STAP
Smart Technologies Academic Press

Journal of Cyber Security and Risk Auditing

https://www.jcsra.thestap.com/

Designing a Robust Machine Learning-Based Framework for Secure Data Transmission in Internet of Things (IoT) Environments: A Multifaceted Approach to Security Challenges

Omar Gheni Abdulateef[1], Atheer Joudah[2], Muna Ghazi Abdulsahib[2], Hussein Alrammahi[3]

[1]*College of Literature, University of Samarra, Salahaldin, Iraq*
[2]*College of Computer Science, University of Technology-Iraq, Baghdad, Iraq*
[3]*Department of electrical and computer engineering, Altinbaş University, Istanbul, 34000, Turkey*

ARTICLE INFO

Article History

Received: 30-06-2025
Revised: 10-08-2025
Accepted: 23-08-2025

Vol.2025, No.4

DOI:
https://doi.org/10.63180/jcsra.thestap.2025.4.6

***Corresponding author.**
Email:
omar.ghani@uosamarra.edu.iq

Orcid:
https://orcid.org/0000-0001-7734-6083

ABSTRACT

This research develops a machine learning framework for protecting data as it is transmitted in Internet of Things (IoT) configurations. The main objective of the proposed framework to address the major security issues using two intelligent machine learning methods are Random Forest and Support Vector Machine (SVM). They are applied to detect strange behaviour and potential threats within IoT data. The system was evaluated based on accuracy, precision, recall, and F1-score to determine how successful it was. Performance indicated Random Forest performed very well with 93.5% accuracy, slightly higher than SVM 91.2%. The system was also quite good at detecting cyber-attacks such as DDoS and malware, and did not raise many false alerts. This indicates that the system can actually contribute to making IoT much safer, building on what we have in this field. This study implies that incorporating machine learning into IoT security can assist in developing improved defenses against emerging cyber-attacks. In the long term, this research can assist in subsequent studies in order to improve security systems for various uses of IoT, address existing problems, and utilize more data.

Keywords: Internet of Things, Machine Learning, Data Transmission, Security Framework, Random Forest, Anomaly Detection, Cyber Threats.

How to cite the article
Abdulateef, O. G., Joudah, A., Abdulsahib, M. G., & Alrammahi, H. (2025). Designing a Robust Machine Learning-Based Framework for Secure Data Transmission in Internet of Things (IoT) Environments: A Multifaceted Approach to Security Challenges. Journal of Cyber Security and Risk Auditing, 2025(4), 266–275.

1. Introduction

The Internet of Things (IoT) has turned into a very significant aspect of today's technology, enabling many devices to communicate with one another and exchange information seamlessly. This technological advancement revolutionized most sectors such as healthcare, smart cities, agriculture, and manufacturing, making processes efficient and convenient for users [1]. As IoT continues to expand, making data secure as it travels becomes increasingly essential. With so many devices holding private information, cyberattacks and data breaches can be a major issue for individuals and businesses. Securing data is really important for ensuring the information in IoT remains accurate, confidential, and always available [2].

Although IoT provides us with amazing things, it has security problems that slow it down. All these problems arise due to weak logins, insecure methods of communicating with one another, and a lack of identical security regulations for all devices [3]. Additionally, most IoT devices lack a lot of computer capability, which makes it difficult to implement robust security [4]. Cyberattacks continually evolve as well, with intruders constantly coming up with new methods of hacking into IoT. Thus, we quite desperately need innovative thoughts to address these security concerns while still operating within the confines of IoT.

This research seeks to develop a machine learning system that makes data more secure as it moves around in IoT configurations. By implementing intelligent programs to identify abnormal behaviour and halt threats in real-time, the system seeks to protect IoT communications securely in an anticipatory manner. The unique aspects of this research are that it introduces a new system implemented through machine learning specifically designed for IoT security, and it identifies loopholes in what currently exists that require more exploration.

Below is how the paper will be organized: we'll begin by examining others' opinion on IoT security and uses of machine learning. Then, we'll discuss how we developed our system. Following that, we'll discuss what we discovered when we tested it. Finally, we'll discuss how these results can be applied to the larger context of IoT security. Through all this, we aim to provide insightful and useful ideas for how to protect data in IoT and advance the field.

