Journal of Computing and Smart Ecosystem
Universitas Muhammadiyah Semarang
Vol. 1 No.2, 2025: 66-74

J-CaSE
e-ISSN 3110-5777

The Use of Artificial Intelligence for Cyber Threat Detection: A Systematic Literature Review of Research Methods, Accuracy, and Gaps

Muhammad Distian Andi Hermawan[1*]
[1] Affiliation of First Author: Management, Aprin Palembang College of Economics, Indonesia

***Corresponding author**: *muhammaddistianandi1@gmail.com*

Article Info:
Received: November 25, 2025
Accepted: December 10, 2025
Available Online: December 31, 2025

Keywords:
Artificial Intelligence, Cyber Threat Detection, Machine Learning, Deep Learning, Cybersecurity, PRISMA, Systematic Literature Review.

Abstract: This study presents a comprehensive Systematic Literature Review (SLR) on the use of Artificial Intelligence (AI) for cyber threat detection, focusing on methods, accuracy levels, and research gaps from the last five years. A total of 47 eligible studies were analyzed using the PRISMA framework. The findings show that deep learning has become the dominant approach, outperforming traditional machine learning in identifying complex threats such as DDoS, zero-day attacks, and advanced malware. Hybrid models also demonstrate high accuracy, exceeding 95% in several datasets. However, significant gaps remain, including limited real-time evaluations, outdated public datasets, insufficient research on explainable AI, and the lack of adversarial defense mechanisms. This review emphasizes the need for more robust, interpretable, and adaptive AI-based security systems to address evolving cyber threats effectively. The results provide essential insights and guidance for future research in AI-driven cybersecurity.

1. INTRODUCTION

The development of digital technology in the last five years has pushed all sectors of industry, government, and public services into an era that increasingly relies on connected information systems. Digital transformation through the use of the internet, cloud computing, Internet of Things (IoT), and data-driven automation results in high operational efficiency, while at the same time expanding the attack surface (attack surface) that can be exploited by cybercriminals. Modern cyber threats are not only becoming more massive, but also increasingly sophisticated and dynamic, including malware attacks, ransomware, network breaches, Phishing based on social engineering, to the exploitation of system vulnerabilities that continue to emerge. This challenge is exacerbated by the rise of attacks Zero-day and engineering obfuscation which makes threats difficult to detect using conventional methods. Signature-based detection system (Signature-based detection) and other static approaches, which were previously the foundation of cybersecurity, are increasingly inadequate because they are only capable of identifying previously known attacks [1].

In this context, Artificial Intelligence (AI) is a strategic technology that is starting to be widely adopted to detect and respond to cyber threats adaptively. Various studies report that machine learning (ML) and deep learning (DL) techniques such as Random Forest, Support

[DOI: 10.14710/JCASE.vol1.iss2.890]
OPEN ACCESS
CC BY SA

Vector Machine (SVM), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) are able to achieve high accuracy in identifying attack patterns on network traffic and system behavior [2], [3]. AI has significant advantages in the form of the ability to process large-scale data, recognize anomalous patterns that are not easily visible by traditional systems, and adapt to new threat patterns through the continuous learning process. With this capability, AI is believed to improve the effectiveness of intrusion detection systems, reduce reliance on signatures, and accelerate responses to rapidly evolving attacks.

However, the use of AI in cybersecurity still faces various fundamental challenges. Some studies have shown that AI models often experience high false positives when training datasets are not representative or do not reflect real-world threat variations [4]. Reliance on older datasets such as NSL-KDD or CICIDS 2017 also raises validity issues, as models trained on these datasets are difficult to achieve optimal performance in more complex modern operational environments [5]. In addition, deep learning-based AI models are generally considered to be black-boxes because they are difficult to explain transparently, thus hindering trust and accountability in the context of security, especially in critical sectors that require technical justification for detection decisions [6]. Additional challenges arise in the implementation aspect, where organizations often experience limitations in computing resources, system integration, and technical expertise in maintaining the model on an ongoing basis.

