WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 4, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v4n1.24
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Artificial Intelligence in Cybersecurity Threat Detection

Zehan Wang

University of Maryland, College Park, United States

ABSTRACT

With the increasing frequency and complexity of cyberattacks, traditional cybersecurity threat detection methods have been difficult to cope with new types of threats. Artificial Intelligence (AI) technology, with its powerful data processing and pattern recognition capabilities, has gradually become an important tool for enhancing cyber security. This paper aims to explore the application of AI in cybersecurity threat detection, firstly outlining the current status of the development of AI technology in cybersecurity, and then focusing on analyzing the application of core methods such as machine learning and deep learning in threat detection, and discussing the advantages of integrated learning and multimodal methods. Finally, this paper summarizes the current challenges faced by AI technology in the field of cyber security and looks forward to the future development direction. Through the research in this paper, it is expected to provide reference for improving the accuracy and efficiency of cybersecurity threat detection.

KEYWORDS

Artificial Intelligence; Cyber Security; Threat Detection; Machine Learning; Deep Learning; Integrated Learning

1. INTRODUCTION

With the rapid development of information technology, networks have become an indispensable and important part of human society [1]. However, the popularization of networks has also brought about increasingly complex and diverse security threats. Traditional means of network security protection, such as rule-based intrusion detection systems (IDS) and firewalls, although effective in the past, have become inadequate in the face of the complexity and stealth of modern network attacks [2]. Attackers use advanced technological means to bypass traditional detection systems, causing serious data leakage, economic losses, and even threatening national security.

Against this backdrop, artificial intelligence (AI) technology has become a key force in addressing cybersecurity challenges by virtue of its advantages in data analysis, pattern recognition and automated processing [3]. By applying technologies such as machine learning and deep learning, cybersecurity threat detection systems are able to more accurately identify anomalous behavior, detect unknown threats, and adapt to new attack patterns in real time. This not only improves the accuracy of threat detection, but also greatly reduces the false alarm rate [4].

The purpose of this paper is to systematically explore the application and development of artificial intelligence technology in cyber security threat detection [5]. First, this paper will outline the overall status of AI application in cybersecurity. Subsequently, it will focus on analyzing the core methods of AI technology in threat detection, including machine learning, deep learning and integrated learning, and introduce the application of multimodal data processing methods. Finally, this paper will discuss the challenges and future development direction of AI technology in network security

applications [6]. Through the research in this paper, we hope to provide theoretical support and practical reference for improving network security protection capability.

2. AN OVERVIEW OF THE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN CYBERSECURITY

Artificial Intelligence (AI), a technology that can mimic human intelligence, has been widely used in several fields in recent years [7]. Its core includes technologies such as machine learning, deep learning, and natural language processing, which are capable of recognizing patterns, predicting future trends, and making automated decisions by analyzing large amounts of data. In the field of cybersecurity, the application of AI technology has significant advantages, especially in handling large-scale, complex data and real-time threat detection [8]. With the advancement of computing power and algorithms, AI technology is gradually moving from academic research to practical application, becoming an indispensable tool in the field of cybersecurity.

As information technology advances, the types and sophistication of cybersecurity threats are evolving. While traditional threats such as viruses and worms rely on signature libraries for detection, modern Advanced Persistent Threats (APTs), zero-day attacks, and social engineering attacks are more insidious and difficult to detect [9]. Attackers are becoming more sophisticated, such as using encrypted traffic, bypassing traditional protections, and exploiting zero-day vulnerabilities to carry out their attacks, posing a huge challenge to traditional cybersecurity protection systems. These changes have prompted the cybersecurity field to continuously seek smarter and more automated solutions [10].

With the increasing sophistication of cyber-attacks, AI technologies are gradually being introduced into cybersecurity to improve threat detection and response capabilities. Initially, AI technology was mainly used in malware detection, spam filtering, and intrusion detection systems (IDS) to automatically classify and detect abnormal behaviors through machine learning algorithms. These initial applications have shown that AI technology can effectively compensate for the shortcomings of traditional security measures and improve detection accuracy and response speed.

Currently, AI technology has penetrated into all aspects of cybersecurity, from the collection and analysis of threat intelligence to real-time threat detection, response and recovery.AI-driven security solutions can automate the processing of massive security data, detect and predict potential threats in real time, and provide intelligent countermeasures. For example, deep learning models can process complex network traffic data and identify anomalous patterns; reinforcement learning algorithms can be used to optimize security policies and dynamically adjust protective measures. As AI technology continues to advance, the cybersecurity industry is also exploring and practicing new application scenarios to address changing security threats.

