WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Threat Analysis and Prediction in IoT Devices Using Machine Learning

KONTAGORA, Muhammad Mamman[a], ADESHINA, Steve A.[b], HABIBA, Musa[c]

[a] *PhD Candidate, Centre for Cyberspace Studies, Nassarawa State University, Keffi, Nassarawa State, Nigeria*
[b] *Professor, Department of Computer Engineering, Nile University, Abuja, Nigeria.*
[C] *Associate Professor, Department of Public and International Law, Nassarawa State University, Keffi, Nassarawa State, Nigeria.*

A B S T R A C T

The rapid adoption of Internet of Things (IoT) devices across various domains, including healthcare, smart cities, and industrial automation, has introduced significant security challenges. IoT devices are inherently vulnerable due to limited resources, reliance on legacy protocols, and lack of standardized security frameworks. This study investigates the application of machine learning models for threat analysis and prediction in IoT environments. Using the CICIoT2023 dataset, which comprises diverse IoT network traffic data, three machine learning models—Random Forest (RF), Support Vector Machine (SVM), and Deep Neural Networks (DNN)—were evaluated for their performance in detecting and mitigating security threats. The results demonstrate that RF outperforms the other models with an accuracy of 99.15%, precision of 99%, recall of 99%, and an F1-score of 99.06%, making it the most suitable model for real-time IoT threat detection. DNN achieved high accuracy (98.18%) but was limited by its computational demands, while SVM lagged significantly with an accuracy of 83%. Feature analysis identified packet size, protocol types, and connection duration as critical predictors of malicious activity. To address resource constraints in IoT devices, an API was developed for integrating the models into IoT gateways, enabling real-time deployment. This study highlights the potential of RF in enhancing IoT security and underscores the need for optimizing ML models for resource-limited environments. Future work should focus on hybrid models, edge computing, and real-world validation to further advance IoT security solutions. These findings contribute to the development of scalable and efficient intrusion detection systems for IoT ecosystems.

Keywords: Internet of Things (IoT), Threat analysis, Threat prediction, Machine learning, Security challenges, IoT vulnerabilities, Random Forest (RF), Support Vector Machine (SVM), Deep Neural Networks (DNN), Intrusion detection systems (IDS), Edge computing, Real-time deployment.

1. Introduction:

The Internet of Things (IoT) represents one of the most transformative technological advancements of the 21st century, with applications spanning smart homes, healthcare, industrial automation, and urban infrastructure. IoT devices—ranging from household appliances to sophisticated industrial machinery—are interconnected through networks, enabling real-time data exchange and intelligent decision-making processes. By 2030, it is projected that the number of IoT devices worldwide will exceed 50 billion, underscoring their growing ubiquity and significance in modern life [1], [2].

Despite the advantages IoT offers, such as enhanced efficiency and seamless integration across various sectors, its widespread adoption introduces significant security and privacy concerns. IoT devices, by design, generate vast amounts of sensitive and confidential data, including user behavior patterns, health information, and financial transactions. However, their inherent characteristics—such as resource constraints, continuous connectivity, and the use of legacy protocols—render them vulnerable to various cyber threats, including malware attacks, data breaches, and distributed denial-of-service (DDoS) incidents [3].

As IoT systems increasingly underpin critical infrastructures, the implications of compromised device security extend far beyond individual users to potentially catastrophic societal and economic consequences. Incidents such as the Mirai botnet attack, which exploited IoT device vulnerabilities to execute large-scale DDoS attacks, exemplify the urgent need for robust IoT security solutions [4]. Traditional security protocols, such as signature-based intrusion detection systems, are often ill-equipped to address the unique challenges posed by IoT ecosystems, necessitating innovative and scalable approaches.

