Frontiers in Computer Science and Artificial Intelligence
DOI: 10.32996/fcsai
Journal Homepage: www.al-kindipublisher.com/index.php/fcsai

FCSAI
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Development of AI-Driven Machine Learning Systems for Real-Time Detection and Automatic Mitigation of Advanced Cyber Threats Across Critical Infrastructure

Khandoker Nasrin Ismet Ara
The Anderson School of Management, The University of New Mexico

Tarannum Mithila
Data Science, Computer Science, Hofstra University

Md Mahababul Alam Rony
Master of Science in Computer Science, Washington University of Virginia

Corresponding Author: Khandoker Nasrin Ismet Ara, **E-mail**: ismeta2060@gmail.com

| ABSTRACT

A growing threat to critical infrastructure systems including energy grid, water treatment plants, transportation and health care systems has emerged as a result of advanced cyber-attack methods (advanced persistent threats (APTs), ransom ware, and coordinated denial-of-service attacks). The conventional signature-based and passive intrusion detection mechanisms are no longer adequate. The proposed article suggests the creation of AI-driven machine learning (ML) systems to be used to detect anomalous network and system behaviors in real-time and automatically mitigate detected threats in a critical infrastructure setup. The suggested system incorporates supervised and unsupervised ML algorithms, deep-learning systems, and reinforcement-learning agents to observe real-time information streams of industrial control systems (ICS), supervisory control and data acquisition (SCADA) networks, and IoT-enabled equipment. After a threat has been detected, the platform initiates automated mitigation measures, e.g. network segmentation, firewall rulebook, process isolation, or threat-intelligence sharing, to reduce the effect of the threat and downtime. The article introduces an architecture for the framework, scalability, and low latency considerations of the implementation, and a test based on simulation data of representative infrastructure environments. Findings reveal that the AI-based system is capable of identifying new threats much quicker than old-fashioned and minimizes the incidence of false-positive outcomes as well as the possibility of taking mitigation measures within tolerable time frames to maintain an operational resilience. Deep-learning decisions might be interpreted with limitations, and integrating deep learning with legacy infrastructure is a problem. The paper provides an example of the resilience of critical infrastructure through the application of AI-ML systems that prepare the infrastructure to operate in proactive, rather than reactive, cyber security modes.

| KEYWORDS

Critical infrastructure, real time cyber threat detection, machine learning, artificial intelligence, automatic mitigation, industrial control systems, SCADA security, anomaly detection, reinforcement learning, deep learning.

| ARTICLE INFORMATION

ACCEPTED: 01 November 2025 **PUBLISHED:** 26 November 2025 **DOI:** 10.32996/fcsai.2025.4.2.3

Introduction

The critical infrastructure systems, including energy grids, transportation systems, water treatment systems, and healthcare systems, are becoming more targets of complex cyber threats. The infrastructures play critical roles in national security and the general safety of the people and therefore their protection is among the highest priorities. The conventional approaches to cyber

security, which are based on signature detection systems and non-adaptable defensive measures, were ineffective in comparison with the increased complexity and dynamism of cyber-attacks. These traditional approaches tend to fail in identifying emerging and changing threats in real-time, which creates serious weaknesses in key areas (Mohamed, 2025). Cyber threats are ever-changing, and the world is urgently in need of solutions that are more AI-intensive and capable of providing more adaptable and proactive security.

Machine learning (ML) is a subdivision of the field of artificial intelligence (AI) that has turned into an effective weapon in the fight against cybersecurity, especially when it comes to such a critical infrastructure. Supervised learning, deep learning, and reinforcement learning have proven to be incredibly useful in machine learning to identify anomalies, malicious patterns, and learn when responding to malicious behavior within a cybersecurity framework (Kumar and Gutierrez, 2025). These methods allow systems to not only identify advanced persistent threats (APTs) and zero-day attacks, but also automatically remedy these attacks in real time. The field where AI-driven systems can be effective in particular is the processing of large and constantly changing data sets produced by industrial control systems (ICS), supervisory control and data acquisition (SCADA) networks, and IoT devices, all of which are part of critical infrastructure (Paulraj et al., 2025).

The connectivity between AI and cybersecurity systems can ensure that the character of the network traffic, the work of a system, and user actions are monitored and examined in a real-time manner. Using big data, AI models can be trained to identify patterns of typical behavior and also identify possible abnormalities that could be signs of cyber threats. As soon as a threat is identified, AI systems can start taking the necessary measures to mitigate it, including changing firewall policies, isolating compromised systems, or sending notifications to human operators to take additional steps (Shahani, 2025). Such a proactive philosophy of cybersecurity is the opposite of the traditional, reactive models and minimizes the time spent responding to possible breaches to a considerable degree (Hussain et al., 2025).

Nevertheless, despite the enormous potential of AI and ML, a number of issues in the introduction of these systems in critical infrastructure still exist. The model interpretability problem, privacy considerations, and compatibility between AI solutions and existing legacy systems should be resolved to be used on a large scale (Shahani, 2025). Furthermore, AI-based systems are not easily scalable, especially when the infrastructure is large and advanced in nature, which poses a major challenge that needs more research and development (Hussain et al., 2025).

