Grenze International Journal of Engineering and Technology, June Issue

Adversarial Robustness of URLLC Traffic Classifiers using Network KPI Features

Rajani P Kurup[1] and Dr. Manoj Kumar T K[2]
[1-2]School of Digital Sciences, Kerala University of Digital Sciences, Innovation and Technology, Trivandrum, India
Email: rajani.dsres22@duk.ac.in, manojtk@duk.ac.in

***Abstract*— Ultra-Reliable Low-Latency Communication (URLLC) services rely on precise monitoring and timely prediction of network Key Performance Indicators (KPIs). In recent years, machine learning models have been widely adopted to interpret these KPIs, enabling automatic classification of network conditions and early detection of reliability degradation. However, an important concern is that such models can be sensitive to subtle, intentional perturbations in the input data. In this paper, we examine the adversarial robustness of a KPI-driven URLLC traffic classifier built using the Unicorn-Genesys dataset. The raw experiment logs are first converted into compact feature representations by aggregating KPIs through statistical measures such as mean, standard deviation, and maximum values. These features are then used to train a multilayer perceptron (MLP) model for reliability prediction. To assess robustness, we subject the classifier to adversarial perturbations generated using the Fast Gradient Sign Method (FGSM) and Projected Gradient Descent (PGD). Our findings indicate that the classifier performs strongly under normal conditions, achieving close to 98% accuracy. However, even small adversarial perturbations lead to noticeable performance degradation, with iterative PGD attacks having the most severe impact. We further explore KPI-level sensitivity using a vulnerability map, which highlights the measurements most prone to manipulation. Adversarial training is also explored to enhance model robustness against attacks. Overall, the results point to potential security risks in data-driven network management and emphasize the need for more robust training and monitoring strategies.**

***Index Terms*— 5G NR, URLLC, Adversarial machine learning, network KPIs, robustness analysis, wireless network intelligence.**

I. INTRODUCTION

Modern cellular networks rely increasingly on data-driven intelligence. Key Performance Indicators (KPIs) are standardized metrics that describe the performance of the connection between the user equipment (UE) and the base station (gNB in 5G). These metrics—ranging from signal quality and buffer occupancy to modulation schemes and resource allocation—are continuously collected through network telemetry systems [1]. When processed effectively, they allow machine learning models to infer network conditions and anticipate reliability outcomes.

For URLLC services, reliability prediction is particularly important. Applications such as industrial automation, autonomous driving, and mission-critical control require extremely low packet error rates and strict latency guarantees.

Grenze ID: 01.GIJET.12.2.319_26

Even minor fluctuations in radio conditions can disrupt performance. As a result, operators are increasingly turning to machine learning (ML) models that can recognize patterns in KPIs and flag potential reliability issues before they escalate.

However, recent research in adversarial machine learning has demonstrated that neural networks are not inherently secure. They can be manipulated through carefully crafted input perturbations [2]. These changes may be extremely small, yet they can significantly alter model predictions. While such attacks have been widely studied in domains like computer vision, their impact on network telemetry analysis remains largely unexplored.

In this work, we focus on understanding how vulnerable KPI-based URLLC classifiers are to such adversarial manipulation. The main contributions of this study are as follows:

- We develop a KPI-based URLLC reliability classifier using experiment logs from the Unicorn Genesys dataset [3].
- We evaluate its robustness under adversarial conditions using FGSM and PGD attacks.
- We analyse KPI-level vulnerabilities through sensitivity analysis and attack heatmaps.
- We demonstrate adversarial training effectiveness for network security.

KPIs capture key aspects of network behavior and help in understanding traffic patterns. The results show that even small perturbations in KPI values can influence reliability predictions. These findings raise important considerations for the safe deployment of ML in network management systems.

II. Related Work

Recent years have witnessed significant interest in applying machine learning techniques to wireless network analytics. Alekseeva et al. [4] evaluated several machine learning models for traffic prediction in operational wireless networks, demonstrating the effectiveness of data-driven approaches for forecasting network KPIs. Similarly, Chen et al. [5] provide a comprehensive overview of neural network applications in wireless communications, highlighting their potential for intelligent network optimization and monitoring.