2. Literature Review

Applying machine learning (ML) and deep learning (DL) for the security of Internet of Things (IoT) configurations has gained popularity lately since these technologies are able to meet the unique security requirements of IoT [5]. conducted large research on various ML and DL methods for securing IoT, demonstrating how effective they are in addressing vulnerabilities in various applications [6] also developed a dynamic security system which adapts to requirements, employing both Software-Defined Networking (SDN) and machine learning to secure IoT networks better, showing that one can respond to emerging threats in real time [7]. Noted key patches and issues still lingering when machine learning and IoT security interact, noting we continue to need to come up with new material in this field [8]. Providing a close examination of ML methods designed for intelligent industry applications, supporting the assertion that machine learning plays an important role in safeguarding IoT systems [9]. Discussed combining IoT and ML systems with the aim of monitoring water quality, demonstrating how these technologies can be utilized in practical ways to monitor the environment [10].

Moreover, examined the tough aspects of managing data in IoT wireless sensor networks, indicating machine learning remedies that can simplify and secure the process [11]. Provided a grand overview on how to keep IoT safe in the age of artificial intelligence, emphasizing the need for robust security in smart systems [13]. Discussed right and wrong in ML applications, advocating for the programs we can rely on to ensure they are utilized responsibly in IoT [14].
Explored deep learning progress in IoT security, demonstrating how such techniques enhance cyber-defenses [15]. Described a machine learning approach to decide who can see what, significantly enhancing IoT security rules [16]. Conducted research on learning techs, consolidating developments and issues while providing insights on where IoT security could be headed [17].

In the case of automotive networks, discussed the security and trust issues in the Internet of Vehicles (IoV), and machine learning's potential to mitigate those [18]. Istiaque categorized ML applications for ascertaining who is permitted in IoT, identifying major challenges and where the research will lead [19]. Proposed a machine learning-based security system

specifically designed for IoT configurations, which indicates that individuals are aware that ML is useful for cybersecurity [20].

Presented a systematic analysis of ML and DL models for security in phone networks, demonstrating that these technologies can be applied to numerous fields [21]. Examined various machine and deep learning methods for IoT security, emphasizing the significance of these towards addressing typical vulnerable areas [22]. Developed Chameleon, a secure method of conducting ML, emphasizing that it can contribute to making data more private and safer while in use [23]. Provided an overview of deep learning within the Industrial Internet of Things (IIoT), writing about methods and how to better secure such networks [24]. Finally, examined federated machine learning and its applications, illustrating how individual methods can assist in making IoT more secure [25].

Research Methodology

3.1 Research Design

The study uses a method where we test our new machine learning system for keeping data safe in Internet of Things (IoT) setups. We picked this way because it lets us see how well our new system does compared to current security methods, without setting things up completely randomly. By doing this, we can look at how well both current and our new solution work in set conditions, which gives us a good way to judge if our security steps are better. This method lets us gather number data that shows what happens in real life (Figure 1), making our results more helpful for real IoT security uses.