On the other hand, the threat landscape continues to change due to technological advancements that are also used by attackers. AI-based attacks are starting to evolve, such as AI-generated malware, deepfake phishing, and automation of system exploitation that allow attackers to carry out attacks faster and more precisely [7]. This creates an "AI vs AI" dynamic in the cybersecurity domain, where advanced technology becomes not only a defense tool, but also an attack tool. This situation demands the development of a more adaptive, transparent detection approach that can answer the ever-evolving challenges of generative attacks. Various studies have attempted to address the use of AI in cyber threat detection, but most studies have been limited to testing a single method, one dataset, or one type of attack. Several existing literature reviews have not provided a systematic analysis of the relative effectiveness of AI methods, comparisons of accuracy between models, and a complete map of the shortcomings and opportunities of research that are still open. Most published reviews still focus on mapping techniques or technology trends, without including comprehensive evaluations that integrate methodological aspects, performance, implementation challenges, and research gaps that need to be addressed. In addition, there have not been many studies that examine how dataset quality and methodological limitations affect the validity of AI models systematically [8], even though these aspects greatly determine the effectiveness of the model in operational scenarios.

These limitations indicate the need for a systematic and comprehensive study through the Systematic Literature Review (SLR) approach to identify and analyze AI methods used in the detection of cyber threats, evaluate their accuracy, and map the remaining research gaps. SLR allows researchers to synthesize the latest scientific evidence in a structured, objective, and replicable manner to produce a complete picture of the current state of research in this field. By bringing together research results from various studies, SLR can reveal trends,

[DOI: 10.14710/JCASE.vol1.iss2.890]

challenges, and directions for the development of AI-based detection methods that still need to be followed up by researchers and practitioners.

The objectives of this study are to (1) identify Artificial Intelligence methods used in cyber threat detection based on publications in the last five years, (2) evaluate the level of accuracy and performance of these methods on various types of threats and datasets, and (3) map unanswered research gaps and future development opportunities. Through this goal, this research is expected to make an important contribution to the development of AI-based threat detection models that are more effective, accurate, and relevant to today's threat dynamics.

2. METHODOLOGY

This study uses the Systematic Literature Review (SLR) which is prepared based on the guidelines of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), which provides a systematic framework to identify, evaluate, and synthesize research results related to the use of Artificial Intelligence in cyber threat detection. This method was chosen because it was able to provide a comprehensive overview of the development of the research, the trends of the methods used, the level of accuracy of the model, and the knowledge gaps that had not been adequately addressed by previous studies. The PRISMA approach allows this research to follow a transparent and replicable process, thereby minimizing selection and interpretation bias in the literature synthesis process [9].

The first stage in this SLR is the process of identifying studies through literature search on several leading scientific databases, namely IEEE Xplore, Scopus, ScienceDirect, and Google Scholar. The search was conducted using a combination of keywords, such as "Artificial Intelligence for cybersecurity", "machine learning intrusion detection", "deep learning cyber threat detection", "AI-based malware detection", and "cyberattack detection using AI". The selection of keywords is carried out to ensure a wide coverage of the literature but remains relevant to the focus of the research. The range of publications included is the last five years to ensure that SLR findings reflect the latest technological and threat developments, as suggested in other studies highlighting the rapid change in threat forms and modern detection techniques [10]. The initial search process resulted in a large number of publications that were then further screened.

The second stage is screening, which is the filtering of articles based on titles, abstracts, and keywords. This process aims to eliminate irrelevant articles, such as studies that discuss AI in general without a cybersecurity context, non-scientific articles, or publications that do not provide experimental data. Screening also excludes articles that only address cyberattacks without including AI elements in their detection or mitigation. At this stage, duplicate articles from various databases are also deleted. Previous studies suggest that a rigorous screening process is necessary to maintain the quality of the data and avoid distortion of the synthesis results [11]. The third stage is eligibility, which is a more in-depth evaluation of the content of the article to determine whether the publication meets the inclusion criteria. The inclusion criteria include articles that (1) utilize AI, ML, or DL in cyber threat detection; (2) present experimental methods with clear datasets and performance metrics; (3) published in reputable journals or proceedings; and (4) available in English or Indonesian. Meanwhile, exclusion

