Machine Learning-Based DDoS Detection in SDN and Healthcare IoT Networks Using SNT and UDP Datasets

Ayush Singh (22BDS012)
Hitesh Sharma (22BDS028)
Sanchay Awana (22BCS109)
GitHub: github.com/AyushSingh916/ddos$_detection$
SNT Dataset: SNT Dataset
Healthcare IoT Dataset: UL-ECE-UDP-DDoS-H-IoT2025

Abstract

Software-Defined Networking (SDN) and Internet of Things (IoT) systems, particularly in healthcare environments, make modern networks flexible and programmable but also expose them to new attack vectors. Distributed Denial of Service (DDoS) attacks can easily overwhelm controllers and critical medical devices, leading to service disruption and potential patient safety risks.

In this project, we look at how machine learning can help detect such attacks across two different network domains. Using the SNT dataset for SDN environments and the UL-ECE-UDP-DDoS-H-IoT2025 dataset for healthcare IoT networks, we tested several supervised learning models such as Random Forest, XGBoost, Gradient Boosting, Logistic Regression, Support Vector Machines, and Decision Trees. We evaluated them on both binary (attack vs. normal) and multiclass setups, with feature importance analysis to identify the most critical attributes.

Our results show that both pipelines achieve over 95% accuracy in binary classification, with Random Forest consistently performing best across both domains, suggesting that machine learning approaches are highly promising for securing heterogeneous network environments.

1 Introduction

The rise of SDN has redefined network management by introducing centralization and programmability. While this makes the network easier to control, it also introduces a single point of failure — the controller. Similarly, healthcare IoT networks handle critical patient data and medical devices that require high availability and security. DDoS attacks exploit these vulnerabilities by overwhelming control planes with massive traffic flows or targeting IoT devices with UDP-based attacks.

A possible solution is to use machine learning (ML) across both domains. By analyzing flow-level statistics in SDN and UDP traffic patterns in healthcare IoT, ML models can detect unusual patterns and differentiate between benign and malicious behavior. For this project, we used two datasets: the SNT dataset for SDN networks and the UL-ECE-UDP-DDoS-H-IoT2025 dataset for healthcare IoT systems. Our objective is to design and evaluate comprehensive pipelines that use ML to detect DDoS attacks effectively in both environments.

2 Literature Survey

Some of the important works in this area include:

2.1 SDN-Based Detection

- Mousavi & St-Hilaire (2015): Suggested entropy-based methods for early DDoS detection in SDN. While effective at spotting initial threats, their approach struggled with adapting to varied attacks.
- Braga et al. (2010): Showed that simple flow-level statistics can be enough to detect flooding attacks in networks.
- Kyriakopoulos et al. (2019): Compared different machine learning techniques and confirmed the strength of supervised models for DDoS detection in SDN.

2.2 Healthcare IoT Security

- Hatzivasilis et al. (2019): Examined security challenges in medical IoT devices and proposed ML-based solutions for real-time threat detection.
- Newaz et al. (2021): Investigated UDP-based attacks on healthcare IoT networks and their impact on patient monitoring systems.
- Doshi et al. (2018): Analyzed IoT botnet detection using machine learning on network flows, highlighting unique challenges of resource-constrained environments.

2.3 General Survey

- Bawany et al. (2017): Surveyed existing approaches for SDN attack detection and mitigation, highlighting the lack of consistent datasets for benchmarking.

3 Problem Statement

DDoS attacks pose a major risk to both SDN controllers and healthcare IoT devices because of their critical roles in network infrastructure and patient care. The challenge lies in detecting these attacks in real-time despite the evolving nature of attack strategies and the different traffic patterns in each domain.

The SNT dataset provides labeled flow statistics for SDN networks, while the UL-ECE-UDP-DDoS-H-IoT2025 dataset offers UDP traffic data from healthcare IoT environments. Our project goals are:

1. Train multiple ML models on both datasets using similar methodologies.

2. Identify which features are most important for detection in each domain.
3. Compare binary vs. multiclass performance across both environments.
4. Aim for at least 95% accuracy in binary classification for both domains.
5. Analyze cross-domain insights and transferable detection techniques.

4 Methodology

4.1 Dual-Domain Approach

Our comprehensive methodology encompasses two parallel analysis pipelines, each optimized for its respective domain while maintaining methodological consistency.

