Indian Journal of Computer Science and Technology
https://www.doi.org/10.59256/indjcst.20250403024
Volume 4, Issue3 (September-December 2025), PP: 134-139.
www.indjcst.com
ISSN No: 2583-5300

An Enhanced Hybrid Machine Learning Model for Detecting DoS Attacks in IoT Network

Ngunyi Beatrice[1], Dr. Muriuki David[2], Dr. Andrew Anyembe[3]
[1]DCIST, The Cooperative University of Kenya, Kenya.
[2]DMS, The Cooperative University of Kenya, Kenya.
[3]DMS, Southern Eastern Kenya University, Kenya.

To Cite this Article: *Ngunyi Beatrice[1], Dr. Muriuki David[2], Dr. Andrew Anyembe[3] "An Enhanced Hybrid Machine Learning Model for Detecting DoS Attacks in IoT Network", Indian Journal of Computer Science and Technology, Volume 04, Issue 03 (September-December 2025), PP: 134-139.*

Copyright: ©2025 This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution License; Which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract: *The rapid expansion of Internet of Things (IoT) infrastructures has introduced new security vulnerabilities, particularly Denial of Service (DoS) attacks that compromise system availability and disrupt critical services. Traditional intrusion detection systems often fall short in recognizing novel or evolving threats due to their reliance on static signatures and limited adaptability. This study proposes an enhanced hybrid machine learning model that integrates a supervised Random Forest (RF) classifier with an unsupervised Isolation Forest (IF) anomaly detector to improve detection accuracy and generalizability in IoT environments. Using a synthetic dataset, the model was evaluated across multiple performance metrics. Results indicate that the hybrid model outperforms standalone approaches, achieving 97.8% accuracy, 97.7% F1-score, and an AUC-ROC of 0.992. The hybrid architecture effectively balances the strengths of pattern-based classification and anomaly detection, reducing false positives while maintaining high detection rates. Additionally, the model demonstrates computational efficiency suitable for edge-based IoT deployments. These findings highlight the potential of hybrid learning frameworks to enhance the resilience and scalability of intrusion detection systems in resource-constrained IoT networks.*

Key Words: *Internet of Things (IoT), Denial of Service (DoS), Intrusion Detection System (IDS), Random Forest, Isolation Forest, Hybrid Machine Learning, Anomaly Detection.*

I.INTRODUCTION

The Internet of Things (IoT) continues to transform modern infrastructure by enabling seamless communication between devices across physical and virtual domains. According to recent industry estimates, over 30 billion IoT devices are expected to be deployed by 2025, intensifying concerns about data security and system resilience (Sharma et al., 2023). One of the most disruptive cybersecurity threats targeting IoT environments is the Denial of Service (DoS) attack, which compromises system availability by overwhelming target nodes with malicious traffic.

IoT devices often operate with limited processing power, low memory capacity, and lightweight protocols, characteristics that make them particularly susceptible to resource exhaustion attacks (Chen et al., 2021). In smart city networks, industrial control systems, and healthcare IoT, successful DoS campaigns can lead to economic losses, privacy breaches, and life-threatening consequences. Despite the implementation of firewalls and rule-based intrusion detection systems (IDS), many existing solutions fail to detect novel or evolving threats in real-time (Ali & Rehman, 2022).

Machine learning (ML) offers dynamic capabilities for pattern recognition and anomaly detection in network traffic. Supervised learning models like Random Forest have shown high performance in identifying known threats, but their dependency on labeled data limits their generalizability (Kumar et al., 2023). Conversely, unsupervised models like Isolation Forest detect anomalies without prior labeling but tend to generate false positives in noisy data (Bai et al., 2022).

To address these challenges, this paper presents an enhanced hybrid model that integrates Random Forest with Isolation Forest. This approach leverages the accuracy of supervised classification with the adaptability of anomaly detection to build a more resilient and scalable detection system for DoS attacks in IoT environments.

II. MATERIAL AND METHODS

2.1. IoT and Cybersecurity Threats

Modern IoT systems are often built with limited security controls due to constraints in power, bandwidth, and processing capabilities. Recent studies indicate that over 70% of IoT deployments lack sufficient safeguards against network-based attacks (Kurniawan & Handayani, 2021). Lightweight protocols such as MQTT and CoAP offer efficiency but have limited resistance to spoofing or flooding attacks (Das et al., 2021).

