2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) | 978-1-6654-0690-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/UEMCON53757.2021.9666744

Tree-based Supervised Machine Learning Models For Detecting GPS Spoofing Attacks on UAS

Ghilas Aissou, Hadjar Ould Slimane, Selma Benouadah, and Naima Kaabouch
School of Electrical Engineering and Computer Science, University of North Dakota
Grand Forks, ND 58202 USA

***Abstract*—The security of Unmanned Aerial System (UAS) networks is becoming crucial as their number and application in several fields are increasing every day. For navigation and positioning, the Global Navigation System (GPS) is essential as it provides an accurate location for the UAS. However, since the civilian GPS signals are open and unencrypted, attackers target them in different ways such as spoofing attacks. To address this security concern, we propose a comparison of several tree-based machine learning models, namely Random Forest, Gradient Boost, XGBoost, and LightGBM, to detect GPS spoofing attacks. In this work, the dataset was built of real GPS signals that were collected using a Software Defined Radio unit and different types of simulated GPS spoofing attacks. The results show that XGBoost has the best accuracy (95.52%) and fastest detection time (2ms), which makes this model appropriate for UAS applications.**

***Index Terms*—UAS, GPS Spoofing Attacks, Detection Techniques, Machine Learning.**

I. Introduction

Nowadays, Unmanned Aerial Systems (UAS) are an essential technology that plays a crucial role in various military and civilian applications. These systems can be fully autonomous or remotely controlled. They are flexible, efficient, and affordable. However, cyber threats targeting these systems are drastically increasing in terms of severity and intensity. These threats are becoming more harmful as attackers are using artificial intelligence techniques to carry out various types of cyberattacks. There are three main types of cyberattacks on UAS: data interception, data manipulation, and denial of service, [1].

UAS rely on the Global Navigation System (GPS) for accurate positioning and velocity estimation. The civilian applications of UAS use the GPS radio frequency link known as the L1 channel and the new civilian signal known as L1C [2]. Unlike military GPS signals, civilian GPS signals are unencrypted, which makes them vulnerable to GPS spoofing and jamming. GPS spoofing is considered the most dangerous type of cyberattack targeting UAS because the attacker can divert the trajectory of the UAS without being detected. A successful GPS spoofing attack can be disastrous not only on material damages or technology stealing but also causing human injuries if the attacker crashes the UAS into populated areas.

Artificial Intelligence (AI) has been used in the cybersecurity field to detect attacks; however, cyber attackers can also use AI-based technologies offensively to implement smart and complex cyber-attacks, called weaponization of AI. For instance, some examples of the applications of AI techniques in cyber-attacks implementation are highlighted in [3]. For example, the IBM Research Laboratory demonstrated how AI can be used in conducting cyberattacks. The research team developed the Deep Locker malware, which is a form of the WannaCry attack. This malware exploits the vulnerabilities of the target system defined by the AI techniques. To the best of the authors' knowledge, no GPS spoofing attack is yet AI-based, but it is very likely to happen soon. AI can be a solution for attackers to bypass classical detection techniques by implementing sophisticated attacks as shown in Fig. 1.

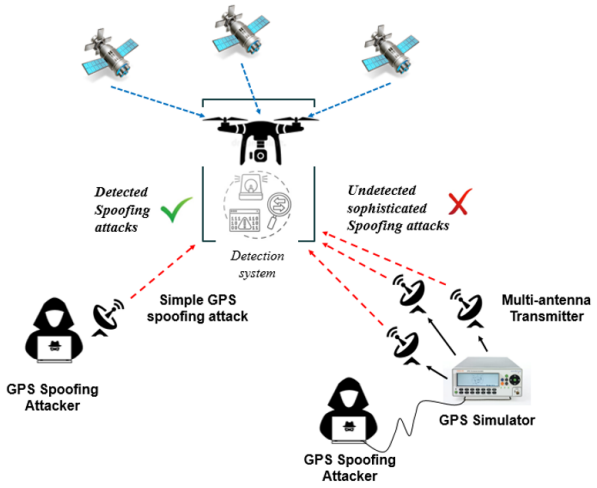

Fig. 1: *GPS Spoofing Attacks.*

A number of works proposed GPS spoofing detection techniques using artificial intelligence. For example, in [4], the authors proposed a GPS Spoofing detection based on the Long Short-Term Memory algorithm. Using existing flight patterns, the authors predicted several UAS flight trajectories and compared them with the received GPS positioning signal. If the GPS signal differs from the predicted data, the algorithm subsequently confirms whether a UAS is under a GPS spoofing attack. However, this is only applicable if the spoofed location causes an error deviation greater than the accepted threshold. In [5], the authors simulated the receiver's clock drift, time derivative of the clock offset, and integrated the spoofed Global Navigation Satellite System (GNSS) dataset

