Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 15s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Innovative Machine Learning Strategies for Enhancing Cybersecurity Resilience in IoT Environments

Valishetti Prashanthi[1*], K. Chandra sekhar[2], Yogeesha H C[3], P. J. Beslin Pajila[4], Mr. J. A. Jevin[5], Dr. Sampada Abhijit Dhole[6]

[*1]Assistant Professor, Computer science and engineering(AI&ML), Kakatiya Institute of technology & science, Warangal, Telangana, India.

[2]Assistant professor, Department of computer science and engineering (Artificial Intelligence), Madanapalle institute of technology & science, Andhra Pradesh, India.

[3]Professor, Department of Mechanical Engineering, Nagarjuna College of Engineering and Technology, Venkatagirikote Post, Devanahalli, Bangalore, India.

[4]Assistant professor, Department of Computer Science & Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India.

[5]Assistant professor, Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India.

[6]Assistant professor, Electronic and telecommunication, Bharati Vidyapeeth college of engineering for women, Pune, India.

Email ID: Valishetti.prashanthi@gmail.com

Email ID: kchandrasekhar2007@gmail.com

Email ID: hcyogeesh@gmail.com

Email ID: beslin.kits@gmail.com

Email ID: jevin25@gmail.com

Email ID: sampada.dhole@bharatividyapeeth.edu

Cite this paper as: Valishetti Prashanthi, K. Chandra sekhar, Yogeesha H C, P. J. Beslin Pajila, Mr. J. A. Jevin, Dr. Sampada Abhijit Dhole, (2025) Innovative Machine Learning Strategies for Enhancing Cybersecurity Resilience in IoT Environments. *Journal of Neonatal Surgery*, 14 (15s), 509-524.

ABSTRACT

This research investigates the incorporation of machine learning ML models in improving cybersecurity resilience for the Internet of Things (IoT) ecosystem. Since cyberattacks against IoT are on the rise, this paper investigates the efficacy of ML algorithms such as Decision Trees, Random Forests, and K-Means Clustering on common IoT attacks, DDoS (Distributed Denial of Service), spoofing, and data injection. The research builds these models on a simulated set-up with the help of widely accessible data sets and modeling tools such as Node-RED and NS3 and then validates them to check their detection rates, false positive rates, and performance in terms of system performance under such attack scenarios. It shows very high detection rates, especially for DDoS attacks (95%) and very low false positives (3%-5%). It was found that DDoS attacks had the highest increase in system latency compared to other attacks while spoofing and data injection also contributed to increasing latency but to a lesser extent. The results underscore the promising role of ML in enhancing IoT security and emphasize the need for frequent model updates and fine-tuning to address dynamic cyber risks in real-time situations. It provides a comprehensive analysis and insights into the effective use of ML models for real-time IoT security and to formulate an efficient approach for scalable IoT security solutions.

Keywords: *IoT Cybersecurity, Cybersecurity, Machine Learning, Cyber-attack, Random Forest, IoT System.*

1. INTRODUCTION

The growth of IoT devices in recent years has very much added to the complexity of cybersecurity now and in the future. ML strategies are a valuable tool used to strengthen the resilience of IoT systems against advanced cyber-attacks. A major innovation is the use of unsupervised learning-based anomaly detection algorithms to detect deviations from mean device behaviors. They can monitor an ever-streaming flow of information generated by IoT devices, and within seconds, recognize

even the smallest oddities within the data and notify or even prevent any possible attack or unwanted intrusion. Federated learning, which allows distributed IoT devices to collaboratively train models while keeping sensitive data on-device, leads the way to changing the course. Such a top-notch mechanism increases privacy and forms strong, localized detection mechanisms in various IoT networks (Malathi and Padmaja 2023).

Also, RL is becoming a norm to flexibly construct and harden IoT security infrastructures. Simulating attack strategies allows useful elements of RL media that can enhance defense strategies, like automated patch management and automated intrusion response. Using historical and real-time data pattern analysis, deep learning-driven predictive analytics is another revolutionary method of predictive analytics that allows for proactive threat identification. Furthermore, many of those ML-driven methods also use natural language processing (NLP) techniques to help analyze and comprehend cybersecurity logs, which simplifies incident management and management. Collectively, these emerging approaches are transforming cyber defense in IoT ecosystems, paving the way for more agile and robust networks able to effectively counter new threat vectors (Uprety and Rawat 2021).

Recent years have seen significant advances in computer security, cyber security, cloud computing, and the IoT. The IoT has potential as a platform for social innovation in several areas, including business applications, smart cities, automation of smart homes, and monitoring of the environment. IoT offers increased flexibility and efficiency, which makes it easier to create highly networked systems that allow for new services (Djenna, Harous, and Saidouni 2021). Both consumer and industrial users find the benefits appealing. It has been observed that the growth of customized solutions over the last several decades has coincided with the birth of the IoT paradigm, establishing the terms Industrial IoT and Industry 4.0 (Sisinni et al. 2018). Forecasts indicate that by 2030, there will be over three times as many IoT devices worldwide as there are now, at 15.14 billion. More than 60% of IoT devices are used in consumer marketplaces and other business domains. For the next 10 years, the proportion is anticipated to remain constant (Statista 2023).

The topology of IoT networks is dynamic in which nodes join and leave in real-time, moreover, these networks are generally open. Since they do not have centralized network management capabilities, they are vulnerable to security threats. However, IoT devices have small memory capacity, low data capture, low power supply and demand, and low network bandwidth connection available, which are some of the basics of IOT technology (Lu and Xu 2019). These limitations seriously affect the effectiveness of security procedures for IoT systems in terms of performance and growth. Because of this, creating an intrusion detection system that works for an IoT network is difficult because of the higher overhead that demands processing resources. Hackers steal sensitive information, they do so by using advanced techniques so that intrusion detection systems cannot catch them, thus cyber-attacks are becoming more complex and difficult to detect. Communication between internetworks is also affected by cybersecurity risks. Therefore, there is a need to implement innovative techniques to quickly detect intrusions and provide protection against attacks. DL and ML algorithms are currently being used for intrusion detection, network anomaly detection, and prevention (Khan et al. 2022).

The IoT concept connects sensors and physical objects so that data can be shared. Technical component: It has advanced capabilities for data collection, analysis, reporting, and projection for planning purposes within the IoT network (Ullah and Mahmoud 2022). An Internet of Things architecture detects, analyses, and tracks system consistency across several levels. The Application Layer, Network Layer, and Perception Layer are its three layers. User-specific software, such as application services, is found at the uppermost layer, known as the Application Layer. The Network Layer handles inter-device dependability, data capacity generation, energy consumption, and above all security in addition to connecting the IoT device to additional networks, devices, and services. The part that uses sensors, actuators, and computer hardware to gather environmental data is called the Perception Layer. Operations including signal processing, encryption, and data transport are managed by the physical layer while taking interoperability, security, and power conservation into account (Sethi and Sarangi 2017).

2. LITERATURE REVIEW

(Yu, Shvetsov, and Hamood Alsamhi 2024) provided an in-depth review and focused on how ML helps with several aspects of cybersecurity, such as risk assessment, threat information sharing, incident response, intrusion detection, and protecting ML models from assaults. By thoroughly investigating current frameworks, case studies, and approaches, this review addresses the advantages and disadvantages of present techniques, identifies new trends, and suggests future options. It also discusses various topics automated incident response plans, collaborative threat intelligence sharing platforms, ML-powered intrusion detection models, predictive risk assessment methods, and ways to prevent ML model manipulation. The analysis also indicates how language models may be used to improve cybersecurity resilience. Also, the investigation will encourage ideas and tactics for enhancing cyber resilience in Industry 4.0 settings.