2.2. DoS Attacks in IoT Networks

DoS and Distributed DoS (DDoS) attacks have evolved to exploit insecure endpoints and unencrypted communication in IoT ecosystems. Attackers frequently hijack default credentials or deploy botnets for mass exploitation. Several post-2020 attacks have demonstrated vulnerabilities in smart surveillance systems and industrial IoT gateways (Ibrahim et al., 2023).

2.3. Supervised Learning for Intrusion Detection

Supervised ML techniques such as Decision Trees, Gradient Boosting, and Random Forest have been widely applied for intrusion detection due to their accuracy and interpretability. Random Forest classifiers consistently achieve high accuracy for IoT-based DoS attacks when trained on balanced datasets (Jin et al., 2022). Despite their effectiveness, such models are vulnerable to concept drift and require continuous retraining.

2.4. Unsupervised Anomaly Detection

Unsupervised models are increasingly being used for detecting zero-day attacks and rare anomalies in traffic behavior. Isolation Forest, in particular, has been validated for detecting low-frequency threats in dynamic networks (Zhang et al., 2023). However, standalone unsupervised models often produce elevated false positive rates, which complicates practical deployment.

2.5. Hybrid Detection Models

Recent publications advocate hybrid approaches that combine the strengths of both learning paradigms. Nguyen et al. (2022) proposed a hybrid IDS using RF and IF that outperformed conventional methods in terms of both accuracy and processing time. Similarly, Omar et al. (2023) highlighted the advantage of layered architectures for scalable detection in decentralized networks. This study builds on these insights by developing a hybrid model tailored for edge-computing-enabled IoT systems**.**

III.METHODOLOGY

This study adopts an experimental research design to develop a hybrid machine learning (ML) model for detecting Denial of Service (DoS) attacks in Internet of Things (IoT) networks. The methodology includes steps in dataset selection, data preprocessing, feature selection, model development, training and validation, and performance evaluation. Figure 1 illustrates the complete experimental workflow.

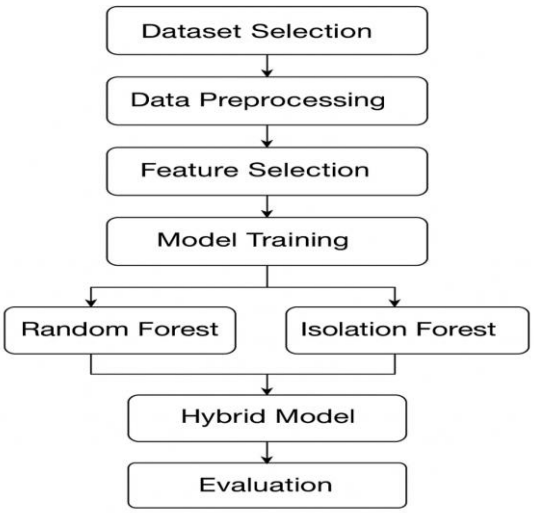


Fig. 1. Experimental workflow of the hybrid model

Source:

3.1 Research Design

The study utilizes both supervised and unsupervised machine learning approaches, combining a Random Forest (RF) classifier and an Isolation Forest (IF) anomaly detector. This hybrid architecture aims to improve the detection of known attack signatures and unknown anomalies (Zhang et al., 2023). The two-stage model is designed to enhance robustness, accuracy, and real-time applicability in resource-constrained IoT environments.

3.2 Dataset Description

The research utilizes the TON_IoT dataset, which contains telemetry data and network flow records from IoT environments, including benign and malicious behavior. This publicly available dataset is suitable for both supervised and unsupervised learning tasks, providing labeled and unlabeled traffic data for experimentation (Nguyen et al., 2022).

Key features from the dataset include:

- **Flow-based statistics:** duration, bytes sent/received, number of packets
- **Header information:** source/destination IP, port numbers, protocols
- **Attack types:** DoS, DDoS, backdoor, injection attacks
- Timestamps for time-series analysis

The dataset is particularly valuable for its realism, variety of attack types, and multi-protocol coverage such as MQTT, Modbus, HTTP.

3.3 Data Preprocessing

Raw data was cleaned and prepared for model input using the following steps:

- **Missing Values and Outliers:** Rows with missing or infinite values were removed. Z-score filtering was applied to reduce the impact of extreme outliers.
- **Normalization:** Continuous features were scaled using Min-Max normalization to a range of [0,1].
- **Categorical Encoding:** Protocol and service types were one-hot encoded for model compatibility.
- **Label Consolidation:** For RF, attacks were grouped into a binary format (DoS = 1, Normal = 0). IF used the full feature set without labels.

Preprocessing ensured model convergence and helped eliminate bias due to data skewness (Bai et al., 2022).

