Advanced Ensemble Techniques for IoT Cybersecurity: A Performance Comparison on the CICIoT2023 Dataset

Simran Kaur Hora, Antanas Čenys, Nikolaj Goranin

Vilnius Gediminas Technical University, Department of Information Systems

Introduction

IoT Growth and Significance

- IoT has revolutionized sectors like healthcare, smart cities and industrial systems.
- Projected IoT-linked device installations to reach 30.9 billion by 2025.

IoT Security Challenges

- IoT devices are resource-constrained and connected to open networks making them vulnerable to attacks.
- Traditional security mechanisms like cryptography and firewalls are insufficient against advanced threats.

Challenges with Traditional Machine Learning (ML) Methods

- Dynamic behavior of IoT attacks poses detection challenges.
- Traditional ML methods struggle with consistency in high-dimensional and imbalanced datasets.

Research Focus

- Compare performance of traditional ML models with advanced ensemble methods for IoT cyberattack detection.
- Leverage the CICIoT2023 dataset to evaluate performance.

Methodology

Dataset Overview

- Controlled environment mimicking a realistic IoT topology using 105 real IoT devices
- Attack Categories (DDoS,DoS,Reconnaissance,Web-based, Brute Force, Spoofing, Mirai Malware)
- Data Formats: Pcap, CSV

Dataset Preprocessing

- Data categorized into:
  a. Binary Class: Benign vs. Malicious.
  b. Multi-Class: 47 attack types + Benign.

Model Categorization

- Traditional ML Methods
  a. Decision Tree (DT)
  b. Random Forest (RF)
- Ensemble Methods
  c. Gradient Boosted Trees (GBT)
  d. XGBoost (Extreme Gradient Boosting)

Dataset Splitting & Validation

- Dataset divided into 80% Training and 20% Testing subsets.
- 5-Fold Cross-Validation
- Equal-Size Sampling applied to balance and prevent biases

Experimental Setup

Tool Used

- Model training and testing conducted using KNIME (Konstanz Information Miner).

Model Hyperparameters

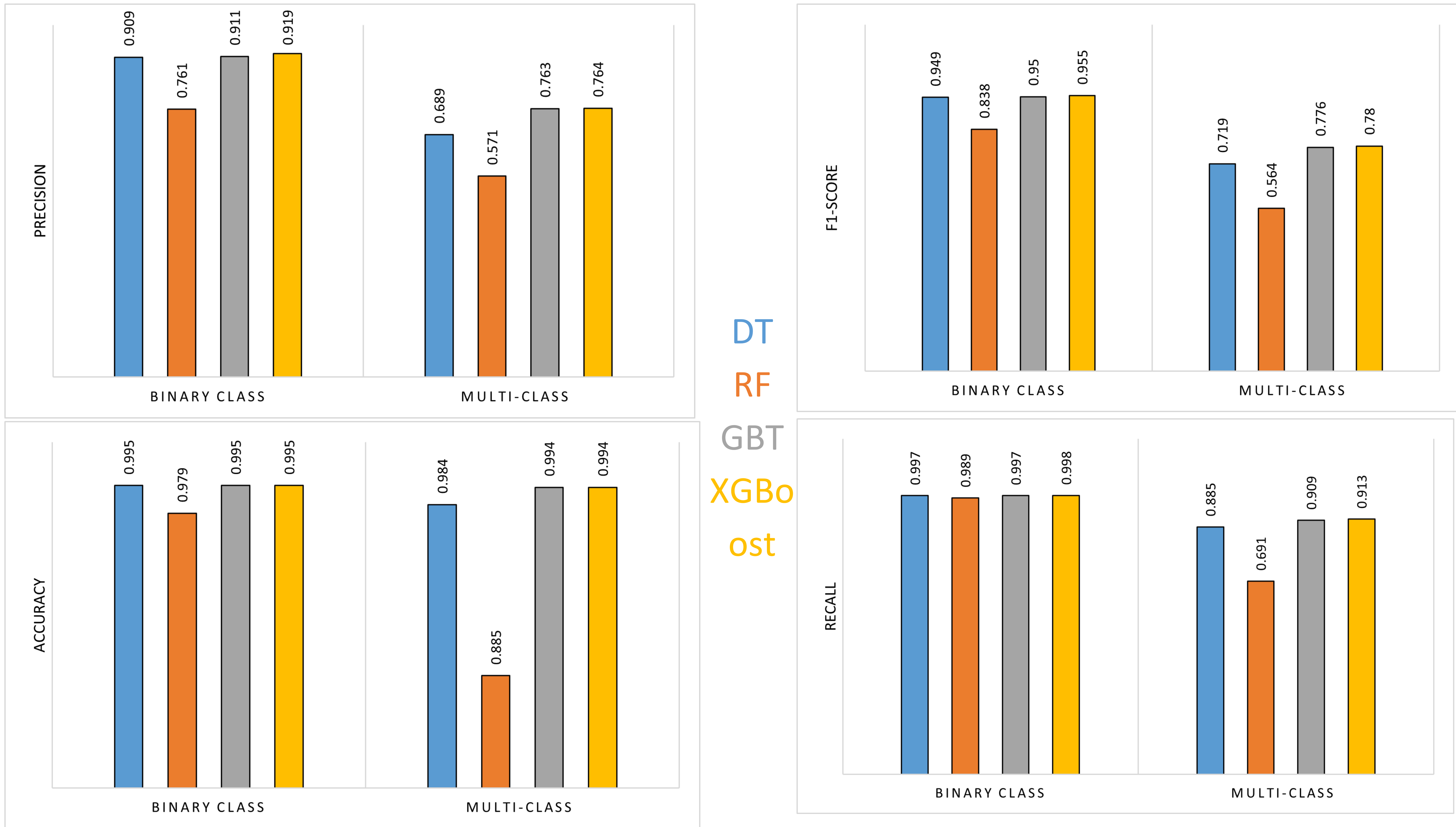
i) Decision Tree
- Gini Index (quality measure)
- Minimum records per node: 50