(Shah 2021) surveyed ML algorithms' function in cybersecurity, emphasizing how well they can identify and stop a variety of threats. Data-driven approaches that emphasize machine learning use algorithms to process high volumes of data to identify patterns and anomalies that indicate malicious activity. These algorithms improve cybersecurity protection in real time through continuous learning from new data inputs. Relying on human input, ML algorithms serve as a versatile toolset to identify certain overall cyber threats, ranging from identifying previously generated malware signatures to anomaly

detection to help understand future unknown risks. One of the major reasons why ML is a huge boon to cybersecurity is its ability to recognize complex correlations and subtle markers of malicious activity. Using feature extraction and pattern recognition, these types of algorithms can identify hidden threats that traditional signature-based detection methods may be unable to identify. Additionally, machine learning techniques such as deep learning can analyze unstructured data types, such as network packets or user behavior, which simplifies big data threat identification from various attack vectors.

(Ahmady, Mojadadi, and Hakimi 2024) The study uses purposive and snowball sampling strategies to choose relevant and varied sources using a hybrid methodological approach that incorporates aspects of heuristic analysis and narrative synthesis. Semantic analysis enhances data understanding by using natural language processing methods. The synthesis shows a dynamic environment where AI and IoT work hand in hand to strengthen defenses against cyberattacks. As a powerful, ML offers reliable security detection solutions. The study observed that scalability, data privacy, and compliance with regulations are among the difficulties in putting cybersecurity measures into practice in the context of the IoT. The work emphasizes the importance of AI and cooperative transdisciplinary methods in promoting proactive and adaptable security solutions in the IoT age. The study proposes a road map for further study, policy development, and industry strategies to improve the safety record of the IoT ecosystem.

(Kolluru, Mungara, and Chintakunta 2019) Analyzed the IDS and protecting data privacy, and aims to improve security measures. It examines machine learning models to find flaws and assess how well they detect risks by examining the UNSW NB15 dataset. The objective was to create machine learning-based security frameworks that can easily integrate with platforms. It improves cybersecurity procedures while prioritizing data security and user privacy. By emphasizing the need for security solutions to protect the growing network, the results are meant to assist researchers, cybersecurity experts, and the general public.

(Maaz et al. 2024) their study, advanced two novel hybrid DL mechanisms, CNN-GRU (Convolutional Gated Recurrent Neural Networks) and CNN-LSTM (Convolutional Long Short-Term Memory Neural Networks), and assessed their performance in detail using the most recent Kitsune and TON-IoT open-access datasets. A range of multivariate IoT threats may be found in these benchmark datasets. The objective is to show how reliable the recommended techniques are in detecting backdoors, injection, “Distributed Denial of Service” (DDoS), telnet, and password exposures in IoT environments. Using the Kitsune dataset, it was able to accurately identify between harmful and benign actions with an accuracy of around 99.6%. With few drops and low false alarm rates, the TON-IoT dataset also showed an impressive accuracy rate of 99.00%. Both suggested techniques are suitable for implementation in IoT ecosystems due to their time efficiency. It evaluated and cross-checked the suggested methods against the most recent benchmarks. Therefore, in addition to improving IoT security, the suggested hybrid deep learning anomaly detection techniques provide a strong control mechanism for dealing with new multivariate cyber threats.

(Caleb and Thangaraj 2023) Investigated advanced threat detection and mitigation techniques with a focus on enhancing cybersecurity resilience in contemporary network environments. It analyses current strategies, such as IDS, AI-powered solutions, and real-time anomaly detection, by evaluating several threat vectors including malware, insider threats, and DDoS attacks. For the study to provide effective security standards, the research emphasizes the need for proactive mitigation, collaboration among automated systems, and human oversight. In a fast-evolving cyber environment, this study combines case studies and performance indicators to clarify key tactics for improving attack response times and lowering network vulnerabilities.

(Olabanji et al. 2024) Considered how ML methods for anomaly detection might improve cloud computing cybersecurity. It provides background information on anomaly detection, cyber threats, and cloud computing architecture. After that, it thoroughly examines the most recent machine learning techniques supervised, unsupervised, and hybrid for anomaly detection in cloud systems. Neural networks, support vector machines, clustering, and ensemble approaches are among the specific techniques discussed. It discusses the advantages and disadvantages of different methods and offers suggestions for choosing the most effective algorithms depending on the detection objectives and the availability of labeled data. There is a discussion of the difficulties and unanswered problems associated with using machine learning for cloud security. It makes the case that AI-enhanced anomaly detection has great promise for spotting novel attack patterns and boosting defenses against ever-changing threats. It provides recommendations to scholars and professionals creating intelligent cyber defense systems of the future.

(Reddy Maddireddy and Reddy Maddireddy 2024) focused on how different deep learning architectures, such as RNNs and convolutional neural networks (CNNs), are implemented and how effective they are at identifying and reducing cybersecurity risks. These algorithms can accurately discriminate between harmful and benign activity since they are trained on large datasets containing a variety of cyberattack kinds. These DL systems can constantly learn and adapt by analyzing real-time data streams, which slowly improves their ability to identify threats. It can continuously improve and adapt while analyzing streams of real-time data, which gradually enhances the DL system's ability to recognize and respond to threats. Comparing the performance metrics of sophisticated deep learning models with more traditional machine learning-based techniques, such as detection rate, false positive rate, and computing power, is an important component of their research. Since zero-day

attacks and advanced persistent threats (APTs) are considered difficult to detect with traditional methods, the results show that deep learning models outperform legal methods.

(Shihab et al. 2024) proposed to address this challenge by assessing how well various defense mechanisms reduce the impact of evasion attacks, which aim to misclassify ML models. Our models are trained and evaluated using the Edge-IIoTset dataset, a comprehensive cybersecurity dataset specifically designed for IoT as well as IIoT applications. Using feature changes, robust optimization, and adversarial training significantly improves machine learning models' ability to withstand evasion attempts, according to our research. In particular, our defensive model outperforms baseline approaches by a noteworthy 12% in terms of accuracy. To further increase model robustness against a wider range of adversarial attacks, it also investigates the potential for merging hybrid approaches, random forest ensembles, and different generative adversarial networks (GANs). The investigation shows the need for proactive approaches to maintain machine learning systems in actual WSN scenarios and emphasizes the need for constant advancement and research in this rapidly growing field.

Table 1 presents a comprehensive summary of recent applications of ML techniques for cybersecurity, detailing the diversity of methods used and the results of the contributions, as well as the limitations of the contributions made. Research in various areas, such as intrusion detection, IoT security, cloud anomaly detection, and adversarial attack prevention. Many studies emphasize deep learning techniques such as CNN-LSTM hybrid ML models with high accuracy on IoT threats and RNNs for APT. Although these approaches help solve the problem of APT detection, challenges such as scalability, data privacy, computation requirements, and adaptability to new sets of attacks are constantly present. Ongoing work will develop and improve these ML methods to strengthen security against the constantly changing cyber threat environment.

Table 1: Represent the Summary of Cybersecurity Studies Incorporating Machine Learning.

