IoT Cyber Attack Detection Using Machine Learning

Celine Al Harake
Effat University / Jeddah, Saudi Arabia
ceaalharake@effat.edu.sa

Dana Al Rijjal
Effat University / Jeddah, Saudi Arabia
daaalrijjal@effat.edu.sa

Jouri Al Daghma
Effat University / Jeddah, Saudi Arabia
joialdaghima@effat.edu.sa

Layal Canoe
Effat University / Jeddah, Saudi Arabia
laacanoe@effat.edu.sa

Abstract

The proliferation of Internet of Things (IoT) devices has introduced new security vulnerabilities due to their lightweight architecture and limited computational capacity. This project presents a machine learning-based Intrusion Detection System (IDS) for detecting cyber attacks in IoT networks. We developed a complete ML pipeline that includes preprocessing noisy telemetry data, balancing imbalanced classes using SMOTE, and applying dimensionality reduction via PCA. Multiple supervised models, including Random Forest, Linear Support Vector Machine (SVM), and XGBoost, were trained to detect various attack types. Evaluation metrics such as precision, recall, F1-score, and confusion matrix were used to assess performance. The results demonstrate that machine learning techniques are effective in identifying malicious activity in IoT networks, offering a scalable and adaptive security solution.

1 Introduction

The Internet of Things (IoT) has revolutionized the digital landscape by enabling seamless interconnectivity among everyday devices. However, the widespread deployment of IoT devices has also expanded the attack surface for cyber threats, exposing networks to vulnerabilities like Denial of Service (DoS), probing, and data theft. Traditional intrusion detection systems often fail to cope with the unique challenges posed by IoT, including high-dimensional data, real-time constraints, and the emergence of novel attack types. To address these limitations, machine learning (ML) offers a promising alternative. ML-based IDS can learn from network traffic patterns and identify both known and previously unseen threats. This project explores the development and evaluation of such a system using the CIC IoT 2023 dataset, highlighting the benefits and challenges of applying ML in this domain.

2 Related Work

The increasing complexity of cyber threats targeting Internet of Things (IoT) environments has driven the development of advanced Intrusion Detection Systems (IDS). Traditional IDS methods often fail to cope with the resource constraints and heterogeneity of IoT devices. Therefore, researchers have explored machine learning (ML) techniques as scalable alternatives.

Supervised ML models such as Support Vector Machines (SVM), Random Forests, and Gradient Boosting classifiers have been widely applied for intrusion detection. These models rely on labeled training data to distinguish between benign and malicious traffic. For example, Goeschel (2023) employed SVM in IoT networks and demonstrated high detection accuracy across multiple attack classes.

Unsupervised learning models like Isolation Forests and Autoencoders have also gained attention for detecting novel or zero-day attacks. These models identify outliers without requiring labeled data. Sohaib et al. (2024) highlighted that Autoencoders can effectively capture complex latent features in IoT network traffic, improving anomaly detection in real-time systems.

Deep learning models have further advanced intrusion detection capabilities. Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks are especially notable for their ability to learn temporal and spatial traffic patterns. Abbas et al. (2024) proposed a Transformer-based multi-class classifier using the CICIoT2023 dataset, outperforming traditional methods in terms of precision and recall.

Benchmark datasets play a crucial role in evaluating IDS performance. The CICIoT2023 dataset from the Canadian Institute for Cybersecurity is one of the most comprehensive IoT intrusion datasets available, containing diverse attack scenarios and realistic network behavior (CIC, 2023). It has been widely used to test and compare IDS models under realistic IoT conditions.

In summary, prior work underscores the promise of both supervised and unsupervised ML approaches in IoT security. However, challenges remain in achieving high accuracy across all attack types and ensuring generalizability to real-world deployments.

3 Data and Methodology

Dataset Description

The dataset used in this study is the CIC IoT 2023 dataset, which consists of real-world network traffic collected from various IoT devices under normal and malicious conditions. The dataset contains numerical features extracted from network telemetry, such as flow statistics and protocol-based metrics, along with corresponding labels indicating whether each instance is benign or malicious.

3.1 Data Preprocessing

Several preprocessing steps were performed to prepare the data for training:

- **Missing Values**: Missing values in the `Std` and `Variance` columns were filled using their respective median values.

- **Encoding**: The target variable `Label` was encoded using `LabelEncoder`, converting categorical labels into numerical form.

- **Infinite Values**: Infinite values were replaced with NaN and subsequently removed to avoid training inconsistencies.

- **Normalization**: Feature scaling was applied using Z-score standardization (`StandardScaler`) to ensure all features have mean zero and unit variance.

- **Dimensionality Reduction**: Principal Component Analysis (PCA) was used to reduce dimensionality while retaining 95% of the variance in the data.

- **Class Balancing**: To handle class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was used to generate synthetic examples of minority classes, resulting in a balanced dataset.

3.2 Supervised Machine Learning Models

Three supervised machine learning models were trained using the preprocessed dataset:

- **Random Forest (RF)**: An ensemble-based classifier with 100 trees.

- **Support Vector Machine (SVM)**: A linear kernel was used to improve training efficiency and scalability, particularly with high-dimensional data.

