Knowledge and Information Systems (2024) 66:6675–6695
https://doi.org/10.1007/s10115-024-02181-9

REGULAR PAPER

Unveiling intrusions: explainable SVM approaches for addressing encrypted Wi-Fi traffic in UAV networks

Sengul Bayrak[1]

Received: 21 March 2024 / Revised: 5 June 2024 / Accepted: 3 July 2024 /
Published online: 15 July 2024

Abstract
Unmanned aerial vehicles (UAVs), also known as drones, have become instrumental in various domains, including agriculture, geographic information systems, media, logistics, security, and defense. These UAVs often rely on wireless communication networks for data transmission, making them vulnerable to cyberattacks. To address these challenges, it is necessary to detect potential threats by analyzing the encrypted Wi-Fi traffic data generated by UAVs. This study aimed to develop a linear SVM model that is enhanced with explainable artificial intelligence (XAI) techniques and fine-tuned using Bayesian optimization for intrusion detection systems (IDSs); the model is specifically designed to identify malware threats targeting UAVs. This research utilized encrypted Wi-Fi traffic data derived from three different UAV networks, namely, Parrot Bebop 1, DBPower UDI, and DJI Spark, while considering unidirectional and bidirectional communication flow modes. SVM-based intrusion detection models have been modeled on these datasets, identified their key features using the local interpretable model-agnostic explanations (LIME) technique, and conducted a cost analysis of the proposed modeling approach. The incorporation of the LIME method enabled to highlight the features that are highly indicative of cyberattacks and provided valuable insights into the importance of each feature in the context of intrusion detection. In conclusion, this interpretable IDS model, fine-tuned with Bayesian optimization, demonstrated its superiority over the state-of-the-art methods, proving its efficacy in detecting and mitigating threats to UAVs while offering a cost-effective solution.

Keywords Unmanned aerial vehicles · Encrypted Wi-Fi traffic classification · Support vector machine · Explainable artificial intelligence · Intrusion detection system

1 Introduction

Unmanned aerial vehicles (UAVs), which are commonly called drones, hold significant importance due to their ability to offer various advantages, including speed, ease of access, cost-effectiveness, and enhanced human safety. They have simplified operations across multiple domains, such as firefighting, emergency responses during natural disasters, search and rescue missions, and even estimating irrigation requirements for agricultural purposes. Fur-

✉ Sengul Bayrak
bayraksengul@ieee.org

[1] Department of Software Engineering, Istanbul Sabahattin Zaim University, 34303 Istanbul, Turkey

thermore, UAVs are extensively employed in border security and counterterrorism scenarios [1–3]. UAVs are capable of transmitting data through a multitude of methods, including radio frequency (RF) communication, Wi-Fi, Bluetooth, mobile data networks (3 G, 4 G, and 5 G), satellite communication, wired or fiber-optic communication, and the use of sensors and cameras [4]. However, during the data transmission process, UAVs may be susceptible to various vulnerabilities, including RF interference, jamming, GPS spoofing, unauthorized data access, malware installation, denial-of-service (DOS) attacks, malicious flight control, and even satellite communication attacks. To proactively prevent malicious UAV intrusions, a UAV anomaly detection framework must effectively handle complex signals in noisy environments, often with minimal labeled data samples [5–7]. As a result, intrusion detection systems (IDSs) have become increasingly vital for safeguarding UAVs against vulnerabilities and potential attacks [8]. Modern anomaly-based IDSs leverage machine learning (ML) methods to detect known and unknown threats [9]. Several noteworthy contributions can be highlighted when reviewing the studies in the literature: Li et al. [10] acquired UAV signals as input data, transformed these data into bispectrum representations, and employed a Siamese network-based adversarial learning model to learn vector encodings. Through unsupervised learning, they achieved an impressive 92.85% accuracy in a UAV type detection task and 91.4% accuracy in terms of identifying out-of-sample UAVs. Anwar et al. [11] utilized Mel-frequency cepstral coefficient (MFCC) and linear predictive cepstral coefficient (LPCC) feature extraction techniques to distinguish amateur drone sounds from various background noises, such as those of birds, airplanes, and thunderstorms, in noisy environments. They effectively employed an SVM with multiple kernels; the SVM cubic kernel coupled with MFCCs outperformed the LPCC method, achieving approximately 96.7% accuracy in an amateur drone detection scenario. Nemer et al. [12] conducted feature extraction on RF signals through finite impulse response filtering. They extracted features using ensemble learning (combining K-nearest neighbors and XGBoosting) to accurately determine whether a UAV was present in a given area, achieving approximately 99% accuracy. Xie et al. [13] introduced an intrusion detection model for UAVs based on a belief rule base (BRB), focusing on reducing the input sample size using Wi-Fi data traffic. They proposed an evidential reasoning (ER) algorithm to address the rule combination explosion issue encountered in BRBs. By merging the capabilities of the ER and BRB methodologies, a novel evaluation model termed the EBRB-based model achieved an impressive 99.50% success rate in terms of predicting UAV intrusion detection, even in the presence of numerous attributes. Alipour-Fanid et al. [14] exclusively extracted features from the packet sizes and interarrival times of encrypted Wi-Fi traffic. To reduce the time required online identification, they implemented a reweighted L1-norm regularization method that considered the numbers of frames and samples, as well as the computational costs of different features. To overcome packet interarrival time uncertainties while optimizing the tradeoff between detection accuracy and delays, they employed maximum likelihood estimation (MLE) for packet interarrival time estimation. Their evaluation results demonstrated the ability of the developed approach to detect and identify UAVs within 0.15−0.35 s with high accuracy, ranging from 85.7 to 95.2%. The UAV detection process covered physical detection ranges of 70 m in line-of-sight (LoS) scenarios and 40 m in nonline-of-sight (NLoS) scenarios. Khan et al. [15] addressed the significant cybersecurity challenge associated with the Internet of Medical Things (IoMT) by using innovative bidirectional simple recurrent units (SRUs) to prevent the vanishing gradient problem and enable fast training through skip connections. They also incorporated explainable artificial intelligence (XAI) into their work. Additionally, the authors successfully detected threats, with a 99.38% accuracy rate. Al-Haija and Badawi [16] employed an encrypted Wi-Fi dataset for UAV intrusion detection, utilizing nine metrics based on three UAVs. They

