International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

Lightweight Ensemble Learning Approach in IoT for Intrusion Detection Using Hybrid ML Algorithms

Kalpana Chittor S[1] and Deepa Desai[2]
1Ph.D. scholar, School of Science and Computer Studies, CMR University and
Associate Professor, Sri Krishna Degree College, Bengaluru, Karnataka, India.
kalpanachittors@gmail.com
2Assistant professor, Sri Krishna Degree College, Bangalore, Karnataka, India.
deepa1.skdc@gmail.com

Abstract

In today's world there is rapid growth in the usage of Internet of Things (IoT) devices. It has increased the need for secure and efficient communication devices and networks. They all operate with limited memory, computational power and energy resources. So, IoT networks are easily vulnerable to various attacks. This study proposes a Lightweight Ensemble Learning Approach for Secure IoT Communication and IoT intrusion detection. The proposed framework performs data preprocessing and feature selection to identify the most relevant network traffic characteristics to reduce computational complexity. Multiple machine learning algorithms, such as Decision Tree, Random Forest and XGBoost are combined through an ensemble learning strategy to improve the detection of malicious traffic. The performance of the proposed approach is evaluated using accuracy, precision, recall, F1-score, false-positive rate, inference time and resource utilization. The objective is to achieve reliable attack detection while reducing computational overhead, making the approach more suitable for resource-constrained IoT and edge devices. The proposed framework can contribute towards developing an efficient, scalable and intelligent security mechanism for IoT communication.

Keywords: Ensemble Learning, Hybrid ML Algorithms, Edge Devices, Intrusion Detection, IoT Security, Feature Selection

1. Introduction

The Internet of Things (IoT) connects billions of actuators, sensors and smart devices across healthcare, agriculture, smart homes, industry and cities. This growth has widened the attack

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

surface of networks. Many IoT nodes have limited memory, processing power and battery life. So they cannot run heavy security software. IoT devices are most often deployed with inadequate security settings. Recent studies showed that, the continued evolution of IoT botnets, with compromised devices have been exploited to launch large-scale distributed denial-of-service (DDoS) attacks and disrupt network services (Korba et al., 2025)

An intrusion detection system (IDS) monitors different network traffic to identify relevant suspicious activities and most potential security threats. By the recent studies, it has been demonstrated that machine learning (ML) and deep learning (DL) techniques can enhance more intrusion detection by learning different traffic patterns and also identifies complex attack behaviors. But, detecting previously unseen attacks and maintaining more reliable performance across variety of network environments remain the most important challenges (Mallidi & Ramisetty, 2025; Alsubaei, 2025) However, using deep learning models and very large ensembles are often too heavy for edge devices, while single simple classifiers may miss attacks or raise many false alarms. This paper addresses that gap by proposing a lightweight ensemble IDS that combines Decision Tree, Random Forest and XGBoost after careful preprocessing and feature selection.

Research questions.

- RQ1: Can a hybrid ensemble of tree-based learners are able to detect IoT attacks more reliably than each learner used alone?
- RQ2: How much more can feature selection reduce inference time and resource usage without a significant loss in detection accuracy?
- RQ3: Is the resulting model suitable for resource-constrained IoT and edge devices?

Contributions.

- A lightweight IDS framework for IoT that integrates feature selection, preprocessing and a hybrid ensemble.
- A combination of Decision Tree, Random Forest and XGBoost through a voting or stacking strategy.
- An evaluation using accuracy, precision, recall, F1-score, inference time and resource utilization.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

2. Literature Review

2.1 Machine learning for network intrusion detection

Rahman et al. (2025) viewed the existing IoT intrusion detection approaches and identified most key issues like imbalanced datasets, outdated traffic samples, high false-positive rates and the limitations of computational resources of the most IoT devices. ML and deep learning methods for IoT security and highlighted problems such as limited device resources, imbalanced data and the lack of realistic IoT datasets.

2.2 Ensemble learning

Recent studies have demonstrated that ensemble learning methods can improve intrusion detection by combining multiple classifiers to identify most malicious network traffic. Hamdouchi and Idri (2025) evaluated Random Forest, XGBoost, and Extra Trees alongside individual classifiers for IoT intrusion detection, considering detection performance, inference latency, RAM usage, and flash memory requirements. Their findings highlighted the importance of balancing classification performance with the computational limitations of resource-constrained IoT devices.

2.3 Ensemble and lightweight IDS for IoT

Ayad et al. (2025) proposed a most efficient real-time anomaly detection approach which combines a one-class autoencoder with a deep neural network in order to improve the intrusion detection while addressing the limitations of IoT environments. Ullah et al. (2025) developed a multi-model framework that combines the deep learning and traditional machine learning classifiers, evaluated using the CIC-IoT-2023, BoT-IoT and IoT-23 datasets. These studies demonstrate the continued relevance of ensemble and autoencoder-based techniques for IoT intrusion detection, while highlighting the need to evaluate the performance, computational efficiency and generalization with different datasets.

