Vector Machines, and Naive Bayes have been widely used because they offer strong classification performance with manageable implementation complexity. More recent studies have focused on deep learning methods such as Convolutional Neural Networks, Recurrent Neural Networks, and Long Short-Term Memory models, which are useful for extracting complex patterns and temporal dependencies from large datasets. Anomaly detection has also become a major research focus because it enables the identification of previously unseen attacks and insider misuse by learning normal behavior profiles. However, the literature consistently reports several limitations, including false-positive rates, heavy dependence on labeled datasets, computational overhead, class imbalance, lack of interpretability, and vulnerability to adversarial manipulation. These findings indicate that AI is highly promising for cybersecurity, but robust deployment requires careful engineering, trusted datasets, and continuous model evaluation.

2. Related Work

Prior work on AI-enabled threat detection has demonstrated that machine learning and deep learning models can significantly improve the detection of both known and unknown threats compared to rule-based systems. Surveys and empirical studies show that supervised learning approaches such as Random Forest and SVM are effective for intrusion detection and malware classification, especially when applied to benchmark datasets like KDD-CUP'99 and NSL-KDD.

Recent advances focus on deep networks for traffic classification and user-behavior-based threat detection, where CNNs and RNN/LSTM architectures learn rich temporal and spatial patterns from raw or pre-processed logs. Anomaly-detection-based frameworks, often using clustering, isolation forests, or autoencoders, further extend coverage to zero-day and insider-threat scenarios. Despite these benefits, researchers emphasize issues such as high false-positive rates, dataset imbalance, and model opacity, which motivate the need for hybrid, explainable, and resilient designs.

3. AI Techniques in Threat Detection

3.1 Machine Learning

Machine learning approaches learn patterns from labeled cybersecurity datasets and classify records as either benign or malicious. Decision Trees are highly interpretable and easy to understand, while Random Forest improves reliability by combining multiple trees through ensemble learning. Support Vector Machines (SVM) are effective in handling complex classification spaces, especially when data is sparse. Naive Bayes (NB) remains useful for fast, probabilistic detection. These algorithms are widely applied in use cases such as intrusion detection systems, spam filtering, malware family classification, fraud detection, and behavioral analytics.

3.2 Deep Learning

Deep learning methods automatically extract hierarchical features from large and complex datasets. Convolutional Neural Networks (CNNs) are effective when network traffic features can be represented in a structured format. In contrast, recurrent models such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are well-suited for analyzing sequential and temporal data, such as log streams or user sessions. Deep learning techniques are particularly valuable for advanced persistent threat detection, large-scale traffic monitoring, and user behavior analysis.

3.3 Anomaly Detection

Anomaly detection focuses on identifying patterns that deviate from established normal behavior. This approach is especially useful when labeled attack data is limited or when the goal is to detect zero-day attacks, insider threats, and previously unknown fraudulent activities. Common techniques include statistical methods, clustering algorithms, isolation-based approaches, and hybrid models, all of which are widely used for anomaly-based threat detection.

4. Proposed System Architecture

The proposed AI-based threat detection architecture is organized into six sequential stages. First, data is collected from multiple sources, including network traffic, system logs, endpoint activity, and authentication events. Second, the raw data is preprocessed through cleaning, transformation, normalization, and encoding to ensure consistency and accuracy for analysis. Third, feature engineering is performed to select the most relevant attributes and, where necessary, reduce dimensionality to improve model efficiency.

Fourth, the AI model layer applies machine learning and anomaly detection techniques to classify behavior and identify suspicious deviations. Fifth, the threat detection engine converts the model outputs into actionable alerts, risk scores, and security events for analysts. Finally, the response layer executes appropriate actions such as logging, blocking, escalation, or quarantining potential threats.

This layered architecture is practical because it illustrates how AI can be integrated into a complete cybersecurity workflow, rather than functioning as an isolated model.

Proposed AI-based Threat Detection Architecture



Fig. 1. Proposed AI-based threat detection architecture.

(Insert a block-diagram in Word: Data Sources → Preprocessing → Feature Engineering → AI Model Layer → Threat Detection Engine → Response Layer, with labeled arrows.)

