Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 8, 2025

INVESTIGATING THE EFFECTIVENESS OF ARTIFICIAL INTELLIGENCE IN DETECTING ZERO-DAY ATTACKS

Nisar Ahmed Memon[*1], Mariya Sultana[2], Engr. Afshan Ahmed Siddiqui[3], Muniba Murtaza[4]

*[*1]Assistant Professor, Department of Telecommunication Engineering*
Faculty of Engineering and Technology, University of Sindh Jamshoro.
[2]BS in Computer Science and Engineering, International Islamic University Chittagong, Bangladesh.
[3]PhD Scholar, Xian University of Science and Technology
[4] Information, and Communication Technology, Future Training and Services Center, Abu Dhabi.

[*1]nisar.memon@usindh.edu.pk, [2]mariasultanafahmida@gmail.com, [3]afshaan_ahmed@hotmail.com, [4]muniba@ftcglobal.ae

DOI: https://doi.org/10.5281/zenodo.16930991

Keywords
Applicability, artificial intelligence, algorithms, detection, zero-day attacks.

Article History
Received: 23 May, 2025
Accepted: 30 July, 2025
Published: 23 August, 2025

Corresponding Author: *
Nisar Ahmed Memon

Abstract
This is an in-depth study on the applicability of artificial intelligence algorithms on the detection of zero-day attacks by using both quantitative experimental studies and qualitative expertise rating. The study tested five different AI-based algorithms Random Forest, Support Vector Machines, Long Short-Term Memory networks, Convolutional Neural Networks and ensemble models on detection performance across synthetically produced network traffic, a real-world organizational dataset and publicly available cybersecurity training datasets. The findings revealed that the best performance of detection accuracy belonged to ensemble techniques (87.3 percent), and LSTM networks (82.1 percent), and conventional machine learning algorithms were relatively low, with Random Forest having 76.8 percent detection accuracy. An analysis of real-world deployment showed that there were formidable obstacles in the form of a high false positive (12.4 to 23.7 percent depending on the algorithm) as well as significant demands on computational resources. The qualitative analysis involving interviews with experts pointed out some important implementation obstacles such as the complexity of integration, employee training requirements and the necessity to retrain the models continuously. The study was able to find that AI systems have demonstrated reasonable potential to detect new and previously unfamiliar discrimination behaviors but, in contrast, hybrid methods based on a combination of various algorithms with human knowledge have resulted in the most consistent breach detection. The results provide new information to the cybersecurity knowledge base on how beneficial the implementation of AI-based zero-day detection solutions will be and on how future research on automated threat detection solutions can be conducted

INTRODUCTION

Since the last decade, cybersecurity has been changing radically due to the contribution of threat actors who have been relying on the modern way of organizing malicious tactics that

are used to avoid common security protocols (El Husseini, Noura et al. 2024). Zero-day attacks usually present some of the highest risks in the contemporary organizations by targeting vulnerabilities in the software systems that are not known to the software developers yet and thus they have not developed and implemented any form of patch to resolve the situation (Ahmad, Alsmadi et al. 2023). The attacks reflect a grave security issue since they exploit the vulnerabilities whose existence has been unidentified, and thus are virtually undetectable by signature-based security systems, which use identifiable patterns to detect the attack. Zero-day attacks have an exponentially increasing financial impact, with organizations in 2022 losing an average of 4.35 million dollars in a data breach and it took on average 277 days to identify and encapsulate a breach (Zoppi, Ceccarelli et al. 2021).

Previous methods of cybersecurity have shown ineffective against zero-day threats because they are based on pre-programmed signatures and known pathways of an attack (Ekong, Etuk et al. 2023). Traditional antivirus programs, IDS and firewall settings are characterized by blacklist choices, that is, recognizing malicious actions by using prior-generated threat tracks. Such reactive measures establish a window of vulnerability where there exists a weakness through which zero-day attacks can go on until an appropriate measure is established to curtail these attacks and hence, may are subject to serious damages in organizational assets and infrastructure. The shortcomings associated with signature-based detection have influenced the need to seek more proactive and dynamic security systems, among cybersecurity researchers and practitioners (Zhou 2022).

Artificial intelligence has become a potential avenue to overcome the weaknesses of the conventional cybersecurity innovations as they involve the possibility of identifying unknown threats based on pattern representation, behavioral representation, and anomaly detection (Jimmy 2021). Machine learning algorithms have the ability to learn through large quantities of information stored in network traffic, system logs and user activity and create baselines of normal operating activities. In the event of deviations of such established patterns, the AI systems can mark potentially malicious activities that should be investigated even in the case when the particular attack vectors have not been previously observed. This is a paradigm shift in terms of being reactive to threats as opposed to being proactive in terms of detecting threats (Zhang, Ning et al. 2022).

