National College of Ireland

Intrusion-and-Impact: Forecasting UAV Communication Attacks and their short-horizon operation degradation

MSc Practicum Part 2
MSc. in Data Analytics

Sahana Lakshmikanthaiah
Student ID: x24172812

School of Computing
National College of Ireland

Supervisor: Prof. Anderson Simiscuka

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Sahana Lakshmikanthaiah
Student ID: x24172812

Programme: Msc Data Analytics **Year:** 2025-2026

Module: MSc Practicum Part 2

Supervisor: Prof. Anderson Simiscuka
Submission Due Date: 27/01/2026

Project Title: Intrusion-and-Impact: Forecasting UAV Communication Attacks and their short-horizon operation degradation

Word Count: ……7546……**Page Count**…20.…….

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Sahana Lakshmikanthaiah

Date: …27/01/2026……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Intrusion-and-Impact: Forecasting UAV Communication Attacks and their short-horizon operation degradation

Sahana Lakshmikanthaiah
x24172812

Abstract

The communication links of Unmanned Aerial Vehicle (UAV) are susceptible to cyber intrusion capable of compromising the Quality of Service (QoS). The dissertation introduces a short horizon forecasting pipeline that pipeline which (i) predicts the likelihood of an intrusion occurring within a 3-minute timeframe and (ii) forecasts near-term degradation in the form of quality of service (RTT, jitter, throughput) over a 3 minute time interval. Temporal causality is imposed by strictly chronologically partitioned and leak-safe feature engineering consisting of current-state KPIs, short lags, as well as compact rolling statistics. Both classification and regression are implemented by gradient-boosting models (XGBoost, LightGBM, CatBoost), and the selected classifier is tuned to give the probabilities, which can be trusted. The weighted severity term derived using the absolute KPI deltas, plus the calibrated intrusion likelihood, produces one operational risk score, which puts imminently large impact windows in focus. The classifier is characterized by a discriminative performance on a held-out test segment, and calibration is improved in terms of quality of probabilities (reduced Brier error, and improved reliability curves are closer to the diagonal). The short-horizon regressors also satisfy low absolute error of RTT and throughput values but jitter is more erratic but directional. All in all, the unified risk score is effective to prioritize pre-emptive intervention, and the method is computationally efficient, interpretable and it is consistent with real-time UAV operations.

Keywords*: UAV communications, intrusion forecasting, QoS degradation, short-horizon prediction, gradient boosting, probability calibration, temporal leakage, risk scoring.*

1 Introduction

Unmanned Aerial Vehicles (UAVs) need permanent and stable communication connections in order to ensure the safety of the flight, manage missions, and transmit sensor data. These connections, however, are susceptible to cyber-attacks, interference, spoofing and degradation because these connections are dynamic and wireless. Conventional Intrusion Detection Systems (IDS) of the UAVs are largely concerned with detection of maliciousness after the occurrence. Although effective in forensic use, in real time, these systems are of

little use since usually, corrective measures could only be effective through prior knowledge and not through an event following detection. The unstable nature of the UAV communication channels which are characterized by RTT, jitter, packet-loss and throughput demands the urgent application of predictive systems that can forecast security threats as well as their performance nearly before they become a reality. This paper has developed a short-horizon forecasting pipeline that forecasts both the probability of a cyber- intrusion and its Quality-of-Service (QoS) effect in the next three minutes. The system uses the combination of calibrated intrusion probabilities with the modeled performance loss to make available a single, explainable risk-of-harm score, which is usable in proactive management of the UAV operations.

1.1 Motivation

Even minor delays in detection of intrusion or loss of communication can be very risky to the operation in mission critical UAV operations. Most of the attacks, including signal interference, or disruption of specific data, have subtle precursors such as increasing RTT instability, jitter bursts, or declining data throughput. All these symptoms can be observed in time before the connection gets seriously damaged. Anticipating these patterns would enable things to be done to prevent them, including changing the air traffic routes, redistributing communication capacities or hitting the defensive measures that would prevent the need to act once the damage is done. Because of this reason, forecasting is more useful than detection. The three minutes forecast window gives a good balance between practical response time and the forecasting horizon: long enough to respond to operations but limited enough to represent significant temporal variations found in UAV information within a few minutes. It is therefore necessary to learn and predict the probability of intrusion as well as the extent of the degradation of the quality of service to effectively engage in proactive UAV communication defense.

1.2 Data and Characteristics of Variables

The data set employed in the study has logs of UAV communication telemetry with a timestamp in benign and attack conditions. The most important variables are network-layer QoS (RTT, jitter, packet loss and throughput), physical-layer and signal-strength (RSSI and transmission rate), and system-level variables (CPU usage, memory usage, hop count and drone velocity). All these variables describe the operational state of the communication link interestingly and are the basis of predictive modelling. Their temporal behaviors, especially fluctuations, short-range instability and multi-metric interactions, give important information in prediction of the occurrence of intrusion and the effect they are likely to cause. The study aim to capture the finer-grained patterns to lead to degradation of communication by utilizing the state-level features as well as the time-based aspects of the study.

1.3 Research Question and Research Objectives

Even though many of the UAV IDS (Intrusion Detection System) models are very accurate on classifying the malicious traffic, they rarely put into consideration the effects of an intrusion on the near-term network performance. This is a limitation that restricts the operational applicability of the IDS outputs in a time-sensitive UAV environment. The main research question that comes up in this dissertation is thus:

Research Question: Is it possible to predict the probability of intrusion and the degradation of QoS at once using short-horizon UAV telemetry and is it possible to integrate these predictions into a single, operationally significant risk score?

Objectives:

The research aims at answering the following questions to address this question:

- build a time-sensitive leak-free forecasting model with crafted temporal features.
- Intrusion likelihood and QoS degradation are both predicted, and the gradient-boosting models are to be trained and tested by a shared three-minute horizon.
- use probability calibration to get good likelihood estimates; and
- consolidate the two outputs into one risk score that can be used to rank forthcoming high-risk windows and target them at the attention of operators.