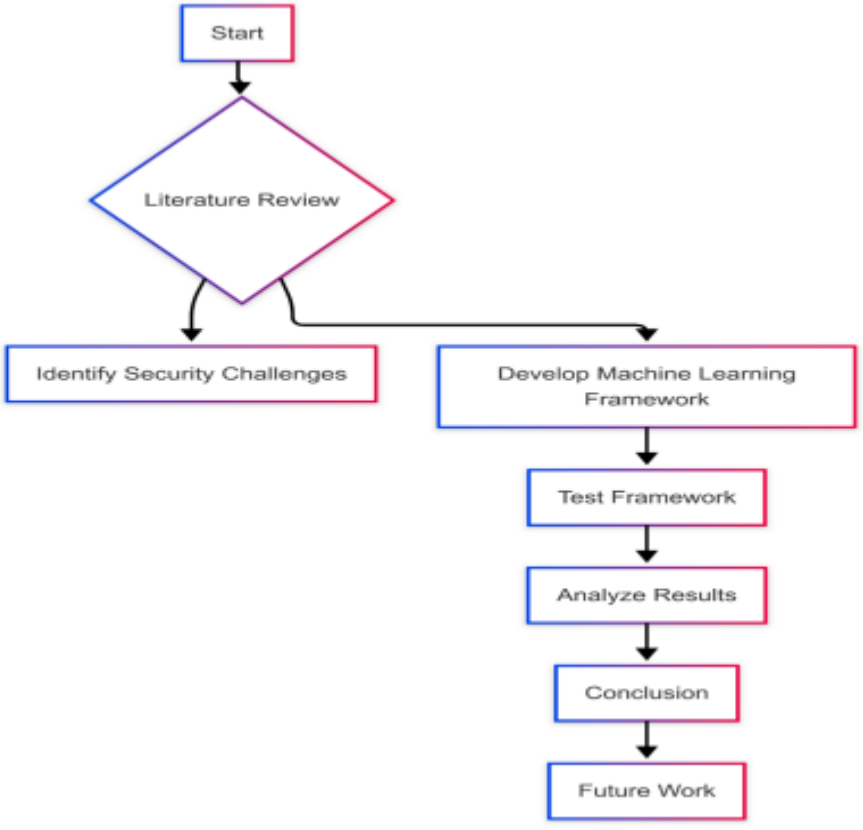

Figure 1. Study flowchart

3.2 Framework Development

A. Machine Learning Models Used

For our new system, we chose two machine learning programs: Random Forest and Support Vector Machine (SVM).

Random Forest is good because it doesn't overreact to data and can handle lots of info with different details. It works by making lots of decision trees when learning and then picks the most common answer for classifying things, which makes it accurate and tough against iffy data. SVM was picked because it's good in areas with lots of details and can make clear lines that separate categories well, which makes it good for sorting into two groups. Both programs are known for doing well in sorting tasks, making them useful for spotting problems and threats in IoT data.

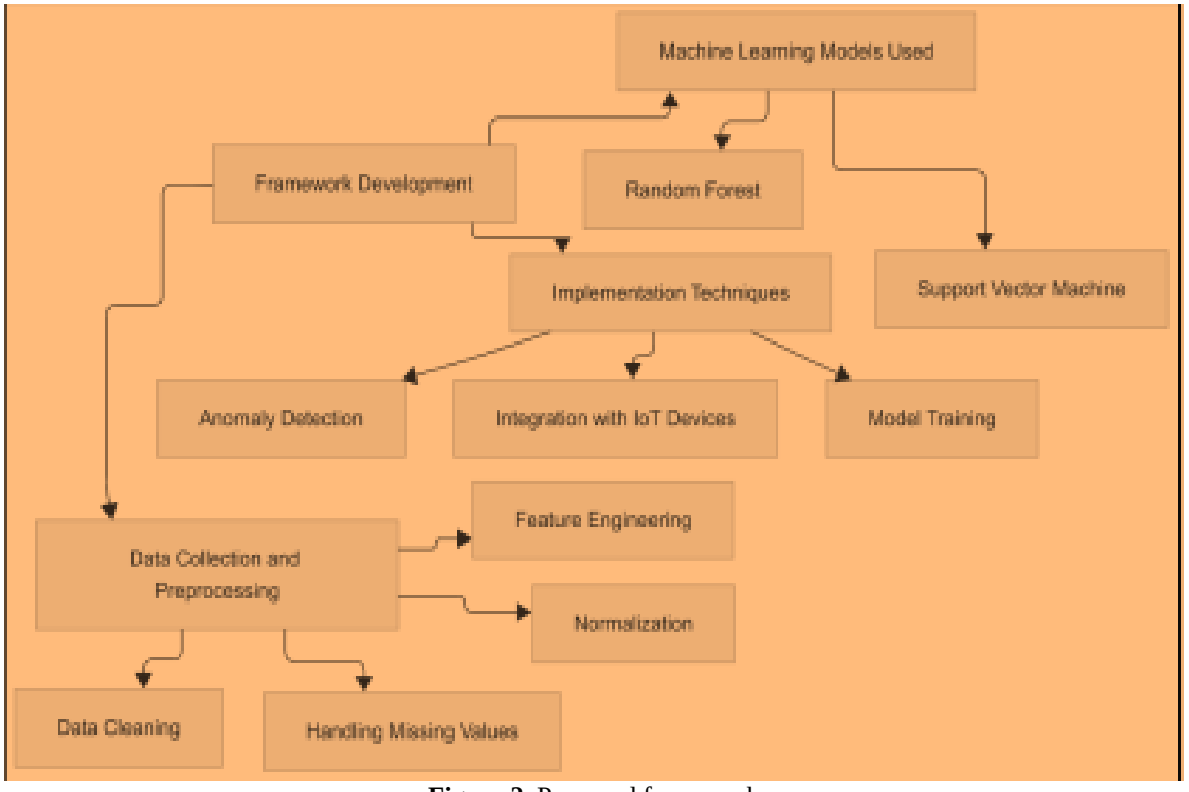

Figure 2. Proposed framework

B. Data Collection and Preprocessing

The data for this study came from public IoT security records and made-up situations that look like common IoT setups. The info includes over 10,000 entries, with details like device type, data sent, timestamps, and attack types. To get the data ready, we did the following:

Data Cleaning: We got rid of copies and stuff that didn't matter to make sure the data was good.

