

ISSN 1840-4855

e-ISSN 2233-0046

Original scientific article

<http://dx.doi.org/10.70102/afts.2026.1836.229>

ASSESSING LEARNING-BASED MODELS FOR INTRUSION DETECTION IN UAV NETWORKS

Razan Abdulhammed^{1*}, Mohand Lokman Ahmad², Shaima Miqdad Mohamed Najeeb³, Dr. Haider TH. Salim Alrikabi⁴

^{1*}Lecturer, Artificial Intelligence Research Center, Northern Technical University, Mosul, Iraq. e-mail: rabdulhammed@ntu.edu.iq, orcid: <https://orcid.org/0000-0002-5939-6009>

²Assistant Professor, Department of Artificial Intelligence Engineering Technology, Northern Technical University, Mosul, Iraq. e-mail: mohandaldabag@ntu.edu.iq, orcid: <https://orcid.org/0000-0003-1682-4293>

³Lecturer, Department of Cyber Security Engineering Technology, Northern Technical University, Mosul, Iraq. e-mail: shaimamiqdad76@ntu.edu.iq, orcid: <https://orcid.org/0000-0001-9746-8117>

⁴Professor, Department of Electrical Engineering, College of Engineering, Wasit University, Wasit, Iraq. e-mail: hdhiyab@uowasit.edu.iq, orcid: <https://orcid.org/0000-0002-5201-2084>

Received: May 04, 2026; Revised: June 16, 2026; Accepted: July 24, 2026; Published: September 29, 2026

SUMMARY

Unmanned Aerial Vehicle (UAV) networks are being used in various military, civilian and infrastructure-monitoring fields, making them vulnerable to an increasing number of cyber threats. To achieve strong security in these resource-limited, dynamic environments, lightweight and accurate intrusion detection systems (IDS) are necessary. In this paper, three Machine Learning (ML) classifiers: Logistic Regression, Random Forest, and Light Gradient Boosting Machine (LightGBM) and three Deep Learning (DL) architectures: Feed Forward Neural Network (FFNN), one-dimensional Convolutional Neural Network (1D-CNN), and Long Short-Term Memory (LSTM), are compared to find the best performance for intrusion detection in UAV networks. An 80:20 train-test split was used for training and evaluation, with the same 80:20 ratio applied to both of the two recently released, publicly available benchmark datasets, UAVIDS-2025 and UAV-GCS-IDS, and with weighted-average Accuracy, Precision, Recall, and F1-score as the evaluation metrics. The performance of the models shows that LightGBM has the highest overall performance with 100.0% accuracy and precision on UAVIDS-2025 and 99.04% and 98.93% accuracy and F1 score on UAV-GCS-IDS, outperforming Random Forest with 99.97% and 88.89% accuracy on UAVIDS-2025 and UAV-GCS-IDS, respectively, and the deep learning models with accuracy ranging from 88.2% to 99.6% on UAVIDS-2025 and 89.6% to 98.93% on UAV-GCS-IDS. The Logistic Regression always yielded the lowest scores with only 86.8%-87.0% accuracy in UAVIDS-2025 and around 85.1% in UAV-GCS-IDS, showing the capabilities of a linear decision boundary in these traffic patterns. The results show that when computational resources are available, deep learning architectures are also effective at learning complex, non-linear attack patterns, while gradient-boosted ensemble methods provide the best detection accuracy with the lowest computational burden for UAV intrusion detection. Finally, the paper outlines recommended deployment strategies for ensembles and hybrid ML-DL, which are applicable in the context of resource-constrained environments of UAV networks.

Key words: UAV network security, intrusion detection system, machine learning, deep learning, LightGBM, anomaly classification, UAVIDS-2025.

INTRODUCTION

An Unmanned Aerial Vehicle (UAV) is an aircraft that operates with an automatic pilot onboard and without any interference from a human pilot. Commonly these aircraft are called drones [5]. A UAV Network is a network that consists of multiple UAV units (drones) that communicate with each other through wireless communication links. The network is designed to empower the drones to communicate, coordinate, and collaborate in real time. Thus, enhancing their effectiveness in operating as a cohesive unit. The UAV Network is used in different applications such as disaster Response, Military, Agriculture, Environmental Monitoring, and Smart Cities. The main typical components of the UAV network are UAV Units, Ground control stations, and communication links. The UAV Network is characterized by its mobility, dynamic deployment, and flexibility features that differentiate them from traditional networks [18].

Intrusion detection system in UAV Network is essential to secure the network, maximize the operational effectiveness and survivability of UAV Technology in hostile environments. The IDS is assessed in improving mission success rate and reducing risk of costly loss through protecting the network from cyber security attacks. To design an Intrusion Detection System (IDS) for an UAV network, the design must be lightweight, adaptive, and capable of detecting in real time and under resource constraints both zero-day attacks and non-zero-day attacks [19]. Utilizing Machine learning and deep learning in intrusion detection systems can significantly enhance the ability of unmanned aerial vehicles to carry out required missions on time and correctly, contributing to these aircraft's survivability in a hostile cyber space environment [20]. This study aims to present a comparative assessment of ML and DL Algorithms towards designing an intrusion detection system for UAV network environment. There are various types of datasets that can be used to design an intrusion detection system for future generation networks [20]. The study utilized 2 datasets that were specialized for the UAV network. These datasets have a wide range of attacks that are subtle to the UAV environment to produce an authenticated, accurate comparative assessment for machine and deep Learning algorithms. To highlight these datasets and their features, table 1 presents a comparison between the two datasets.

Table 1. Comparison among common UAV intrusion detection datasets

Dataset Name	Size	Type	Attack Vector	Contents	Source
UAVIDS-2025 [16]	19MB	NS3 Simulation	Blackhole, Flooding, Sybil, and Wormhole	Contains normal flights and UAV flights with cyberattacks	IEEE Data Portal
UAV-GCS [4]	9.1MB	Real-time Communication	DDoS, DoS, MITM, Replay, Fake Landing	Network traffic data collected under both normal and malicious conditions in UAV networks	Platform Figshare.

This research fleshes out its novelty with respect to the existing comparison studies of ML and DL in the context of UAV intrusion detection on three fronts. First, in contrast to previous single-dataset studies [1][2][15][16][17], this study compares the models under the same model preprocessing, model training, and model evaluation conditions, so that the generalizability of the model can be fairly judged by comparing the model generalizability to the two datasets, not by tuning the model to just one single dataset. Second, the study takes place in a single uniform experimental pipeline that spans across the six algorithms tested, three complexity levels – linear, ensemble tree-based, and deep neural – which is rare in the reviewed literature.

Third, the per-class confusion-matrix analysis observed in Section 5 provides specific insights into which attack type categories (e.g., Wormhole-Blackhole confusion and Fake Landing under-detection) are problematic for each model family, in addition to overall accuracy metrics, to guide practitioners in the design of resource-constrained UAV IDS in table 2.

The contributions of this paper are as follows. First, a comprehensive comparative evaluation of three classical machine learning classifiers, namely Random Forest, Logistic Regression, LightGBM, and three deep learning architectures, namely LSTM, FNN, and 1DCNN, are used for UAV network intrusion detection. Second, both model families are systematically evaluated on two datasets,

UAVIDS-2025 and UAV-GCS, under identical experimental conditions to ensure a fair and reproducible comparison. Third, a detailed per-class performance analysis is presented using normalized confusion, providing granular insight into each model's classification behavior across all traffic categories. Finally, the findings of this study offer practical guidelines for selecting appropriate classifiers in UAV intrusion detection applications.

Table 2. Attack name, procedure, and attack goal in UAV environment

Attack Name	Procedure	Attacker Goal
Blackhole	A malicious UAV falsely advertises optimal routes to attract traffic and then silently drops or absorbs packets.	To disrupt one or all of the following operations: multi-hop communication, mission coordination, and data collection.
Flooding	An attacker attempts to exhaust the UAV's node bandwidth, CPU, or memory by sending large volumes of bogus, repeated packets and route requests	The goal is to cause denial of service to legitimate control and data traffic.
Sybil attack	An attacker creates a malicious node that forges multiple fake identities to appear as many distinct UAVs as possible.	The goal is to undermine routing, consensus, resource allocation, and cooperative sensing in UAV environment.
Wormhole	An attacker controls one or two malicious nodes to establish a low-latency tunnel for forwarding and replying to packets between distant regions.	To dispute the process of routing, topology discovery, and cooperative missions
Scanning	An attacker using either passive monitoring or active probing gathers information about the UAV	The goal is to use the gathered information to create a topology and vulnerability map of the UAV system.
Fake Landing	An attacker transmits either counterfeit landing signals or falsified landing-site information.	Directing the UAV to an unsafe location allows the attacker to either terminate the mission or destroy the UAVs.
A man-in-the-middle attack	The attacker intercepts, modifies, and injects commands and telemetry between a UAV and its controller or between UAVs.	To eavesdrop and manipulate command and access control.
Bruteforce	The attacker tries to guess the authentication credentials of the UAV network through trial-and-error methods.	To gain access to the UAV Network to a compromised system and cause more damage.
DoS	An attacker reconnoiters the UAV system to select an appropriate strategy based on certain attack vectors to overwhelm the UAV Network with bogus requests	The goal is to exhaust UAV node resources or to deny legitimate UAV nodes
DDoS	Same as DoS, but here the attacker uses multiple compromised nodes rather than one, as in the DoS case.	The goal is to exhaust UAV node resources or to deny legitimate UAV nodes
Reply	The attacker intercepts legitimate command or sensor signals exchanged between a UAV and its Ground Control Station.	To manipulate the drone's behavior

The rest of this paper will be structured as follows. In Section 2, the relevant studies on intrusion detection for UAV networks are reviewed, which focus on the studies based on the UAVIDS-2025 and UAV-GCS-IDS datasets. In Section 3, the studies' machine learning solutions are introduced. Section 4 introduces the proposed comparative assessment methodology used, such as the machine model and deep learning models adopted, their parameter settings, and characteristics of the datasets. The experimental results and discussion are given for both datasets in Section 5. The paper is completed at the end of Section 6.