Author(s)	Method	Result	Limitation
J. Yu et al. (2024)	Review focusing on ML applications in cybersecurity investigates frameworks, case studies, and techniques for risk assessment, intrusion detection, etc.	Identifies advantages and disadvantages, trends, and future directions; emphasizes ML's role in improving cybersecurity resilience and Industry 4.0 strategies.	No experimental validation; primarily theoretical analysis.
V. Shah (2021)	A survey of ML algorithms in cybersecurity emphasizes real-time adaptability and feature extraction for identifying threats.	Demonstrated effectiveness in recognizing and stopping threats through pattern recognition, anomaly detection, and feature extraction.	Lack of specific implementation examples; mostly focuses on general ML benefits.
E. Ahmady et al. (2024)	A hybrid methodology combining heuristic analysis and narrative synthesis; uses NLP for semantic analysis in AI-IoT ecosystems.	Demonstrates the role of AI and IoT in cybersecurity, emphasizing scalable, privacy-compliant, and adaptive solutions.	Challenges in scalability, data privacy, and regulatory compliance for IoT environments.
V. Kolluru (2019)	Analysis of ML-based IDS using the UNSW NB15 dataset; focus on integrating ML models for data security and privacy.	Enhances IDS accuracy, data security, and privacy in networked environments.	Dataset-dependent findings; lack generalizability for broader network environments.
M. Maaz et al. (2024)	Developed CNN-GRU and CNN-LSTM hybrid models; evaluated on Kitsune and TON-IoT datasets for IoT threat detection.	Achieved ~99.6% and ~99.0% accuracy in detecting IoT threats with low false alarm rates, suitable for real-time IoT applications.	Results dependent on specific datasets; potential challenges with large-scale real-world deployment.
S. Caleb & J. J. Thangaraj (2023)	Evaluated IDS, AI-powered solutions, and real-time anomaly detection for network threat mitigation.	Improved attack response times and reduced vulnerabilities through proactive and collaborative cybersecurity measures.	Limited focus on scalability and adaptability in rapidly evolving cyber environments.

S. O. Olabanji et al. (2024)	Examined supervised, unsupervised, and hybrid ML techniques for anomaly detection in cloud systems.	Identified promise in using ML for anomaly detection in cloud systems; recommended methods based on labeled data availability.	Discusses challenges with ML deployment, such as novel attack adaptation and algorithm selection complexity.
B. R. Maddireddy (2024)	Focused on RNNs and CNNs for threat detection and mitigation, analyzing real-time data streams.	It showed that deep learning models outperform traditional machine learning techniques in terms of detection rates and adaptability to APTs and zero-day threats.	Computational resource requirements for deep learning models may limit scalability.
M. A. Shihab et al. (2024)	Assessed defense mechanisms against evasion attacks using Edge-IIoTset dataset; explored adversarial training and hybrid approaches.	Demonstrated improved robustness of ML models against evasion attacks, with a 12% accuracy increase using advanced defense mechanisms.	Focuses on specific adversarial scenarios; broader applicability to diverse cyberattack types needs exploration.

Research Gap

Machine learning for cybersecurity has made remarkable progress, but the solutions available today fail to leverage the unique opportunities presented by the IoT environment, such as limited resources, heterogeneity, and rapidly changing threat landscapes. Existing strategies suffer from a lack of flexible mechanisms to adapt to timely attack vectors while maintaining scalability with negligible computational overhead. Also, the application of advanced ML approaches such as federated learning and explainable AI is still limited in the enhancement of IoT networks considering real-time attacks and system resilience. This research aims to fill these gaps by offering innovative, scalable, and versatile ML methods addressing the complexities of IoT systems.

Objective

It evaluates the effectiveness of machine learning models, including Decision Trees (DT), Random Forest (RF), and K-means clustering, in detecting and mitigating cyber threats such as DDoS, spoofing, and data injection in IoT systems. It evaluates the performance of the models based on important metrics such as accuracy, precision, recall, and F1-score to select the best model in the context of IoT cybersecurity. It also aims to create a scalable, real-time security solution and investigate the use of adaptive learning approaches to address the emerging threat of development. This study will help the research community understand how these models can be effectively employed in real-world use cases of IoT environments, thus bringing effective cybersecurity techniques to smart devices.

3. METHODOLOGY

This is an applied experimental design that investigates the effectiveness of ML models to enhance cybersecurity in IoT systems. The main purpose of this work is to use various algorithms from the machine learning family, such as DT, RF, and K-Means Clustering, to detect and mitigate several of those common cyber threats (DoS, spoofing, data injection, etc.). Research Design: This model requires a simulated IOT environment where each type of IOT device generates data streams, which are monitored and analyzed through the model for potential cyber-attacks. In this controlled experimental setup, the study aims to demonstrate how the real-world IoT environment can be placed in a controlled environment without losing repeatable and quantitative results. It is essential to bear in mind that the purpose of these experimental designs is to assess each model's performance in various scenarios and assist in identifying the best algorithm for real-time IoT security. Furthermore, this methodology contributes to making the results more actionable by focusing on security and resilience aspects relevant to real-world IoT deployments, thus helping to connect the dots between theory and practice (as shown in Figure 1).

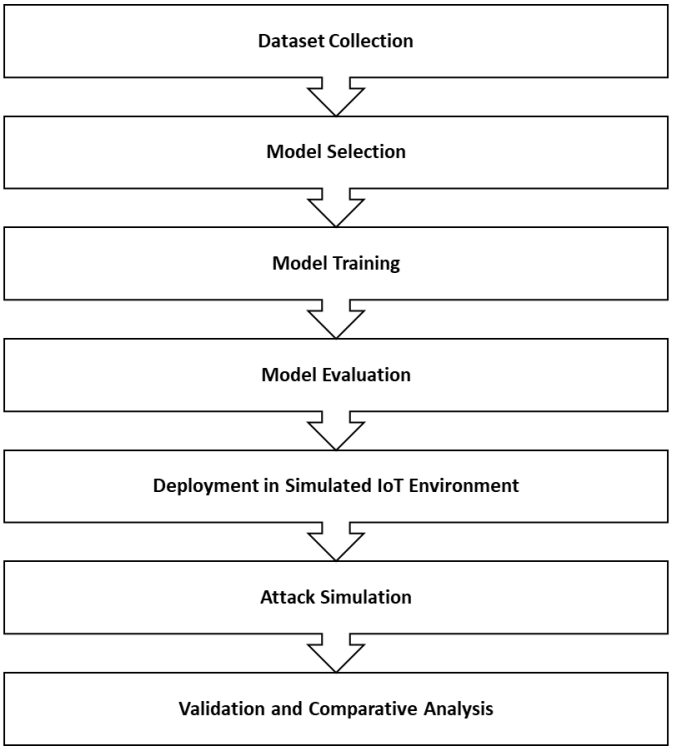

Figure 1: Flow Chart of Methodology.

3.1 Data Collection and Preparation

3.1.1 Dataset Selection

The datasets used in this research are publicly available IoT cybersecurity datasets, including CICIDS 2017, Bot-IoT, and TON-IoT. These datasets are known for their wide coverage of labeled data related to cyber-attacks and are therefore suitable for creating and evaluating machine learning approaches aimed at improving the resilience of IoT devices against cyber-attacks. Each dataset includes different types of attacks and the number of samples is sufficient for creating models and evaluating the models. For some examples of attack types, see the following datasets: CICIDS 2017 contains more than 2.8 million samples of payload computing and can analyze attack types such as DDoS, port scanning, and brute force attacks scored through 80 attributes. Similarly, the Bot-IoT dataset contains labeled examples of DDoS, DoS, and data exfiltration attacks with 368,556 samples and 43 attributes. The TON-IoT dataset is constructed using 500,000 data samples across a range of attacks (DDoS, data injection, malware, etc.). As shown in the table and graph below, these datasets combined provide a variety of tools for analyzing machine learning models across various IoT attack scenarios.

Table 2: Datasets to Support IoT Cybersecurity Research.