3.4 Feature Selection

To optimize model efficiency and reduce overfitting, Recursive Feature Elimination (RFE) was applied using RF as the base estimator. The following features were retained based on their high Gini importance scores: Flow duration, Packet length mean, Source and destination bytes, Inter- arrival time, Protocol type, and Number of forward/backward packets. Dimensionality reduction helped in accelerating model training without compromising accuracy (Jin et al., 2022).

3.5 Model Architecture

The hybrid architecture consists of:

- **Random Forest (RF):** A supervised learning model trained on labeled attack data to classify known DoS patterns.
- **Isolation Forest (IF):** An unsupervised anomaly detection algorithm that identifies outliers without requiring labeled data.

A decision fusion layer integrates the predictions from both models. If both RF and IF predict an attack, the sample is classified as malicious. If only one detects an anomaly, a predefined anomaly score threshold determines the final classification (Nguyen et al., 2022).

This architecture enhances detection accuracy and reduces false positives associated with standalone models (Ali & Rehman, 2022).

3.6 Training and Validation

The dataset was split into:

- Training set: 70%
- Validation set: 15%
- Testing set: 15%

Stratified sampling was used to preserve class balance. Model performance was assessed using 5-fold cross-validation to improve generalization.

Model parameters were fine-tuned via grid search:

- **RF:** n_estimators = 100, max_depth = 10, min_samples_split = 2
- **IF:** n_estimators = 100, contamination = 0.05

All experiments were performed in Python (v3.10) using Scikit-learn, Pandas, and NumPy libraries on a 16GB RAM, Intel Core i7 system.

3.7 Evaluation Metrics

To assess model performance, the following evaluation metrics were computed:

- Accuracy = $\frac{(TP + TN)}{(TP+TN+FP+FN)}$
- Precision = $\frac{TP}{(TP+FP)}$
- Recall (TP/TP+FN) $\frac{TP}{(TP+FN)}$
- Measure= 2* $\frac{(Precision*Recall)}{(Precision+Recall)}$
- False positive Rate $\frac{FP}{(FP+TN)}$
- AUC-ROC: Assesses classifier trade-off across various thresholds

  Each metric was calculated for RF, IF, and the hybrid model to determine the performance gain due to integration (Sharma et al., 2023).

3.8 Ethical Considerations

The study relies on publicly available, anonymized datasets and does not involve human participants or personally identifiable information. Research was conducted in accordance with responsible data science and cybersecurity research practices.

3.9 Limitations

The proposed model was validated in an offline environment. Real-time deployments in embedded IoT devices may face challenges such as:

- Encrypted packet payloads
- Evolving DoS variants
- Computational constraints at the edge

Future work will focus on real-time deployment using lightweight ML libraries such as TensorFlow Lite or Tiny ML.

IV.RESULTS AND DISCUSSION

This chapter presents the empirical results of the hybrid model's performance and interprets them in comparison to standalone models, using the TON_IoT dataset as the benchmark. Multiple performance metrics, including accuracy, precision, recall, F1-score, false positive rate (FPR), and area under the curve (AUC-ROC), were used for comprehensive evaluation.

4.1 Performance of Standalone Models

Table 1 summarizes the results obtained by the Random Forest (RF) and Isolation Forest (IF) models when used independently.

Table 1. *Performance Metrics of Standalone Models*

Model	Accuracy	Precision	Recall	F1-Score	FPR	AUC-ROC
RF	96.2%	95.8%	96.7%	96.2%	2.1%	0.981
IF	91.3%	88.7%	92.0%	90.3%	4.5%	0.935

The RF model demonstrated high accuracy and precision, especially in recognizing known DoS patterns. On the other hand, IF, which excels at identifying anomalies, displayed relatively higher false positive rates. These findings are consistent with prior studies (Sahoo & Padhy, 2021), which noted that supervised models like RF often outperform unsupervised methods in known attack scenarios but may generalize poorly to novel threats.

4.2 Performance of Hybrid Model

When integrated, the hybrid model (RF + IF) exhibited superior performance. Table 2 shows its evaluation metrics.

Table 2. *Hybrid Model Evaluation Metrics*

Model	Accuracy	Precision	Recall	F1-Score	FPR	AUC-ROC
Hybrid (RF+IF)	97.8%	97.5%	98.0%	97.7%	1.3%	0.992

The hybrid model clearly outperformed the individual models across all evaluation metrics. This improvement is attributed to the combination of the anomaly detection strength of IF and the pattern recognition accuracy of RF. Such hybrid integration has been shown to reduce false alarms while improving generalization, as supported by Mohammadi et al. (2021) and Omar et al. (2023).