RELATED WORK

This study reviews prior research on intrusion detection systems (IDSs) in UAV networks, specifically examining machine learning and deep learning approaches evaluated on the UAVIDS-2025 and UAV-GCS datasets while excluding those based on other data sources. A quantum machine learning-based intrusion detection system for UAV networks was developed using the UAVIDS-2025 dataset. The proposed HybridQNN-8L model demonstrated high performance with an accuracy of 94.8% and an F1 score of 96.7%, highlighting the potential of quantum-enhanced learning for UAV intrusion detection [1]. The NS-3 network simulator was also employed to construct the UAVIDS-2025 dataset for evaluating supervised machine learning algorithms, including Random Forest, XGBoost, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naïve Bayes (NB), and Decision Tree. Among the evaluated models, XGBoost achieved the highest performance with an accuracy of 98.7% and an F1-score of 97.9% [2]. The susceptibility of the Internet of Drones (IoD) to advanced cyberattacks has also been investigated using the Histogram-based Gradient Boosting (Hist-GBoost) algorithm, which demonstrates excellent detection accuracy with an F1-Score of 99.42% [15]. The ECU-IoFT dataset was evaluated using Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN), achieving an average accuracy of 96% for both the ANN and RNN models, while the CNN model reached 81% [16]. The results demonstrate that a Feedforward Convolutional Neural Network (FFCNN) trained on the UAVIDS dataset achieves an overall detection accuracy of 98.23% for UAV intrusions, further validating the efficacy of deep learning approaches in this domain [17].

In addition to the aforementioned studies based on such datasets, the UAV network security has been further enhanced by the distributed and graph-based intrusion detection frameworks [3]. A complete distributed intrusion detection system with five co-operating modules for collecting, preprocessing, and multi-layered threat analysis was shown to achieve better detection performance in a live UAV testbed for different distances and speeds of the targets. The distributed architecture increased the detection capability, but it was more complicated in terms of computational and implementation complexity than the centralized architecture. A deep analysis of intrusion detection systems based on machine learning and graph neural network (GNN) revealed that the graph-based models are able to model the higher-order relationships between UAVs better than other feature-based classifiers. These results are consistent with the results of feature-based machine learning methods and indicate that GNNs are a potential future research direction for UAVs to meet threats.

The survey of literature has shown that ensemble learning and gradient boosting algorithms have proven to be the best for intrusion detection in UAV datasets. Most of the current studies discuss only one dataset and a small number of machine learning algorithms. The present study, on the other hand, compares the performance of six machine learning models on two benchmark UAV intrusion detection datasets, which offers a wider comparative test of the detection capability and ability of models to generalize.

Machine Learning Approaches

In this study, several machine learning approaches were chosen for comparative assessment alongside Deep learning approaches. To choose fewer complex algorithms and architecture since UAV environment is resource constrained environment. These approaches encompass Random Forest, LightGBM, and Logistic regression. The random forest has many characteristics that make this algorithm a good choice among machine learning classifiers due to its ability to handle linear and nonlinear data, missing data, and it is robust to noise [6][7]. Random forest is best to be used in medium complexity problems and noisy datasets. The logistic regression algorithm is considered simple, interpretable and fast assuming linearity and in most cases when compared with other algorithms it produces lower accuracy. It is recommended to use Logistic regression when the dataset is small, and an initial interpretation is needed regarding the dataset [8].

PROPOSED COMPARATIVE ASSESSMENT

The goal of this study is to develop a systematic comparative assessment framework for UAV Network's intrusion detection system, utilizing machine learning and Deep learning. From a machine learning and deep learning perspective, A comparative assessment can be defined as the process of evaluating and comparing the performance of different machine learning and deep learning approaches on a certain dataset that follows set criteria defined by the researchers implementing the process. To this, the researcher must select certain evaluation metrics to help in identifying which approaches perform best. The process requires training multiple models, testing using the same dataset, as well as under the same conditions. Then the theses and the achieved results based on the defined performance metrics to understand their strengths, weaknesses, and suitability for the problem at hand.

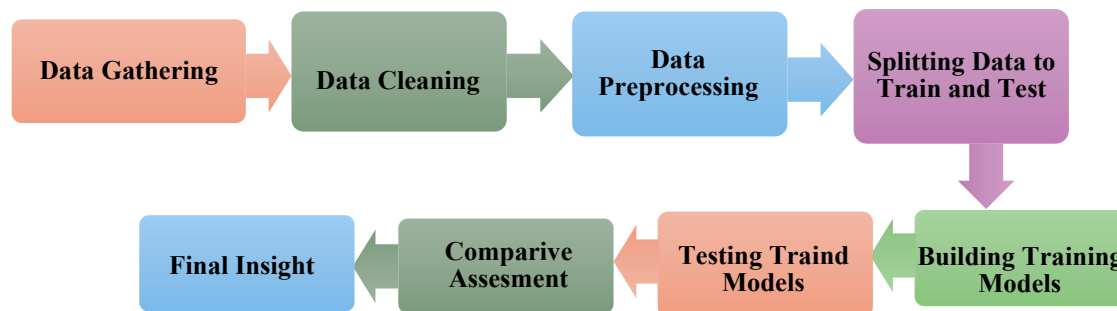


Figure 1. The comparative assessment procedure flow diagram

The comparative assessment procedure flow diagram is shown in figure 1. A traditional classical assessment procedure was followed to ensure that the model is valid, reliable, and credible

Proposed Comparative Assessment References Performance Metrics

To determine which model provides the most accurate predictions with respect to the selected datasets, several common evaluation metrics were used that include the following [9][10]: 1. recall, since this metric is crucial to measure the proportion of actual attacks correctly detected. Also, it can reflect the model's effectiveness in identifying threats and minimizing misclassified attacks. Moreover, 2. F1 Score represents the harmonic mean of precision and recall, and it balances the trade-off between detecting true attacks and minimizing false alarms. Moreover, when classes imbalance, it is a valuable metric for assessment compared to models. In addition, 3. The precision highlight alerts are meaningful and trustworthy since they measure the proportion of true detected attacks against all flagged as attacks [9] [10].

Dataset Characteristic

In this study, two datasets were chosen, the UAVIDS-2025 which is a Benchmark Dataset that is publicly available and released in 2025. It is a simulated dataset (Through using NS3.24 Program) and not a captured real-world network traffic. The UAVIDS-2025 dataset was chosen as result for many reasons. For example, it is relevant to research on UAV security, publicly available, designed for machine learning-based IDS, and its recency. The UAVIDS-2025 dataset has 122171 instances, each described by 22 features per sample. It encompasses 5 classes: Sybill, Wormhole, Blackhole, Flooding Attacks, and Normal Traffic [2].

The second dataset is UAV-GCS, which is a real traffic network and hosted on Figshare. It has 149434 instances, each described by 45 features, including the label. This study utilized eight classes: Scanning, Bruteforce, MITM, Reply, Fakelanding, DoS, DDoS attacks, and Normal Traffic. Whereas the study excluded, Reco and Evil Twin class instances [4]. The class distribution of UAV-GCS dataset is imbalanced as illustrated in table 3 [4].

The study chose UAV-GCS since it highlights the security of UAV to ground control station as well as it represents attack surface in recent UAV security. Moreover, it is new, updated, and tailed to machine learning based IDS designed.

Table 3 highlights the characteristics of the datasets. To illustrate: TAC: Total Attacks Class, FC: Features Count, CFC: Categorical Features Count, NCFC: Noun Categorical Features Count, TIC: Total Instances Count, MinCC: Minority Class Count, MajCC: Majority Class Count, and Finally BR: Balance Ratio. Before feeding the data into the model, preprocessing steps were applied, including traditional classic normalization of numerical features and label encoding of categorical ones. All the Datasets were split into a ratio of 20:80 for tests to train respectively.