The role of AI in cybersecurity has received a lot of attention both in the academic circles and with practitioners and the research has already exhibited considerable potential in different threat detection environments. Neural networks, and indeed deep learning algorithms more generally, have proven the most effective at processing high-dimensional and high-dimensionality data and, in particular, finding subtle patterns that may otherwise elude human analysts. Supervised learning has shown promise in the presence of an adequate amount of labeled training data, and unsupervised learning has potential in identifying entirely new (i.e., never seen) kinds of attack signatures without training based on particular threat signatures (Aslam, Aslam et al. 2025).

The deployment of AI-driven solutions to detecting criminals is not without a number of challenges that need to be overcome to guarantee feasible efficacy (Aslam, Aslam et al. 2025). False positive rates are also paramount since scheduling genuine actions as maliciously caused actions might flood the security team and makes them experience alert fatigue (Aslam, Aslam et al. 2025). Advanced attackers started creating adversarial methods with the sole intention of being able to pass through the AI-based detection methods making it an arms-race between malicious actors and defenders. The complex AI algorithms may also present resource-constrained organizations with practical deployment complications because of the computational demands of complex AI algorithms (Nkongolo and Tokmak 2023).

The evolving characteristics of network environments in the modern world make AI-based threat detection another level of

complexity (Aslam, Aslam et al. 2025). The topology of the networks, the use of the applications or user habits are changing continuously, and hence AI systems must continually adjust their models of detection. Tasks such as needing to constantly learn something new, and updating the models thus create an operational overhead and also requires specialization which is not available to most organizations. Moreover, it is also essential to note that the interpretability of AI decisions is taken so seriously because security teams are expected to know why their certain activities can be labeled as suspicious and malicious in order to make informed decisions regarding how to respond (Guo 2023).

The advantages of AI-driven zero-day detection systems are enough to persuade researchers to keep working on further research. Possibility to detect hitherto undetected threats may help greatly shorten the window of vulnerability that rendered organizations susceptible whenever new exploits are discovered. Besides, AI-based systems are able to process and analyze a magnitude and rate of information that a human cannot and can, theoretically, detect and respond to the threat in real-time. Indeed, the combination of various AI approaches via ensemble methods has proven especially promising in enhancing detection accuracy at the same time being able to control false positive rates (GUPTA, AKTER et al. 2023).

The technology has seen more investments in the segment of security solutions powered with AI and more vendors are integrating machine learning to their products. Nevertheless, the quality of these commercial solutions is very different, and it is necessary to conduct independent research that would focus on their performance during the real operation. Theoretical research that has been conducted in academia has taught us a lot about the theoretical performance of different algorithms of AI, but implementation issues in the real world are sometimes different to those that are controlled in a laboratory (Igugu 2024).

Areas of existing research gaps are the lack of assessment of AI systems under powerful adversarial attacks, the lack of study of long-term behavior in dynamic settings, and a lack of attention to the costs of implementation and operational specifications. Also, many prior works have worked on a particular AI algorithm with less comparative research among other strategies under the same conditions. The use of quantitative parameters about performance in combination with qualitative reasoning conducted by the practitioners of cybersecurity is also absent in existing literature (Adavelli, Mittapelly et al. 2025).

Filling such gaps, this study evaluates several AI algorithms regarding zero-day attacks detection and offers a scientific, yet applicable, assessment of the currently available machine learning algorithms, including the expertise of the industry specialists. The purpose of the study is to enlighten both academic scientists and professionals working in the field of cybersecurity about the present-day level of the AI-based threat detection possibilities, along with the prospects and the restrictions it has on the profitable implementation. Combined with technical performance ratings with real-world implementation issues, this study has helped contribute to the fuller picture of AI and its role in the current cybersecurity defensive practices.

Research Objectives

1. To measure and assess the accuracy of detection of five separate artificial intelligence algorithms (Random Forest, Support Vector Machines, Long Short-Term Memory networks, Convolutional Neural Networks, and ensemble methods) on the ability to detect zero-day attacks in different network settings as well as conditions of the attack.
2. To examine the challenges and demands of AI-based zero-day detection systems with the help of experts and implementation tests, such as resources consumption, false positive handling, and difficulty in the integration process.
3. To formulate evidence-based decision rules on the best AI algorithm to choose and implementation strategies to use that could balance off between the effectiveness of the algorithm in detection and operational ease to

organizations interested in implementing AI-based cybersecurity solutions.

Research Questions

1. Which AI algorithms have proven to be most effective in identifying zero-day attacks and what are the performance measures of these algorithms in identifying different attacks in different environments of the network and the complexity of attacks.
2. What are the major issues and obstacles faced by cybersecurity experts in deploying the AI-driven zero-day detection systems in the real world in organizations?
3. What can organizations do to maximize the effectiveness of their AI-driven cybersecurity defenses to detect zero-day attacks as stealthily as possible without falsely reporting attacks and overhead costs?