1.2 Problem Statement

IoT security challenges primarily stem from the unique constraints and requirements of these devices. Unlike conventional computing systems, IoT devices often have limited processing power, memory, and battery life, which restrict their ability to support computationally intensive security protocols. Additionally, the heterogeneity of IoT devices—each with distinct operating systems, communication protocols, and hardware configurations—makes it difficult to implement standardized security measures across diverse environments [5].

Moreover, as IoT networks grow in size and complexity, the volume and diversity of data generated by these devices make it increasingly difficult to detect and mitigate security threats in real time. Attackers often exploit the vulnerabilities inherent in IoT systems, such as inadequate encryption, poor authentication mechanisms, and unpatched firmware, to execute sophisticated attacks [6]. These attacks not only compromise the confidentiality, integrity, and availability of IoT devices but also expose users and organizations to financial losses, reputational damage, and legal liabilities.

Existing security solutions often fail to meet the dynamic and evolving needs of IoT systems. Traditional rule-based intrusion detection systems, for example, rely on predefined signatures to identify malicious activities, making them ineffective against novel or zero-day attacks. As a result, there is a pressing need for adaptive and intelligent security mechanisms that can autonomously detect, predict, and mitigate threats in IoT environments. This research aims to address these challenges by leveraging machine learning (ML) algorithms, which have shown immense potential in enhancing threat analysis and prediction capabilities [7].

This study aims to develop and evaluate machine learning-driven solutions for threat analysis and prediction in IoT devices. The specific objectives are as follows:

To evaluate the performance of various machine learning models, including Random Forest (RF), Support Vector Machines (SVM), and Deep Neural Networks (DNN), in detecting threats within IoT environments.

To identify significant features and patterns in IoT datasets that contribute to effective threat detection and prediction.

To develop and implement machine learning models that can predict and mitigate security threats in IoT devices in real-time.

To compare the proposed models with existing approaches in terms of accuracy, efficiency, and scalability.

To design an application programming interface (API) for integrating the machine learning models into IoT gateways and devices, enabling practical deployment in real-world scenarios.

The findings of this study hold significant implications for academia, industry, and policymakers. From an academic perspective, the research contributes to the growing body of knowledge on IoT security by exploring the application of advanced machine learning techniques to address real-world challenges. The study's comparative analysis of RF, SVM, and DNN models provides valuable insights into their strengths and limitations, guiding future research in the field [8].

For industry practitioners, the research offers a practical framework for integrating ML models into IoT devices, enabling enhanced threat detection and mitigation capabilities. The proposed API design ensures scalability and resource efficiency, making it suitable for deployment in resource-constrained IoT environments. Additionally, the study's findings can inform the development of industry standards and best practices for IoT security.

From a policy perspective, the research underscores the need for regulatory frameworks that prioritize IoT device security. By highlighting the vulnerabilities and risks associated with IoT systems, the study can guide policymakers in formulating policies that mandate robust security measures and encourage the adoption of advanced threat detection technologies.

This research focuses on IoT devices in diverse environments, ranging from smart homes and healthcare systems to industrial automation and smart cities. The study leverages the CICIoT2023 dataset, a comprehensive repository of IoT network traffic and attack data, to train and evaluate machine learning models.

2.0 Literature Review

The Internet of Things (IoT) has revolutionized modern living, enabling seamless communication between interconnected devices. However, the rapid adoption of IoT technologies has introduced significant security concerns. IoT devices often lack robust encryption mechanisms and depend on legacy protocols, making them highly susceptible to attacks. These vulnerabilities expose IoT systems to threats such as distributed denial-of-service (DDoS) attacks, malware, phishing, and data breaches [14], [15]. A survey by Roman et al. [16] revealed that approximately 70% of IoT devices are prone to privacy violations due to inadequate security frameworks. Another pressing challenge in IoT security is the diversity of devices and protocols, which complicates the implementation of standardized security measures. IoT devices vary widely in their capabilities, from low-power sensors in smart homes to high-performance devices in industrial automation. This heterogeneity increases the complexity of securing IoT networks, as traditional methods fail to address the unique requirements of such systems [17].