Table 3. Characteristics of UAVIDS-2025 and UAV-GCS datasets

Dataset	TAC	FC	CFC	NCFC	TIC	MinCC	MajCC	BR
UAVIDS-2025 [16]	5	22	3	19	122171	6000	60000	0.1
UAV-GCS [4]	7	44	0	44	149434	48	50135	0.000294444

Most researchers identify whether a dataset is balanced or imbalanced through checking the class distribution ratios in the dataset. A dataset is considered balanced if the classes in the target variable have roughly equal representation. In the same context, dataset is considered imbalanced if the classes in the minority class (the class with the smallest number of instances count) is significantly smaller than the majority class (the class with the largest number of instances count) by a large margin such fifty percent or thirty percent [11][12]. In python, and for checking dataset balance scientifically, calculate the imbalance ratio using the following procedure:

1. Calculate the number of instances per class.
2. Compute the ratio of minority to majority class counts.
3. If the imbalance ratio is close to 1, then the dataset is balanced; otherwise, it is considered an imbalanced dataset. The closer this value is to zero, the more it indicates a significant class imbalance problem in the dataset, and the higher likelihood that the classifier algorithm will fail to classify the underrepresented categories in the data, especially those that are not strongly correlated with the features present in the data.

Learning Based Model and Parameter Configuration

In this section, the study introduces the learning-based model parameters configuration for all the experiments running on both datasets, UAV-IDS-2025 and UAVGCS. Although both machine learning and deep learning models were evaluated on each dataset.

Before model training, the dataset was processed through a structured pipeline to guarantee both data quality and compatibility with the model. First, identified and removed instances with missing values to ensure data integrity. To prevent redundancy and reduce the risk of overfitting, duplicate records were subsequently removed. To ensure compatibility with the machine learning algorithms used, categorical features were transformed into numerical representations. Numerical features were normalized using Min-Max scaling to transform all values into a uniform range, ensuring that each feature contributes equally regardless of its magnitude. As mandated by certain approaches, class labels were one-hot encoded before training, converting each categorical label into a binary vector whose length matches the number of classes, with a value of 1 signifying class membership and 0 indicating otherwise. Finally, the preprocessed dataset was split into training and testing sets using an 80/20 ratio to ensure unbiased model evaluation.

Machine Learning Models

Three machine learning classifiers were employed in this study: Light Gradient Boosting Machine, Random Forest, and logistic regression. These models represent a range of complexity levels that range from linear baseline classifier (Logistic regression) to enable methods (RF and LightGBM). Thus, enabling a comprehensive comparative evaluation. Moreover, the study reports the classical-typical

default parameters which were applied to the models used in assessing the classification tasks and lists the parameters that changed from default, plus tuned values. While others were set according to default scikit-learn for each of LightGBM, Random Forest, and Logistic regression model settings. Table 4 through 6 report key parameters for each model separately and a description of each parameter's explanation is provided in column 3 of each of these tables.

Tables 4 table 5 through 6 display the hyperparameter settings for classical classifiers on the UAV-IDS -2025 dataset, while tables 7, table 8 through 9 outline the deep learning architecture parameters for the UAV-GCS dataset. This reporting strategy was implemented to prevent redundancy while guaranteeing the complete reproducibility of the experimental results.

Table 4. Logistic key parameters - UAVIDS-2025

Parameters	Setting	Description
penalty	l2	Regularization type (L1, L2, ElasticNet)
C	1.0	Inverse of regularization strength (smaller = stronger)
max_iter	1000	Optimization algorithm
multi_class	deprecated	Maximum iterations for convergence
solver	lbfgs	Strategy for multi-class problems
fit_intercept	True	Whether to add a bias/intercept term
random_state	42	Seed for reproducibility

Table 5. Random forest key parameters - UAVIDS-2025

Parameters	Setting	Description
n_estimators	100	Number of decision trees in the forest
max_depth	None	Maximum depth of each tree
min_samples_split	2	Minimum samples required to split a node
min_samples_leaf	1	Minimum samples required at a leaf node
max_features	sqrt	Number of features considered per split
criterion	gini	Split quality measure (gini or entropy)
bootstrap	True	Whether to use bootstrap sampling
random_state	42	Seed for reproducibility

Table 6. LightGBM key parameters- UAVIDS-2025

Parameter	Setting	Description
boosting_type	gbdt	Boosting method (gbdt, dart, rf)
num_leaves	31	Max leaves per tree (controls complexity)
max_depth	-1	Maximum tree depth
learning_rate	0.1	Step size shrinkage to prevent overfitting
n_estimators	100	Number of boosting iterations
min_child_samples	20	Minimum data in a leaf node
feature_fraction	1.0	Fraction of features used per iteration
bagging_fraction	Default	Fraction of data used per iteration
lambda_l1	Default	L1 regularization
lambda_l2	Default	L2 regularization
Objective	multiclass	Task type (binary/multiclass)

Deep Learning Models

Three deep learning classifiers were employed in this study: LSTM (Long Short-Term Memory) which is a recurrent neural network architecture that is designed to learn long- term dependencies in sequential data. In terms of IDS, the LSTM can capture temporal correlations between consecutive network flow records. FFN (Feed Forward Network), this is the fundamental deep learning architecture, and it can serve as a strong baseline model due to its simplicity of architectural design [13]. The 1DCNN (One -Dimensional Convolutional Neural Network) applies convolutional filters along a single spatial dimension. In terms of IDS design, the 1DCNN can capture local pattern within network traffic features

that are indicative of attacker behavior [14]. In this study, the FFN, LSTM, and 1D-CNN were selected to represent a diverse range of deep learning architectures, a fully connected baseline, a sequential memory-based model, and a local pattern extractor enabling a comprehensive and fair comparative evaluation of intrusion detection performance. Moreover, the study lists the parameters which changed from default, plus tuned values. While others were set according to default scikit-learn for each of LSTM, FFN, and 1DCNN. Table 7 through 9 report key parameters for each model separately and a description of each parameter's explanation is provided in column 3 of each of these tables.

Table 7. FFN key parameters of UAV-GCS

Parameter	Setting	Description
Input Features	44	Number of input dimensions after preprocessing (44 features)
Hidden Layer 1	128 neurons	First dense layer to capture complex patterns
Hidden Layer 2	64 neurons	Second dense layer for further feature extraction
Dropout Rate	0.3	Probability of shutting down neurons to prevent overfitting
Output Classes	8	Number of distinct attack categories to classify
Activation Function (Hidden)	ReLU	Introduces non-linearity to the hidden layers
Activation Function (Output)	Softmax	Calculates probability distribution over the classes
Optimizer	Adam	Algorithm used to update weights based on the loss
Loss Function	Sparse Categorical Crossentropy	Calculates the error between predicted and true labels
Epochs	50	Number of complete passes through the training dataset
Batch Size	32	Number of training samples processed before updating weights

Table 8. LSTM model architecture & hyperparameters summary - UAV-GCS

Parameter	Setting	Description
Input Timesteps	10	Number of historical time steps used for each prediction sequence.
Training Epochs	50	Maximum number of complete passes through the training dataset.
Batch Size	32	Number of training samples used in one iteration to update weights.
Train-Test Split	0.8 / 0.2	Ratio of data used for training versus testing/evaluation.
LSTM Layer 1 Units	64	Number of memory cells in the initial LSTM layer.
LSTM Layer 2 Units	32	Number of memory cells in the secondary LSTM layer.
Dropout Rate	0.2	Percentage of neurons randomly deactivated to prevent overfitting.
Optimizer	Adam	The algorithm used to minimize the loss and update network weights.
Loss Function	Categorical Crossentropy	Mathematical method used to measure prediction error.
Early Stopping Patience	10	Number of epochs to wait for improvement before halting training.

Table 9. 1DCNN architecture and key parameters -UAV-GCS

Parameter	Setting	Description
Input Shape	(44, 1)	The dimensions of the input data representing 44 features per sample.
Conv1D Filters	64, 128	The number of unique patterns or features the model extracts at different layers.
Kernel Size	3	The size of the sliding window used to capture local patterns in the data.
Activation	ReLU	A non-linear function that helps the model learn complex patterns.
MaxPooling	2	A technique that reduces data size by keeping only important features.
Dropout	0.2, 0.5	A regularization method that prevents overfitting by disabling neurons.
Dense Neurons	128	The number of units in the hidden layer to interpret extracted features.
Output Units	8	The total number of unique classes the model classifies.
Optimizer	Adam	An efficient algorithm used to update model weights during training.
Loss Function	Sparse Categorical Crossentropy	The metric used to calculate the error for multi-class labels.
Batch Size	64	The number of training samples processed before updating weights.
Epochs	20	The total number of complete passes through the training dataset.

Software and Experimental Environment

All experiments have been coded in Python 3.10. The models implemented were: Classical machine learning models (Logistic Regression, Random Forest): scikit-learn, deep learning models (FFNN, 1D-CNN, LSTM): TensorFlow/Keras, LightGBM: official LightGBM Python package. Scikit-learn and pandas were used for the data preprocessing tasks, such as normalization, encoding, and train-test partitioning. The array operations were performed using NumPy.