In parallel, the field of adversarial machine learning has revealed vulnerabilities in deep learning models. Goodfellow et al. [6] introduced the Fast Gradient Sign Method (FGSM), showing that small input perturbations can drastically alter neural network predictions. Madry et al. [7] later proposed Projected Gradient Descent (PGD), a stronger iterative attack widely used for evaluating adversarial robustness.

Despite extensive work in adversarial machine learning [8] [9], its application to wireless network telemetry remains relatively unexplored. Existing wireless ML studies have primarily focused on traffic prediction, resource allocation, or network optimization rather than adversarial robustness [10] [11]. For instance, reinforcement learning based resource allocation techniques have been proposed for vehicular communication systems, demonstrating the potential of AI-driven network management [12] [13]. Similarly, edge intelligence frameworks integrate machine learning into communication infrastructures to cater intelligent networking functions [14]. Theoretical foundation for adversarial attacks in deep learning models is established [15] and models proposed for performance study and simulation. Although prior studies have investigated adversarial vulnerabilities in wireless signal classification [16] and applications like V2X (Vehicle to Everything) [17]; limited work has examined the adversarial robustness of URLLC traffic classifiers, particularly using real communication datasets. Earlier works also typically assume trustworthy telemetry inputs.

In contrast, the present study investigates how adversarial manipulation of KPI measurements can affect URLLC classifiers reliability. It investigates adversarial robustness in network measurements, where manipulation could potentially mislead intelligent network management systems. Unlike prior works focusing on signal-level adversarial attacks, this work investigates feature-space adversarial robustness.

III. Adversarial Attack Methodology

The adversarial attacks used here are Fast Gradient Sign Method (FGSM) and Projected Gradient Descent (PGD), which are Gradient-Based, White-Box Evasion Attacks. They occur during the inference phase. The model is already trained and "frozen." The attacker modifies the input data to trick the model into a wrong classification. This needs knowledge of the model's internal architecture, specifically the gradients. In the primary form, the attacks simply move the data point away from the correct class, regardless of what the new (incorrect) class is. From a security perspective, this problem can be interpreted through the STRIDE threat model, originally introduced by Microsoft. STRIDE categorizes threats into spoofing, tampering, repudiation, information disclosure, denial of service and elevation of privilege. In the context of URLLC traffic classification, adversarial perturbations align most closely with tampering attacks.

A. FGSM Attack

The Fast Gradient Sign Method generates adversarial examples by perturbing input features in the direction of the gradient of the loss function. FGSM is a "one-shot" attack designed for speed rather than subtlety. It takes a single step in the direction of the gradient to maximize the loss. Given an input vector x, the adversarial example is generated as:

$$x\text{adv} = x + \epsilon \cdot sign(\nabla \text{x}\, J(\theta, x, y)) \quad (1)$$

where
x = original KPI feature vector
$x$adv = adversarial example
$J(\theta, x, y)$ = model loss function
θ = model parameters
ϵ = perturbation magnitude / attack strength
∇x = gradient with respect to input features
In the experiments, epsilon values between 0.03 and 0.3 are used to simulate realistic measurement distortions and y is the true class label (0 = normal, 1 = failure).

B. Projected Gradient Descent

Projected Gradient Descent is widely regarded as a stronger attack since it iteratively searches for adversarial inputs that maximize model error. It extends FGSM by performing multiple gradient steps - a series of small, calculated steps. After every step, the data is "projected" back into a specified range (the epsilon-ball) to ensure the change remains small and "invisible" to a human or a secondary validator:

$$x_{adv}^{t+1} = \Pi_{B\epsilon(x)}(x_{adv}^{t} + \alpha \cdot sign(\nabla_x J(\theta, x_{adv}^{t}, y))) \quad (2)$$

where
x_{adv}^{t} = adversarial sample at iteration ttt
α = step size
$\Pi_{B\epsilon(x)}$ = projection operator restricting the perturbation within the L_∞ ball
$J(\,.\,)$ = loss function.
Hence PGD performs multiple iterative perturbations while constraining the perturbation within epsilon-ball around the original input.