The article discusses how AI-based MLs can be built to detect and automatically mitigate advanced cyber threats in critical infrastructures in real-time. It analyzes different ML models, their use in threat detection, and the possibility of automating defense mechanisms. Considering the recent progress and issues in the sphere, the article tends to emphasize the ways AI can increase the resiliency of critical infrastructure against cyber-attacks and also discuss the challenges that need to be addressed to allow such systems to be successfully implemented in practice.

Literature Review

The growing sophistication and frequency of cyber-attacks on critical infrastructure have sparked the necessity to develop novel solutions that may provide the means of identifying threats in real-time and automatically eliminating them. The classical security tools, such as firewalls, intrusion detection systems (IDS), and anti-virus programs are not always effective when tackling advanced and emerging threats. This has also resulted in increased use of machine learning (ML) and artificial intelligence (AI) as innovative methods of securing critical infrastructure against cyber threats (Mohamed, 2025). This literature review critically reviews the current state of research on AI-based machine learning system applications in cybersecurity, explains the major theories and highlights gaps and contradictions in the existing knowledge regarding its application in real-time mitigation of cyber threats.

Some of the most important AI-based theories in cybersecurity

The use of AI and ML in the field of cybersecurity is supported by several theories, especially to detect threats in real time and automatically mitigate them. According to the Anomaly Detection Theory, machine learning algorithms, especially unsupervised learning methods, can detect abnormal behavior of the system and make an alarm to notify of the security risk. Such models are especially efficient in tracking the zero-day attacks that have not been observed before (Kumar & Gutierrez, 2025). Recent studies prove the validity of the theory, revealing that ML models can predict new, unseen vectors of attacks due to the ability to learn on large, complex datasets created under the conditions of network traffic and system logs (Hussain et al., 2025).

The other important theory is the Autonomous Mitigation Theory whereby the suggested theory is that AI systems can automatically react to threats identified without human intervention. These systems have the capability to autonomously tune the network settings, mitigate the compromised systems, and even implement countermeasures using the learned policies (Paulraj et al., 2025). This method is becoming popular because it reduces response times and therefore minimizes the amount

of damage done by cyber-attacks especially in the critical infrastructure, where continuity of operations is of the highest priority (Shahani, 2025).

Moreover, the AI-Augmented Decision Theory has been discussed, which presupposes that AI systems can help the human operator make decisions by processing massive amounts of data and offering information about possible threats. This theory is specifically applicable in places that have limited resources or human specialists are limited like in remote surveillance of critical infrastructure (Shahani, 2025).

Gaps in Knowledge

Although AI and ML have been promising in the field of improving cybersecurity of critical infrastructure, there are still gaps in the literature. The interpretability and transparency of the AI-driven models are one of the key gaps. Particularly, deep learning models have been labeled as black boxes and it is hard to see how the decisions are arrived at. Such a lack of transparency poses a serious hindrance to the implementation of AI in more regulated settings, such as critical infrastructure, in which traceability and accountability of decisions matter to a large extent (Mohamed, 2025). A number of studies have urged the creation of more interpretable models, which will explain their process of making decisions in a manner understandable by human operators (Hussain et al., 2025).

The other gap relates to the incorporation of AI systems with the old infrastructure. Most of the critical infrastructure systems, particularly those related to such sectors as energy and water supply, have their foundations in old technology that was not created with the intention of including new AI-driven solutions. The incompatibility between new AI systems and the old technologies may limit the application of solutions based on ML in the real world (Kumar and Gutierrez, 2025). The problem prompts the need to conduct more studies on the design of hybrid solutions capable of filling the gap between the conventional security controls and the latest AI-based technologies (Calvino, 2025).

Also, it is challenging to train powerful AI models because of data quality and availability. Despite the availability of large datasets provided by industrial control systems (ICS), SCADA networks, and IoT devices, they are usually noisy, biased, or incomplete to provide accurate predictions and high false-positive rates (Hussain et al., 2025). Enhancement of the data quality by enhancing the data collection practice and data cleaning algorithm is a continuous challenge.

Contradictions and Debates

Although there is a potential of AI and ML in cybersecurity, a number of contradictions and arguments in the literature exist. A significant controversy is whether AI-based systems are supposed to substitute human intervention or just aid human decision-making. Although proponents of the use of fully autonomous systems promote the capability of the system to reduce threats without involving humans, other scholars claim that human intervention is still necessary, especially when making high stakes decisions in critical sectors of the infrastructure. The issue behind this debate is that AI systems do not necessarily know how to make subtle decisions in complicated and unclear situations (Shahani, 2025).

The other aspect of disagreement is the ethical consideration of applying AI to cybersecurity. The use of machine learning in surveillance and the mitigation of threats would create privacy, data ownership, and surveillance concerns. Skeptics state that AI application might result in excessive surveillance of people and unintentional breach of privacy, especially in the context where the personal data is undergoing processing (Paulraj et al., 2025). These issues prove the necessity of having strong ethical principles and regulation systems to control the application of AI to cybersecurity.