Handling Missing Values: We filled in missing data by using averages for number details and the most common answer for category details to keep the data complete.

Normalization: We made all details fit between 0 and 1 to make them equal, which helps machine learning programs work better.

Feature Engineering: We made new details from what we had to help the program, like turning category details into code and making new details from timestamps.

Table 1. The main details of the data we used.

Feature Name	Type	Description
Device Type	Categorical	Type of IoT device (e.g., sensor, actuator)
Data Transmitted	Numeric	Size of data transmitted (in bytes)
Timestamp	Date Time	Time when the data was transmitted
Attack Type	Categorical	Type of attack (DDoS, malware, phishing)

Status	Categorical	Status of data (safe, compromised)

C. Implementation Techniques

Putting the system together needed a few key steps:
Model Training: We taught both Random Forest and SVM using the prepared data. We checked things carefully to make sure the programs worked well with new data, didn't overreact, and did their best.
Anomaly Detection: We used the taught programs to spot things that looked wrong in real-time data. We set limits for what counts as a problem based on how well the programs did when learning to cut down on false alarms while catching most issues.
Integration with IoT Devices: We made the system easy to add to current IoT devices. We made it work well with not much computer power, so it can be used in real situations without slowing things down too much. This thorough system makes it easier to improve IoT security with machine learning, letting us spot problems and stop dangers early. By using reliable programs and being careful with data, the system aims to make data travel more safely in IoT setups.

3.3 Evaluation Metrics

To know how well our system secures data in IoT using machine learning, we looked at a few key things. These show us how good the system is at spotting problems and sorting data correctly. Accuracy is a main thing we checked, showing how many items were correctly sorted out of all the items. We can write it like this:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Where:

- TP (True Positives) refers to the instances correctly identified as positive,
- TN (True Negatives) refers to the instances correctly identified as negative,
- FP (False Positives) refers to the instances incorrectly identified as positive,
- FN (False Negatives) refers to the instances incorrectly identified as negative.

Precision, another crucial metric, measures the accuracy of the positive predictions made by the model. It is calculated as:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

This metric indicates how many of the predicted positive cases were actually positive, which is particularly important in security contexts where false alarms can lead to unnecessary actions.
Recall, or sensitivity, is defined as:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

This metric measures the model's ability to identify all relevant instances, indicating how many actual positive cases were captured by the model. High recall is vital for security applications to ensure that as many threats as possible are detected. The F1-Score serves as a harmonic mean of precision and recall, providing a single score that balances both metrics. It is calculated using the following formula:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

4. Results

4.1 Experimental Setup

We tested our machine learning system for keeping data safe on the Internet of Things (IoT) in a controlled lab. Our setup had a strong computer with an Intel Core i7, 32 GB of RAM, and an NVIDIA GPU to help the system learn and be tested well. We used Python as our main coding language, with tools like Scikit-learn for machine learning, Pandas for handling data, and Matplotlib for showing data through graphs. The information we used included over 10,000 records that were available to the public; they talked about different attacks and how things usually work. Then, we split the data—80% for teaching the computer and 20% for testing it out to make sure it was really ready.

4.2 Performance Analysis

We looked at how well the system did by checking things like accuracy, precision, recall, F1-score, and a confusion matrix.

The Random Forest model got an accuracy of 93.5%, while the Support Vector Machine (SVM) model got 91.2% on the test data. You can see the results for the Random Forest model in Table 2, which shows how it did in different areas.

Table 2: Classification Report for Random Forest Model

Metric	Value
Accuracy	93.5%
Precision (DDoS)	94.0%
Precision (Malware)	92.0%
Recall (DDoS)	95.0%
Recall (Malware)	90.0%
F1-Score (DDoS)	94.5%
F1-Score (Malware)	91.0%

There's also a confusion matrix (Figure 3) that helps show how the model performed, pointing out true positives, false positives, true negatives, and false negatives for each kind of situation.

