MSC RESEARCH POSTER | RESEARCH METHODS MODULE | DEPARTMENT OF COMPUTING

LSTM-based DDoS Detection in Telecom IoT: Evaluation on CIC IoT-DIAD 2023 Dataset

SETU
South East Technological University

STUDENT: Chioma Debbie Okoye
STUDENT NO.: C00326756
PROGRAMME: MSc Cybersecurity, Privacy & Trust
SUPERVISOR: Jason Baron
DATE: April 2026

105 Real IoT Devices in Dataset | **33** Attack Types Captured | **47M+** Labeled Network Flows | **7** Attack Categories (incl. DDoS) | **≥95%** Primary Target F1-Score | **16 wks** Research Timeline

1 Introduction to Research Area

Telecommunications IoT networks—spanning smart meters, 5G User Equipment, edge sensors, and environmental controllers—generate high-volume sequential traffic with heterogeneous protocols (MQTT, CoAP, TCP/UDP). This expanded connectivity dramatically widens the attack surface for Distributed Denial of Service (DDoS) attacks, which disrupt availability through volumetric floods, reflection amplification, and slow-rate exhaustion.

Manual Security Operations Centers (SOCs) cannot match machine-speed DDoS propagation: attacks achieve network-wide compromise in **10–30 seconds** while human triage requires **5–15 minutes**. This gap demands automated, data-driven detection.

Long Short-Term Memory (LSTM) neural networks process sequential data using gated memory cells, making them theoretically ideal for capturing DDoS staging (scan → amplification → saturation). This research evaluates LSTM for binary DDoS detection against a Random Forest baseline using the **CIC IoT-DIAD 2023** dataset.

5G IoT Security | Deep Learning | CIC IoT-DIAD 2023 | Telecom SOC

2 Research Questions & Hypotheses

RQ1 — DETECTION PERFORMANCE
How accurately can an LSTM model detect DDoS attacks in CIC IoT-DIAD 2023 telecom IoT traffic using precision, recall, F1-score, and ROC-AUC?

RQ2 — BASELINE COMPARISON
Does LSTM outperform a Random Forest baseline for DDoS detection under the same 5-fold cross-validation setting?

RQ3 — FEATURE SELECTION
Which reduced feature subset, identified using Recursive Feature Elimination (RFE), gives the best balance between detection performance and model simplicity?

Hypothesis: An LSTM model using a reduced RFE-selected feature subset will outperform Random Forest on DDoS detection in CIC IoT-DIAD 2023, showing the value of sequential modelling for telecom IoT traffic.

3 Literature Review Summary

MODEL/STUDY	F1	GAP
Bi-LSTM + Attention *CIC IoT 2023*	99.3%	Generic features
Random Forest *CIC benchmarks*	98–99%	Stateless
CNN-LSTM Hybrid *Flow features*	96–98%	Compute heavy
LSTM-Autoencoder *Anomaly detect*	94–97%	No telecom eval

Three Focused Gaps Identified:

1. **Limited Telecom Context:** Many prior studies use general IoT feature sets and do not explicitly evaluate telecom-style traffic characteristics such as Flow Bytes/s saturation and connection asymmetry.
2. **Limited Baseline Comparison:** Few studies compare LSTM directly with a strong traditional baseline such as Random Forest under the same validation setting.
3. **Limited Feature Optimisation:** Few studies identify a smaller, practical feature subset that preserves DDoS detection performance while reducing model complexity.

4 Research Methodology

Design: Quantitative experimental — CRISP-DM on public secondary data (CIC IoT-DIAD 2023, ~47M flows, DDoS subset).

P1 Data Preparation
Weeks 1–3
Download CIC 2023 DDoS CSV → extract binary labels → select 12 candidate features (Flow Duration, Fwd/Bwd Bytes/s, Pkt Len Stats, Flag Counts) → SMOTE 1:1 → StandardScaler → 80/20 stratified split.

P2 Model Training
Weeks 4–7
LSTM: 50 units, Dropout(0.2), Adam (lr=0.001), early stopping (p=5), binary crossentropy. RF baseline: n=100, max_depth=10. 5-fold stratified CV, batch=64, 20 epochs max.

P3 Evaluation & Feature Selection
Weeks 8–11
Metrics: F1, Precision, Recall, ROC-AUC, Confusion Matrix. Paired t-test (p<0.05) on CV F1-scores. RFE (k=5–8) identifying optimal feature subset.

P4 Analysis & Write-up
Weeks 12–16
Supporting interpretability analysis on the final model. Ablation validation on RFE subset. GitHub repo (notebooks + requirements.txt). Dissertation chapters drafted.

ML Pipeline:

CIC IoT 2023 CSV Download → DDoS Subset Extract → SMOTE + Scaler → LSTM Training → RF Baseline → RFE + analysis → F1 / ROC Results

5 Technologies, Tools & LSTM Architecture

Python 3.11 | TensorFlow 2.15 | Scikit-learn 1.4 | JupyterLab | SHAP | GitHub | imbalanced-learn | Matplotlib

LSTM Model Architecture:

Telecom-Relevant Features (12 Candidate):

Flow Bytes/s	Flow Duration
Fwd/Bwd Pkts Ratio	SYN Flag Count
Pkt Len Mean/Std	ACK / PSH Flags

6 Target Metrics & Expected Outcomes

≥0.95 F1-SCORE (DDOS CLASS)	**≥0.95** ROC-AUC
≥0.90 RECALL (CATCH RATE)	**5–8** EXPECTED RFE FEATURES

Remaining 8-Week Plan:

- **WEEKS 1–3** Dataset download, DDoS subset extraction, preprocessing (SMOTE, scaling), feature candidate selection
- **WEEKS 4–7** LSTM model construction, RF baseline training, 5-fold CV execution, early stopping validation
- **WEEKS 8–11** F1/ROC-AUC evaluation, paired t-test statistical comparison, RFE feature selection (k=5–8)
- **WEEKS 12–16** SHAP interpretability, ablation analysis, GitHub repo finalisation, dissertation write-up & submission

Early Indications & Next Steps:

- **Confirmed:** CIC IoT-DIAD 2023 is publicly available from UNB with 47M+ labeled flows across 105 devices and 33 attacks — dataset is accessible and well-documented.
- **Immediate:** Download CIC IoT 2023 Flow CSV files and verify DDoS subset completeness. Set up Jupyter + TensorFlow 2.15 environment on Google Colab (GPU).
- **Phase 1 Work:** Conduct exploratory data analysis on DDoS flows to confirm class imbalance ratio and validate 12 candidate features against dataset columns.
- **Anticipated finding:** LSTM is expected to show measurable improvement over RF on sequential DDoS patterns, while RFE may reduce the feature set to a smaller practical subset with limited performance loss.
- **Contribution:** A focused telecom-context benchmark on CIC 2023 with RFE analysis — informing SOC practitioners on model selection and practical feature engineering.

KEY REFERENCES

- Al-Hawawreh et al. (2024) LSTM-based IoT attack detection, CIC 2023. *Computers & Security*, 132.
- Canadian Institute for Cybersecurity (2023) *CIC IoT-DIAD 2023 Dataset*. University of New Brunswick.
- Ngo et al. (2023) Bi-LSTM with attention for IoT DDoS classification. *J. Information & Intelligence*.
- ENISA (2023) *Threat Landscape for 5G & IoT Critical Infrastructure*. European Union Agency for Cybersecurity.
- Empl et al. (2022) SOAR4IoT: Securing IoT with digital twins. *Proc. IEEE S&P*.