ii) Random Forest
- Gini Index (split criterion)
- Tree depth: 5

iii) Gradient Boosted Trees
- Tree depth: 5
- Learning rate: 0.1

iv) XGBoost
- Boosting rounds: 100
- Base score: 0.5

Comparison of Performance Metrics on CICIoT2023 Dataset



Key Finding and Future Directions

Key Findings

- Ensemble Methods Superiority
- XGBoost achieved the highest metrics across Recall, Precision, and F-Measure in both Binary and Multi-Class settings.
- Tackled challenges like class imbalance effectively, demonstrating the robustness of ensemble learning for IoT security.
- Highlights the critical role of ensemble models (especially XGBoost) in bolstering IoT network defenses against cyber threats.

Future Directions

- Hybrid Approaches: Combining multiple ensemble techniques for enhanced accuracy and robustness.
- Advanced Models: Explore algorithms like LightGBM or hybrid ensemble-deep learning models.
- Hyperparameter Tuning: Optimize parameters (e.g., learning rate, tree depth) to further boost model performance.

Core References

- Goranin, N.; Hora, S.K.; Čenys, H.A. A Bibliometric Review of Intrusion Detection Research in IoT: Evolution, Collaboration, and Emerging Trends. Electronics 2024, 13, 3210. https://doi.org/10.3390/electronics13163210
- Verma, P.; Dumka, A.; Singh, R.; Ashok, A.; Gehlot, A.; Malik, P.K.; Gaba, G.S.; Hedabou, M. A Novel Intrusion Detection Approach Using Machine Learning Ensemble for IoT Environments. *Appl. Sci.* 2021, *11*, 10268. https://doi.org/10.3390/app112110268
- https://www.unb.ca/cic/datasets/iotdataset-2023.html
- Abdullahi, M.; Baashar, Y.; Alhussian, H.; Alwadain, A.; Aziz, N.; Capretz, L.F.; Abdulkadir, S.J. Detecting Cybersecurity Attacks in Internet of Things Using Artificial Intelligence Methods: A Systematic Literature Review. *Electronics* 2022, *11*, 198. https://doi.org/10.3390/electronics11020198
- Chhikara, Rita, and Neeti Kashyap. "A Comparative Analysis of Machine Learning Prediction Algorithms for Detecting IoT Botnet Activities." *2024 International Conference on Intelligent Systems for Cybersecurity (ISCS)*. IEEE, 2024.
- Neto, E.C.P.; Dadkhah, S.; Ferreira, R.; Zohourian, A.; Lu, R.; Ghorbani, A.A. CICIoT2023: A Real-Time Dataset and Benchmark for Large-Scale Attacks in IoT Environment. Sensors 2023, 23, 5941. https://doi.org/10.3390/s23135941