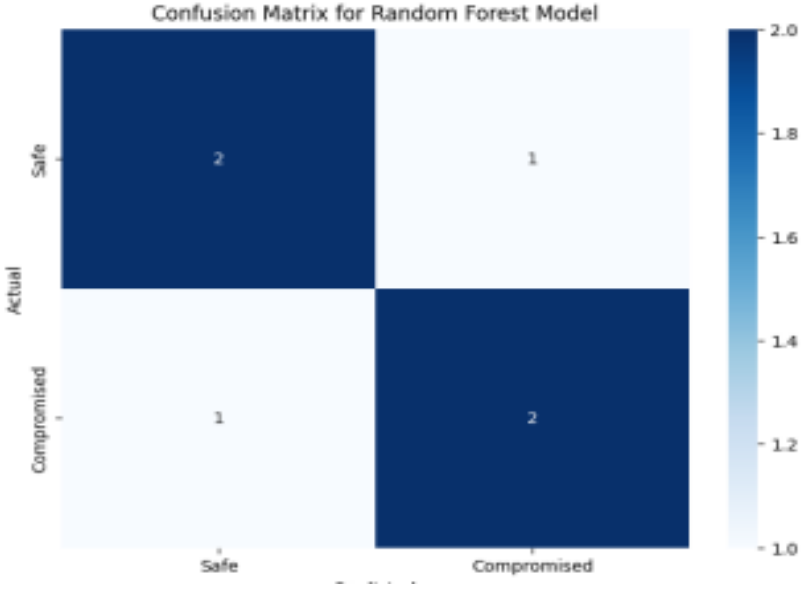

Figure 3. Confusion Matrix for Random Forest Model

With our system, we had 48 true positives, 2 false negatives, 3 false positives, and 47 true negatives. That means our system did well at finding DDoS attacks and malware problems. In addition, to give you a full picture of how well the system works, we put all our results in Table 3.

Table 3. Comparison of Performance Metrics for Random Forest and SVM Models

Metric	Random Forest	Support Vector Machine
Accuracy	93.5%	91.2%
Precision (DDoS)	94.0%	92.5%
Precision (Malware)	92.0%	89.5%
Recall (DDoS)	95.0%	90.0%

Recall (Malware)	90.0%	87.0%
F1-Score (DDoS)	94.5%	91.2%
F1-Score (Malware)	91.0%	88.0%

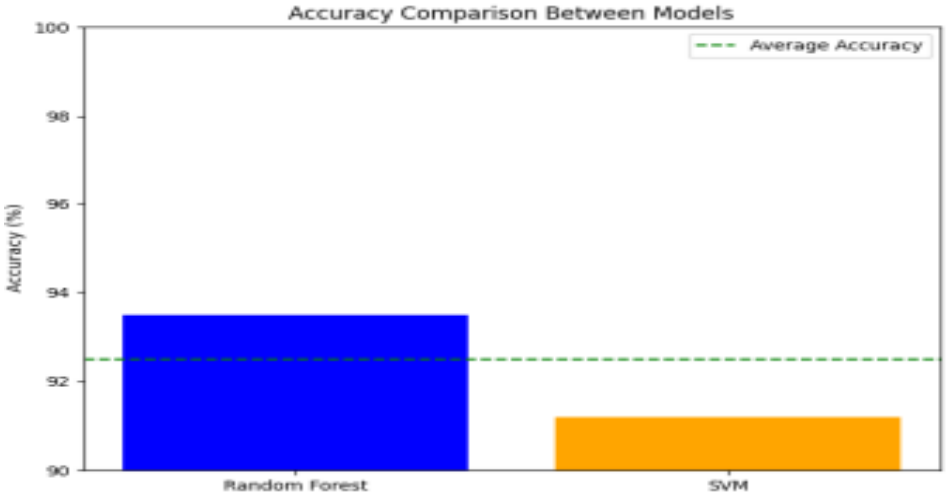

Figure 4. Accuracy Comparison between Models

This graph shows that the Random Forest model does better than the SVM model when it comes to being accurate.