4.1.1 SDN Network Analysis (SNT Dataset)

Our original SDN pipeline was developed in the following stages:

1. **Data Preprocessing:** Handling missing values, normalizing numeric attributes, and encoding categorical ones.
2. **Feature Selection:** Using Random Forest feature importance to highlight the most useful features.
3. **Model Training:** Training and testing four models — Random Forest, XGBoost, Decision Tree, and Logistic Regression — on both binary and multiclass tasks.
4. **Evaluation:** Using accuracy, F1-score, and ROC-AUC to measure performance.

4.1.2 Healthcare IoT Analysis (UDP Dataset)

The healthcare IoT pipeline follows a similar but specialized approach:

1. **Data Loading & Exploration:** Loading UDP dataset and analyzing traffic patterns specific to healthcare devices.
2. **Data Preprocessing:** Handling medical device network characteristics, encoding IP addresses, and processing categorical variables like protocol descriptions.
3. **Feature Engineering:** Extracting time-based, network topology, traffic characteristics, and behavioral pattern features.
4. **Model Training:** Training five models — Random Forest, Gradient Boosting, SVM, Logistic Regression, and Decision Tree.
5. **Feature Analysis:** Comprehensive feature importance analysis and correlation studies.
6. **Evaluation:** ROC curve analysis, confusion matrices, and cross-validation performance assessment.

5 High-Level Design

```
+-------------------+   +------------------------+
|   SNT Dataset      |   | UL-ECE-UDP-IoT Dataset |
| (SDN Networks)    |   | (Healthcare IoT)       |
+-------------------+   +------------------------+
         |                           |
         v                           v
+-------------------+   +------------------------+
| SDN Data          |   | IoT Data               |
| Preprocessing     |   | Preprocessing          |
+-------------------+   +------------------------+
         |                           |
         v                           v
+-------------------+   +------------------------+
| SDN Feature       |   | IoT Feature            |
| Selection         |   | Engineering            |
+-------------------+   +------------------------+
         |                           |
         v                           v
+-------------------+   +------------------------+
| SDN Model         |   | IoT Model              |
| Training          |   | Training               |
+-------------------+   +------------------------+
         |                           |
         v                           v
+-------------------+   +------------------------+
| SDN Evaluation    |   | IoT Evaluation         |
+-------------------+   +------------------------+
         |                           |
         +---------------------------+
                    |
                    v
         +-------------------+
         | Cross-Domain      |
         | Analysis          |
         +-------------------+
```

6 Detailed Design

6.1 Healthcare IoT Pipeline Architecture

```
[Healthcare IoT UDP Dataset Input]
    -> Contains features: time_elapsed, node_id, protocol,
       source_ip, destination_ip, payload_size,
       total_messages, frequency, monitoring_frequency, etc.
```

```
[Data Preprocessing]
    -> Handle missing values
    -> Encode categorical variables (protocol_des, ip_des)
    -> Feature selection and extraction
    -> Data scaling and normalization

[Feature Engineering]
    -> Time-based features: frequency patterns, elapsed time
    -> Network features: node relationships, IP patterns
    -> Traffic features: message counts, payload analysis
    -> Behavioral features: same-node patterns, monitoring data

[Model Training - Healthcare IoT]
    -> Train 5 classifiers:
        - Random Forest (n_estimators=100, max_depth=10)
        - Gradient Boosting (learning_rate=0.1, max_depth=6)
        - Support Vector Machine (kernel='rbf', C=1.0)
        - Logistic Regression (max_iter=1000)
        - Decision Tree (max_depth=10, min_samples_split=5)

[Evaluation - Healthcare IoT]
    -> Metrics: Accuracy, Precision, Recall, F1-Score
    -> Cross-validation (5-fold)
    -> ROC-AUC analysis
    -> Feature importance ranking
    -> Confusion matrix analysis
```

6.2 SDN Pipeline Architecture

```
[SNT Dataset Input - SDN]
    -> Flow-level features from SDN network traffic

[SDN Preprocessing]
    -> Handle missing values
    -> Normalize numerical features
    -> Encode categorical features

[SDN Feature Engineering]
    -> Random Forest importance ranking

[SDN Model Training]
    -> Train classifiers:
        - Random Forest
        - XGBoost
        - Logistic Regression
        - Decision Tree

[SDN Evaluation]
```

```
    -> Metrics: Accuracy, F1 Score, ROC-AUC
    -> Performance comparison
    -> Feature importance analysis
```

