Safe-UAV: An Explainable AI-assisted Framework for Securing UAV Communication Networks

Dev Mehta*, Kathan Patel†, Janam Patel‡, Rajesh Gupta§, Sudeep Tanwar¶, Ankur Gupta‖, Isaac Woungang**
*‡§¶Department of Computer Science and Engineering, Institute of Technology, Nirma University, Gujarat, India
†Department of IT, G H Patel College of Engg. and Tech., Charutar Vidya Mandal (CVM) University, India
‖Department of Computer Science & Engineering, Model Institute of Engineering & Technology, J&K, India
**Department of Computer Science, Toronto Metropolitan University, Toronto, ON, Canada
Emails: *22bcm015@nirmauni.ac.in, †kathanpatelkp10@gmail.com, ‡21bce197@nirmauni.ac.in, §rajesh.gupta@nirmauni.ac.in, ¶sudeep.tanwar@nirmauni.ac.in, ‖ankurgupta@mietjammu.in, **iwoungan@torontomu.ca,

***Abstract*—Technology is progressing, leading to the widespread adoption of Unmanned Aerial Vehicles (UAVs) in multiple industries. Nevertheless, security remains the top priority, especially when it comes to identifying and stopping drone attacks. This paper suggests using machine learning and ensemble learning in system model to effectively detect and prevent attacks before they happen. A major obstacle is reducing both false positives and false negatives, since incorrectly identifying valid nodes as intruders, or vice versa, can impact the security system's efficiency. In response to this issue, the paper presents a *Safe-UAV* framework for detecting and categorizing attacks, using XGBoost, which achieves an accuracy of 92.87% as well as strong precision, F1 score, and recall measures. Moreover, the research utilizes LIME and SHAP, important tools in XAI, to improve the transparency and reliability of the model's forecasts. LIME focuses on explaining the model's actions for individual cases, while SHAP provides consistent explanations rooted in Shapley values. This mix of advanced detection methods and interpretability tools enhances the transparency and dependability of the *Safe-UAV* framework.**

***Index Terms*—Ensemble learning, LIME, SHAP, UAV networks, XAI**

I. Introduction

In the recent years, many revolutionary computing devices and communication networks have been evolved. Unmanned Aerial Vehicles (UAV) are aerial technology which interact with the environment and provide value added services across various domains. Additionally, UAVs solves the problem of depending only upon fixed terrestrial IoT infrastructure [1]. UAVs are intelligent Cyber-Physical Systems (CPS) that rely on functional sensors, real-time information communication, and flight control systems [2]. These are now widely used in the fields of Remote Sensing and mapping, surveillance, agriculture, environmental monitoring [3]. Due to rapid growth in UAV technology, it brings immense risks of cyber-attack, physical-attack, robust detection and mitigation strategies. Malicious entities can attack UAV's Air to Air or Air to Ground wireless channels, leading to potential information leakage [4].

The communication links in UAV networks can be attacked by hackers in two ways: jamming attack and spoofing attack. A jamming attack breaks UAV communication by flooding the channel with radio signals, violating mac layer protocols, and causing a Denial of Service (DoS) [5]. A spoofing attack targets positioning and address systems, altering signals to distort UAV navigation. Hackers spoof GPS to mislead UAVs. These attacks compromise data integrity and network availability. The architecture of the UAV should be robust to resist cyber-attacks. Based on past data several Machine Learning (ML) and Deep Learning (DL) algorithms can spot anomaly patterns. Modern Intrusion Detection Systems (IDS) adapt to evolving attack patterns [6]. Network security is enhanced by classifying network traffic as benign or malicious. This allows for timely actions to prevent or mitigate attacks.

Advanced ML-DL algorithms are complex and involve a huge number of weights. This complexity makes it difficult for users to understand the results [7]. Making decisions without a proper explanation of the results can lead to security, privacy, and legal concerns. Hence, there is a need for Explainable AI (XAI). XAI can help in understanding the complex AI models of network traffic, identifying attacking nodes, and can quantify feature influence, provide attack certainty [8]. By enhancing the interpretability of ensemble learning models, XAI helps identify model weaknesses, leading to more accurate and reliable outcomes. Confluence of Ensemble Learning (EL) and XAI has wide use in healthcare [9], finance [10], autonomous driving [11]. In UAV communication networks, XAI can identify and mitigate cyber attacks by explaining anomalies in network traffic and UAV co-ordination.