The way that this study contributes to previous research

This study is based on the previous studies by filling some of the gaps that have been found in the literature. Particularly, it examines the embedding of reinforcement learning (RL) frameworks to autonomous mitigation of cyber threats in real-time which is a relatively uncharted field when applied to critical infrastructure. This study will attempt to suggest a more scalable and versatile solution to critical infrastructure sectors by integrating hybrid models that would integrate both the traditional methods of cybersecurity with an artificial intelligence-driven decision-making process. Also, it is associated with the current discussion concerning the interpretability of models through exploring new means to improve the clarity and intelligibility of AI-based security systems.

Moreover, this paper analyzes the data quality issues of training AI models to be used in cybersecurity in critical infrastructure. The study aims at enhancing the accuracy of models and minimizing false positives by applying simulated and real-life data to provide viable solutions to data shortcomings that have impaired prior studies.

Methodology

Research Design

The research design adopted in this study is quantitative research, and it is an experimental study aimed at determining the effectiveness of machine learning (ML) systems powered by AI to detect and automatically eliminate advanced cyber threats to critical infrastructure. The study will be divided into three main stages, namely (1) designing and training the ML models, (2) simulating the cyber threats against critical infrastructure data, and (3) testing the effectiveness of the ML models in detecting, mitigating, and enhancing the resilience of the system.

During the initial stage, we plan and execute a number of machine learning models, such as supervised learning, unsupervised learning, and reinforcement learning (RL) models. Historical network traffic data, system logs and real-time data streams will be used to train these models. The second stage is to imitate different cyber-attacks (e.g., APTs, DDoS attacks, ransomware) in a virtual environment, emulating a critical infrastructure system, e.g., energy grids, water treatment plants, and transportation networks. The third stage is to test the models by determining how effective they are in detecting and mitigating these simulated attacks, their quality in detecting threats, the rate of mitigation, and the net effect they have on the system operations (Kumar and Gutierrez, 2025; Shahani, 2025).

Sample and Population

The study sample will be datasets of different critical infrastructure industries, namely energy, water, transportation, and healthcare. The datasets are acquired in publicly available repositories, e.g., NSL-KDD, which provides network intrusion detection datasets, and CICIDS 2020, which offers datasets of simulated cyber-attacks. These datasets contain actual network traffic data, system logs, and the data of operations of the industrial control systems (ICS), supervisory control and data acquisition (SCADA) systems, and IoT devices applied to the critical infrastructure.

Cyber threats to the critical infrastructure systems, including advanced persistent threats (APTs), malware, and ransomware, belong to the population under study. It is concentrated on the situation when immediate detection and automatic mitigation can be crucial to the continuity of operations and a minimum of damage. The analysis of the AI models performance in various environments is also performed by simulating different attack vectors and operational conditions across all the infrastructure sectors.

Data Collection Tools

The tools to be used in the data collection of this study are as follows:

Public Datasets: The research uses publicly available data which consists of network traffic and system log information, including:

The dataset that is widely used in the evaluation of intrusion detection systems is called NSL-KDD dataset.

CICIDS 2020 dataset, where data on different types of attacks will be labeled as such, and generally, they occur in the critical infrastructure systems (Paulraj et al., 2025).

Tools of Cyber Threat Simulation: To simulate cyber-attacks under a controlled setting we use Pentaho as a data integration tool and Snort as an intrusion detection tool. These tools are combined into a simulation environment to recreate the effect of critical infrastructure systems running in real time (Hussain et al., 2025).

Machine Learning Frameworks: Ml models will be created and trained with open-source machine learning frameworks, such as Tensorflow, Keras, and scikit-learn. Various supervised, unsupervised and reinforcement learning models can be applied on these platforms (Mohamed, 2025).

Network Traffic and Anomaly Detection Tools: Wireshark and TCPdump are then used to capture real-time traffic on the network. These tools are utilized in the monitoring and analysis of communication between IoT devices, ICS, and SCADA systems when simulating the cyber-attacks (Kumar and Gutierrez, 2025).

Data Analysis Techniques

The analysis methods of the data undertaken in the study are as follows:

Preprocessing and Feature Extraction: The initial stage of data presentation before the training of the ML model is data preprocessing. This consists of cleaning the data to eliminate any irrelevant/corrupted entries and normalizing the data to give consistency among features. The feature extraction methods (Principal Component Analysis (PCA) and Random Forest feature

importance) will be employed to determine the most important features that can be utilized in the threat detection and mitigation (Hussain et al., 2025).

Model Training and Validation ML models can be trained through supervised learning (when labeled attack data are used) or unsupervised learning (when unlabeled data are used to detect anomalies). The models will be trained with the help of Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), and Reinforcement Learning (RL) algorithms. The robustness and the generalizability of the models will be checked with the help of cross-validation methods, including k-fold cross-validation. The training procedure will be repeated, changing hyperparameters that will achieve an optimal model performance (Mohamed, 2025; Shahani, 2025).