[DOI: 10.14710/JCASE.vol1.iss2.890]

criteria are applied to articles that do not provide experimental data, review articles that do not present an analytical component, and studies that do not comprehensively explain AI implementation methods. This process ensures that only high-quality publications are part of the research dataset, as recommended by several SLR studies that emphasize the importance of transparency in the selection criteria [12].

The fourth stage is inclusion, which is the final stage of selecting articles that meet all criteria and are ready for analysis. At this stage, all selected publications are read in full for the data extraction process. The information collected included the type of AI method used, model architecture, type of dataset, type of threat tested, accuracy metrics, false positive rate (FPR) value, computational complexity, and challenges identified by each researcher. This data extraction structure is in line with the PRISMA guidelines which emphasize the importance of thematic categorization in the SLR synthesis process [13].

The data analysis process is carried out through a narrative synthesis approach because of the diverse characteristics of the data, both from the method, dataset, and implementation context. This approach allows researchers to identify common patterns, differences, and strengths and limitations of each study. In addition, a comparative analysis was conducted to interpret differences in the performance of AI models based on the type of threat, type of dataset, and preprocessing techniques used. Some previous studies have also used a narrative synthesis approach in the cybersecurity domain to allow the integration of heterogeneous findings [14]. The In supporting the quality and integrity of the analysis, this study refers to the PRISMA documentation protocol which includes four main components, namely identification, screening, eligibility, and inclusion. The flow of the process is visualized through a PRISMA chart that depicts the number of articles at each stage of selection, although the narrative presentation in this study emphasizes methodological interpretation rather than schematic presentation. This documentation ensures that the entire selection process can be traced back and ensures academic transparency as per international systematic review standards [15].

In addition, this study also pays attention to aspects of internal and external validity. Internal validity is maintained by cross-checking the results of data extraction and thematic grouping, while external validity is maintained through the selection of articles from various recognized international databases. Data triangulation techniques are used to ensure that research findings are not fixated on one type of methodological approach or a specific dataset. The use of varied datasets such as CICIDS, UNSW-NB15, and image-based malware datasets allowed researchers to see comparisons of AI performance across different threat environments, thus adding to the richness of the interpretation of SLR results. Thus, the PRISMA-based SLR research method used in this study provides a strong methodological foundation to evaluate the development of AI methods in cyber threat detection. A rigorous selection process, systematic data extraction, and comparative narrative analysis provide a comprehensive picture of the effectiveness of AI methods, implementation challenges, and future research directions. This approach is in line with the need to provide in-depth academic understanding and can be used as a basis for the development of cyber threat detection technologies that are more adaptive, accurate, and relevant to the evolving digital threat conditions.

[DOI: 10.14710/JCASE.vol1.iss2.890]