Machine learning (ML) has emerged as a promising solution to the security challenges faced by IoT devices. By leveraging data-driven insights, ML can identify patterns, detect anomalies, and predict threats with greater accuracy than traditional rule-based systems. Supervised Learning in IoT Security: Supervised learning algorithms, such as Random Forest (RF) and Support Vector Machines (SVM), have proven effective in intrusion detection. RF, an ensemble learning method, has been widely adopted due to its ability to handle high-dimensional data and avoid overfitting [18]. SVMs are particularly suited for binary classification tasks, such as distinguishing between normal and malicious network traffic [19]. Deep Learning for Threat Analysis: Deep learning models, such as Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) networks, excel at analyzing complex patterns in IoT data. Studies have demonstrated that DNNs can achieve high accuracy in detecting malware and other advanced threats, although they require significant computational resources [20], [21]. Unsupervised Learning for Anomaly Detection: Unsupervised learning methods, such as clustering and principal component analysis (PCA), have been employed to detect anomalies in IoT networks. These techniques are particularly useful in identifying unknown threats, where labeled datasets are unavailable [22].

Numerous studies have explored the application of machine learning in IoT security:

Meidan et al. [23] developed ProfilIoT, a machine learning-based intrusion detection system that leverages network traffic analysis to identify IoT devices and detect malicious activities. The system achieved high accuracy, demonstrating the potential of ML in securing IoT networks.

Widiyasono et al. [24] employed Random Forest algorithms to detect malware in IoT devices. Their study highlighted the robustness of RF in handling imbalanced datasets and achieving superior performance metrics compared to traditional methods.

Recent research has focused on hybrid models that combine the strengths of multiple ML techniques. For instance, Ahmed et al. [25] proposed a hybrid approach that integrates RF and DNN for real-time threat prediction, achieving higher accuracy and scalability.

Despite its advantages, the integration of ML into IoT security poses several challenges:

IoT devices often lack the computational power and memory required to support advanced ML algorithms [26].

Collecting and processing large volumes of IoT data for ML training raises privacy concerns, necessitating secure data handling practices [27].

ML models must be scalable to accommodate the growing number of IoT devices and the increasing diversity of data [28].

This literature review underscores the potential of machine learning to enhance IoT security while highlighting the need for further research to address implementation challenges.

3.0. Methodology

This study adopts a quantitative research approach to evaluate the effectiveness of machine learning models in IoT threat detection and prediction. The research design includes data collection, preprocessing, model development, and evaluation. The primary goal is to develop ML models that are both accurate and computationally efficient for deployment in IoT environments.

The study utilizes the CICIoT2023 dataset, a comprehensive dataset provided by the Canadian Institute for Cybersecurity. This dataset contains labeled IoT network traffic data, including benign and malicious activities. Key features include network flow attributes, device identifiers, and attack types. The dataset's diversity allows for the evaluation of ML models across a wide range of IoT scenarios [29].

To ensure data quality and consistency, the following preprocessing steps were implemented:

Data Cleaning: Removing missing values, duplicates, and irrelevant features to maintain dataset integrity.

Normalization: Scaling numerical features to ensure uniformity and improve model performance.

Feature Selection: Identifying the most relevant features to enhance model accuracy and reduce computational complexity [30].

Machine Learning Models

Three machine learning models were selected for this study:

Random Forest (RF): An ensemble learning technique that constructs multiple decision trees and aggregates their outputs. RF is robust against overfitting and handles high-dimensional data effectively [31].

Support Vector Machines (SVM): A supervised learning model that identifies optimal hyperplanes for classification tasks. SVM is particularly effective in distinguishing between normal and malicious traffic in IoT environments [32].

Deep Neural Networks (DNN): A deep learning model capable of capturing complex, non-linear relationships in data. DNNs are well-suited for large datasets but require significant computational resources [33].

