JIATIS
Journal of International Accounting, Taxation and Information Systems
https://jiatis.com/index.php/journal
Online ISSN 3048-085X

Artificial Intelligence in Cybersecurity: A Comparative Study of Threat Detection Algorithms

Shir Ahmad Hamidi[1], Ali Mohammad Amiri[2], Hedayatullah Shujaee[3]*

[1,3]Computer Science Faculty, Karwan University, Kabul, Afghanistan

[2]Computer Science Faculty, Zawul Institute of Higher Education, Kabul, Afghanistan

E-mail: [1)] shirahmad1603@gmail.com, [2)] alimohammad.amiri1212@gmail.com, [3)] hedayatullahshujaee@gmail.com

ARTICLE INFO

Article History
Received : 05.05.2025
Revised : 05.06.2025
Accepted : 11.06.2025
Article Type: Literature Review

*Coresponding author:
Shir Ahmad Hamidi
shirahmad1603@gmail.com

ABSTRACT

This paper presents a systematic literature review (SLR) on AI-based algorithms for cybersecurity threat detection, aiming to evaluate the effectiveness and performance differences of various artificial intelligence techniques. The purpose of this study is to provide a comprehensive overview of the most effective AI models for detecting cyber threats and to examine their practical applications across various cybersecurity domains, including IoT, critical infrastructure, and cyber-physical systems. The review includes studies published between 2021 and 2025, sourced from prominent academic databases such as MDPI, SpringerLink, and IEEE Xplore.The methodology employed involved the selection of peer-reviewed articles using inclusion and exclusion criteria, followed by thematic analysis of the AI techniques used in the studies. Key themes such as supervised learning, unsupervised learning, deep learning, and hybrid approaches were explored. Performance metrics including accuracy, precision, recall, F1-score, and false positive rates were used to evaluate algorithm effectiveness. The results highlight the comparative performance of different AI models and provide insights into the strengths and weaknesses of each approach, as well as their suitability for specific cybersecurity applications.The findings emphasize the importance of dataset quality, algorithm transparency, and the need for reducing false positives in real-world applications. The review concludes by recommending the continued development of hybrid AI approaches and the need for more transparent, explainable models.

Keywords: AI-Based Algorithms, Cybersecurity Threat Detection, Machine Learning, Deep Learning, Systematic Literature Review

1. Introduction

In the era of digital transformation, cybersecurity has become a cornerstone of information technology systems. With the rising frequency and sophistication of cyberattacks, traditional rule-based security mechanisms are proving inadequate to address the complexities of modern cyber threats. As a result, artificial intelligence (AI) has emerged as a powerful tool in enhancing cybersecurity through intelligent threat detection, rapid response, and system resilience (Gopalsamy, 2022; Kilincer et al., 2021). The application of AI, including machine learning (ML) and deep learning (DL), has significantly improved the ability to detect anomalies and identify patterns that signify potential cyber intrusions.

AI-driven cybersecurity systems leverage vast amounts of data to learn from previous attacks and adapt to new threat environments. Machine learning models such as Support Vector Machines (SVM), Decision Trees, Random Forests, and Neural Networks have been widely adopted for intrusion detection and threat classification tasks (Barik et al., 2022; Zhang et al., 2022). These models provide high accuracy and scalability, particularly in network-based intrusion detection systems (NIDS) and host-based intrusion detection systems (HIDS). Furthermore, deep learning techniques such as Convolutional Neural Networks (CNNs) and

Recurrent Neural Networks (RNNs) have demonstrated superior performance in handling complex and high-dimensional cybersecurity datasets (Hesham et al., 2024; Salem et al., 2024).

However, the deployment of AI in cybersecurity is not without significant challenges. Issues such as dataset imbalance, vulnerability to adversarial attacks, lack of transparency and explainability, and high computational costs limit the widespread adoption and effectiveness of AI solutions in real-world scenarios (Lysenko et al., 2024; Otoum et al., 2021). Additionally, operational challenges such as integration with existing security infrastructures, scalability across diverse environments, and the ability to operate in real-time pose substantial barriers to practical implementation. Many organizations struggle to seamlessly incorporate AI systems into their current security workflows without disrupting ongoing operations or compromising performance.

Moreover, concerns around model generalizability and the risk of overfitting to specific datasets raise questions about the robustness of AI-based defenses against novel or evolving threats. These limitations highlight the need for cautious and critical integration of AI technologies in cybersecurity frameworks, with attention to both technological capabilities and operational realities.

Despite promising results from individual studies, there remains a need for comprehensive comparative analyses to evaluate the performance, efficiency, and applicability of different AI algorithms across diverse cybersecurity use cases (Al-Suqri & Gillani, 2022; Hernández-Rivas et al., 2024). Furthermore, increased emphasis on real-world implementations and case studies is essential to better understand the practical challenges and benefits of AI adoption in operational cybersecurity environments.

To achieve this, the study sets out three primary objectives. First, it aims to compare the effectiveness of various AI-based algorithms in cyber threat detection. Second, it evaluates their performance using key metrics across different scenarios. Third, it seeks to recommend the most suitable AI techniques for practical cybersecurity applications. These objectives will guide the analysis and contribute to advancing intelligent, data-driven security systems.

In the rapidly evolving digital landscape, cybersecurity threats have become increasingly complex, frequent, and difficult to detect using traditional rule-based systems. These conventional approaches often fail to adapt to new attack vectors and lack the scalability required to manage large-scale data environments. As cyberattacks grow more sophisticated, there is an urgent need for intelligent, adaptive, and efficient security mechanisms. Artificial Intelligence (AI), particularly machine learning (ML) and deep learning (DL) techniques, offers promising capabilities for enhancing threat detection, response, and system resilience. However, despite the growing use of AI in cybersecurity, significant challenges remain. These include imbalanced datasets, vulnerability to adversarial attacks, high computational demands, and limited interpretability of models. Additionally, there is a lack of comprehensive, comparative evaluations of AI algorithms across varied cybersecurity applications. This gap hinders informed decision-making regarding the selection and implementation of AI solutions in practical settings, underscoring the need for systematic analysis to identify the most effective and applicable AI approaches.