3. CORE APPROACHES TO ARTIFICIAL INTELLIGENCE TECHNOLOGY IN THREAT DETECTION

In cybersecurity threat detection, AI technology greatly improves the accuracy and efficiency of detection through its powerful data analysis and pattern recognition capabilities. The core methods mainly include machine learning, deep learning, and integrated learning and multimodal methods. Each of these techniques has unique advantages, with machine learning being able to learn threat features from large amounts of labeled data, deep learning processing complex unstructured data through multi-layer neural networks, and integrated learning and multimodal approaches further improving the robustness and adaptability of detection by combining multiple data sources and algorithms. These three core approaches are discussed in detail below. Binary Classification Loss (Cross-Entropy Loss):

$$\text{Loss} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \quad (1)$$

3.1. Machine Learning in Threat Detection

Supervised learning is one of the most common approaches in machine learning that relies on labeled datasets to train models to identify known threat types in new data. In cybersecurity, supervised learning is widely used in scenarios such as malware detection, spam filtering, and intrusion detection. With a large amount of labeled normal and anomalous data, the model is able to learn and recognize the characteristics of malicious behaviors, thus effectively distinguishing between normal traffic and threat activities in real-time detection. The advantages of this approach are its high accuracy and interpretability, but its effectiveness depends on the quality and quantity of labeled data, and may have some limitations for new or unknown threats.

Unsupervised learning does not rely on labeled datasets, but rather analyzes the intrinsic structure of the data to discover hidden patterns and anomalous behaviors. In cybersecurity, unsupervised learning is mainly used for anomaly detection and intrusion detection, and is especially suitable for detecting threat types that are not predefined. Common unsupervised learning methods include cluster analysis and anomaly detection techniques such as K-means, isolated forests, and selfencoders. These methods are able to identify anomalous activities that differ from normal behavioral patterns and thus detect potential security threats. The advantage of unsupervised learning lies in its ability to detect unknown threats, but it also suffers from a high rate of false positives, which needs to be optimized in combination with other techniques.

Reinforcement learning is a dynamic decision-making machine learning method for scenarios that require continuous learning and adaptation to environmental changes. In network security threat detection, reinforcement learning continuously adjusts the detection strategy to optimize the security protection effect by interacting with the environment. For example, reinforcement learning can be used to dynamically adjust the alert thresholds of an intrusion detection system or to optimize the defense strategy in the face of changing attack techniques. Compared with traditional methods, reinforcement learning is highly adaptive and can update detection strategies in real time according to new threat intelligence, improving the flexibility and real-time network security protection, showed in Figure 1:

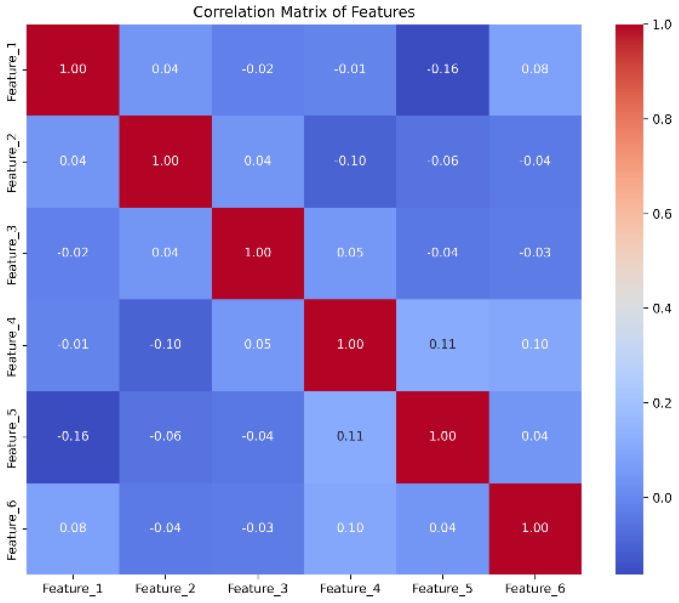

Figure 1. Correlation Matrix of Features

Transfer learning is a technique that utilizes learned knowledge to cope with new tasks, and is particularly suitable for application in situations where data is scarce or training costs are high. In the