Hong et al. in [12] addressed the limitation of previous DL based studies on spoofing attacks. It proposes EL model for intrusion detection, with accuracy of 96%. Bayrak et al. in [13] developed a linear SVM model, that is enhanced with LIME based XAI technique. They fine-tuned model using Bayesian optimization to protect UAVs from malware threats. In [14], Wei et al. introduced a AI-based model for Urban Air Mobility(UAM). Model is decentralised for safe data access and sharing between AVs and ground stations. This model promotes low latency and high explainable using SHapley Additive exPlanations (SHAP) XAI. Shafique et al. [15] employed SVM and K-folded ML techniques and GPS based features to identify spoofing. In [16], Thaker et al. proposed an EL based IDS that identifies malicious CAN traffic of the AVs. This model achieves highest accuracy of

98.57%. Motivated from the current researches in UAV network intrusions, we propose a robust and scalable EL model and introduce the explainablity using LIME and SHAP techniques.

A. Research Contributions

Research contribution of our model is listed below as:

- We propose a lightweight *Safe-UAV* framework for Intrusion detection and then Attack classification on UAV communication network data. We used a standard Cyber-Physical UAV simulation dataset, and employed Ensemble classifiers such as XGBoost, Bagging, Random forest.
- We used different XAI techniques to give explanation of *Safe-UAV* framework to the network administrator. LIME model gives explanation for a local instance, whereas SHAP features explain the overall model.
- The proposed *Safe-UAV* framework is evaluated on various parameters of model training and testing. Accuracy, precision, recall, f1 score, and ROC-AUC evaluates the effectiveness of model.

B. Paper Organization

The findings of our paper are organised into various sections. Section II explains in detail the proposed approach along with the system model and problem formulation. Results analysis and discussion of our model is in Section III. Section IV provides the conclusion for our research model.

II. The Proposed *Safe-UAV* Framework

Fig. 1 shows the proposed *Safe-UAV* framework with UAV networks, data assembling and transformations, attack classification, and XAI layer. The icons used in diagram are downloaded from [17].

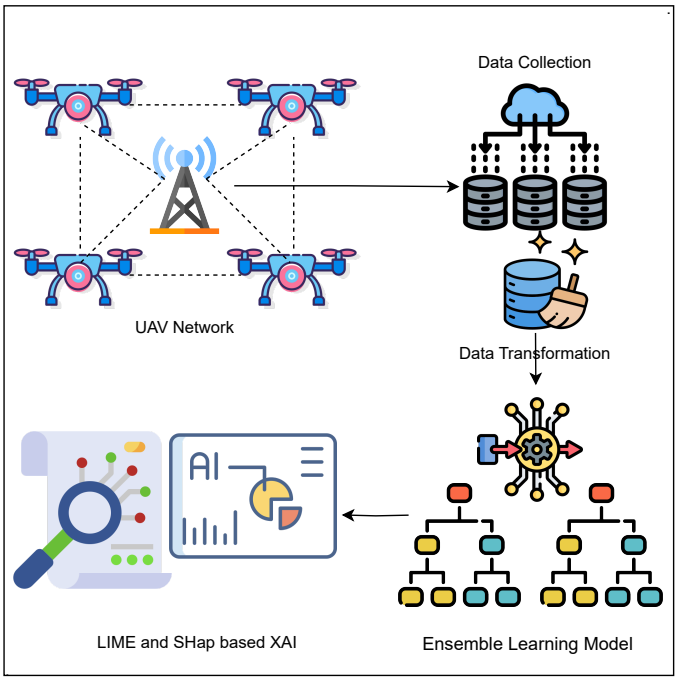

Fig. 1: The proposed *Safe-UAV* framework.

A. System Model

The system model comprises of different UAVs such that $\{\delta_1,\delta_2,\delta_3,...,\delta_n\} \in \Delta$. To assist the Δ network, their exist certain ground stations such that $\{\gamma_1,\gamma_2,\gamma_3,...,\gamma_m\} \in \Gamma$. The UAVs can directly exchange information ϕ with each other, and it can also communicate with ground station Δ.
Primarily, ML model μ_1 is used to detect whether the UAV network is exploited or not based on physical attributes eta_1. Further, a robust EL model μ_2 is used to further classify the attack A using the network attributes η_2.

$$A \in \{\text{Benign, DoS, Replay, Evil Twin, FDI}\} \tag{1}$$

The local instance explaining model LIME Υ and global context interpreter model SHAP Ω assist the decision making of μ_1 and μ_2. The overall objective of the proposed scheme is to maximise the prediction accuracy of attack type, and explainability Π of overall system.