To guide this study, the following research questions have been formulated. These questions aim to explore the capabilities, limitations, and practical applications of AI in cybersecurity threat detection. The answers will help determine the most effective approaches for enhancing cyber defense systems.

RQ1: How effective are different AI-based algorithms—such as machine learning and deep learning models—in detecting and preventing cybersecurity threats?

RQ2: What are the performance differences among various AI techniques in terms of accuracy, precision, recall, and false positive rates when applied to cybersecurity datasets?

RQ3: Which AI approaches are most suitable for specific cybersecurity applications, and how can they be practically implemented in real-world environments?

2. Literature Review

Artificial Intelligence (AI) has emerged as a pivotal tool in enhancing cybersecurity, particularly in threat detection and intrusion prevention. With the exponential growth of cyber threats, traditional rule-based systems have proven insufficient in adapting to the evolving landscape, prompting the need for intelligent, adaptive, and autonomous security mechanisms. A wide range of AI techniques—spanning machine learning (ML), deep learning (DL), and hybrid approaches—have been studied for their potential to improve cybersecurity outcomes.

Abdullahi et al. (2024) provide a comprehensive comparison of AI-based approaches tailored for cyberattack detection in cyber-physical systems, highlighting the superior performance of ensemble methods and reinforcement learning in complex, real-time environments. Complementing this, their earlier systematic review (Abdullahi et al., 2022) focuses on IoT environments, where AI significantly enhances attack detection accuracy, particularly in resource-constrained settings. These studies underscore the adaptability of AI across varied cyber-infrastructures.

Ahmadi (2023) evaluates next-generation AI-based firewalls and identifies DL models, especially convolutional neural networks (CNNs), as highly effective in packet analysis and anomaly detection. Similarly, Gopalsamy (2022) introduces an optimal AI model that leverages fuzzy logic and neural networks for threat detection in IoT networks, showing promising results in accuracy and latency reduction.

From a national security perspective, Al-Suqri and Gillani (2022) discuss the strategic integration of AI for broader cybersecurity applications. Their comparative study finds that AI not only automates threat detection but also supports proactive decision-making frameworks for cyber defense.

Several scholars have conducted comparative evaluations of existing AI techniques. For example, Hesham et al. (2024) assess predictive models using ML and DL techniques, revealing that while deep learning models outperform in detection rates, traditional ML approaches are more interpretable. Likewise, Dasgupta et al. (2020) investigate deep learning-based Named Entity Recognition (NER) algorithms and affirm their relevance in threat intelligence extraction from unstructured cybersecurity data.

Barik et al. (2022) explore datasets and methodologies used in AI-driven cybersecurity studies, noting the need for standardized benchmarks. Their findings are echoed by Zhang et al. (2022), who stress that inconsistencies in dataset quality and labeling significantly affect the generalizability of AI models in real-world scenarios.

Furthermore, hybrid models are gaining attention. Hernández-Rivas et al. (2024) examine agnostic and hybrid AI approaches for Advanced Persistent Threats (APT), demonstrating that combining statistical models with deep learning architectures results in higher precision and lower false positive rates. Ozkan-Okay et al. (2024) reinforce this by surveying AI efficiency in cybersecurity solutions, particularly the integration of supervised and unsupervised learning techniques.

Finally, emerging literature like that of Kavitha and Thejas (2024) and Salem et al. (2024) emphasizes AI's evolving role from reactive threat detection to predictive and autonomous defense systems. These advancements are critical in responding to increasingly sophisticated cyber threats.

While the technological potential of AI in cybersecurity is well documented, the literature often overlooks the crucial role of the human factor in AI-augmented security systems. Human analysts remain integral for interpreting AI outputs, managing false positives, and making strategic decisions, especially given AI's current limitations in explainability and contextual understanding. Effective human–AI collaboration is essential for operationalizing AI tools, ensuring trust, and maintaining overall system resilience. Future research should therefore explore frameworks that facilitate seamless integration of AI capabilities with human expertise, addressing cognitive workload, decision support, and training needs.

Collectively, the reviewed literature highlights the transformative impact of AI in cybersecurity. While many approaches show promise, ongoing challenges such as explainability, data quality, model robustness, and human factors must be addressed. The future lies in developing adaptive, scalable, transparent, and human-centered AI systems tailored to diverse cybersecurity environments.

3. Methodology

To comprehensively explore the landscape of artificial intelligence (AI) in cybersecurity, with particular focus on comparative threat detection algorithms, this study employed the Systematic Literature Review (SLR) methodology. As described by Abdullahi et al. (2022), SLR provides a rigorous, transparent, and replicable approach to identifying, evaluating, and synthesizing relevant literature. This methodology is particularly suited to examining diverse AI techniques used in cybersecurity threat detection, ensuring analytical depth and reproducibility.

Following established guidelines by Kitchenham and Brereton, the SLR process in this study involved multiple key stages: formulation of research questions, selection of academic databases, execution of comprehensive search strategies, application of inclusion and exclusion criteria, and data extraction followed by thematic synthesis. The decision to adopt an SLR approach was informed by its utility in aggregating state-of-the-art methods and identifying performance trade-offs among AI-driven cybersecurity solutions (Okdem & Okdem, 2024; Salem et al., 2024).