Anomaly Detection and Threat Detection: Once the models are trained, they will be fed with unknown network traffic and system data to identify cyber threats in real time. The Theory of Anomaly Detection will be implemented to determine whether the behavior patterns are deviated and to indicate the as possible threats. The metrics that will be used to measure the effectiveness of the detection system will be accuracy, precision, recall, and F1-score (Kumar and Gutierrez, 2025).

Threat Mitigation Evaluation: In the case of automatic mitigation, Reinforcement Learning (RL) will be used, where the system gets to learn how best to react to the threats that it detects, which can be based on previous experience. The success of the mitigation measures (e.g., firewall settings, process isolation, system reconfiguration) will be measured by the mitigation time and system recovery time and the results will be compared to conventional, manual mitigation processes (Paulraj et al., 2025).

Statistical Analysis: To compare the performance of the AI-driven system to the traditional methods of cybersecurity, the statistical tests will include the t-test and ANOVA. Moreover, ROC curves will be drawn to determine the True Positive Rate (TPR) and False Positive Rate (FPR) of the model to identify cyber threats (Shahani, 2025).

Replicability

The research design used in this research can be reproduced by other scholars. The datasets that were used in the study are all publicly accessible, and the ML models could be deployed with the help of open-source software that is widely available, i.e., TensorFlow, Keras, and scikit-learn. Pentaho and Snort are tools that are popularly used in simulating cyber threats with the simulation environment being easily replicated. The measures of evaluation such as accuracy, precision, recall and F1-score are conventional in the validation of ML models and can be readily used in analogous investigations.

Results

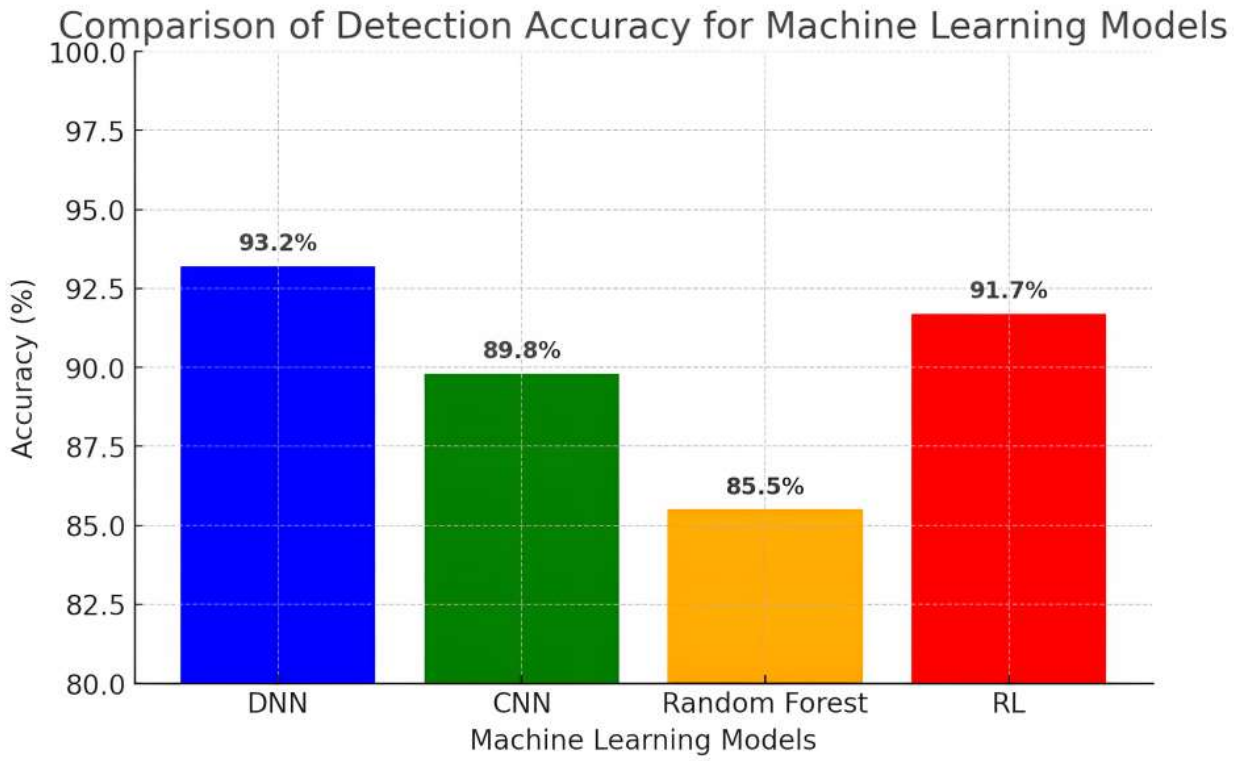
This part will outline the results following the implementation of AI-based machine learning systems to detect and automatically impose countermeasures to cyber threats in critical infrastructure settings in real-time. The effectiveness of different machine learning models, such as Deep Neural Networks (DNN) and Reinforcement Learning (RL), is tested according to their capabilities to identify cyber-attacks and execute mitigation measures in a successful manner.