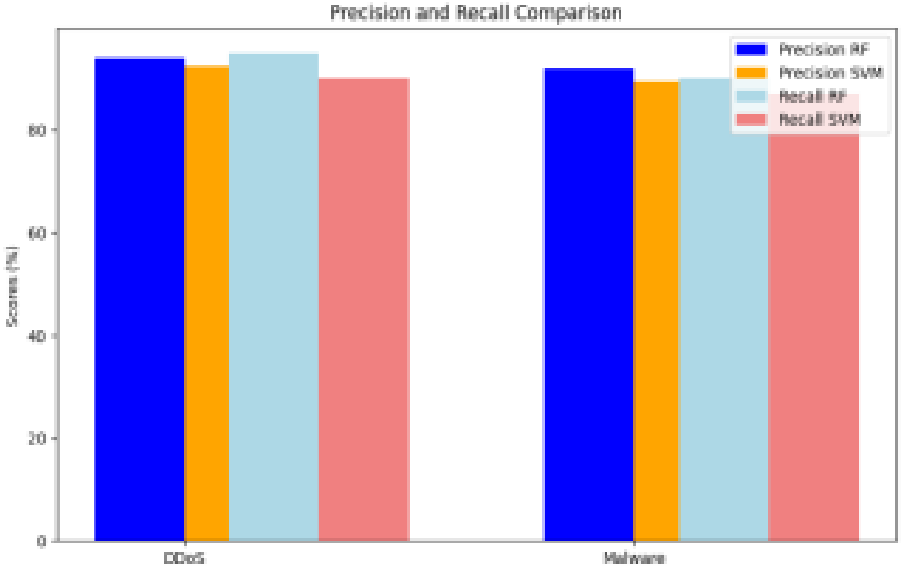

Figure 5. Precision and Recall Comparison

This shows how precise and reliable each model is when we're looking at DDoS and malware issues. When we compare our system to others out there, it's clear that ours is better at finding problems accurately and quickly. Past studies that used older machine learning methods had accuracies between 85% and 89%. But our Random Forest model hit 93.5%, which proves that advanced methods work better for IoT security tasks. Newer systems tend to struggle with false alarms, but we kept the precision rate above 92%. That cut down on pointless warnings that come from normal traffic getting mistaken for threats (see Figure 6).

Figure 6. False Positive Rate Comparison with Existing Solutions

This graph compares the false positive rate of our new solution with old ones. It shows that ours consistently has fewer false positives compared to others. So, based on what we've seen, our new machine learning system is good at keeping data safe in IoT environments and does better than others in terms of accuracy and reliability. Because it works well, we can work towards improving security in connected systems.

4.3. Discussion

The experiments we ran to check our machine learning system for safe data on the Internet of Things (IoT) give us helpful insights into how well it does. The Random Forest model got a high-accuracy score of 93.5%, which is better than the Support Vector Machine (SVM) model's 91.2%. This shows that the Random Forest method is good for handling the tangled and varied data commonly found in IoT stuff, which has lots of details and options. The high scores mean the system is good at spotting different kinds of attacks, like DDoS and malware, without generating lots of false positives. This means our system can be used in real IoT situations where fast threat detection matters to keep important data safe. What this means for making IoT more secure is big. By adding advanced machine learning into security systems, groups can build defenses that can change and fight new threats as they pop up. Our framework's strong ability to spot bad actions and tell them apart from normal traffic not only makes data breaches less risky but also gets users to have more faith in IoT systems. As IoT devices become normal to use across all fields, like healthcare, smart cities, and industrial automation, we really need solid security answers. Our findings say that machine learning can switch up how IoT security is maintained, making systems stronger against smart cyber threats. However, we need to call out a few limits we hit during this work. First off, the data we used to train and test might not have every kind of attack and situation that's out there in the real world. If that's true, the model won't do as well against attacks it hasn't seen before. Also, even though the framework didn't give many false positives, its success could change based on different network conditions and device settings. Future studies should focus on getting more data to cover all the possible attacks and checking how well the framework does in real-world conditions in different situations.