cybersecurity domain, migration learning can rapidly improve the effectiveness of detection models by applying threat detection knowledge learned from one domain (e.g., financial security) to another (e.g., healthcare security). Migration learning helps to build efficient threat detection models with insufficient data, while also reducing model training time and computational resource consumption. However, the challenge of migration learning lies in the need to carefully select the similarity between the source and target domains to ensure the effectiveness of knowledge migration. Gradient Descent Update Rule:

$$\theta_{t+1} = \theta_t - \alpha \nabla_\theta J(\theta) \tag{2}$$

3.2. Deep Learning in Threat Detection

Deep learning is able to automatically extract and learn features of complex data through the architecture of multi-layer neural networks, giving it a significant advantage in cybersecurity threat detection. Convolutional Neural Networks (CNNs) are commonly used to process image and traffic data, and can effectively detect features such as malicious code or anomalous network traffic. Recurrent Neural Networks (RNN) and its improved version, Long Short-Term Memory Networks (LSTM), on the other hand, are good at processing time-series data and are suitable for application scenarios such as analyzing network behavior logs and detecting persistent threats. Through deep learning of large-scale data, these neural network models are able to identify threats that are difficult to detect by traditional methods, improving detection accuracy and response speed.

An auto-encoder is an unsupervised learning model that learns a latent representation of data by compressing and reconstructing the input data. In cybersecurity, auto-encoders are commonly used for anomaly detection, such as identifying abnormal network traffic or user behavior patterns. When the input data differs significantly from normal data patterns, the reconstruction error increases significantly, thus suggesting possible threats. Generative Adversarial Networks (GANs), on the other hand, consist of a generator and a discriminator for generating fake data similar to real data and enhancing the capability of the detection model through adversarial training. GANs can be used in cybersecurity not only for generating attack samples to enhance the robustness of the detection model, but also for spoofing the detection system to conduct adversarial tests to evaluate and improve the system's protection capability.

While deep learning models typically require large amounts of data and computational resources for training, migration learning can quickly build effective threat detection models with less data by utilizing pre-trained models. Pre-trained models are typically trained on large-scale datasets and then applied to specific cybersecurity tasks such as malware classification or anomalous traffic detection. Through migration learning, deep learning models are able to inherit the knowledge of pre-trained models, quickly adapt to new threat scenarios, and improve detection results. This approach not only saves training time and resources, but also improves the generalization ability of the model, enabling it to cope with diverse security threats.

Graph Neural Network (GNN) is a deep learning model specialized in processing graph-structured data, which is particularly suitable for network topology analysis and threat detection. In the field of cybersecurity, GNNs are able to process complex structured data such as social networks and computer networks to identify potential threats by capturing the relationships between nodes. For example, GNN can be used to detect anomalous communication patterns in a network and discover latent malicious nodes or attack paths. Compared to traditional methods, GNN can better understand the relationships in complex network structures and provide more accurate threat detection results, providing new ideas and tools for network security protection.

3.3. Integrated Learning and Multimodal Approaches

Integrated learning improves the overall detection accuracy and robustness by combining the prediction results of multiple models. In cybersecurity threat detection, integrated learning can synthesize the advantages of different algorithms, such as integrating multiple models such as decision trees, random forests, gradient boosting trees, etc., in order to improve the detection of complex threats. The integrated learning approach can effectively reduce the false alarm rate of a single model and enhance the adaptability to various threats through the complementarity of different models. For example, in malware detection, integrated learning can synthesize multiple feature extraction methods and classifiers to improve the accuracy and generalization ability of detection.

The data involved in network security threat detection is usually multimodal, including network traffic data, system logs, user behavior, geolocation, etc. Multimodal approaches enable a more comprehensive understanding and detection of potential threats by fusing information from different data sources. For example, combining network traffic data with user behavior data can more accurately identify anomalous behavior and potential attack patterns. Multimodal data fusion can improve the detection system's perceptual ability and decision-making accuracy, and form a multidimensional understanding of complex threats by comprehensively analyzing information from different modalities, thus enhancing the overall effect of threat detection.