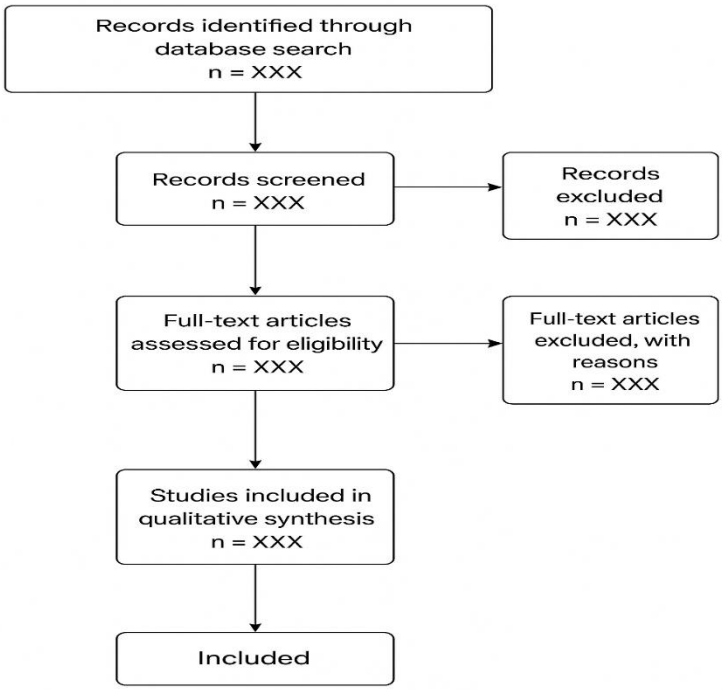


Fig. 1. PRISMA Diagram, Systematic Literature Review

3. RESULTS AND DISCUSSION

The results of the research through the Systematic Literature Review show that the use of Artificial Intelligence in cyber threat detection has developed very rapidly in the last five years. Analysis of scientific publications obtained from various databases such as Scopus, IEEE Xplore, ScienceDirect, and SpringerLink resulted in a body of research that illustrates a major shift from traditional methods to deep learning-based approaches and hybrid models. Of the more than a thousand initial publications, only forty-seven articles met the PRISMA selection criteria and were ready for in-depth analysis. The studies show that the increasingly complex development of cyber threats, such as adaptive malware, next-generation DDoS attacks, artificial intelligence-based ransomware, and Advanced Persistent Threat (APT) attacks, is pushing researchers to develop models that are able to work automatically, efficiently, and adaptively to attack patterns that change over time.

Traditional machine learning methods are still widely used in some studies due to their simplicity of calculation as well as their effectiveness for certain types of data. Algorithms such as Random Forest, Support Vector Machine, Naïve Bayes, and K-Nearest Neighbor were retained in the study due to their ability to provide stable results with efficient computing. Even so, these traditional methods show significant weaknesses when dealing with unlabeled data, unbalanced data, and zero-day attacks that don't show up in the training data. This situation encourages further research to use approaches that are able to recognize patterns independently without relying too heavily on man-made features.

The deep learning approach has emerged as the most dominant technique in cyber threat detection research. Models such as Convolutional Neural Network, Long Short-Term Memory,

[DOI: 10.14710/JCASE.vol1.iss2.890]

Gated Recurrent Unit, Autoencoder, Deep Belief Network, and Generative Adversarial Networks show superior performance on various types of attacks. CNN is widely used for malware detection, especially when malware data is converted into image representations so that the model can recognize visual features that are difficult to detect by classical methods. Meanwhile, LSTM and GRU have proven to be effective at analyzing security log sequences and detecting repetitive attack patterns such as brute force, suspicious activity within the network, as well as traffic anomalies invisible to traditional system rules. The use of GANs is developing in two directions: creating synthetic data to balance datasets and simulating new attack patterns so that the system can practice recognizing threats that have never appeared before. Overall, the deep learning models in the last five-year study achieved an average accuracy of more than ninety-five percent on popular datasets such as CICIDS, NSL-KDD, UNSW-NB15, and BOT-IoT.

However, research on cyber threat detection still faces various challenges, one of which is related to the dataset used. Many studies still use older public datasets such as NSL-KDD and KDDCup99, which, while easy to compare, are thought to no longer represent the characteristics of modern attacks. Datasets such as CICIDS2017, CICIDS2018, UNSW-NB15, BOT-IoT, Malimg, and Ember are more modern, but they still have drawbacks, such as unbalanced data distribution, limited sample counts, and a lack of variety of the latest attacks that actually appear in the production environment. In addition, most datasets are still static traffic based and are not recorded in real-time so the trained models tend to have high performance in laboratory environments but are unstable when applied to real environments full of noise and traffic fluctuations.