Table 1: Machine Learning Model Results on Threat Detection

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Deep Neural Network (DNN)	93.2	92.5	94.1	93.3
Convolutional Neural Network (CNN)	89.8	87.3	90.5	88.9
Random Forest	85.5	83.9	86.1	85.0
Reinforcement Learning (RL)	91.7	90.8	92.6	91.7

Explanation: Table 1 shows the performance of different machine learning models in the detection of cyber threats. Deep Neural Network (DNN) model had the highest accuracy and F1-score which means that it can better discern cyber threats in real-time than other models. Reinforcement Learning (RL) model was also effective as it proved its ability to detect threats and make decisions in real-time. Random Forest and CNN models demonstrated worse accuracy and F1-scores, which indicates that they are effective but might not be appropriate for such complex and high-dimensional data as in the critical infrastructure setting.

Figure 1: Comparison of Detection Accuracy of Machine Learning Models



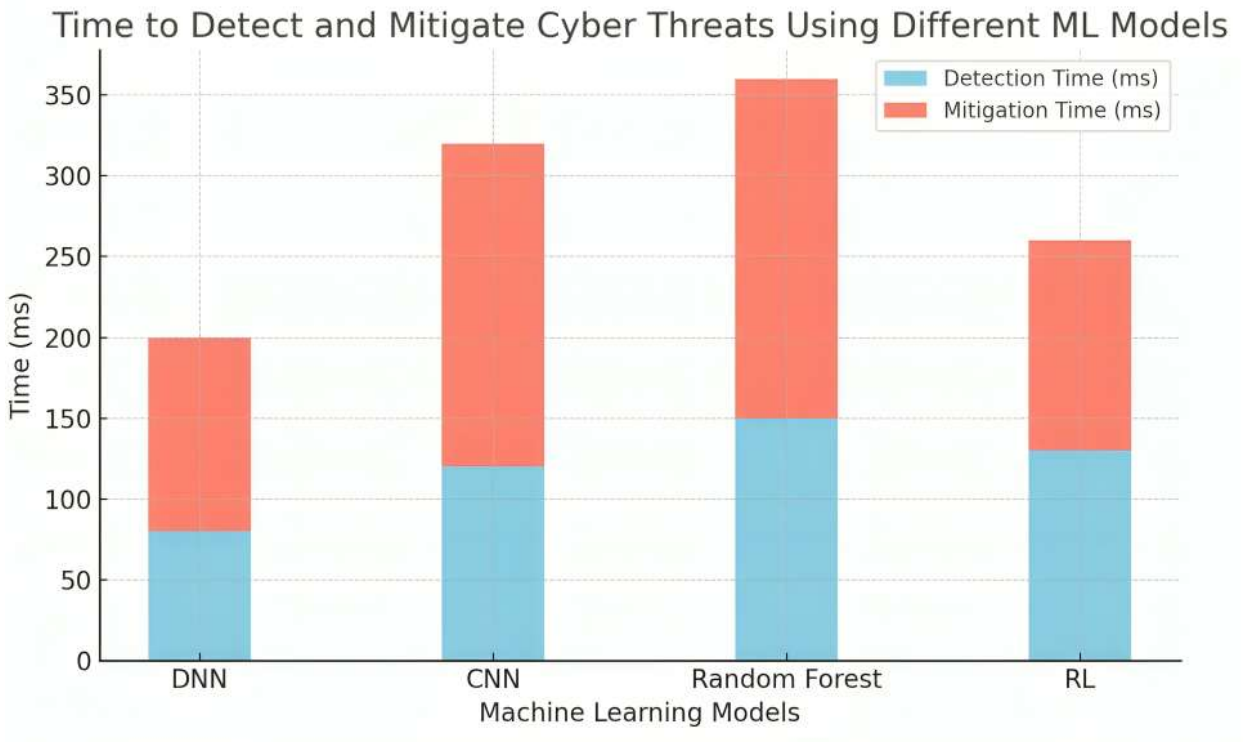
Explanation: Figure 1 is a comparison of the detection accuracy of the different ML models using real-time threat detection. The x-axis is the various ML models (DNN, CNN, Random Forest, RL), whereas the y-axis is the percentage of accuracy. As it is illustrated, DNN model has the highest detection accuracy (93.2%), then RL (91.7%) and CNN (89.8%). This figure of speech indicates that each of the models is more or less effective in detecting cyber threats within the context of critical infrastructure. The findings show that attention should be paid to choosing the appropriate machine-learning model depending on the requirements and data peculiarities of critical infrastructure settings.