Heterogeneous integration methods are an extension of integration learning that combines different types of models and diverse data sources to significantly improve the effectiveness of threat detection. Compared with traditional isomorphic integration methods (e.g., the integration of multiple similar models), heterogeneous integration methods have more processing power by integrating the advantages of different models (e.g., deep learning models and traditional machine learning models), especially when dealing with unknown threats and complex attacks. For example, by using deep learning models for feature extraction and feeding their outputs into integrated learning models for final decision making, this heterogeneous integration strategy can fully utilize the powerful feature representation capability of deep learning and the decision making capability of integrated learning to achieve more accurate threat detection, showed in Figure 2 :

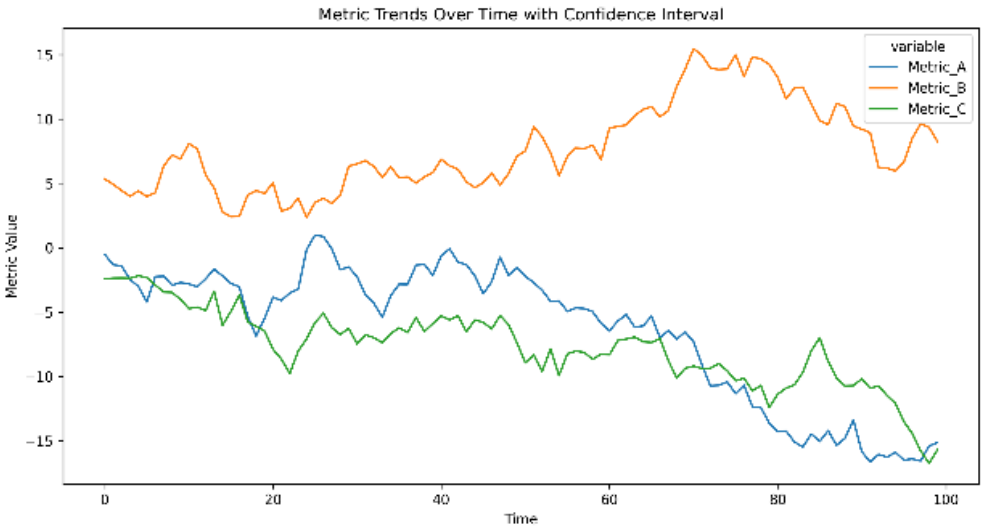

Figure 2. Metric Trends Over Time with Confidence Interval

Multimodal deep learning can better address complex cybersecurity threats by processing multiple data modalities simultaneously. In practical applications, multimodal deep learning models can fuse multiple data types such as text, images, audio, video, etc. to comprehensively analyze and detect threats. For example, in APT (Advanced Persistent Threat) detection, multimodal deep learning can combine network traffic, log analysis, and user behavior to provide a global view and identify long-lurking threats. The advantage of multimodal deep learning is that it can fuse multiple information

sources to form a comprehensive judgment of the threat and improve the accuracy and robustness of detection, especially in the face of complex attacks and multi-dimensional data.

4. CHALLENGES AND FUTURE DEVELOPMENTS OF ARTIFICIAL INTELLIGENCE IN CYBERSECURITY THREAT DETECTION

In the process of utilizing AI for cybersecurity threat detection, data privacy and security become the primary challenges. Since threat detection relies on a large amount of data such as network traffic, user behavior, system logs, etc., these data may contain sensitive information. How to protect these data from leakage or misuse during model training and deployment is an urgent problem to be solved. Currently, privacy-preserving techniques such as differential privacy and federated learning are gradually being applied to the field of cybersecurity, but how to ensure the accuracy and efficiency of the detection model while safeguarding data privacy is still a direction that needs to be further investigated.

Artificial intelligence models excel in threat detection, but also face the threat of adversarial attacks. An attacker can induce the model to make wrong judgments by generating adversarial samples or otherwise, leading to the failure of the security system. Such adversarial attacks not only weaken the effectiveness of the detection system, but may also be exploited to circumvent security defenses. Therefore, enhancing the robustness of AI models. Researchers are exploring how to improve the reliability of threat detection systems by enhancing the model's resistance to attacks, designing more secure architectures, and developing protection mechanisms. Anomaly Detection Using Autoencoder (Reconstruction Error):

$$\text{Reconstruction Error} = \frac{1}{N}\sum_{i=1}^{N}|x_i - \hat{x}_i|^2 \tag{3}$$

Artificial intelligence models, especially deep learning models, are often regarded as "black boxes" whose decision-making processes are difficult to explain. In cybersecurity, where threat detection results often need to be validated and interpreted by security experts, model interpretability is critical. In addition, as more and more industries and domains begin to adopt AI for security detection, compliance requirements are becoming more stringent. How to improve the performance of models while enhancing their interpretability to comply with various types of security and privacy regulations is a key issue to be addressed in the future. Explainable Artificial Intelligence (XAI) and model visualization tools are gradually being introduced to help security personnel understand and trust detection results.