modeled the dataset using convolutional neural networks (CNNs), achieving an accuracy of approximately 99.50%. Utilizing the KDD Cup 99 dataset, Tan et al. [17] achieved a success rate of 92.44% by employing a particle swarm-optimized deep belief network (DBN) for UAV intrusion detection in networks. Alheeti et al. [18] employed ICMetric technology for UAV detection using signals acquired from an encrypted Wi-Fi network. ICMetric numbers represented additional features integrated into the dataset used for drone detection. This study employed a deep neural network (DNN) for classification and achieved an impressive performance rate of 99.99%. Medaiyese et al. [19] utilized Wi-Fi and Bluetooth data to extract distinctive signatures from the transient and steady states of signals. By employing RF control, signals from UAVs, the discrete wavelet transform, the continuous wavelet transform, and the wavelet scattering transform, they conducted feature extraction under varying signal-to-noise ratio (SNR) levels. They constructed different models using these features for signal feature extraction and trained on the obtained dataset using SqueezeNet, achieving an accuracy of 98.9%. Ezuma et al. [20] detected RF signals from 15 UAVs using a naive Bayes decision mechanism based on Markov models. Additionally, signals from Wi-Fi and Bluetooth emitters were detected based on the bandwidth and modulation characteristics of the identified RF signals. After recognizing the input signals as UAV control signals, they achieved a classification accuracy of 98.13% using the k-nearest neighbors (knn) method. In recent studies, SVMs have been used to detect injection attacks on automatic dependent surveillance-broadcast (ADS-B) devices in UAVs [21] and to protect against global positioning system (GPS) signal spoofing attacks [22, 23]. In this study, however, the SVM method was used for two purposes. First, attacks conducted over an encrypted Wi-Fi network were modeled, and the obtained model was validated with a black-box model called the LIME method. When analyzing prior research, it became evident that intrusion detection studies involving Wi-Fi-based data have predominantly embraced ML methodologies. However, it is noteworthy that these studies have not rigorously explored model interpretability similarly to LIME. In the contemporary era of burgeoning computational power, ML approaches have garnered extensive use in UAV intrusion detection scenarios. Even so, ML models, which are often shrouded in opacity, need help regarding their comprehensibility. Very few studies in this domain have adequately addressed the imperative of comprehending their constructed models. The deficiencies of this oversight may encompass pitfalls such as data overfitting, the absence of confidence intervals surrounding point estimates, and arbitrary variable selection results. Khan et al. [24] proposed an autoencoder-based detection framework utilizing convolutional and recurrent networks to identify and explain cyber threats in Industrial Internet of Things (IIoT) networks, using an SVM for error construction. By applying a two-step sliding window approach, the framework effectively extracted the temporal and spatial features of malicious events for classification and explanation purposes. The empirical results showed that this framework outperformed the state-of-the-art methods, demonstrating its robustness and suitability for real-world IIoT-based networks. This study utilized encrypted Wi-Fi traffic data from three UAV networks: Parrot Bebop 1, DBPower UDI, and DJI Spark. This made it possible to evaluate the performance of various UAV models and determine their general validity, unlike in previous studies. The LIME technique was utilized to improve the performance and interpretability of SVM models. This enabled the identification of the features that are critical for intrusion detection and made the decision-making processes of the models transparent. Such an approach is rare in the literature and is an innovative aspect of this work. This study trained SVM models considering both one-way and two-way communication flow modes. The communication mode is an essential factor that was often overlooked in previous works, and it helped us better understand how UAVs behave in real-world scenarios. This study demonstrates the performance superiority and effectiveness of the IDS model devel-

a) Parrot Bebop I b) DBPower UDI c) DJI Spark

Fig. 1 The UAVs utilized in this study [9]

oped with Bayesian optimization-based fine-tuning through a comprehensive comparison. This thorough analysis makes the scientific contributions and practical applications of the study more evident.

The principal contributions of this research can be succinctly summarized as follows.

- A linear SVM model was developed and engineered for UAV attack risk detection.
- In this investigation, the local LIME method, which is an XAI technique, was used to elucidate the results obtained from the proposed model. This approach allowed us to interpret the significance of each feature, shedding light on the performance of the model.
- This study represents a pioneering effort of introducing an interpretable SVM model for risk detection. The model is predicated on encrypted Wi-Fi traffic records sourced from the Parrot Bebop 1, DBPower UDI, and DJI Spark UAVs, encompassing bidirectional and unidirectional communication flow modes.
- Furthermore, this research delved into unraveling the intricate nonlinear relationships among features, marking a departure from the conventional variable selection approach employed in classic statistical methods.

The structure of this paper is as follows. Section 2 elucidates the employed materials and methods, providing a detailed architectural overview. Section 3 presents the experimental results obtained from the proposed framework. Finally, Section 4 offers concluding remarks and provides a comprehensive discussion of the paper.

2 Materials and methods

2.1 Communication flow modes

UAVs can establish direct connections with devices such as smartphones or tablets by creating Wi-Fi networks. The data utilized in this study were collected through direct Wi-Fi connections between UAVs and their controllers. Specifically, three types of UAVs, namely, Parrot Bebop, DBPower UDI, and DJI Spark, as illustrated in Fig. 1, were employed to capture Wi-Fi data under both the bidirectional and unidirectional communication flow modes [25].

(a) The bidirectional communication flow mode pertains to communication modes that facilitate the bidirectional exchange of data and commands between a drone and the control station. Information is transmitted from the ground control station (GCS) to the drone and from the drone to the control station. This mode encompasses three categories of data sources: the uplink flow, downlink flow, and total traffic flow. Nine statistical parameters were calculated for each data source, culminating in 54 features. (b) The unidirectional communication flow mode delineates a scenario wherein data and commands flow solely in one direction,

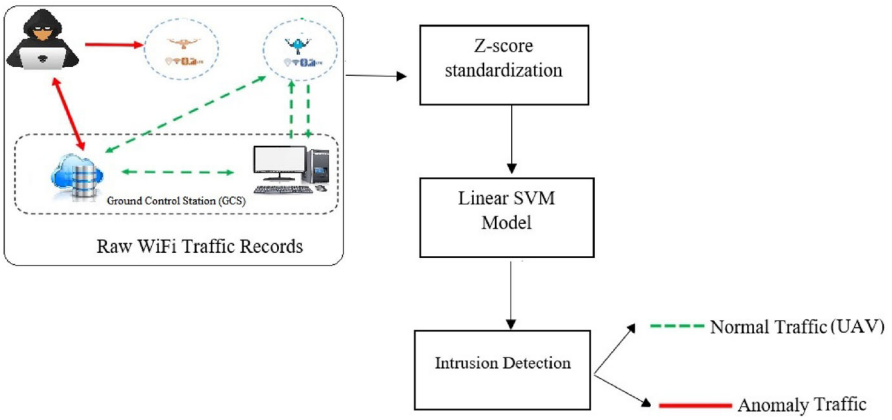

Fig. 2 Proposed model architecture

predominantly from the ground control station (GCS) to a drone. In such an arrangement, the drone is adept at receiving instructions and commands from the operator but may not furnish real-time data or feedback for the operator. This mode encompasses solely the total traffic flow data source, from which 9 statistical parameters were computed.

2.2 Methodology for the ML

In this investigation, a binary classification approach for UAV intrusion detection involving the operation of normal and anomalous UAV traffic was modeled utilizing a linear SVM. The modeling steps are depicted in Fig. 2.

2.2.1 Z score standardization

The packet sizes and interarrival times between the data packets obtained in bidirectional and unidirectional communication flow modes served as raw data sources. The dataset derived from these sources underwent processing utilizing the z score standardization method, as outlined in Eq. 1 [26].

$$X^* = \frac{X - \bar{X}}{\sigma_X} \tag{1}$$

2.2.2 SVM model

A dataset D, consisting of n elements, is structured as (X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n), where $y\epsilon(+1, -1)$. Selecting the element that maximizes the margin between two hyperplanes is necessary for data classification. The midpoint of hyperplanes H_1 and H_2 is represented as H_0. H_0 is the optimal hyperplane that linearly separates the two classes. H_0 can be expressed as in Eq. 2 [27, 28].

$$H_0 = \sum_{i=1}^{n} (w_i x_i + b) = 0 \tag{2}$$

In accordance with Eq. 2, where n represents the number of attributes, $W = w_1, w_2, ..., w_n$ denotes the weight vector, and b is a constant. In a case with having two attributes, the hyperplanes H_1 and H_2 for the values $X = x_1, x_2, ..., x_n$ are formulated as demonstrated in Eqs. 3 and 4, respectively.

$$H_1 : W^T X + b = 1 \tag{3}$$

$$H_2 : W^T X + b = -1 \tag{4}$$

The points on the upper side of each hyperplane are calculated using Eq. 5.

$$W^T X + b > 0, y_1 = +1 \tag{5}$$

The points at the bottom of each hyperplane are calculated according to Eq. 6.

$$W^T X + b < 0, y_2 = -1 \tag{6}$$

The observations on hyperplanes H_1 and H_2 are designated support vectors. z denotes a point on a hyperplane, and the distance between a support vector and H_0 is computed as demonstrated in Eq. 7.

$$d = \frac{|W X_z^{'} \mp b|}{\|w\|} \tag{7}$$

2.3 Fine-tuning with Bayesian optimization

This study used the Bayesian optimization method to estimate the optimal hyperparameter combinations for the model developed by the SVM method, using the priority function and an observation function [29, 30].