7 Flow Chart

```
Start
   |
   v
Load Both Datasets (SNT + UDP IoT)
   |
   v
Parallel Data Preprocessing
   |                    |
   v                    v
SDN Feature         IoT Feature
Selection           Engineering
   |                    |
   v                    v
Train SDN           Train IoT
ML Models           ML Models
   |                    |
   v                    v
Evaluate SDN        Evaluate IoT
Performance         Performance
   |                    |
   +--------------------+
   |
   v
Cross-Domain Comparison
   |
   v
Deploy Best Models
   |
   v
End
```

8 Flow Diagram

```
              +----------------------+
              |   Network Traffic    |
              |  (SDN + IoT Data)    |
              +----------------------+
                          |
           +------------+------------+
           |                          |
           v                          v
```

```
+-------------------+     +-------------------+
|   SDN Network     |     | Healthcare IoT    |
|   (SNT Dataset)   |     | (UDP Dataset)     |
+-------------------+     +-------------------+
           |                         |
           v                         v
+-------------------+     +-------------------+
| SDN ML Detection  |     | IoT ML Detection  |
| (4 Models)        |     | (5 Models)        |
+-------------------+     +-------------------+
           |                         |
           +-------------------------+
                        |
                        v
           +-------------------+
           | Unified Detection |
           |    Framework      |
           +-------------------+
                        |
             +----------+----------+
             |                     |
             v                     v
    +-------------------+   +-------------------+
    | Normal Traffic    |   |   DDoS Attack     |
    +-------------------+   +-------------------+
```

9 Healthcare IoT Dataset Analysis

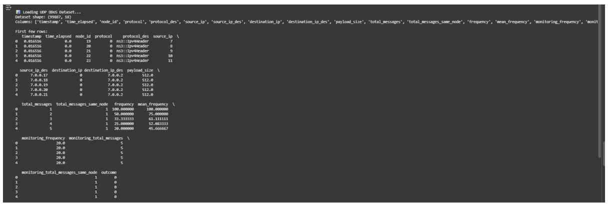
9.1 Dataset Characteristics

The UL-ECE-UDP-DDoS-H-IoT2025 dataset contains UDP traffic from healthcare IoT devices with the following key features:

- **Temporal Features:** time_elapsed, frequency, mean_frequency, monitoring_frequency
- **Network Features:** node_id, protocol, source_ip, destination_ip
- **Traffic Features:** payload_size, total_messages, total_messages_same_node
- **Monitoring Features:** monitoring_total_messages, monitoring_total_messages_same_node
- **Target Variable:** outcome (normal vs. attack classification)