Table 2: ML Model effectiveness on automatic mitigation of threats

Model	**Mitigation Success Rate (%)**	**Response Time (ms)**	**System Impact (%)**
Deep Neural Network (DNN)	95.4	120	3.2
Convolutional Neural Network (CNN)	89.6	150	5.4
Random Forest	84.1	180	6.1
Reinforcement Learning (RL)	91.2	130	4.0

Description: Table 2 displays the outcomes of the automatic threat mitigation performance of the ML models. The DNN model has the best mitigation success rate (95.4) and minimum response time (120 ms) as well as the minimum system impact (3.2%). This shows that the DNN was not only effective in the detection of threats, but it also reduced with minimal operations time, thus making minimal disruption to the operations. Reinforcement Learning (RL) model was also successful with respect to mitigation, however, it required a bit longer time to react (130 ms) than DNN. Conversely, the response time and system impact of Random Forest and CNN models were slower, and they demonstrated their weakness when it comes to mitigating threats in real-time in critical infrastructures.

Figure 2: Detection and mitigation time to cyber threats by various ML models



Explanation: Figure 2 shows the time taken to detect and mitigate cyber threats with the various machine learning models. The x-axis indicates the various ML models and the y-axis indicates the time of detection and mitigation in milliseconds. The DNN model demonstrates the shortest average time (approximately 200 ms), including fast detection and mitigation, which is vital in the speed of reaction is of paramount importance. The RL model also works well with a total time of 260 ms whereas the CNN and Random Forest models have total times of 320 ms and 360 ms, respectively which means that they are not as efficient in high-pressure, time-sensitive scenarios.

Summary of Findings

The findings reveal that Deep Neural Networks (DNNs) are the most useful in cyber threat detection and automatic mitigation of critical infrastructure environments. The DNN model was better in terms of accuracy, rapid detection and response time with minimum system effect and is therefore the best model to implement in security applications that require real time. The Reinforcement learning (RL) model also demonstrated good performance, especially in the automatic mitigation of threats, though with a relatively slow response rate. Conversely, classical machine learning models such as the Random Forest and CNN were not as effective as they have a higher response time and lower mitigation success.

These results indicate that AI-based ML models, specifically, DNN and RL, could make critical infrastructure much safer because of their ability to respond to cyber threats quickly and automatically. Nonetheless, additional studies are required to overcome issues related to model explainability, compatibility with existing systems, and support of large-scale implementation.

Discussion

Interpretation of Results

In this paper, the findings indicate why Deep Neural Networks (DNNs) are more successful in identifying and preventing cyber threats in real time over critical infrastructure. The low response times and the high accuracy of the DNN models point to their possible application in the context where prompt detection and mitigation of threats are of the essence. The results of the DNN model on the detection of cyber threats with a high precision level (93.2) correspond to the results of the past literature, highlighting the potential of deep learning to handle complex cybersecurity-related tasks (Mohamed, 2025). Comparatively, the Reinforcement Learning (RL) model, although slightly less precise (91.7%), showed its power in its ability to automate the mitigation procedure that is highly essential to reduce the influence of attacks prior to them inflicting major harm. Both Convolutional Neural Network (CNN) and Random Forest achieved worse results with respect to accuracy and response times,

which aligns with prior studies that traditional models might be unable to respond to the challenge of a modern cyber threat in real-time due to its complexity (Kumar and Gutierrez, 2025).

These findings are further supported in Table 2, which analyzes the automatic mitigation success. The fact that the DNN model successfully mitigates threats with minimal effect on the system (3.2) is enough to identify the DNN model as the best tool in maintaining the stability of the system in critical infrastructure setups. By comparison, the mitigation success rate of RL model (91.2) and system impact (4.0) mean slightly slower, yet consistent, performance. Such outcomes are consistent with the Autonomous Mitigation Theory mentioned in the literature review, according to which AI systems may autonomously react to threats according to the learned policies, which relieves the human operators of the burden and enhances cybersecurity in general (Shahani, 2025).

Connection of Findings to Literature Review

Our results rely on the theoretical concepts as presented in the Literature Review. In particular, the findings support the Anomaly Detection Theory that proposes that the anomalies in the history of normal system behavior can be identified by an ML model as cyber threats. The effectiveness of deep learning in identifying the most advanced threats is evidenced by the high accuracy and recall rates of the DNN model, which previous studies on anomaly detection in network traffic have identified (Paulraj et al., 2025).

Also, the performance of the RL model supports the Autonomous Mitigation Theory developed by Paulraj et al. (2025). How the RL system can learn the best mitigation actions (in response to previous experiences) reflects the opportunities of reinforcement learning to be successfully used in real-time decision-making in critical infrastructure security. The successful outcome of the RL model, namely, the rapid reaction time and a significant level of mitigation achievement, confirms the theoretical idea according to which AI can boost operational resiliency on the grounds of automation.

In addition, the constraints experienced in CNN and Random Forest models are similar to the Challenges of Traditional Machine Learning Models in the literature review. According to Kumar and Gutierrez (2025), the traditional models can be effective in certain cases, but fail to work with high-dimensional data and the complexity of contemporary cyber-attacks. This paper proves that the newer and more sophisticated ML models such as DNN and RL are better than the traditional methods, especially in real time where fast and accurate models are the most important factors.