$$\Psi = \max \sum_{j=1}^{n} \mu_i(\eta_i), \ \ (i = 1, 2) \tag{2}$$

$$\Pi = \max(\Upsilon(\mu_2) + \Omega(\mu_2)) \tag{3}$$

Algorithm 1 Data Preparation, Model Training, and Interpretation

Input: Main dataset D
$\hat{D}_{\text{physical}} \leftarrow \text{SelectFeatures}(D, \text{Attributes} = \text{Physical})$
$\hat{D}_{\text{cyber}} \leftarrow \text{SelectFeatures}(D, \text{Attributes} = \text{Cyber})$
$\hat{D}_1 \leftarrow \text{ExtractFeatures}(\hat{D}_{\text{physical}})$
$\hat{D}_2 \leftarrow \text{ExtractFeatures}(\hat{D}_{\text{cyber}})$
$M_{\text{ML\_physical}} \leftarrow \text{TrainModel}(\hat{D}_1, \text{ModelType} = \text{ML})$
$M_{\text{EL\_cyber}} \leftarrow \text{TrainModel}(\hat{D}_2, \text{ModelType} = \text{Ensemble})$
$\text{Lm}_{\text{xai}} \leftarrow \text{LIME}(M_{\text{Best\_EL}}, \hat{D}_{\text{test}})$
$\text{Sp}_{\text{xai}} \leftarrow \text{SHAP}(M_{\text{Best\_EL}}, \hat{D}_{\text{test}})$
Return: $M_{\text{ML\_physical}}, M_{\text{EL\_cyber}}, \text{Lm}_{\text{xai}}, \text{Sp}_{\text{xai}}$

B. Dataset Description

The data for training our model is downloaded from IEEE dataports [18]. Our dataset consists of normal flights of UAV and UAV flight subjected to attack. The dataset contains physical as well as network attributes. There are four cyber attacks which are considered in the dataset namely denial-of-service (DoS) attacks, replay attacks, false data injection (FDI) attacks, and evil twin (ET) attacks. We split the main data based on physical and cyber attributes. When we processed the data we found certain common features within physical dataset and cyber dataset. Therefore, only common attributes of respective datasets are considered for attack detection. The physical attributes like timestamps which indicate the exact moment the the data was recorded, the speed in X, Y, and, Z dimensions. The sensor data which includes temperature and barometer which can identify about the drones environment. The flight time that is the total time the drone is flying since takeoff,

and battery of the drone. The cyber attributes dataset records network details like frame number, frame length, transmitter address, source address, receiver address, destination address, fragment number, sequence number, frame control type, and frame control sub type. Based on these characteristics, samples are categorised as benign, DoS, Replay, ET, and FDI, which is essential for examining network traffic and detecting malicious behaviour. By integrating these datasets, our model adopts a comprehensive approach to UAV cybersecurity, ensuring robust detection of attacks across both physical dynamics and network communications. This dataset not only supports the development of more effective IDS for UAVs but also contributes to the broader field of cyber-physical system security, highlighting the unique challenges and opportunities in protecting autonomous aerial vehicles.

C. ML and EL Model Training

After splitting the data as mentioned in above section, we preprocessed the data (steps shown in Algorithm 1). The data is checked for redundancy and extra or duplicate rows are dropped. Then, label encoding is done and the categorical data is converted to numerical data. For physical attributes data, the attack class is converted to benign and malicious class from multi-class. This transformation will help for simple attack detection. Principal component analysis (PCA) is done for removing dependencies among various attributes. After data transformation, entire dataset is divided into training and testing set. 70% of data is given for model training and the rest of the 30% unseen data is provided for model testing for evaluation of the model.

In first instance, lightweight ML model like Naive Bayes, K-Nearest Neighbour, Logistic Regression and Decision Tree are trained on training dataset and detection is done on whether it is benign or malicious. It is performed on the situations where only physical features are seen i.e. on the physical dataset. We used above models based on the dataset characteristics, model performance and evaluation and efficiency.

For further validating the detection based on physical features, we evaluated cyber features. The cyber data is preprocessed to remove redundancy and unimportant features. Label encoding is performed on the cyber features dataset. The multi-class categorical features are converted to numbers, to make it feasible for training complex EL models. The data is split into 70-30, train-test split. XGboost, Random Forest, Decision Tree, KNN, Bagging Classifier models for training and evaluation. The selection of the above models is done for ensemble model because of their unique strength to create a robust, high-performing ensemble. By combining these models, it reduces variance and bias, improved generalization, and better handling of different data characteristics. This diversity in modeling helps in achieving high accuracy and reliability in predictions.