1.4 Limitations

The research is conducted with a few limitations that define the limits and overall applicability of the results. Limitations in datasets such as using one publicly available dataset on UAV communications, which is not broad enough to encompass the entire range of operational situations, styles of attack, or hardware compositions. Limitations of modelling it arises because the prediction horizon is short, and predicting volatile KPIs like jitter is difficult as it can be impacted by factors that are not completely reflected in the telemetry. Although the gradient boosting models are interpretable and efficient, they are fixed learners; they are not able to adapt to concept drift automatically like in UAV environments. Evaluation limitations also exist: all experiments are based on chronological splits which simulate deployment. not containing real-time feedback loops or adaptive thresholding policies. Nonetheless, the system demonstrates strong evidence of concept of proactive risk prediction in UAV communication despite these limitations.

1.5 Contribution and Structure of the report

This dissertation has three key contributions. First, it trains a single time-conscious forecasting pipeline of UAV communication telemetry predicting the probability of intrusion and short-horizon QoS deterioration using a common pool of designed time-related features. Second, it applies probability calibration to the modelling task, which generates interpretable and operationally valid likelihood estimates, which is frequently absent in UAV IDS research. Third, it presents a risk-scoring system that combines both the intrusion likelihoods in a controlled way and the forecasted KPI degradation to form one, operationalizable signal of the impending communication risk. Combined, these contributions indicate the success

and utility of anticipating both security and service risks within UAV networks, which provide a basis on which more complex and mission-conscious proactive defense models may be constructed.

The rest of this report is then structured as follows. Section 2 is a review of related research in the field of UAV communication security, intrusion detection, and QoS forecasting to position the proposed approach within the scope of existing research. Section 3 provides the research methodology, including the preparation of the data, feature-engineering approach, model design, and evaluation. Section 4 gives the design specification and high-level architecture of integrated intrusion- and impact prediction pipeline. Section 5 outlines how it was implemented, the tools, libraries and the code structure pipeline. The results of the experiment are reported and analyzed in Section 6 in both intrusion prediction and short- horizon QoS degradation forecasting, and a critical discussion of the implications is provided. Section 7 finally closes the study and presents the future work.

2 Related Work

2.1 UAV communication security and IDS survey and taxonomies.

Variously, the surveys and taxonomies have examined the security of UAV communication links, which have been found to be vulnerable to command-and-control channels, telemetry streams, and mission-critical channels (Islam et al., 2025; Choudhary et al., 2018). Three systemic challenges highlighted in prior analyses include (i) channel conditions that change rapidly because of mobility, (ii) onboard computer and power is limited, and (iii) evaluation datasets are constrained, non-standardized, and not reproducible (Choudhary et al., 2018; Islam et al., 2025). Although these works map threats and defenses, they usually consider intrusion detection as an independent classification problem (Choudhary et al., 2018), without considering the short-term consequences of operations (e.g. RTT/jitter/throughput degradation) that occur after an attack.

2.2 ML-Based IDS for UAVs: Algorithms, Efficiency, and Deployment

Recent papers examine ML to provide flexible and precise intrusion detection in aerial and ad-hoc networks (Abdulghani et al., 2022; Abu Al-Haija & Al Badawi, 2022; Alharthi, 2024; Choudhary et al., 2018), such as tree-based ensembles and small deep networks to run on a limited platform. New accomplishments like distributed detection, model compression and protocol-aware design enhance deploy ability (Choudhary et al., 2018). Nevertheless, there are three gaps: the use of IDS evaluations is rarely concerned with the impact of performance, often random (leak-prone) splits on time-varying telemetry are taken, and the calibration of the predicted probabilities are rarely considered, restricting (Bergmeir et al., 2018; Niculescu-Mizil & Caruana, 2005; Zadrozny & Elkan, 2002) the direct use of IDS.

2.3 Short-Horizon QoS Forecasting (RTT, Jitter, Loss, Throughput)

Parallel work demonstrates that it is possible to forecast RTT, jitter, and packet loss and throughput on a short-time basis with tabular ML based on rolling statistics, lagged features and temporal encodings (Wassie et al., 2023; Jagmagji et al., 2024); these predictions can be

used to proactively manage latency compressed systems. However, much of this literature is security-agnostic (Wassie et al., 2023): they are trained to learn normal performance dynamics, but not on intrusion state or risk, thus they are unable to attribute or predict degradation caused by an attack.

2.4 Calibration and Decision Thresholds for Operational Use

The raw scores of the classifier can be transformed into credible probabilities through calibration (Zadrozny & Elkan, 2002; Niculescu-Mizil & Caruana, 2005) and this allows cost sensitive thresholding and risk aware decision making. Investigations and recent research demonstrate that the uncalibrated models are usually either over- or under-confident (Niculescu-Mizil & Caruana, 2005), which impedes the practical application. Calibration is rarely mentioned in UAV security (Kim et al., 2024), which is an asymmetrically costly task with interpretable and mission-time indicators required.

2.5 Dataset and Evaluating Protocols

Open UAV/IoD and network-QoS datasets are too small and incomparable (Bergmeir et al., 2018) to make it difficult to replicate evaluation and external validity. Random train/test splits of time-recorded telemetry are a common pitfall of methodology as it causes time leakage and inflated accuracy reporting. In the case of streaming telemetry, splits that are chronological (time-sensitive) and have leak-safe (Bergmeir et al., 2018). feature construction (lags/rolling windows that are computed solely based on historical data) are necessary to estimate the deployment condition. In severe cases of class imbalance (rare intrusions) resampling or weighting the classes is also required to boost recall on minority classes without overwhelming the operators with false alarms (Choudhary et al., 2018; Shanmugam et al., 2025). These practices are the immediate stimulus of the time-honest handling of imbalance which is employed in this work.

2.6 Concept Drift, Adversarial Dynamics and Online Adaptation

Operational telemetry is nonstationary (Gama et al., 2014) The distributions of KPIs change with the environment, mission profile, traffic load and attacker behaviour. Offline trained models do not have a good performance without regular re-calibration, residual/KPI drift checking, and periodic re-training (Gama et al., 2014). In places where it is possible, roll- based updates or streaming learners can be used to reduce the performance degradation, but these also create verification and stability issues with safety-critical applications. This research foresees drift through proposals of basic drift tests, frequent threshold re- examination and low-profile re-training in accordance to the actualities of UAV operations.

