

Performance Evaluation of Machine Learning Models Using IODA Data

Mohid Hasan

Student ID: 301588534

School of Engineering Science

Simon Fraser University

Under the supervision of Prof. Ljiljana Trajković

Abstract—Internet anomalies may result from malware outbreaks, infrastructure failures, and conflict-driven connectivity disruptions. Early detection is important for network stability, cybersecurity visibility, and resilience during crises. This project evaluates eight machine learning models for anomaly detection on Internet Outage Detection and Analysis (IODA) outage datasets using a sliding-window framework. Two country-level case studies are considered: Iran and Cuba. The models are compared using ROC-AUC, PR-AUC, Accuracy, F1-score, confusion-matrix-based counts, Precision, Sensitivity, and training time. Results show that temporal models provide strong detection capability, while Broad Learning System variants offer competitive performance at very low training cost.

I. INTRODUCTION

Internet connectivity disruptions are important indicators of cyber incidents, infrastructure failures, and policy-driven shutdowns. Examples include malware outbreaks, large infrastructure failures, and conflict-driven connectivity loss. Detecting these anomalies early is important for operational awareness and cyber defense.

The Internet Outage Detection and Analysis (IODA) platform provides a multi-signal view of Internet connectivity using routing visibility, active probing, and traffic-related measurements. In this project, IODA outage data is used to evaluate the performance of multiple machine learning models under a unified anomaly detection framework.

The main goal is to detect anomalies in IODA outage datasets using machine learning, with case studies based on Iran and Cuba.

II. OBJECTIVE AND CONTRIBUTIONS

The objectives of this work are:

- to apply eight machine learning models for anomaly detection on IODA data;
- to use a sliding-window framework to capture temporal dependencies;
- to evaluate performance using ROC-AUC, PR-AUC, Accuracy, F1-score, TP/FP/FN/TN, Precision, Sensitivity, and training time; and
- to compare deep sequence models, broad learning systems, and a tree-based baseline.

The key contributions are:

- an eight-model benchmark suite for IODA datasets with consistent preprocessing;

- sliding-window-based anomaly detection on outage signals;
- updated and extended code support for CNN, RNN-LSTM, and RNN-Bi-LSTM on IODA data;
- per-model optimal F1-score threshold selection; and
- full confusion-matrix and metric reporting.

III. DATASETS

Two country-level datasets are used, both spanning Dec. 2025 to Mar. 2026, with 2,161 hourly samples each. Each dataset contains five normalized IODA signals:

- Google Search,
- BGP Routing,
- Mozilla City Count,
- Telescope Unique IPs,
- Active Probing.

Each time step is represented as

$$x_t = [\text{Google}_t, \text{BGP}_t, \text{Mozilla}_t, \text{Telescope}_t, \text{Probing}_t].$$

A. Iran Dataset

Iran contains two total Internet shutdowns: Jan. 8–26 and Feb. 28–Mar. 19, 2026. The most significant drop is observed in Active Probing, which falls from approximately 97% to 3%. In total, 900 rows are labeled anomalous, corresponding to 41.6% of the dataset.

B. Cuba Dataset

Cuba contains five partial disruptions between Feb. and Mar. 2026. Active Probing drops to about 81–88%, with a baseline already degraded to 94–98%. In total, 103 rows are labeled anomalous, corresponding to 4.8% of the dataset.

IV. METHODOLOGY

A. Evaluated Models

Eight models are evaluated:

Deep models

- **CNN**: Convolutional Neural Network; extracts local temporal patterns.
- **RNN-GRU**: Gated Recurrent Unit; provides strong balanced performance.
- **RNN-LSTM**: Long Short-Term Memory; captures long-range dependencies.

- **RNN-Bi-GRU**: Bidirectional GRU; processes sequences in both directions.
- **RNN-Bi-LSTM**: Bidirectional LSTM; strongest deep learning model on Cuba.

Analytical models

- **BLS**: Broad Learning System; closed-form solution without gradient descent.
- **VFBS**: Variant-Feature Broad Learning System with $\tanh$ activation.