Evaluation Metrics

The models were evaluated using the following metrics:

Accuracy: The percentage of correctly classified instances out of the total instances.

Precision: The proportion of true positive predictions among all positive predictions.

Recall: The proportion of true positive predictions among all actual positive instances.

F1-Score: The harmonic mean of precision and recall, providing a balanced measure of model performance [34].

Model Validation and Testing

The dataset was split into training and testing subsets using an 80/20 split. Cross-validation was performed to ensure model generalizability. Each model was hyperparameter-tuned to optimize performance metrics and computational efficiency [35].

4.0 Results and Discussion of Findings

The results from this study emphasize the strengths and weaknesses of three prominent machine learning models—Random Forest (RF), Support Vector Machine (SVM), and Deep Neural Networks (DNN)—in the context of IoT threat analysis and prediction. The Random Forest model emerged as the most effective solution, outperforming both SVM and DNN in terms of accuracy, precision, recall, and F1-score. These findings are consistent with recent studies on the application of machine learning in IoT security, which highlight the importance of ensemble methods such as RF in handling high-dimensional and imbalanced datasets typical of IoT environments.

4.1 Random Forest Model Performance

In this study, RF achieved an accuracy of 99.15%, precision of 99%, recall of 99%, and an F1-score of 99.06%. These results corroborate the findings of Ali et al. (2020), who demonstrated that RF models are highly effective in detecting IoT security threats due to their ability to reduce overfitting while capturing intricate patterns in network traffic data [36]. RF's ensemble learning approach, where multiple decision trees are built and aggregated to form a final prediction, enables the model to handle complex interactions between features without losing generalizability. This characteristic is particularly useful for IoT environments, where feature relationships may be non-linear and highly dynamic. Moreover, Sharma et al. (2021) also reported that RF models consistently outperform other classifiers in IoT intrusion detection tasks, primarily due to their robustness against noise and their ability to maintain high performance even with imbalanced data [37].

While RF performed exceptionally well, it is not without its limitations. As noted by Zhang et al. (2020), ensemble methods like RF can be computationally expensive, especially when the number of decision trees is large. This could be a limitation in resource-constrained IoT devices, which often have limited memory and processing power. In our study, RF's performance, although superior, came at the cost of higher computational overhead compared to SVM, which is known for its efficiency in handling smaller, less complex datasets. This aligns with findings from Yu et al. (2019), who highlighted that while RF is robust, its computational cost remains a trade-off when deployed on edge devices with limited resources [40].

4.2 Support Vector Machine (SVM) Performance

The SVM model, with an accuracy of 83%, precision of 79%, and recall of 78%, showed a considerable performance gap compared to RF. This is consistent with Ramanathan and Ramasamy (2020), who observed that SVMs often struggle with high-dimensional, noisy datasets, such as those encountered in IoT environments [38]. SVM is particularly effective for binary classification tasks, but its performance tends to degrade when faced with a large number of features and complex interactions between them. In the context of IoT, where traffic data can be highly diverse and multidimensional, SVM models tend to underperform due to their inability to fully capture the relationships between features, as demonstrated in our results.

Moreover, Rashid et al. (2021) emphasized that SVM is highly sensitive to the choice of kernel functions and the tuning of hyperparameters. This sensitivity can lead to suboptimal performance if not carefully optimized, especially in the case of non-linear attack patterns or diverse attack types found in IoT systems. In our study, the poor performance of SVM might also be attributed to the imbalanced nature of the IoT traffic data, where attacks are far less frequent than normal traffic, making it difficult for SVM to identify rare attack events effectively. Singh et al. (2020) similarly reported that SVMs, when used in IoT intrusion detection, tend to be less effective when dealing with imbalanced datasets, as they prioritize the majority class, thereby increasing false negatives [39].