2.7 Cost-Sensitive and Human-in-the-Loop Decision Thresholds

To implement it effectively, there is a need to link model outputs with actionable, cost-sensitive (Niculescu-Mizil & Caruana, 2005; Zadrozny & Elkan, 2002) decisions. The asymmetric costs reflected in calibrated probabilities can include things like false alarming to

cause unwarranted manoeuvres and missed alarming to cause mission instability. The combination of a human-in-the-loop review process and ranked risk lists helps in minimizing the alert fatigue (Kim et al., 2024) because the analyst attention is focused on windows that are a high probability of intrusion and are likely to cause a significant impact on QoS. This drives joint risk score in this work (likelihood x severity), where model estimates are converted to operational prioritisation to pre-emptive mitigation. This joint security-vs-QoS type of thinking is also present in 5G networks, where the intrusion detection decision- making is optimised based on the severity of threats and on the service degradation (Bozorgchenani et al., 2023).

2.8 Summary and Research Gap

The literature has two robust capabilities: high-accuracy intrusion detection based on ML and accurate short-horizon QoS prediction. Nevertheless they are mostly independent: IDS research focuses on discriminating attacks, whereas QoS research is used to model the dynamics of performance without security considerations; calibration and cost-sensitive thresholding (Choudhary et al., 2018; Abu Al-Haija & Al Badawi, 2022; Wassie et al., 2023; Niculescu-Mizil & Caruana, 2005). are poorly reported. The current research paper fills the gap by creating a single intrusion-and-impact pipeline that makes predictions (a) whether an intrusion is most likely to be detected within the next three minutes, and (b) the anticipated degraded QoS within the same horizon based on a unified engineered feature space and calibrated outputs to aid the real-time operational decision-making process. Recent research makes it clear that calibration and class-imbalance processing are important to real-time forecasting of intrusions. Our study follows this line with the addition of isotonic calibration and short-window regression of indicators of QoS by establishing a connection between detection and actual operational impact.

3 Research Methodology

3.1 Dataset

The dataset used Tokyo Drone Communication and Security Dataset from Kaggle open source (DatasetEngineer, 2018). It has rich features of links, signal, and platform, including RTT, jitter, throughput, both packet loss and SNR, signal strength, hop count, packet size, transmission rate, battery, CPU/memory utilizations, altitude, velocity, GPS, and distance to base station and security/context fields, protocol, encryption status and frequency of control command. All the records are either marked as normal, DDoS, Malware, or Anomaly to permit multi-class IDS (an optional binary Intrusion vs Normality fall). When the experiment has short-horizons, the time data may be chronologically resampled into rolling windows to position the inputs in place of the future 3-minute targets (e.g. DRTT/DJitter/DThroughput) without introducing leakage. The breadth and realism of the dataset are suitable in assessing

calibrated intrusion likelihood models and QoS-impact forecasters during dense and operating urban environments.

3.2 Exploratory Data Analysis (EDA)

An exploratory research was performed to know the statistical behaviour, volatility and interdependencies of the telemetry variables. Distribution of RTT, jitter and throughput showed strong variance and dips and spikes that are short-lived and attest to the dynamic character of the UAV communication links. These variations underscore the importance of modelling approaches that can help encompass the micro-temporal variations and not the long-term seasonal variations. Analysis of time-series visualisation showed that sequences of RTT inflation and sudden drops in throughput would often be observed to be near intrusion- labelled intervals, implying that those KPIs may include precursors of malicious interference. The correlation pattern also demonstrated that there are strong correlations that exist between RTT and jitter and an inverse association with throughput, which suggests the possibility of predictive ability when using a combination of both measures in engineered time features.

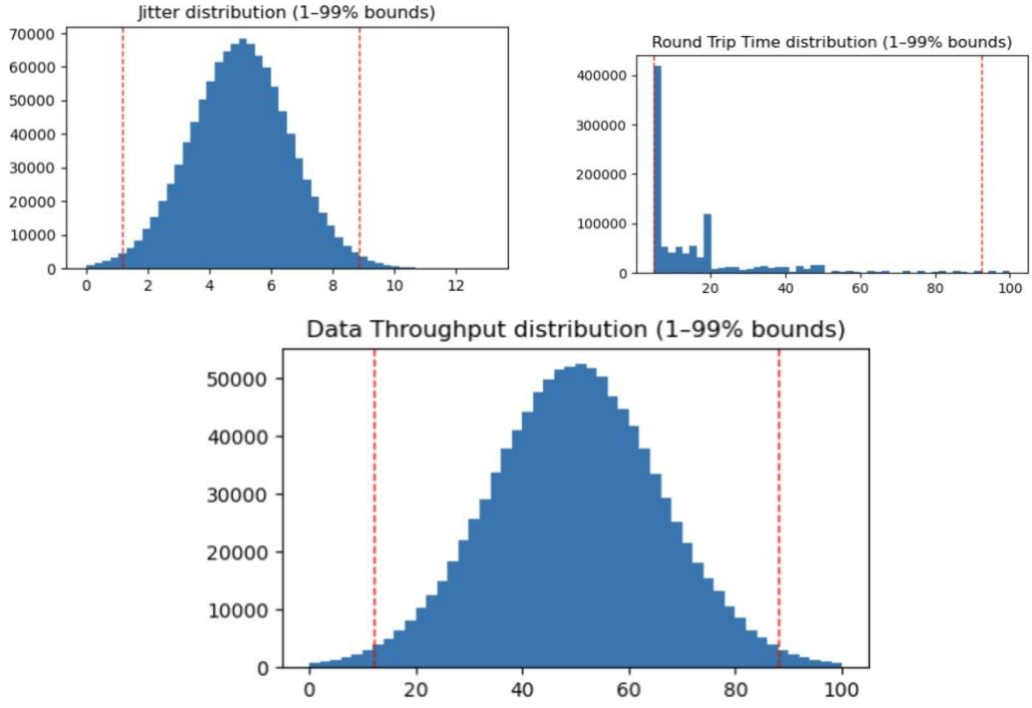

Figure 1. Distribution of RTT, Jitter, Data Throughput values in the UAV telemetry dataset.