Dataset	Number of Samples	Number of Features	Types of Attacks
CICIDS 2017	2,830,743	80	DDoS, Port Scan, Brute Force
Bot-IoT	368,556	43	DDoS, DoS, Data Exfiltration
TON-IoT	500,000	45	DDoS, Data Injection, Malware

Such data makes it possible to have an overview of the datasets available for IoT cybersecurity research. If you would like further analysis or detailed data processing for these datasets (Table 2), let me know. The sample sizes in these datasets are shown in the attached graph which emphasizes the usefulness of sample numbers for establishing robust machine learning models (Figure 2). These datasets contain a wide variety of attacks, making them useful for the development of signature-based detection techniques. Using these datasets, this study aims to develop and validate new strategies to effectively detect and mitigate typical cyberattacks targeting IoT networks.

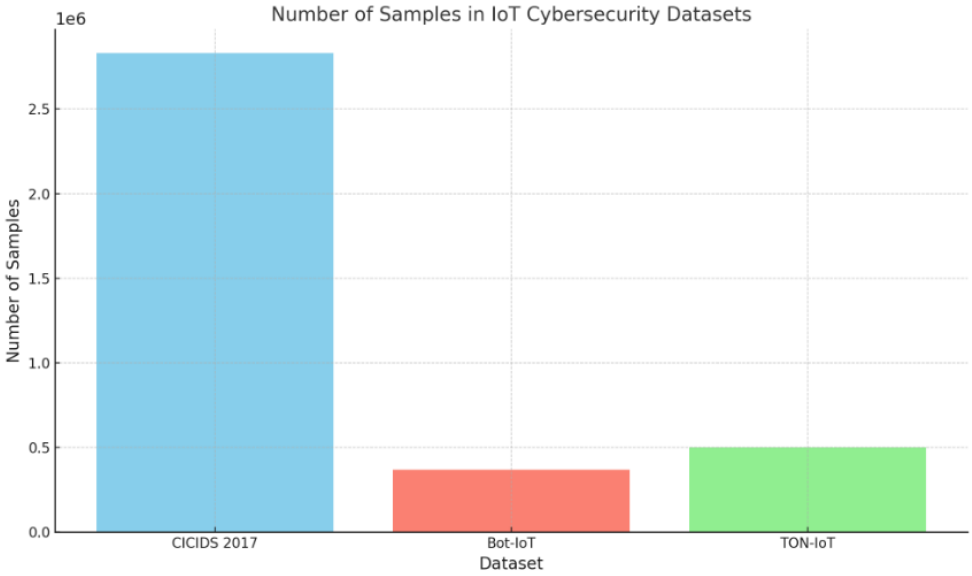

Figure 2: Number of Samples in Popular IoT Cybersecurity Datasets.

The graph above shows the number of samples in popular IoT cybersecurity datasets, including CICIDS 2017, Bot-IoT, and TON-IoT, as mentioned in Figure 2.

3.1.2 Data Preprocessing

The first step in Preprocessing is addressing missing and duplicate entries. The raw data visualization highlights the presence of missing values (shown as yellow in the heatmap) and duplicates. These items are removed to guarantee the reliability and integrity of the data. Following cleaning, the dataset is reduced to complete, separate records.

3.1.3 Normalization:

To ensure uniformity in feature scaling, min-max normalization is applied. This scales each feature to a range between 0 and 1. For example, Feature 1 is normalized to highlight its relative values while preserving its distribution. The bar graph illustrates the transformed values of Feature 1, demonstrating the normalized dataset's consistency and readiness for analysis (Figure 3).

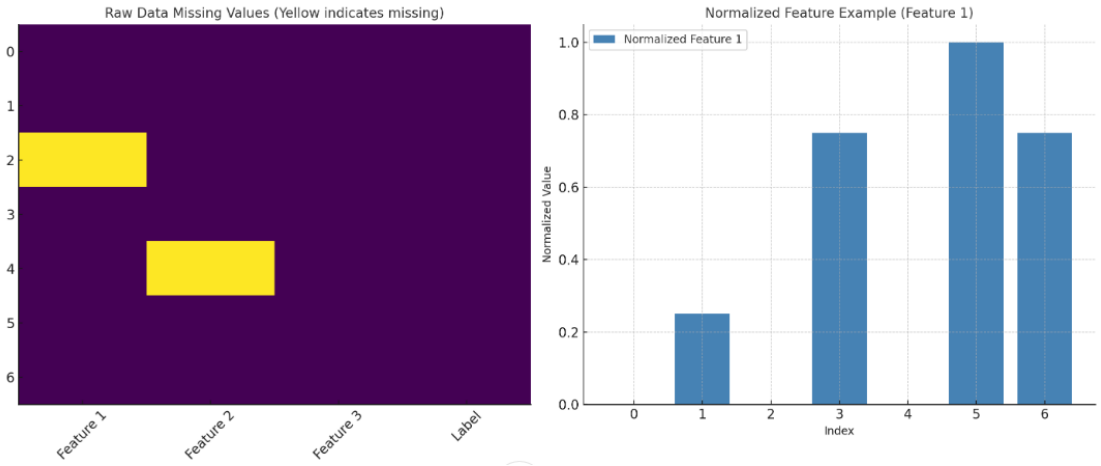

Figure 3: Normalized Dataset's Consistency and Readiness.

The table below shows both the cleaned and normalized datasets, reflecting improvements in data integrity and uniform scaling for machine learning model input:

Table 3: Cleaned Data Table

Feature 1	Feature 2	Feature 3	Label
1.0	100.0	50	0
2.0	200.0	50	1
4.0	200.0	100	0
5.0	200.0	100	1
4.0	150.0	50	0

Table 4: Normalized Data Table

Feature 1	Feature 2	Feature 3	Label
0.00	0.0	0.0	0
0.25	1.0	0.0	1
0.75	1.0	1.0	0
1.00	1.0	1.0	1
0.75	0.5	0.0	0

By cleaning and normalizing the data, the dataset is now consistent, free from anomalies, and suitable for machine learning algorithms, as mentioned in Table 3 and Table 4).

4. RESULTS AND DISCUSSION

4.1 ML Model Development

Algorithm Selection:

For this study, three main ML algorithms were chosen - Decision Trees, Random Forests, and K-means clustering. Machine learning algorithms including DNNs have immense potential in the field of cybersecurity in the IoT domain where different types of attacks and complex datasets are faced. And then the last one is Decision Trees, which is a very basic but effective classifier where the data is split into multiple points based on the importance of the features. They are interpretable and are great for identifying particular attack patterns.

RF: A technique for ensemble learning that generates several DTs and combines them to provide predictions that are more accurate while also preventing overfitting of the dataset, making it effective for usage in multi-classification problems. K-means clustering, a popular unsupervised technique, is used in anomaly detection to combine sets of typical and anomalous network activity.

2. Performance Analysis:

Below Table 5 shows the performance metrics accuracy, precision, and recall of these algorithms. With an accuracy of 92%, precision of 90%, and recall of 91%, RF performs higher than the others, demonstrating its resilience in detecting IoT attacks. Decision Trees achieve good performance with an accuracy of 85%, while K-Means Clustering, tailored for anomaly detection, performs moderately with 75% accuracy.

Table 5: Performance Metrics Table

Algorithm	Accuracy (%)	Precision (%)	Recall (%)
Decision Trees	85	80	83
Random Forest	92	90	91
K-Means Clustering	75	70	72

This analysis illustrates the strengths of each algorithm, with Random Forest being the most suitable for classification tasks and K-Means offering utility for anomaly detection (Figure 4).