D. LIME Interpretation

XAI (Explainable AI) over traditional machine learning because it offers transparency and interpretability, essential for applications like UAV cybersecurity. By providing insights into decision-making processes, XAI enhances trust, accountability, and compliance, ensuring safer and more reliable AI system adoption. Local Interpretable Model-Agnostic Explanations (LIME) is a popular method to perform interpretability of any kind of machine learning model. By learning a simple linear model around prediction, it explains ML prediction. The model is trained on data points, which are sampled from the training dataset distribution and weighted based on their distance from the reference point being explained by LIME [19]. To comprehend the variables that affect decision-making regarding cyber attack prediction that can be accounted for by model-agnostic techniques. Fig. 2 shows that the classifier predicts 3rd class and the probability of it is 0.87. The prediction for the 2nd class is very less compared to 3rd class with probability of 0.13. The explanation for the above probability is based on following features and there contribution for prediction: source address,less than or equal to 0.0, has contributed 0.20, frame length, less than or equal to 26.0, has contributed -0.19, reciever address, between 0.0 and equal to 1.0, has contributed -0.18, wlan sequence, greater than 1739.50, has contributed 0.04, fc type, less than or equal to 0.0, has contributed 0.03, destination address, between 0.0 and equal to 1.0, has contributed -0.03, wlan.ta, less than or equal to 0.0, has contributed 0.03, frame number, greater than 2120.0, has contributed 0.02, and the fc sub-type, greater than 8.0, has contributes 0.01. The generic equation for the LIME model is:

$$\xi(x) = \underset{a \in A}{\arg\min}\ P(\mu_2, a, \rho_x) + \nu(a) \tag{4}$$

where $a \in A$ (family of interpretable models), $\mu_2 : R^d \rightarrow R$, d is number of features, ρ_x is neighbourhood of x or proximity, $\nu(a)$ is the complexity of the model a, x is the original representation of the instance being explained.

- Weighting the new samples according to their proximity to the interested instance.

$$\pi(x) = \exp\left(-\frac{K(x,z)^2}{\sigma^2}\right) \tag{5}$$

  where K is the distance function, σ is the kernel width which should range between [0,1].
- Training a weighted interpretable model on the dataset with the perturbed instances.

$$P(h, a, \rho_x) = \sum_{z, z^{|} \in Z} \rho_x(z)(\mu_2(z) - a(z^{|}))^2 \tag{6}$$

where Z is the dataset at an instance, z,$z^{|} \in Z$ and z,$z^{|}$ is the subset Z, $\mu_2(z)$ is the complex model prediction i.e EL model, $a(z^{|})$ is the simple model prediction.

E. SHAP Features

The output of machine learning models is interpreted using the SHAP framework. SHAP values fundamental principle is derived from Shapley values and cooperative game theory. We can clearly see how each feature affects predictions with SHAP, unlike with other approaches. By doing this, equity is guaranteed and comprehension is facilitated for everybody.The