C. Defense Mechanism

Adversarial training is employed by augmenting the training dataset with FGSM-generated perturbed samples. This enables the classifier to learn robust feature representations and improves resilience against adversarial attacks. The training objective is formulated as a min–max optimization problem:

$$\min_{\theta} \mathbb{E}_{(x,y)} \left[\max_{\|\delta\|_\infty \le \epsilon} \mathcal{L}(f_\theta(\mathbf{x} + \delta), y) \right] \quad (3)$$

where δ denotes adversarial perturbations constrained within an ℓ_∞-ball of radius ϵ. This formulation enforces robustness by optimizing model parameters against worst-case perturbations during training.
Unlike prior work that applies unconstrained perturbations in normalized feature space, this study adopts a KPI-aware adversarial framework. The key idea is to enforce domain-specific constraints during attack generation. This ensures that perturbations remain physically meaningful, such as respecting limits on signal quality, buffer levels and resource allocation.
As a result, the approach better reflects real wireless network behavior rather than purely theoretical settings. It bridges the gap between adversarial ML theory and practical deployment. This is especially important for URLLC systems, where unrealistic perturbations could otherwise lead to misleading robustness conclusions.

IV. Dataset and Feature Engineering

A. Unicorn Genesys URLLC Dataset

The experiments in this study are based on the Unicorn Genesys dataset containing URLLC experiment logs [3]. It is created by collecting real-world 5G traffic from smartphone applications and this captured traffic was replayed through the emulator (an O-RAN digital twin) to generate specific network KPIs. Each log file records

time-series measurements of multiple network KPIs captured during controlled network experiments. For each experiment file, the following KPIs are extracted:

- Downlink modulation and coding scheme (dl_mcs)
- Downlink buffer size (dl_buffer)
- Downlink bitrate (tx_brate)
- Downlink packet counts (tx_pkts)
- Downlink packet error percentage (tx_errors)
- Downlink channel quality indicator (dl_cqi)
- Uplink modulation and coding scheme (ul_mcs)
- Uplink buffer size (ul_buffer)
- Uplink bitrate (rx_brate)
- Uplink packet counts (rx_pkts)
- Uplink packet error percentage (rx_errors)
- Uplink RSSI (ul_rssi)
- Uplink SINR (ul_sinr)
- Requested physical resource blocks (requested_prbs)
- Granted physical resource blocks (granted_prbs)
- Power headroom report (phr)
- Slice PRB allocation (slice_prb)
- Power multiplier

The dataset consists of 491 experiment files. Each file contains a time sequence of measurements that capture the dynamic behaviour of the network during the experiment. The complete KPI dataset, which comprises 491 KPI traces equally distributed among six distinct application classes, amounts to 85 MB of data. Each trace is a rectangular matrix of N time samples collected from 18 KPIs. The lengths of the traces vary, averaging to approximately 1500-time samples per trace.

B. Feature Aggregation

Instead of using raw time-series data, we convert each experiment log into a fixed-length feature vector. For every KPI, three statistical descriptors are computed: Mean value μ, Standard deviation σ & maximum value.
This aggregation reflects both the overall trend and the variation in the measurements. It also ensures that information from one time segment does not unintentionally leak into the training or testing data. The resulting feature vector contains: 18 KPIs × 3 statistics = 54 features.

C. Reliability Label Generation

To construct a classification task, each experiment file is assigned a reliability label based on the average uplink packet error percentage. Experiments with error rates above the dataset median are labelled as reliability degradation events, while the rest are labelled as normal operation. The resulting dataset contains 491 samples with a natural class imbalance, where failure events constitute approximately 8% of the observations.

D. Threat Mapping

Threat mapping for various KPIs is shown in table I. By subtly manipulating KPI values, an attacker could mislead the classifier and potentially disrupt latency-critical network operations.