- **XGBoost**: A gradient boosting model configured with `use_label_encoder=False` and `eval_metric='mlogloss'`.

The dataset was split into 70% training and 30% testing sets using stratified sampling to maintain class proportions. Each model was trained on the balanced dataset and evaluated using standard performance metrics.

3.3 Unsupervised Anomaly Detection

In addition to supervised learning, unsupervised models were optionally applied to identify previously unseen or novel attacks. These models do not rely on labeled data and are useful in real-world scenarios where new attack types may emerge.

- **Isolation Forest**: This tree-based model detects anomalies by isolating observations in randomly generated partitions. Data points that require fewer splits to be isolated are considered anomalous.

- **Autoencoder**: A neural network-based approach trained to reconstruct normal traffic. High reconstruction error indicates potential anomalies, such as novel attacks that deviate from learned patterns.

These models help enhance the system's robustness by flagging unfamiliar behavior that may not have been seen during training, contributing to a proactive intrusion detection strategy.

4 Results and Discussion

To evaluate the effectiveness of the proposed Intrusion Detection System (IDS), five machine learning models were trained and assessed: Random Forest, Linear Support Vector Machine (SVM), XGBoost, Multilayer Perceptron (MLP), and a 1D Convolutional Neural Network (CNN). Each model was trained on the same preprocessed dataset and evaluated using accuracy, precision, recall, F1-score, and confusion matrix.

4.1 Evaluation Metrics

1. **Random Forest** achieved the highest accuracy at **95%**, with strong precision and recall across most classes. It particularly excelled in identifying class 1 (F1-score: 1.00) and maintained balanced performance even on minority classes.

2. **XGBoost** followed with **92% accuracy**, demonstrating robust generalization. It showed slightly lower performance for class 3 (F1-score: 0.80) but performed well on the rest.

3. **Linear SVM** underperformed with an accuracy of only **76%**. It struggled significantly on classes 0, 2, and 3, while achieving near-perfect precision and recall only on class 1.

4. **MLP Neural Network** reached a validation accuracy of **84.7%**. It performed decently on frequent classes but showed reduced capability on less frequent ones, especially class 3.

5. **CNN (1D Convolutional)** had the lowest performance with a peak validation accuracy of **74.9%**. The model failed to effectively learn class boundaries, indicating that the tabular structure of the dataset was not well-suited for convolutional approaches.

4.2 Classification Performance Summary

Model	Accuracy	Best F1-score Class	Worst F1-score Class
Random Forest	95%	Class 1 (1.00)	Class 3 (0.90)
XGBoost	92%	Class 1 (0.98)	Class 3 (0.80)
Linear SVM	76%	Class 1 (1.00)	Class 2 (0.29)
MLP	84.7% (val)	Class 1 (0.95)	Class 3 (0.71)
CNN	74.9% (val)	Class 1 (0.93)	Class 2 (0.00)

