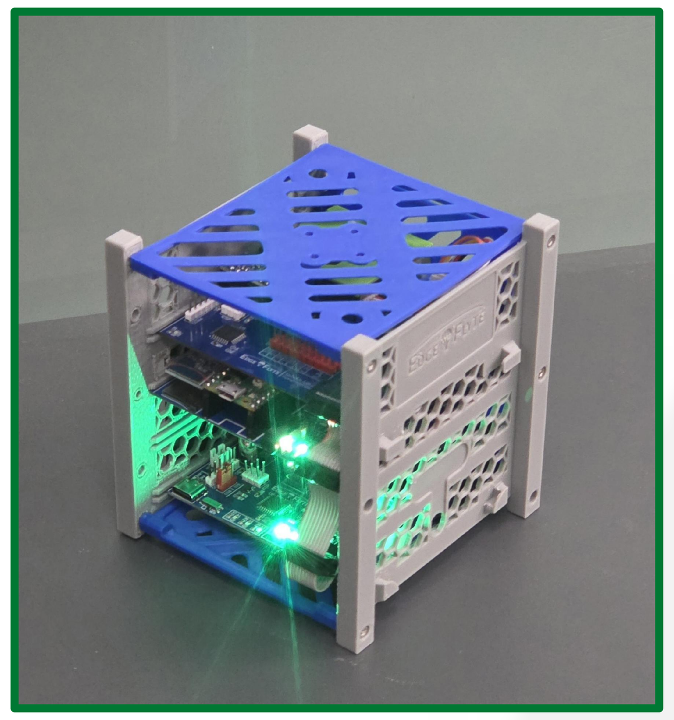
Evaluating the Validity of the CuCD-ID Dataset for Machine Learning Intrusion Detection

Brogan Oberhaus, *Farragut High School* | **Dr. Luke Koch**, *Oak Ridge National Laboratory*

Center for Artificial Intelligence Security Research

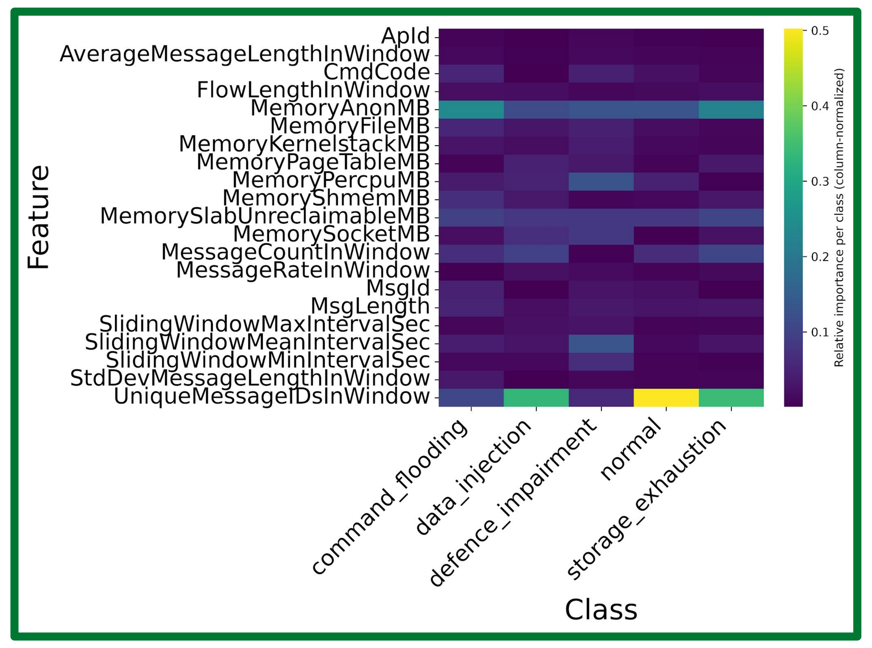
Introduction

- Low Earth Orbit (LEO) satellites often lack basic security features despite performing high-importance functions.
- The CubeSat Cybersecurity Dataset for Intrusion Detection (CuCD-ID)[1] consists of simulated network traffic captured during four malicious and one baseline scenario.
- Each record contains a network packet, a sliding window describing the last 20 seconds of traffic, and metric information.