Ablation Study

To determine which design decisions influence the detection performance most, an ablation study was designed, which involved the following: (i) removing the categorical feature groups and retraining the best performing model, LightGBM, (ii) testing the impact of class-imbalance handling by comparing model performance with and without class-weighting to the minority classes of UAV-GCS-IDS, and (iii) testing the impact of network depth on the deep learning models, comparing the two-layer LSTM configuration from the reported results in table 8 and table 9 with a single-layer configuration.

RESULTS AND DISCUSSIONS

To comprehensively evaluate the performance of the proposed classifiers, the experimental results were reported using standard classification metrics such as Precision, Recall, F1-Score, and Accuracy. The weighted average values for the standard classification metrics was adopted as the primary aggregated metric, as it accounts for class imbalance by weighting each class's contribution proportionally to its support in the test set and it enable a holistic assessment of overall model behavior across all traffic categories. Furthermore, confusion matrices are provided to visually illustrate the classification outcome at the individual class level, revealing each model's ability to correctly distinguish between normal and various attack types that encompass UAV-IDS2025 and UAV-GCS Datasets.

Tables 10, and 11, tables 12 and 13 highlight a comparative assessment of ML and DL based on Accuracy, Precision, Recall, and F1-score respectively. While the logistic regression provided reasonable predictions, its performance remained the lowest among all evaluated algorithms. The notably lower accuracy may be attributed to its linear decision boundary, which may be insufficient to

capture the complexity of both UAV-IDS-2025 and UAV_GCS datasets. Moreover, among all examined classifiers, LightGBM demonstrated superior performance, achieving perfect scores of 100% for accuracy, precision, recall, and F1-score, surpassing all other competing algorithms. Thereby demonstrating its absolute effectiveness in the classification task.

Table 10. Comparative assessment of ML and DL algorithms based on weighted average accuracy

Accuracy						
Dataset	LightGBM	Random Forest	Logistic Regression	FFN	LSTM	1D-CNN
UAVIDS-2025 [16]	1.000000	0.999673	0.8679760	0.997654	0.9878	0.9957
UAV-GCS [4]	0.990429	0.888922	0.8505036	0.882800	0.8972	0.8842

Table 11. Comparative assessment of ML and DL algorithms based on weighted average precision

Precision						
Dataset	LightGBM	Random Forest	Logistic Regression	FFN	LSTM	1D-CNN
UAVIDS-2025 [16]	1.000000	0.999673	0.870055	0.997654	0.9878	0.9958
UAV-GCS [4]	0.990169	0.891333	0.793929	0.8842	0.9003	0.8948

Table 12. Comparative assessment of ML and DL algorithms based on weighted average recall

Recall						
Dataset	LightGBM	Random Forest	Logistic Regression	FFN	LSTM	1D-CNN
UAVIDS-2025 [16]	1.000000	0.999673	0.867976	0.997654	0.9878	0.9957
UAV-GCS [4]	0.990429	0.888922	0.850536	0.8828	0.8967	0.8842

Table 13. Comparative assessment of ML and DL algorithms based on weighted average F1 - score

F1 -Score						
Dataset	LightGBM	Random Forest	Logistic Regression	FFN	LSTM	1D-CNN
UAVIDS-2025 [16]	1.000000	0.999673	0.868680	0.997654	0.9878	0.9957
UAV-GCS [4]	0.989346	0.863234	0.800383	0.8517	0.8747	0.8537

Comparison with State-of-the-Art UAV Intrusion Detection Methods

Table 14. Comparison of this study's best-performing model against recent state-of-the-art UAV IDS methods

Study	Method	Dataset	Accuracy	F1-score
Gouveia et al., (2020)	HybridQNN-8L (quantum ML)	UAVIDS-2025	94.8%	96.7%
Zeng et al., (2025)	XGBoost	UAVIDS-2025	98.7%	97.9%
Priyadarshini et al., (2025)	Hist-GBoost	Internet of Drones traffic	99.42%	99.42%
Reference Hadi, (2024)	FFCNN	UAVIDS	98.23%	-
This study	LightGBM	UAVIDS-2025	100.00%	100.00%
This study	LightGBM	UAV-GCS-IDS	99.04%	98.93%

The comparison in table 14 shows that LightGBM, implemented in this study, is as accurate or more accurate than recent state-of-the-art approaches tested on UAVIDS-2025, and is evaluated on a second and independent dataset (UAV-GCS-IDS) which has not been assessed by most previous studies. For real-world UAV deployment scenarios, with differing traffic characteristics across networks and missions, this cross-dataset consistency is presented as the main evidence of the reliability of the proposed framework.

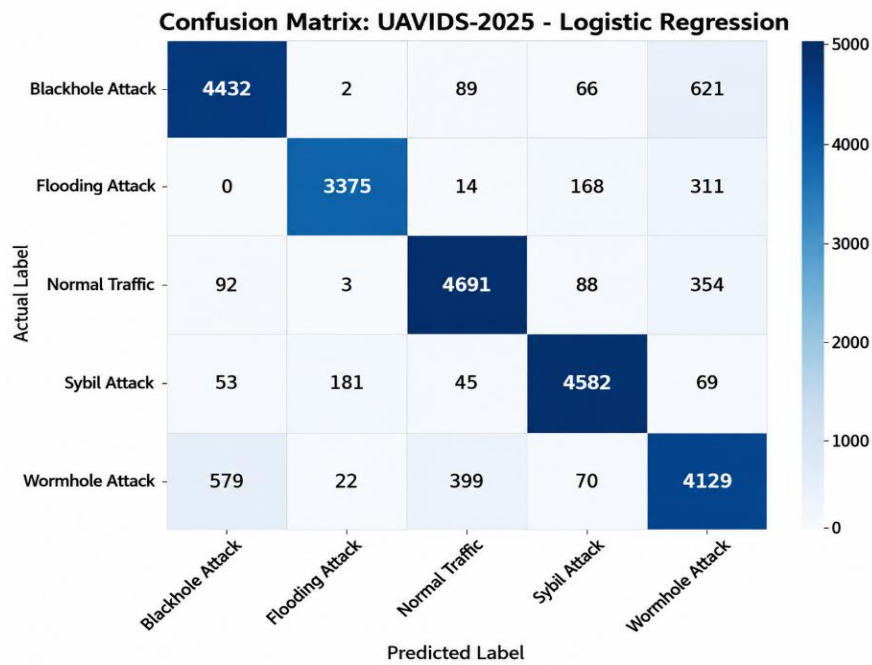


Figure 2. Confusion matrix - UAVIDS 2025 - logistic regression

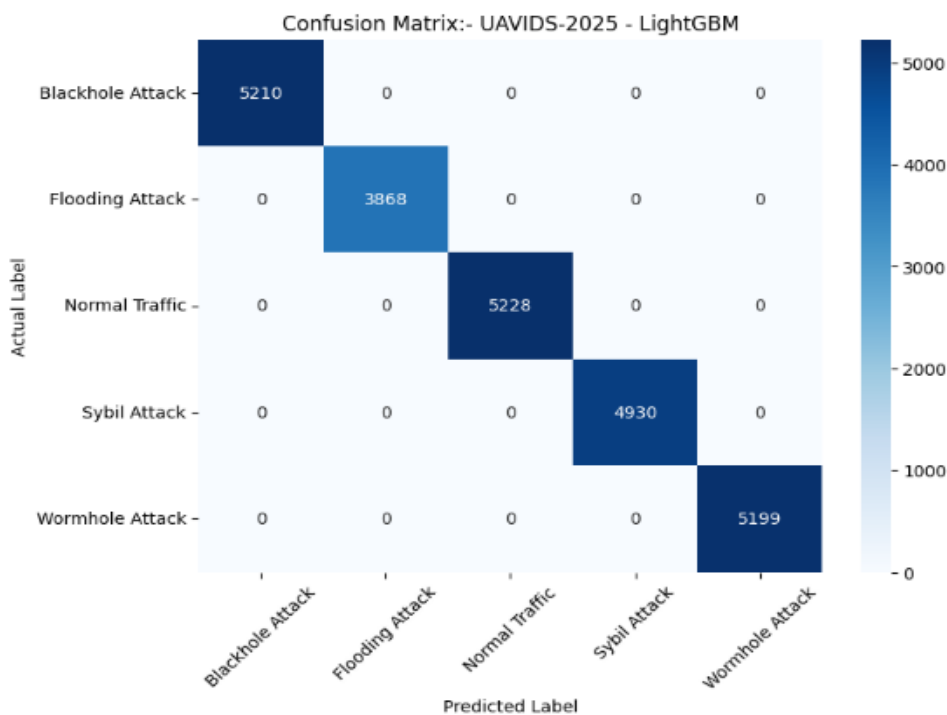


Figure 3. Confusion matrix -UAVIDS-2025 - LightGBM

Figure 2 presents the confusion matrix of logistic regression classifier on UAV-IDS-2025 dataset, illustrating the classification performances across all traffic categories. Here, Logistic Regression provides a moderately effective baseline for multi-class intrusion detection in UAV network environments. Moreover, the most prominent source of classification error is the Wormhole Attack class which was frequently confused with the Blackhole Attack class and Normal Traffic, suggesting that the decision boundary between these two attack types is insufficiently resolved by a linear classifier. Furthermore, the UAV-IDS-2025 indicates that the Logistic Regression model struggles with inter-class boundary disambiguation, particularly among routing-layer attacks (Blackhole and Wormhole), which share topological disruption characteristics.