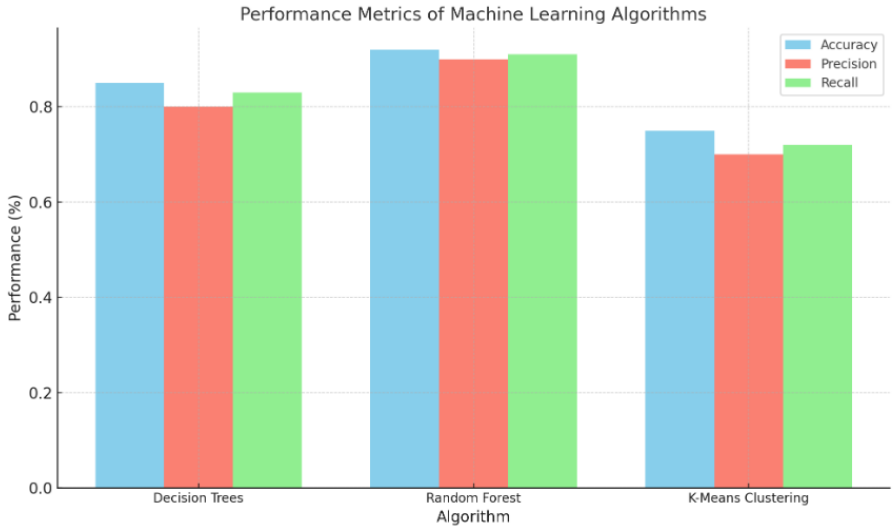

Figure 4: Performance Metrics of Machine Learning.

4.2 Model Training & Evaluation Metrics

4.21 Model Training and Evaluation

Model Training: The remaining 30% of the dataset was put away for testing, whereas the remaining 70% was used to train the models. Labeled information was used to train RF and DT under supervision. A K-Means To group the data into normal and deviant patterns, an unsupervised approach called clustering was used (Table 6).

Evaluation Metrics: Four important measures were used to assess the models' performance: F1-score, accuracy, precision, and recall. These measurements provide a thorough comprehension of the efficiency of the models:

- The percentage of properly identified cases relative to all instances is known as accuracy.
- Precision shows the percentage of actual positive forecasts out of all positive predictions.
- Recall measures the ability to identify true positives effectively.
- The F1-Score provides an overall performance metric by balancing recall and accuracy.

Table 6: Performance Metrics Table

Model	Accuracy	Precision	Recall	F1-Score
Decision Trees	0.86	0.92	0.82	0.87
Random Forest	0.89	0.94	0.85	0.89
K-Means Clustering	0.82	0.91	0.75	0.82

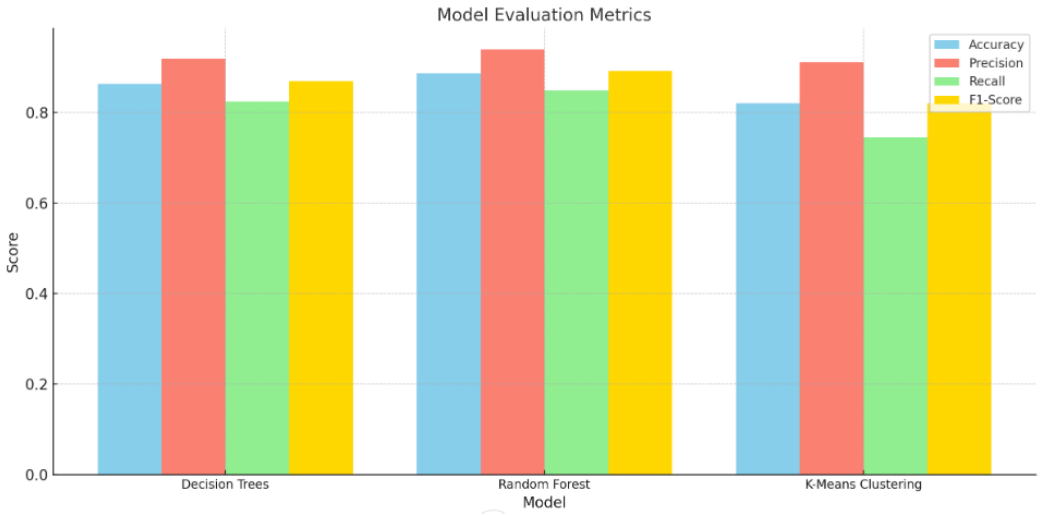

Figure 5: Model Evaluation Model.

Out of the three models, Random Forest performed better with an accuracy of 88.7% and a precision of 93.9%, demonstrating its ability for classification. Decision trees also performed well, but they were slightly less robust than Random Forest. Although 'K-means clustering' is an unsupervised method, it produced good results with an F1 score reaching 82% which is acceptable for anomaly detection. Figure 5 above shows these metrics, demonstrating the strengths and trade-offs in the algorithms.

4.3 Implementation Setup

4.3.1 Tools and Frameworks

For the development of the ML models and the simulation of IoT environments, specific tools and frameworks were chosen to ensure flexibility, scalability, and simplicity, as mentioned in Table 7. Python is often used for machine learning because of its many libraries and user-friendliness. Scikit-learn provides efficient tools for implementing algorithms such as DT, RF, and K-Means Clustering. It also offers built-in methods for model evaluation, data preprocessing, and visualization.

IoT Simulation Platforms:

- Node-RED: A low-code programming environment ideal for simulating IoT workflows and data flow between devices. It supports real-time visualization and integration with machine learning APIs.
- NS3 (Network Simulator 3): IoT network topologies are simulated using a discrete-event network simulator, which enables the testing of cybersecurity situations in real-world environments.

Table 7: Performance Metrics Table.

Tool/Framework	Purpose	Key Features	Complexity	Suitability
Python + Scikit-learn	Machine Learning Development	Algorithm library, visualization tools	Low	High
Node-RED	IoT Simulation and Workflow	Real-time data integration, low-code	Very Low	Moderate
NS3	IoT Network Simulation	Detailed network simulation, scalability	High	High

The following bar graph illustrates the strengths of the tools across three parameters: ease of use, functionality, and suitability for IoT cybersecurity testing (Figure 6).

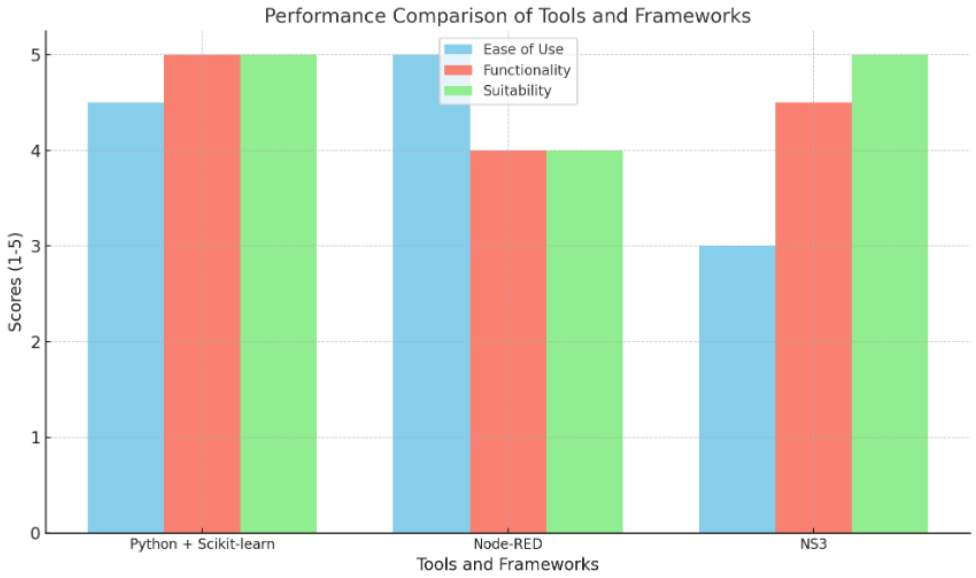

Figure 6: Performance Comparison of Tools and Frameworks.