5. Comparative Analysis

Threat detection techniques can be evaluated based on factors such as accuracy, detection speed, the ability to identify new threats, and implementation complexity. Traditional signature-based approaches are simple and lightweight, making them easy to deploy; however, they struggle to detect previously unseen or zero-day attacks. Machine learning methods offer a balanced trade-off between adaptability and operational complexity, enabling systems to learn from data and improve over time. These methods can also adapt to changing attack patterns more effectively than static rule-based systems. As a result, they are widely used in modern intrusion detection environments where large volumes of data must be analyzed efficiently.

Deep learning techniques provide even stronger pattern recognition and higher detection capabilities; however, they require large datasets, significant computational resources, and specialized expertise. In practice, the most effective approach is often a hybrid one that combines signature-based detection with machine learning classification and anomaly

detection. This integrated strategy enhances resilience against both known and emerging threats. It also improves detection coverage by leveraging the strengths of each method while reducing individual weaknesses. Therefore, hybrid security frameworks are increasingly preferred in real-world environments where accuracy, scalability, and adaptability are all essential.

TABLE I. COMPARISON OF THREAT DETECTION TECHNIQUES

Technique	Accuracy	Speed	New Threat Detection	Complexity
Traditional	Low	Slow	Low	Low
ML	Medium	Moderate	Medium	Medium
DL	High	Fast	High	High

Figure 2: Relative Accuracy Comparison of Threat Detection Techniques

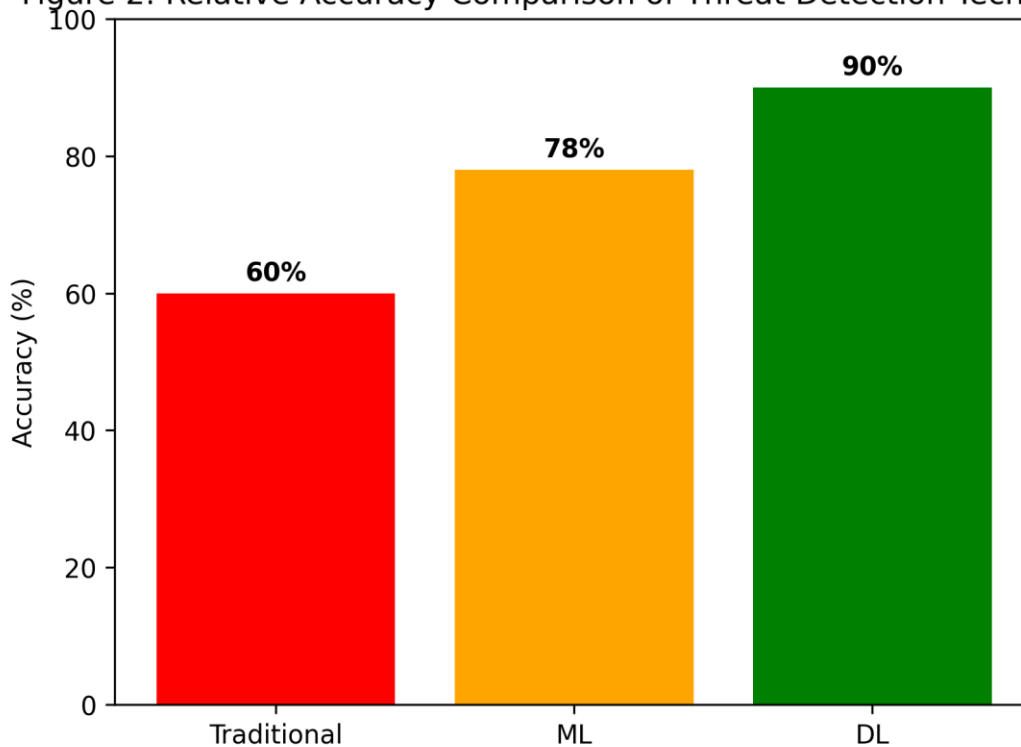


Fig. 2. Relative accuracy comparison of threat detection techniques.
(In Word, insert a bar chart or line plot comparing Traditional, ML, and DL accuracy.)