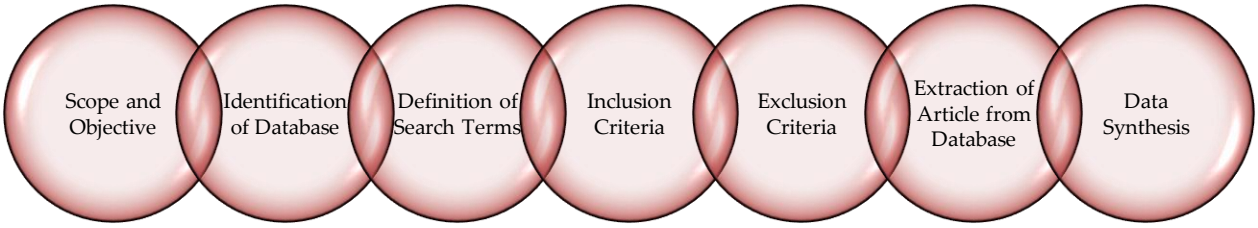

Figure 1. Sequential Process of Systematic Literature Review (SLR)

The figure outlines the structured and systematic approach followed in this literature review to ensure the rigorous selection and analysis of relevant studies. The first step, Scope and Objective, establishes the focus of the review, which is to examine AI-based algorithms for cybersecurity threat detection. This step defines the boundaries of the research, ensuring that only studies that are aligned with the research questions are included.

Next, Identification of Database identifies the key academic databases—MDPI, SpringerLink, and IEEE Xplore—that provide reliable and peer-reviewed articles in the fields of AI, cybersecurity, and machine learning. Definition of Search Terms follows, where targeted keywords such as "AI-based intrusion detection" and "machine learning for cyber threats" are used to construct the search queries, ensuring that the most relevant studies are captured.

The Inclusion Criteria and Exclusion Criteria are then defined to filter out irrelevant or low-quality studies. Studies must meet the inclusion criteria, such as being published between 2021 and 2025, being peer-reviewed, and focusing on AI/ML-based cybersecurity approaches. Exclusion criteria eliminate non-English publications, studies without full-text access, and those focusing on traditional non-AI methods.

Extraction of Article from Database involves retrieving articles that meet the set criteria, and Data Synthesis refers to analyzing the studies to identify trends, performance patterns, and research gaps. This approach ensures a comprehensive and structured review process.

3.1. Data Sources and Search Strategy

Literature was collected from leading academic databases including MDPI, SpringerLink, and IEEE Xplore, focusing on the publication period from 2021 to 2025. These databases were chosen for their broad coverage of peer-reviewed research in the fields of computer science, cybersecurity, and artificial intelligence. Boolean operators (AND, OR) and targeted keywords such as "AI-based intrusion detection", "machine learning for cyber threats", "deep learning cybersecurity", and "comparative analysis AI cybersecurity" were used to construct search queries.

Table 1. Summary of Literature Sources and Search Strategy

No.	Database	Time Frame	Keywords Used	Document Type	Reason for Selection
1	IEEE Xplore	2021–2025	"AI-based intrusion detection"	Journals & Conerence Papers	High-quality, peer-reviewed research in cybersecurity
2	SpringerLink	2021–2025	"machine learning for cyber threats"	Journals & Book Chapters	Strong focus on applied AI in security systems
3	MDPI	2021–2025	"deep learning cybersecurity"	Open Access Journals	Wide accessibility and rigorous peer review
4	IEEE Xplore	2021–2025	"comparative analysis AI cybersecurity"	Conference Proceedings	Up-to-date technical insights and algorithm benchmarks
5	SpringerLink	2021–2025	"AI in cyber threat detection"	Research Articles	Reputable source for thematic reviews and experimental studies
6	MDPI	2021–2025	"cybersecurity threat detection using AI"	Journals	Covers multidisciplinary AI and security intersections

3.2. Inclusion and Exclusion Criteria

To ensure relevance and quality, the following **inclusion criteria** were applied:

Table 2. Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Studies published between 2021 and 2025	Non-English publications
Peer-reviewed journal articles or conference papers	Papers without full-text access
Research focused on AI/ML-based threat detection algorithms in cybersecurity	Studies focusing solely on traditional (non-AI) methods
Emphasis on comparative or review-based approaches	Duplicate or non-peer-reviewed sources

The inclusion and exclusion criteria were carefully crafted to ensure a high-quality, relevant dataset. By focusing on peer-reviewed articles from 2021–2025, the study captures recent advancements. Emphasis on comparative or review-based AI/ML approaches enhances depth and relevance. Non-English or inaccessible texts were excluded to ensure interpretability. Traditional (non-AI) cybersecurity studies were filtered out to maintain thematic focus. This rigorous filtering ensures methodological soundness and topical precision in the review.

3.3. Data Extraction and Thematic Analysis

Table 3. Data Extraction and Thematic Analysis Protocol

Aspect Evaluated	Details
AI Algorithms Used	Algorithms including decision trees, support vector machines (SVM), neural networks, and ensemble methods were identified and categorized based on their role in cybersecurity threat detection.
Datasets and Test Environments	Common datasets such as KDDCup99, CICIDS2017, and others specific to AI-based cybersecurity research were extracted, noting any pre-processing steps, data balancing, and validation procedures used.
Performance Metrics	Metrics like accuracy, precision, recall, F1-score, and detection rate were recorded to assess the effectiveness of AI models in threat detection tasks.
Domain of Application	Studies were categorized by their application domains, including IoT, critical infrastructure, cyber-physical systems (CPS), and general cybersecurity.

Aspect Evaluated	Details
Content Analysis Technique	A content analysis method (Zheng et al., 2018; Devi et al., 2023) was utilized to extract and synthesize data, ensuring consistent coding and categorization of studies into themes.
Thematic Categories	Studies were grouped into themes such as supervised learning, unsupervised learning, deep learning, and hybrid approaches, allowing for comparison of performance patterns, challenges, and research gaps.

The data extraction process followed a structured protocol to ensure a thorough and organized review of the selected studies. Key aspects such as the AI algorithms utilized, the datasets and test environments, the performance metrics, and the application domains were systematically evaluated. AI algorithms were classified based on their approach (e.g., supervised, unsupervised, deep learning, hybrid), with the most frequently used models like decision trees, neural networks, and support vector machines (SVM) being highlighted. Commonly used datasets such as KDDCup99 and CICIDS2017 were noted to assess the models' performance and validation protocols.