TABLE I. STRIDE THREAT MAPPING FOR NETWORK KPI GROUPS

KPI Group	Primary Category	Threat Manifestation
PRB / Slice Metrics	Denial of Service	Resource Exhaustion / Slicing starvation
Error Rates / SINR	Denial of Service	RF Jamming / Signal Interference
MCS / CQI / PHR	Tampering	Link Adaptation manipulation / Power exhaustion
Throughput /Packet Counts	Information Disclosure	Traffic Analysis / User Profiling

Here is the detailed classification of the KPI columns into STRIDE categories:
Denial of Service (DoS) - These metrics track resource exhaustion, interference, and signal degradation. Sudden spikes or drops here usually indicate a jamming attack or a resource exhaustion attack (e.g., a "Signaling Storm").

- requested_prbs / granted_prbs: High requested vs. low granted PRBs indicates resource starvation or a DoS on the scheduler.

- dl_buffer / ul_buffer [bytes]: Large buffer sizes suggest the network is unable to clear traffic, potentially due to a flood or exhaustion attack.
- tx_errors / rx_errors (%): High error rates are a classic sign of RF jamming or intentional interference.
- ul_sinr / ul_rssi: If the Signal-to-Interference-plus-Noise Ratio (SINR) drops while RSSI (Received Signal Strength Indicator) stays high, it strongly suggests a malicious jammer is present.
- slice_prb: If a specific network slice is starved of PRBs, it indicates a DoS targeting a specific service (like URLLC).

Tampering - Tampering in a wireless context often involves manipulating the power control or link adaptation logic to degrade performance or force a handover to a rogue base station.

- power_multiplier / phr: Manipulation of these can force a UE to drain its battery or interfere with other users by over-transmitting.
- dl_mcs / ul_mcs : If a device is forced to a lower MCS despite good signal, an attacker may be tampering with the link adaptation feedback to throttle throughput.
- dl_cqi: Attackers can spoof/tamper with CQI reports to trick the base station into allocating resources inefficiently.

Information Disclosure (Side-Channel) - These metrics don't contain "cleartext" passwords, but they can be used for traffic pattern analysis—a form of information disclosure where an attacker deduces what a user is doing based on metadata.

- tx_brate / rx_brate [Mbps]: Patterns in bitrates can reveal the type of application being used (e.g., a video call vs. a banking transaction).
- tx_pkts / rx_pkts: Packet counts and timing can be used for fingerprinting specific user behaviors or identifying the presence of a target user in a cell.

V. URLLC TRAFFIC CLASSIFICATION MODEL

A supervised learning-based URLLC traffic classification framework is considered, where network KPIs are used to infer traffic states under normal and adversarial conditions.
Let: $\mathsf{K}(t) = [k_1(t), k_2(t), \ldots, k_d(t)]$ represent the KPI vector at time t, where d is the number of KPIs.
Each trace is aggregated into a feature vector:

$$\mathsf{x} = [\mu(\mathsf{K}), \sigma(\mathsf{K}), \max(\mathsf{K})] \in \mathbb{R}^{3d} \quad (4)$$

The classifier learns a mapping:

$$f_\theta : \mathbb{R}^{3d} \rightarrow \{0,1\}$$

where θ are model parameters and the output represents traffic class (normal / anomalous).
This aims to enable real-time classification of known URLLC traffic without accessing sensitive user data or requiring user equipment (UE) interaction. By looking at the relationship between PRBs, MCS, Packet Counts, etc. the AI classifier tells if the required service is mission critical.

A. Neural Network Architecture

A multilayer perceptron (MLP) is used as the baseline classifier. The network architecture is shown in Fig.1. The network is trained using the Adam optimizer for 40 epochs with a mini-batch size of 32. For an input KPI feature vector x ∈ Rd, the multilayer perceptron computes the output through successive nonlinear transformations:

$$h^{(l)} = \phi\left(W^{(l)} h^{(l-1)} + b^{(l)}\right) \quad (5)$$

where
$h^{(l)}$ = hidden representation at layer l
$W^{(l)}$ = weight matrix
$b^{(l)}$ = bias vector
$\phi(\cdot)$ = ReLU activation function
The final classifier output probability is obtained using softmax:

$$\hat{y} = softmax\ (W_o h^{(L)} + b_o) \quad (6)$$

where yˆ represents the predicted class probability.