978-1-6654-0690-1/21/$31.00 ©2021 IEEE

with C-Support Vector Machine (C-SVM). Afterward, real-world spoofing data was used to verify and validate the results. The model was trained with an unbalanced dataset, which in most cases lead to misleading performance result of the supervised learning model. The authors of [6] proposed a computer architecture that uses neural network to detect spoofing and meaconing attacks. Once an attack is detected the computer evaluate the severity of the attack using Bayesian inference subsystem.

In [7], the authors compared the performance of different machine learning models, namely Support Vector Machine (SVM), Random Forest (RF), decision trees, Naïve Bayesian (NB), and Linear regression (LR), in detecting actuator-based GPS Spoofing attacks. The authors extracted several features from the GPS signal to create the dataset. They conducted a K-fold analysis to select the best machine learning algorithm. Based on their results, SVM (with the polynomial kernel) outperforms the other methods. In [8], Secure Code Estimation and Replay spoofing attacks were studied for end-user GPS receivers. A comparison based primarily on the probability of detection between five machine learning models was conducted. The implemented machine learning algorithms were Radial Basis Function and linear kernel SVM, Ada Boost, decision trees, Nearest Neighbors, and RF. The obtained results show that algorithms based on decision trees give better detection rates. The authors in [9] proposed GPS-Probe, a machine learning GPS spoofing detection algorithm, that analyzes the Received Signal Strength Indicator and the Time of Arrival of Air Traffic Control messages that are periodically broadcasted by aerial vehicles. They adapted the K-Nearest Neighbors (K-NN) classifier to estimate the location of the vehicle and the Extreme Gradient Boost model (XGBoost) for attack detection.

This paper proposes a systematic and detailed performance evaluation of four learning tree-based supervised learning models, namely, Gradient boost, XGBoost, and Light Gradient Boosting Machine (LightGBM). These models are trained and tested using a dataset described in section II. This dataset contains 10,055 legitimate and simulated attacks samples. Three different types of GPS spoofing attacks with different complexity levels were simulated.

The remainder of this paper is organized as follows: Section II discusses the methods implemented to generate the dataset and the used machine learning models in this work. The obtained results are presented and discussed in Section III. Finally, the conclusion is stated in Section IV.

II. METHODOLOGY

In this section, we discuss the dataset generation steps along with the feature extraction process. Afterward, we describe the machine learning models used to detect GPS spoofing attacks.

A. Dataset Generation

The application of machine learning-based GPS spoofing detection techniques in UAS requires an accurate, reliable, and energy-efficient solution since the UAS is limited in terms of computational resources and power consumption. For the above reasons, the use of supervised learning models is more suitable for such applications.

1) Authentic GPS Signal Collection: In this study, we performed two dynamic scenarios for real-time authentic GPS signal acquisition and feature extraction. The hardware used to build the GPS receiver is a universal Software Defined Radio (SDR) Peripheral (USRP) unit. The open-source GNSS-SDR software is based on GNU radio blocks running in an I5-4300U laptop with 8G RAM running with Ubuntu 16.04.7 LTS version. The actual receiver is an 8-channel receiver capable of parallel tracking of GPS signals.

2) Feature Extraction: We did a real-time features extraction from different receiver stages, starting from the acquisition, tracking to the observable block. During this process, 13 features were extracted as listed below:

Carrier to Noise Ratio ($C/N_\circ$): $C/N_\circ$ is the ratio of the signal power to the noise. The actual fluctuation in the ($C/N_\circ$) is due to the variation in the signal power. The signal-to-noise ratio is estimated to be in the $\pm$ 2.8dB range [10].

Prompt Correlator (PC): PC is the operation used in synchronizing with the incoming GPS signal. This operation is the code-tracking process, which is done by multiplying the local pseudorandom spreading code replica generated at the local code phase generator with the complex digital signal outputs [11].