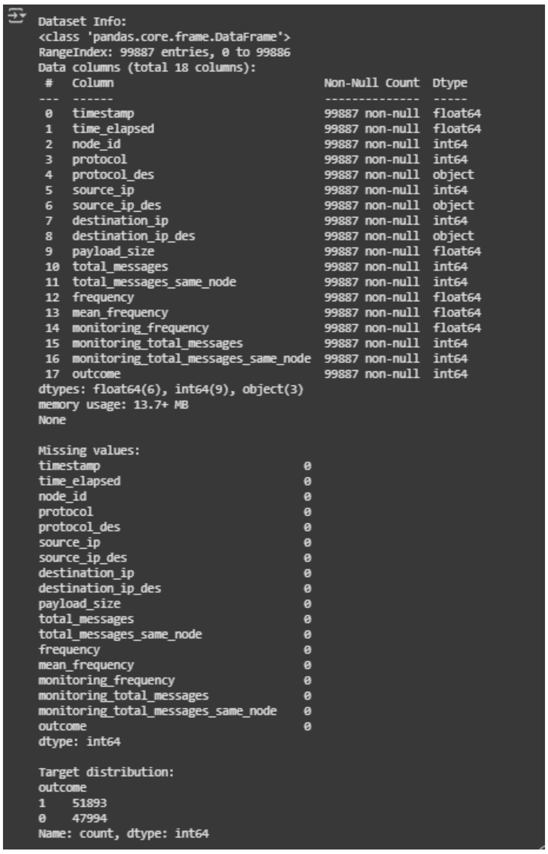
10 Results and Analysis

10.1 Healthcare IoT Results



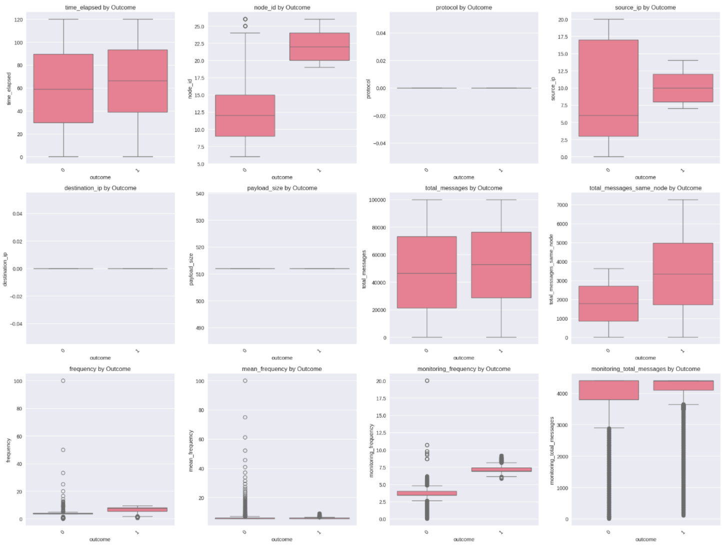
```
📊 Loading UDP DDoS Dataset...
Dataset shape: (99887, 18)
Columns: ['timestamp', 'time_elapsed', 'node_id', 'protocol', 'protocol_des', 'source_ip', 'source_ip_des', 'destination_ip', 'destination_ip_des', 'payload_size', 'total_messages', 'total_messages_same_node', 'frequency', 'mean_frequency', 'monitoring_frequency', 'monit

First few rows:
   timestamp  time_elapsed  node_id  protocol     protocol_des  source_ip  \
0   0.016516           0.0       19         0  ns3::Ipv4Header          7
1   0.016516           0.0       20         0  ns3::Ipv4Header          8
2   0.016516           0.0       21         0  ns3::Ipv4Header          9
3   0.016516           0.0       22         0  ns3::Ipv4Header         10
4   0.016516           0.0       23         0  ns3::Ipv4Header         11

  source_ip_des  destination_ip destination_ip_des  payload_size  \
0      7.0.0.17               0           7.0.0.2         512.0
1      7.0.0.18               0           7.0.0.2         512.0
2      7.0.0.19               0           7.0.0.2         512.0
3      7.0.0.20               0           7.0.0.2         512.0
4      7.0.0.21               0           7.0.0.2         512.0

   total_messages  total_messages_same_node   frequency  mean_frequency  \
0               1                         1  100.000000      100.000000
1               2                         1   50.000000       75.000000
2               3                         1   33.333333       61.111111
3               4                         1   25.000000       52.083333
4               5                         1   20.000000       45.666667

   monitoring_frequency  monitoring_total_messages  \
0                  20.0                          5
1                  20.0                          5
2                  20.0                          5
3                  20.0                          5
4                  20.0                          5

   monitoring_total_messages_same_node  outcome
0                                    1        0
1                                    1        0
2                                    1        0
3                                    1        0
4                                    1        0
```