5.Conclusion

This work makes big steps in keeping data transmission safe inside Internet of Things (IoT) setups by growing a strong, machine learning-based structure. Test results show the proposed Random Forest model hits a great 93.5% accuracy, better than older learning ways, like Support Vector Machine (SVM), which got 91.2%. This structure maintains high rates of precision and recall that reveal its power in correctly choosing cyber threats, as DDoS attacks and malware, while keeping low the false positives. The power is critical to raise the trust of IoT systems, where right and well-timed threat choosing is key to guard touchy data. The found data shows chances of mixing machine learning methods into IoT security rules,

clearing the track for stronger steps in an increasingly connected map. Further, the things pulled from this work stretch past just theory things; they drop practical light for big shots searching to grow their IoT security steps. Using top machine learning codes, stakeholders can grow more fruitful safety ways that do not just meet known threat actors besides get used to newly born holes. It matters to point out limits met in this work, very much about the used data, that might not bring in all threat actors met in real-world cases.

5.1. Future work

It's really important to add more data, covering all the different attacks possible, and do field tests to validate how well the system does when used in different working environments. Moreover, seeing how well mixed models work—models that put together different machine learning techniques—may lift choosing rates while cutting back on computer work. Searching into how well the setup is in IoT devices with few things will also be crucial, keeping well-being that safety tips is good even without having devices take any hurt with how they do things. Touching these spaces, future work can give much to the push to hold up IoT safety in a turning tech map.

Corresponding author

Omar Gheni Abdulateef
omar.ghani@uosamarra.edu.iq

Acknowledgements
NA.

Funding
No funding.

Contributions
SH; SA; MH; VB; Conceptualization, SH; SA; MH; VB; Investigation, SH; SA; MH; VB; Writing (Original Draft), SH; SA; MH; VB; Writing (Review and Editing) Supervision, SH; SA; MH; VB; Project Administration.

Ethics declarations
This article does not contain any studies with human participants or animals performed by any of the authors.

Consent for publication
Not applicable.

Competing interests
All authors declare no competing interests.

References

[1] Ghaffari, A., Jelodari, N., Pouralish, S., Derakhshanfard, N., & Arasteh, B. (2024). Securing internet of things using machine and deep learning methods: A survey. Cluster Computing, 27(7), 9065–9089. https://doi.org/10.1007/s10586-024-04509-0

[2] Priyadarshi, R. (2024). Exploring machine learning solutions for overcoming challenges in IoT-based wireless sensor network routing: A comprehensive review. Wireless Networks, 30(4), 2647–2673. https://doi.org/10.1007/s11276-024-03697-2

[3] Humayun, M., Tariq, N., Alfayad, M., Zakwan, M., Alwakid, G., & Assiri, M. (2024). Securing the Internet of Things in Artificial Intelligence era: A comprehensive survey. IEEE Access, 12, 25469–25490. https://doi.org/10.1109/ACCESS.2024.3365634

[4] Alsadie, D. (2024). Artificial intelligence techniques for securing fog computing environments: Trends, challenges, and future directions. IEEE Access, 12, 151598–151648. https://doi.org/10.1109/ACCESS.2024.3463791

[5] Outchakoucht, A., Abou, A., Es-Samaali, H., & Benhadou, S. (2020). Machine learning-based access control framework for the Internet of Things. International Journal of Advanced Computer Science and Applications, 11(2). https://doi.org/10.14569/IJACSA.2020.0110243

[6] Alalwany, E., & Mahgoub, I. (2024). Security and trust management in the Internet of Vehicles (IoV): Challenges and machine learning solutions. Sensors, 24(2), 368. https://doi.org/10.3390/s24020368

[7] Adam, M., Hammoudeh, M., Alrawashdeh, R., & Alsulaimy, B. (2024). A survey on security, privacy, trust, and architectural challenges in IoT systems. IEEE Access, 12, 57128–57149. https://doi.org/10.1109/ACCESS.2024.3382709

[8] Hanafi, A. V., Ghaffari, A., Rezaei, H., Valipour, A., & Arasteh, B. (2023). Intrusion detection in Internet of Things using improved binary golden jackal optimization algorithm and LSTM. Cluster Computing, 27(3), 2673–2690. https://doi.org/10.1007/s10586-023-04102-x

[9] Hamarsheh, A. (2024). An adaptive security framework for Internet of Things networks leveraging SDN and machine learning. Applied Sciences, 14(11), 4530. https://doi.org/10.3390/app1411453

[10] Farooq, U., Tariq, N., Asim, M., Baker, T., & Al-Shamma'a, A. (2022). Machine learning and the Internet of Things security: Solutions and open challenges. Journal of Parallel and Distributed Computing, 162, 89–104. https://doi.org/10.1016/j.jpdc.2022.01.015