Significance of the Study

This study is of immense importance to various stakeholders in the cybersecurity landscape that it delivers important insights on the further development of theoretical knowledge and the application of smart threat detection systems powered by AI. In context of academic researchers, it will cover the expanding universe of knowledge within the arena of AI applied to cybersecurity as it provides a deep comparative analysis of several algorithms under controlled conditions closing the existing gaps in terms of empirical assessment techniques. The methodology of the present research (mixed-methods) provides a structure that interested researchers can use in the future when trying to determine the blend between technical performance of a protocol and real world constraints within implementation, resulting in a more holistic approach to evaluations in cybersecurity research. The findings offer evidence to guide cybersecurity practitioners or an organization decision-maker about the necessary alternatives of AI technologies and implementation strategies that they identify as suitable that could alleviate the trial-and-error cycle that is usually witnessed when implementing a new technology. The applied approach with regard to challenges of practical deployment helps to fill important knowledge gaps existing between laboratory research and the reality of operations, providing insights about practical ways to face the general impediments to adoption. The research also helps to solve the problem of enhanced security postures in organizations by offering validated techniques of identifying previously unknown threats, which may minimize destruction of finances and brands that are affected by the success of a zero-day attack. The theory in relevance to the wider cybersecurity sector is that the results could shape the approach of security vendors in developing their products, as well as contribute to premises advanced in policy-related decision-making with references to the AI governance of critical infrastructure protection.

Literature Review

Artificial intelligence in cybersecurity has recently become an area of fast growth as scholars consider different machine learning models to overcome the shortcomings of conventional threat detection systems. Initial work in this field concentrated mainly on supervised learning approaches, in which algorithms were trained using labeled data on known malicious and benign traffic patterns in the network. The earliest contribution of Denning and Neumann laid down fundamental ideas of intrusion detection systems, which was further refined by other researchers by adding statistical based analysis and using methods of pattern recognition. The introduction of machine learning methods into the study of cybersecurity marked a turning point because the research study showed the increased detection of known attack patterns (Okoli, Obi et al. 2024).

Later in the late 1990s and early 2000s, neural networks emerged in the field of cybersecurity where they attracted the attention of researchers who began investigating its applicability in detecting anomalies and recognizing patterns. The findings of the above research by Cannady revealed that employment of artificial neural networks in intrusion detection was feasible to detect intrusion attempts and misuse to systems. This work was subsequently extended in studies

by Mukkamala and colleagues that compared various network architectures and training strategies, in the context of cybersecurity. The works of these pioneers set crucial principles of data preprocessing, feature selection, and network topology optimization that survive to impact the state of AI-based cybersecurity systems research today (Zhang, Ning et al. 2022).

Deep learning technologies completely changed the face of AI-based threat detection, and researchers looked into convolutional neural networks, recurrent neural networks, and complex cybersecurity data processing architectures. Comprehensive studies on the applications of deep learning in network intrusion detection were carried out by Wang and colleagues, which prove to perform better than conventional methods of machine learning. The paper that they published stressed the need to use feature engineering and data representation to obtain maximal detection results. At the same time, the Long Short-Term Memory networks started being explored to detect cybersecurity attacks based on sequences as most cybersecurity activities exhibit a temporal aspect (Kavitha and Thejas 2024).

This is when ensemble methods were proposed as a solution to address the problem of accuracy of detection with control of the trade-offs existing between sensitivity and specificity for cybersecurity purposes. Research by Chen and Liu showed that combining two or more learning algorithms would perform better than single classifier and especially under the circumstance where there are a variety of different attack types and variations in network conditions. The idea of ensemble learning resonated with the uniqueness of cybersecurity spaces where one algorithm was not sufficient to serve the just amount all the threat scenarios. This field of research has tried several combination methods, such as voting, stacking methods, and dynamic ensemble selection (Hossain, Afrin et al. 2024).

The issue of zero-day attack detection drew particular attention because researchers have already learned about the inefficiency of signature-based solutions to deal with previously unknown threats. The foundations of intrusion detection systems as per the assumptions made were critically addressed by Sommer and Paxson studies and demonstrated the inherent challenges of detecting attacks patterns that are new. The findings of their work had impacts on the directions of the future work that revealed the importance of using anomaly-based detection mechanisms that would recognize abnormal behavior as opposed to searching known attack patterns. This change in the worldview allowed thorough research of unsupervised learning methods and methods of the behavioral analysis (Yaseen 2023).

The scale of anomaly detection research increased when it was realized that in cases of zero-day attacks the behavior patterns of the attack may be uncharacteristic of typical network usage even when the exact technical implementation is unidentified. Chandola et al. developed elaborated surveys of anomaly detection methods where taxonomy and evaluation schemes were proposed to impact later trends in cybersecurity studies. Their contribution emphasized the need to study various forms of anomalies and choose a suitable algorithm of detecting anomalies depending upon the user cases and environmental limitations. The combination of domain knowledge with statistical analysis and machine learning became a key focus of an anomaly detection research (Dari, Thool et al. 2023).

The professional application of artificial intelligence-based cybersecurity showed that there are many obstacles that do not manifest themselves well in closed laboratory conditions. Axelsson and others have studied the issue of base-rate fallacy in intrusion detection showing that when attack probabilities are very low in the real-world, false positive rates may be unacceptable with very high accuracy classifiers. The study noted the operational context and cost-benefit analysis as one of the most crucial factors to consider during the evaluation of the cybersecurity systems. The results of the studies impacted other studies to include more realistic assessment conditions and take into account possible deployment limitations (Raja 2025).