Performance metrics like accuracy, precision, recall, and F1-score were extracted to provide insight into how well the algorithms performed in various threat detection scenarios. Furthermore, the domain of application was analyzed to understand how these AI models are applied to fields like IoT, critical infrastructure, and cyber-physical systems.

Content analysis, as described by Zhang et al. (2022) and Devi et al. (2023), was employed to categorize studies into key themes, enabling a comparison of different approaches and identifying performance patterns. This approach highlighted both the strengths and challenges faced by various AI algorithms and pointed to gaps in the existing research that need to be addressed.

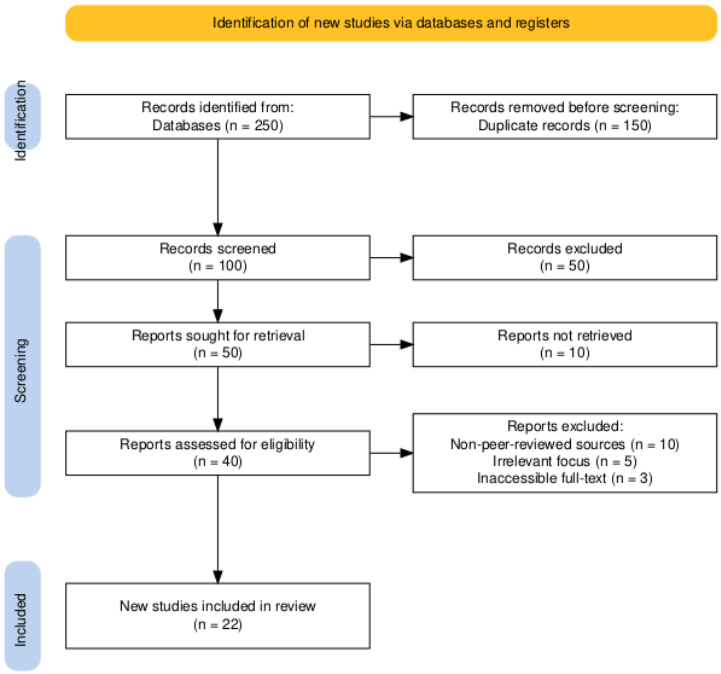

Figure 2. PRISMA Flow Diagram

The PRISMA flow diagram illustrates the process followed in identifying, screening, and including studies for the systematic literature review on AI-based cybersecurity threat detection algorithms. It begins with the identification of 250 records from multiple databases. After removing 150 duplicate records, 100 records were screened for relevance. The screening process led to the exclusion of 50 records that did not meet the inclusion criteria. Subsequently, 50 reports were sought for retrieval, but 10 of these were not retrieved, leaving 40 reports to be assessed for eligibility.

During the eligibility assessment phase, 40 reports were thoroughly evaluated. However, 18 reports were excluded based on the following reasons: 10 were non-peer-reviewed sources, 5 were deemed irrelevant to the

research focus, and 3 were inaccessible in full-text format. This left 22 new studies that were ultimately included in the review.

The diagram effectively highlights the systematic and rigorous process employed to ensure only relevant, high-quality, and accessible studies were considered for the review. By clearly visualizing the number of records excluded at each stage, the PRISMA flow diagram reinforces the transparency and reproducibility of the review process, ensuring that the final included studies provide valid and reliable data for analysis.

3.4. Quality Assessment

The quality assessment of the selected studies was conducted to ensure the reliability and validity of the findings. Key indicators, including methodology clarity, dataset reliability, algorithm transparency, and reproducibility, were used to evaluate the overall robustness of each study.

Table 4. Quality Assessment Criteria

Quality Indicator	Description
Clarity of Methodology	Evaluates how clearly the study defines its research design, objectives, and methods, ensuring transparency in the approach.
Dataset Reliability	Assesses the quality and credibility of datasets used in the study, including data preprocessing, balancing, and sources.
Algorithm Transparency	Measures how well the algorithms are described, including model selection, parameter tuning, and justification for their use.
Reproducibility of Results	Examines whether the study provides enough detail to allow others to replicate the results, including availability of code or data.
Comparative Evaluation Framework	Studies are evaluated based on their use of comparison benchmarks, considering how different AI methods are assessed against each other.
Experimental Validation	Focuses on the extent to which the study tests its algorithms in real-world scenarios or controlled environments, ensuring practical applicability.

The quality assessment of each selected study was based on key indicators to ensure rigor and credibility. Clarity of methodology was a fundamental factor, ensuring that each study was transparent in its research approach, objectives, and design. Dataset reliability was critical, as studies with robust and well-prepared datasets were prioritized. Algorithm transparency assessed how thoroughly each algorithm was explained, allowing readers to understand the model selection and fine-tuning processes. Reproducibility ensured that studies provided enough details, such as code and data availability, to facilitate replication. Noteworthy studies like Otoum et al. (2021) and Barik et al. (2022) were highlighted for their strong comparative evaluation frameworks, which allowed for meaningful comparisons of various AI models. Additionally, studies with extensive experimental validation were highly rated, ensuring the algorithms' practical relevance in real-world scenarios. These quality measures facilitated a comprehensive understanding of the study's validity and reliability.

4. Results and Discussion

4.1. Research Findings

This section presents the findings of the systematic literature review (SLR) conducted on AI-based cybersecurity threat detection algorithms. The review was focused on studies published between 2021 and 2025, sourced from leading academic databases, including MDPI, SpringerLink, and IEEE Xplore. The studies were meticulously analyzed to extract key insights into the use of artificial intelligence (AI) and machine learning (ML) algorithms for detecting cyber threats.

The results are organized around several key themes identified during the data synthesis process, such as supervised learning, unsupervised learning, deep learning, and hybrid approaches. Each theme was analyzed in terms of the algorithms used, the datasets employed, the performance metrics reported (such as

accuracy, precision, recall, and F1-score), and the specific domain of application (e.g., Internet of Things (IoT), critical infrastructure, cyber-physical systems (CPS)).