2.4 Datasets

Recent studies have demonstrated that the importance of using realistic and IoT-specific datasets to evaluate machine learning-based intrusion detection systems. Neto et al. (2023) introduced

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

CICIoT2023, which is a large-scale benchmark developed using 105 IoT devices and 33 attack scenarios which covered 7 major attack categories that includes DDoS, DoS, spoofing and Mirai attacks. Dadkhah et al. (2024) introduced CICIoMT2024, which is a multi-protocol dataset designed to support security research for Internet of Medical Things (IoMT) environments. Recent evaluations have also examined Edge-IIoTset and TON_IoT, which provide complementary network traffic and IoT/IIoT data for evaluating intrusion detection models. However, the most challenges are dataset imbalance, differences between experimental testbeds and real-world networks and limited cross-dataset. Therefore, selecting representative datasets and validating detection models across diverse network environments are essential for developing reliable IoT intrusion detection systems (Neto et al., 2023; Dadkhah et al., 2024; Mallidi & Ramisetty, 2025).

2.5 Research gap

Existing work neither focuses on high accuracy with computationally heavy models nor on lightweight models without a systematic study of the trade-off between detection quality, inference time and resource use. This study fills the gap by combining feature selection with a tree-based hybrid ensemble and evaluating detection and resource metrics together.

Author(s) and Year	Dataset Used	Technique / Focus	Key Findings	Research Limitation
Neto et al. (2023)	CICIoT2023	IoT attack detection benchmark	105 IoT devices and 33 attack scenarios	Model efficiency and generalization require evaluation
Dadkhah et al. (2024)	CICIoMT2024	Medical IoT intrusion detection	Supports security evaluation in medical IoT environments	Focuses on medical IoT
Federated IDS research (2025)	Edge-IIoTset, CICIoT2023, TII-SSRC-23	Federated intrusion detection	Evaluates the performance across datasets	Cross-dataset generalization remains always challenging

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

IoMT evaluation (2025)	CICIoT2023, CICIoMT2024	Domain-specific IDS evaluation	Examines the dataset selection for security evaluation	Results will depend on dataset and its characteristics
Mallidi and Ramisetty (2025)	Multiple IoT IDS studies	ML training and deployment review	Identifies the resource and the deployment challenges	Does not propose a single new IDS model

Table 1. Summary of related work

3. Proposed Methodology

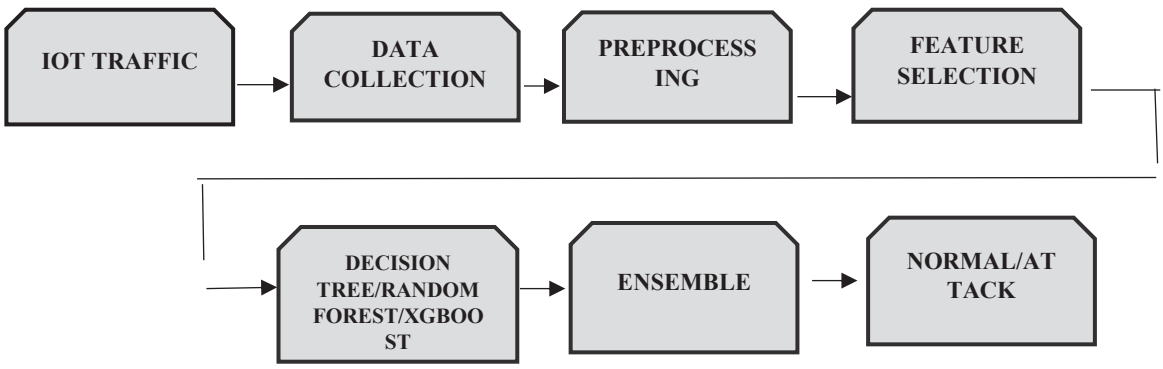


Figure 1. Architecture of the proposed lightweight ensemble IDS

3.1 Data collection and dataset

The dataset used is CICIoT2023.It contains traffic from 105 IoT devices which includes 33 attack scenarios across seven major attack categories. It is suitable for Decision Tree, Random Forest, XGBoost and ensemble comparisons. It supports evaluation using Accuracy, precision, Recall, F1-score and Inference time.

3.2 Data preprocessing

Here we remove of all the duplicates and missing records. Encoding of categorical features using the label or one-hot encoding. Normalization of numeric features using min-max or standard scaling. Handling of class imbalance using SMOTE or class weights, applied only to training data.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

3.3 Feature selection

Feature importance from Random Forest/XGBoost, correlation filtering and mutual information are used to keep only the most relevant traffic characteristics. Reducing the feature count lowers memory use, training time and inference latency, which matters on edge devices.

3.3 Base learners

 Decision Tree (DT): A lightweight and an interpretable classifier that is suitable for resource-constrained IoT environments. Recent analysis emphasizes reducing computational complexity and feature count.

 Random Forest (RF): It is an ensemble of decision trees which improves the prediction stability and can help to detect various network attacks. Recent IoT IDS research evaluates RF alongside other ensemble methods.