Adversarial machine learning New attackers developed techniques specifically designed to exploit detection systems developed with AI; the byproduct was that adversarial machine learning became a critical area of research (Aslam, Aslam et al. 2025). The research of Huang et al. was the first of its kind to attack machine learning classifiers in applications to cybersecurity by performing adversarial attacks in which they experimented with different tricks to hide input features and not be caught. Their paper discovered fundamental weaknesses of AI based systems of security, and stimulated a twofold going into the defensive mechanisms and solid learning algorithms. The oppositional machine learning paradigm added new measures of performance and testing methods that became the common rule in the study of AI in cybersecurity (Ajagbe, Adigun et al. 2023).

The research on the aspect of feature selection and engineering solved the important problem of defining feasible data attributes to accurately detect a threat without worrying about computational costs and interpretability of the model. Guyon and Elisseeff developed theory and cybersecurity scientists used it in analysis of network traffic, processing of system logs, and behavior modeling. Cybersecurity data was high-dimensional, which presented distinct problems to feature selection causing specialized methods to deal with temporal sequences, categorical variables, and low data densities (Salem, Azzam et al. 2024).

The areas of study in respect to AI-based cybersecurity systems based on real-time processing capabilities and scalability issues have become more and more topical in the most recent researches. Vasilomanolakis and colleagues studied the computational needs and processing delays that have been linked to various machine learning algorithms when it comes to practical operating conditions. Inventoried work in their field revealed the trade-offs inherent between detection accuracy and system performance that were driving the optimization of algorithms and distributed processing architectures. Prioritizing real-time capability signifies the practical necessity to ensure that the cybersecurity systems have the ability to advance timely threat detection and action, and they can offer accurate protection and threat detection (JothiShri, Upender et al. 2024) .

Multimodal analysis and the incorporation of several data sources have emerged as a significant research direction as the accuracy of detection becomes a major concern of an organization to utilize comprehensive monitoring approaches (Mohamed 2025). Researchers have investigated networks traffic analysis, system logs correlations, user behavior analytics, and threats intelligence integration as a means to achieve better detection systems. This study has identified multiple advantages and opportunities related to multi-modal strategies as well as disadvantages and limitations such as high detection rates at the cost of complexity and investments.

Recent literature has been placing more importance on explainable AI in cybersecurity use cases where it has become clear that security analysts require a method of comprehending and trusting AI-generated alerts so that the correct decisions can be made on how to respond. Pieces done by Arrieta and others have defined inferable machines in high-stakes usage and threat surveys, and cybersecurity experts have applied the ideas to the idea of threat identification. Creation of explainable AI methods is a new research problem that would be balanced between model efficacy and interpretability needs, especially in regulated industries and civil infrastructure security situations.

Research Methodology

The researchers have adopted mixed-method research strategy as it entailed both quantitative analysis on the basis of experiments as well as qualitative evaluation of AI detection systems in various controlled settings. Datasets were gathered by researchers in three main categories: synthetic network traffic with zero-day attacks simulation, real network logs implemented at participating organizations during six months, and publicly available data sets of cybersecurity such as DARPA and KDD Cup archives. The methodology used a comparative framework in which the researchers used five different AI

algorithms Random Forest, Support Vector Machine, Long Short-Term Memory network, Convolutional Neural Network and ensemble methods on the same testbeds to measure detection accuracy, false positive rates and response time. In order to define baseline levels of performance, researchers trained each algorithm on known attack patterns and then applied previously unknown patterns representing zero-day attacks and measured the ability to detect them. The methodology involved use of cross-validation techniques where 20% test and 80% training sets were used with ten iterations of this data to make the methodologically significant. Network topology, volume of traffic, and the complexity of the types of attacks served as control variables whereas the dependent variables that were measured included accuracy of the detection, processing latency, and resource consumption. Also, the study has used expert interviewing approach of fifteen cybersecurity specialists with the purpose to determine the real-life deployment issues, and received feedback on false positive management tactics which eventually triangulated quantitative performance statistics with qualitative responses of industry experts in order to present a complete evaluation.

Results and Data Analysis

Quantitative Analysis

The quantitative analysis revealed significant variations in detection performance across the five AI algorithms tested in this study. The ensemble method demonstrated the highest overall detection accuracy, achieving 87.3% correct classification of zero-day attacks across all test scenarios. This superior performance was consistent across different network environments and attack complexity levels, indicating robust generalization capabilities. The Long Short-Term Memory network achieved the second-highest detection rate at 82.1%, demonstrating the effectiveness of sequential learning approaches for temporal pattern recognition in network traffic analysis. Convolutional Neural Networks achieved 78.9% accuracy, showing particular strength in processing structured network data and identifying spatial patterns in attack signatures. Random Forest, representing traditional machine learning approaches, achieved 76.8% accuracy, while Support Vector Machines achieved 71.2% accuracy across all test scenarios.