Early Correlators (EC): EC is $\frac{1}{2}$ chip spacing before the PC [12].

Late Correlator (LC): LC is $\frac{1}{2}$ chip spacing after the PC [12].

Prompt In-Phase Component (PIP): PIP is the in-phase component of the Prompt correlator amplitude.

Prompt Quadrature component (PQP): PQP is the quadrature component of the prompt correlator amplitude. PC can be expressed in terms of PIP and PQP components as shown in (1).

$$PC = \sqrt{PIP^2 + PQP^2} \tag{1}$$

Carrier Phase Cycles (CP): It is given as the accumulated number of cycles expressing the beat frequency difference between the receiver's generated carrier frequency and the satellite's received carrier frequency [13].

Time Of the Week (TOW): It is the decoded information from the GPS signal; it provides the number of seconds elapsed since the start of each week (from 0s to 604,799s) [14].

Receiver Time (RX): It is the receiver reference time, in our receiver, we use the GPS DO-kit for time reference.

Pseudorange (PD): It is the signal transit time between transmission and reception (satellite to the receiver) [15]. It is expressed in terms of meter as shown in (2).

$$P_s = c(t_r - t_s) \tag{2}$$

Where t_r is the reception time and t_s is the transmission time.

Carrier Doppler (DO): It is the result of relative motion of the satellite with respect to the receive known as Doppler

effect [16], as expressed in (3).

$$f = (\frac{c + v_r}{c + v_s}) \times f_i \tag{3}$$

Tracking Carrier Doppler (TCD): It is the Doppler shift measured during the correlation stage, in this work, it is used for defining the upper and lower bounds of the accepted DO.

Pseudorandom Noise (PRN): The satellite identification number.

3) GPS Spoofing Attack Simulation: The authors in [17] divided spoofing attacks into three main categories depending on the complexity and sophistication of the attack: simple attacks, intermediate attacks, and sophisticated attacks. As mentioned earlier, most detection techniques can only detect non-complex spoofing attacks, while they cannot detect sophisticated or even intermediate spoofing attacks. Based on the literature review, we summarize all the attack signatures and simulate the above attacks by modifying the authentic GPS signals.

a) Simplistic Spoofing Attacks: The spoofer generates a fake GPS signal that is unsynchronized with the authentic signal since the attacker is unaware of the receiver's position. This leads to a higher Doppler shift exceeding the normal range of ±20 Hz thus a large deviation in the pseudo-range measurement. In this type of attack, the attacker transmits the spoofed signal at a high power level, resulting in a higher $C/N_\circ$ value, which can easily be detected. Yet, many low-cost commercial GPS spoofing devices are available to orchestrate such an attack [18].

b) Intermediate Spoofing Attacks: In intermediate spoofing attacks, the spoofing attacker knows the position of the target receiver, resulting in code phase alignment between the real and spoofed signals. Unlike a simple spoofing attack, the intermediate attacker can take control of the UAS by precisely controlling the GPS-generated signal. Moreover, a detection system based on signal characteristics cannot detect the temptation since the spoofer takes the detection system into account. The Doppler shift and pseudorange are always kept under control to avoid exceeding the normal ranges. However, close monitoring of the TOW information, carrier phase shift, and correlator amplitude can reveal the effects of the received spoofing signal.

c) Sophisticated Spoofing Attacks: In sophisticated attack scenarios, the attacker uses multiple synchronized antennas to emulate the GPS constellation. This way, the attacker can spoof different channels in parallel gaining complete control over the system. The synchronization between multiple antennas is challenging to achieve but can be easily overcome with advanced SDR technologies. The authors in [19] showed the distortions and the quadrature component shift effect in the tracking correlators during a spoofing attack. However, for such a critical type of attack, a correlator-based detection system may be misleading due to the effect of multipath signals and the motion of the satellite and receiver.

B. Data Preprocessing

Our dataset contains 10,055 samples, of which 55% are real GPS signals and 45% are simulated GPS spoofing attacks. Data preprocessing aims to clean and prepare the data before feeding it into the learning models. In this work, we perform feature correlation analysis using the Spearman technique to remove correlated features and thus reduce the size of the dataset resulting in a less complex and accurate learning model.