Bayesian optimization mathematically calculates where the objective function $\mathcal{F}(x)$ is maximized:

- Bayesian optimization starts with a probabilistic model of the objective function. This model is usually represented by Gaussian processes (GPs). Gaussian processes provide a probability distribution for $\mathcal{F}(x)$ with respect to any combination of hyperparameters x.
- Initially, the objective function is evaluated on several hyperparameter combinations $x_1, x_2, \ldots, x_n$ and observations $y_1 = \mathcal{F}(x_1), y_2 = \mathcal{F}(x_2), \ldots, y_n = \mathcal{F}(x_n)$ are obtained for these points.
- As new observations arrive, the Gaussian process updates the posterior probability distribution. This is computed using Bayes' theorem, as in Eq. 8.

$$P(\mathcal{F}|X, Y) = \frac{P(Y|X, \mathcal{F})P(\mathcal{F})}{P(Y|X)} \tag{8}$$

  Here $X = [x_1, x_2, \ldots, x_n]$ and $Y = [y_1, y_2, \ldots, y_n]$ are observation points.
- Bayesian optimization uses an acquisition function to select the points to be observed next. A commonly used acquisition function is given in Eq. 9.

$$EI(x) = \mathbb{E}[\max(0, \mathcal{F}(x) - \mathcal{F}^*)] \tag{9}$$

  Here, $\mathcal{F}^*$ is the best available observation.

$$\mathrm{UCB}(x) = \mu(x) + \kappa\sigma(x) \tag{10}$$

Here $\mu(x)$ is the estimated mean, and $\sigma(x)$ is the estimated uncertainty.
- A new hyperparameter combination x_{new} is chosen at the location where the acquisition function is maximal, and the objective function is evaluated at this point. This new observation is used to update the posterior distribution.
- This process is repeated until the stopping criterion of the objective function or a certain number of iterations is reached.

2.3.1 Evaluational metrics

In binary classification models, accuracy, specificity, sensitivity, and precision are employed to evaluate the achieved modeling performance [31]. These metrics are calculated using Eqs. 11 - 14, respectively. Minimizing the false-negative (FN) rate is crucial for optimizing the accuracy of a classification model. A ROC analysis represents the comprehensive success curve of the tested classification model, which is generated from different precision- false-positive rate pairs. The area under the curve (AUC) of the ROC curve reflects the accuracy of the classification model [32]. The area under the ROC curve, which is based on the true-positive (TP) rate and false-positive (FP) rate, is calculated according to Eq. 15.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{11}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{12}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{13}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{14}$$

$$\text{ROC} = \int_0^1 \left(\frac{\text{TP}}{\text{TP} + \text{FN}}\right) d\left(1 - \frac{\text{TN}}{\text{TN} + \text{FP}}\right) \tag{15}$$

2.4 Explanation of the proposed model with LIME

The original Wi-Fi data in D and (x) are segmented into samples via an SVM. The annotation model is $g \epsilon G$, where G denotes the set of interpretable models that can be visually presented to a user. $\pi_x(z)$ is employed to represent the proximity between instances z and x and to establish locality around x. An objective function $\xi(x)$ is established, and the L-function applied to $\xi(x)$ elucidates how the interpretability g of the local definition approximates f through $\pi_x(z)$. The L-function is minimized from the perspective of human comprehension to attain the optimal solution for the objective function when $\Omega(g)$, the complexity of the explanatory model, is adequately low. The explanation function $\xi(x)$ acquired through the LIME method is expressed as in Eq. 16 [33].

$$\xi(x) = \text{argmin}_{g \epsilon G} L(f, g, \pi_x(z)) + \Omega(g)) \tag{16}$$

Equation 17 elaborates on the extent of the similarity, $\pi_x(z)$.

$$\pi_x(z) = \exp\left(-\frac{D(x, z)^2}{\sigma^2}\right) \tag{17}$$

The formula for the similarity degree $\pi_x(z)$ in Eq. 17 and the objective function are presented in Eq. 18. Here, $g(z)^{'}$ signifies the estimated value in a d-dimensional space, and

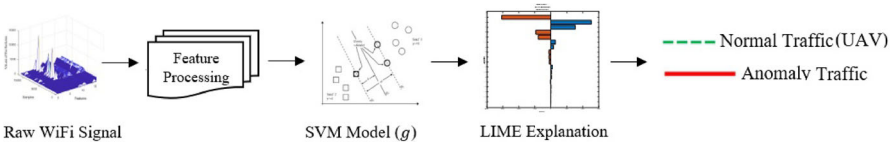

Fig. 3 Architecture of the proposed model

Table 1 Features extracted from Source 1 under the bidirectional communication flow mode

Source 1 Uplink Flow	Downlink Flow	Total Traffic Flow
ID: Name	*ID: Name*	*ID: Name*
X1: Mean	X10: Mean	X19: Mean
X2: Median	X11: Median	X20: Median
X3:MAD	X12:MAD	X21:MAD
X4: STD	X13: STD	X22: STD
X5:Skewness	X14:Skewness	X23:Skewness
X6:Kurtosis	X15:Kurtosis	X24:Kurtosis
X7:Max	X16:Max	X25:Max
X8:Min	X17:Min	X26:Min
X9:Mean Square (ms)	X18:Mean Square (ms)	X27:Mean Square (ms)

$f(z)$ represents the estimated value in the d-dimensional space. The perturbed $z^{'}$ values are generated by toggling features on and off.

$$\xi(x) = \sum_{z,z^{'} \epsilon Z} \pi_x(z) \left(f(z) - g(z^{'}) \right)^2 \tag{18}$$

The perturbed data $z^{'}$ are mapped back to the original input, resulting in z being the new input for f. The estimates of z correspond to the labels of the perturbed samples. The combination of $z^{'}$ and $f(z)$ constitutes the dataset. The SVM model, as defined, can be enhanced by incorporating a locally trained interpretable model [34, 35].

3 Experimental results

The procedural steps of the model presented in this study are illustrated in Fig. 3.

According to Fig. 3, the processing steps of the proposed architecture in this study were as follows.

Step 1 Encrypted raw Wi-Fi recordings were collected from Parrot Bebop 1, DBPower UDI, and DJI Spark UAV sources in bidirectional and unidirectional communication flow modes. Tables 1 and 2 show that nine features × 2 sources × 3 directional flows = 54 features were obtained from the UAVs with the bidirectional communication flow. Similarly, as indicated in Tables 3, 9 features × 2 sources = 18 features were obtained from the UAVs with the unidirectional communication flow mode.

The formulas for the mean ($\bar{x}$), median, mean absolute deviation (MAD), standard deviation (STD; (σ)), skewness (γ), kurtosis (β), maximum, minimum, and mean square (MS) are provided in Eqs. 19–26, respectively.