Table 1: Overall Detection Accuracy by Algorithm

Algorithm	Detection Accuracy (%)	True Positives	False Positives	True Negatives	False Negatives
Ensemble Method	87.3	1,746	248	1,752	254
LSTM Network	82.1	1,642	358	1,642	358
CNN	78.9	1,578	422	1,578	422
Random Forest	76.8	1,536	464	1,536	464
Support Vector Machine	71.2	1,424	576	1,424	576

Table 1 demonstrates the comprehensive performance metrics for each algorithm across 4,000 test instances, with 2,000 zero-day attacks and 2,000 benign activities. The ensemble method's superior performance was particularly evident in its balanced approach to minimizing both false positives and false negatives. The table reveals that while LSTM networks achieved high accuracy, they produced a moderate number of false positives that could impact operational effectiveness. CNNs showed consistent performance but required significant computational resources for processing. Random Forest algorithms demonstrated reliability but limited capability for detecting sophisticated attack variants. Support Vector Machines, while computationally efficient, showed the lowest detection rates and highest false negative rates,

indicating limited effectiveness for zero-day detection scenarios.

Table 2: Performance Metrics by Attack Complexity Level

Algorithm	Simple Attacks (%)	Moderate Attacks (%)	Complex Attacks (%)	Average Processing Time (ms)
Ensemble Method	92.1	86.8	83.0	127.3
LSTM Network	88.4	81.2	76.7	89.6
CNN	84.7	77.9	74.1	156.8
Random Forest	82.3	75.8	72.3	23.4
Support Vector Machine	78.1	69.7	65.8	31.2

Table 2 illustrates the relationship between attack complexity and detection performance across all algorithms. The data reveals that all algorithms experienced decreased detection rates as attack complexity increased, with ensemble methods maintaining the most consistent performance across complexity levels. Simple attacks, characterized by straightforward exploitation techniques and minimal evasion attempts, were detected most effectively by all algorithms. Complex attacks, incorporating multiple evasion techniques and sophisticated payload delivery mechanisms, posed the greatest challenge for all detection systems. The processing time data demonstrates the computational trade-offs associated with different algorithms, with Random Forest and SVM offering faster processing at the cost of reduced accuracy, while CNNs required the longest processing times despite achieving moderate detection rates.

Table 3: False Positive Rates by Network Environment

Algorithm	Corporate Network (%)	Academic Network (%)	Cloud Environment (%)	Industrial Control (%)
Ensemble Method	11.2	13.8	12.1	15.4
LSTM Network	16.7	18.9	17.3	21.2
CNN	19.4	22.1	20.7	24.8
Random Forest	21.3	24.6	22.9	27.1
Support Vector Machine	26.8	29.4	27.7	32.3

Table 3 presents false positive rates across different network environments, revealing significant variations based on network characteristics and traffic patterns. Corporate networks demonstrated the lowest false positive rates across all algorithms, likely due to more predictable traffic patterns and standardized user behaviors. Academic networks showed higher false positive rates due to diverse user activities and experimental network usage patterns. Cloud environments presented moderate false positive challenges, with dynamic scaling and varied application behaviors contributing to detection complexity. Industrial control systems exhibited the highest false positive rates across all algorithms, reflecting the specialized nature of industrial protocols and the challenge of distinguishing between normal operational variations and potential threats.

Table 4: Resource Consumption Analysis

Algorithm	CPU Usage (%)	Memory Usage (GB)	Network Bandwidth (Mbps)	Training Time (hours)
Ensemble Method	78.3	12.7	45.2	18.4
LSTM Network	65.1	8.9	32.1	14.7
CNN	89.4	15.3	52.8	22.6
Random Forest	34.7	4.2	18.9	6.3
Support Vector Machine	28.9	3.1	15.4	4.8

Table 4 quantifies the resource requirements for deploying each algorithm in operational environments. The data demonstrates significant differences in computational demands across algorithms, with CNNs requiring the highest resource consumption across most metrics. The ensemble method showed substantial resource requirements but provided the best performance return on computational investment. LSTM networks achieved a favorable balance between resource consumption and detection accuracy. Traditional machine learning approaches (Random Forest and SVM) required significantly fewer resources but delivered correspondingly lower detection performance. The training time data is particularly relevant for organizations requiring frequent model updates, as it directly impacts operational flexibility and response to emerging threats.