Table 1: Comparison of classification performance across models

4.3 Confusion Matrix Visualization

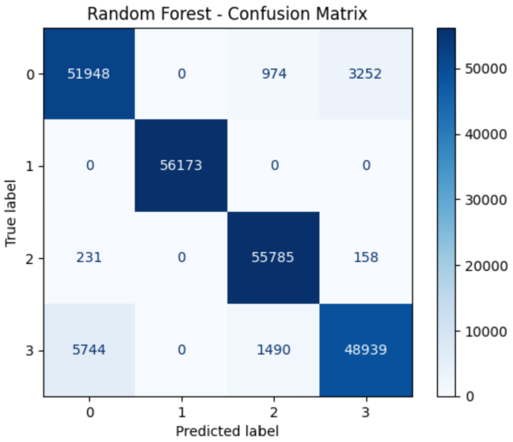

Figure 1: Confusion Matrix - Random Forest

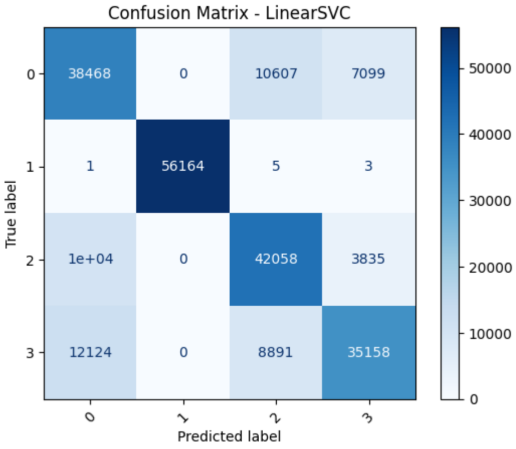

Figure 2: Confusion Matrix - Linear SVM

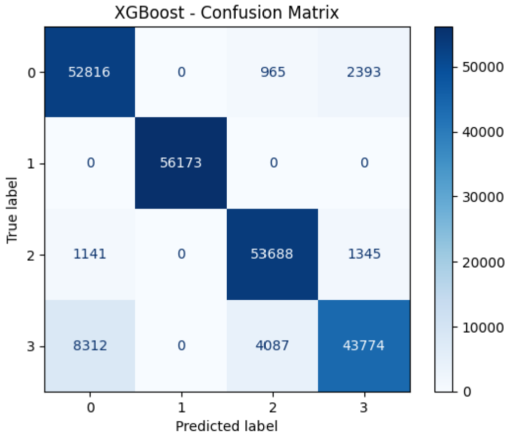

Figure 3: Confusion Matrix - XGBoost

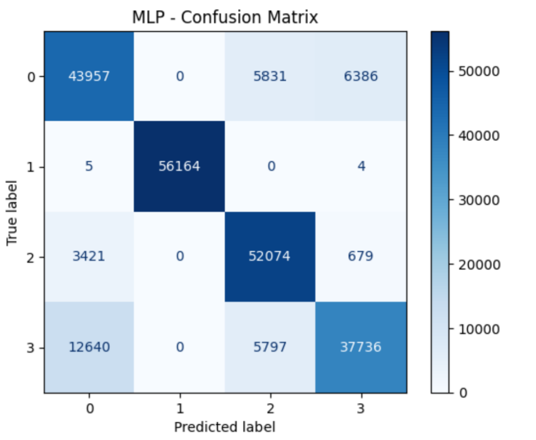

Figure 4: Confusion Matrix - MLP Neural Network

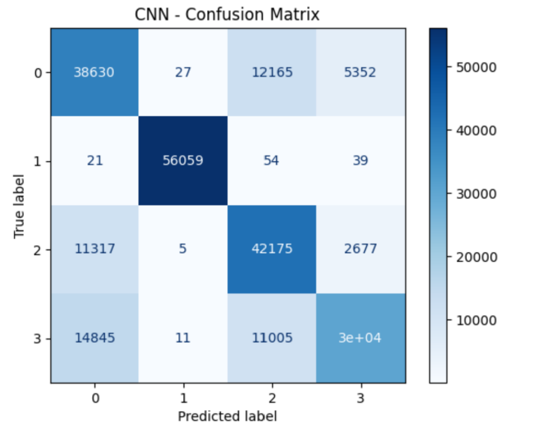

Figure 5: Confusion Matrix - CNN Neural Network

4.4 Discussion

The results show that tree-based ensemble models outperform linear and neural architectures on this tabular dataset. Random Forest was the top performer in terms of accuracy, consistency, and per-class recall. XGBoost was competitive with slightly better generalization and less overfitting. Linear SVM showed high precision for class 1 but failed on others. Neural networks, especially CNN, performed poorly—likely due to the absence of spatial features in the tabular input.

These findings highlight the importance of model selection based on data type, and reinforce the value of preprocessing techniques like SMOTE and PCA in improving overall performance.

5 Conclusion and Future Work

This project presented the development of a machine learning-based Intrusion Detection System (IDS) tailored for Internet of Things (IoT) networks. Using the CIC IoT 2023 dataset, we implemented a full ML pipeline involving data cleaning, dimensionality reduction via PCA, and class balancing through SMOTE. Five machine learning models were evaluated: Random Forest, XGBoost, Linear SVM, MLP, and CNN.

The results clearly demonstrate the superiority of ensemble methods, particularly Random Forest, which achieved the highest accuracy (95%) and showed balanced detection performance across all attack types. XGBoost was a close second, while linear and deep models, especially CNN, underperformed due to the nature of tabular data and class imbalance challenges.

5.1 Future Work

- **Hyperparameter Tuning:** Apply advanced techniques like Grid Search or Bayesian Optimization to fine-tune model performance.
- **Hybrid Architectures:** Combine supervised models (e.g., Random Forest) with unsupervised approaches (e.g., Autoencoders) for detecting novel or zero-day attacks.
- **Explainable AI (XAI):** Use SHAP or LIME to interpret model predictions, enabling transparent and trustworthy decisions in security environments.
- **Deployment in Real-Time Environments:** Test latency, scalability, and resilience of models in real-world IoT settings.
- **Cross-Dataset Evaluation:** Validate the pipeline on different IoT attack datasets to assess generalizability and improve robustness.

In conclusion, this study demonstrates that carefully designed machine learning pipelines—when paired with appropriate preprocessing and model selection—can significantly enhance cybersecurity in IoT ecosystems.

References

Michael Goeschel. Machine learning-based intrusion detection methods in IoT systems. *Electronics*, 13(18):3601, 2023. URL: https://www.mdpi.com/2079-9292/13/18/3601.

Muhammad Sohaib et al. A survey on intrusion detection system in IoT networks. *Internet of Things and Cyber-Physical Systems*, 4:100048, 2024. URL: https://www.sciencedirect.com/science/article/pii/S2772918424000481.

Noman Abbas et al. Multi-class intrusion detection based on transformer for IoT networks using CIC-IoT-2023 dataset. *Future Internet*, 16(8):284, 2024. URL: https://www.mdpi.com/1999-5903/16/8/284.

Canadian Institute for Cybersecurity. CICIoT2023 Dataset, 2023. URL: https://www.unb.ca/cic/datasets/iotdataset-2023.html.