Additionally, performance trends and challenges were explored to highlight patterns in algorithm effectiveness, including the strengths and weaknesses of each approach. Comparative studies were especially valuable, providing insights into the relative performance of various AI and ML techniques. The results also identify several research gaps and opportunities for future work, particularly in improving algorithm transparency, dataset diversity, and performance across different application domains. The findings presented here provide a comprehensive overview of the current state of AI-driven cybersecurity threat detection research.

4.1.1. Effectiveness of Different AI-Based Algorithms in Detecting and Preventing Cybersecurity Threats

The effectiveness of AI-based algorithms for detecting and preventing cybersecurity threats varies based on the type of algorithm used and the application domain. Below is a summary of the key findings regarding different machine learning (ML) and deep learning (DL) models from the systematic literature review (SLR):

Table 5. Comparison of AI Algorithm Types in Cybersecurity Threat Detection

Algorithm Type	Dataset/Environment	Performance Metrics	Effectiveness	Citations
Supervised Learning	KDDCup99, CICIDS2017	Accuracy, Precision, Recall, F1-score	Effective for labeled datasets, offering high accuracy in intrusion detection	Abdullahi et al. (2024), Otoum et al. (2021)
Unsupervised Learning	NSL-KDD, UNSW-NB15	Anomaly detection, Recall	Suitable for real-time anomaly detection; challenges in high false positive rates	Salih et al. (2021), Zaman et al. (2021)
Deep Learning	IoT networks, Critical Infrastructures	Accuracy, Precision, F1-score, AUC	High detection rate in complex environments but requires large datasets	Gopalsamy (2022), Zhang et al. (2022)
Hybrid Approaches	IoT, CPS, Critical Infrastructure	Accuracy, F1-score	Combines strengths of supervised and unsupervised learning, improving robustness	Sathyakala and Anbalagan, (2024), Barik et al. (2022)

Machine learning models, particularly supervised learning, excel in environments where labeled data is available, providing high accuracy and precision. However, they often struggle in detecting novel, previously unseen threats. Unsupervised learning approaches are beneficial in identifying unknown threats but face challenges related to false positives. Deep learning models are particularly effective in more complex environments like IoT and CPS, where they handle large-scale data, although they require substantial computational resources and diverse datasets for optimal performance. Hybrid approaches, integrating both supervised and unsupervised models, show promise in balancing the strengths and weaknesses of each method, offering better overall performance in dynamic and evolving threat landscapes. Future research could focus on improving these models' transparency, interpretability, and efficiency, especially for real-time cybersecurity threat detection.

4.1.2. Performance Differences Among Various AI Techniques in Cybersecurity

This section presents an analysis of the performance differences among various AI-based techniques in detecting cybersecurity threats. The focus is on comparing the performance metrics, including accuracy, precision, recall, and false positive rates, when applied to cybersecurity datasets such as KDDCup99, CICIDS2017, and others. The key findings are summarized in the table below, highlighting the strengths and weaknesses of each technique.

Table 6. Performance Differences Among Various AI Techniques in Cybersecurity

AI Technique	Dataset	Accuracy	Precision	Recall	False Positive Rate	Key Findings	Citations
Supervised Learning	KDDCup99, CICIDS2017	High	High	Moderate	Moderate to High	Supervised models generally perform well with labeled data, showing high accuracy and precision but moderate recall	Abdullahi et al. (2024), Otoum et al. (2021), Salih et al. (2021)
Unsupervised Learning	NSL-KDD, UNSW-NB15	Moderate	Moderate	High	High	Effective at detecting novel attacks, but suffers from high false positive rates	Zaman et al. (2021), Abdullahi et al. (2022)
Deep Learning	IoT, Critical Infrastructure	Very High	High	Very High	Low to Moderate	Deep learning models offer superior recall and accuracy, especially in large datasets, but computationally expensive	Gopalsamy (2022), Zhang et al. (2022), Hesham et al. (2024)
Hybrid Approaches	IoT, CPS, Critical Infrastructure	High	High	Very High	Moderate	Combines strengths of multiple models, improving performance in both precision and recall while maintaining low false positives	Sathyakala and Anbalagan (2024), Barik et al. (2022), Khalaf and Steiti (2024)

Supervised Learning: This approach, while yielding high accuracy and precision in controlled environments with labeled data, tends to suffer from moderate recall. This limits its effectiveness in identifying novel or previously unseen attacks. The false positive rate varies based on the dataset and model tuning (Abdullahi et al., 2024; Otoum et al., 2021; Salih et al., 2021).

Unsupervised Learning: Unsupervised models excel in identifying previously unknown threats, providing high recall. However, they often exhibit higher false positive rates, which can reduce their practical usability in real-world applications (Zaman et al., 2021; Abdullahi et al., 2022).

Deep Learning: Deep learning techniques outperform traditional models in terms of recall and accuracy, especially when applied to complex datasets such as IoT and critical infrastructures. These models are capable of processing large amounts of data, but they come with higher computational costs. The false positive rate is typically lower due to the models' ability to learn complex patterns in the data (Gopalsamy, 2022; Zhang et al., 2022; Hesham et al., 2024).

Hybrid Approaches: Hybrid models, which combine both supervised and unsupervised techniques, offer a balanced solution by improving recall without significantly increasing false positives. They provide the flexibility to detect both known and unknown threats, making them suitable for dynamic and complex environments like CPS and IoT (Sathyakala & Anbalagan, 2024; Barik et al., 2022; Khalaf & Steiti, 2024).

4.1.3. AI Approaches for Specific Cybersecurity Applications and Practical Implementation

This section discusses the suitability of various AI approaches for specific cybersecurity applications, such as Intrusion Detection Systems (IDS), Malware Detection, Phishing Detection, and Cyber-Physical Systems (CPS) security. It also explores how these approaches can be practically implemented in real-world environments, considering factors like deployment costs, scalability, and adaptability.