3.3 Feature Engineering

The idea of feature engineering was created to capture not only the current state of the communication link but also its immediate temporal dynamic because prediction over the short term is based on recent behavior, not long historical windows. There are three feature types that were built. First, current-state variables are the values of KPIs at time t, which offer a blunt view of the conditions of links in the present moment. Second, lag features (e.g. t-1 and t-2) represent persistence and directionality of changes in telemetry over the short term. Third, rolling-window statistics summaries micro-fluctuations on small windows - such as rolling means, differences and ratio-based statistics such as jitter/RTT and throughput deltas. The transformations enable the models to identify the onset of instability without the need to look far into the past, which becomes less informative when forecasting three minutes. All engineered features were calculated in a strictly leak-safe way, based solely on previous observations, to make sure that the modelling process is in line with the real-time working conditions.

3.4 Label Construction

A common three-minute prediction horizon was used to specify two supervised learning tasks. In the case of intrusion forecasting, a binary classification label was created by labelling timestamps in which an intrusion happened in the next three minutes. This refocuses the issue as being more proactive than reactive, where the operators can predict and react before the communication link becomes strained. In the case of QoS degradation forecasting, RTT, jitter and throughput change parameters were predicted at a future time, and the value was seen to be calculated in the same three minutes, giving three continuous regression targets. Any positive change in RTT or jitter and a negative change in throughput signify deterioration in the link conditions. With an aligned horizon, both prediction tasks are defining the same period of the future so that they can be integrated into one unified risk- scoring framework.

3.5 Model Strategy

To maintain temporal causality and avoid leakage, a chronological train-validation- test split was used. The first part of the dataset was used as the training partition, the middle part was the validation partition, and the last part was the test part, which was the most recent unseen information. This splitting strategy ensures that model choice and analysis resemble actual use, where choices can be made only based on historical and current data. The entirety of the preprocessing measures, such as scaling, encoding and temporal feature construction, was only trained on the training split and then applied to the validation and test sets without refining. This helps in the elimination of contamination of observations in future and builds a sound forecasting based on tasks.

3.6 Model Training

Three families of gradient-boosting models (XGBoost, LightGBM and CatBoost) were taken into consideration to perform both tasks because of their good performance in structured telemetry data and the possibility to model the nonlinear interactions between two or more KPIs. These models are efficient in dealing with heterogeneous feature types and provide intrinsic facilities to deal with class imbalance as well as give interpretable feature-

importance scores, which are essential in understanding the behaviors of models in safety-critical applications. Following some initial experimentation, the highest-performing classifier was calibrated using probability to make sure that the scores that it produced were true estimates of likelihood. Details of calibration and design justification will be discussed later, whereas this section forms the methodological basis of the choice of candidate models and their evaluation.

3.7 Imbalance Management

Intrusion events are a minor portion of the data, establishing a strong imbalance in the size of the classes that may bias models to predict the majority class. To reduce this, baseline trials were used to examine the effects of oversampling on sensitivity, and the final classification models were based on class-weighting to enhance recall on minority-class data without artificially inflating the false positives. Such methods enable the classifier to more efficiently identify delicate short-term trends that are occurring before intrusions whilst retaining viable alerting behaviors. In the final model, class weighting is applied.

3.8 Evaluation Strategy

According to the task, the evaluation was designed using the right measures. The performance of intrusion forecasting was evaluated based on ROC-AUC, accuracy, recall, F1-score and confusion-matrix analysis to describe both discriminative and error composition. The test of probability calibration was done through reliability curves and through the Brier score, such that the predicted probabilities could be used in operational thresholding. In degradation forecasting of QoS, short-range prediction accuracy was quantified using Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and coefficient of determination (R2). To give unbiased estimates of the actual performance in the real world, all results were computed on the held-out test segment after selection of the model. These measures combined constitute a holistic assessment system that is consistent with the two-prediction aims of the study.

4 Design Specification

4.1 System Overview

The architecture is designed as a unified chain of forecasting, which takes the UAV communication telemetry and converts it into two congruent values the probability of the imminent intrusion in the next three minutes and the anticipated severity of the temporarily reduced QoS in the identical horizon. The pipeline is planned as a series of stages that constitute modules starting with the data ingestion, then preprocessing, feature construction and label generation. To maintain causality between time, feature engineering relies on

lagged and rolling statistics that are only calculated using previous samples. All transformations, such as scaling, encoding and calibration, are adapted on the training span and applied over, making zero look-ahead leakage. Chronological train–validation–test divisions are simulated as real time deployment. The components are used to feed individual classification and regression model paths whose results are then integrated to a calibrated, integrated risk score. This architecture will make sure that raw telemetry is converted to actionable and operator ready insights in a consistent and reproducible pattern.