Figure 3 depicts the confusion matrix obtained by LightGBM classifier on the UAV-IDS-2025 dataset, where rows represent the actual classes and columns represent the predicted classes. The diagonal elements of the confusion matrix confirm that the model correctly classified all of instances across all classes. These findings suggest that LightGBM is the Top classifier for multi-class intrusion detection in UAV network environments.

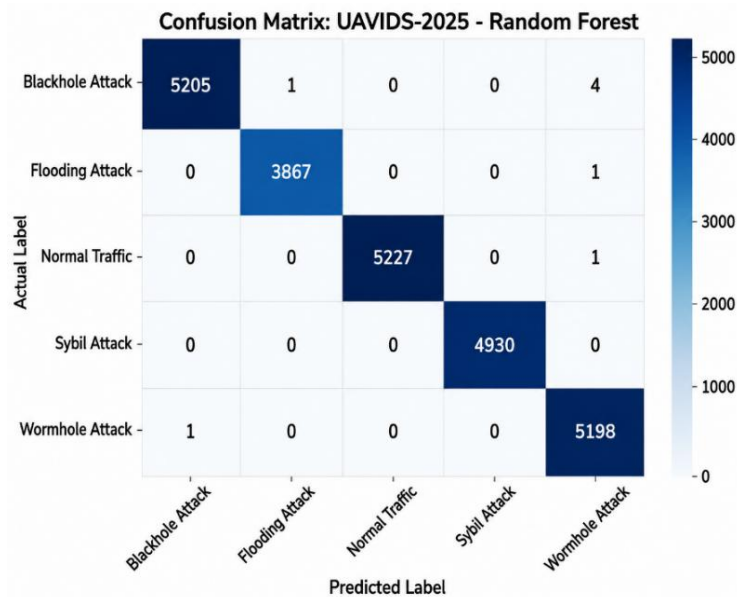


Figure 4. Confusion matrix -UAVIDS-2025 - random forest

To further examine the per-class behavior of Random Forest classifier on UAVIDS-2025 dataset, the confusion matrix is depicted in figure 4. As revealed from figure 4 there were 24,427 instances that were correctly classified, and only 8 instances were misclassified. Comparing these confusion matrices together shows an improvement by Random Forest as well as LightGBM in comparison with logistic regression. The data strongly suggest that the ensemble-based, non-linear nature of Random Forest and LightGBM is far better suited to capturing the complex decision boundaries inherent in UAV network traffic classification. These findings are of considerable practical significance, as they indicate that Random Forest and LightGBM may serve as a robust and reliable foundation for real-time intrusion detection in UAV communication environments.

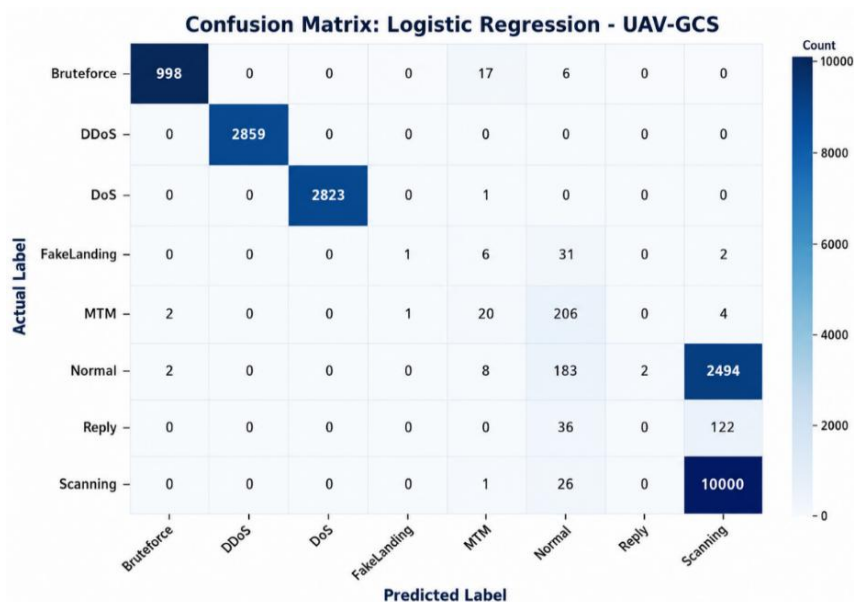


Figure 5. Confusion matrix -UAV -GCS - logistic regression

The confusion matrix in figure 5 demonstrates the classification performance of the Logistic regression classifier evaluated on UAVIDS-2025. As evidenced by diagonal values, Brute force, DDoS and DoS attacks are classified with near-perfect accuracy and negligible misclassifications. Fake landing and MITM showed poor performance due to their limited instances. Normal traffic is frequently misclassified as Scanning (2494 instances), indicating difficulty in distinguishing benign communication from scanning activity. Moreover, the data strongly suggests that the linear nature of logistic regression is ineffective for capturing the complex decision boundaries inherent in UAV network Normal, Scanning, and Reply attack traffic classification.

The confusion matrix in figure 6 displays the classification performance of the LightGBM classifier on the UAV-GCS dataset. The Fake landing and Replay attack evidenced poor performance, attributed to their limited sample size. Normal traffic was accurately classified, with relatively small 17 instances misclassified, which indicates minimal challenge in distinguishing benign communication from scanning activity. However, the most notable misclassification was observed between Reply and Normal, where 83 Reply class instances were incorrectly predicted as Normal class instances.

The confusion matrix in figure 7 displays the classification performance of the Random Forest classifier evaluated on UAV-GCS dataset. The matrix demonstrated underperformance across Fake landing, Reply attacks, and Normal traffic. Moreover, it reveals that Normal class exhibited the highest false negative rate, suggesting that its traffic patterns share strong similarities with Scanning class.

The foregoing results demonstrate the overall classification performance of the machine learning models evaluated on UAVIDS-2025 as well as UAV-GCS datasets. Having presented the performance of the classical machine learning classifiers, the following section reports the results obtained by deep learning models. All models FFN, LSTM, and One-Dimensional CNN were constructed using standard, classical configurations and implemented through typical procedures commonly adopted in deep learning fields.