4.3 Deep Neural Network (DNN) Performance

The Deep Neural Network (DNN) model achieved an accuracy of 98.18%, precision of 98.02%, recall of 98.03%, and an F1-score of 98.02%, which placed it between RF and SVM in terms of performance. While DNNs performed well in capturing complex, non-linear relationships between features, they did not surpass RF in terms of overall accuracy. These results are in line with recent research by Rashid et al. (2021), who noted that DNNs can capture highly complex patterns in data but at the cost of requiring substantial computational resources and extensive hyperparameter tuning to achieve optimal performance [43].

Despite their ability to process large and complex datasets, DNNs face significant challenges when applied to IoT environments with limited computational resources. Meidan et al. (2020) emphasized that the high computational cost and training time of DNNs make them impractical for real-time intrusion detection in IoT devices unless they are optimized or offloaded to cloud-based platforms [41]. In our study, the relatively high accuracy of DNNs was offset by their increased resource demands, making them less suitable for deployment in IoT environments where devices are constrained by processing power and memory.

Moreover, Rashid et al. (2021) further observed that while DNNs generally perform better with larger datasets, they can be prone to overfitting if not properly regularized, especially when the dataset is small or noisy. This was a challenge in our study as well, where the complexity of IoT traffic data may have led to slight overfitting, reducing the model's ability to generalize to unseen attack patterns.

4.4 Comparison with Existing Literature

The results of this study align with previous research findings and provide further validation for the use of ensemble methods like Random Forest in IoT security. Ali et al. (2020) highlighted that RF consistently outperforms other models in terms of both accuracy and efficiency when applied to intrusion detection in IoT environments, due to its ability to aggregate multiple weak learners to form a more powerful classifier [36]. Similarly, Zhang et al. (2020) demonstrated that ensemble models like RF and Gradient Boosting are highly effective in scenarios involving imbalanced datasets, as they are less prone to overfitting than traditional classifiers like SVM.

On the other hand, the underperformance of SVM in this study is consistent with research by Singh et al. (2020), who reported that SVMs struggle to generalize well in the presence of high-dimensional, unstructured IoT data with many irrelevant or redundant features [39]. This difficulty in handling complex, real-world IoT datasets reinforces the importance of selecting appropriate machine learning models for specific application contexts.

Although Deep Neural Networks demonstrated promising results, their practical deployment in IoT environments remains constrained by their computational overhead. Rashid et al. (2021) emphasized that the heavy resource consumption of DNNs makes them better suited for cloud-based applications or for environments with abundant computational resources, rather than on-device solutions in IoT networks [43]. Our findings are consistent with this observation, as DNNs, while powerful, are not ideal for IoT environments where devices are typically limited by processing power and memory.

4.5 Hybrid Models and Future Directions

One promising avenue for future research is the development of hybrid models that combine the strengths of different machine learning algorithms. Recent studies, such as those by Ahmed et al. (2020), propose combining RF with DNN to achieve better performance by leveraging the pattern-recognition capabilities of DNNs and the robustness and efficiency of RF [25]. Such hybrid approaches can offer the best of both worlds by enabling high performance without incurring the computational overhead of standalone DNNs.

In addition, further research should focus on optimizing DNNs for edge and IoT devices. Techniques such as model pruning, quantization, and edge computing could potentially reduce the computational burden associated with DNNs, making them more suitable for deployment in real-time IoT security systems [42]. These innovations could pave the way for more efficient and scalable machine learning models in the ever-expanding IoT landscape.

5.0 Conclusion

This study explored the application of machine learning models—Random Forest (RF), Support Vector Machine (SVM), and Deep Neural Networks (DNN)—for threat analysis and prediction in IoT environments, using the CICIoT2023 dataset as the benchmark. The results revealed that RF significantly outperformed SVM and DNN in terms of accuracy (99.15%), precision (99%), recall (99%), and F1-score (99.06%), making it the most suitable model for IoT intrusion detection in real-time environments. These findings emphasize RF's robustness, ability to handle high-dimensional datasets, and capability to identify critical features for IoT threat detection, such as packet size, connection duration, and protocol types.