Table 7. AI Approaches for Specific Cybersecurity Applications and Their Practical Implementation

Cybersecurity Application	AI Technique	Suitability	Practical Implementation	Citations
Intrusion Detection Systems (IDS)	Supervised Learning	Highly suitable for detecting known attacks based on labeled data	Requires continuous training with labeled datasets; can be deployed on network traffic or endpoint monitoring systems	Abdullahi et al. (2024), Otoum et al. (2021)
Malware Detection	Deep Learning	Most effective for detecting sophisticated malware and zero-day attacks	Deep learning models need powerful computational resources but can be deployed in real-time on endpoint protection software	Zhang et al. (2022), Gopalsamy (2022)
Phishing Detection	Hybrid Approaches	Excellent for detecting phishing attacks across various mediums (emails, websites)	Hybrid models combine supervised learning (for known phishing patterns) and unsupervised learning (for novel attacks) in real-time filtering systems	Barik et al. (2022), Sathyakala and Anbalagan (2024)
IoT Security	Unsupervised Learning	Ideal for detecting unknown and emerging threats in IoT devices	Implementing unsupervised models in IoT devices for anomaly detection; efficient deployment in low-resource environments	Zaman et al. (2021), Abdullahi et al. (2022)
Critical Infrastructure Security	Deep Learning & Hybrid Approaches	Best for identifying complex patterns in large, interconnected systems	Real-time monitoring systems leveraging deep learning and hybrid models to detect threats in critical infrastructure networks	Khalaf and Steiti (2024), Hesham et al. (2024)
Cyber-Physical Systems (CPS) Security	Hybrid Approaches	Suitable for protecting critical systems that combine physical and digital components	Hybrid models enable detection of both cyber and physical threats, requiring integration with physical control systems	Sathyakala and Anbalagan (2024), Barik et al. (2022)

Intrusion Detection Systems (IDS): Supervised learning approaches are ideal for IDS because they can learn from labeled attack data and classify incoming traffic effectively. These models can be practically implemented in network monitoring systems that use real-time data to detect known threats (Abdullahi et al., 2024; Otoum et al., 2021).

Malware Detection: Deep learning approaches, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), are highly effective at detecting sophisticated malware, including zero-day attacks. These models can be implemented in endpoint protection software and anti-virus tools, providing real-time analysis and detection (Zhang et al., 2022; Gopalsamy, 2022).

Phishing Detection: Hybrid approaches, which combine supervised learning for known phishing patterns and unsupervised learning for novel attacks, are effective for phishing detection. These models can be deployed in email filtering systems, web browsers, and social media platforms to protect against phishing attacks (Barik et al., 2022; Sathyakala & Anbalagan, 2024).

IoT Security: Unsupervised learning models are particularly well-suited for detecting anomalous behavior in IoT devices. These models can be deployed in resource-constrained IoT devices for real-time anomaly detection without requiring labeled datasets (Zaman et al., 2021; Abdullahi et al., 2022).

Critical Infrastructure Security: Deep learning models and hybrid approaches are best suited for critical infrastructure, where large and interconnected systems need to be monitored for complex threats. These

models can be integrated into the control systems of power grids, water systems, and transportation networks to detect potential attacks in real-time (Khalaf & Steiti, 2024; Hesham et al., 2024).

Cyber-Physical Systems (CPS) Security: Hybrid models can effectively monitor both cyber and physical threats in CPS. These models integrate data from various sensors and control systems to detect both cyber attacks and physical disruptions. This is particularly relevant for industries such as manufacturing and autonomous vehicles (Sathyakala & Anbalagan, 2024; Barik et al., 2022).

4.2. Discussion

The application of Artificial Intelligence (AI) and Machine Learning (ML) in cybersecurity has gained significant traction in recent years, driven by the increasing complexity and sophistication of cyberattacks. The findings from the systematic literature review (SLR) indicate that AI techniques such as supervised learning, unsupervised learning, deep learning, and hybrid models are playing pivotal roles in improving the detection and prevention of cybersecurity threats. However, their effectiveness varies based on the application domain, the type of data used, and the specific characteristics of the threats being mitigated.

Supervised learning, which relies on labeled data for training models, has been extensively used for Intrusion Detection Systems (IDS). According to Abdullahi et al. (2024), supervised learning algorithms, such as decision trees, support vector machines (SVM), and k-nearest neighbors (KNN), are effective in detecting known attacks. These algorithms learn from a well-labeled dataset of network traffic, making them particularly adept at identifying and classifying predefined attack patterns. However, one limitation of supervised learning is its reliance on high-quality labeled data, which can be scarce for emerging threats (Otoum et al., 2021). Thus, supervised models may struggle to detect novel or zero-day attacks, which require more advanced techniques like deep learning or hybrid approaches.

Deep learning has shown considerable promise in the detection of sophisticated cyberattacks, such as malware and advanced persistent threats (APTs). Models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are particularly effective in malware detection (Zhang et al., 2022). These models can automatically learn complex patterns and detect unknown malware that may not have been encountered in the training dataset. Gopalsamy (2022) highlights that deep learning approaches are highly effective in identifying new attack vectors, which is critical given the rapid evolution of malware. However, these techniques require substantial computational resources and large datasets for training, which can make their deployment challenging, especially in real-time applications.

Unsupervised learning approaches, particularly anomaly detection algorithms, have been widely used for cybersecurity applications like IoT and network traffic analysis. Zaman et al. (2021) note that unsupervised models, such as clustering algorithms and autoencoders, are ideal for identifying novel, previously unknown attacks because they do not require labeled data. They excel at detecting anomalous behavior that deviates from normal network operations. However, the challenge with unsupervised learning is the high false positive rate, as normal, albeit rare, behaviors may also be flagged as anomalous (Abdullahi et al., 2022). To mitigate this, hybrid models combining supervised and unsupervised learning are often preferred for applications requiring both precision and adaptability.