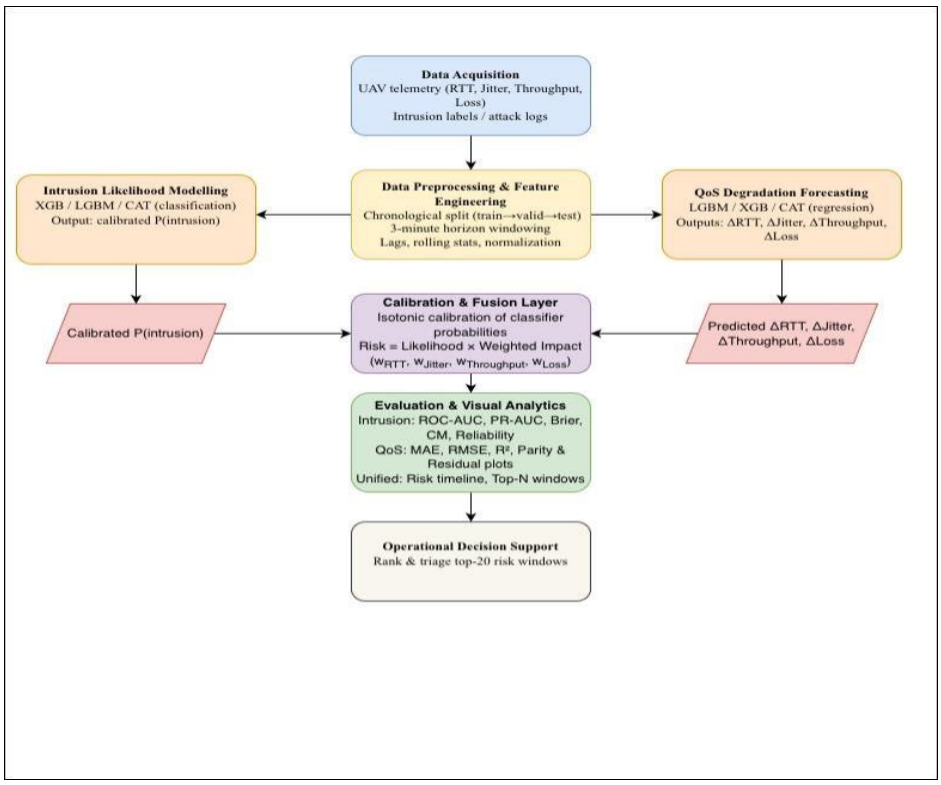

Figure 2. Overview

4.2 Core Components

The data ingestion layer determines the chronological sequence of the telemetry and commits the type consistency in the process which maintains the temporal structure required in short-horizon forecasting. The preprocessing layer does deterministic transformations with one-hot encoding and robust numerical scaling and makes sure that only the historical data is fitted to prevent time leakages. The module of feature construction creates a rich temporal representation of every timestamp by adding short lags, compact rolling-window statistics as well as ratio-based indicators. These features are addressed in terms of the system architecture as opposed to methodological justification. The label generator builds the binary intrusion target, the regression-based targets on QoS degradation of a given three minutes prediction window. The model layer comprises of two prediction branches namely a gradient- boosting classifier (forecasting of intrusion) and three gradient-boosting regressors (DRTT, DJitter and DThroughput). One of the modules, a specific calibration module, selects the reliable probabilities between the scores of the classifiers with Platt scaling or isotonic

regression. The risk composer is lastly the combination of predictive degradation and calibrated intrusion probabilities to generate a single scalar risk score, which is produced by the reporting layer in the form of a prioritized list of high-risk windows about to happen.

4.3 Data Movement and Data Leakage Controls

The data flow in the system has strict temporal integrity. The entire preprocessing and feature-engineering process are fitted on the training partition and subsequently transferred forward without any changes to the validation and test partitions. The temporal characteristics can be calculated based on what has been observed in the past avoiding any unintentional introduction of future knowledge. These architectural controls provide resistance to temporal leakage and also guarantee that offline test is a good simulation of actual real-time performance.

4.4 Calibration and Training Pipeline

Model training is carried out according to a progressive flow in accordance with the architectural blueprint. The first training and hyperparameter choice are done on the validation window which follows the training data in a chronological manner. When the best configuration is found the model is re-trained and evaluated on the test window. Calibration of probabilities is designed to be part of the design and not an after-the-fact addition because the scoring of risks and operational thresholding can only be done based on calibrated scores. The selected models are also used to produce feature-importance outputs by the system to facilitate interpretability.

4.5 Operational Risk Scoring

The pipeline ends in the determination of the single risk score that combines the adjusted risk of intrusion with the expected scale of QoS deterioration. Multiplicative composition of this type puts emphasis on timestamps where damaging degradation and high likelihood of intrusion can coincide so that UAV operators may focus limited attention on the more crucial windows. It has a severity component weighted based on mission priorities providing flexibility to various operational situations. Its calibrated output of the classifier is used to compute all risk scores, which are reliable in terms of probability. This expression can be associated with new collaborative Cybersecurity-vs-QoS models with an explicit expression of the weighting of security events by their anticipated performance impact to enable risk- aware prioritisation (Bozorgchenani et al., 2023).

4.6 Non-Functional Requirements

The system design focuses on the low-latency inference that is applicable to near-real-time UAV communication monitoring. The deterministic preprocessing, fixed seeds and chronologically consistent evaluation can ensure reproducibility. Scalability is facilitated by modularity which enables the integration of extra features or models into the architecture without causing major disruption, robustness is facilitated by outlier-tolerant scaling and leak-safe architecture. This is done by keeping interpretability by using calibrated probabilities, and easy to interpret feature-importance summaries.

5 Implementation

5.1 Tools and Languages

It is implemented in Python with the help of Pandas and NumPy to manipulate data, scikit-learn to preprocess data, split and calibrate; XGBoost, LightGBM and CatBoost to train models. Matplotlib is utilized in the generation of diagnostic plots. Random seeds and the fixed configuration settings are used throughout the environment to guarantee the reproducibility.

5.2 Preprocessing Implementation

At load time, chronological ordering is imposed, and the downstream operations contain the telemetry natural flow. A ColumnTransformer is the preprocessing logic, which is an implementation of one-hot encoding of categorical variables and robust scaling applied to numerical features. The transformer is only applied to the training part of the dataset and applied on the validation and test sets. This ensures that no information is put in the future which affects the normalisation of data.

5.3 Feature Engineering Implementation

The temporal features are constructed based on the vectorised operations, which refer only to past rows. Lag features can record the instantaneous history of KPIs, whereas rolling-window features can summarise the short-term average, difference and instability metrics. Ratio based features like jitter-to-RTT relationships give further normalised measures of link behaviour. The computation of all features is done with the indices aligning closely to ensure that there is no leakage.

5.4 Label Generation in Code

The intrusion label is created by checking every timestamp and recording whether an intrusion is detected within the next three-minute interval. In the case of regression targets, RTT, jitter and throughput at time t + 3 minute are compared with the values at time t to calculate DRTT, DJitter and DThroughput. To make these operations effective, they are done with the help of effective DataFrame indexing to align them and ensure correct temporal alignment.

5.5 Time-Aware Split

The data is separated into training, validation and test windows sequentially. The trained all preprocessing components and windowing functions are applied to the training window and then are applied to the following windows. This temporal discipline is also strict to make sure that development and evaluation of model simulates real-world deployment conditions.

5.6 Model Training

Members of the three gradient-boosting families are trained on the classification and regression problems. The validation performance is used to pick hyper parameters and once this is finalised the models are retrained and tested on the test window. The intrusion forecasting task is mitigated against class imbalance through the use of class-weighting. Each boosting library has its own Python API to model training implemented. CatBoost,

LightGBM, and XGBoost are tested on the pipeline in terms of intrusion and QoS forecasting. In the case of intrusion, isotonic calibration (through CalibratedClassifierCV) is used to transform the raw outputs of the model to trustworthy probabilities. Regarding the QoS regressions, the LightGBM provides the lowest MAE and the highest R 2, which proves its applicability in multivariate KPI forecasting.