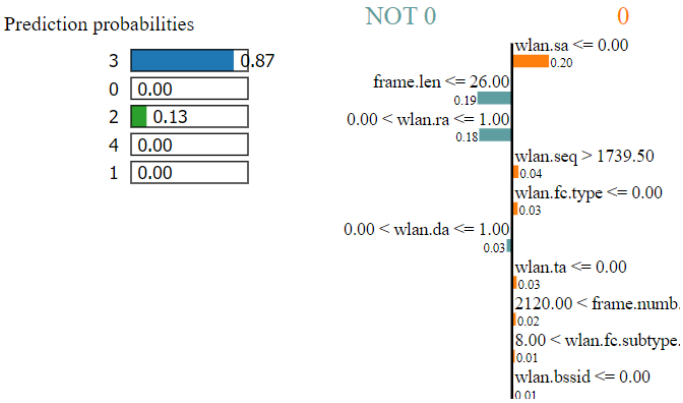

Fig. 2: LIME report of a test sample

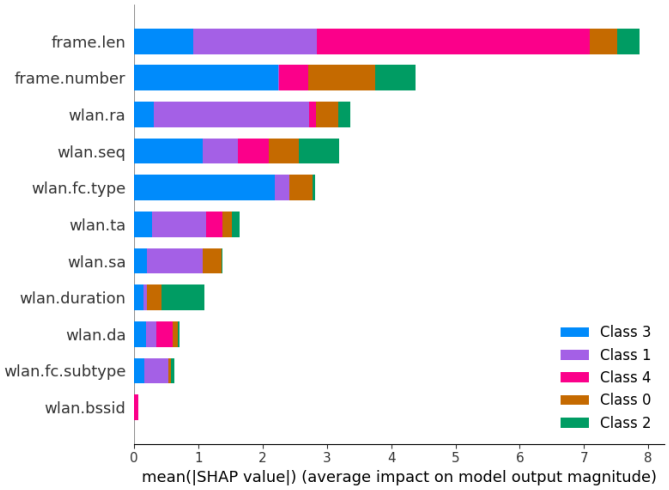

Fig. 3: ShAP based feature importance

benefit of SHAP is that it highlights the significance of every trait in predicting outcomes. The availability of Shapley values aids in our comprehension of intricate models and the ways in which input features impact predictions.In the above graph we have used colour coding technique of SHAP in which features are colour coded and plot will show the spread of the contributions of each feature. Let us understand it through the SHAP equations. Our dataset contains various features like frame number, time stamp, frame length, transmitter address, source address, receiver address , destination address, fragment number, sequence number, frame control type, and frame control subtype . And using these we identify the class like benign, DoS, Replay, evil twin, and FDI which we named it class 0 to class 5.

The SHAP value for a specific feature x_i in our dataset, for a given prediction, is calculated as follows:

$$\phi_{x_i} = \sum_{S \subseteq F \setminus \{x_i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup \{x_i\}) - f(S)] \tag{7}$$

Where F is the set of all features,S is a subset of F that does not contain feature x_i; $f(S)$ is the machine learning model's prediction using only the features in subset S; $f(S \cup \{x_i\})$ is the prediction using the features in S plus feature x_i; $|S|$ is the cardinality (size) of the subset S; and $|F|$ is the total number of features. This equation computes the average marginal contribution of feature x_i across all possible subsets S of the feature set F that do not include x_i. The SHAP values allowed us to quantify the impact of each cyber feature on the prediction of drone attack types. For example, features such as `frame.len` were found to have significant contributions to the model's prediction of class 4, while `frame.number` was more influential in predicting class 3 as seen from Fig. 3. Fig. 4 is beeswarm plot of the xgboost model trained. It show the impact of a particular feature instance on the output. The blue part shows negative impact on output, while red part shows the positive impact on output.

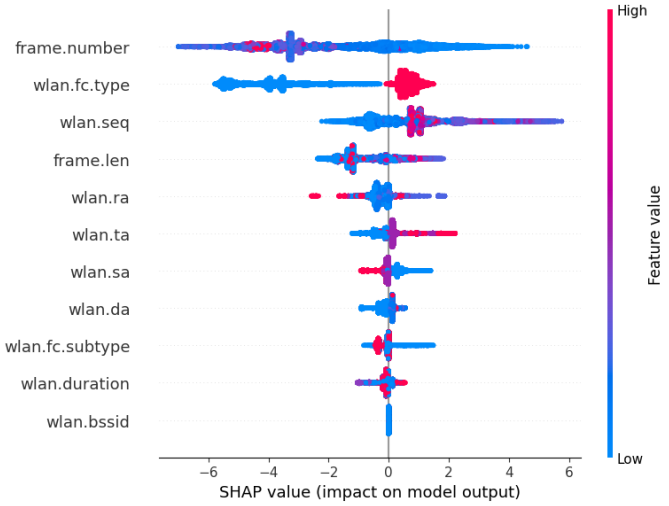

Fig. 4: Shap based impact on model output

III. Results and Discussion

A. Simulation Setup and Tools

The system model has been simulated on Kaggle Notebook using Python v3.10.13 for our approach. For data preprocessing Pandas v2.1.1 is used. Numpy v1.26.4 is used to perform operations on arrays. The data analysis and result visualization are carried by Matplotlib v3.7.5. For the purpose of Simulation, the system used is Apple Mac M2 Air which has an 8GB RAM, 8-Core CPU and 8-Core GPU. Keeping the same configurations throughout the testing increases reliability. We trained our proposed model and other ML models using default parameters.