Tree-based baseline

- **GBDT-LightGBM**: Gradient Boosted Decision Trees using flattened windows.

B. Processing Pipeline

The processing steps are:

- 1) Load and sort the data by timestamp.
- 2) Label disruption windows: hours inside verified outages are assigned LABEL = 1, and all others LABEL = 0.
- 3) Impute missing values using the median of each column, then normalize features using Z-score normalization.
- 4) Perform feature selection using `ExtraTreesClassifier` to retain informative signals.
- 5) Create sliding windows of size $W = 10$ hours; each window label is taken from the final hour.
- 6) Split the data chronologically into 70% training and 30% testing without shuffling.

C. Sliding-Window Representation

Using a window size of $W = 10$, the input samples are formed as

$$[x_1, \dots, x_{10}], [x_2, \dots, x_{11}], \dots, [x_{N-9}, \dots, x_N].$$

Each window is assigned the label of its final hour.

D. Threshold Selection

For probabilistic classifiers, a decision threshold τ is selected to maximize the F1-score:

$$\hat{y} = \begin{cases} 1, & p \geq \tau \\ 0, & p < \tau \end{cases}$$

V. PERFORMANCE METRICS

The following metrics are used:

- ROC-AUC,
- PR-AUC,
- Accuracy,
- F1-score,
- Precision,
- Sensitivity (Recall),
- Training Time,
- TP, FP, FN, and TN.

ROC-AUC measures discrimination across thresholds, while PR-AUC is particularly informative when anomalies are rare. Precision quantifies the correctness of flagged anomalies, and Sensitivity measures the proportion of true anomalies detected.

VI. RESULTS AND DISCUSSION

The framework generalizes across countries, disruption severities, and time periods. Signal severity strongly affects task difficulty: total shutdowns in Iran are easier to detect, whereas subtle degradations in Cuba better reveal true model differences.

Among deep learning models, RNN-GRU and RNN-Bi-LSTM provide the strongest and most balanced performance across datasets. CNN underperforms on shorter and subtler disruptions, especially when training data is limited. BLS and VFBS consistently achieve near-top accuracy with training times under 0.1 seconds, making them strong candidates for real-time deployment. The sliding-window approach with window size 10 hours successfully captures temporal disruption patterns across all datasets.

Overall, temporal models outperform purely static approaches, multi-signal fusion is essential, and threshold tuning has a significant effect on F1-score.

VII. CONCLUSION

This project evaluated eight machine learning models for anomaly detection using IODA outage datasets from Iran and Cuba. The results show that severe outages are relatively easy to detect, while subtle degradations require stronger temporal modeling. RNN-GRU and RNN-Bi-LSTM provide the best overall deep learning performance, whereas BLS and VFBS offer highly efficient alternatives with minimal training cost. The proposed sliding-window framework is effective for detecting temporal connectivity anomalies across different countries and outage conditions.

ACKNOWLEDGMENT

The author acknowledges the supervision and guidance of Prof. Ljiljana Trajković.

REFERENCES

- [1] Z. Li and L. Trajković, "Enhancing cyber defense," in *Proc. IEEE SMC*, 2023.
- [2] L. F. Oliveira *et al.*, "Internet outages during times of conflict," in *Proc. IEEE SMC*, 2024.
- [3] Z. Li *et al.*, "Machine learning for anomaly detection," *IEEE Journal on Selected Areas in Communications*.
- [4] CyberDefense GitHub. [Online]. Available: <https://github.com/zhida-li/CyberDefense>
- [5] RouteViews. [Online]. Available: <http://www.routeviews.org/>
- [6] Internet Outage Detection and Analysis (IODA). [Online]. Available: <https://ioda.inetintel.cc.gatech.edu/>
- [7] L. F. Oliveira, "paper_smc_2024" GitHub repository. [Online]. Available: https://github.com/luizsoliveira/paper_smc_2024
- [8] M. Hasan, "anomaly_detection" GitHub repository. [Online]. Available: https://github.com/mohid-hasan/anomaly_detection
- [9] Takhar *et al.*, SFU, 2024.