5.7 Calibration Procedure

It is implemented with scikit-learn calibration tools in order to perform probability calibration. Validation predictions are subject to platt scaling/isotonic regression to model mapping functions that scale outputs of classifiers to calibrated probabilities. To check the quality of calibration, reliability curves and Brier scores are generated and then the probabilities are incorporated into the risk-scoring step.

5.8 Evaluation Outputs

The notebook generates ROC and precision recall curves, confusion matrices, F1/precision recall tables and calibration curves of the classifier. MAE, RMSE, R2 and residual diagnostics are generated by regression models. The end result of all these computations is an integrated risk score achieved through adding the calibrated intrusion probability and the projected QoS degradation to generate a risk timeline listing the future windows of critical regions.

5.9 Reproducibility

It is reproducible (using fixed seeds), deterministic preprocessing, chronologically consistent data division and a consistent structure of notebook cells. Every configuration parameter is explicitly specified, allowing the pipeline to be rerun in full.

6 Evaluation

The experiments were performed in strictly chronological train-validation-test division to simulate the real conditions of operation in the UAV communication system. Preprocessing components, feature transformations and time windows were only trained on the training segment and extrapolated in real time to the validation and test segments. This guaranteed that no information was leaked to the modelling process in the future hence maintaining the integrity of short-horizon forecasting. Testing was carried out on two mutually complementary activities, predicting the probability of intrusion and forecasting short-term QoS performance. The classifier was also tuned on validation material in such a way that the predicted probabilities would represent real probabilities- a prerequisite of the unified risk- scoring mechanism. The results that are reported below are calculated on the held-out test window only to give unbiased estimates of the generalization performance.

6.1 Experiment 1: Intrusion Forecasting

The gradient-boosting classifier exhibits limited discrimination on the held-out window (V2-LGBM: accuracy = 0.7504, F1 = 0.1475, ROC-AUC = 0.4985, PR-AUC = 0.1510). This is natural when there is extreme imbalance in the classes, due to the ROC/PR curves displaying little or no separation between intrusion-positive and intrusion-negative windows, and the confusion matrix exhibiting few errors that occur during transitional periods when precursors are weak-characteristic of short-horizon forecasting. More importantly, probability calibration results in a better reliability even in cases of a low degree of discrimination: the Brier score decreases to 0.1952 (calibrated V2-LGBM) compared to 0.2037 (uncalibrated BASE-LGBM) and the reliability curve shifts toward the diagonal, which implies a better- calibrated likelihoods. The calibrated model can also serve as a helpful proactive risk indicator since in the safety-critical environment the operational decisions are based on well- calibrated probabilities (not the raw scores) therefore, the probability output of the model can be combined with the predicted impact on QoS (probability x impact) to rank and triage near- term high-risk windows within the integrated risk framework.

Name	Accuracy	F1	ROC_AUC	PR_AUC	Brier
BASE-LGBM	0.732	0.165	0.507	0.153	0.2037
BASE-XGB	0.630	0.200	0.495	0.150	0.2284
BASE-CAT	**0.601**	**0.211**	**0.498**	**0.152**	**0.2345**

Table 1. Baseline -Intrusion Result

Name	Accuracy	F1	ROC_AUC	PR_AUC	Brier
V2-CAT	**0.623**	**0.206**	**0.504**	**0.151**	**0.2297**
V2-XGB	0.664	0.197	0.498	0.152	0.2185
V2-LGBM	0.750	0.148	0.498	0.151	0.1952

Table 2. V2 -Intrusion Result

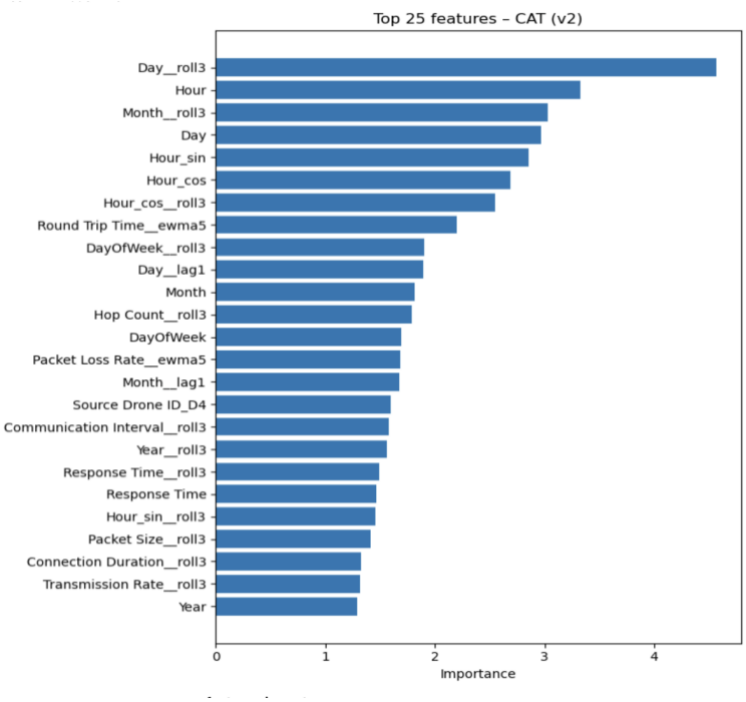

Figure 3. Feature Importance Analysis

6.2 Experiment 2: QoS Degradation Prediction Performance

Regression pipeline provides good short-term predictions in all KPIs, but LightGBM remains the best in terms of the test window. The performance MAE = 0.0628, RMSE = 0.2768, R2 = 0.9714 in the case of RTT, MAE = 0.0175, RMSE = 0.0361, R2 = 0.9343 in the case of Jitter, MAE = 0.1794, RMSE = 0.3618, R2 = 0.9342 in the case of Data Throughput, and the absolute loss values. These findings suggest that near KPI dynamic are effectively represented, although jitter is the most volatile even though overall fit is high. The predictions in conjunction with the calibrated probabilities of intrusion give sound estimates of the level of degradation and confirm the application of engineered temporal characteristics and gradient-boosting to the second aim of the study.