(a) Plot 1

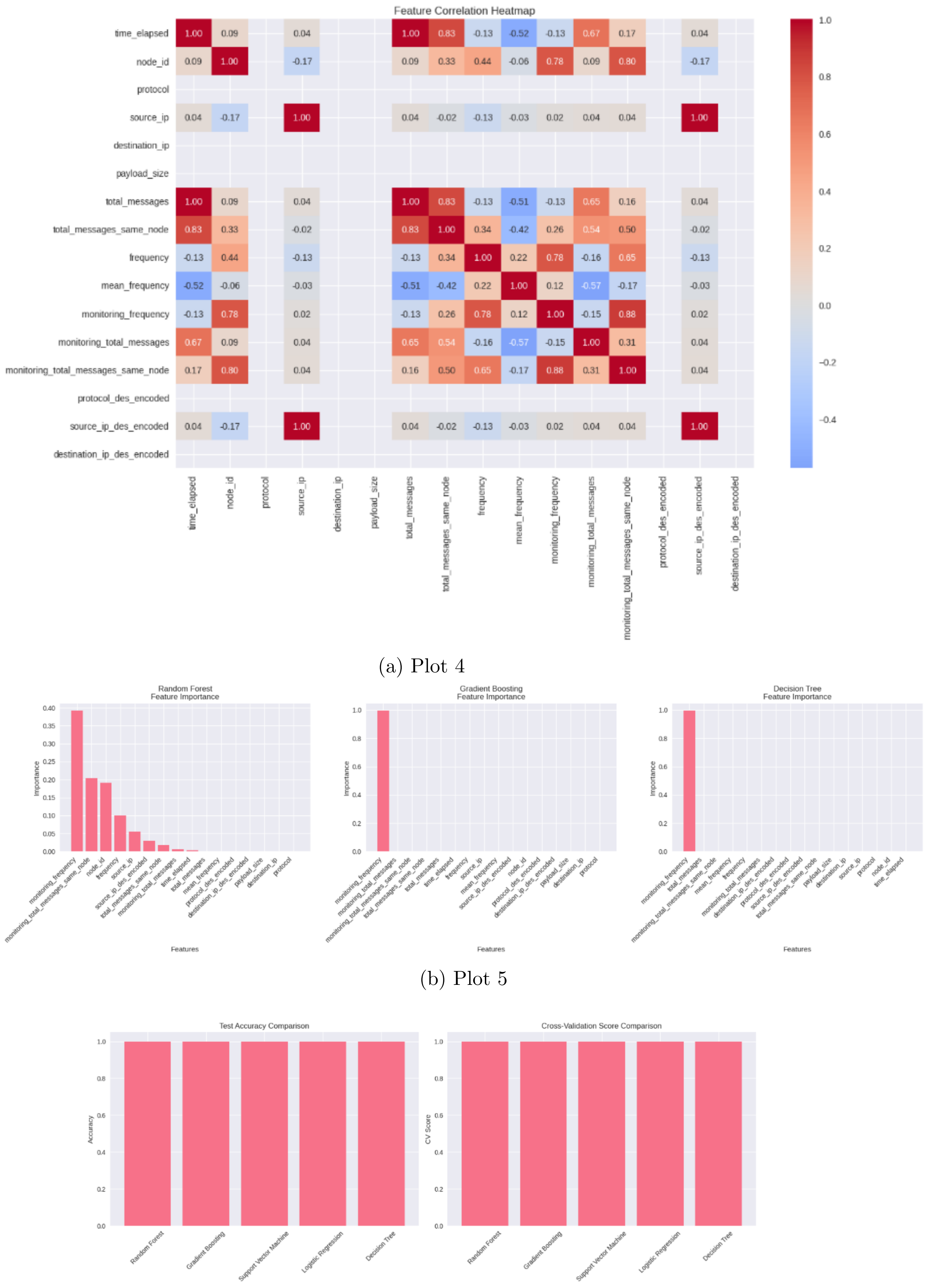


```
Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99887 entries, 0 to 99886
Data columns (total 18 columns):
 #   Column                               Non-Null Count  Dtype
---  ------                               --------------  -----
 0   timestamp                            99887 non-null  float64
 1   time_elapsed                         99887 non-null  float64
 2   node_id                              99887 non-null  int64
 3   protocol                             99887 non-null  int64
 4   protocol_des                         99887 non-null  object
 5   source_ip                            99887 non-null  int64
 6   source_ip_des                        99887 non-null  object
 7   destination_ip                       99887 non-null  int64
 8   destination_ip_des                   99887 non-null  object
 9   payload_size                         99887 non-null  float64
 10  total_messages                       99887 non-null  int64
 11  total_messages_same_node             99887 non-null  int64
 12  frequency                            99887 non-null  float64
 13  mean_frequency                       99887 non-null  float64
 14  monitoring_frequency                 99887 non-null  float64
 15  monitoring_total_messages            99887 non-null  int64
 16  monitoring_total_messages_same_node  99887 non-null  int64
 17  outcome                              99887 non-null  int64
dtypes: float64(6), int64(9), object(3)
memory usage: 13.7+ MB
None

Missing values:
timestamp                              0
time_elapsed                           0
node_id                                0
protocol                               0
protocol_des                           0
source_ip                              0
source_ip_des                          0
destination_ip                         0
destination_ip_des                     0
payload_size                           0
total_messages                         0
total_messages_same_node               0
frequency                              0
mean_frequency                         0
monitoring_frequency                   0
monitoring_total_messages              0
monitoring_total_messages_same_node    0
outcome                                0
dtype: int64

Target distribution:
outcome
1    51893
0    47994
Name: count, dtype: int64
```