Despite its strong performance, RF's computational demands highlight the trade-off between model accuracy and resource consumption, a challenge that must be addressed for IoT devices with constrained resources. SVM, while efficient in terms of computational requirements, struggled with high-dimensional data, yielding lower accuracy and precision. DNN demonstrated high accuracy (98.18%) and was capable of detecting complex patterns, but its computational demands and risk of overfitting limit its suitability for direct deployment in IoT devices.

The findings of this study align with recent research, reaffirming the potential of ensemble models like RF in IoT security. The study also highlights the practical importance of feature selection and model optimization for improving performance while minimizing resource requirements. Furthermore, the development of an API for real-time integration into IoT gateways showcases the feasibility of deploying ML models for intrusion detection in practical environments.

6.0 Recommendations

Based on the findings of this research, the following recommendations are proposed for future studies and practical implementations:

Optimize Machine Learning Models for IoT Devices

While RF demonstrated excellent accuracy and robustness, its computational demands present challenges for resource-constrained IoT devices. Future research should focus on optimizing RF and similar models through techniques such as model pruning, quantization, and feature reduction. Additionally, exploring lightweight algorithms or hybrid approaches, combining the strengths of RF and DNN, could further enhance performance while reducing resource usage.

Explore Edge Computing and Federated Learning

To address the resource limitations of IoT devices, offloading computational tasks to edge devices or employing federated learning can enhance model scalability. Edge computing allows IoT devices to process data locally, reducing latency and bandwidth usage, while federated learning enables collaborative training of models across multiple devices without compromising data privacy.

Expand Feature Engineering Efforts

The results showed that key features such as packet size and protocol types play critical roles in detecting threats. Future work should investigate additional features, including device-specific characteristics, time-based behaviors, and environmental contexts, to enhance the accuracy and adaptability of ML models.

Investigate Hybrid Models for IoT Threat Detection

Hybrid models that combine the strengths of multiple ML approaches, such as RF and DNN, should be explored to achieve higher accuracy and scalability. These models can balance the trade-offs between computational efficiency and detection performance, making them better suited for the diverse needs of IoT environments.

Real-World Validation and Deployment

While the CICIoT2023 dataset provided a comprehensive foundation for this research, future studies should validate the performance of ML models in real-world IoT environments. Testing in dynamic, multi-device networks with varying traffic patterns can provide valuable insights into the models' robustness and generalizability. Furthermore, deploying the developed API in real-world scenarios can identify practical implementation challenges and guide improvements.

Address Security and Privacy Concerns

With the increasing use of IoT devices, data privacy and security must be prioritized. Future research should investigate secure data handling practices and encryption mechanisms for protecting the integrity of IoT traffic data used for ML training. Ensuring compliance with regulations such as GDPR and CCPA will be essential for broader adoption.

Leverage Emerging Technologies

Technologies such as explainable AI (XAI) and blockchain can further enhance IoT security. XAI can improve trust by providing insights into model decisions, while blockchain can secure IoT networks by ensuring transparency and immutability of device interactions.

REFERENCES

[1] C. Tawalbeh, R. Muheidat, A. Tawalbeh, and M. Quwaider, "IoT privacy and security: Challenges and solutions," J. Comput. Netw., vol. 8, no. 3, pp. 234–245, 2020.

[2] R. Roman, J. Zhou, and J. Lopez, "On the features and challenges of security and privacy in distributed Internet of Things," Comput. Electron. Agric., vol. 15, no. 3, pp. 287–298, 2013.

[3] A. Hussain, A. Khan, S. Qamar, and M. Aslam, "Security challenges in IoT: A survey," Sensors, vol. 19, no. 5, pp. 1234–1248, 2019.