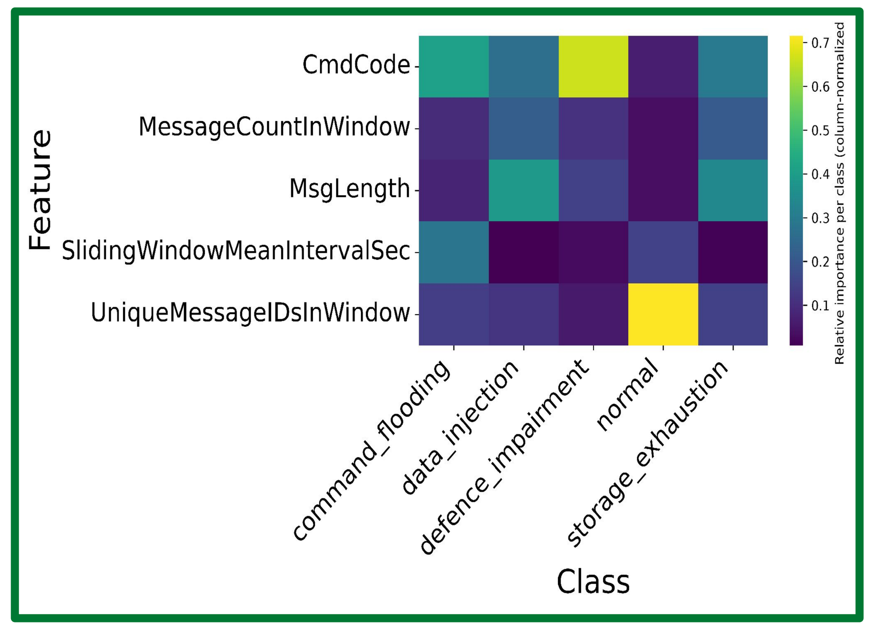


EdgeFlyte CubeSat, a LEO Nanosatellite

Model Training



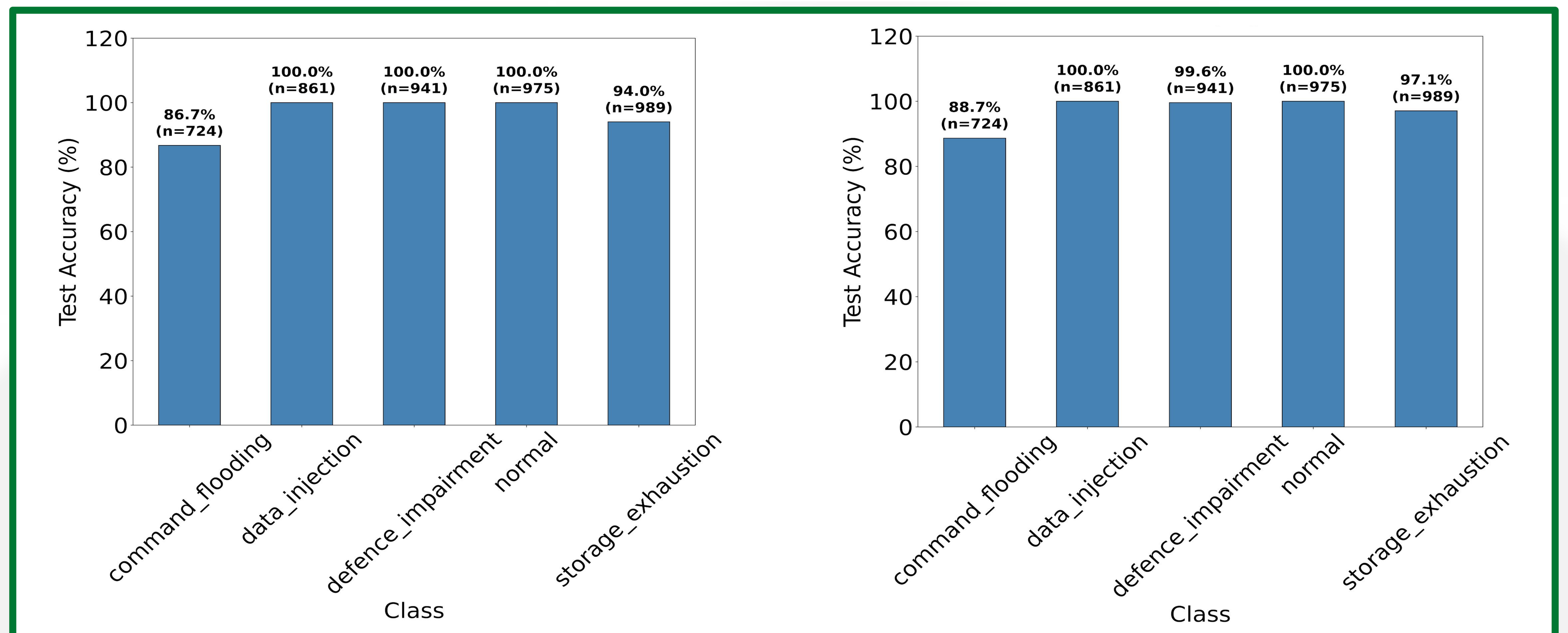
Feature Importance (Integrated Gradients) 21 Features



