British Journal of Computer, Networking and Information Technology
ISSN: 2689-5315
Volume 8, Issue 1, 2025 (pp. 14-29)
www.abjournals.org

RISK MITIGATION APPROACH TO CYBER THREAT USING AI-DRIVEN MODELS FOR THE EVOLVING THREAT LANDSCAPE

Jesufemi Olanrewaju[1], Matthias Oluloni Togunde[2], and Oyebola Akande[3]

[1]SLWC Inc, Winnipeg, Canada.
Email: jolanrewaju@springs.ca

[2]Interswitch Limited, Nigeria.
Email: matthias.togunde@interswitchgroup.com

[3]Computer Science Department, School of Computing, Babcock University, Ilishan, Ogun State, Nigeria.
Email: akandeo@babcock.edu.ng

Cite this article:

Olanrewaju, J., Togunde, M. O., Akande, O. (2025), Risk Mitigation Approach to Cyber Threat using AI-Driven Models for the Evolving Threat Landscape. British Journal of Computer, Networking and Information Technology 8(1), 14-29. DOI: 10.52589/BJCNIT-1HH9NPSN

Manuscript History

Received: 18 Nov 2024

Accepted: 5 Jan 2025

Published: 17 Jan 2025

ABSTRACT: *This systematic review examines the effectiveness of AI-driven models in mitigating evolving cyber threats, using the PRISMA framework to analyze studies published between 2019 and 2024. The review focuses on machine learning techniques, including supervised, unsupervised, and deep learning. Findings show that deep learning excels in detecting complex threats like Advance Persistent Threats (APTs) and zero-day vulnerabilities, while supervised learning (deep learning is also a supervised type of supervised learning, so be specific) is effective for known threats but struggles with new attack types. Unsupervised learning adapts well to dynamic environments but has higher false positive rates. The review proposes a multi-layered framework combining AI models with traditional security measures for enhanced threat detection and response. A hybrid approach is recommended as the most effective strategy, though challenges like data quality and algorithmic bias must be addressed for optimal implementation.*

KEYWORDS: Advance Persistent Threats, zero-day, PRISMA framework, multi-layered framework, AI-driven

Article DOI: 10.52589/BJCNIT-1HH9NPSN
DOI URL: https://doi.org/10.52589/BJCNIT-1HH9NPSN

INTRODUTION

The increasing complexity and frequency of cyber threats have highlighted the limitations of traditional cybersecurity measures like Firewalls, Antivirus Software, Password-Based Authentication [31]. Their limitations lead to a growing interest in using Artificial Intelligence (AI) to enhance cyber risk mitigation. AI-driven models offer advanced capabilities in threat detection, response, and prediction, enabling more proactive and adaptive cybersecurity strategies. These models are particularly effective in handling complex and evolving threats through the use of machine learning (ML) techniques, including supervised, unsupervised, and deep learning methods. Such techniques help identify patterns and anomalies in large datasets, improving detection accuracy and response times for malware, phishing, and Advanced Persistent Threats (APTs) [2]. This systematic review aims to comprehensively evaluate AI-driven models in cybersecurity, focusing on their ability to mitigate evolving risks. It compares different ML approaches to assess their relative strengths and weaknesses in various threat scenarios. Additionally, the review proposes a framework for integrating AI-driven models into existing cybersecurity strategies, emphasising improving real-time detection and automated response mechanisms. By synthesising findings from various studies, the review offers insights into the most effective AI techniques for addressing the dynamic nature of modern cyber threats.

Statement of the Problem

The rapid evolution of cyber threats has exposed critical vulnerabilities in traditional cybersecurity measures such as firewalls, antivirus programs, and password-based authentication systems. These conventional approaches struggle to address the dynamic nature and sophistication of modern threats, including Advanced Persistent Threats (APTs) and zero-day vulnerabilities. The escalating complexity of cyberattacks has driven a growing need for advanced, adaptive, and proactive solutions. While Artificial Intelligence (AI)-driven models, employing machine learning (ML) and deep learning (DL) techniques, offer promise in identifying and mitigating such threats, challenges like data quality issues, algorithmic biases, and high false-positive rates remain barriers to their effective adoption. Addressing these gaps is critical for developing robust cybersecurity frameworks that can withstand evolving threats in dynamic environments such as IoT and cloud computing systems.

Objectives of the Study

This study aims to systematically evaluate the effectiveness of AI-driven models in mitigating evolving cyber threats. Specifically, it seeks to:

1. Assess the strengths and limitations of various machine learning approaches, including supervised, unsupervised, and deep learning, in threat detection and response.
2. Propose a multi-layered framework that integrates AI-driven techniques with traditional cybersecurity measures for enhanced threat management.
3. Identify practical solutions to challenges such as data quality, algorithmic bias, and high computational costs to improve the adaptability and scalability of AI models.
4. Provide insights into real-time threat detection, anomaly identification, and mitigation strategies to strengthen cybersecurity systems against modern threats.

Article DOI: 10.52589/BJCNIT-1HH9NPSN
DOI URL: https://doi.org/10.52589/BJCNIT-1HH9NPSN

LITERATURE REVIEW

AI-Driven Cybersecurity Models

AI-driven models are central to cybersecurity threat detection and response mechanisms, enhancing accuracy and response times in real-time threat environments. AI models are widely applied in Intrusion Detection Systems (IDS), malware detection, phishing prevention, threat intelligence, and endpoint protection [1]. In IDS, AI techniques such as supervised and unsupervised learning are deployed to detect malicious traffic and previously unknown attack patterns more effectively than traditional methods. In malware detection, AI-based static and dynamic analysis surpasses signature-based approaches, identifying complex behaviors and unknown malware. Phishing detection has also benefited from AI, significantly reducing incidents by improving email filtering and URL analysis, though false positives remain a challenge [3]. Additionally, AI is critical in threat intelligence, where it automates data analysis and applies predictive analytics to uncover hidden threats and anticipate attacks.