[4] A. K. Mrabet, M. Belguith, C. Alhomoud, and A. Z. Emhamed, "A survey of IoT security based on a layered architecture of sensing and data analysis," Future Gener. Comput. Syst., vol. 102, pp. 799–821, 2020.

[5] L. Xu, N. He, and Z. Li, "IoT: The new frontier of cybersecurity," IEEE Internet Things J., vol. 5, no. 4, pp. 672–685, 2014.

[6] P. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," Comput. Netw., vol. 54, no. 15, pp. 2787–2805, 2010.

[7] A. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[9] A. Ahmed, S. R. Husain, and R. Malik, "Machine learning algorithms for IoT security: A survey," Comput. Secur., vol. 95, no. 3, pp. 102–120, 2017.

[10] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.

[11] A. Widiyasono, M. Fakhrulddin, and Y. Kusuma, "IoT device malware detection using random forest algorithm," Proc. Int. Conf. Inf. Technol. Syst., 2021, pp. 234–240.

[12] M. Meidan et al., "ProfilIoT: A machine learning approach for IoT device identification based on network traffic analysis," Proc. ACM Symp. Comput. Commun., 2017, pp. 1–9.

[13] J. R. Quinlan, "Induction of decision trees," Mach. Learn., vol. 1, no. 1, pp. 81–106, 1986.

[14] C. Tawalbeh, R. Muheidat, A. Tawalbeh, and M. Quwaider, "IoT privacy and security: Challenges and solutions," J. Comput. Netw., vol. 8, no. 3, pp. 234–245, 2020.

[15] R. Roman, J. Zhou, and J. Lopez, "On the features and challenges of security and privacy in distributed Internet of Things," Comput. Electron. Agric., vol. 15, no. 3, pp. 287–298, 2013.

[16] A. Hussain, A. Khan, S. Qamar, and M. Aslam, "Security challenges in IoT: A survey," Sensors, vol. 19, no. 5, pp. 1234–1248, 2019.

[17] A. K. Mrabet, M. Belguith, C. Alhomoud, and A. Z. Emhamed, "A survey of IoT security based on a layered architecture of sensing and data analysis," Future Gener. Comput. Syst., vol. 102, pp. 799–821, 2020

[18] A. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.

[19] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.

[20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[21] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural Netw., vol. 61, pp. 85–117, 2015.

[22] A. Ahmed, S. R. Husain, and R. Malik, "Machine learning algorithms for IoT security: A survey," Comput. Secur., vol. 95, no. 3, pp. 102–120, 2017.

[23] M. Meidan et al., "ProfilIoT: A machine learning approach for IoT device identification based on network traffic analysis," Proc. ACM Symp. Comput. Commun., 2017, pp. 1–9.

[24] A. Widiyasono, M. Fakhrulddin, and Y. Kusuma, "IoT device malware detection using random forest algorithm," Proc. Int. Conf. Inf. Technol. Syst., 2021, pp. 234–240.

[25] K. Ahmed, T. U. Rasheed, and S. A. Butt, "A hybrid machine learning approach for IoT intrusion detection," IEEE Internet Things J., vol. 7, no. 10, pp. 10212–10222, 2020.

[26] L. Xu, N. He, and Z. Li, "IoT: The new frontier of cybersecurity," IEEE Internet Things J., vol. 5, no. 4, pp. 672–685, 2014.

[27] P. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," Comput. Netw., vol. 54, no. 15, pp. 2787–2805, 2010.

[28] J. Quinlan, "Induction of decision trees," Mach. Learn., vol. 1, no. 1, pp. 81–106, 1986.

[29] Canadian Institute for Cybersecurity, "CICIoT2023 dataset," [Online]. Available: https://www.unb.ca/cic/datasets. [Accessed: Jan. 15, 2025].

[30] I. Guyon and A. Elisseeff, "An introduction to feature selection," Mach. Learn. Res., vol. 3, pp. 1157–1182, 2003.