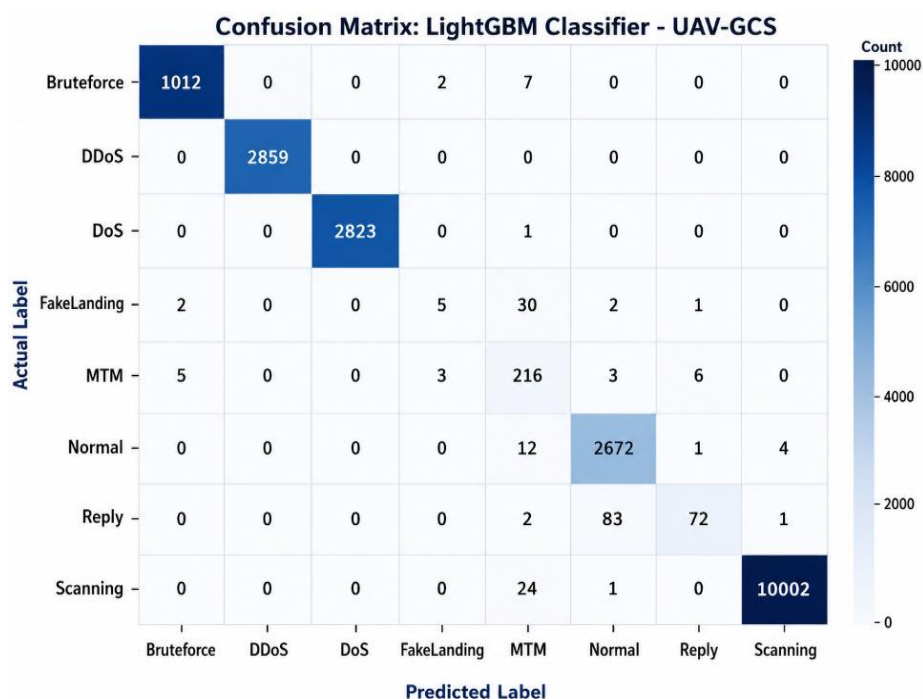


Figure 6. Confusion matrix -UAV -GCS - LIGHTGBM

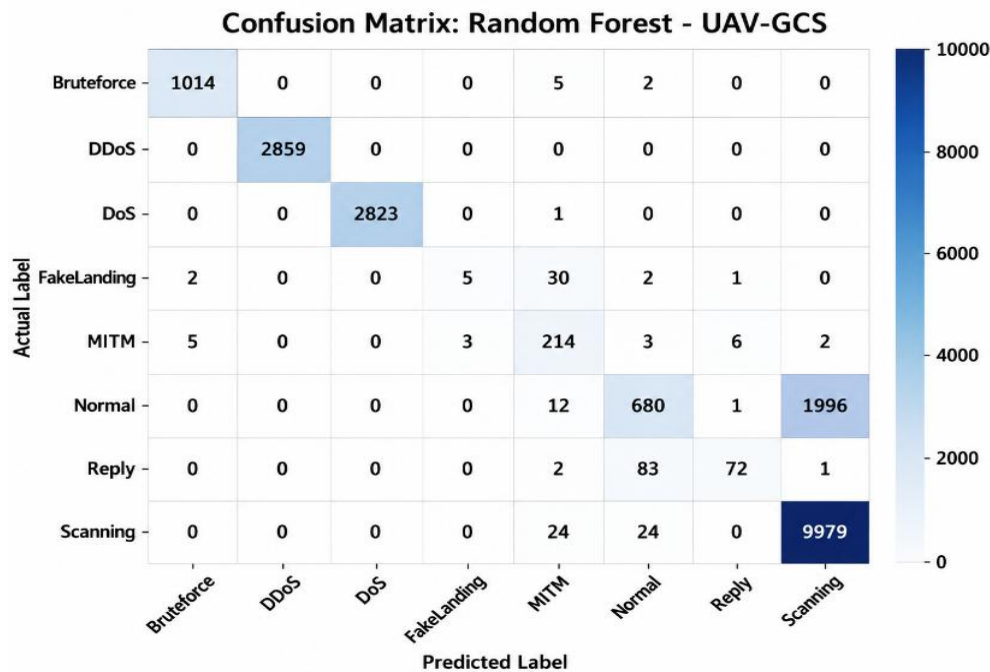


Figure 7. Confusion Matrix -UAV -GCS - random forest

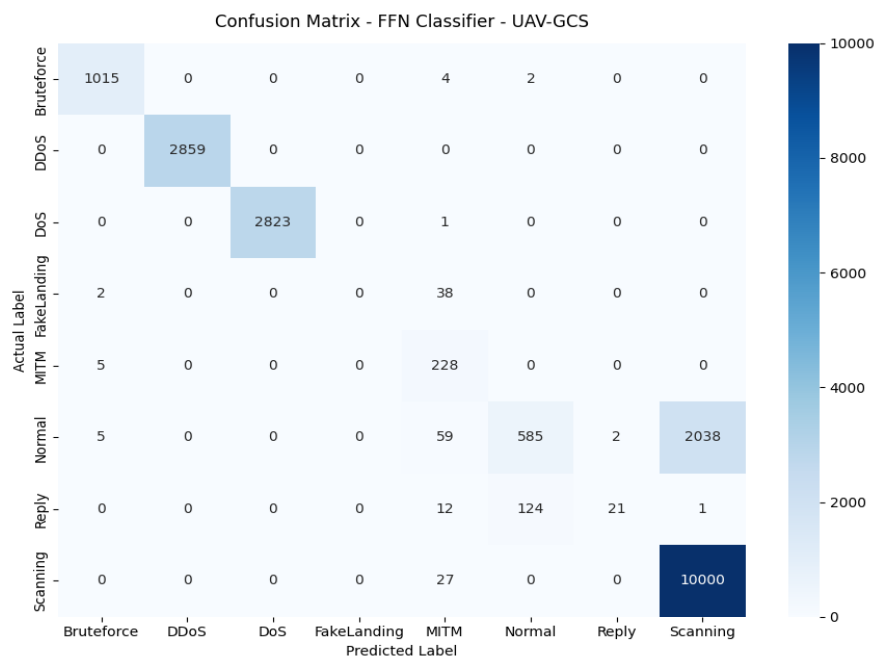


Figure 8. Confusion matrix -UAV -GCS FFN

Figure 8, figure 9 and figure 10 illustrate the confusion matrices of FNN, LSTM. and 1DCNN evaluated on UAVGCS dataset, while figures 11 and figure 12 and figure 13 present the corresponding result on UAVIDS-2025 dataset respectively, illustrating the per-class classification behavior across all traffic categories. Consistent with the results observed on UAVGCS, FFN, 1DCNN, and LSTM models maintained high true positive rates across all classes when evaluated on UAVIDS-2025, demonstrating stable generalization capability. Thus, indicating a highly effective classification capability.