Feature Importance (Integrated Gradients) 5 Features

- A neural network with 4 neurons was trained to classify the data.
- 16 input features were algorithmically culled to improve efficiency.
- The model was shown to rely heavily on sliding window features.

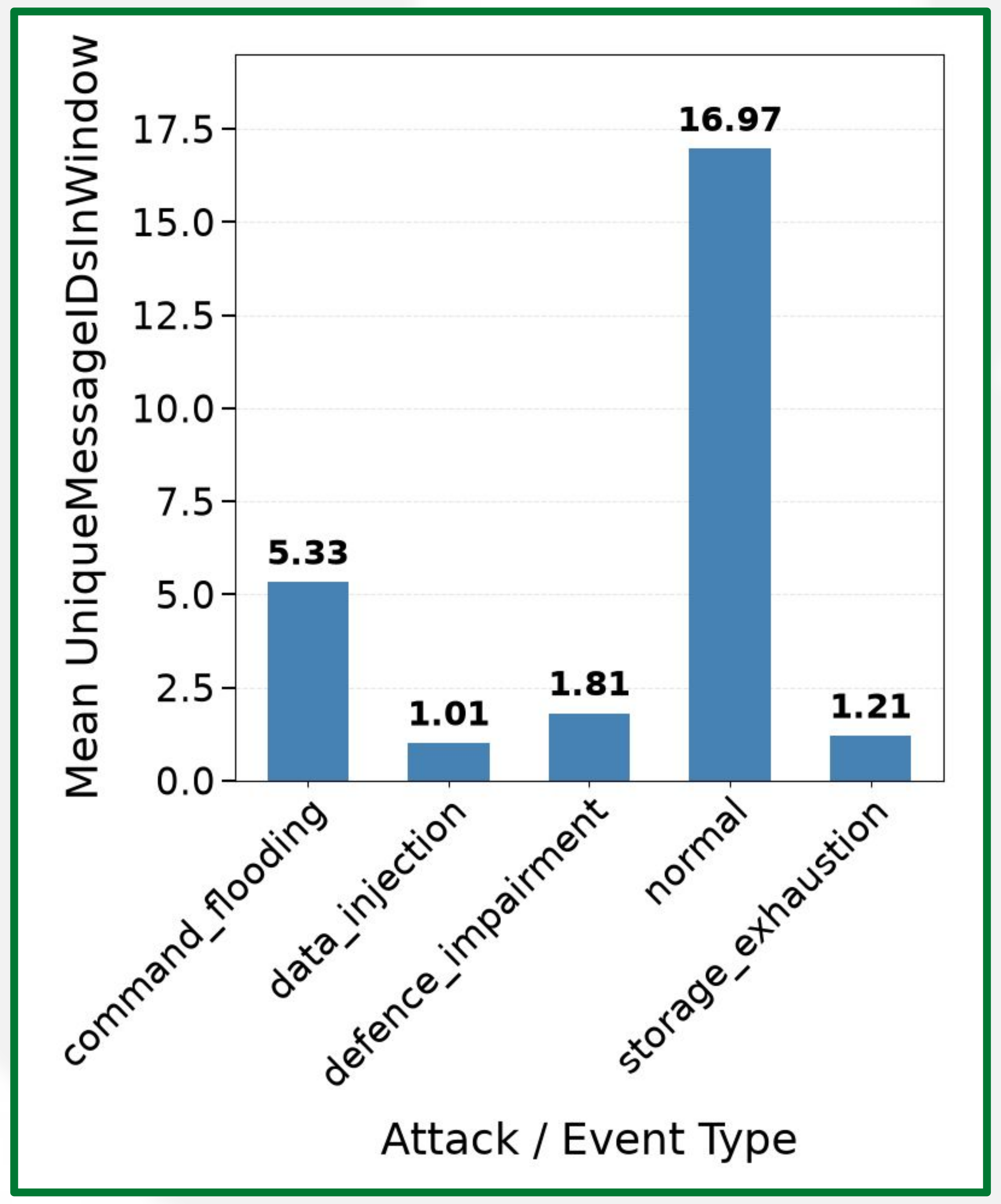
Model Performance



Test Accuracy 21 Features | **Test Accuracy 5 Features**

The model using only 5 features was able to achieve an overall accuracy of over 97 percent, equivalent to using all inputs.