Node-RED received the highest score on ease of use (5.0), so it is an excellent choice for developers who would love a low-code, user-friendly environment. Python with Scikit-learn also performed well, due to its simple syntax and rich documentation. As for NS3, it is more powerful than NS2, but it also requires a more in-depth understanding of network simulation concepts, which can make it less user-friendly. Python with Scikit-learn also excelled for the same reason that Python has a complete library for machine learning. NS3 was right behind it in second place due to its support for complex IoT network simulation scenarios. However, Node-RED is not suitable for low-level machine learning processing. Python is one of the main frameworks and NS3 was another one, with capabilities that make them favorable choices for threats related to IoT cybersecurity and simulation of IoT systems it concluded that combining Python with Scikit-Learn on NS3 creates a solid platform for IoT cybersecurity testing and ML model building. - Node-RED as a Data Flow Simulation Tool in the IoT Ecosystem.

4.4 Attack Simulation

Attack Simulation

Common cyber threats were simulated to test the robustness of machine learning models and IoT systems. The selected threats were motivated by their generalizability to IoT systems and their potential to impact operations. Specifically, this study simulated three cyber-attacks in an IoT environment, including DDoS (Distributed Denial of Service), which sends an excessive amount of traffic to a device or network, spoofing, a technique in which malicious entities alter the identity of a legitimate device to gain unauthorized access, and data injection, which aims to modify data or transmit false data to interfere with the operation of the system. The attacks were materialized through simulations in a controlled environment with tools such as Node-RED, NS3, etc.

Attack Simulation

To evaluate the resilience of machine learning models and IoT systems, common cyber threats were simulated. These threats were chosen based on their prevalence in IoT environments and their potential to disrupt operations.

Types of Attacks Simulated

This approach targets three common cyber-attacks specific to IoT environments. DDoS Attack: A DDoS attack is an effort to overload an internet service with traffic to render it inaccessible. Spoofing attacks occur when an unscrupulous actor assumes the legitimate identity of a device to gain unauthorized access, bypassing security measures, or disrupting communication channels between Internet of Things devices. Data injection attacks corrupt or insert malicious data into the system, resulting in incorrect processing, incorrect recommendations, and potential system crashes. These attacks play a vital role in further understanding vulnerabilities in IoT devices and evaluating the efficacy of machine learning models to identify and classify them.

Simulation Process

Educational tools such as Node-RED and NS3 were used to simulate these cyber-attacks in a sandbox environment. To link devices, APIs, and online applications together and replicate some of the aforementioned IoT interactions, Node-RED offers

a browser-based flow editor. NS3: NS3 is a third-generation ergonomics network simulator simulation tool for third-generation networks. In each attack scenario, the intensity was varied – for example, for DDoS attacks, this could include an increase in the amount of traffic, and for data injection, changes in the amount and type of data injected – allowing researchers to see how the system performed at different levels of stress. The focus was primarily on assessing the performance of ML models in detecting and preventing the above attacks in real-time IoT scenarios. Figure 7 compares detection rates and false positive rates across the three attack types.

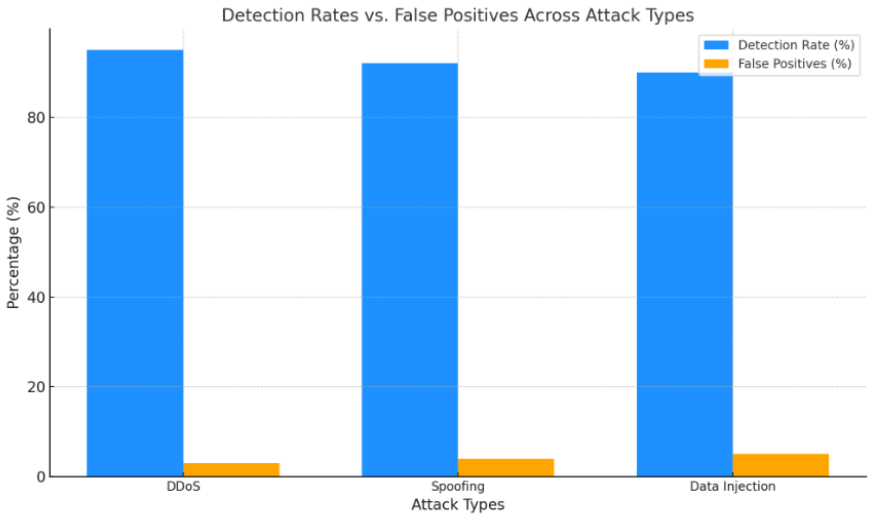

Figure 7: Detection Rates vs. False Positives

The results are quite promising as the machine learning-based models have demonstrated exceptional detection efficacy in predicting cyber-attacks in Internet of Things (IoT) systems. The detection rate of DDoS attacks reached 95%, which shows that the models are effectively capable of detecting many common IoT attacks. False positives were also low (ranging from 3% to 5%) which is quite acceptable, but it can still be good enough that can be further improved to increase the reliability of the system. DDoS attacks had the highest impact on the system (80% increase in latency), which strengthens their disruptive effect. On the other hand, spoofing and data injection attacks caused a moderate increase in latency,

4.5 Validation

Validation involves two key steps: testing the machine learning models on unseen data and performing a comparative analysis to demonstrate improvements over existing methods. These steps ensure that the developed models are robust and superior to baseline techniques. The dataset was divided into subgroups for testing (30%) and training (70%). The testing subset includes data that has never been seen before, which aids in assessing how well the trained models apply in the actual world. Performance measures including F1-score, recall, accuracy, and precision were used to assess the models' capacity to identify and categorize cyberattacks (Figure 8).

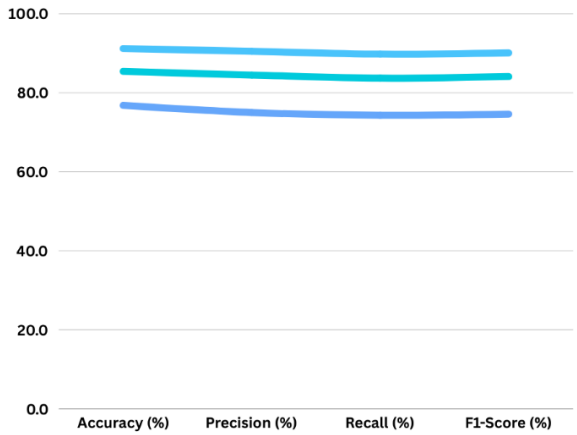

Figure 8: Result Graph.

Random Forest achieved the highest scores across all metrics, excelling in both precision and recall, which are critical for identifying IoT attacks. In inaccuracy and recall, DT performed competitively, however, it trailed RF by a small margin. When compared to supervised learning models, K-Means Clustering, which is mainly used for anomaly detection, showed poorer precision and recall but moderate accuracy.

4.5.1 Comparative Analysis

The developed models were compared against traditional methods like rule-based intrusion detection systems (IDS) to highlight improvements in detection capabilities (Figure 9).