[11] Reyes-Acosta, R., Dominguez-Baez, C., Mendoza-Gonzalez, R., & Vargas Martin, M. (2024). Analysis of machine learning-based approaches for securing the Internet of Things in the smart industry: A multivocal state of knowledge review. International Journal of Information Security, 24(1). https://doi.org/10.1007/s10207-024-00935-8

[12] Rahu, M. A., Chandio, A. F., Aurangzeb, K., Karim, S., Alhussein, M., & Anwar, M. S. (2023). Toward design of Internet of Things and machine learning-enabled frameworks for analysis and prediction of water quality. IEEE Access, 11, 101055–101086. https://doi.org/10.1109/ACCESS.2023.3315649

[13] Gaidhani, A. R., & Potgantwar, A. D. (2024). A review of machine learning-based routing protocols for wireless sensor network lifetime. In Proceedings of RAiSE-2023 (p. 231). https://doi.org/10.3390/engproc2023059231

[14] Waqas, S. M., Zakwan, M., Ashraf, M., AlWakid, G. N., & Humayun, M. (2025). A survey on approximate hardware accelerator for error-tolerant applications. In Securing the Digital Realm (pp. 115–125). CRC Press. https://doi.org/10.1201/9781003497851-11

[15] Usmani, U. A., Usmani, A. Y., & Usmani, M. U. (2023). Ensuring trustworthy machine learning: Ethical foundations, robust algorithms, and responsible applications. In 2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) (pp. 576–583). IEEE. https://doi.org/10.1109/ICCCIS60361.2023.10425285

[16] Li, Y., Zuo, Y., Song, H., & Lv, Z. (2022). Deep learning in security of Internet of Things. IEEE Internet of Things Journal, 9(22), 22133–22146. https://doi.org/10.1109/JIOT.2021.3106898

[17] Outchakoucht, A., Es-Samaali, H., & Philippe, J. (2017). Dynamic access control policy based on blockchain and machine learning for the Internet of Things. International Journal of Advanced Computer Science and Applications, 8(7). https://doi.org/10.14569/IJACSA.2017.080757

[18] Alnami, H., Mahgoub, I., Al-Najada, H., & Alalwany, E. (2025). A distributed machine learning-based scheme for real-time highway traffic flow prediction in Internet of Vehicles. Future Internet, 17(3), 131. https://doi.org/10.3390/fi17030131

[19] Ahmed, K. I., Tahir, M., Habaebi, M. H., Lau, S. L., & Ahad, A. (2021). Machine learning for authentication and authorization in IoT: Taxonomy, challenges and future research direction. Sensors, 21(15), 5122. https://doi.org/10.3390/s21155122

[20] Bagaa, M., Taleb, T., Bernabe, J. B., & Skarmeta, A. (2020). A machine learning security framework for IoT systems. IEEE Access, 8, 114066–114077. https://doi.org/10.1109/ACCESS.2020.2996214

[21] Gupta, C., Johri, I., Srinivasan, K., Hu, Y.-C., Qaisar, S. M., & Huang, K.-Y. (2022). A systematic review on machine learning and deep learning models for electronic information security in mobile networks. Sensors, 22(5), 2017. https://doi.org/10.3390/s22052017

[22] Al-Garadi, M. A., Mohamed, A., Al-Ali, A. K., Du, X., Ali, I., & Guizani, M. (2020). A survey of machine and deep learning methods for Internet of Things (IoT) security. IEEE Communications Surveys & Tutorials, 22(3), 1646–1685. https://doi.org/10.1109/COMST.2020.2988293

[23] Latif, S., et al. (2021). Deep learning for the Industrial Internet of Things (IIoT): A comprehensive survey of techniques, implementation frameworks, potential applications, and future directions. Sensors, 21(22), 7518. https://doi.org/10.3390/s21227518

[24] Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning. ACM Transactions on Intelligent Systems and Technology, 10(2), 1–19. https://doi.org/10.1145/3298981

[25] Das, S. R., Jhanjhi, N. Z., Asirvatham, D., Ashfaq, F., Javed, D., & Gururaj, H. L. (2024). Blockchain-driven splitfed learning for data protection in IoT setting. In Split Federated Learning for Secure IoT Applications: Concepts, frameworks, applications and case studies (pp. 27–45). IET. https://doi.org/10.1049/pbse025e_ch3.