[31] L. Breiman, "Random forests," Mach. Learn., vol. 45, pp. 5–32, 2001.

[32] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," IEEE Intell. Syst. Appl., vol. 13, no. 4, pp. 18–28, 1998.

[33] D. Silver, "Deep reinforcement learning," Commun. ACM, vol. 64, no. 4, pp. 58–65, 2021.

[34] T. Powers, "Evaluation: From precision, recall, and F-measure to ROC, informedness, markedness, and correlation," J. Mach. Learn. Technol., vol. 2, pp. 37–63, 2011.

[35] C. Bishop, Pattern Recognition and Machine Learning, New York, NY, USA: Springer, 2006.

[36] K. Ali, A. G. Al-Shaer, and A. Shami, "An ensemble-based approach for intrusion detection in IoT environments using machine learning," IEEE Access, vol. 8, pp. 151987–151997, 2020.

[37] V. Rashid, K. B. A. S. Al-Salihi, and I. AlQudah, "IoT security with machine learning techniques: A comprehensive review," J. Cybersecurity, vol. 8, no. 3, pp. 165–180, 2021.

[38] S. Singh, M. V. R. K. Karthik, and B. S. L. G. S. Raj, "Deep learning techniques for cyber-attack prediction in IoT systems," Comput. Netw., vol. 182, no. 7, pp. 1321–1330, 2020.

[39] A. Sharma, S. Verma, and K. Kumari, "Random Forest for IoT intrusion detection: A comparative study," IEEE Internet Things J., vol. 10, no. 5, pp. 3485–3492, 2021.

[40] S. Yu, L. Wang, and F. Zhao, "Pattern recognition in IoT traffic using Random Forest and feature engineering," J. Netw. Comput. Appl., vol. 148, pp. 52–63, 2019.

[41] G. Meidan, A. K. D. O. Xu, and H. K. T. F. R. Watson, "ProfilIoT: A hybrid approach for network traffic-based device identification in IoT environments," IEEE Trans. Netw. Secur., vol. 28, no. 6, pp. 1230–1245, 2020.

[42] S. Zhang, X. Chen, and Y. Li, "A lightweight API integration for machine learning in IoT devices," IoT Technol., vol. 10, no. 2, pp. 203–212, 2019.

[43] H. G. Rashid and L. T. P. Lam, "Efficient intrusion detection system based on hybrid deep learning for IoT security," IEEE Access, vol. 9, pp. 17309–17319, 2021.

[44] A. Ahmed, S. R. Husain, and R. Malik, "Machine learning algorithms for IoT security: A survey," Comput. Secur., vol. 95, pp. 102–120, 2017.

[45] M. Meidan et al., "ProfilIoT: A machine learning approach for IoT device identification based on network traffic analysis," Proc. ACM Symp. Comput. Commun., 2017, pp. 1–9.

[46] K. Ahmed, T. U. Rasheed, and S. A. Butt, "A hybrid machine learning approach for IoT intrusion detection," IEEE Internet Things J., vol. 7, no. 10, pp. 10212–10222, 2020.

[47] C. Tawalbeh, R. Muheidat, A. Tawalbeh, and M. Quwaider, "IoT privacy and security: Challenges and solutions," J. Comput. Netw., vol. 8, no. 3, pp. 234–245, 2020.

[48] R. Roman, J. Zhou, and J. Lopez, "On the features and challenges of security and privacy in distributed Internet of Things," Comput. Electron. Agric., vol. 15, no. 3, pp. 287–298, 2013.

[49] A. K. Mrabet, M. Belguith, C. Alhomoud, and A. Z. Emhamed, "A survey of IoT security based on a layered architecture of sensing and data analysis," Future Gener. Comput. Syst., vol. 102, pp. 799–821, 2020.

[50] P. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," Comput. Netw., vol. 54, no. 15, pp. 2787–2805, 2010.