B. Performance analysis

A few of the most fundamental, lightweight machine learning models are employed in the graph above to determine if the UAV signals are malicious or not. Fig. 5 shows the decision tree and K nearest neighbour models are more resilient to data outliers than all other models combined, we may

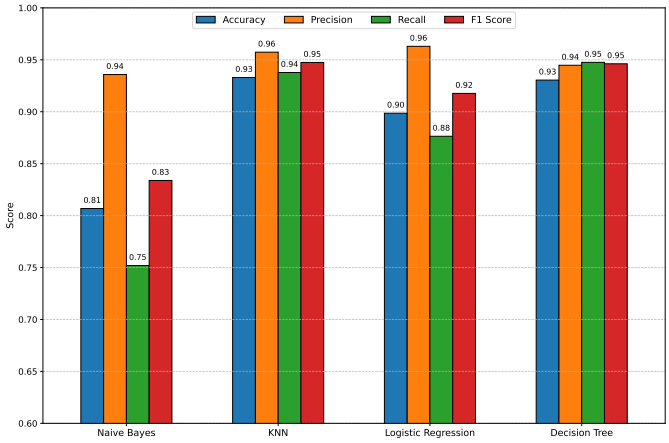

Fig. 5: Primary detection of Attack

observe that they perform better and produce more accurate results.The precision is centered on the affirmative predictions, indicating that a drone detected in a specific class is a member of that class. The accuracy displays the total proportion of drone actions that were correctly classified. The recall is the percentage of drones that the model successfully classified as belonging to a specific class .The harmonic mean of recall and precision is the F1 metric. Because of its limited modeling power and the way it treats each attribute as a single one, Naive Byes produces the least accurate and unstable results. In this case, however, because the speed itself has multiple directions that may be correlated, Naive Byes is unable to capture this effectively.

We assessed the outcome of the implemented machine

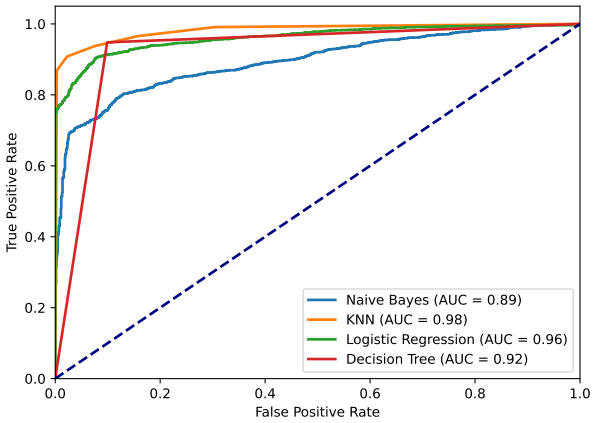

Fig. 6: Model performance for attack detection

learning models for drone attack detection using Receiver Operating Characteristic (ROC) curves. The ROC curve shows the trade-off between True Positive Rate (TPR) and False Positive Rate (FPR) for different categorization levels. TPR is the proportion of drone attacks properly recognized by the algorithm, and FPR reflects the fraction of routine drone flights misclassified as attacks. An ideal ROC curve would be located in the upper left corner of the graph, indicating a high TPR (effectively identifying most attacks) and a low FPR (few false positives). The Area Under the Curve (AUC) measures this performance. A higher AUC suggests more overall efficacy in distinguishing drone attacks from typical drone activity.As clear from the Fig. 6 KNN model achieved the highest AUC of 0.98, followed by logistic regression, decision tree, and naive Bayes .This shows that the KNN model can efficiently detect a major fraction of drone assaults while decreasing false alarms for normal drone operations. The dataset determines

TABLE I: Model Accuracy

Model	Accuracy
XGBoost	92.87
Random Forest	92.2
Decision Tree	91.65
KNN	87.67
Bagging	92.36

the cyber attacks which happen on the drone and this graph represents how accurately we can determine which type of attack will happen and as we can see the XGBoost shows the most promising accuracy with high recall and f1 score and this will identify which type of attack would happen Dos, Replay,Evil Twin or FDI.