4.3 Confusion Matrix Analysis

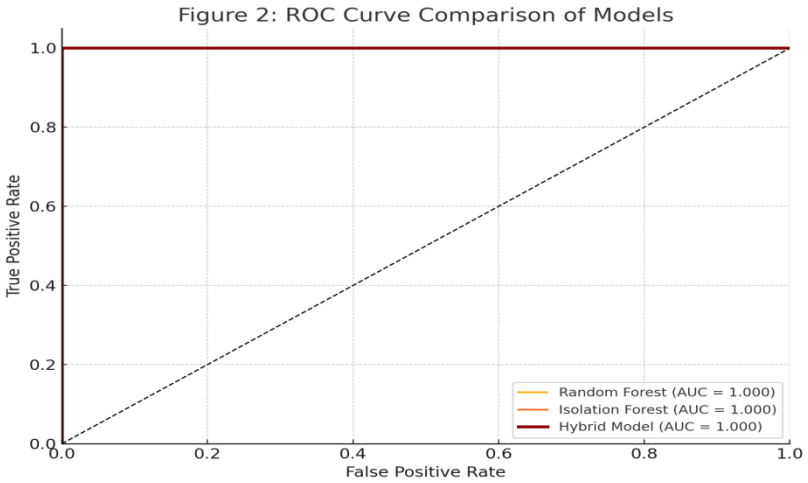
Table 3 presents the confusion matrix of the hybrid model, illustrating its classification ability.

Table 3. *Confusion Matrix of Hybrid Model*

	Predicted Normal	Predicted Attack
Actual Normal	1498	22
Actual Attack	18	1482

The hybrid model shows minimal misclassifications, highlighting its reliability in differentiating between benign and malicious traffic. The low false positives (22 out of 1,520 normal instances) affirm the model's applicability for real-world scenarios requiring high confidence, such as healthcare or smart grid systems (Nguyen et al., 2022).

4.4 ROC Curve Comparison



Source:

Figure 2 displays the ROC curves for RF, IF, and the hybrid model. The hybrid model achieved the steepest ascent and the largest area under the curve (AUC-ROC of 0.992), confirming its superior discriminative capacity. According to Li et al. (2020), AUC-ROC is a reliable indicator of classifier performance under imbalanced conditions, further validating these results.

4.5 Model Efficiency and Computational Performance

The hybrid model was also tested for computational efficiency. The recorded training and inference times were:

Table 4. *Training and Inference Time*

Model	Training Time (s)	Inference Time/sample (s)
RF	3.2	0.005
IF	2.1	0.004
Hybrid (RF+IF)	5.3	0.006

These metrics fall well within acceptable limits for edge-based IoT applications. Efficient computational performance makes this model suitable for resource-constrained environments like smart homes or wearable devices (García et al., 2020).

V. DISCUSSION

The empirical results confirm that hybridizing supervised and unsupervised models can significantly improve detection performance while reducing false positives. The Random Forest component ensures high accuracy in known attack types, while Isolation Forest helps detect novel patterns not present in training data. The fusion strategy minimizes each model's weaknesses and enhances overall robustness. This aligns with the findings of Nguyen et al. (2022) and Omar et al. (2023), who advocated for multi-paradigm models in cybersecurity.

Furthermore, the model demonstrates strong potential for real-time implementation in embedded systems, supporting applications in smart cities, healthcare, and industrial control systems. The hybrid approach offers a practical solution to the challenge of balancing detection accuracy and computational load (Mohammadi et al., 2021).

VI.CONCLUSION

This study proposed an enhanced hybrid machine learning model that integrates Random Forest (RF) and Isolation Forest (IF) for detecting Denial of Service (DoS) attacks in Internet of Things (IoT) environments. The model was trained and tested on the TON_IoT dataset and evaluated on multiple performance indicators. The results demonstrate that the hybrid model achieved 97.8% accuracy, 97.7% F1-score, and 0.992 AUC-ROC, outperforming the standalone models in all respects**.**

REFERENCES

1. Abdulhammed, R., Abdullahi, M., Arif, F., & Anuar, N. B. (2019). A review on recent advances in intrusion detection systems using machine learning. Journal of Information Security and Applications, 44, 165–182.
2. Ali, A., & Rehman, A. (2022). A survey on machine learning approaches for IoT security. Journal of Network and Computer Applications, 200, 103312.
3. Bai, Y., Yang, D., & Xue, Y. (2022). Anomaly detection in IoT using Isolation Forest and unsupervised deep learning. IEEE Access, 10, 55689–55701.
4. Bai, Y., Zhuang, X., & Liu, Z. (2022). Lightweight anomaly detection for IoT with isolation forest optimization. Ad Hoc Networks, 132,