The application of AI in cybersecurity threat detection remains promising, and in the future, as technology continues to advance, threat detection systems will become more intelligent and automated, and will be able to respond to new and complex attacks faster and more accurately. Cross-field multidisciplinary cooperation will facilitate the emergence of more innovative applications, such as threat detection solutions that combine artificial intelligence with quantum computing and blockchain technology. In addition, the application of AI in cybersecurity will become safer, more reliable, and more efficient with the continuous development of counter-attack protection techniques, privacy protection mechanisms, and interpretable models. In the future, AI will not only be a tool for threat detection, but will also become a core component in cybersecurity strategy, pushing the entire industry toward a higher level of security.

5. CONCLUSION

The application of artificial intelligence technology in cybersecurity threat detection offers new possibilities for improving the efficiency and accuracy of detection. By utilizing a variety of methods such as machine learning, deep learning, and integrated learning, AI is able to effectively respond to

increasingly complex cyberattacks, discover potential threats in a timely manner, and provide more intelligent solutions for security protection. However, with the in-depth application of AI technology, challenges such as data privacy, anti-attacks, and interpretability are gradually emerging, requiring us to pay attention to security and compliance issues along with technological innovation.

The role of artificial intelligence in cybersecurity will become more important. With the development of technology and cross-disciplinary cooperation, threat detection systems will become more intelligent, automated and interpretable, thus providing stronger security in the complex and changing cyber environment. To achieve this goal, there is a need to continuously explore new technologies, optimize existing methods, and strengthen comprehensive research on AI applications to ensure its continued development and effective application in the field of cybersecurity. Through sustained efforts, AI will become a core force in the network security protection system, laying a solid foundation for building a more secure and reliable network environment.

REFERENCES

[1] Khan N F, Ikram N, Murtaza H, et al. Social media users and cybersecurity awareness: predicting self-disclosure using a hybrid artificial intelligence approach [J]. Kybernetes, 2023, 52(1):401-421. DOI:10.1108/K-05-2021-0377.

[2] Cheng K S, Pan R, Pan H, et al. ALICE: a hybrid AI paradigm with enhanced connectivity and cybersecurity for a serendipitous encounter with circulating hybrid cells [J]. Theranostics, 2020, 10(24):11026. DOI:10.7150/thno.44053.

[3] Iwendi C, Rehman S U, Javed A R, et al. Sustainable Security for the Internet of Things Using Artificial Intelligence Architectures [J]. ACM Transactions on Internet Technology, 2021, 21(3):1-22. DOI:10.1145/3448614.

[4] Zhao L, Zhu D, Shafik W, et al. Artificial intelligence analysis in cyber domain: A review: [J]. International Journal of Distributed Sensor Networks, 2022, 18(4):121-131. DOI:10.1177/15501329221084882.

[5] Perdisci R, Giacinto G, Roli F. Alarm clustering for intrusion detection systems in computer networks [J]. Engineering Applications of Artificial Intelligence, 2006, 19(4):429-438. DOI:10.1016/j.engappai.2006.01.003.

[6] DeBenedictis, Erik P. Plotting a Socially Responsible Course for Computers Using Cybersecurity as an Example [J]. Computer, 2017, 50(12):86-90. DOI:10.1109/MC.2017.4451217.

[7] Sedjelmaci H, Guenab F, Senouci S M, et al.Cyber Security Based on Artificial Intelligence for Cyber-Physical Systems [J]. IEEE Network, 2020, 34(3):6-7. DOI:10.1109/MNET.2020.9105926.

[8] Silvestri S, Islam S, Amelin S C M. Cyber threat assessment and management for securing healthcare ecosystems using natural language processing [J]. International Journal of Information Security, 2024, 23(1):31-50.

[9] Alohali M A, Al-Wesabi F N, Hilal A M, et al.Artificial intelligence enabled intrusion detection systems for cognitive cyber-physical systems in industry 40 environment [J]. Cognitive neurodynamics, 2022, 16(5):1045-1057. DOI:10.1007/s11571-022-09780-8.

[10] Xu S, Qian Y, Hu R Q. Data-Driven Network Intelligence for Anomaly Detection [J]. IEEE Network, 2019, 33(3):88-95. DOI:10.1109/MNET.2019.1800358.