Similarly, the role of AI-driven models in cybersecurity, particularly through the automation of threat detection and anomaly identification, has been transformative, enabling faster response times, improved accuracy in identifying complex attack patterns, and reducing the burden on human analysts by handling large-scale data analysis in real-time [34]. These models significantly reduce the manual workload by analyzing large data sets and identifying patterns that indicate potential security breaches, enabling organizations to act proactively. A notable strength of AI models is their adaptability, as they continuously learn from new threats, enhancing their defense mechanisms over time. AI-based cybersecurity models (Artificial Neural Networks (ANNs), agent-based systems, and Genetic-fuzzy IDSs) have transformed the fight against cybercrime because they offer greater flexibility and robustness than traditional methods [32]. These advanced systems enhance the detection of unauthorised access and cyber threats, adapting to an ever-evolving threat landscape.

Artificial Intelligence (AI) Machine Learning (ML) in Cybersecurity or rather Artificial Intelligence-Machine Learning (AIML) in Cybersecurity

The application of Artificial Intelligence (AI) and Machine Learning (ML) in cybersecurity has emerged as a critical factor in improving threat detection and response mechanisms. AI-driven models, particularly those utilising ML techniques, are crucial in identifying anomalies, detecting malicious activities, and enhancing response times [1]. Machine learning algorithms include supervised, unsupervised, and reinforcement learning essential for analysing complex data patterns and improving the accuracy of intrusion detection systems and malware analysis. However, integrating AI into cybersecurity also faces challenges, such as ethical concerns and data privacy issues, which must be addressed to fully harness its potential. Similarly, the significance of AI for system security assurance is that AI enables proactive measures against cyber threats [2]. The intelligence-based approaches involving AI allow faster detection and mitigation of threats, reducing the potential damage caused by cyberattacks [3]. These studies collectively highlight the growing role of AI in advancing cybersecurity practices, emphasising the importance of real-time detection and dynamic response.

Furthermore, integrating Machine Learning (ML) into cybersecurity frameworks is increasingly recognized as essential for managing evolving cyber threats. ML's ability to analyse vast amounts of data enables organizations to identify threats before they escalate,

Article DOI: 10.52589/BJCNIT-1HH9NPSN
DOI URL: https://doi.org/10.52589/BJCNIT-1HH9NPSN

thereby improving threat intelligence and overall system security [4]. Their review highlights the successful implementation of ML in real-world case studies, demonstrating its transformative impact on cybersecurity practices.

A broader perspective on the gaps in ML research related to cybersecurity, identifying the need for more robust frameworks and standardized auditing practices [5]. Although ML offers promising solutions, its current implementation remains immature, with gaps in its ability to address all cyber threats. ML, especially advanced techniques like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), is crucial in detecting and responding to zero-day vulnerabilities [6]. Ensemble learning methods like Random Forest and Support Vector Machines (SVMs) further enhance the detection capabilities of cybersecurity systems.

The inadequacy of traditional methods in addressing sophisticated cyber threats is evidenced but ML and AI are crucial in enabling systems to learn from data, detect potential threats early, and mitigate risks before significant harm occurs [7]. However, challenges such as data quality issues, security vulnerabilities, and potential bias in ML algorithms can undermine the effectiveness of these technologies. Also, while ML is effective in early detection, it is also susceptible to false positives and adversarial attacks, underscoring the need for continuous innovation and refinement in ML models [8].

Deep Learning (DL)

Deep learning (DL) has become a pivotal technology in enhancing cybersecurity across various domains. Its capacity to process large datasets and detect complex patterns makes it highly effective in identifying evolving cyber threats. Deep learning, particularly through artificial neural networks, outperforms traditional machine learning algorithms by identifying nonlinear patterns essential for detecting sophisticated cyber-attacks [9]. This advantage is critical in modern cybersecurity environments where threats are more frequent and increasingly complex. The versatility of DL extends beyond cybersecurity, as it has shown success in fields like medical data processing and automation, demonstrating its robustness in handling intricate datasets. Deep learning has a transformative role in network security, where traditional methods often fail to keep pace with the evolving nature of cyberattacks [10]. Through automation and the ability to process unstructured data, DL models enable real-time detection of potential threats. This is particularly important as cyber threats become more sophisticated and interconnected with other emerging technologies like IoT and cloud computing.

Focusing on Intrusion Detection Systems (IDS), DL's application detects many cyber threats, including malware and phishing attacks [11]. Their review highlights the increasing need for advanced IDS solutions as organisations face growing cyber vulnerabilities, particularly through e-learning platforms and other digital systems with weak security frameworks. DL techniques, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have proven effective in identifying these threats, offering enhanced detection capabilities compared to traditional methods. The paper provides a comparative analysis of DL techniques, shedding light on their advantages in improving network security through real-world implementations.

Article DOI: 10.52589/BJCNIT-1HH9NPSN
DOI URL: https://doi.org/10.52589/BJCNIT-1HH9NPSN

Deep Learning in IoT Cybersecurity

The role of DL in IoT cybersecurity is essential. The interconnected nature of IoT devices introduces unique security challenges that traditional IDS solutions often cannot address [12]. DL models, with their capacity to analyze large volumes and non-linear patterns of data, are particularly suited for detecting cyber threats in IoT environments. These models excel in identifying a range of attacks specific to IoT, including DDoS (Distributed Denial-of-Service) and botnet attacks.