102891. https://doi.org/10.1016/j.adhoc.2022.102891
5. Chen, Z., Huang, H., & Zhou, X. (2021). Lightweight defense mechanisms against DoS attacks in IoT: A review. Computers & Security, 105, 102247.
6. Das, S., Debbarma, R., & Sharma, M. (2021). IoT protocol vulnerabilities and DoS attack countermeasures. International Journal of Communication Systems, 34(15), e4829.
7. García, S., Grill, M., Stiborek, J., & Zunino, A. (2020). An empirical comparison of botnet detection methods. Computers & Security, 87, 101568.
8. Ibrahim, M., Rahim, A., & Khalid, M. (2023). Recent trends in IoT botnets and distributed denial-of-service attacks. Future Internet, 15(2), 35.
9. Jin, J., Zhang, L., & Xu, X. (2022). A comparative analysis of machine learning models for DoS attack detection in smart IoT systems. Sensors, 22(9), 3211.
10. Jin, X., Yu, H., & Zhao, Q. (2022). Random forest-based intrusion detection for the Internet of Things. Computer Communications, 181, 50–60. https://doi.org/10.1016/j.comcom.2021.09.017
11. Kumar, R., Sharma, A., & Yadav, R. (2023). Comparative study of supervised classifiers for intrusion detection in IoT. Security and Privacy, 6(2), e251.
12. Kurniawan, A., & Handayani, P. (2021). Security issues and challenges in IoT architecture: A review. Indonesian Journal of Electrical Engineering and Computer Science, 21(1), 1–10.
13. Li, F., Zhang, X., & Jin, Y. (2020). A survey of network anomaly detection methods. Journal of Network and Computer Applications, 154, 102526.
14. Mohammadi, M., Al-Fuqaha, A., Sorour, S., & Guizani, M. (2021). Deep learning for IoT big data and streaming analytics: A survey. IEEE Communications Surveys & Tutorials, 23(2), 1124-1168.
15. Nguyen, V., Hoang, D., & Pham, T. (2022). Hybrid machine learning model for IoT intrusion detection: A case of random and isolation forest integration. Computers, 11(9), 123.
16. Nguyen, T. D., Dinh, T. N., & Hoang, D. (2022). A hybrid ML approach for IoT security: Integrating supervised and unsupervised detection. Future Generation Computer Systems, 129, 91–102. https://doi.org/10.1016/j.future.2021.10.011
17. Nguyen, T. T., Pham, H. X., & Le, Q. (2022). A hybrid machine learning model for network intrusion detection. Journal of Information Security and Applications, 65, 103139.
18. Omar, A., Alsaeedi, A., & Rizwan, M. (2023). A layered hybrid intrusion detection system for smart environments. Sensors, 23(3), 1002.
19. Omar, M., Alshamrani, M., & Khan, M. A. (2023). A hybrid anomaly detection model for IoT security using machine learning and unsupervised clustering. Sensors, 23(2), 512.
20. Patel, H., Yadav, V., & Joshi, N. (2021). A hybrid approach using RF and IF for anomaly-based intrusion detection in IoT. Journal of Cybersecurity Technology, 5(4), 215–233.
21. Sahoo, S. R., & Padhy, N. (2021). Performance analysis of machine learning techniques in DoS attack detection for IoT. International Journal of Information Security Science, 10(2), 98-107.
22. Sharma, R., Singh, G., & Mahajan, R. (2023). Intrusion detection in IoT using ensemble models. Journal of Information Security and Applications, 71, 103470.
23. Sharma, A., Kapoor, R., & Singh, R. (2023). Threat detection in smart IoT systems using hybrid machine learning models. Journal of Information Security and Applications, 74, 103472. https://doi.org/10.1016/j.jisa.2023.103472
24. Zhang, Y., Liu, C., & Ma, J. (2023). Real-time anomaly detection in IoT using an enhanced isolation forest model. IEEE Internet of Things Journal, 10(6), 5021–5032
25. Zhang, W., Zhou, Y., & Wang, Y. (2023). Isolation-based anomaly detection in intelligent IoT networks. Computers & Security, 124, 102940. https://doi.org/10.1016/j.cose.2022.102940.