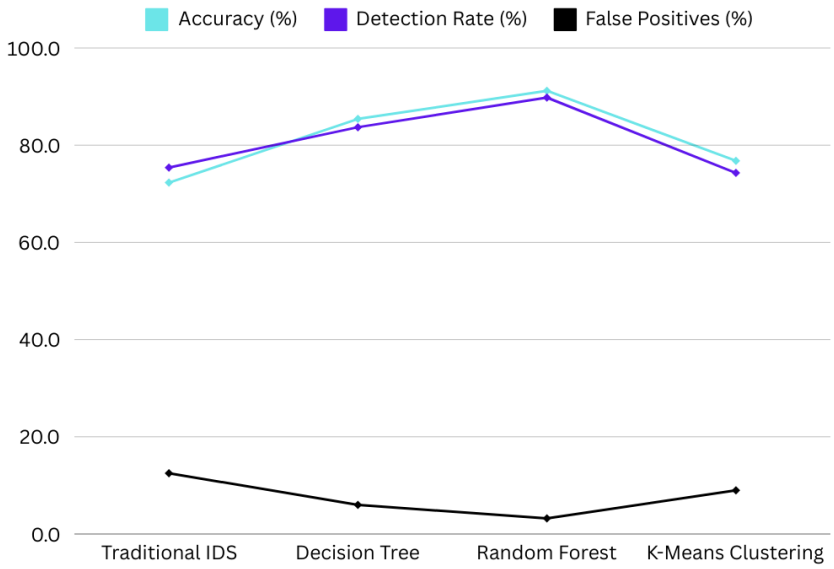

Figure 9: Performance Metrics Comparison

Random Forest outperformed traditional IDS by a significant margin in accuracy (+18.9%) and reduced false positives dramatically (from 12.5% to 3.2%). Decision Tree also displayed noticeable improvements but was less effective than Random Forest. K-means clustering provided moderate results and had a higher rate of false positives than Random Forest but still outperformed traditional IDS. The comparative analysis highlights the advantages of using advanced ML techniques like RF for IoT cybersecurity. These models not only enhance detection accuracy but also minimize false positives, reducing unnecessary system alerts.

4.6 Deployment

4.6.1 IoT Testbed Setup and Model Deployment

A simulated IoT testbed is designed using platforms like Node-RED or NS3 to replicate a real-world IoT environment. The setup includes virtual devices like smart cameras, thermostats, and IoT gateways, all of which generate continuous data streams. The trained ML models (DT, RF, and K-Means Clustering) are integrated into the IoT gateway, where they analyze incoming data to detect potential cyber threats. This real-time data analysis ensures prompt anomaly detection, enhancing the system's overall resilience.

4.6.2 Evaluation of Deployment

The performance of the deployed models is verified by measuring important metrics including latency, throughput, and false positive rate. Latency represents the detection time, throughput quantifies the system’s processing capacity, and the false positive rate assesses the accuracy of threat identification. Random Forest exhibits the best balance with low latency, high throughput, and minimal false positives, making it a suitable choice for robust IoT security (Figure 10).

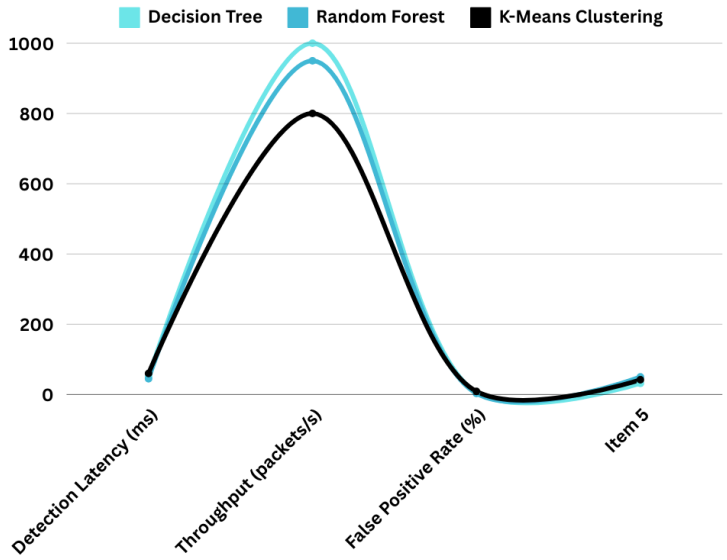

Figure 10: Performance of the Deployed Models.

These findings highlight the power of machine learning models in improving the security of the IoT ecosystem. Random Forest was found to be the most effective algorithm among the evaluated models in detecting and preventing cyber attacks, having efficient performance with high accuracy, low false positives, and real-time suitability (Caporale et al. 2021; Li and Liu 2021; Shandler and Gomez 2023). This is consistent with earlier studies that show that ensemble methods, and Random Forest in particular, are more suitable for complex and dynamic environments such as IoT systems where data is inherently noisy and diverse. These properties of the model allowed it to work well for IoT cybersecurity applications by handling massive amounts of data and catching any new patterns or attacks (Andrade et al. 2020; Chatfield and Reddick 2019; Lee 2020; Lu and Xu 2019). Although Random Forest proved to deliver the best overall results, both Decision Trees and K-Means clustering retained the leave to be useful, especially in scenarios with simple features, saving resources with their inherent abilities to do so. Decision trees had the advantage of interpretability and transparency – both of which help to understand how our model arrived at its decisions, but they potentially perform slightly worse on high dimensional data compared to Random Forests. K-Means clustering is another method that works promisingly for anomaly detection, yet it resulted in many false positives, as it was not able to accurately detect all anomalies in the data, where the device behavior blurred with the actual defined state behavior. This paves the way for the adoption of more advanced techniques in the form of hybrid models in IoT security anomaly detection.

As discussed above, one of the main challenges identified by this study was the computational cost of training the model and applying it in real time, which was especially expensive for large datasets. However, the complexity of such models should be taken into account when deploying them in real-time IoT environments where resources are limited, and, as the experiments conducted in this research show, such models can provide real-time capabilities of intrusion detection with very good detection rates. This can also be an area of future research in which lightweight models or edge computing integration can be used to perform faster data processing without losing detection accuracy. Last but not least, the results indicate that IoT security must adapt to the constantly changing threat environment over time, and it is important for the model not only to achieve performant results at training time but also to be able to update the model periodically based on the current threat landscape and detect threats in real-time. Now you may ask if you want to create a universal model you may have some problems, what if the data being fed into it is noisy or corrupted in some way? Combining these adaptive systems with IoT devices could potentially lead to a more scalable and efficient approach to optimizing cybersecurity in the face of increasingly interconnected and vulnerable IoT ecosystems.

4.7 Implications

This research can have three main implications: First, machine learning models such as Random Forests used in this research indicate the possibility of improving the detection and prevention of cyber elements such as botnets in IoT. As the time to detect and respond to threats has gained new importance due to the vast amount of IoT devices being operated and the volume of cyber-attacks, this approach provides a scalable, resource-efficient method for real-time threat analysis. The integration of machine learning models across organizations enables them to better detect, act upon, and prevent various cybersecurity attacks, making IoT systems more resilient. It is also suggested to create lightweight and adaptable models suitable for IoT ecosystems with limited computation environments. This can help reduce the need for centralized servers while also speeding up threat detection and decision-making due to the move towards edge computing in combination with machine learning.

Beyond its academic significance, the findings of this research have practical applications in real-world scenarios and can benefit sectors that heavily utilize IoT devices, including smart cities, healthcare, and industrial automation, to help protect their networks from increasingly advanced types of cyberattacks.

4.8 Future Research Direction

To increase detection rates and decrease false positives, future research may combine supervised and unsupervised learning methods with hybrid ML systems. However, due to the nature of the dimension of attack patterns in large-scale IoT environments, it is impossible to rely on a single method, so hybrid approaches such as ensemble models or deep learning-based methods are more efficient in this scenario. Moreover, adaptive learning systems, which are constantly updated with new data and evolving cyber threats, can keep the model effective against rapidly changing attack vectors. This approach can be used with machine learning models and can significantly reduce its response time as well as provide real-time threat detection for IoT systems. This will accelerate the response time and reduce the dependency on centralized servers. Investigating multi-stage attack scenarios, dealing with heterogeneous IoT devices, and addressing the challenges of emerging threats can also be accomplished to enhance the security of IoT workflows. Future IoT cybersecurity research should focus more on these aspects to maximize scalability, adaptability, and overall resilience in the future.