Integration and Hybrid Approaches

The integration of Artificial Intelligence (AI) into various sectors, including cybersecurity and education, has opened up new opportunities for innovation while presenting unique challenges. A recurring theme in the literature is the potential of hybrid approaches, which combine different AI techniques to address complex problems more effectively. A hybrid approach to modelling reality and perception emphasizes the importance of cognitive frameworks such as cyber semiotics and cognitive metaphor in understanding the dynamic interaction between digital and physical realities [13]. This approach, particularly relevant in educational contexts, fosters integrative metathinking, which is critical for addressing cognitive distortions and safety concerns induced by digitalisation. The authors argue that this methodology can be applied to Generative AI (GenAI) [33], providing creative and critical thinking strategies for navigating cognitive science and educational environments.

In cybersecurity, a hybrid genetic algorithm (GA) and neural network-based approach enhance the detection of DDoS and malware attacks [14]. Their research highlights the effectiveness of GAs in feature selection, which optimises neural networks by identifying the most relevant data attributes. The integration of GAs with Swarm Intelligence (SI) and other nature-inspired algorithms, such as Artificial Bee Colony (ABC), improve both the accuracy and efficiency of cybersecurity systems. This hybrid method enhances precision detection and reduces the computational burden of processing large datasets, offering a robust solution for wide-area networks.

METHODOLOGY

The methodology for conducting systematic review of AI-driven models for cyber risk mitigation is designed to comprehensively assess and evaluate the current landscape of AI applications in cybersecurity. This systematic approach ensures rigorous data collection, minimises bias, and provides a clear structure for analyzing the effectiveness of AI-driven models in mitigating evolving cyber threats. The primary objectives of this review include evaluating the effectiveness of AI-driven models in addressing advanced threats. Additionally, the review aims to compare the performance of different machine learning (ML) approaches—supervised, unsupervised, and deep learning—to determine their relative effectiveness in diverse cyber risk scenarios. Finally, the review proposed a comprehensive framework integrating AI-driven models into cybersecurity strategies, enhancing threat detection and response mechanisms.

Article DOI: 10.52589/BJCNIT-1HH9NPSN
DOI URL: https://doi.org/10.52589/BJCNIT-1HH9NPSN

Review Protocol

The review was conducted according to the guidelines provided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework (Moher et al., 2009). The PRISMA framework ensures transparency and reproducibility in systematic reviews, guiding the process from planning to reporting. The protocol was pre-defined, including objectives, research questions, eligibility criteria, and the data extraction process. It was registered in PROSPERO to enhance transparency and prevent duplication.

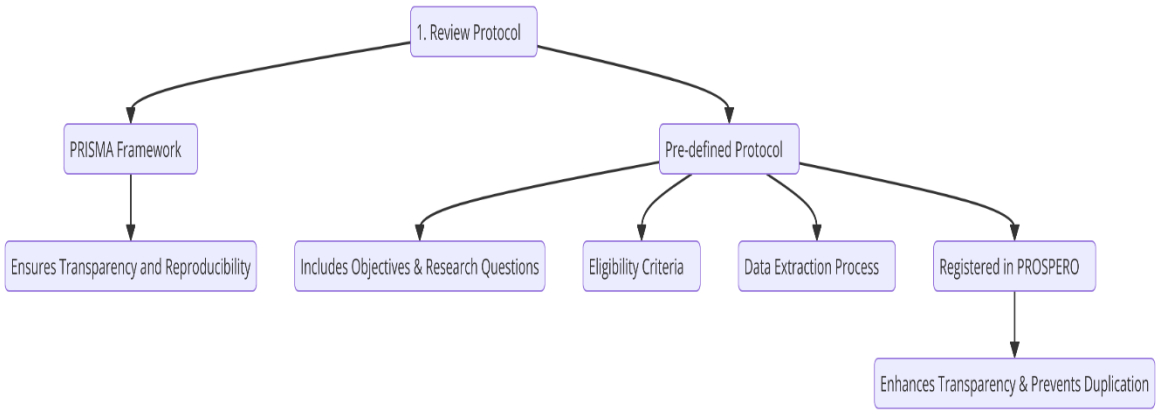

Figure 1: Review Protocol Diagram

Search Strategy

A comprehensive search strategy was developed to identify relevant studies on AI-driven models for cyber risk mitigation. The search was conducted using multiple academic databases, including IEEE Xplore, ACM Digital Library, Scopus, and Web of Science, to ensure a broad range of sources. A combination of Boolean operators, such as "AND," "OR," and "NOT," was employed to refine the search terms. The keywords included "Artificial Intelligence," "cybersecurity," "risk mitigation," "machine learning," and "cyber threats." For example, a typical search string used was: ("Artificial Intelligence" OR "AI" OR "Machine Learning" OR "Deep Learning") AND ("Cyber Risk" OR "Cybersecurity" OR "Risk Management").

Table 1: Search Strategy Component	Description
Objective	To identify relevant studies focusing on AI-driven models for cyber risk mitigation.
Databases	The search was cond ucted using Web of Science, IEEE Xplore, ACM Digital Library
Search Strategy	A comprehensive approach using Boolean operators like "AND," "OR," and "NOT" was employed to filter and refine search results. This ensured the inclusion of relevant studies and the exclusion of unrelated ones.
Keywords	Artificial intelligence, Deep Learning, Cybersecurity, Cyber Risk, Cyber Threats

Article DOI: 10.52589/BJCNIT-1HH9NPSN
DOI URL: https://doi.org/10.52589/BJCNIT-1HH9NPSN

Boolean Operators	("Artificial Intelligence" OR "AI" OR "Machine Learning" OR "Deep Learning") AND ("Cyber Risk" OR "Cybersecurity" OR "Risk Management")
Study Languages	Only studies published in English were included to ensure accessibility and consistency in data interpretation.