Model	MAE	RMSE	R^2
Round Trip Time — LGBM	**0.0628**	0.2768	**0.9714**
Round Trip Time — XGB	0.0662	0.2811	0.9705
Round Trip Time — CAT	0.1497	0.5129	0.9018
Jitter — LGBM	**0.0175**	0.0361	**0.9343**
Jitter — XGB	0.0193	0.0397	0.9203
Jitter — CAT	0.0247	0.0550	0.8474
Data Throughput — LGBM	**0.1794**	**0.3618**	**0.9342**
Data Throughput — XGB	0.1937	0.3912	0.9230
Data Throughput — CAT	0.2438	0.5394	0.8537
Packet Loss Rate — LGBM	**0.000064**	**0.000303**	**0.9707**
Packet Loss Rate — XGB	0.000072	0.000307	0.9700
Packet Loss Rate — CAT	0.000102	0.000452	0.9348

Table 3. QoS Impact Result

6.3 Experiment 3: Combined Risk Score Evaluation.

Our normal construction is of Risk = Likelihood x Impact. Likelihood: This is the calibrated posterior probability of intrusion of the optimal IDS model at each 3-minute window. Impact is a normalized weighted magnitude of the forecasted degradation of QoS in the next three minutes. It sums the magnitude of the forecasted changes in each KPI-RTT, Jitter, the absolute reduction in the Throughput, and Loss- using the best regressor in each KPI. The operational weights are wRTT = 1.0, wJitter = 0.7, wThroughput = 0.9, wLoss = 0.8. Such a score is high in only cases where both likelihood of intrusion and QoS impact are high and offer a clear-cut ranking to the triage. The highest Risk-of-Harm scores were 2.77-1.34; violation-based (probability x per-KPI violation probabilities) ones were 0.434-0.301. The high-risk intervals brought out by both methods were similar. Operationally, the single, explainable unified score provides a single, near-term response predictor, as opposed to several independent predictions.

Top 10 risky 3-minute windows		Top 10 risky windows	
948914	2.765646	917819	0.436682
948918	2.74407	886061	0.434665
948919	2.603439	886062	0.433476
977513	2.58789	944279	0.32803
977516	2.585247	1009839	0.314335
977514	2.583538	1023818	0.303384
977512	2.5537	843879	0.301396
977515	2.483674	904539	0.301396
977510	2.461179	914759	0.301394
977509	2.436062	914758	0.301394

Table 4. Risk Score Windows

6.4 Analyses

Experiments on variant determined the effect of certain design decisions on performance. The use of class-weighting significantly enhanced recall in intrusion forecasting models over unweighted models, which is why imbalance-aware learning is relevant. The use of baseline oversampling had a slight advantage but created an extra amount of variance, which made class-weighting more appropriate to be deployed. In the final pipeline evaluated on the test window, no oversampling was applied; only class-weighting was used Temporal characteristics, in particular, the short lags, and compact rolling windows, were proved to be decisive both in classification and regression, their absence significantly reduced predictive performance, which justified the methodological decisions undertaken during the feature engineering. The forecast horizon was also investigated using sensitivity analysis. The shorter the horizon, the higher the accuracy was at the expense of the lower operational lead time, and the longer the horizon, the worse was the predictive signal owing to the higher uncertainty of UAV telemetry. The chosen three-minute window is thus a compromise to doable forecasting.

6.5 Error Analysis and Negative Results

The system did not carry out all its functions in a similar manner, and it is worth analyzing its shortcomings to understand where the system can be improved. The intrusion classifier had problems with borderline windows where telemetry gradually crept prior to an attack and which generated conservative probabilities, which occasionally slow down alerts. The QoS regression models, especially DJitter, were found to have a problem with sudden micro-outbursts of volatility that is typical of interference-rich UAV environments. Also, there were a few false-positive risk spikes which were caused by benign congestion that caused KPI degradation, but calibration reduced their operational effects, by giving them low likelihood. Such results indicate that some contextual characteristics or adjustment mechanisms are required to further stabilize predictions when extreme circumstances are considered.

6.6 Practical Implications

The analysis has shown that the suggested system delivers interpretable predictions that can be used in operational UAVs. The system will be able to help the operators handle the communication risk in advance, instead of responding to it. The single risk score provides a succinct decision support device that meshes well with the human-in-the-loop process of making informed decisions on mitigation measures like route deviations, switch to redundancy or a switch to different communication channels. Gradient-boosting models are low-latency, and they exhibit methodological transparency, which makes the system susceptible to real-life implementation where there can be a limited number of computational resources.

6.7 Limitations

The main constraints of the assessment are due to the variety of the datasets and the diversity of the environments. This system was trained and tested on one publicly available dataset of UAV communication and its efficiency might not be consistent on other UAV platforms, communication equipment, and environments or attacker behaviors. Also, none of the models are trained online and are not naturally updated to concept drift, i.e. it can begin to perform poorly in the long run without being recalibrated. The fixed weighting that is used in the severity component of the risk score might not indicate the different priorities of the mission. Despite these limitations, the assessment gives a strong proof-of-concept that justifies the existence of unified intrusion-and-impact forecasting.

6.8 Summary

The fact that the short-horizon UAV telemetry can be utilized to predict the intrusion probability and the quality-of-service deterioration and that the integration of the two predictions into one risk measure can substantially improve the operational preparedness. The intrusions classifier was shown to have a discriminative performance with the aid of the use of the temporal features and calibrated probabilities whereas the QoS regressors showed a more precise degradation pattern over a short period of time which could be used to make risk judgments. The integrated score was able to rank high-risk windows effectively, which exhibited undoubted practical applicability. The results can be used as the direct answers to the research question and lay a groundwork towards the more sophisticated proactive defense systems within the UAV communication settings and generates a measurable prioritization scale to operators, with levels of detection sensitivity and tolerance to false alarms.