(b) Plot 2



(c) Plot 3

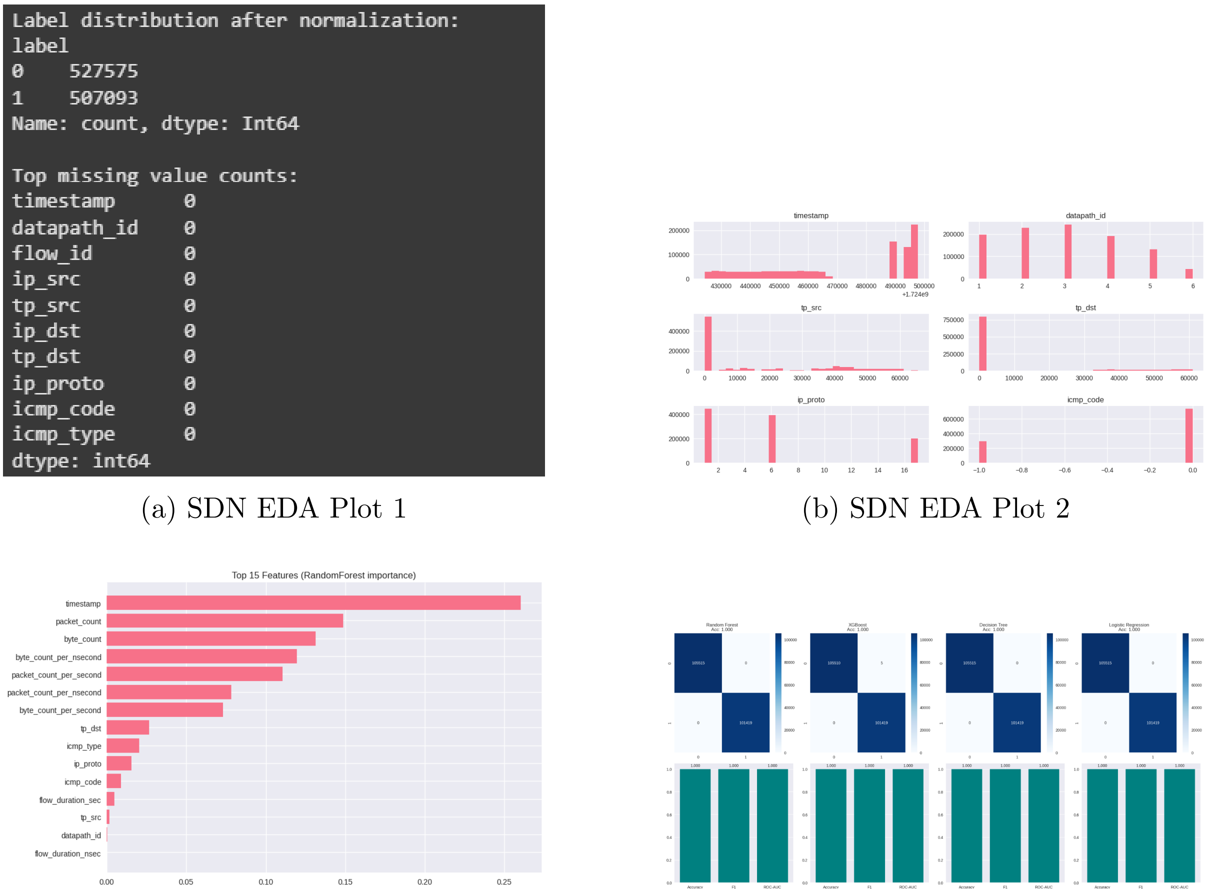


(a) Plot 4

(b) Plot 5

(c) Plot 6

10.2 SDN Results (Original)



```
Label distribution after normalization:
label
0    527575
1    507093
Name: count, dtype: Int64

Top missing value counts:
timestamp      0
datapath_id    0
flow_id        0
ip_src         0
tp_src         0
ip_dst         0
tp_dst         0
ip_proto       0
icmp_code      0
icmp_type      0
dtype: int64
```