Analyzing the Dataset

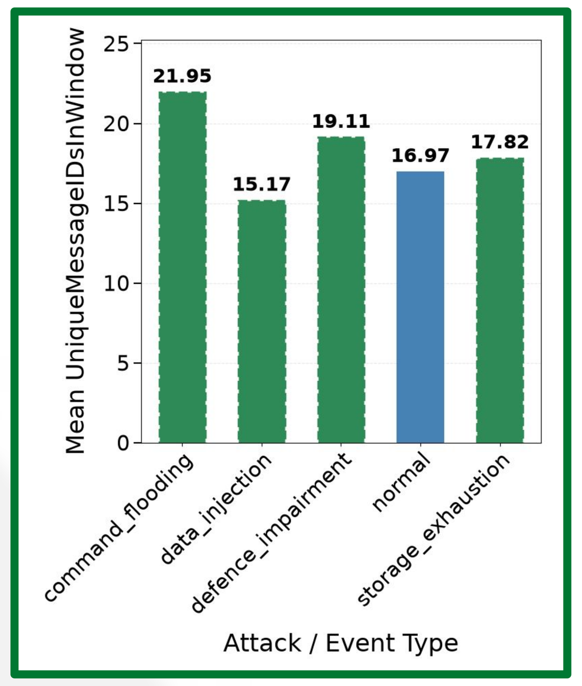


Sliding Window Command Variation

The bar graph shows the attack scenarios contain less variation than the baseline, revealing benign traffic is not present in the malicious data's sliding window.

An examination of the data generation code confirmed the only scenario containing benign traffic was the normal operations script.