Table 2 Features extracted from Source 2 under the bidirectional communication flow mode

Source 2 Uplink Flow	Downlink Flow	Total Traffic Flow
ID: Name	*ID: Name*	*ID: Name*
X28: Mean	X37: Mean	X46: Mean
X29: Median	X38: Median	X47: Median
X30:MAD	X39:MAD	X48:MAD
X31: STD	X40: STD	X49: STD
X32:Skewness	X41:Skewness	X50:Skewness
X33:Kurtosis	X42:Kurtosis	X51:Kurtosis
X34:Max	X43:Max	X52:Max
X35:Min	X44:Min	X53:Min
X36:Mean Square (ms)	X45:Mean Square (ms)	X54:Mean Square (ms)

Table 3 Features extracted under the unidirectional communication flow mode

Source 1 Total Traffic Flow *ID: Name*	Source 2 Total Traffic Flow *ID: Name*
X1: Mean	X10: Mean
X2: Median	X11: Median
X3:MAD	X12:MAD
X4: STD	X13: STD
X5:Skewness	X14:Skewness
X6:Kurtosis	X15:Kurtosis
X7:Max	X16:Max
X8:Min	X17:Min
X9:Mean Square (ms)	X18:Mean Square (ms)

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x(i) \tag{19}$$

$$\text{MAD} = \text{median}(|x(i) - \text{median}(x)|) \tag{20}$$

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x(i) - \text{mean}(x)^2)} \tag{21}$$

$$\gamma = \frac{1}{N}\sum_{i=1}^{N}(x(i) - \text{mean}(x)/\sigma)^3 \tag{22}$$

$$\beta = \frac{1}{N}\sum_{i=1}^{N}(x(i) - \text{mean}(x)/\sigma)^4 \tag{23}$$

$$H = (\max(x(i)) \mid i = 1 \cdots N) \tag{24}$$

$$L = (\min(x(i)) \mid i = 1 \cdots N) \tag{25}$$

$$MS = \frac{1}{N}\sum_{i=1}^{N}(x(i))^2 \tag{26}$$

Table 4 Sizes of the datasets acquired from different UAV sources

	Parrot Bebop	DBPower UDI	DJI Spark
Bidirectional	19,380x54 (UAV:10.717 Anomaly:8663)	17,256x54 (UAV:8664 Anomaly:8592)	5500x54 (UAV:2735 Anomaly:2765)
Unidirectional	11,663x18 (UAV:4982 Anomaly:6681)	14,864x18 (UAV:7877 Anomaly:6987)	73x18 (UAV:33 Anomaly:40)

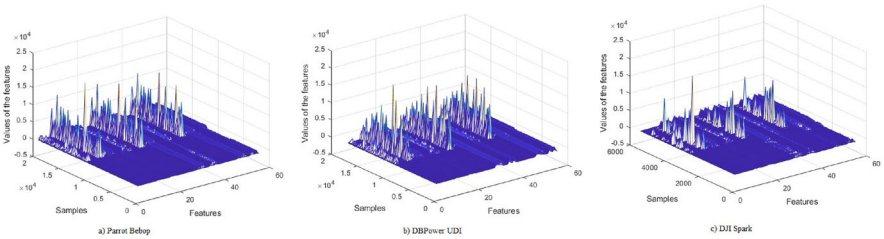

Fig. 4 Mesh visualizations produced under the bidirectional communication flow mode

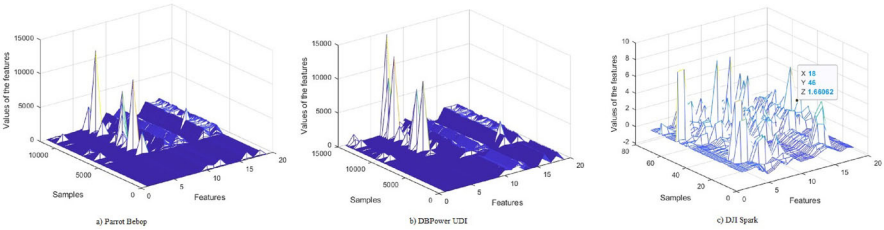

Fig. 5 Mesh visualizations produced under the unidirectional communication flow mode

The interdependencies of the nine features are delineated in Tables 1 through 3. Each node symbolizes a distinct feature, while each hypersegment denotes a feature component shared among various features.

Step 2 The gathered data were subjected to feature processing by implementing the z score normalization technique.

Step 3 The datasets were evaluated using a linear model formulated through the SVM methodology.

Step 4 The integrity of the performance of the model constructed via a linear SVM was elucidated using a black-box explanatory model (LIME).

3.1 Dataset

The processed datasets obtained from the Parrot Bebop 1, DBPower UDI, and DJI Spark UAVs and their sizes are presented in Table 4.

The three-dimensional visual representations of the datasets procured from various UAV sources are shown in Figs. 4 and 5.

Table 5 Training and validation results obtained under the bidirectional communication flow mode

UAV Types	Validation Accuracy	Validation Total Cost	Prediction Speed (obs/s)	Training Time (s)
Parrot Bebop	100	0	66,000	5.73
DBPower UDI	100	0	63,000	5.41
DJI Spark	100	1	33,000	2.98

Table 6 Test results obtained under the bidirectional communication flow mode

UAV Types	Accuracy (%)	Specificity (%)	Sensitivity (%)	Precision (%)
Parrot Bebop	100	100	100	100
DBPower UDI	100	100	100	100
DJI Spark	99.94	99.88	100	99.88

3.2 Linear SVM modeling

When modeling on the datasets acquired from disparate UAV sources via the linear SVM, 70% of the data were earmarked for training purposes, with the remaining 30% being designed for testing. The training data were divided into tenfold using the cross-validation method against overfitting. For classification, the UAV category was assigned a label of '1', while the anomaly category was denoted with a label of '0'.

3.2.1 SVM modeling results obtained under the bidirectional communication flow mode

The encrypted Wi-Fi records produced by the Parrot Bebop, DBPower UDI, and DJI Spark UAVs under the bidirectional communication flow mode were modeled using a linear SVM. The training and validation results obtained with the linear SVM are summarized in Table 5.

As delineated in Table 5, the validation accuracies achieved for the trio of UAVs, specifically, the DJI Spark, DBPower UDI, and Parrot Bebop UAVs, were recorded at values of 100%. During the comparative analysis of the parameters, i.e., the prediction velocity and training duration, gradation trends were discernible, extending from the minimal values to the maximal values in order of the above UAVs. Furthermore, Table 6 provides a detailed account of the testing efficacy achieved within the scope of the bidirectional communication flow modality.

The confusion matrix produced for the training and validation datasets is presented in Fig. 6.

The confusion matrices produced for the test dataset are presented in Fig. 7.

In the intrusion detection task, when operating within the bidirectional communication flow mode, the encrypted Wi-Fi signals from the Parrot Bebop and DBPower UDI UAVs exhibited paramount efficacy. Despite attaining a test performance level of 99.94%, DJI Spark manifested a marginally diminished model efficacy value relative to the other UAVs, as mentioned above; this outcome is attributable to the use of a less voluminous dataset. Empirical trials executed throughout the testing phase culminated in a 100% success metric with respect to the ROC analyses performed for all three classes of UAVs.