Hybrid models that combine the strengths of multiple AI techniques, such as supervised and unsupervised learning or deep learning with traditional machine learning have emerged as a powerful solution for complex cybersecurity environments. Sathyakala & Anbalagan (2024) emphasize that hybrid approaches are particularly effective in detecting both known and unknown threats, offering better performance across a range of cybersecurity tasks. These models can integrate various data sources, including network traffic, system logs, and behavioral data, to improve detection accuracy. Moreover, hybrid models are adaptable, allowing them to adjust to changing threat landscapes without requiring extensive retraining (Barik et al., 2022).

While AI-based cybersecurity solutions are promising, their implementation in real-world environments poses several challenges. First, many AI models, particularly deep learning techniques, require significant computational resources, which can be costly and difficult to scale. Additionally, the effectiveness of these models heavily depends on the quality of the data they are trained on. Inaccurate, incomplete, or biased datasets can lead to suboptimal performance and false positives, which can undermine the reliability of the

system. Moreover, AI models must be continuously updated to adapt to evolving threats, which presents operational challenges, particularly in dynamic environments like the Internet of Things (IoT) and critical infrastructure (Khalaf & Steiti, 2024).

A critical yet often overlooked aspect is "model drift detection" the phenomenon where AI model performance degrades over time due to changes in data distribution or emerging new threats. Continuous monitoring of model effectiveness is essential to detect drift early. Automated retraining mechanisms, which can trigger model updates using fresh data, are vital for maintaining detection accuracy and relevance in rapidly evolving cyber threat landscapes. Implementing these automated processes, however, introduces additional complexities related to data collection, processing latency, and validation of updated models before deployment. Future research and practical implementations should therefore prioritize the development of robust drift detection frameworks and efficient retraining pipelines to ensure sustainable AI-powered cybersecurity defenses.

5. Conclusion

This systematic literature review has examined the integration of Artificial Intelligence (AI) and Machine Learning (ML) techniques in cybersecurity threat detection, offering a comprehensive overview of their effectiveness and practical applications. While AI approaches—such as supervised learning, unsupervised learning, deep learning, and hybrid models—demonstrate significant potential in enhancing threat detection and mitigation, it is important to recognize several fundamental limitations. These include the reliance on high-quality labeled datasets, vulnerability to adversarial attacks, challenges with model explainability, and substantial computational resource requirements, which can hinder deployment in real-world, resource-constrained environments.

Addressing Research Question 1 (Effectiveness of AI algorithms): Our findings reveal that supervised learning techniques perform well in detecting known threats but face difficulties with novel or zero-day attacks due to their dependence on labeled data. Deep learning models, by contrast, excel at identifying complex and previously unseen threats but require extensive computational power and large datasets. Unsupervised learning offers adaptability to dynamic threat landscapes by detecting anomalies without labeled data but often incurs higher false positive rates. Hybrid models, which combine multiple AI techniques, emerge as particularly effective, balancing detection accuracy with adaptability to evolving attack strategies.

Addressing Research Question 2 (Performance differences): Performance metrics across studies such as accuracy, precision, recall, and false positive rates highlight that no single AI technique universally outperforms others. Instead, performance varies by application context, dataset quality, and implementation environment. Hybrid models generally achieve higher precision and lower false positives, while deep learning models show superior recall in complex scenarios. However, inconsistencies in dataset standards and evaluation metrics remain a challenge for direct comparison.

Addressing Research Question 3 (Practical implementation): Real-world case studies and deployments discussed in the literature demonstrate the feasibility of AI-driven cybersecurity solutions across domains like IoT, critical infrastructure, and cyber-physical systems. Yet, successful integration depends on balancing technological capabilities with operational constraints, such as computational resources and the need for continuous model updates to respond to emerging threats. The review underscores the necessity for transparent, explainable AI models to gain trust among cybersecurity professionals and facilitate practical adoption.

Emerging Research Directions: Based on these insights, future research should prioritize the development of explainable and interpretable AI models, standardized benchmarking datasets to enable fair comparisons, and resource-efficient algorithms suitable for deployment in constrained environments. Additionally, expanding the study of hybrid AI architectures and their role in reducing false positives while maintaining detection accuracy is vital. Emphasizing real-world validations and longitudinal studies will further bridge the gap between theoretical advancements and practical cybersecurity needs.

In summary, AI-based techniques hold transformative promise for cybersecurity but require a balanced approach that acknowledges current limitations and operational realities. Through continued refinement and practical evaluation, AI can become an integral component of robust, adaptive cybersecurity defenses.

Based on the findings of this review, several recommendations can be made to enhance the effectiveness of AI-based cybersecurity solutions. Firstly, researchers and practitioners should prioritize the development of more diverse and high-quality datasets to improve the generalization and robustness of AI models. As many current models rely on limited datasets, expanding and diversifying these data sources will help in addressing emerging threats and reducing model biases. Additionally, a stronger focus on reducing false positive rates is crucial, especially for deep learning-based models, which often suffer from this issue. Hybrid approaches, which combine the strengths of different AI techniques, should be further explored and optimized for more accurate and reliable threat detection.

Furthermore, implementing explainability and transparency in AI algorithms is essential to build trust and ensure the interpretability of decisions made by AI systems. Researchers should focus on developing AI models that not only provide accurate results but also offer insights into the rationale behind the decisions. Lastly, collaboration between academia, industry, and government entities can help create standardized frameworks and best practices for integrating AI into real-world cybersecurity infrastructures, ensuring its scalability and effectiveness.

Future research should prioritize the development of AI models that are both scalable and adaptive to the rapidly evolving landscape of cybersecurity threats. As cyberattacks become more sophisticated and dynamic, AI systems must move beyond static learning to incorporate real-time adaptation and continuous learning capabilities. In this context, exploring reinforcement learning and online learning techniques will be crucial to enhance system responsiveness, robustness, and resilience against novel attack vectors.