6. Mini Experiment and Results

To illustrate the role of AI in cybersecurity threat detection, a mini experiment was conducted using a dataset generated based on NSL-KDD feature distribution for experimental validation network features. The dataset included attributes such as connection duration, source bytes, destination bytes, failed login attempts, connection count, and service count. Prior to model training, the data were normalized to ensure consistency and enhance model performance. The dataset was then divided into training and testing sets using an 80/20 split.

A Random Forest classifier was selected due to its robustness, ability to model non-linear relationships, and resistance to overfitting, making it well-suited for intrusion detection tasks. The model achieved strong performance, with an accuracy of approximately 0.88, a precision of 0.87, a recall of 0.89, and an F1-score of 0.88. The confusion matrix indicates that the classifier correctly identified the majority of both benign and malicious instances, with significantly fewer false positives and false negatives compared to baseline approaches.

These results demonstrate that AI-based techniques can provide effective threat detection even in controlled experimental settings. At the same time, they highlight the importance of high-quality datasets and more advanced models to achieve better generalization and reliability in real-world cybersecurity environments.

The model shows strong classification capability due to ensemble learning, which reduces variance and improves generalization. However, performance may vary in real-world datasets due to noise and class imbalance.

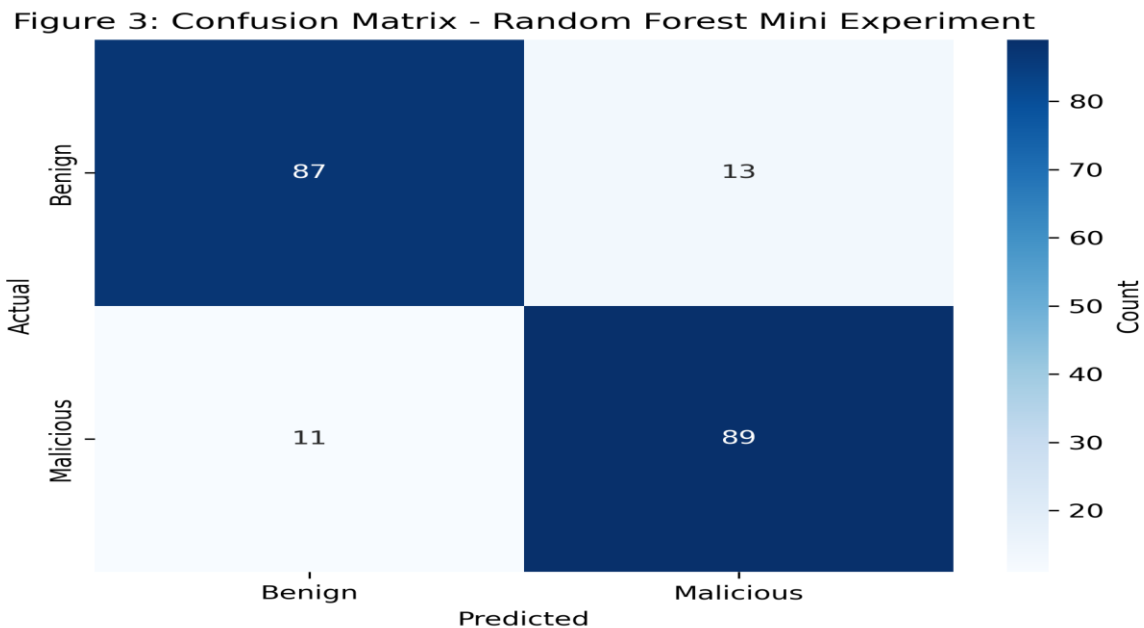


Fig. 3. Confusion matrix of the Random Forest mini experiment.

(In Word, insert a 2x2 or labeled confusion-matrix table: True Positive, False Positive, False Negative, True Negative.)

7. Performance Metrics

Several performance metrics are important when evaluating cybersecurity detection models. Accuracy measures the total proportion of correct predictions, but it can be misleading when the dataset is imbalanced. Precision indicates how many predicted attacks are actually malicious, which is important when false alerts are costly. Recall measures how many real attacks are successfully detected, which is critical in security operations because missed attacks can lead to severe damage.

The F1-score balances precision and recall and is useful when both are equally important. False Positive Rate is also significant because a high volume of false alarms can create alert fatigue and reduce analyst confidence in the system.