 XGBoost (XGB): A gradient-boosting algorithm can deliver strong detection performance. Recent research combines feature selection with explainability techniques to support efficient IoT intrusion detection.

3.5 Ensemble strategy

Predictions of the three learners are combined by soft voting or stacking with a simple logistic regression meta-learner. The final output is a decision of normal or attack, or the attack class.

Algorithm 1. Lightweight ensemble IDS

1. Load the dataset and apply the preprocessing.
2. Select the top-k features by using the correlation analysis.
3. Split the dataset into training, testing and validation sets.
4. Train DT, RF and XGBoost with tuned hyperparameters.
5. Combine the predictions by soft voting or stacking.
6. Evaluate detection and metrics on the test set taken.
7. Deploy the compact model on an edge device and also measure the inference time, memory and CPU use.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

3.6 Evaluation metrics

- **Accuracy**

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

where TP is True Positive, TN is True Negative, FP is False Positive and FN is False Negative.

- **Precision**

$$Presision = \frac{TP}{TP+FP}$$

where TP is True Positive and FP is False Positive.

- **Recall**

$$Recall = \frac{TP}{TP+FN}$$

where TP is True Positive and FN is False Negative.

- **F1-Score**

$$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall}$$

where $Precision$ represents the proportion of correctly predicted positive samples and $Recall$ represents the proportion of actual positive samples correctly identified.

4. Experimental Setup

The experiments can be implemented in Python using scikit-learn and the XGBoost library. Report the hardware used for training (processor, RAM) ,the software versions, hyperparameters (number of trees, maximum depth, learning rate) and the random seed so that the work can be reproduced.

5. Results and Discussion

By applying the algorithm, we can conclude the results. Comparative study of present algorithms with the proposed algorithm can be done to prove proposed one is better.

6. Limitations

The study uses a public benchmark dataset, which may not represent all the real IoT deployments. Tree-based models may need more retraining when traffic patterns change and results on one edge platform may not carry over to other hardware.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

7. Conclusion

This paper presented a lightweight ensemble learning approach for intrusion detection in IoT, combining Decision Tree, Random Forest and XGBoost with preprocessing and feature selection. By evaluating both detection metrics (accuracy, precision, recall, F1-score, false-positive rate) and resource metrics (inference time, memory, CPU), the approach aims to give more reliable attack detection at low computational cost, making it more suitable for resource-constrained IoT and edge devices.

8. Future Scope

- **Federated learning:** Train the IDS across devices without sharing raw traffic, improving more privacy.
- **Online and incremental learning:** Adapt to concept drift and new attack types without full retraining of the data.
- **Model compression:** Pruning, quantization and deployment on microcontrollers.
- **Explainable AI:** Use SHAP-style explanations, so that administrators understand alerts.
- **Zero-day attack detection:** combine the supervised ensemble with anomaly detection to detect Zero-day attacks.
- **Real-time testbed:** Deploy on a real IoT gateway with live traffic and measure energy consumption.
- **Blockchain and 5G/6G integration:** Secure alert sharing and extend the work to next-generation networks.

References

1.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.

2.Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.E. C. P. Neto et al., “CICIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment,” Sensors, vol. 23, no. 13, Art. no. 5941, 2023.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

3. S. K. R. Mallidi and R. R. Ramisetty, “Advancements in training and deployment strategies for AI-based intrusion detection systems in IoT: A systematic literature review,” Discover Internet of Things, vol. 5, Art. no. 8, 2025.

4. “Dataset-centric evaluation of federated intrusion detection models in IoT networks,” Scientific Reports, 2025.

5. “Evaluating and enhancing intrusion detection systems in IoMT: The importance of domain-specific datasets,” Internet of Things, vol. 32, Art. no. 101631, 2025.

6.Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Multiple Classifier Systems* (Lecture Notes in Computer Science, Vol. 1857, pp. 1–15). Springer.

7.Kolias, C., Kambourakis, G., Stavrou, A., & Voas, J. (2017). DDoS in the IoT: Mirai and other botnets. *Computer, 50*(7), 80–84.

8.Koroniotis, N., Moustafa, N., Sitnikova, E., & Turnbull, B. (2019). Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset. *Future Generation Computer Systems, 100*, 779–796.

9.Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30*.

10.Meidan, Y., Bohadana, M., Mathov, Y., Mirsky, Y., Shabtai, A., Breitenbacher, D., & Elovici, Y. (2018). N-BaIoT: Network-based detection of IoT botnet attacks using deep autoencoders. *IEEE Pervasive Computing, 17*(3), 12–22.

11.Moustafa, N., Turnbull, B., & Choo, K.-K. R. (2019). An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of Internet of Things. *IEEE Internet of Things Journal, 6*(3), 4815–4830.

12.Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

13.Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. In *Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP)* (pp. 108–116).

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14, Issue 11, CONFLUENCE - 2026

14.Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. In *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications* (pp. 1–6). IEEE.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)