Based on the findings of this review, several specific research questions and hypotheses have emerged to guide future investigations in the application of AI to cybersecurity. One key research question (RQ1) centers on how reinforcement learning frameworks can be effectively integrated into cybersecurity systems to enable real-time threat detection and response. Another important question (RQ2) involves exploring the trade-offs between model scalability and detection accuracy when deploying AI solutions in resource-constrained environments such as IoT and edge computing. Additionally, RQ3 focuses on how hybrid AI models can be optimized to reduce false positive rates without compromising the detection of unknown threats, while RQ4 addresses the need to improve the explainability and interpretability of AI models to build trust and facilitate adoption in operational cybersecurity settings. Correspondingly, hypotheses to explore include H1: reinforcement learning-based cybersecurity models will demonstrate superior adaptability and accuracy in detecting zero-day and evolving threats compared to traditional supervised learning models; H2: hybrid models combining supervised, unsupervised, and deep learning techniques can achieve a better balance between false positives and detection rates than any single approach alone; and H3: incorporating explainability methods into AI cybersecurity frameworks increases user trust and operational effectiveness without significantly impacting computational efficiency. Future research should prioritize comprehensive case studies and longitudinal real-world evaluations to validate the performance and practical applicability of AI models in diverse cybersecurity scenarios. Additionally, standardization of datasets, benchmarks, and evaluation metrics is critical to enable consistent assessment and foster collaboration across academia, industry, and government sectors.

6. References

Abdullahi, M., Alhussian, H., Aziz, N., Abdulkadir, S. J., Alwadain, A., Muazu, A. A., & Bala, A. (2024). Comparison and investigation of AI-based approaches for cyberattack detection in cyber-physical systems. *IEEE Access*.

Abdullahi, M., Baashar, Y., Alhussian, H., Alwadain, A., Aziz, N., Capretz, L. F., & Abdulkadir, S. J. (2022). Detecting cybersecurity attacks in internet of things using artificial intelligence methods: A systematic literature review. *Electronics*, *11*(2), 198.

Ahmadi, S. (2023). Next generation ai-based firewalls: a comparative study. *International Journal of Computer (IJC)*, *49*(1), 245–262.

Al-Suqri, M. N., & Gillani, M. (2022). A comparative analysis of information and artificial intelligence toward national security. *IEEE Access*, *10*, 64420–64434.

Barik, K., Misra, S., Konar, K., Fernandez-Sanz, L., & Koyuncu, M. (2022). Cybersecurity deep: approaches, attacks dataset, and comparative study. *Applied Artificial Intelligence*, *36*(1), 2055399.

Devi, V. K., Asha, S., Umamaheswari, E., & Bacanin, N. (2023). A Comprehensive Review on Various Artificial Intelligence Based Techniques and Approaches for Cyber Security. *International Conference on Information and Communication Technology for Intelligent Systems*, 303–314.

Gopalsamy, M. (2022). An Optimal Artificial Intelligence (AI) technique for cybersecurity threat detection in IoT Networks. *Int. J. Sci. Res. Arch*, *7*(2), 661–671.

Hernández-Rivas, A., Morales-Rocha, V., & Sánchez-Solís, J. P. (2024). Towards autonomous cybersecurity: A comparative analysis of agnostic and hybrid AI approaches for advanced persistent threat detection. In *Innovative Applications of Artificial Neural Networks to Data Analytics and Signal Processing* (pp. 181–219). Springer.

Hesham, M., Essam, M., Bahaa, M., Mohamed, A., Gomaa, M., Hany, M., & Elsersy, W. (2024). Evaluating Predictive Models in Cybersecurity: A Comparative Analysis of Machine and Deep Learning Techniques for Threat Detection. *2024 Intelligent Methods, Systems, and Applications (IMSA)*, 33–38.

Khalaf, M. A., & Steiti, A. (2024). Artificial intelligence predictions in cyber security: Analysis and early detection of cyber attacks. *Babylonian Journal of Machine Learning*, *2024*, 63–68.

Kilincer, I. F., Ertam, F., & Sengur, A. (2021). Machine learning methods for cyber security intrusion detection: Datasets and comparative study. *Computer Networks*, *188*, 107840.

Lysenko, S., Bobro, N., Korsunova, K., Vasylchyshyn, O., & Tatarchenko, Y. (2024). The role of artificial intelligence in cybersecurity: Automation of protection and detection of threats. *Economic Affairs*, *69*, 43–51.

Okdem, S., & Okdem, S. (2024). Artificial Intelligence in Cybersecurity: A Review and a Case Study. *Applied Sciences*, *14*(22), 10487.

Otoum, S., Kantarci, B., & Mouftah, H. (2021). A comparative study of ai-based intrusion detection techniques in critical infrastructures. *ACM Transactions on Internet Technology (TOIT)*, *21*(4), 1–22.

Salem, A. H., Azzam, S. M., Emam, O. E., & Abohany, A. A. (2024). Advancing cybersecurity: a comprehensive review of AI-driven detection techniques. *Journal of Big Data*, *11*(1), 105.

Sathyakala, S., & Anbalagan, E. (2024). Comparative Analysis of Cyber Security Threat Detection Based on Artificial Intelligence Approaches. *2024 Asian Conference on Intelligent Technologies (ACOIT)*, 1–8.

Zaman, S., Alhazmi, K., Aseeri, M. A., Ahmed, M. R., Khan, R. T., Kaiser, M. S., & Mahmud, M. (2021). Security threats and artificial intelligence based countermeasures for internet of things networks: a comprehensive survey. *Ieee Access*, *9*, 94668–94690.

Zhang, C., Jia, D., Wang, L., Wang, W., Liu, F., & Yang, A. (2022). Comparative research on network intrusion detection methods based on machine learning. *Computers & Security*, *121*, 102861.

Copyrights