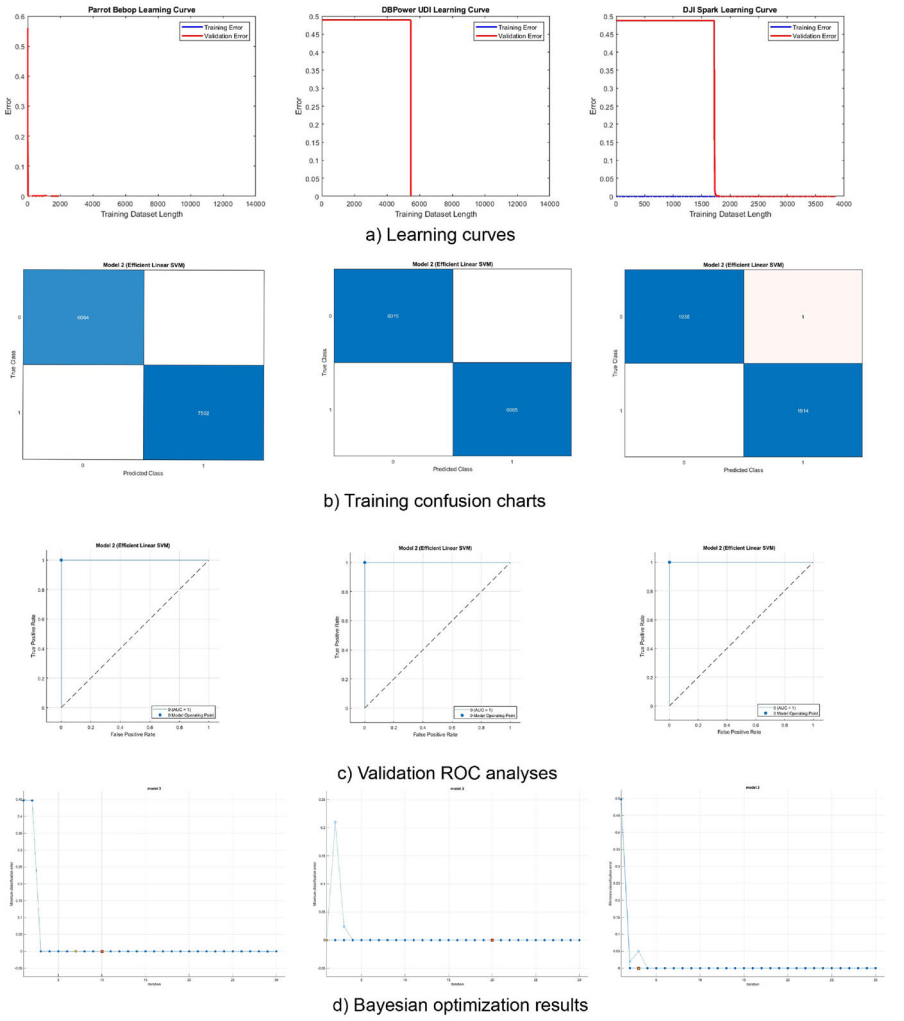

Fig. 6 Train and validation confusion chart produced under the bidirectional communication flow mode

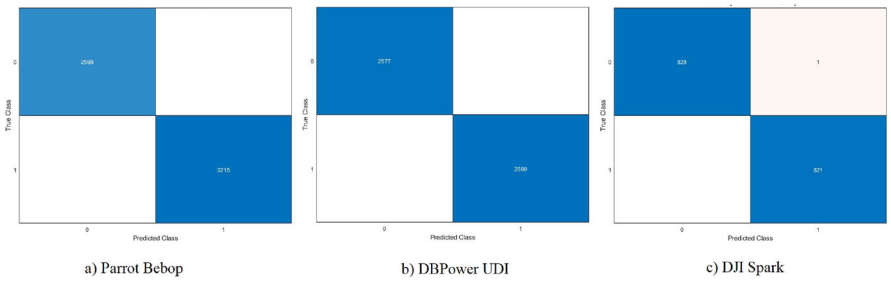

Fig. 7 Test confusion chart produced under the bidirectional communication flow mode

Table 7 Training and validation results obtained under the unidirectional communication flow mode

UAV Types	Validation Accuracy	Validation Total Cost	Prediction Speed (obs/s)	Training Time (s)
Parrot Bebop	100	1	16,000	48.66
DBPower UDI	100	0	46,000	9.46
DJI Spark	98.1	1	750	12.48

Table 8 Test data performance evaluation metrics produced under the unidirectional communication flow mode

UAV Types	Accuracy (%)	Specificity (%)	Sensitivity (%)	Precision (%)
Parrot Bebop	100	100	100	100
DBPower UDI	100	100	100	100
DJI Spark	100	100	100	100

3.2.2 SVM modeling results obtained under the unidirectional communication flow mode

The encrypted Wi-Fi transmission logs of the Parrot Bebop, DBPower UDI, and DJI Spark UAVs operating under the unidirectional communication flow mode were subjected to modeling through a linear SVM methodology. The outcomes associated with the training and validation exercises utilizing the linear SVM are concisely encapsulated in Table 7.

Referencing Table 7, the validation accuracies attained for the Parrot Bebop and DBPower UDI UAVs were 100%. In comparison, the validation accuracy achieved for the DJI Spark UAV was 98.1%. In the context of the total costs associated with these three UAVs, Parrot Bebop and DJI Spark yielded values of 1, whereas DBPower UDI returned a value of 0. In descending order, the prediction speeds of the UAVs were ranked in the following order: DJI Spark > Parrot Bebop > DBPower UDI. Concerning the training time, the rankings of the durations from shortest to longest were as follows: DBPower UDI < DJI Spark < Parrot Bebop. The DJI Spark UAV possessed fewer data samples within the dataset than did the other two UAVs. Owing to the reduced sample size, its performance was lower than that of the other two UAVs. When evaluating the prediction speed and training time parameters, the following order was discerned from lowest to highest: DJI Spark, DBPower UDI, and Parrot Bebop. Table 8 presents the test performance achieved under the unidirectional communication flow mode.

The confusion matrices obtained for the training and validation datasets are presented in Fig. 8.

The confusion matrices obtained for the test dataset is presented in Fig. 9.

In the intrusion detection task conducted under the unidirectional communication flow mode, the encrypted Wi-Fi signals procured from all three UAVs attained an accuracy of 100%. Under empirical trials executed throughout the testing phase, the success rates derived from the ROC analyses performed for all three categories of UAVs were uniformly recorded as 100%.

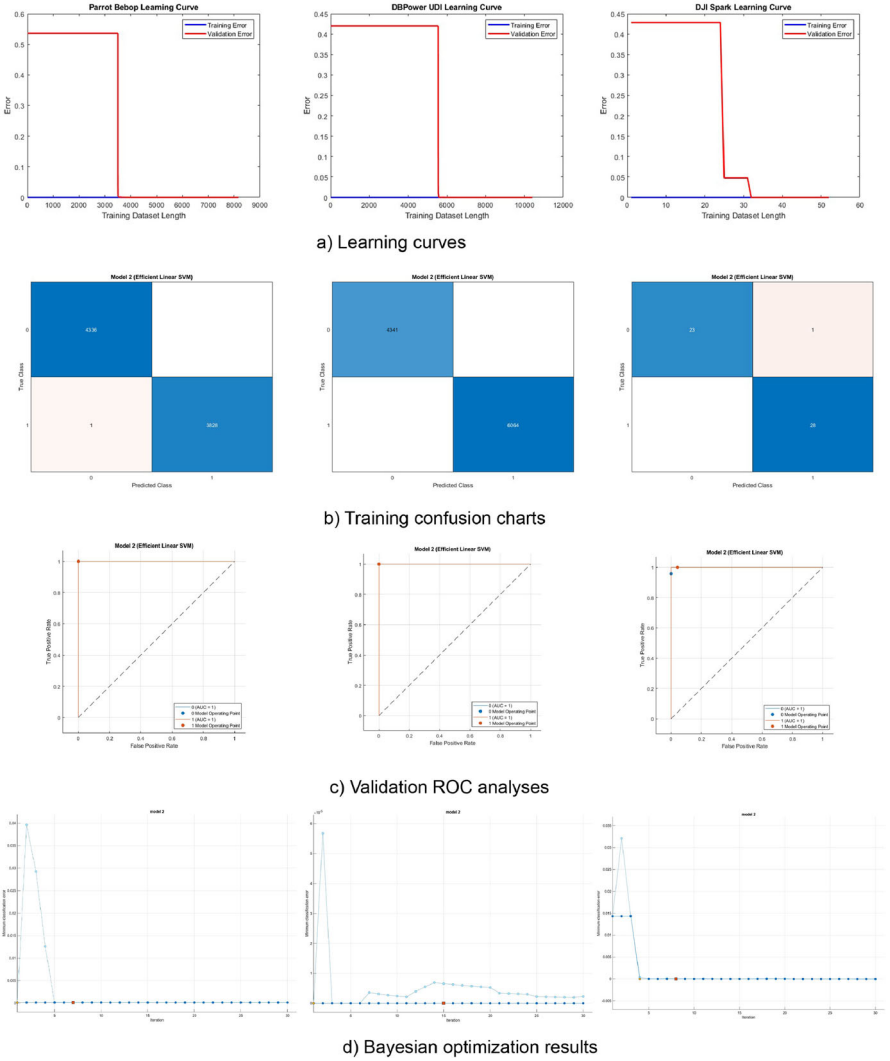

Fig. 8 Taining and validation confusion matrices produced under the unidirectional communication flow mode