1) Spearman Correlation Coefficient: The Spearman correlation coefficient expresses how strong is the monotonic relation between the features as shown in (4). The score is given on a scale of [-1:1], where a score close to the limits (1 or -1) stands for strongly correlated features.

$$S(f_i, f_j) = 1 - 6 \sum_l \frac{(f_i - f_j)^2}{N(N^2 - 1)} \tag{4}$$

Where S is the Spearman rank correlation coefficient, N represents the length of the vector, f represents the rank of feature i.

2) Non-Stationary Data Modification: Machine-learning algorithms cannot handle non-stationary data; a static relationship is needed for a correct learning model. We look for features that follow a non-stationary distribution and convert the data into stationary data using differencing. This process involves calculating the consecutive differences between samples, as it is expressed in (5).

$$R = \frac{x_{i+1} - x_i}{n_{i+1} - n_i} \tag{5}$$

Where R is the instantaneous rate of change and $n_{i+1} - n_i$ is the distance between two samples, which in our case is equal to 1.

C. Machine Learning Classification Models

In this study, we compare the performance of four different supervised tree-based machine-learning algorithms, namely, Random Forest, Gradient Boost, XGBoost, and LightGBM. Random Forest is based on the bagging method of ensemble learning, which combines predictions from multiple decision trees using a majority voting method to improve the predictive accuracy and control over-fitting. On the other hand, Gradient Boost is an ensemble model of boosting where trees are constructed and trained sequentially to minimize the error from the previous trees using the gradient descent algorithm.

XGBoost is an optimized Gradient Boosting algorithm that utilizes parallelization, tree pruning, and regularization to avoid overfitting and improve computational performance [20]. LightGBM is also based on the Gradient Boosting algorithm; however, it uses Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) to reduce its complexity. GOSS reduces the size of data by removing instances with small gradients, while EFB is used to decrease the number of features. This process will yield a lower storage consumption and a higher performance speed [21].

III. Results and Discussion

In this study, the dataset was divided into 70% and 30% for training and testing, respectively. The dataset contains 10,055 samples where 55% are authentic signals and 45% are spoofed GPS signals. The results of the investigation are shown in Fig. 2 through Fig. 6.

The Spearman correlation results show that the pairs ([RX,TOW], [TCD,DO]) are highly correlated with a score of 0.94 as shown in Fig. 2. Therefore, we removed the RX and TCD features. This was quite expected since the RX is the reference time of the receiver and the TCD is the doppler measurement in the tracking blocks.

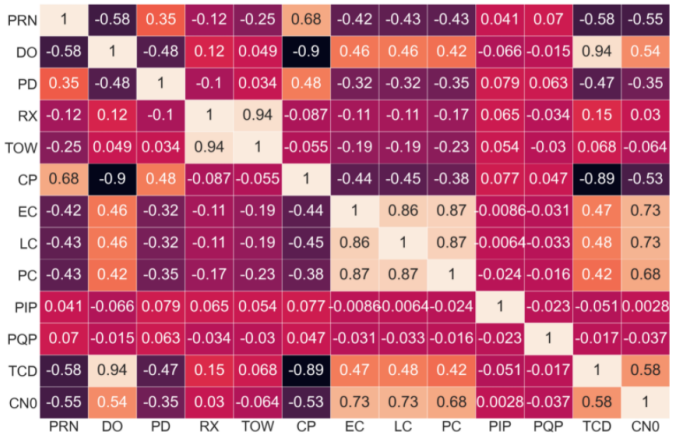

Fig. 2: *Spearman's Correlation Coefficient Heatmap.*

Fig. 3 shows raw TOW information, which has a non-stationary distribution, while first-order differencing results are represented in Fig. 4.

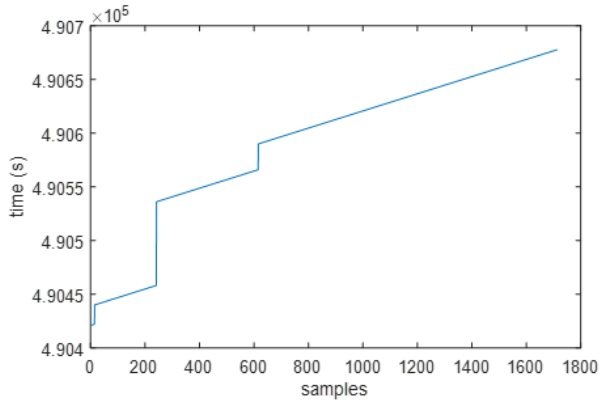

Fig. 3: *Example of Non-stationary Data.*

Fig. 5, Fig. 6 and Table I show the results of the machine learning models. As one can see in Fig. 5, XGBoost has the best accuracy (95.52%) followed by LightGBM (95.23%), Random Forest (94.07%), and Gradient Boost (91.45%). Fig. 6 shows the probability of misdetection, which is considered as also an important parameter in the cybersecurity arena since it