Identification of Databases and Keywords

The databases selected for this systematic review were chosen based on their relevance to computer science and cybersecurity research. Primary databases included IEEE Xplore, ACM Digital Library, Scopus, Web of Science, and Google Scholar, covering peer-reviewed journals, conference proceedings, and industry reports. Keywords and search terms were iteratively refined during preliminary searches to ensure the inclusion of all relevant studies. Keywords were selected based on the thematic focus of the review, such as "neural networks," "anomaly detection," "threat intelligence," and "risk prediction."

Inclusion Criteria

The inclusion criteria were designed to ensure that only studies with a focus on AI-driven models for cyber risk mitigation were selected. Specifically, studies were included if they met the following criteria. Only literature published between 2019 and 2024 were included to ensure recent developments in AI technologies and also, literature that focuses on the application of AI in cybersecurity, particularly in the context of risk assessment, threat detection, or mitigation were included.

Table 2: Inclusion Criteria

Inclusion Criteria	Description
Publication Date	Studies published between 201 and 2024 were included to ensure the review covers recent developments in AI technologies.
Language	Only studies published in English were considered, ensuring that the findings are accessible and interpretable.
Study Type	Empirical studies were included, including those with experimental results or case studies. These studies are needed to demonstrate the application of AI in real-world scenarios or simulations.

Exclusion Criteria

The exclusion criteria aimed to filter out irrelevant or low-quality studies, ensuring that the review focuses on significant contributions. Studies focusing solely on traditional, non-AI-based cybersecurity approaches were not used. Articles without empirical data, such as opinion pieces, editorials, and review papers, were excluded. Research focused on theoretical models without practical validation or application in cyber risk scenarios was not included. Also, duplicate studies from different databases were identified and excluded using reference management software, ensuring a unique set of studies for analysis.

Article DOI: 10.52589/BJCNIT-1HH9NPSN
DOI URL: https://doi.org/10.52589/BJCNIT-1HH9NPSN

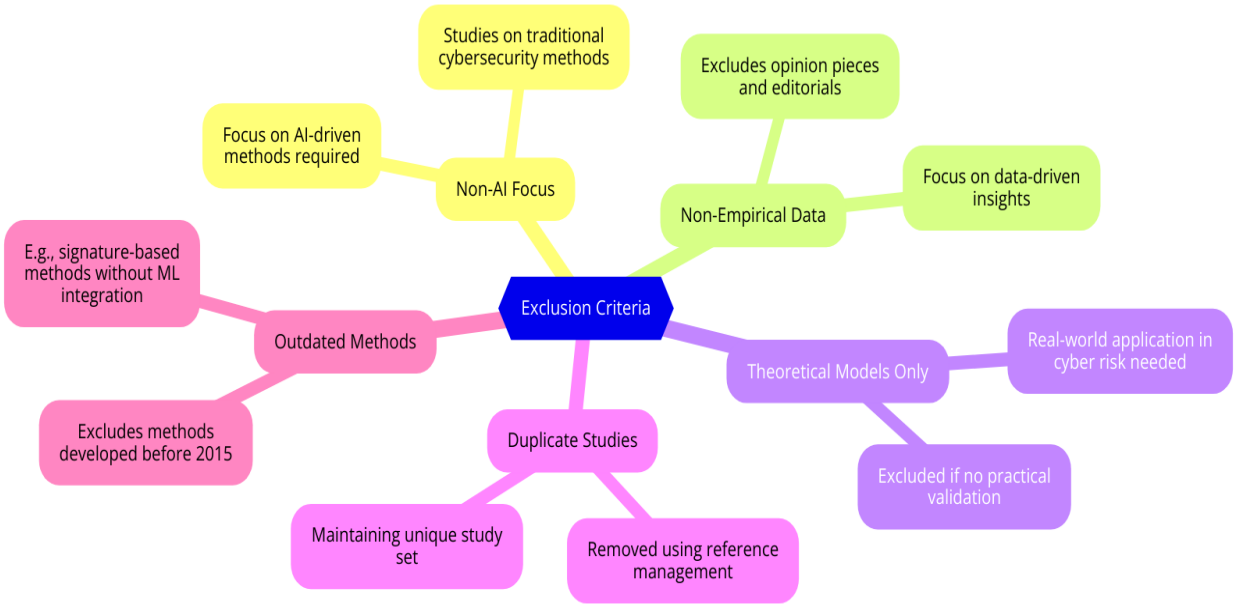

Figure 2: Exclusion Criteria Diagram

Data Extraction Criteria

Data extraction is the process of systematically collecting and categorising data from the included studies [16]. Data extraction was conducted systematically using a predefined template to ensure consistency. The extracted information included the study's title, authors, publication year, AI techniques (e.g., machine learning algorithms, neural networks), type of cyber risks addressed (e.g., phishing, malware detection, data breaches), evaluation metrics, and key findings. Two reviewers independently verified data extraction to minimise errors and ensure accuracy. Discrepancies between reviewers were resolved through discussion and consultation with a third reviewer.

Quality Assessment

To assess the quality of the included studies, the Critical Appraisal Skills Programme (CASP) checklist for quantitative research was used [15]. This checklist evaluates studies based on criteria such as clarity of research objectives, appropriateness of study design, validity of methods, and robustness of results.

Data Synthesis and Analysis

Data synthesis was performed using both narrative synthesis and meta-analysis, where appropriate. A narrative synthesis helped identify common themes, AI techniques, and application areas across the studies. At the same time, a meta-analysis was conducted to statistically aggregate findings related to the effectiveness of different AI models in mitigating specific cyber risks.

Article DOI: 10.52589/BJCNIT-1HH9NPSN
DOI URL: https://doi.org/10.52589/BJCNIT-1HH9NPSN

RESULT AND DISCUSSION

Effectiveness of AI Models in Cybersecurity

In this narrative analysis, we explore the effectiveness of AI models in cybersecurity, synthesising insights from multiple studies. The discussion will focus on machine learning approaches, their adaptability to evolving threats, and their success in threat mitigation.