7 Conclusion and Future Work

This dissertation demonstrates that jointly predicting the probability of intrusion and the anticipated magnitude of imminent degradation of QoS is possible with short-horizon UAV telemetry and that the resulting predictions can be effectively combined to produce a single operationally useful risk score. Two complementary datasets (calibrated intrusion likelihoods and predicted KPI changes (DRTT, DJitter, DThroughput)) on a shared, engineered feature space on a shared three-minute horizon were provided by a leak-safe time-aware pipeline that has a shared gradient-boosting model space provided. The classifier had a discriminative performance and calibration enhanced probability reliability on a chronologically held out test window and the regressors modelled short-range dynamics on both RTT and throughput

with small absolute error (and directionally informative results on jitter). The unified risk score, most importantly, always prioritized risk where high likelihood and high impact coincide, transforming the predictions of the model into a simple, take-action, prioritization of imminent risk. These results give a solid and interpretable line of action to proactive UAV communication defense, instead of reactive response, within the confines of the scope of a single dataset and fixed horizon. All these results show that the simultaneous intrusion-forecasting and QoS-degradation-forecasting can be combined to a single, functional risk indicator that can be used to respond to the research question and promote proactive triage.

Future Work

The future development will aim at increasing the adaptability, strength, and in-field preparedness of the proposed framework. To maintain reliability, concept-drift monitoring and regular recalibration will be adopted to ensure drift does not increase as UAV telemetry trends and attackers' behavior changes with time. It is possible to add more contextual indicators to the feature space like mobility, retransmission, and interference measures to enhance the stability of the jitter and throughput predictions, though this may come at a small computation cost. Expansions of the system to multi-horizon prediction e.g. 1-, 3-, and 5-minute would permit operators to tradeoff between lead time and accuracy and enable more graded and proactive responses. In addition, optimization of severity metric by normalizing KPI scales and providing mission-specific weights (where latency- or throughput-critical missions are run) will bring the unified risk score to consistent operation costs. Lastly, model validation on various datasets and platforms and on- device or edge deployments with organized human-in-the-loop supervision will bring the framework into closer relation to field operations. In general, the calibrated short-horizon forecasting model is capable of correlating intrusion probability with the predicted QoS degradation, and all these extensions are expected to enhance its flexibility, accuracy, and implementation in autonomous UAV control to pre-empt route or channel switching.

References

Abdulghani, A.M., Abdulghani, M.M., Walters, W.L. and Abed, K.H. (2022). Improving Intrusion Detection in UAV Communication Using an LSTM-SMOTE Classification Method. Journal of cyber security, 4(4), pp.287–298. doi:https://doi.org/10.32604/jcs.2023.042486.

Abu Al-Haija, Q. and Al Badawi, A. (2022). High-performance intrusion detection system for networked UAVs via deep learning. Neural Computing and Applications. doi:https://doi.org/10.1007/s00521-022-07015-9.

Aldossary, M., Alzamil, I. and Almutairi, J. (2025). Enhanced Intrusion Detection in Drone Networks: A Cross-Layer Convolutional Attention Approach for Drone-to-Drone and Drone-to-Base Station Communications. Drones, 9(1), p.46. doi:https://doi.org/10.3390/drones9010046.

Arash Bozorgchenani, Zarakovitis, C.C., Su Fong Chien, Tiew On Ting, Ni, Q. and Wissam Mallouli (2023). Novel modeling and optimization for joint Cybersecurity-vs-QoS Intrusion Detection Mechanisms in 5G networks. Computer Networks, 237, pp.110051–110051. doi:https://doi.org/10.1016/j.comnet.2023.110051.

Bergmeir, C., Hyndman, R.J. and Koo, B. (2018). A note on the validity of cross-validation for

evaluating autoregressive time series prediction. Computational Statistics & Data Analysis, 120, pp.70–83. doi:https://doi.org/10.1016/j.csda.2017.11.003.

Choudhary, G., Sharma, V., You, I., Yim, K., Chen, I.-R. and Cho, J.-H. (2018). Intrusion Detection Systems for Networked Unmanned Aerial Vehicles: A Survey. [online] IEEE Xplore. doi:https://doi.org/10.1109/IWCMC.2018.8450305.

DatasetEngineer (2018). Tokyo Drone Communication and Security Dataset. [online] Kaggle.com. Available at: https://www.kaggle.com/datasets/datasetengineer/tokyo-drone-communication-and-security-dataset?select=drone_communication_dataset.csv [Accessed 2025].

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M. and Bouchachia, A. (2014). A survey on concept drift adaptation. ACM Computing Surveys, 46(4), pp.1–37. doi:https://doi.org/10.1145/2523813.

Islam, M.S., Mahmoud, A.S. and Tarek Rahil Sheltami (2025). AI-Enhanced Intrusion Detection for UAV Systems: A Taxonomy and Comparative Review. Drones, [online] 9(10), pp.682–682. doi:https://doi.org/10.3390/drones9100682.

Jagmagji, A.S., Zubaydi, H.D., Molnár, S. and Alzubaidi, M. (2024). Utilizing Machine Learning as a Prediction Scheme for Network Performance Metrics of Self-Clocked Congestion Control Algorithm. Infocommunications journal, 16(3), pp.2–17. doi:https://doi.org/10.36244/icj.2024.3.1.

Kim, Y., György Dán and Zhu, Q. (2024). Human-in-the-loop Cyber Intrusion Detection Using Active Learning. IEEE Transactions on Information Forensics and Security, 19, pp.8658–8672. doi:https://doi.org/10.1109/tifs.2024.3434647.

Niculescu-Mizil, A. and Caruana, R. (2005). Predicting good probabilities with supervised learning. Proceedings of the 22nd international conference on Machine learning - ICML '05. doi:https://doi.org/10.1145/1102351.1102430.

Raed Alharthi (2024). Enhancing unmanned aerial vehicle and smart grid communication security using a ConvLSTM model for intrusion detection. Frontiers in Energy Research, 12. doi:https://doi.org/10.3389/fenrg.2024.1491332.

Shanmugam, V., Roozbeh Razavi-Far and Ehsan Hallaji (2024). Addressing Class Imbalance in Intrusion Detection: A Comprehensive Evaluation of Machine Learning Approaches. Electronics, [online] 14(1), pp.69–69. doi:https://doi.org/10.3390/electronics14010069.

Wassie, G., Ding, J. and Wondie, Y. (2023). Traffic prediction in SDN for explainable QoS using deep learning approach. Scientific Reports, 13(1). doi:https://doi.org/10.1038/s41598-023-46471-8.

Zadrozny, B. and Elkan, C. (2002). Transforming classifier scores into accurate multiclass probability estimates. Transforming classifier scores into accurate multiclass probability estimates. doi:https://doi.org/10.1145/775047.775151.