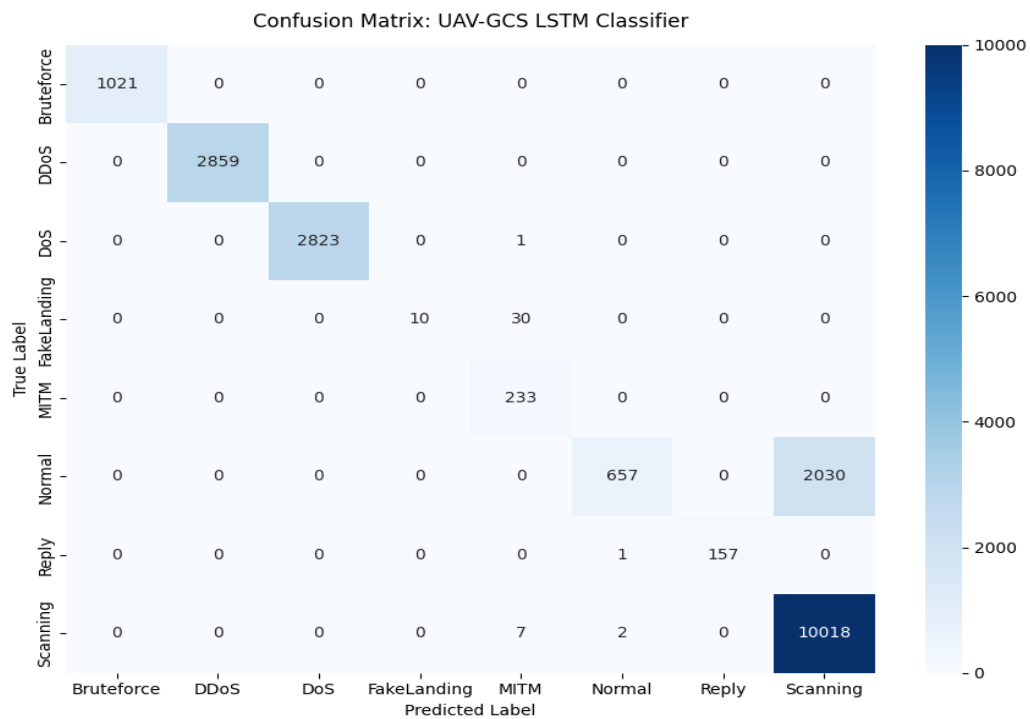


Figure 9. Confusion matrix -UAV -GCS LSTM

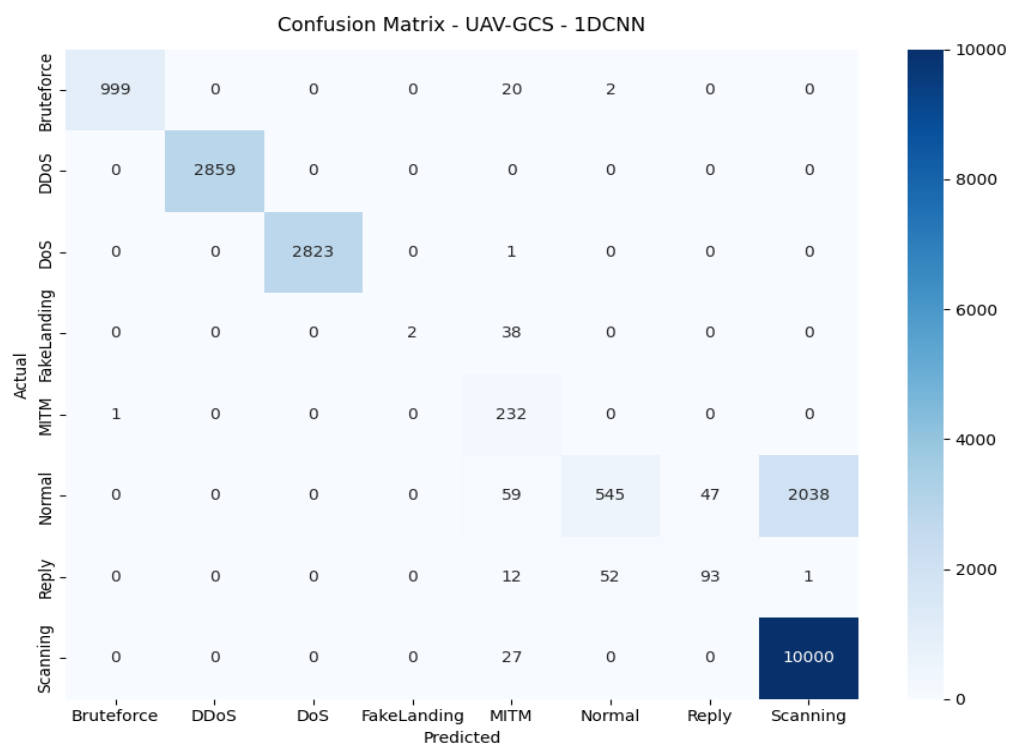


Figure 10. Confusion matrix -UAV -GCS 1DCNN

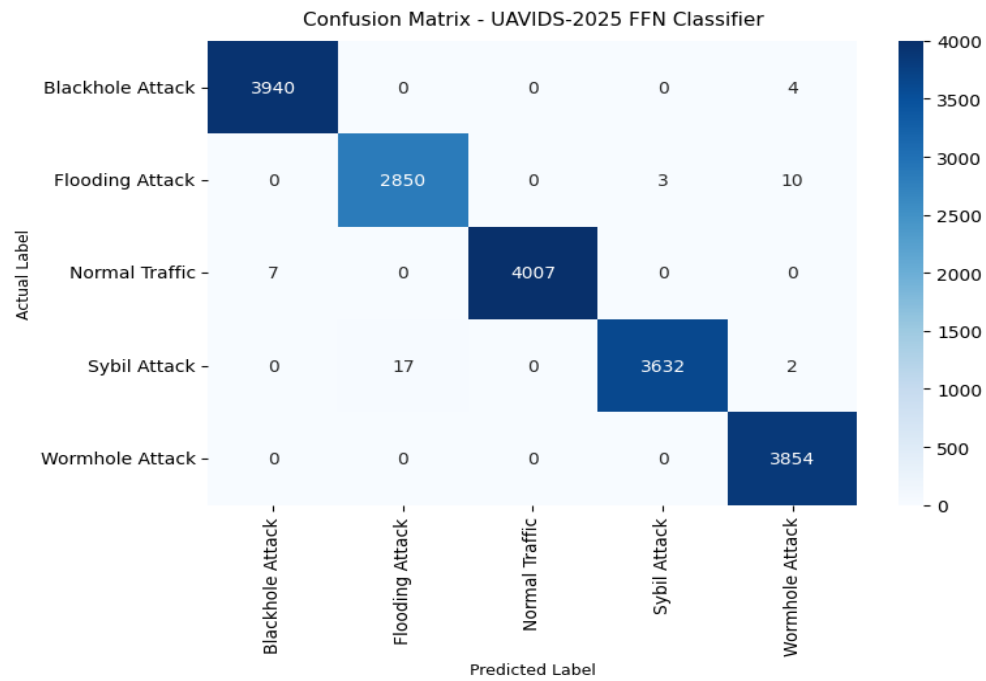


Figure 11. Confusion matrix -UVIDS -2025 FNN

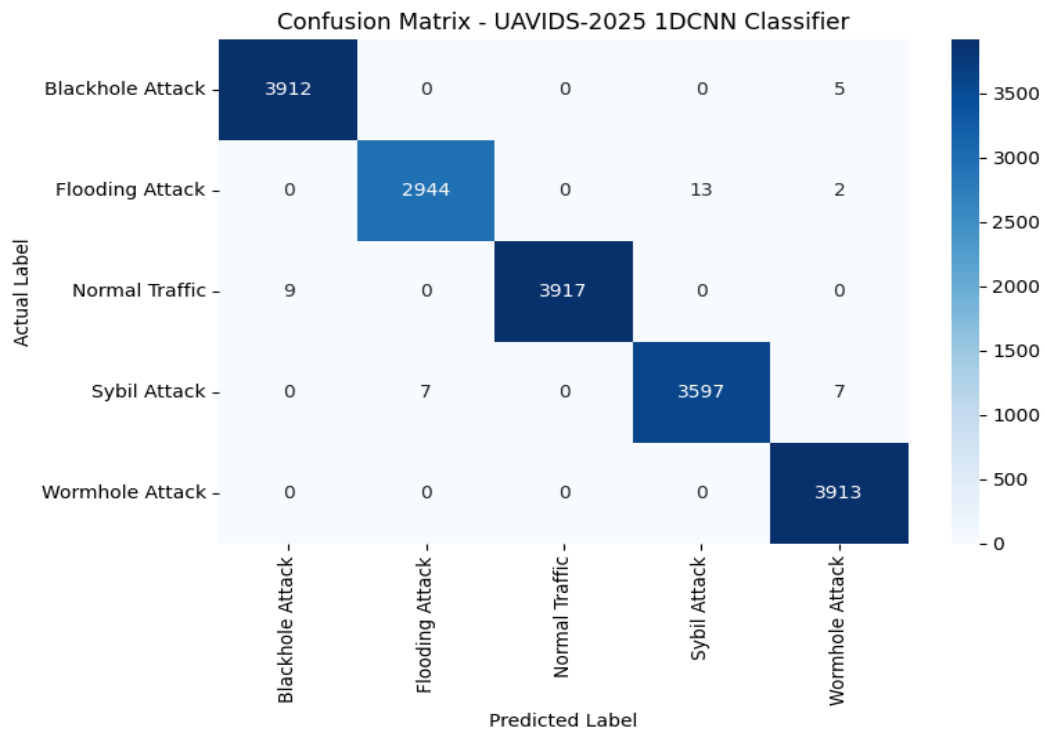


Figure 12. Confusion matrix -UVIDS -2025 LSTM

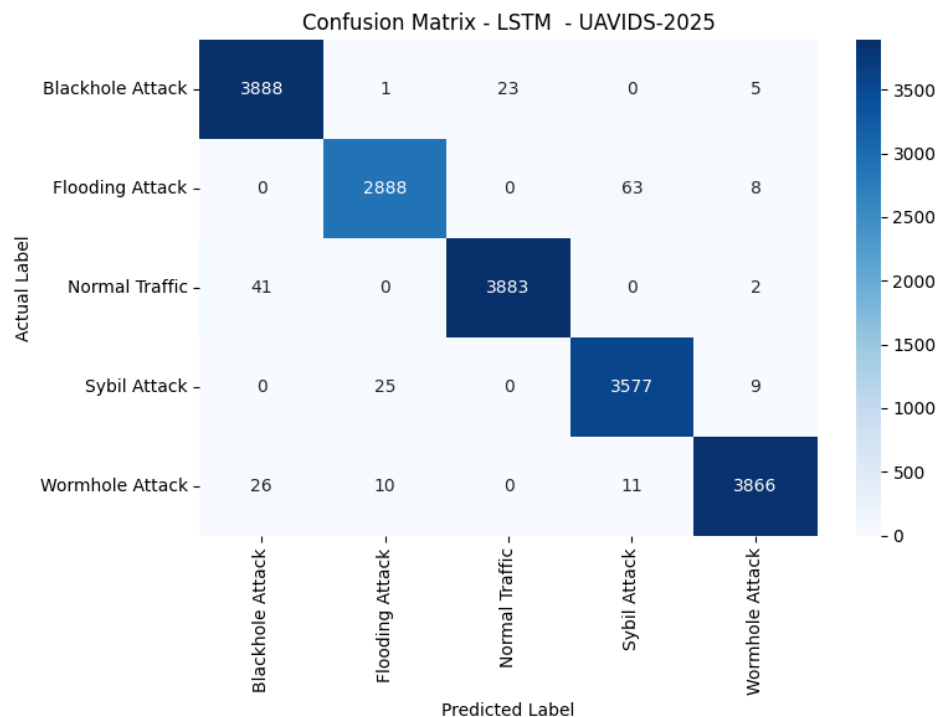


Figure 13. Confusion matrix -UVIDS -2025 IDCNN

Limitations

There are a number of limitations associated with this study that need to be taken into account when interpreting the results. The validity of the results derived from the UVIDS-2025 dataset is limited, as it is a simulated dataset, and thus may not fully reflect the performance of a live UAV deployment, where there will be hardware noise and variable wireless channel conditions. Second, the deep learning architectures were tested with standard, non-optimized architecture and hyperparameter settings, which could mean that they are not as capable as they could be if they had undergone more extensive hyperparameter optimization, and were compared with the ensemble methods, which are also more heavily optimized. Third, some attack classes in both datasets have few instances and therefore, per-class metrics for these classes are not sufficiently statistically reliable. Lastly, the study assesses the accuracy of the classification and relevant statistical performance measures but does not test inference latency, memory footprint, and energy consumption on embedded UAV hardware—factors that are key to deployment options for resource-limited platforms.

Implications

Results have implications for the design of UAV IDS. With a very consistent performance on both datasets, gradient-boosted ensemble methods are the best method of choice when computational resources are limited on the UAV platform, because they offer higher accuracy compared to deep neural architectures and their training and inference costs are comparatively lower. The sub-optimal performance of Logistic Regression on both the datasets suggests that linear classifiers are not the only classifiers to be used in UAV networks when attacks like Blackhole and Wormhole are taking place, which are non-linear, topologically structured. Based on the confusion patterns identified in Section 5.2, especially for the Replay-Normal and Wormhole-Blackhole classes, an operational system that uses a fast ensemble pre-filter coupled with a secondary dedicated classifier for ambiguous classes (Replay-Normal and Wormhole-Blackhole) could further enhance the reliability of the operational system.