Supervised Learning

Supervised Machine Learning (SL) demonstrates high effectiveness in cybersecurity, particularly in controlled environments. SL often achieves over 90% accuracy in detecting cyberattacks, although these results can be misleading in real-world conditions, where complexities may reduce effectiveness [16].

In Intrusion Detection Systems (IDS), SL models such as Random Forest (RF), Decision Tree (DT), and Support Vector Machine (SVM) have achieved over 99% classification accuracy, highlighting their reliability in detecting known attacks (Aamir et al., 2021) [18]. However, SL models struggle with novel threats like zero-day vulnerabilities, as they rely heavily on labelled datasets (Meaad et al., 2024) [19].

SL performs well in well-defined environments but lacks adaptability in dynamic areas like IoT and cloud systems, where new attack vectors frequently emerge without historical data. This limits its effectiveness in detecting evolving threats [20].

Unsupervised Learning

Unsupervised learning, which does not require labelled datasets, is more suited for anomaly detection in dynamic environments where new threats continuously emerge. This model can identify unusual patterns that may indicate new or unknown threats, making it particularly useful in environments like IoT and cloud systems, where cyber threats evolve rapidly [19]. This approach's strength lies in its ability to uncover hidden structures in data, allowing it to detect zero-day attacks and Advanced Persistent Threats (APTs) without relying on predefined signatures [21]. Unsupervised learning can lead to a higher rate of false positives, as benign anomalies may be mistakenly flagged as threats. These false positives require further investigation by human analysts, which can delay response times and reduce the model's overall efficiency.

Deep Learning

Deep learning has dramatically improved real-time threat detection in cybersecurity, with models demonstrating high effectiveness in identifying anomalies (98.5%) and malware (99.2%) [23]. These models also perform well in phishing detection (95.8%) and threat intelligence analysis, allowing organisations to respond quickly to evolving threats. However, challenges such as interpretability and robustness must be addressed for further optimisation.

Deep learning models, particularly CNNs, are highly effective in detecting regular malware (nearly 100%) and obfuscated malware (93.55%), enhancing precision in real-time applications [24]. Their study used real-world malware samples, improving cybersecurity performance. Deep learning's ability to process complex data and recognise patterns makes it especially useful in environments like cloud computing and critical infrastructure for detecting

Article DOI: 10.52589/BJCNIT-1HH9NPSN
DOI URL: https://doi.org/10.52589/BJCNIT-1HH9NPSN

sophisticated threats such as APTs and zero-day vulnerabilities [25].

Comparison of Machine Learning Techniques in Cybersecurity

In this thematic analysis, the researcher explores the strengths and weaknesses of different machine learning (ML) methods—supervised, unsupervised, and deep learning—in cybersecurity, focusing on their performance regarding false positives, detection accuracy, and computational cost. By comparing these methods across environments like cloud computing, IoT, and critical infrastructure, we can assess their suitability for addressing evolving cyber threats.

Supervised Learning Strengths and Weaknesses

Strengths: High Accuracy: Supervised learning provides high detection accuracy for known threats, as shown [26],[27]. These models excel in recognising malware and phishing attacks by drawing on labelled datasets.

Weaknesses: Inflexibility for Novel Threats: Supervised learning struggles with novel or zero-day attacks that do not appear in the training data. This lack of adaptability makes it less effective in dynamic environments like IoT and cloud systems.

Data Dependency: It relies heavily on labelled datasets, which can be challenging to obtain, especially for new or rare attack types. This limitation reduces the model's generalisation ability in unknown contexts [22].

Unsupervised Learning Strengths and Weaknesses

Strengths: Effective for Novel Threats: Unsupervised learning excels at detecting zero-day attacks and anomalies in environments where the threat landscape is constantly changing, such as IoT and cloud computing [27].

Adaptability: This technique can adapt to new, previously unseen attack patterns without relying on predefined threat signatures [28].

Weaknesses: High False Positives: One significant drawback is the higher rate of false positives, as benign deviations are sometimes flagged as threats. This can lead to alert fatigue and inefficient use of security resources [22].

Deep Learning Strengths and Weaknesses

Strengths: High Complexity Handling: Deep learning models can process intricate data structures and identify patterns that simpler models might miss. This makes them highly effective in detecting sophisticated attacks like Advanced Persistent Threats (APTs) and botnets [27].

Weaknesses: High Computational Cost: Training deep learning models requires substantial computational resources and time, making them less suitable for real-time applications where speed is crucial [28].

Data Dependency: Similar to supervised learning, deep learning models rely on large amounts of data for training, which can be a limitation in environments with limited data availability or

Article DOI: 10.52589/BJCNIT-1HH9NPSN
DOI URL: https://doi.org/10.52589/BJCNIT-1HH9NPSN

where labelled data is scarce.

Table 3: Comparison of Models

Technique	Strengths	Weaknesses	Best-Suited Usage
Supervised Learning	High accuracy for known threats	Inflexible for novel threats, data-dependent	Network security, well-defined threat spaces
Unsupervised Learning	Adaptable to novel threats, suitable for dynamic environments	A high false positive rate requires validation	IoT, cloud computing, anomaly detection
Deep Learning	Handles complex data, scalable for large datasets	High computational cost, data-heavy	Critical infrastructure, APT detection

Framework for Integrating AI-Driven Models with Traditional Security Measures

In today's cybersecurity landscape, addressing both known and unknown threats requires a defence system that combines the strengths of AI-driven models with traditional security measures. This integrated, multi-layered approach offers resilience against a wide range of cyber threats, from common malware attacks to sophisticated zero-day exploits and Advanced Persistent Threats (APTs). The framework proposed here leverages machine learning (ML) and deep learning (DL) models for dynamic detection and response while maintaining traditional security systems' robust, rule-based mechanisms.