Conclusion

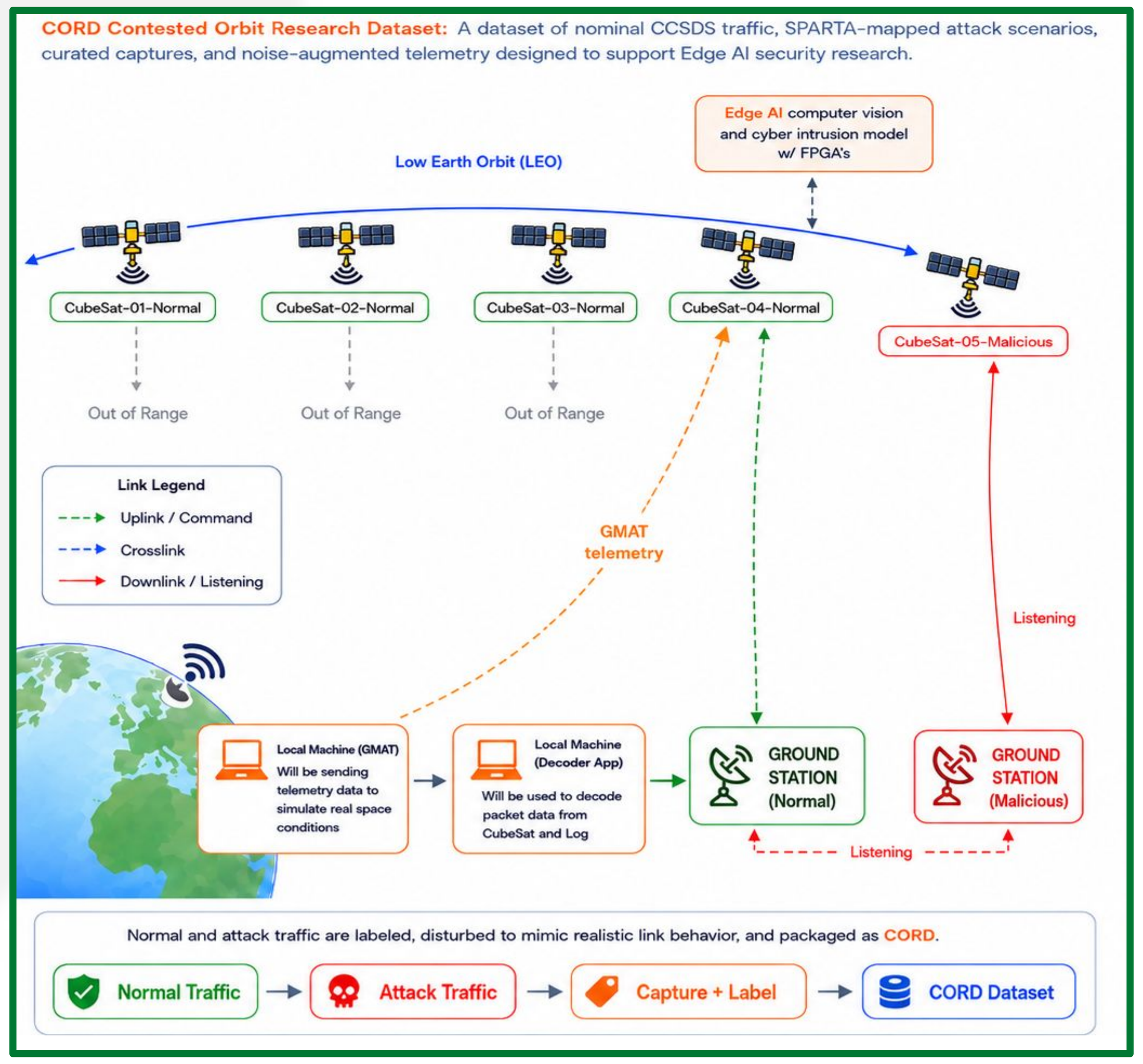


Improved Command Variation Example (Theoretical)

This dataset lacks sufficient realism due to the absence of health-checks and other normal traffic during attacks, saturating the context-aware sliding window with malicious packets.

Future Work

We have built our own hardware-in-the-loop range and are building a cyber intrusion dataset where attack windows include normal traffic.



SCAIR: Satellite Cyber AI Range – Lab Constellation Topology

References

[1]Yasamin Fayyaz, et al. CubeSat cybersecurity dataset for intrusion detection (CuCD-ID): Labelled NOS3/cFS telemetry (raw + augmented) with COSMOS reproduction scripts (2026)

Oak Ridge National Laboratory | Oak Ridge Institute for Science and Education | U.S. Department of Energy

This work was supported in part by the U.S. Department of Energy, Office of Science, Next Generation STEM Internship Program (NGSI), hosted at Oak Ridge National Laboratory (ORNL) and administered by the Oak Ridge Institute for Science and Education (ORISE).