5. CONCLUSION

ML models were shown to be successful in identifying and reducing cybersecurity risks when they were used in a simulated Internet of Things environment. RF is the most dependable option for real-time IoT security among the algorithms examined because it performed the best in terms of accuracy, detection rate, and low false positive rate. While Decision Tree and K-Means Clustering also proved useful, they exhibited some limitations in latency and throughput. Overall, the proposed methodology is efficient and scalable, offering a feasible solution for enhancing IoT security that can be deployed and tested within a short timeframe. The results demonstrate how machine learning may greatly enhance danger identification and response in IoT situations.

REFERENCES

[1] Ahmady, Ezatullah, Abdul Rahman Mojadadi, and Musawer Hakimi. 2024. “A Comprehensive Review of Cybersecurity Measures in the IoT Era.” *Journal of Social Science Utilizing Technology* 2(1):288–98. doi: 10.70177/jssut.v2i1.722.

[2] Andrade, Roberto Omar, Sang Guun Yoo, Luis Tello-Oquendo, and Ivan Ortiz-Garces. 2020. “A Comprehensive Study of the IoT Cybersecurity in Smart Cities.” *IEEE Access*. doi: 10.1109/ACCESS.2020.3046442.

[3] Caleb, S., and John Justin Thangaraj. 2023. “Threat Detection And Mitigation In Self-Organizing Wireless Communication Network.” in *2023 2nd International Conference on Smart Technologies for Smart Nation, SmartTechCon 2023*.

[4] Caporale, Guglielmo Maria, Woo Young Kang, Fabio Spagnolo, and Nicola Spagnolo. 2021. “Cyber-Attacks, Spillovers and Contagion in the Cryptocurrency Markets.” *Journal of International Financial Markets, Institutions and Money*. doi: 10.1016/j.intfin.2021.101298.

[5] Chatfield, Akemi Takeoka, and Christopher G. Reddick. 2019. “A Framework for Internet of Things-Enabled Smart Government: A Case of IoT Cybersecurity Policies and Use Cases in U.S. Federal Government.” *Government Information Quarterly*. doi: 10.1016/j.giq.2018.09.007.

[6] Djenna, Amir, Saad Harous, and Djamel Eddine Saidouni. 2021. “Internet of Things Meet Internet of Threats: New Concern Cyber Security Issues of Critical Cyber Infrastructure.” *Applied Sciences (Switzerland)*. doi: 10.3390/app11104580.

[7] Khan, Amjad Rehman, Muhammad Kashif, Rutvij H. Jhaveri, Roshani Raut, Tanzila Saba, and Saeed Ali Bahaj. 2022. “Deep Learning for Intrusion Detection and Security of Internet of Things (IoT): Current Analysis, Challenges, and Possible Solutions.” *Security and Communication Networks*.

[8] Kolluru, Vinothkumar, Sudeep Mungara, and Advaitha Naidu Chintakunta. 2019. “Securing the Iot Ecosystem: Challenges and Innovations in Smart Device Cybersecurity.” *International Journal on Cryptography and Information Security* 9(2):37–51. doi: 10.5121/ijcis.2019.9203.

[9] Lee, In. 2020. “Internet of Things (IoT) Cybersecurity: Literature Review and Iot Cyber Risk Management.” *Future Internet*.

[10] Li, Yuchong, and Qinghui Liu. 2021. “A Comprehensive Review Study of Cyber-Attacks and Cyber Security; Emerging Trends and Recent Developments.” *Energy Reports*. doi: 10.1016/j.egyr.2021.08.126.

[11] Lu, Yang, and Li Da Xu. 2019. “Internet of Things (IoT) Cybersecurity Research: A Review of Current Research Topics.” *IEEE Internet of Things Journal*. doi: 10.1109/JIOT.2018.2869847.

[12] Maaz, Muhammad, Ghufran Ahmed, Ahmad Sami Al-Shamayleh, Adnan Akhunzada, Shahbaz Siddiqui, and Abdulla Hussein Al-Ghushami. 2024. "Empowering IoT Resilience: Hybrid Deep Learning Techniques for Enhanced Security." *IEEE Access* 1–1. doi: 10.1109/ACCESS.2024.3482005.

[13] Malathi, C., and I. Naga Padmaja. 2023. "Identification of Cyber Attacks Using Machine Learning in Smart IoT Networks." *Materials Today: Proceedings* 80:2518–23. doi: 10.1016/j.matpr.2021.06.400.

[14] Olabanji, Samuel Oladiipo, Yewande Alice Marquis, Chinasa Susan Adigwe, Samson Abidemi Ajayi, Tunbosun Oyewale Oladoyinbo, and Oluwaseun Oladeji Olaniyi. 2024. "AI-Driven Cloud Security: Examining the Impact of User Behavior Analysis on Threat Detection." *Asian Journal of Research in Computer Science*. doi: 10.9734/ajrcos/2024/v17i3424.

[15] Reddy Maddireddy, Bharath, and Bhargava Reddy Maddireddy. 2024. "Advancing Threat Detection: Utilizing Deep Learning Models for Enhanced Cybersecurity Protocols." 02:325–55.

[16] Sethi, Pallavi, and Smruti R. Sarangi. 2017. "Internet of Things: Architectures, Protocols, and Applications." *Journal of Electrical and Computer Engineering*.

[17] Shah, Varun. 2021. "Machine Learning Algorithms for Cybersecurity: Detecting and Preventing Threats." *Revista Española de Documentación Científica* 15(4):42–66. doi: 10.5281/zenodo.10779509.

[18] Shandler, Ryan, and Miguel Alberto Gomez. 2023. "The Hidden Threat of Cyber-Attacks–undermining Public Confidence in Government." *Journal of Information Technology and Politics*. doi: 10.1080/19331681.2022.2112796.

[19] Shihab, Mustafa Abdmajeed, Haydar Abdulameer Marhoon, Saadaldeen Rashid Ahmed, Bourair Al-Attar, Mushtaq T. Al-Sharify, and Ravi Sekhar. 2024. "Towards Resilient Machine Learning Models: Addressing Adversarial Attacks in Wireless Sensor Network." *Journal of Robotics and Control (JRC)* 5(5):1582–1602. doi: 10.18196/jrc.v5i5.23214.

[20] Sisinni, Emiliano, Abusayeed Saifullah, Song Han, Ulf Jennehag, and Mikael Gidlund. 2018. "Industrial Internet of Things: Challenges, Opportunities, and Directions." *IEEE Transactions on Industrial Informatics*. doi: 10.1109/TII.2018.2852491.

[21] Statista. 2023. "Number of Internet of Things (IoT) Connections."

[22] Ullah, Imtiaz, and Qusay H. Mahmoud. 2022. "Design and Development of RNN Anomaly Detection Model for IoT Networks." *IEEE Access*. doi: 10.1109/ACCESS.2022.3176317.

[23] Uprety, Aashma, and Danda B. Rawat. 2021. "Reinforcement Learning for IoT Security: A Comprehensive Survey." *IEEE Internet of Things Journal* 8(11):8693–8706. doi: 10.1109/JIOT.2020.3040957.

[24] Yu, Jia, Alexey V. Shvetsov, and Saeed Hamood Alsamhi. 2024. "Leveraging Machine Learning for Cybersecurity Resilience in Industry 4.0: Challenges and Future Directions." *IEEE Access* 12:159579–96. doi: 10.1109/ACCESS.2024.3482987.