Layered Security Architecture

The framework is based on a layered security architecture that incorporates AI-driven models at various levels of the security stack, working alongside traditional measures like firewalls, intrusion detection systems (IDS), and antivirus solutions. The goal is to ensure that AI and traditional methods complement each other, improving overall defence effectiveness.

Layer 1: Perimeter Defense with Traditional Methods

Objective: Establish a solid first line of defence using conventional measures like firewalls, VPNs, and IDS to block known threats.

Mechanisms: Firewalls and Antivirus Software: These systems handle signature-based threat detection, which is highly effective for known attacks. They use databases of known malware signatures and rules to block unauthorised access [28].

Intrusion Detection Systems (IDS): Traditional IDS tools monitor network traffic for known attack patterns or signature-based anomalies. Though effective against common attacks, they struggle with zero-day threats.

Layer 2: AI-Powered Anomaly Detection and Behavioral Analytics

Objective: Use AI-driven models to monitor network behaviour and detect anomalies that may indicate unknown threats.

Article DOI: 10.52589/BJCNIT-1HH9NPSN
DOI URL: https://doi.org/10.52589/BJCNIT-1HH9NPSN

Mechanisms: Unsupervised Learning for Anomaly Detection: AI models using unsupervised learning are integrated with traditional IDS to detect deviations from normal behaviour that might indicate emerging or unknown threats. These models are highly adaptable and can identify zero-day attacks and insider threats in real time [27].

Layer 3: Machine Learning-Enhanced Threat Detection

Objective: Leverage supervised learning models to detect and respond to known threats more efficiently.

Mechanisms: Supervised Learning Models: By continuously learning from labelled datasets, these models offer high detection accuracies for known threats, such as malware and phishing attacks [22].

Hybrid Threat Detection: AI models are designed to work in tandem with signature-based detection systems. While signature-based models handle known attacks, AI models provide an additional layer of scrutiny for behaviours that deviate from known patterns.

Layer 4: Deep Learning for Advanced Threats and Big Data Analysis

Objective: Address complex threats like APTs and botnets through deep learning techniques.

Mechanisms: Deep Learning for Large-Scale Threats: Deep learning models such as CNNs and RNNs are employed to analyse vast amounts of network traffic and complex data patterns, particularly in cloud environments and IoT networks. These models can detect multi-stage attacks like APTs by recognising patterns that evolve over time [22].

Adversarial Machine Learning: To combat attacks aimed at exploiting AI vulnerabilities, adversarial machine learning is integrated into the system. This technique enhances the resilience of AI-driven models against attacks that seek to manipulate ML algorithms, such as adversarial inputs.

Real-Time Response and Automation

This framework emphasises the automation of AI-powered response mechanisms and human oversight for critical decisions to ensure effective real-time threat mitigation.

Autonomous Defense Systems with Reinforcement Learning

Reinforcement learning (RL) models are introduced to automate real-time attack responses. RL agents continuously adapt their defence strategies by learning from past attacks and evolving threats. In environments where quick decision-making is essential (e.g., critical infrastructure), these models can autonomously adjust security protocols (firewalls, access controls) to neutralise threats while reducing the workload on human operators.

Hybrid Security: Combining AI and Human Expertise

A crucial aspect of this framework is the collaboration between AI models and human analysts. While AI excels in processing large amounts of data and detecting complex patterns, human intuition and contextual understanding remain irreplaceable in specific scenarios.

Article DOI: 10.52589/BJCNIT-1HH9NPSN
DOI URL: https://doi.org/10.52589/BJCNIT-1HH9NPSN

Human-AI Collaboration

AI models handle the bulk of threat detection, filtering out false positives and identifying potential threats, while human analysts focus on high-priority incidents and complex decision-making (Onuh et al., 2024). This hybrid approach ensures the defence system benefits from AI's efficiency without sacrificing human operators' insight and expertise.

Table 4: Framework Overview: Multi-Layered Defense System

Layer	Objective	Technology/Method	Key Features
Layer 1: Perimeter Defense	Block known threats	Firewalls, IDS, Antivirus systems	Signature-based detection, rule-based protection
Layer 2: Anomaly Detection	Detect unknown threats via behavioural analysis and anomaly detection	Unsupervised learning, Behavioral Analytics	Real-time anomaly detection, insider threat detection
Layer 3: Threat Identification	Identify known threats with high accuracy	Supervised learning models	High accuracy for known threats, reduced false positives
Layer 4: Advanced Threat Detection	Detect complex threats like APTs	Deep learning, CNNs, RNNs note that CNN and RNN are types of deep learning architecture. CNN handles images, audio better while RNN is for natural language processing.	Complex data processing, scalable threat detection
Layer 5: Real-Time Response	Automate defence actions based on real-time threats	Reinforcement learning, Automated response systems	Autonomous threat mitigation, real-time adaptability
Layer 6: Human-AI Collaboration	Ensure accurate and contextual decision-making	Explainable AI, Human-AI collaboration	Human oversight, transparent AI decisions

CONCLUSION

The analysis of various AI models in cybersecurity reveals that each approach, like supervised, unsupervised, and deep learning, has distinct strengths and limitations that influence their effectiveness in different environments. Supervised learning is best suited for detecting well-documented threats in stable environments due to its high accuracy, but it lacks flexibility when addressing novel attacks. Unsupervised learning offers adaptability for emerging threats and dynamic environments like IoT, though it struggles with high rates of false positives that require human validation. Deep learning excels in processing complex data and detecting sophisticated threats, making it suitable for critical infrastructure and cloud computing.