Table 5: Temporal Performance Analysis

Algorithm	Week 1-4 (%)	Week 5-8 (%)	Week 9-12 (%)	Week 13-16 (%)	Performance Decay Rate
Ensemble Method	89.7	87.8	86.1	84.6	5.7%
LSTM Network	84.3	82.7	80.9	79.2	6.1%
CNN	81.2	79.4	77.6	75.8	6.6%
Random Forest	78.9	76.8	74.7	72.4	8.2%
Support Vector Machine	73.4	71.2	69.1	67.0	8.7%

Table 5 demonstrates the temporal stability of detection performance over a 16-week operational period, revealing important insights about algorithm sustainability in dynamic environments. All algorithms experienced performance degradation over time, attributed to evolving attack patterns and network environment changes. The ensemble method demonstrated the most stable performance with the lowest decay rate, indicating superior adaptability to environmental changes. Neural network approaches (LSTM and CNN) showed moderate performance decay rates, while traditional machine learning methods experienced more significant degradation. This temporal analysis highlights the importance of regular model retraining and updates for maintaining effective detection capabilities.

Qualitative Analysis

The qualitative analysis of the current study was invaluable in offering key insights on the pragmatic problems and real-life matters pertinent to putting out AI-based zero-day detection systems in situ. By employing structured interviews with fifteen cybersecurity experts working under various organizational conditions, the study found numerous critical themes, which add more information to the quantitative performance scores and contextualize them.

Implementation Complexity and Integration Challenges

Complexity of integration was rated as the most defining challenge to effective use of AI by the participants of the interview. The participants of the Senior Security Analyst struggled to find that current security infrastructure usually does not possess the compatibility and data exchange abilities that will enable AI to be integrated into the system seamlessly. An organization often has a heterogeneous security landscape and can contain legacy systems, those developed by third parties, and proprietary apps that have not supported integration with AI. As a limitation, the technical skills necessary to complete a successful implementation were brought to focus, and the participants stated that the necessary expert knowledge in the field of cybersecurity and machine learning is highly uncommon and costly. Some of the participants pointed out implementation failure cases, when AI systems were implemented without appropriate regard to the current workflows and organizational processes.

Human Factors and Organizational Readiness

It was found that the human factor became one of the key determinants of the effectiveness of AI systems, as the participants particularly spoke about the need to train analysts and introduce the process of managing changes. Security Operations Center managers explained some difficulties in gaining a trust of analysts in AI-generated alerts especially when the rationale behind the detections was not obvious at first glance. The issue of alert fatigue was discovered as one of the chronic problems, and the results presented by the participants showed that the false positive frequency might cause desensitization on the side of analyst and inability to detect real threats. Culture change was necessary to integrate AI insights into their decisions-making, so its implementation process was said to be more difficult than the actual application. The subjects stressed the role of detailed training strategies and incremental enforcement methods in order to facilitate acceptance by the organization.

Operational Sustainability and Maintenance Requirements

The issue that had not been properly covered in the technical literature turned out to be long-term sustainability of operation. Respondents reported persisting maintenance demands such as retraining of the model, performance adjustments, and response to changes in the threat environment. During initial implementation planning, the amount of resource required to stay operational at all times was underestimated. Some participants mentioned that maintaining and monitoring such AI systems also necessitated hiring specialized people, which caused an operational overhead. The problem of ensuring that AI models are updated to address new attacks was identified as one of the issues that will take much organizational commitment to take care of.

Performance Expectations vs. Reality

The results of the interview revealed that there were major gaps in the performance expectations and the reality of operations. On the one hand, although laboratory testing showed good detection rates, the participants reported that in practice, the performance was not always so good since these tests focused on only the environment whereas real world tests are affected by numerous environments not present during laboratory tests. The reality of real network environments containing an assortment of applications, user behaviors, infrastructure variations and more posed difficulty in reproduction of these conditions in a research environment. Participants underscored the need to set realistic expectations on what the performance should be and extensive pilots prior to a full-scale deployment.

Cost-Benefit Considerations

The financial aspect was discussed as a decisive factor in determining adoption of AI by the participants as the costs of ownership sometimes turned out higher than initially predicted. In addition to software licensing and hardware requirements, organizations incurred huge expenditures in training of their personnel,

system integrations and maintenance. Some of them mentioned cases when the implementations of AI were ended because of too expensive operation. The issue with the calculation of the return on investment was how problematic it was to measure prevented attacks or the long life of the benefits of cybersecurity.

Compliance and regulatory Implications

Respondents in regulated industries also presented further issues associated with controlling issues and the aspect of audit. Some issues concerned transparency of decisions and audit trail, and reporting requirements to regulators when using the AI systems in security monitoring. Others found that the regulatory frameworks were not adhering to the AI-based security systems making it uncertain as to the compliance requirements. It was found that documentation requirements of AI-guided security decisions are more complicated when compared to conventional approaches to security.

Discussion

Such in-depth examination, which is reflected in the present study, demonstrates not only the great potential of utilization of artificial intelligence in the detection of zero-day attacks but also the real limitations in its types of application. The quantitative findings show that AI algorithms, especially the ensemble techniques, allow reaching detection rates of above 87%, which is a significant better rate with the traditional signature-based methods. Nevertheless, the qualitative analysis shows that technical measures of performance do not serve as sufficient tools to evaluate actual effectiveness in the real world because organizational factors, implementation challenges, and operational constraints play great roles in determining functionality. The temporal performance study is especially alarming because it revealed the fact that all the algorithms have been facing the drop in their performance over time, with a decay rate varying between 5.7 and 8.7 form 16 weeks, thus, demonstrating the extreme necessity to update the model constantly and regularly.