B. Training and Evaluation

The dataset is divided using an 80–20 hold-out split. Feature normalization is applied using z-score scaling.
Under clean conditions, the trained classifier achieves approximately 98% classification accuracy. Precision, recall, and F1-score metrics confirm that the model effectively distinguishes reliability degradation events despite the dataset imbalance.
We then model an adversary that perturbs input KPIs:

$$\mathbf{x}_{adv} = \mathbf{x} + \delta \text{ subject to: } \| \delta \|_{\infty} \leq \epsilon. \quad (7)$$

Fig. 2. illustrates the confusion matrix for the clean model predictions and after adversarial manipulations of input.
Adversarial training is implemented by generating FGSM-based perturbed samples during training and augmenting the original dataset with these adversarial examples. The model is then trained on a mixture of clean and adversarial inputs to improve robustness.

VI. Experimental Results

The URLLC traffic classifier achieved 98% accuracy under clean conditions, indicating that the model effectively learns the underlying traffic characteristics of the Unicorn Genesys dataset. However, on introducing adversarial perturbations, the model's robustness varied depending on the attack strength.
The FGSM attack caused only a marginal degradation, reducing the accuracy from 0.98 to 0.969. This suggests that the classifier exhibits a certain degree of resilience to single-step gradient perturbations. In contrast, the PGD attack resulted in a significant performance drop to 0.796 accuracy. Since PGD applies iterative perturbations, it is a stronger adversarial attack.
The results indicate that such strong manipulations may lead to incorrect traffic classification in critical network scenarios.

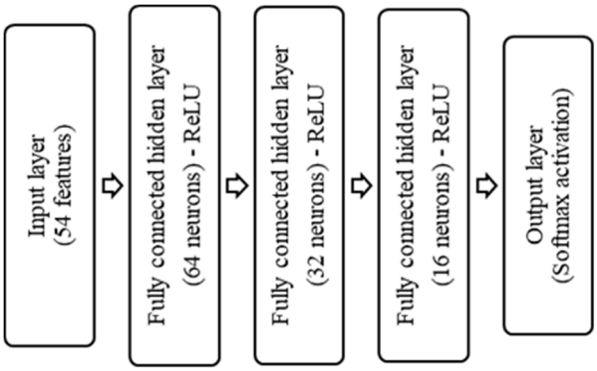

Figure 1. MLP architecture of the URLLC traffic classifier

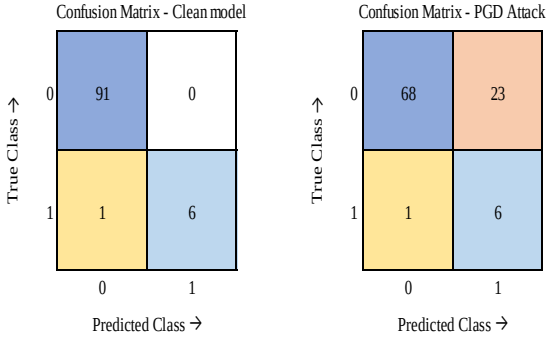

Figure 2. Confusion Matrix of Classifier under different scenarios

A. Classification Performance

Table II summarizes the classification performance under different conditions. We define Robustness Degradation (RD) metric as:

$$RD = \frac{Acc_{clean} - Acc_{attack}}{Acc_{clean}} \quad (8)$$

TABLE II. Impact of Adversarial Attacks on URLLC Traffic Classification Accuracy

Scenario	Accuracy	Robustness Degradation %
Clean data	0.980	0.00
FGSM attack	0.969	1.12
PGD attack	0.769	18.80

The results show that single-step adversarial perturbations have a limited effect on model accuracy. However, iterative attacks cause significant performance degradation. To evaluate model robustness, classification accuracy is measured across increasing perturbation strengths. Accuracy gradually decreases as epsilon increases, confirming that larger perturbations lead to higher misclassification rates.

B. Random Noise vs Adversarial Noise

To differentiate adversarial effects from random measurement noise, Gaussian noise with varying standard deviations is applied to the input features. Fig. 3. compares classifier performance under random noise and FGSM. The results show that adversarial perturbations degrade performance more effectively than random noise of comparable magnitude.