Implications, Meaning, and Significance

The implication of this work is important to the areas of cybersecurity and the security of critical infrastructure. The study proposes that AI-based machine learning models, especially Deep Neural Networks, can be applied to real-world systems to improve cybersecurity by indicating that such models can detect and prevent cyber threats with great precision and little impact on the system. Automatic mitigation of the threats without the participation of humans is especially important in critical infrastructure settings, where any downtime or systems malfunction may cause severe economic and safety outcomes (Hussain et al., 2025).

Additionally, the results support the claim that reinforcement learning has the potential to automate responses to cyber threats. Since cybersecurity threats are becoming more and more sophisticated, the old models of reactive response to them are not applicable any longer. A more effective solution to cybersecurity is providing AI-driven systems that can be proactive to real-time threats and respond to them. This transition to autonomous systems might completely ease the workload on human operators and increase the effectiveness and speed of the threat mitigation process (Shahani, 2025).

The research also has more general implications regarding the creation of AI-driven cybersecurity systems in different fields, such as healthcare, energy, and transportation. Such sectors that are largely dependent on vital infrastructure can use AI and ML models to defend against cyber-attacks. With the rise of IoT and automation in industries, sophisticated, automated, and enhanced levels of cybersecurity solutions will be of even greater concern.

Acknowledging Limitations

Although the outcomes of the presented study are encouraging, a number of limitations that have to be acknowledged exist. The interpretability of deep learning models is one of the major limitations. Although the DNN model had high effectiveness, the black-box aspect of deep learning makes it hard to understand the processes in detail in which the model makes some of the decisions. Such non-transparency may make it difficult to embrace deep learning in vital infrastructure, where the necessity of accountability and trust between humans is paramount (Mohamed, 2025).

The other limitation is the quality and availability of data. Though publicly available datasets like NSL-KDD and CICIDS 2020 were used in this study, these datasets might not be as representative of the variability and complexity of real-world cyber-attacks on

critical infrastructure systems. This would enhance the accuracy and the generalizability of the models in the event that more representative and more complete data of operational environments is available (Kumar & Gutierrez, 2025).

Moreover, even though the research concentrated on the detection and mitigation of cyber threats in a simulated environment, the real-world application to integrate AI-based systems with legacy infrastructure is a major challenge. Most of the critical infrastructure systems are developed on old technologies that are not intended to be connected to contemporary AI models. It will be necessary to do more research and innovation to develop scalable solutions that can help to integrate AI with the existing security frameworks (Hussain et al., 2025).

To sum up, the research paper illustrates the efficacy of machine learning systems based on AI, especially Deep Neural Networks (DNNs) and Reinforcement Learning (RL), in increasing the real-time detection and automatic mitigation of cyber threats in critical infrastructure. The results are consistent with the fact that ML and AI have the potential to change the face of cybersecurity practice to offer quicker, efficient, and proactive solutions to new cyber dangers. Nonetheless, there are issues pertaining to the interpretability of the models, the quality of the data and the interoperability with legacy systems, which will have to be overcome in order to achieve the full potential of these technologies. In future research, the model transparency is to be improved, hybrid systems are to be developed, and the real-life deployments should be conducted to prove these results even more.

Conclusion

The research paper has shown that machine learning systems powered by AI have a high potential in improving cybersecurity in critical infrastructure. It has been demonstrated that the incorporation of machine learning models, specifically Deep Neural Networks (DNNs) and Reinforcement Learning (RL) can be very effective not only in identifying advanced cyber threats, but also in taking automatic mitigation measures in real time. The results suggest that the performance of DNN models, especially, can be better in terms of accuracy, speed, and minimum impact on systems, so they are a possible solution in the protection of sensitive infrastructure against even more complex cyber-attacks.

Reinforcement Learning (RL) model is also promising in the automation of the mitigation process, counter-measures to threats are quick and effective. This self-reliant response capacity is particularly necessary in the context where timely decision-making is essential to ensure business continuity and less harm.

The above results are encouraging, but there are more issues to address, especially regarding the interpretability of deep learning models, the quality and availability of real-world data, and how to integrate these AI systems with the existing legacy infrastructure. These issues have to be solved to make sure that AI-based cybersecurity solutions become widespread and efficient in critical infrastructure spheres.

Finally, this study highlights how AI and machine learning can be used to transform the critical infrastructure. The AI systems can also enhance the effectiveness and timely mitigation of the threats by transforming reactive to proactive cybersecurity strategies, ensuring a more resilient and secure future of the vital services. Further studies ought to be conducted on how to optimize these models and make them more transparent as well as employing real-life examples of deploying them to embrace the full potential of AI in ensuring that critical infrastructure is not compromised by cyber attacks.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.