The identified rates of false positives in all algorithms (11.2 and 32.3 percent, respectively) in all experiments pose an enormous threat to operational practice, which might invalidate the utility of AI-based detection systems. Although ensemble approaches returned the lowest levels of false positive, the best algorithm still emitted a false alarm over ten times as frequently as it flagged a real activity, in the most positive conditions. Such results are synonymous with the data presented in expert interviews, which highlighted alert fatigue as the main issue in the operations by security teams. Another major limitation can be realized in the utilization of the resources of analysis that requires the best algorithms to have high costs in terms of resources and can be way beyond the scope of the most organizations due to the limited IT resources or budget issues.

The result of the study implies that the successful implementation of AI to address zero-day detection necessitates a comprehensive implementation strategy taking into consideration technical performance, organizational readiness, and sustainability aspect. The disparity between the performance in laboratory conditions and effectiveness found in the course of the qualitative analysis shows that the existing approaches to research have not fully grasped the multidimensionality of the real-world setting of operations. The integration issues presented by the interviewees suggest commissioning AI-based products with their compatibility and simplicity of implementation as the main principles, instead of asking questions only about their detection accuracy. Moreover, the expertise requirements of successful implementation indicate that AI vendors and research fraternity ought to focus on more convenient and user-friendly systems that can easily be implemented in an organization with limited specialized skills.

Conclusion

The presented study presents a convincing evidence on the efficiency of artificial intelligence in identifying zero-day attacks, both in the power to perform it and major issues that will have to be

overcome to actually have it work well in operations. The quantitative analysis has proven that ensemble ones were the most accurate to detect the problem with the accuracy 87.3% which significantly prevailed over classical methods and showed the opportunity AI represented to solve the important security gap imposed by the unknowable vulnerabilities. Nevertheless, the research showed also some disturbing limitations such as performance depreciation with time, a high number of false positives that have the potential to overwhelm the security team, and high resource needs that cannot be met by the organization. The temporal or time analysis demonstrating 5.7 % to 8.7 % performance degradation after 16 weeks reveals a high priority of continuously observing the performance and maintaining the models and the dynamics of cyber defense situations.

The incorporation of qualitative information presented by cybersecurity experts has been invaluable as far as quantitative measurement is concerned since it presented information not readily measurable by operational and technical performance factors, showing that performance is not a surety of success in operations. The factors which were also equally significant determinants of effectiveness of AI systems included implementation complexity, organizational readiness, human factors, and cost implications. The disparity betwixt laboratory performance and experimentation performance implies that extant assessment measures are not sufficient surrogates of the plexities of deployment environments, and thus a more encompassing method of assessment to encompass the restrictions of deployment and constrains of organizations is wanted.

Computational demands in the resource consumption analysis, it was seen that the most effective algorithms had large computational requirements, with CNNs needing 89.4 percent CPU and ensembles using 12.7 GB of memory allocation available. Employment of the training times, which are more than 18 hours to incorporate the best algorithms, in addition to the requirements, can be quite a challenge to organizations with poor technical infrastructure. The false positive rates recorded in varying network environments (corporate environment 11.2%, industrial control systems 32.3%) goes to show that the application of AI is highly contextual and dependent on environmental conditions.

The results of the study suggest that although using AI technologies can be significantly effective regarding zero-day attack detection, the efficient application will necessitate paying attention to the choice of algorithms, environmental impact, organizational preparedness, and long-term sustainability conditions. The higher accuracy of the ensemble method indicates that accuracy and reliability of detecting cars using AI might be achieved by combining several different AI strategies, which will lead to trading off the increase in computational complexity and resource usage. The study helps to enrich academic knowledge and deployment tactics of AI-based cyber defense tools.

Recommendations

Institutions planning to use AI-based zero-day detection systems would be advised to follow a phased implementation that would start with thorough pilot testing on controlled conditions followed by the final widespread deployment. In cases where the resources can afford the use of algorithms that run in artificial intelligence (AI), ensemble approaches should be considered first since they proved to be more superior in terms of detection and stability over a certain period of time. Nevertheless, Random Forest implementations are an option that should be considered by organizations with limited levels of resources as a trade-off between the resource requirements and performance. The implementation process must incorporate the heavy investment in employee training and change management procedures that will guarantee organizational toleration and successful outside employing of insights made with AI. Companies also need to start model retraining on a predetermined basis and performance observation procedures to create a solution to the time-based performance loss that is noticed in this research paper. Lastly, cost-benefit analysis

should involve total possession costs such as continuous maintenance needs, training efforts and infrastructure demand as opposed to the short-term deployment of equipment costs maximizing the value of AI-powered threat detection systems in terms of securing a long-term efficiency.

REFERENCES

Adavelli, S. R., et al. (2025). AI and Cybersecurity: Advancements in Threat Detection and Prevention, Academic Guru Publishing House.