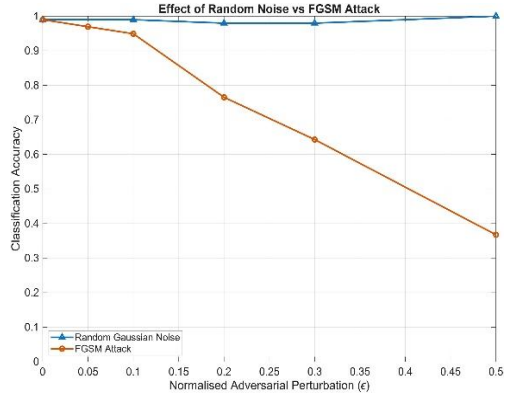

Figure 3. Random Noise vs FGSM Attack Impact

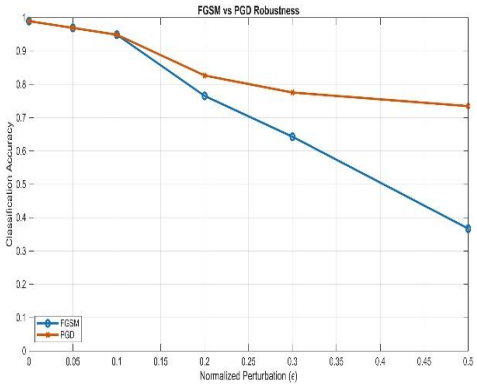

Figure 4. FGSM vs PGD Robustness

C. FGSM vs PGD Comparison

Fig. 4. compares the robustness curves for FGSM and PGD attacks. The PGD attack causes significantly larger accuracy degradation, indicating that iterative perturbations can more effectively exploit model vulnerabilities.

D. KPI Vulnerability Analysis and Heatmap

To understand which KPIs contribute most to adversarial vulnerability, we analyze the magnitude of adversarial perturbations across features. Fig.5. shows the aggregated sensitivity of each KPI. Radio-layer measurements such as SINR, RSSI and PRB allocation exhibit higher sensitivity compared to traffic-level metrics.

KPI attack map is constructed to visualize adversarial sensitivity across aggregated features and is depicted in Fig.6. The heatmap highlights how adversarial perturbations propagate across mean, variance and maximum KPI statistics. The results indicate that signal-quality measurements and resource allocation parameters are particularly susceptible to manipulation.

E. Adversarial training defence

The adversarial-trained model maintains significantly higher classification accuracy, under increasing perturbation compared to the baseline model as shown in table III.

The results demonstrate improved robustness against FGSM attacks. Deep learning based KPI classifiers are highly vulnerable to adversarial perturbations, even at low magnitudes. By incorporating adversarial samples during training, the classifier learns invariant, robust feature representations that remain more stable. This leads to a more gradual degradation in performance under increasing perturbation strength.

TABLE III. PERFORMANCE COMPARISON OF BASELINE AND ADVERSARIALLY TRAINED MODELS

Epsilon	Baseline	Defended
0	0.9898	1
0.05	0.9693	0.9898
0.1	0.9489	0.9898
0.2	0.7653	0.9898
0.3	0.6428	0.9898
0.5	0.3673	0.9796

VII. DISCUSSION

The results show that KPI-based URLLC classifiers achieve high performance under clean conditions. The MLP model attains ~0.98 accuracy, with strong precision and recall. The confusion matrix indicates minimal errors, confirming that aggregated KPI features capture network behavior effectively. However, this performance degrades under adversarial conditions