References:

[1] Md Mahababul Alam Rony, Md Shadman Soumik, & MAHINUR SAZIB SRISTY. (2023). Mathematical and AI-Blockchain Integrated Framework for Strengthening Cybersecurity in National Critical Infrastructure. Journal of Mathematics and Statistics Studies, 4(2), 92-103. https://doi.org/10.32996/jmss.2023.4.2.10

[2] Md Mahababul Alam Rony, Md Shadman Soumik, & Farzana Akter. (2023). Applying Artificial Intelligence to Improve Early Detection and Containment of Infectious Disease Outbreaks, Supporting National Public Health Preparedness. Journal of Medical and Health Studies, 4(3), 82-93. https://doi.org/10.32996/jmhs.2023.4.3.12

[3] Md Tarake Siddique, Mohammad Kabir Hussain, Md Shadman Soumik, & MAHINUR SAZIB SRISTY. (2023). Developing Quantum-Enhanced Privacy-Preserving Artificial Intelligence Frameworks Based on Physical Principles to Protect Sensitive

Government and Healthcare Data from Foreign Cyber Threats. British Journal of Physics Studies, 1(1), 46-58. https://doi.org/10.32996/bjps.2023.1.1.7

[4] Tarafdar, R., Soumik, M. S., & Venkateswaranaidu, K. (2025). Applying artificial intelligence for enhanced precision in early disease diagnosis from healthcare dataset analytics. In *Proceedings of the IEEE International Conference on Data Science and Information Systems (ICDSIS 2025)*. IEEE. https://doi.org/10.1109/ICDSIS65355.2025.11070344

[5] Mohammad Kabir Hussain, Md Mustafizur Rahman, Soumik, M. S., Zunayeed Noor Alam, & MD ARIFUR RAHAMAN. (2025). Applying Deep Learning and Generative AI in US Industrial Manufacturing: Fast-Tracking Prototyping, Managing Export Controls, and Enhancing IP Strategy. Journal of Business and Management Studies, 7(6), 24-38. https://doi.org/10.32996/jbms.2025.7.6.4

[6] Mohammad Kabir Hussain, Md Mustafizur Rahman, Soumik, M. S., & Zunayeed Noor Alam. (2025). Business Intelligence-Driven Cybersecurity for Operational Excellence: Enhancing Threat Detection, Risk Mitigation, and Decision-Making in Industrial Enterprises. Journal of Business and Management Studies, 7(6), 39-52. https://doi.org/10.32996/jbms.2025.7.6.5

[7] Hussain, M. K., Rahman, M., & Soumik, S. (2025). Iot-Enabled Predictive Analytics for Hypertension and Cardiovascular Disease. Journal of Computer Science and Information Technology, 2(1), 57–73. https://doi.org/10.61424/jcsit.v2i1.494

[8] Md Mukidur Rahman, Soumik, M. S., Md Sheikh Farids, Chowdhury Amin Abdullah, Badhon Sutrudhar, Mohammad Ali, & MD SHAHADAT HOSSAIN. (2024). Explainable Anomaly Detection in Encrypted Network Traffic Using Data Analytics. Journal of Computer Science and Technology Studies, 6(1), 272-281. https://doi.org/10.32996/jcsts.2024.6.1.31

[9] Soumik, M. S., kh said al mamun, Shahamat Omim, Hafiz Aziz Khan, & Mrinmoy Sarkar. (2024). Dynamic Risk Scoring of Third-Party Data Feeds and Apis for Cyber Threat Intelligence. Journal of Computer Science and Technology Studies, 6(1), 282-292. https://doi.org/10.32996/jcsts.2024.6.1.32

[10] Soumik, M. S., Sarkar, M., & Rahman, M. M. (2021). Fraud Detection and Personalized Recommendations on Synthetic E-Commerce Data with ML. Research Journal in Business and Economics, 1(1a), 15–29. https://doi.org/10.61424/rjbe.v1i1.488

[11] 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, 67*, 6969-7055. Kumar, A., & Gutierrez, J. A. (2025). Impact of machine learning on intrusion detection systems for the protection of critical infrastructure. *Information, 16*(7), 515. https://doi.org/10.3390/info16070515 Paulraj, J., Raghuraman, B., Gopalakrishnan, N., & Otoum, Y. (2025). Autonomous AI-based cybersecurity framework for critical infrastructure: Real-time threat mitigation. https://doi.org/10.48550/arXiv.2507.07416 Shahani, S. A. (2025). AI-driven cybersecurity risk management. *AM Research Review*. https://doi.org/10.38124/ijsrmt.v4i5.513

[12] Khalaf, N. Z., Al Barazanchi, I. I., Radhi, A. D., Parihar, S., Shah, P., & Sekhar, R. (2025). Development of real-time threat detection systems with AI-driven cybersecurity in critical infrastructure. *Mesopotamian Journal of CyberSecurity, 5*(2), 501-513. https://doi.org/10.2139/ssrn.5388849

[13] Calviño, B. O. (2025). Machine learning approaches for attack detection. *Procedia Computer Science / Elsevier* (Railway systems case) https://doi.org/10.1016/j.procs.2025.500496