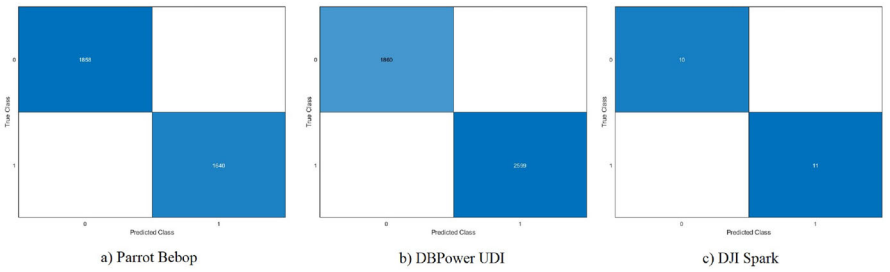

Fig. 9 Test confusion matrices produced under the unidirectional communication flow mode

3.3 Explanation results obtained with LIME

Interpretation results were derived from the dataset employed to train the intrusion detection system for identifying the discordant instances of each UAV. The LIME method was utilized to extract descriptions of normal UAVs and anomalous data records from the dataset. The process of extracting rules from the training dataset using the LIME explainer is delineated in Algorithm 1.

Initially, the training data ($X\_$ train) were annotated with the target ($Y_$train) by employing the SVM model, and a descriptor of this dataset encompassing all its features was acquired. The prediction function of the trained model was ascertained. By leveraging the LIME descriptor and the prediction function of the model, a description of the normal (UAV)/anomalous records in the dataset was obtained. Subsequently, the software generated samples of a synthetic dataset and fit a simplistic model for the query point with the critical predictors. By specifying the number of features N (where N was 54 for the bidirectional communication flow mode and 18 for the unidirectional communication flow mode), the most salient features to be utilized were identified. If more than one feature did not pertain to set z, this input traffic was categorized as anomalous. If all features pertained to the dataset, the input traffic was classified as normal and necessitated no further action (as per Algorithm 1).

Algorithm 1 Extract significant parameters and detect UAV attacks

Data: X_{Train}
Result: Significant UAV parameters
while X_{Train} **do**
  $Explainer \leftarrow$ LIME_Explainer(X_{Train}, Y_{Train}, All Parameters)
  Number of significant parameters: N (54 for bidirectional, 18 for unidirectional)
  if $Decision \leftarrow$ UAV **then**
    return UAV
  else
    return Anomaly
  end if
end while

Once the number of features was selected, the explication of the normal (UAV)/anomalous data records in the dataset was finalized using the prediction function of the model. The annotation process yielded a set containing the N most significant features among all selected normal (UAV)/anomalous data records. The algorithm then identified the z features that constituted the normal (UAV)/anomalous data descriptions. In the detection phase, the intrusion detection system analyzed the network traffic under surveillance. Explanation results were obtained by using LIME on all data, including both the attack and normal test data, to evaluate the detection performance of the model. Subsequently, an analysis was conducted to ascertain whether the extracted features belonged to set z. If all the features of the data pertained to the set, they were deemed normal. If a feature did not pertain to set z, the data were classified as an attack.

In Fig. 10a, in the anomaly class detection results obtained for the Parrot Bebop UAV, the coefficient values for parameters X41, X23, X5, and X50 are the top four parameters that had negative impacts. At the same time, the significant anomaly detection parameter values for Parrot Bebop were X24, X42, and X6. In Fig. 10b, in the UAV class detection results obtained for the DBPower UDI UAV, the coefficient values for parameters X14, X46, and

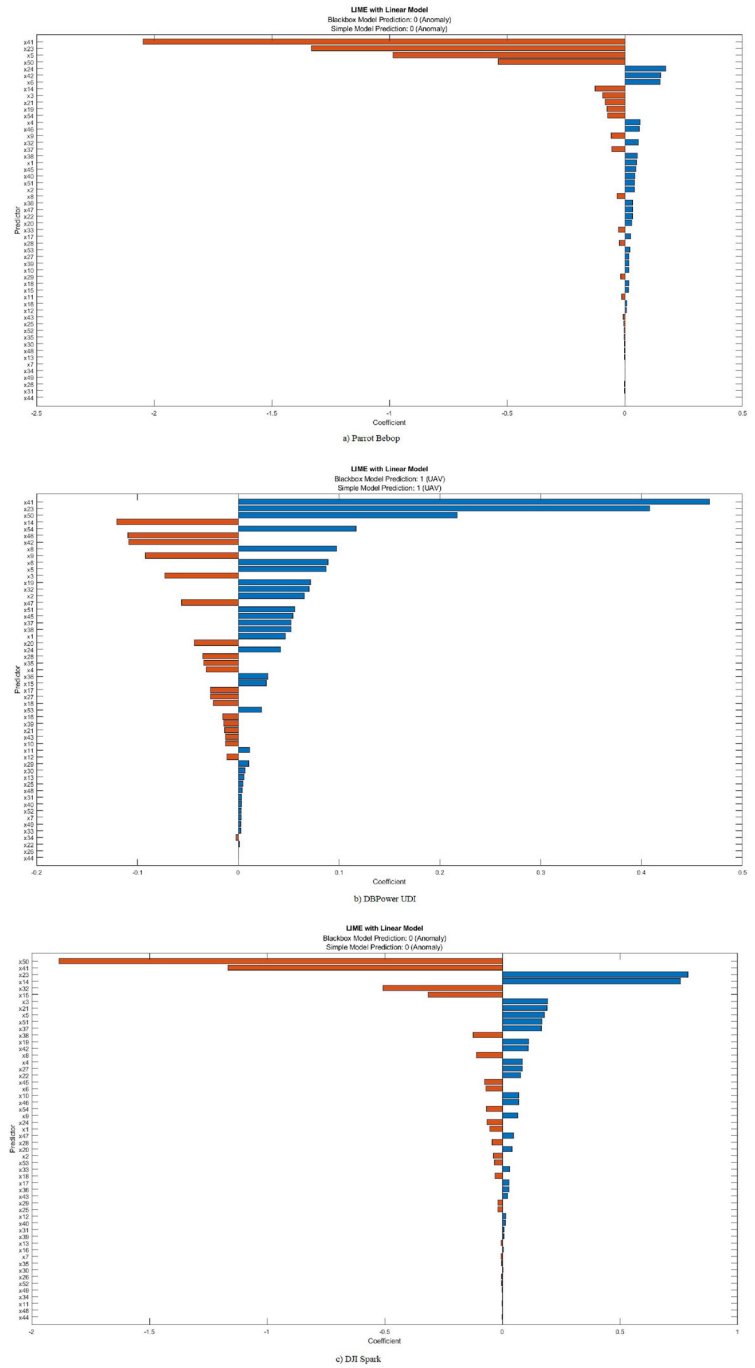

Fig. 10 The LIME model for the bidirectional communication flow mode

X42 are the top parameters with negative impacts. At the same time, the significant UAV detection parameter values for DBPower UDI were X41, X23, and X50. In Fig. 10c, the anomaly detection results obtained for DJI Spark shows that parameters X50 and X41 had negative impacts. At the same time, the significant anomaly detection parameter values for DJI Spark were X23 and X14.