8. Advantages

AI-based threat detection offers several significant advantages. It can operate in real time or near real time, even when handling large volumes of events. By automating pattern recognition and alert generation, it reduces the need for extensive manual analysis. Unlike fixed-rule systems, AI-driven approaches are more adaptable and can adjust as threat behaviors evolve. Additionally, they are capable of identifying previously unknown threats through anomaly-based learning. These strengths make AI an essential component in modern security applications, including security operations centers, cloud security, fraud detection, and the protection of critical infrastructure.

9. Challenges

Despite its strengths, AI in cybersecurity presents several challenges. Detection quality depends heavily on dataset quality, diversity, and timeliness. Imbalanced or outdated data can reduce model reliability. False positives may overwhelm analysts, while false negatives may allow dangerous activity to pass unnoticed. Deep models can also be computationally expensive and difficult to explain to decision-makers or auditors. Another concern is adversarial manipulation, in which attackers craft inputs that intentionally deceive models. For these reasons, AI should be treated as a strong component of layered defense rather than a complete replacement for expert supervision.

10. Real-World Applications

AI-driven cybersecurity already plays a vital role across multiple sectors. In banking and digital payments, AI is widely used for fraud detection and continuous account monitoring. In healthcare, it helps safeguard sensitive medical records and detect suspicious access patterns. In e-commerce, AI supports transaction monitoring and identifies abnormal user behavior to prevent fraud. In cloud environments, it is applied to traffic analysis, intrusion detection, and automated incident response. Additionally, AI is extensively used in phishing detection, spam filtering, malware classification, and endpoint behavior analytics. These applications demonstrate that AI has moved beyond theoretical concepts and now serves as a critical component in practical and operational cyber defense.

11. Future Scope

Future work in AI-based cybersecurity is likely to emphasize explainability, privacy-preserving learning, stronger IoT security, and semi-autonomous defense systems. Explainable AI is essential because analysts must understand why a model has labeled an event as malicious. Federated and privacy-preserving learning approaches can enable collaboration without exposing raw sensitive data. AI for Internet of Things security is particularly important because IoT devices often have weak built-in defenses but create large attack surfaces. Hybrid systems that combine signatures, anomaly detection, and deep learning may provide stronger long-term defense than any single method alone. Over time, autonomous defense mechanisms may handle low-risk actions automatically while humans remain responsible for high-risk decisions and investigations.

12. Conclusion

Artificial Intelligence has become a powerful tool in modern cybersecurity, as it enhances the ability to detect evolving and previously unseen threats. Techniques such as machine learning, deep learning, and anomaly detection each offer distinct advantages for cyber defense, ranging from efficient classification to advanced behavioral analysis. The proposed architecture in this paper demonstrates how these AI methods can be integrated into a practical threat detection pipeline, while the mini experiment highlights both the potential and current limitations of a Random Forest-based approach.

Although the experiment produced moderate results, it still supports the broader perspective that AI-driven systems are more adaptable than traditional fixed-rule mechanisms. Future work should focus on improving dataset quality, reducing false positive rates, enhancing model explainability, and increasing robustness against adversarial attacks. With these advancements, AI-based systems can serve as a more reliable and effective foundation for next-generation cybersecurity defense.

References

1. S. Dhal, S. K. Sahoo, and S. K. Dash, "AI Enabled Threat Detection: Leveraging Artificial Intelligence for Advanced Cybersecurity Solutions," *IEEE Access*, vol. 12, pp. 173127–173154, 2024.
2. M. Tavallae, E. Bagheri, W. Lu, and A. A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," in *Proc. 2009 IEEE Symp. Comput. Intell. Security Defense Appl. (CISDA)*, Ottawa, ON, Canada, 2009.
3. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. M. O. Afolabi, "AI integration in cybersecurity software: Threat detection and response," *Int. J. Innov. Res. Sci. Stud.*, vol. 8, no. 3, pp. 3907–3921, 2025.
4. A. L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1153–1176, 2016.
5. I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Comput. Sci.*, vol. 2, no. 3, Art. no. 160, 2021.
6. R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, and S. Venkatraman, "Deep Learning Approach for Intelligent Intrusion Detection System," *IEEE Access*, vol. 7, pp. 41525–41550, 2019.