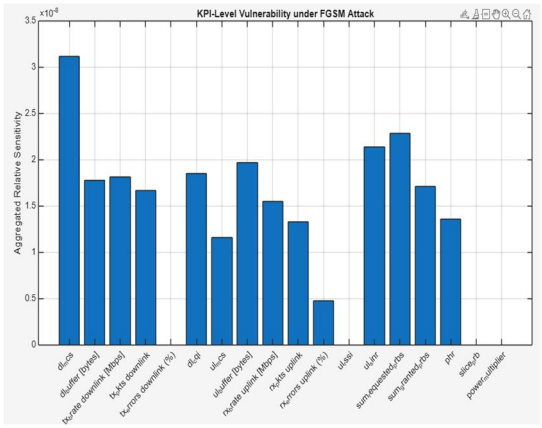

Figure 5. KPI Sensitivity under FGSM Attack

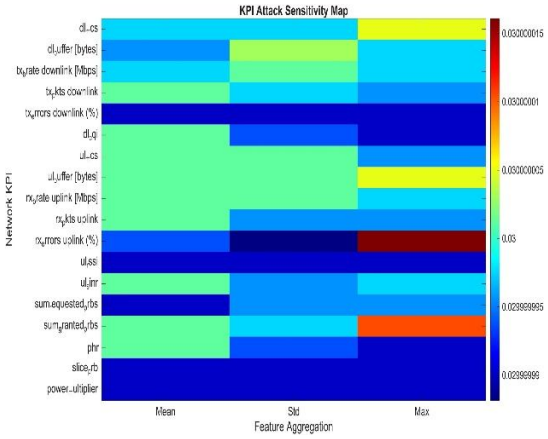

Figure 6. KPI Attack Sensitivity Map

To isolate adversarial effects, Gaussian noise was introduced. Under random noise, accuracy degrades gradually and remains relatively stable. In contrast, adversarial perturbations cause targeted degradation. This confirms that the impact is driven by structured, gradient-based attacks rather than stochastic variations.

FGSM has a limited effect, indicating robustness to single-step perturbations. PGD causes a larger degradation, reflecting higher sensitivity to iterative attacks. However, accuracy alone does not fully capture this impact. Under clean conditions, failure detection achieves precision of 1.0 and recall of 0.8. Under FGSM, these drop to (0.75, 0.6), and further to (0.6, 0.6) under PGD. This indicates a clear increase in missed failures.

KPI sensitivity analysis shows that physical-layer metrics are more vulnerable than traffic-level features. Signal quality indicators such as SINR and CQI, along with resource-related KPIs, have the highest impact. Perturbations affect all statistical descriptors, including mean, variance, and maxima.

From a security perspective, these vulnerabilities align with tampering and denial-of-service threats. Manipulated KPI inputs can lead to incorrect classification and suboptimal resource allocation. In URLLC systems, such errors directly affect reliability and latency guarantees.

Adversarial training improves robustness across perturbation levels. The defended model shows more stable performance with increasing ε, while maintaining near-clean accuracy. This highlights a practical trade-off between robustness and accuracy. Overall, the results demonstrate that accuracy alone is insufficient. Reliability-oriented metrics, particularly recall and false negatives, are critical for evaluating URLLC robustness under adversarial conditions.

VIII. Conclusions

This paper investigated the adversarial robustness of KPI-based URLLC traffic classifiers using the Unicorn Genesys dataset. Time-series KPI traces were aggregated into statistical feature representations and used to train an MLP classifier for reliability prediction. The model achieved high performance under clean conditions (~0.98 accuracy), showing that aggregated KPI features effectively capture network behavior.

However, the performed adversarial evaluation exposes important shortcomings. While accuracy degradation under FGSM and PGD attacks is moderate, class-wise analysis shows a drop in failure detection performance. Recall decreases from 0.8 (clean) to 0.6 under adversarial perturbations, showing an increase in missed reliability violations. This highlights that accuracy alone is insufficient to assess robustness in URLLC type systems. Reliability-oriented metrics such as recall, and false negatives add more meaningful evaluation.

The study also shows that adversarial perturbations are more effective than random noise, confirming that structured, gradient-based attacks exploit model vulnerabilities. KPI sensitivity analysis indicates that physical-layer and resource-related features are more susceptible to manipulation, emphasizing their influence on model decisions.

A KPI-aware adversarial framework with domain constraints is introduced, to enable more practical robustness evaluation. Adversarial training improves stability across perturbation strengths with minimal impact on clean accuracy, demonstrating a favorable robustness–accuracy trade-off.

Despite these contributions, several limitations remain. The attacks are applied in feature space and do not explicitly model signal-level distortions. The dataset size and aggregation may limit generalization across

diverse deployments. In addition, the defense strategy is based on FGSM and may not fully protect against stronger adaptive attacks.