In Fig. 11a, the anomaly class detection results obtained for the Parrot Bebop UAV show that the coefficient values for parameters X3 and X4 ranked among the top four parameters exerting negative influences. Concurrently, X1, X17, and X5 were identified as significant anomaly detection parameters for the Parrot Bebop UAV. In Fig. 11b, which shows the anomaly class detection results obtained for the DBPower UDI UAV, the coefficient values for parameters X4 and X5 emerged as the paramount parameters manifesting negative impacts. Simultaneously, X17, X2, and X3 were ascertained to be significant anomaly detection parameters for the DBPower UDI UAV. In Fig. 11c, which concerns the anomaly detection results obtained for the DJI Spark UAV, the X1, X9, X2, X4, and X3 parameters were shown to have negative impacts. Concurrently, X7 was recognized as a significant anomaly detection parameter for the DJI Spark UAV.

3.4 Cost analysis of the proposed model

The cost analysis steps of the model developed in this study, including evaluations of its temporal complexity, spatial complexity, computational resource costs, and practical implementation costs, were as follows. The temporal complexity of the linear SVM with LIME for explainability was calculated as shown below.

- Linear SVM training is $O(nd)$, where n is the number of samples and d is the number of features.
- LIME explainability is $O(n \cdot d^2)$.

The spatial complexity of the linear SVM with LIME for explainability was calculated as follows.

- Linear SVM is $O(nd)$ for storing the input data and model parameters.
- LIME is $O(nd)$ for storing perturbations and the local models.
- The computational resource cost of the linear SVM with LIME for explainability was calculated as follows.
- Training is 2 CPU hours.
- Explainability is 3 CPU hours.

4 Conclusion

A linear SVM model was developed to detect attacks on UAVs through encrypted Wi-Fi data. For this purpose, data from Parrot Bebop 1, DBPower UDI, and DJI Spark drones were utilized. The data were processed separately for each UAV under bidirectional and unidirectional communication flow modes. The efficacy of the developed linear model in terms of predicting anomalies in UAVs was elucidated utilizing the black-box LIME method. As shown in Table 9, a comparison with prior methods presented in the literature underscores the significance of the success achieved in this investigation [13, 14, 16, 18]. Moreover, under each communication flow mode, the LIME method computed crucial parameters for UAV and anomaly prediction. The model was fine-tuned using Bayesian optimization, enhanc-

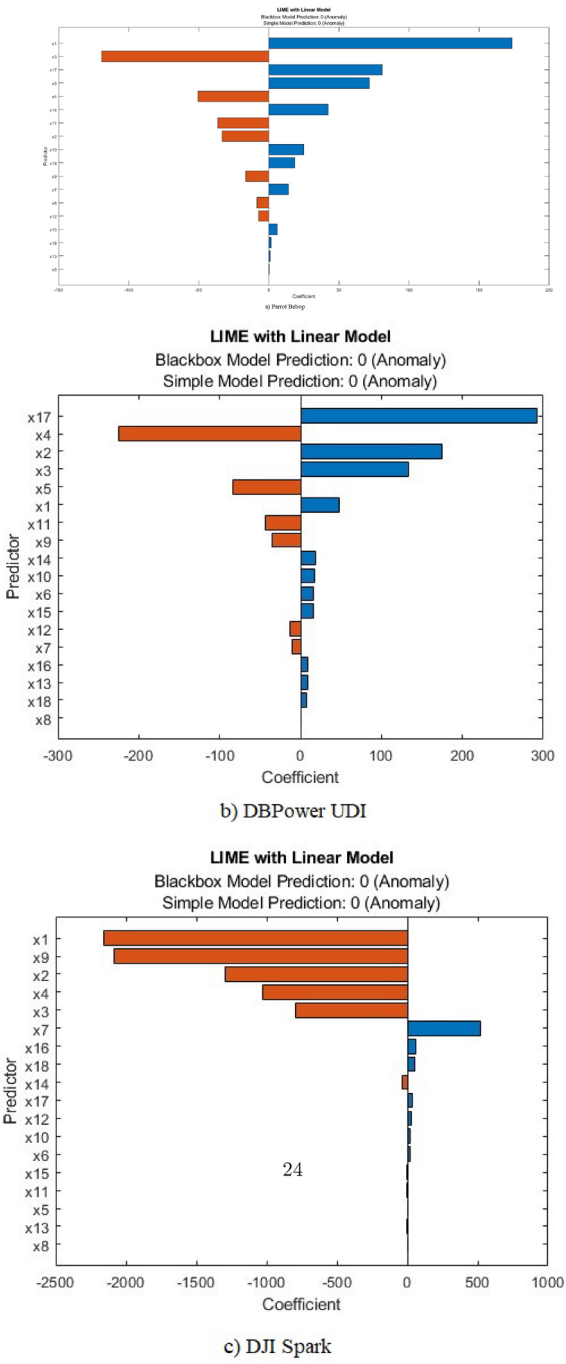

Fig. 11 The LIME model for the unidirectional communication flow mode

Table 9 Literature comparison intrusion detection system with Wi-Fi encrypted data for UAVs

Authors	Model	Test Accuracy
Xie vd. [13]	EBRB model	99.50%
Alipour-Fandid et al. [14]	MLE	88.50%
Al-Haija and Badawi [16]	CNN	99.50%
Alheeti et al. [18]	DNN	99.99%
This study	Linear SVM + LIME	Bidirectional flow mode: Parrot Bebop 1: 100% DBPower UDI: 100% DJI Spark: 99.94% Unidirectional flow mode: 100% (for 3 types UAVs)

ing its predictive accuracy and robustness. A cost analysis of the proposed model was also performed, demonstrating its practicality and efficiency. In the future, particularly with the augmentation of encrypted Wi-Fi data derived from DJI Spark drones, the objective will be to attain more auspicious outcomes. This study provided a wide range of encrypted Wi-Fi traffic data from three different UAV networks. Using the LIME technique, the decision-making process of the developed model was made transparent, and critical features were identified. The model was trained and evaluated while considering both unidirectional and bidirectional communication flow modes, proving that the developed model performs better than existing methods and that it is more suitable for practical applications. These modifications highlight the place of this study in the literature and its original contributions, emphasizing the efficacy and cost-effectiveness of the enhanced model, which are achieved through Bayesian optimization.

Author Contributions Şengül Bayrak is the sole author.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

Data Availability No datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare no Conflict of interest.