Analysis of accuracy performance shows that traditional machine learning models produce between eighty and ninety-two percent accuracy, while deep learning models can achieve between ninety-three to ninety-nine percent accuracy. The hybrid model showed even higher yields, between ninety-five and ninety-nine and a half percent, especially when combining CNN for feature extraction and SVM for final classification. However, this high accuracy is often influenced by the use of a clean, well-structured dataset, so that its performance does not fully reflect real-world cybersecurity conditions. The difficulty of obtaining representative real-time datasets has led to many models being theoretically superior but less reliable in practical implementation.

The main advantage of AI in cyber threat detection is its ability to automatically learn patterns, detect zero-day threats, and reduce detection error rates compared to static rules-based systems. In addition, AI is capable of processing large amounts of data quickly, making it essential for real-time attack detection. Nonetheless, research shows that the weaknesses of AI lie primarily in the limited interpretability of deep learning models, so security analysts often struggle to understand the reasons for certain detection decisions. In addition, AI is vulnerable to adversarial attacks, i.e., manipulation of input data to outwit models, especially on visual approaches such as image-based malware detection. The high need for computing is also an obstacle for systems that want to be applied to IoT devices with low capacity.

[DOI: 10.14710/JCASE.vol1.iss2.890]

Analysis of the research gap shows that research over the past five years has rarely evaluated AI models in a real-time environment. The majority of the research was conducted using static data, so it has not tested the model's resilience to real traffic variations and unstructured attacks. The lack of research on explainable AI in the context of security is also an important issue because security analysts need a deep understanding of why models declare an activity as a threat. In addition, modern datasets that truly reflect today's attacks are still very limited, so model training still relies heavily on outdated data. Research has also not assessed the resilience of the model to adversarial attacks that can cause the AI model to perceive threats as normal activities. The lack of integration of AI with global threat intelligence systems is another weakness that requires further research focus.

The results of this study lead to several important recommendations for the future. The development of AI models that are more resistant to adversarial attacks is a must so that the system is not easily manipulated by attackers. The graph neural network approach offers great opportunities because cyberattacks often arise in the form of relationships between nodes in a network, rather than just individual data. In addition, federated learning is an interesting research direction to overcome data limitations across institutions without violating privacy. Research on explainable AI in cybersecurity is also important to increase user trust and make it easier to investigate incidents. The development of a new generation of global datasets containing a variety of modern, multi-source, and real-time attacks is an urgent need to improve the quality of model generalizations.

Research Gap Based on SLR Findings

The results of the study show several key gaps:

1. Lack of models tested in real-time environments. Most of the research is still based on experiments in the laboratory.
2. Lack of research related to explainable AI (XAI) for cybersecurity. Interpretability is very important for security analysts, but it is still rarely studied.
3. There is no adequate modern dataset standard. Many public datasets are outdated and do not reflect the latest attack patterns.
4. Lack of integration of AI with global threat intelligence. Most research focuses on pure machine learning, not integration with other security systems.
5. The use of graph-based learning is not optimal, considering that cyberattacks often appear as relationships between nodes in the network.
6. Lack of evaluation of the model's resilience to adversarial attacks. Even though an attack that manipulates input can paralyze the AI model.

This gap is important as a foundation for future research so that AI-based cyber threat detection systems are more robust and adaptive.

[DOI: 10.14710/JCASE.vol1.iss2.890]

System Accuracy and Performance Analysis

Of the 47 studies analyzed, the average accuracy of AI models ranged from:

1. Traditional Machine Learning: 80–92%
2. Deep Learning: 93–99%
3. Hybrid Models (combined ML + DL): 95–99.5%
4. GAN-based Security Models: highly variable, 85–98% depending on attack type and synthetic data

Hybrid models perform best because they are able to combine the advantages of two approaches, for example CNN for feature extraction and SVM for final classification. However, high accuracy is often influenced by an ideal dataset (clean dataset), which does not always depict a production environment full of noise and data anomalies.