As we can see from the Fig. 7, KNN does not yield

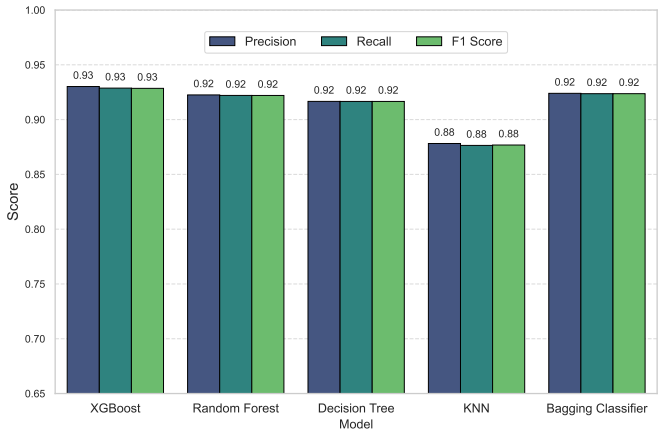

Fig. 7: Attack classification by network features

the best results in this case. This is due to two factors: first, the enormous dataset makes it computationally expensive; second, the model is highly sensitive to noise and irrelevant characteristics.In contrast, XGBoost is scalable and versatile with huge datasets and can manage missing values.Here, the XGBoost model exhibits exceptional performance in all of the factors, including accuracy (the total percentage of drone classes correctly identified), precision (the true positive percentage), recall (which measures how well the model would identify positive cases), and f1 score (the harmonic mean of precision and recall).

The above image Fig. 8 is of ROC-AUC curve for XG Boost

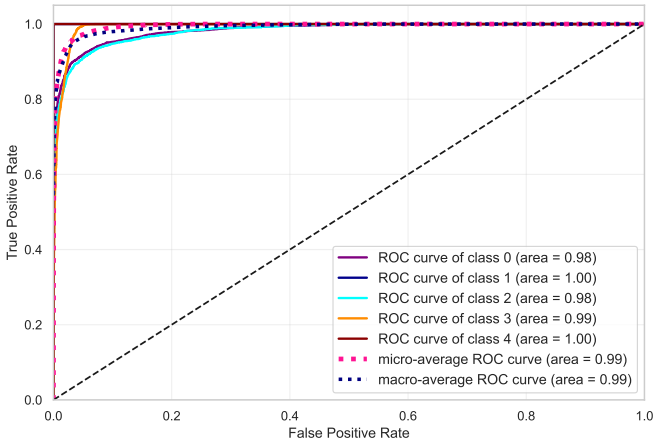

Fig. 8: ROC-AUC curve for XG Boost model

represents the Receiver Operating characteristics (ROC) curve for our class detection problem .The ROC curve estimates how well the classifier performs when determining which thresholds to use for categorizing instances into classes. It plots the True Positive Rate (TPR) against the False Positive Rate (FPR) for different threshold values. The Area Under the Curve (AUC) measures the overall performance of the model. Model's higher AUC value signifies it's ability to distinguish the class types.

IV. Conclusion

In this paper, a ensemble model is proposed to classify the parodic and authentic signals recieved by UAVs. For this purpose, data from IEEE DataPorts were utilized. The feature selection is done from the dataset and two different dataset is formed classifying physical and cyber attack. Prior to training, the dataset was preprocessed with redundancy. After that, label encoding was done to convert the categorical attributes. The attacks are classified based on the features shown by the UAVs. Then, lightweight ML models like KNN, naive Bayes, logistic regression and decision tree were trained on physical attributes dataset to detect the attack. This models provide accuracy of 93%. Further, EL models such as XGBoost, random forest, decision tree, bagging classifier, and, KNN were trained on data with cyber attributes of UAVs. Finally, the proposed model was evaluated for various performance parameters such as accuracy, precision, recall, f1-score and ROC-AUC curve. This models achieved an accuracy of 92.8%. In addition, LIME and SHAP models were used for explaining the predictions of this model. This models provided explanation for local instances as well as the global interpretation. They provide insights about the contribution of the features in the evaluation.

For future works, we will use Federated learning based approach and edge intelligence for enhancing UAV defence capabilities and ensuring efficiency in dynamic and adversarial environments. The framework could be adapted to protect other cyber-physical systems, such as self-driving cars and industrial IoT networks, extending its impact on critical infrastructure. Expanding the dataset to include attacks like GPS spoofing or MitM would further enhance the model's robustness and threat detection capabilities.

References

[1] N. S. Labib, M. R. Brust, G. Danoy, and P. Bouvry, "The rise of drones in internet of things: A survey on the evolution, prospects and challenges of unmanned aerial vehicles," *IEEE Access*, vol. 9, pp. 115466–115487, 2021.

[2] J. Xiao and M. Feroskhan, "Cyber attack detection and isolation for a quadrotor uav with modified sliding innovation sequences," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 7, pp. 7202–7214, 2022.