References

1. Gupta L, Jain R, Vaszkun G (2016) Survey of important issues in UAV communication networks. IEEE Commun Surv Tutor 18(2):1123–1152. https://doi.org/10.1109/comst.2015.2495297
2. Stöcker C, Bennett R, Nex F, Gerke M, Zevenbergen J (2017) Review of the current state of UAV regulations. Remote Sens 9(5):459. https://doi.org/10.3390/rs9050459
3. Mohsan SAH, Othman NQH, Khan MA, Amjad H, Żywiołek J (2022) A comprehensive review of micro UAV charging techniques. Micromachines 13(6):977. https://doi.org/10.3390/mi13060977
4. Li Z, Zhang Y (2022) Constrained ESKF for UAV positioning in indoor corridor environment based on IMU and Wi-Fi. Sensors 22(1):391. https://doi.org/10.3390/s22010391
5. Abro GEM, Zulkifli SABM, Masood RJ, Asirvadam VS, Laouti A (2022) Comprehensive review of UAV detection, security, and communication advancements to prevent threats. Drones 6(10):284. https://doi.org/10.3390/drones6100284
6. Rachmawati TSN, Kim S (2022) Unmanned aerial vehicles (UAV) integration with digital technologies toward construction 4.0: a systematic literature review. Sustainability 14(9):5708. https://doi.org/10.3390/su14095708
7. Mangewa LJ, Ndakidemi PA, Munishi LK (2019) Integrating UAV technology in an ecological monitoring system for community wildlife management areas in Tanzania. Sustainability 11(21):6116. https://doi.org/10.3390/su11216116
8. Fu R, Ren X, Li Y, Wu Y, Sun H, Al-Absi MA (2023) Machine learning-based UAV assisted agricultural information security architecture and intrusion detection. IEEE Internet Things J. https://doi.org/10.1109/JIOT.2023.3236322
9. Abu Al-Haija Q, Al Badawi A (2022) High-performance intrusion detection system for networked UAVs via deep learning. Neural Comput Appl. https://doi.org/10.1007/s00521-022-07015-9
10. Li T, Hong Z, Cai Q, Yu L, Wen Z, Yang R (2021) BISSIAM: Bispectrum siamese network based contrastive learning for UAV anomaly detection. IEEE Trans Knowl Data Eng. https://doi.org/10.1109/tkde.2021.3118727
11. Anwar MZ, Kaleem Z, Jamalipour A (2019) Machine learning inspired sound-based amateur drone detection for public safety applications. IEEE Trans Veh Technol 68(3):2526–2534. https://doi.org/10.1109/tvt.2019.2893615
12. Nemer I, Sheltami T, Ahmad I, Yasar AU-H, Abdeen MAR (2021) RF-based UAV detection and identification using hierarchical learning approach. Sensors 21(6):1947. https://doi.org/10.3390/s21061947
13. Xie Y, He W, Zhu H, Yang R, Mu Q (2022) A new unmanned aerial vehicle intrusion detection method based on belief rule base with evidential reasoning. Heliyon 8(9):e10481. https://doi.org/10.1016/j.heliyon.2022.e10481
14. Alipour-Fanid A, Dabaghchian M, Wang N, Wang P, Zhao L, Zeng K (2019) Machine learning-based delay-aware UAV detection and operation mode identification over encrypted Wi-Fi traffic. IEEE Trans Inf Forensics Secur 15:2346–2360. https://doi.org/10.1109/tifs.2019.2959899
15. Khan IA, Moustafa N, Razzak I, Tanveer M, Pi D, Pan Y, Ali BS (2022) XSRU-IoMT: explainable simple recurrent units for threat detection in internet of medical things networks. Future Gener Comput Syst 127:181–193
16. Abu Al-Haija Q, Al Badawi A (2022) High-performance intrusion detection system for networked UAVs via deep learning. Neural Comput Appl. https://doi.org/10.1007/s00521-022-07015-9
17. Tan X, Su S, Zuo Z, Guo X, Sun X (2019) Intrusion detection of UAVs based on the deep belief network optimized by PSO. Sensors 19(24):5529. https://doi.org/10.3390/s19245529
18. Alheeti A, Khaled Alarfaj F, Alreshoodi M, Naif Almusallam, Al Dosary Duaa (2023) A hybrid security system for drones based on ICMetric technology. PLOS ONE 18(3):e0282567–e0282567. https://doi.org/10.1371/journal.pone.0282567
19. Medaiyese O, Ezuma M, Lauf AP, Guvenc I (2021) Wavelet transform analytics for RF-based UAV detection and identification system using machine learning. arXiv:2102.11894 [eess]. Available: arXiv:2102.11894
20. Ezuma M, Erden F, Kumar Anjinappa C, Ozdemir O, Guvenc I (2020) Detection and classification of UAVs using RF fingerprints in the presence of Wi-Fi and bluetooth interference. IEEE Open J Commun Soc 1:60–76. https://doi.org/10.1109/ojcoms.2019.2955889
21. Slimane HO, Benouadah S, Al Shamaileh K, Devabhaktuni V, Kaabouch N (2022) ADS-B message injection attack on UAVs: assessment of SVM-based detection techniques. In: 2022 IEEE international conference on electro information technology (eIT). IEEE, pp 405–410
22. Panice G, Luongo S, Gigante G, Pascarella D, Di Benedetto C, Vozella A, Pescapè A (2017) A SVM-based detection approach for GPS spoofing attacks to UAV. In: 2017 23rd international conference on automation and computing (ICAC). IEEE, pp 1–11

23. Shafique A, Mehmood A, Elhadef M (2021) Detecting signal spoofing attack in UAVs using machine learning models. IEEE Access 9:93803–93815
24. Khan IA, Moustafa N, Pi D, Sallam KM, Zomaya AY, Li B (2021) A new explainable deep learning framework for cyber threat discovery in industrial IoT networks. IEEE Internet Things J 9(13):11604–11613
25. KDD (2018)—Prediction-time Efficient Classification Using Feature Computational Dependencies. www.kdd.org. https://www.kdd.org/kdd2018/accepted-papers/view/prediction-time-efficient-classification-using-feature-computational-depend. Accessed 17 Sep 2023
26. Hafeez B, Kabir MH, Li X (2021) Measuring bank risk: forward-looking z-score. Social Science Research Network
27. Joachims T (2006) Training linear SVMs in linear time. Knowledge Discovery and Data Mining. https://doi.org/10.1145/1150402.1150429
28. Schuldt C, Laptev I, Caputo B (2004) Recognizing human actions: a local SVM approach. In: Proceedings of the 17th international conference on pattern recognition, 2004. ICPR 2004. https://doi.org/10.1109/icpr.2004.1334462
29. Gelbart MA, Snoek J, Adams RP (2014) Bayesian optimization with unknown constraints. arXiv preprint arXiv:1403.5607
30. Victoria AH, Maragatham G (2021) Automatic tuning of hyperparameters using Bayesian optimization. Evol Syst 12(1):217–223
31. Bayrak S, Yucel E, Takci H (2022) Epilepsy radiology reports classification using deep learning networks. Comput Mater Contin 70(2):3589–3607. https://doi.org/10.32604/cmc.2022.018742
32. Obuchowski NA (2005) ROC analysis. Am J Roentgenol 184(2):364–372. https://doi.org/10.2214/ajr.184.2.01840364
33. Ribeiro MT, Singh S, Guestrin C (2016) Why Should I Trust You?': Explaining the Predictions of Any Classifier. arXiv.org. arXiv:1602.04938
34. Bayrak S (2023) Interpretation of deep network predictions on various data sets using LIME. Explainable, Interpretable, and Transparent AI Systems. Taylor and Francis CRC Press, pp 137–150. https://doi.org/10.1201/9781003442509
35. Khan IA, Razzak I, Pi D, Zia U, Kamal S, Hussain Y (2024) A novel collaborative SRU network with dynamic behaviour aggregation, reduced communication overhead and explainable features. IEEE J Biomed Health Inform. https://doi.org/10.1109/JBHI.2024.3352013

Sengul Bayrak received her PhD in Computer Engineering from Istanbul University - Cerrahpasa in 2021. She has worked as an Assistant Professor in the Istanbul Sabahattin Zaim University Software Engineering Department since 2021. Her fields of study include Machine Learning, Artificial Intelligence, Natural Language Processing, Data Mining, and Image Processing.