Article DOI: 10.52589/BJCNIT-1HH9NPSN
DOI URL: https://doi.org/10.52589/BJCNIT-1HH9NPSN

However, its computational intensity can be a drawback in real-time responsiveness scenarios. Overall, the choice of AI model should be aligned with the specific cybersecurity challenges and the nature of the environment to maximise threat detection and mitigation capabilities.

REFERENCES

[1]X. Meng, "Advanced AI and ML techniques in cybersecurity: Supervised and unsupervised learning, reinforcement learning, and neural networks in threat detection and response," *Applied and Computational Engineering*, vol. 82, no. 1, pp. 1–5, Jul. 2024, doi: https://doi.org/10.54254/2755-2721/82/2024glg0054.

[2]R. Kaur, D. Gabrijelčič, and T. Klobučar, "Artificial Intelligence for Cybersecurity: Literature Review and Future Research Directions," *Information Fusion*, vol. 97, no. 101804, p. 101804, 2023, doi: https://doi.org/10.1016/j.inffus.2023.101804.

[3]A. Jarrar, R. Tahril, A. Lasbahani, and Y. Balouki, "Using Deep Learning Algorithm in Security Informatics," *International Journal of Innovative Science and Research Technology (IJISRT)*, pp. 2933–2944, May 2024, doi: https://doi.org/10.38124/ijisrt/ijisrt24apr2271.

[4]S. Han, H. Yun, and Y. Park, "Deep Learning for Cybersecurity Classification: Utilizing Depth-Wise CNN and Attention Mechanism on VM-Obfuscated Data," *Electronics*, vol. 13, no. 17, p. 3393, Aug. 2024, doi: https://doi.org/10.3390/electronics13173393.

[5]O. Alshaikh, S. Parkinson, and S. Khan, "On the Variability in the Application and Measurement of Supervised Machine Learning in Cyber Security," *Communications in computer and information science*, pp. 545–555, Jan. 2023, doi: https://doi.org/10.1007/978-981-99-0272-9_38.

[6]A. S. Ahanger, S. M. Khan, and F. Masoodi, "An Effective Intrusion Detection System using Supervised Machine Learning Techniques," *IEEE Xplore*, Apr. 01, 2021. https://ieeexplore.ieee.org/document/9418291 (accessed Oct. 11, 2024).

[7]M. Sankaram, M. Roopesh, S. Rasetti, and N. Nishat, "A COMPREHENSIVE REVIEW OF ARTIFICIAL INTELLIGENCE APPLICATIONS IN ENHANCING CYBERSECURITY THREAT DETECTION AND RESPONSE MECHANISMS," *Deleted Journal*, vol. 3, no. 5, pp. 1–14, Jul. 2024, doi: https://doi.org/10.62304/jbedpm.v3i05.180.

[8]S.-F. Wen, A. Shukla, and B. Katt, "Artificial intelligence for system security assurance: A systematic literature review," *Research Square (Research Square)*, Jul. 2024, doi: https://doi.org/10.21203/rs.3.rs-4589465/v1.

[9]M. Yakubu Bala, S. Badara, and H. Dan'azumi, "An Intelligence-Based Cybersecurity Approach: A Review," *Journal of Intelligent Communication*, vol. 4, no. 1, pp. 32–43, May 2024, doi: https://doi.org/10.54963/jic.v4i1.232.

[10]S. Basak, P. Chatterjee, D. Biswas, R. Das, and P. Bhadra, "Introduction to AI and ML Technologies and Their Potential Applications in Cybersecurity," *Advances in web technologies and engineering book series*, pp. 277–309, Aug. 2024, doi: https://doi.org/10.4018/979-8-3693-6557-1.ch012.

[11]H. S. Venter and N. Rananga, "A comprehensive review of machine learning applications in cybersecurity: identifying gaps and advocating for cybersecurity auditing," *Research Square (Research Square)*, Aug. 2024, doi: https://doi.org/10.21203/rs.3.rs-4791216/v1.

[12]P. Prakriti, "Cyber Threat Detection Using Machine Learning," *INTERANTIONAL*

Article DOI: 10.52589/BJCNIT-1HH9NPSN
DOI URL: https://doi.org/10.52589/BJCNIT-1HH9NPSN

JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT, vol. 08, no. 07, pp. 1–15, Jul. 2024, doi: https://doi.org/10.55041/ijsrem36799.

[13]E. Liu, "Early detection and mitigation of cyber attacks with machine learning and artificial intelligence," *Applied and Computational Engineering*, vol. 73, no. 1, pp. 261–266, Jul. 2024, doi: https://doi.org/10.54254/2755-2721/73/20240409.

[14]C. S. Kodete, B. Thuraka, V. Pasupuleti, and S. Malisetty, "Determining the Efficacy of Machine Learning Strategies in Quelling Cyber Security Threats: Evidence from Selected Literatures," *Asian Journal of Research in Computer Science*, vol. 17, no. 8, pp. 24–33, Jul. 2024, doi: https://doi.org/10.9734/ajrcos/2024/v17i7487.

[15]N. Varatharajan, S. Lavanya, A. Suganya, and R. Vikkram, "Deep Learning," *Auerbach Publications eBooks*, pp. 1–8, Jul. 2024, doi: https://doi.org/10.1201/9781003433309-1.

[16]Kaythry. P., "A Review of Deep Learning Strategies for Enhancing Cybersecurity in Networks," *Journal of Scientific & Industrial Research*, vol. 82, no. 12, Dec. 2023, doi: https://doi.org/10.56042/jsir.v82i12.1702.

[17]A. Makhlouf, B. Sedraoui, A. Benmachiche, and C. Chemam, "Intrusion Detection with deep learning: A literature review," vol. 7, pp. 1–8, Apr. 2024, doi: https://doi.org/10.1109/pais62114.2024.10541191.