[3] D. R. Green, J. J. Hagon, C. Gómez, and B. J. Gregory, "Chapter 21 - using low-cost uavs for environmental monitoring, mapping, and modelling: Examples from the coastal zone," in *Coastal Management* (R. Krishnamurthy, M. Jonathan, S. Srinivasalu, and B. Glaeser, eds.), pp. 465–501, Academic Press, 2019.

[4] G. K. Pandey, D. S. Gurjar, H. H. Nguyen, and S. Yadav, "Security threats and mitigation techniques in uav communications: A comprehensive survey," *IEEE Access*, vol. 10, pp. 112858–112897, 2022.

[5] F. Alrefaei, A. Alzahrani, H. Song, and S. Alrefaei, "A survey on the jamming and spoofing attacks on the unmanned aerial vehicle networks," in *2022 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, pp. 1–7, 2022.

[6] M. Soltani, B. Ousat, M. Jafari Siavoshani, and A. H. Jahangir, "An adaptable deep learning-based intrusion detection system to zero-day attacks," *Journal of Information Security and Applications*, vol. 76, p. 103516, 2023.

[7] P. P. Angelov, E. A. Soares, R. Jiang, N. I. Arnold, and P. M. Atkinson, "Explainable artificial intelligence: an analytical review," *WIREs Data Mining and Knowledge Discovery*, vol. 11, no. 5, p. e1424, 2021.

[8] C. S. Kalutharage, X. Liu, and C. Chrysoulas, "Explainable ai and deep autoencoders based security framework for iot network attack certainty (extended abstract)," in *Attacks and Defenses for the Internet-of-Things* (W. Li, S. Furnell, and W. Meng, eds.), (Cham), pp. 41–50, Springer Nature Switzerland, 2022.

[9] L. Zou, H. L. Goh, C. J. Y. Liew, J. L. Quah, G. T. Gu, J. J. Chew, M. P. Kumar, C. G. L. Ang, and A. W. A. Ta, "Ensemble image explainable ai (xai) algorithm for severe community-acquired pneumonia and covid-19 respiratory infections," *IEEE Transactions on Artificial Intelligence*, vol. 4, no. 2, pp. 242–254, 2023.

[10] T. Martins, A. M. de Almeida, E. Cardoso, and L. Nunes, "Explainable artificial intelligence (xai): A systematic literature review on taxonomies and applications in finance," *IEEE Access*, vol. 12, pp. 618–629, 2024.

[11] S. Nazat, L. Li, and M. Abdallah, "Xai-ads: An explainable artificial intelligence framework for enhancing anomaly detection in autonomous driving systems," *IEEE Access*, vol. 12, pp. 48583–48607, 2024.

[12] Y.-W. Hong and D.-Y. Yoo, "Multiple intrusion detection using shapley additive explanations and a heterogeneous ensemble model in an unmanned aerial vehicle's controller area network," *Applied Sciences*, vol. 14, no. 13, p. 5487, 2024.

[13] S. Bayrak, "Unveiling intrusions: explainable svm approaches for addressing encrypted wi-fi traffic in uav networks," *Knowledge and Information Systems*, pp. 1–21, 2024.

[14] S. Wei, Z. Fan, G. Chen, E. Blasch, Y. Chen, and K. Pham, "Tadad: Trust ai-based decentralized anomaly detection for urban air mobility networks at tactical edges," in *2024 Integrated Communications, Navigation and Surveillance Conference (ICNS)*, pp. 1–10, IEEE, 2024.

[15] A. Shafique, A. Mehmood, and M. Elhadef, "Detecting signal spoofing attack in uavs using machine learning models," *IEEE access*, vol. 9, pp. 93803–93815, 2021.

[16] J. Thaker, N. K. Jadav, S. Tanwar, P. Bhattacharya, and H. Shahinzadeh, "Ensemble learning-based intrusion detection system for autonomous vehicle," in *2022 Sixth International Conference on Smart Cities, Internet of Things and Applications (SCIoT)*, pp. 1–6, IEEE, 2022.

[17] "Flaticon." https://www.flaticon.com/, 2024. Accessed on July 22, 2024.

[18] S. Hassler, U. Mughal, and M. Ismail, "Cyber-physical dataset for uavs under normal operations and cyber-attacks," 2023.

[19] G. Visani, E. Bagli, and F. Chesani, "Optilime: Optimized lime explanations for diagnostic computer algorithms," *arXiv preprint arXiv:2006.05714*, 2020.