Future work will focus on stronger attack models, adaptive defense mechanisms and real-time adversarial detection. Extensions to multi-class traffic classification, cross-layer attack modeling and validation on large-scale 5G/6G datasets will further enhance the applicability of the proposed approach.

ACKNOWLEDGMENT

The authors acknowledge the use of the AI-based language model ChatGPT to assist with drafting and language refinement of selected sections of this manuscript. All technical content, experimental results and interpretations were verified and finalized by the authors.

REFERENCES

[1] 3GPP, “NR; Physical Channels and Modulation,” 3GPP TS 38.211, Release 17, 2023.
[2] N. Carlini and D. Wagner, “Towards evaluating the robustness of neural networks,” in Proc. IEEE Symp. Security Privacy (SP), 2017.
[3] N. Soltani et al., “UNICORN: URLLC network traffic classification and OOD detection for O-RAN,” in Proc. IEEE Int. Conf. Mach. Learn. Commun. Netw. (ICMLCN), 2025.
[4] D. Alekseeva, N. Stepanov, A. Veprev, A. Sharapova, E. S. Lohan, and A. Ometov, “Comparison of machine learning techniques applied to traffic prediction of real wireless network,” IEEE Access, vol. 9, pp. 159495–159514, 2021.
[5] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, “Machine learning for wireless networks with artificial intelligence: A tutorial on neural networks,” IEEE Commun. Surveys Tuts., vol. 21, no. 4, pp. 3039–3071, 2019.
[6] I. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and harnessing adversarial examples,” in Proc. Int. Conf. Learn. Represent. (ICLR), 2015.
[7] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, “Towards deep learning models resistant to adversarial attacks,” in Proc. Int. Conf. Learn. Represent. (ICLR), 2018.
[8] Wang, Yulong, et al. "Adversarial attacks and defenses in machine learning-empowered communication systems and networks: A contemporary survey." IEEE Communications Surveys & Tutorials 25.4 (2023): 2245-2298.
[9] Adesina, Damilola, et al. "Adversarial machine learning in wireless communications using RF data: A review." IEEE Communications Surveys & Tutorials 25.1 (2022): 77-100.
[10] Su, Jian, et al. "Traffic prediction for 5G: A deep learning approach based on lightweight hybrid attention networks." Digital Signal Processing 146 (2024): 104359.
[11] Koursioumpas, Nikolaos, et al. "Network traffic anomaly prediction for beyond 5G networks." 2022 IEEE 33rd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). IEEE, 2022.
[12] H. Ye, G. Y. Li, and B. H. Juang, “Deep reinforcement learning based resource allocation for V2V communications,” IEEE Trans. Veh. Technol., vol. 68, no. 4, pp. 3163–3173, Apr. 2019.
[13] Habib, Md Arafat, et al. "Traffic steering for 5G multi-RAT deployments using deep reinforcement learning." 2023 IEEE 20th Consumer Communications & Networking Conference (CCNC). IEEE, 2023.
[14] X. Wang, Y. Han, C. Wang, Q. Zhao, X. Chen, and M. Chen, “In-edge AI: Intelligentizing mobile edge computing, caching and communication by deep learning,” IEEE Networking, vol. 33, no. 5, pp. 156–165, 2019.
[15] Y. E. Sagduyu, T. Erpek, and Y. Shi, “Adversarial machine learning for 5G communications security,” in Game Theory and Machine Learning for Cyber Security. Hoboken, NJ, USA: Wiley, 2021, pp. 270–288.
[16] B. Kim, Y. E. Sagduyu, K. Davaslioglu, T. Erpek, and S. Ulukus, “Channel-aware adversarial attacks against deep learning-based wireless signal classifiers,” IEEE Trans. Wireless Commun., vol. 21, no. 6, pp. 3868–3880, Jun. 2022.
[17] R. Sedar, C. Kalalas, F. Vázquez-Gallego, L. Alonso, and J. Alonso-Zarate, “A comprehensive survey of V2X cybersecurity mechanisms and future research paths,” IEEE Open J. Commun. Soc., vol. 4, pp. 325–391, 2023.