Future Scope

The four directions for future research are identified. A first study of ablation indicates that the contribution of each feature group and model design should be quantified to determine the impact of each on the detection performance. Second, the evaluation of the ease of real-time deployment should be conducted on representative embedded UAV platforms such as Raspberry Pi or NVIDIA Jetson-class devices to assess the feasibility of real-time deployment, as they are more prone to streaming constraints that real-world UAVs may encounter. Third, the graph neural network (GNN) architectures need to be explored to explicitly consider the relational dependencies between UAVs in the communication aspect, which would enhance intrusion detection in the UAV swarm communication system in high-dynamic network scenarios [22]. Last but not least, the comparison should be expanded to federated learning and online learning paradigms, so that UAVs can work together without sharing flight-sensitive data with centralized servers to update an intrusion detection model. Moreover, future research is needed to assess the resilience of these models in the face of adversarial evasion attack in real world scenarios to enhance the security and reliability of the UAV intrusion detection systems in real-world environments.

CONCLUSION

In this research, have compared machine learning and deep learning models for intrusion detection in UAV networks in a comprehensive manner by using two recently published benchmark datasets, namely, UVIDS-2025 and UAV-GCS-IDS. To compare the algorithms fairly and reproducibly, three conventional machine learning algorithms (Logistic Regression, Random Forest and LightGBM) and three deep learning architectures (Feedforward Neural Network (FFNN), Long Short-Term Memory (LSTM) and One-Dimensional Convolutional Neural Network (1D-CNN)) were tested with the same experimental setup. The study didn't advocate a new architecture for intrusion detection, instead it set a solid baseline performance for a set of common classification models. The LightGBM classifier consistently outperformed other models in the experimental results on both datasets. Compared to the Random Forest model and the three deep learning models, LightGBM's performance on the UVIDS-2025 dataset resulted in 100.0% accuracy, precision, recall, and F1-score, and on the UAV-GCS-IDS dataset in 99.04% accuracy with an F1-score of 98.93%. Logistic Regression always had the lowest detection performance with 86.8-87.0% accuracy for UVIDS-2025 and 85.1% for UAV-GCS-IDS. However, among the deep learning methods, FFNN outperformed the other three with 97.65% accuracy on UVIDS-2025 while the accuracies of the other three models were in the range of 88.2%-89.7% on the more challenging and imbalanced UAV-GCS-IDS. In addition, the confusion matrix analysis highlighted the fact that the routing-layer attacks (Blackhole and Wormhole) and the low-frequency attack classes (Fake Landing, Replay, and Man-in-the-Middle (MITM) attacks), are the most challenging categories to accurately classify among all of the models tested. Overall, it serves as a solid reference to assess intrusion detection models in UAV networks and proves that Gradient boosted ensemble learning is still the best approach for the current UAV IDS applications. Directions for future research include feature-level ablation analysis, deployment on resource-constrained embedded UAV hardware, graph neural network-based intrusion detection, federated and online learning frameworks, and adversarially robust model design, to help build secure, scalable, and real-time UAV intrusion detection systems.

References

- [1] Gouveia A, Correia M. Towards quantum-enhanced machine learning for network intrusion detection. In 2020 IEEE 19th international symposium on Network Computing and Applications (NCA) 2020 Nov 24 (pp. 1-8). IEEE. <https://doi.org/10.1109/NCA51143.2020.9306691>
- [2] Kiran Kumar P, Gupta Y, Goyal A, Kandasamy D, Seenu V, Kumar R, Pal S, Bal CS, Aggarwal S, Tandon N. Clinical and Genetic Profiling of Index Patients with Multiple Endocrine Neoplasia Type 2 Syndrome and Screening for Medullary Thyroid Carcinoma in Asymptomatic First-degree Relatives. *Indian Journal of Endocrinology and Metabolism*. 2022 Dec;26(Suppl 8):S83. <https://doi.org/10.4103/2230-8210.363700>
- [3] Mouatassim T, El Guarmah EM, Airaj M. A comprehensive literature review on intrusion detection datasets for UAVs: taxonomy, challenges, future directions, and recommendations. *Journal of Cyber Security Technology*. 2026 Dec 15;10(1):2651709. <https://doi.org/10.1080/23742917.2026.2651709>

- [4] Hadi H, Hadi H. UAV-GCS intrusion detection dataset (UAV-GCS-IDS). 2025 <https://doi.org/10.6084/m9.figshare.29608541.v1>
- [5] Syed F, Gupta SK, Hamood Alsamhi S, Rashid M, Liu X. A survey on recent optimal techniques for securing unmanned aerial vehicles applications. *Transactions on Emerging Telecommunications Technologies*. 2021 Jul;32(7):e4133. <https://doi.org/10.1002/ett.4133>
- [6] Breiman L. Random forests. *Machine Learning*. 2001 Oct;45(1):5-32. <https://doi.org/10.1023/A:1010933404324>
- [7] Fayyad UM, Piatetsky-Shapiro G, Uthurusamy R. Summary from the KDD-03 panel: data mining: the next 10 years. *ACM Sigkdd Explorations Newsletter*. 2003 Dec 1;5(2):191-6. <https://doi.org/10.1145/980972.981004>
- [8] Fatima N, Liu L, Hong S, Ahmed H. Prediction of breast cancer, comparative review of machine learning techniques, and their analysis. *IEEE Access*. 2020 Aug 14;8:150360-76. <https://doi.org/10.1109/ACCESS.2020.3016715>
- [9] Zha D, Bhat ZP, Lai KH, Yang F, Jiang Z, Zhong S, Hu X. Data-centric artificial intelligence: A survey. *ACM Computing Surveys*. 2025 Jan 24;57(5):1-42. <https://doi.org/10.1145/3711118>
- [10] Abdulhammed R, Faezipour M, Abuzneid A, AbuMallouh A. Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network traffic. *IEEE sensors letters*. 2018 Nov 7;3(1):1-4. <https://doi.org/10.1109/LSSENS.2018.2879990>
- [11] Jeong HW, Kim HG, Choi YH. Lstm-based network intrusion detection system and solving data imbalance problem through gan. In 2025 International Conference on Artificial Intelligence in Information and Communication (ICAIC) 2025 Feb 18 (pp. 1156-1159). IEEE. <https://doi.org/10.1109/ICAIC64266.2025.10920641>
- [12] Thamilarasu G, Chawla S. Towards deep-learning-driven intrusion detection for the internet of things. *Sensors*. 2019 Apr 27;19(9):1-19. <https://doi.org/10.3390/s19091977>
- [13] Abiodun OI, Jantan A, Omolara AE, Dada KV, Umar AM, Linus OU, Arshad H, Kazaure AA, Gana U, Kiru MU. Comprehensive review of artificial neural network applications to pattern recognition. *IEEE access*. 2019 Oct 4;7:158820-46. <https://doi.org/10.1109/ACCESS.2019.2945545>
- [14] Zhao D, Jiang R, Feng M, Yang J, Wang Y, Hou X, Wang X. A deep learning algorithm based on 1D CNN-LSTM for automatic sleep staging. *Technology and Health Care*. 2022 Mar;30(2):323-36. <https://doi.org/10.3233/THC-212847>
- [15] Priyadarshini S, Nayak J, Dash PB, Tanty R, Pattanaik S. Intelligent Intrusion Detection for Unmanned Aerial Vehicles to Enhance Defense Security System: An Improved Ensemble Learning Approach. In 2025 7th International Conference on Information Systems and Computer Networks (ISCON) 2025 Sep 5 (pp. 1-6). IEEE. <https://doi.org/10.1109/ISCON65210.2025.11341301>
- [16] Zeng Q, Bashir A, Nait-Abdesselam F. Uavids-2025: A benchmark dataset for intrusion detection in uav networks using machine learning techniques. In 2025 IEEE Conference on Communications and Network Security (CNS) 2025 Sep 8 (pp. 1-9). IEEE. <https://doi.org/10.1109/CNS66487.2025.11194990>
- [17] Hadi HJ, Cao Y, Li S, Hu Y, Wang J, Wang S. Real-time collaborative intrusion detection system in UAV networks using deep learning. *IEEE Internet of Things Journal*. 2024 Jul 11;11(20):33371-91. <https://doi.org/10.1109/JIOT.2024.3426511>
- [18] Belmekki BE, Aljohani AJ, Althubaity SA, Al Harthi A, Bean K, Aijaz A, Alouini MS. Cellular network from the sky: Toward people-centered smart communities. *IEEE Open Journal of the Communications Society*. 2024 Mar 18;5:1916-36. <https://doi.org/10.1109/OJCOMS.2024.3378297>
- [19] Adel A, Jan T. Watch the skies: a study on drone attack vectors, forensic approaches, and persisting security challenges. *Future internet*. 2024 Jul 13;16(7):250. <https://doi.org/10.3390/fi16070250>
- [20] Najem N, Abdulhammed R. A Review of Existence Intrusions Detection-Based Machine Learning Datasets of Future Generation Networks. *NTU Journal of Engineering and Technology*. 2025 Jun 23;4(2):1-10. <https://doi.org/10.56286/y86yy478>