[18]M. Kaniški, J. Dobša, and D. Kermek, "Deep Learning within the Web Application Security Scope – Literature Review," May 2023, doi: https://doi.org/10.23919/mipro57284.2023.10159847.

[19]M. A. Khalaf and A. Steiti, "Artificial Intelligence Predictions in Cyber Security: Analysis and Early Detection of Cyber Attacks," *Babylonian Journal of Machine Learning*, vol. 2024, pp. 63–68, May 2024, doi: https://doi.org/10.58496/bjml/2024/006.

[20]H. N. Fakhouri, K. Omar, F. Hamad, N. Z. Halalsheh, B. Alhadidi, and S. Makhadmeh, "AI-Driven Solutions for Social Engineering Attacks: Detection, Prevention, and Response," Feb. 2024, doi: https://doi.org/10.1109/iccr61006.2024.10533010.

[21]A. Patel, P. Pali, S. Verma, S. Tiwari, V. Vishwakarma, and S. Chourasiya, "Artificial Intelligence Strategies for Combating IoT-Centric Cyber Attacks," *International Journal of Innovative Research in Science Engineering and Technology*, vol. 12, no. 05, pp. 8050–8056, Nov. 2023, doi: https://doi.org/10.15680/ijirset.2023.1205496.

[22]J. Govea, W. Naranjo, and W. Villegas, "Transforming Cybersecurity into Critical Energy Infrastructure: A Study on the Effectiveness of Artificial Intelligence," *Systems*, vol. 12, no. 5, p. 165, May 2024, doi: https://doi.org/10.3390/systems12050165.

[23]S. Kumar, H. Pathak, L. Awasthi, A. Rai, K. Awasthi, and A. Bajpai, "Cyber Crime Prevention Model Using Artificial Intelligence," Nov. 2023, doi: https://doi.org/10.53555/jchr.v13.i4s.1660.

[24]M. Onuh, I. Idoko , E. Godslove , O. Frederick , O. Timilehin , and U. Chukwunonso , "Harnessing adversarial machine learning for advanced threat detection: AI-driven strategies in cybersecurity risk assessment and fraud prevention," *Open Access Research Journal of Science and Technology*, vol. 11, no. 1, pp. 001–004, May 2024, doi: https://doi.org/10.53022/oarjst.2024.11.1.0060.

[25]Al. Sukhvinder, "Neural Networks and Cyber Resilience: Deep Insights into AI Architectures for Robust Security Framework," *Deleted Journal*, vol. 19, no. 3, pp. 78–95, Jan. 2024, doi: https://doi.org/10.52783/jes.653.

[26]F. Alwahedi, A. Aldhaheri, M. Ferrag, and A. Battah, "Deep learning for cyber threat detection in IoT networks: A review," *Internet of Things and Cyber-Physical Systems*, vol. 4, pp. 110–128, Jan. 2024, doi: https://doi.org/10.1016/j.iotcps.2023.09.003.

[27]V. Mygal, G. Mygal, and S. Mygal, "AI: Unique Opportunities and Global Challenges –

Article DOI: 10.52589/BJCNIT-1HH9NPSN
DOI URL: https://doi.org/10.52589/BJCNIT-1HH9NPSN

A Hybrid Approach to Modeling Reality and Its Perception," *Qeios*, Jul. 2024, doi: https://doi.org/10.32388/gij3ri.4.

[28]A. Anusooya, N. Revathi, S. P, A. N. Duraivel, and S. Prabu, "A Hybrid Genetic Algorithm and Neural Network-Based Cyber Security Approach for Enhanced Detection of DDoS and Malware Attacks in Wide Area Networks," *Journal of Cybersecurity and Information Management*, vol. 14, no. 2, pp. 253–262, Jan. 2024, doi: https://doi.org/10.54216/jcim.140217.

[29]D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA Statement," *PLoS Medicine*, vol. 6, no. 7, Jul. 2009, doi: https://doi.org/10.1371/journal.pmed.1000097.

[30]L. Schmidt, N. Finnerty, R. W. Elmore, B. Olorisade, J. Thomas, and J. P. T. Higgins, "Data Extraction Methods for Systematic Review (semi)automation: Update of a Living Systematic Review," *F1000Research*, vol. 10, no. 401, pp. 401–401, Oct. 2023, doi: https://doi.org/10.12688/f1000research.51117.2.

[31]Z. E. Rasjid and R. Setiawan, "Performance Comparison and Optimization of Text Document Classification using k-NN and Naïve Bayes Classification Techniques," *Procedia Computer Science*, vol. 116, pp. 107–112, 2017, doi: https://doi.org/10.1016/j.procs.2017.10.017.

[32]A. Al-Subaiey, M. Al-Thani, N. Abdullah Alam, K. Fatema Antora, and A. Khandakar, "Novel interpretable and robust web-based AI platform for phishing email detection," *Computers & Electrical Engineering*, vol. 120, pp. 109625–109625, Sep. 2024, doi: https://doi.org/10.1016/j.compeleceng.2024.109625.

[33]I. Pérez-Martínez, M. Martínez-Rojas, and J. M. Soto-Hidalgo, "A methodology for urban planning generation: A novel approach based on generative design," *Engineering Applications of Artificial Intelligence*, vol. 124, p. 106609, Sep. 2023, doi: https://doi.org/10.1016/j.engappai.2023.106609.

[34]U. Sakthivelu and C. N. S. Vinoth Kumar, "Advanced Persistent Threat Detection and Mitigation Using Machine Learning Model," *Intelligent Automation & Soft Computing*, vol. 36, no. 3, pp. 3691–3707, 2023, doi: https://doi.org/10.32604/iasc.2023.036946.

Article DOI: 10.52589/BJCNIT-1HH9NPSN
DOI URL: https://doi.org/10.52589/BJCNIT-1HH9NPSN