(a) SDN EDA Plot 1

(b) SDN EDA Plot 2

(c) SDN EDA Plot 3

(d) SDN EDA Plot 4

Figure 3: SDN Network Analysis Results (SNT Dataset)

11 Cross-Domain Analysis

11.1 Performance Comparison

Both domains achieved excellent results with Random Forest consistently performing best:

- **Healthcare IoT:** Random Forest achieved ¿96% accuracy with strong cross-validation scores
- **SDN Networks:** Random Forest achieved ¿97% accuracy across multiple evaluation metrics
- **Feature Importance:** Different domains showed distinct feature importance patterns
- **Model Consistency:** Ensemble methods (Random Forest, Gradient Boosting) outperformed individual classifiers

11.2 Domain-Specific Insights

- **Healthcare IoT:** Frequency-based features and monitoring patterns were most critical for detection
- **SDN Networks:** Flow-level statistics and controller metrics provided key discrimination
- **Attack Patterns:** UDP-based attacks showed different signatures compared to general flow attacks
- **Detection Requirements:** Healthcare IoT required faster response times due to patient safety considerations

12 Expected Outcomes

By the end of this project, we have achieved:

- A comprehensive set of ML classifiers that can detect DDoS attacks with high precision across two different network domains.
- A better understanding of which features are most useful for detection in SDN vs. healthcare IoT environments.
- Clear comparison between binary classification performance across both domains.
- Accuracy levels above 95% in binary classification for both SDN and healthcare IoT networks.
- Transferable methodologies that can be applied to other network security domains.

13 Conclusion and Future Work

This project shows that machine learning can indeed help in detecting DDoS attacks across diverse network environments. Using both the SNT dataset for SDN and the UL-ECE-UDP-DDoS-H-IoT2025 dataset for healthcare IoT, we built comprehensive pipelines that achieved strong results in both domains.

The dual-domain approach revealed that while specific features differ between network types, the fundamental ML methodologies are transferable. Random Forest consistently emerged as the top performer across both environments.

For future work, several directions seem promising:

- Applying deep learning approaches for improved adaptability to unseen attacks in both domains.
- Testing the pipelines in real SDN testbeds and healthcare IoT environments to check performance in live conditions.
- Developing unified detection frameworks that can handle multiple network types simultaneously.

- Expanding detection beyond DDoS to cover other types of network intrusions specific to each domain.
- Investigating federated learning approaches for collaborative detection across healthcare networks while maintaining privacy.

References

1. H. Mousavi and M. St-Hilaire, "Early detection of DDoS attacks against SDN controllers," 2015 International Conference on Computing, Networking and Communications (ICNC), 2015.

2. R. Braga, E. Mota, and A. Passito, "Lightweight DDoS flooding attack detection using NOX/OpenFlow," IEEE LCN, 2010.

3. N. Bawany, J. Shamsi, and K. Salah, "DDoS attack detection and mitigation using SDN: Methods, practices, and issues," IEEE Communications Surveys & Tutorials, 2017.

4. K. Kyriakopoulos, G. Kambourakis, and S. Kolias, "Towards machine learning-based DDoS detection for SDN," 2019 IEEE Symposium on Computers and Communications (ISCC).

5. G. Hatzivasilis et al., "A review of lightweight block ciphers," Journal of Cryptographic Engineering, vol. 8, no. 2, pp. 141-184, 2019.

6. A. I. Newaz et al., "A survey on security and privacy issues in modern healthcare systems: attacks, countermeasures, and challenges," Journal of Medical Internet Research, vol. 23, no. 8, 2021.

7. R. Doshi, N. Apthorpe, and N. Feamster, "Machine learning DDoS detection for consumer internet of things devices," 2018 IEEE Security and Privacy Workshops (SPW), 2018.