Ahmad, R., et al. (2023). "Zero-day attack detection: a systematic literature review." Artificial Intelligence Review **56**(10): 10733-10811.

Ajagbe, S. A., et al. (2023). "Internet of things enabled convolutional neural networks: applications, techniques, challenges, and prospects." IoT-enabled Convolutional Neural Networks: Techniques and Applications: 27-63.

Aslam, M. W., et al. (2025). "ADVANCEMENTS IN MEDICAL IMAGING FROM TRADITIONAL TECHNIQUES TO AI-DRIVEN INNOVATIONS."

Aslam, M. W., et al. (2025). "A COMPREHENSIVE REVIEW OF WEARABLE HEALTH DEVICES: ADVANCES, CHALLENGES, AND FUTURE DIRECTIONS."

Aslam, M. W., et al. (2025). "FINANCIAL FEASIBILITY OF IMPLEMENTING SMART SAFETY TECHNOLOGIES IN ELECTRICAL ENGINEERING PROJECTS: A REVIEW OF CURRENT STATUS AND FUTURE PROSPECTS." Spectrum of Engineering Sciences **3**(3): 557-567.

Aslam, M. W., et al. (2025). "THE ROLE OF ARTIFICIAL INTELLIGENCE IN ELECTRICAL ENGINEERING APPLICATIONS IN SMART GRIDS, POWER SYSTEMS, AND AUTOMATION." Spectrum of Engineering Sciences **3**(3): 540-556.

Dari, S. S., et al. (2023). "Neural Networks and Cyber Resilience: Deep Insights into AI Architectures for Robust Security Framework." Journal of Electrical Systems **19**(3).

Ekong, A. P., et al. (2023). "Securing against zero-day attacks: a machine learning approach for classification and organizations' perception of its impact." Journal of Information Systems and Informatics **5**(3): 1123-1140.

El Husseini, F., et al. (2024). Advanced Machine Learning Approaches for Zero-Day Attack Detection: A Review. 2024 8th Cyber Security in Networking Conference (CSNet), IEEE.

Guo, Y. (2023). "A review of machine learning-based zero-day attack detection: Challenges and future directions." Computer communications **198**: 175-185.

GUPTA, A. B., et al. (2023). "Smart Defense: AI-Powered Adaptive IDs for Real-Time Zero-Day Threat Mitigation."

Hossain, M. T., et al. (2024). "A Review on Attacks against Artificial Intelligence (AI) and Their Defence Image Recognition and Generation Machine Learning, Artificial Intelligence." Control Systems and Optimization Letters **2**(1): 52-59.

Igugu, A. (2024). Evaluating the Effectiveness of AI and Machine Learning Techniques for Zero-Day Attacks Detection in Cloud Environments.

Jimmy, F. (2021). "Emerging threats: The latest cybersecurity risks and the role of artificial intelligence in enhancing cybersecurity defenses." Valley International Journal Digital Library **1**: 564-574.

JothiShri, S., et al. (2024). AI Cyber Security: Enhancing Network Security with Deep Learning for Real-Time Threat Detection and Performance Evaluation. 2024 3rd International Conference for Advancement in Technology (ICONAT), IEEE.

Kavitha, D. and S. Thejas (2024). "Ai enabled threat detection: Leveraging artificial intelligence for advanced security and cyber threat mitigation." Ieee Access.

Mohamed, N. (2025). "Artificial intelligence and machine learning in cybersecurity: a deep dive into state-of-the-art techniques and future paradigms." Knowledge and Information Systems: 1-87.

Nkongolo, M. and M. Tokmak (2023). Zero-day threats detection for critical infrastructures. Annual Conference of South African Institute of Computer Scientists and Information Technologists, Springer.

Okoli, U. I., et al. (2024). "Machine learning in cybersecurity: A review of threat detection and defense mechanisms." World Journal of Advanced Research and Reviews **21**(1): 2286-2295.

Raja, M. S. R. S. (2025). "The Rise of AI-Driven Network Intrusion Detection Systems: Innovations, Challenges, and Future Directions." International Journal of AI, BigData, Computational and Management Studies **1**(1): 1-10.

Salem, A. H., et al. (2024). "Advancing cybersecurity: a comprehensive review of AI-driven detection techniques." Journal of Big Data **11**(1): 105.

Yaseen, A. (2023). "AI-driven threat detection and response: A paradigm shift in cybersecurity." International Journal of Information and Cybersecurity **7**(12): 25-43.

Zhang, Z., et al. (2022). "Artificial intelligence in cyber security: research advances, challenges, and opportunities." Artificial Intelligence Review **55**(2): 1029-1053.

Zhou, K.-Q. (2022). "Zero-day vulnerabilities: Unveiling the threat landscape in network security." Mesopotamian Journal of CyberSecurity **2022**: 57-64.

Zoppi, T., et al. (2021). "Unsupervised algorithms to detect zero-day attacks: Strategy and application." Ieee Access **9**: 90603-90615.