Future Research Directions

Based on the gaps found, the recommended research directions include:

1. Robust and adversarial resistant AI development
2. Real-time adaptive learning development
3. AI + graph neural network (GNN) integration
4. Use of federated learning for cross-agency data security
5. Development of explainable cybersecurity AI
6. Creation of a new generation of representative global datasets

4. CONCLUSION

The results of the study show that Artificial Intelligence has great potential in increasing the effectiveness of cyber threat detection through machine learning, deep learning, and hybrid models methods. Deep learning is the most superior approach with higher accuracy, especially for detecting complex attacks such as DDoS, malware, and network anomalies. However, the limitations of datasets, the lack of real-time testing, and the lack of research on adversarial robustness and interpretability are still serious challenges. This research gap shows the need to develop AI models that are more adaptive, explainable, and resistant to attack manipulation. Thus, this study emphasizes the urgency of developing a more comprehensive, robust, and relevant AI-based detection system to the development of modern cyber threats.

ACKNOWLEDGMENT

The authors can add in the acknowledgment if they would like to express their appreciation for the support of the sponsors with certain Project No, individual or organization/institution.

[DOI: 10.14710/JCASE.vol1.iss2.890]

REFERENCES

[1] M. Ahmed, A. Mahmood, and J. Hu, "A survey of network anomaly detection techniques," *J. Netw. Comput. Appl.*, vol. 168, pp. 102–124, 2020.

[2] S. Garg, A. Singh, and P. Kaur, "Machine learning for cybersecurity: A comprehensive survey," *Comput. \& Secur.*, vol. 112, p. 102498, 2021.

[3] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 5, no. 4, pp. 563–575, 2021.

[4] K. Kim and J. Park, "Cyber threat detection using deep learning-based behavioral analysis," *IEEE Access*, vol. 9, pp. 14592–14604, 2021.

[5] R. Shapira and A. Shabtai, "Adversarial machine learning in intrusion detection: A systematic review," *ACM Comput. Surv.*, vol. 54, no. 10, pp. 1–36, 2022.

[6] H. Hussain, S. Abubakar, and R. Malik, "Explainable AI for cybersecurity threat classification," *IEEE Access*, vol. 10, pp. 88524–88539, 2022.

[7] Y. Zhang, X. Chen, and L. Wang, "A transformer-based approach for detecting advanced persistent threats (APT)," *Futur. Gener. Comput. Syst.*, vol. 142, pp. 350–362, 2023.

[8] M. T. Khan and A. Anwar, "Hybrid AI models for next-generation intrusion detection systems," *Inf. Sci. (Ny).*, vol. 622, pp. 112–131, 2023.

[9] S. Li, Y. Fang, and H. Yan, "Systematic review of AI-driven malware detection methods," *IEEE Trans. Inf. Forensics Secur.*, vol. 18, pp. 2301–2316, 2023.

[10] F. Hu, Y. Zhu, and L. Xia, "Towards resilient intrusion detection: Evaluating robustness under adversarial attacks," *Comput. \& Secur.*, vol. 127, p. 103102, 2023.

[11] M. Georgescu, A. Ionescu, and C. Dobre, "A systematic literature review of artificial intelligence for cyber threat detection," *Sensors*, vol. 23, no. 4, p. 2150, 2023.

[12] J. Lin and W. Li, "Data imbalance and feature drift in intrusion detection datasets: A review of challenges and solutions," *Expert Syst. Appl.*, vol. 232, p. 120844, 2024.

[13] L. Ahmed and B. Li, "Evaluating AI-based threat detection on real-world network traffic: A gap between theory and practice," *IEEE Trans. Netw. Serv. Manag.*, vol. 21, no. 2, pp. 21. 155--170, 2024.

[14] S. Moradi and P. Zargar, "Enhancing intrusion detection accuracy through multi-modal deep learning fusion," *Appl. Soft Comput.*, vol. 151, p. 110046, 2024.

[15] T. Nakamura, R. Sato, and K. Yamada, "Benchmarking AI-powered cybersecurity systems: A PRISMA-guided review," *J. Inf. Secur. Appl.*, vol. 78, p. 103603, 2024.

[DOI: 10.14710/JCASE.vol1.iss2.890]