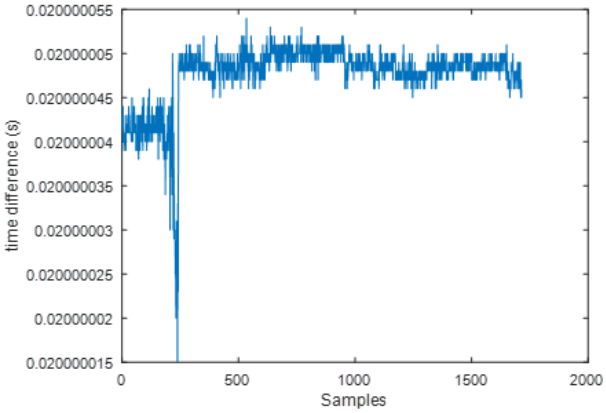

Fig. 4: *Example of Stationary Data.*

defines the number of undetected attacks, Random Forest has the best results followed by XGBoost and LightGBM then Gradient Boost.

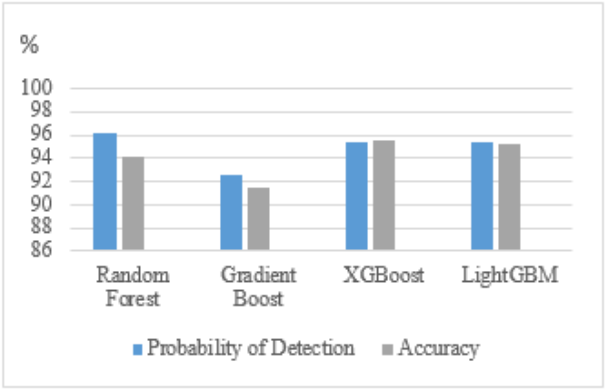

Fig. 5: *Probability of Detection and Accuracy.*

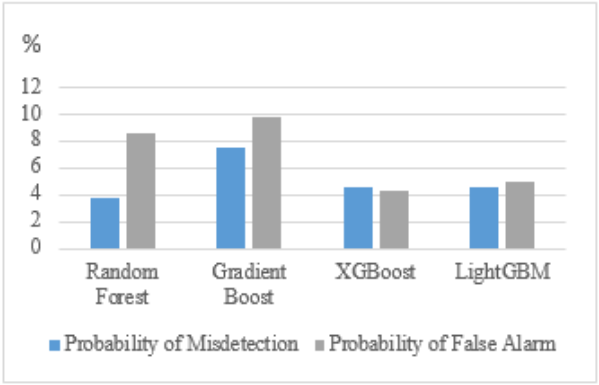

Fig. 6: *Probabilities of Misdetection and False Alarm.*

The detailed results of the evaluation metrics, probability of detection, probability of misdetection, probability of false alarm, and accuracy of the machine learning models are represented in Table I.

TABLE I: *Evaluation Metrics Results.*

	RF	GBM	XGB	LGBM
Probability of Detection	96.23%	92.52%	95.38%	95.38%
Probability of Misdetection	3.77%	7.48%	4.62%	4.62%
Probability of False Alarm	8.53%	9.84%	4.30%	4.96%
Accuracy	94.07%	91.45%	95.52%	95.23%

As shown in Table II, XGBoost is at least three times faster and occupies two times less memory than the other models in the detection stage. While Random Forest shows the worst results in the processing time, but it maintains good memory size in both the training and detection. Since UAS are limited in computational power and memory, we are mostly interested in a fast model that occupies less memory. To conclude, we can say that XGBoost is the most suitable model for UAS applications to detect GPS spoofing attacks.

TABLE II: *Performance Metrics Results.*

		RF	GBM	XGB	LGBM
Processing Time (ms)	**Training**	617.01	265.01	395.98	269.00
	detection	21.01	6.99	2.00	8.99
Memory Size (MiB)	**Training**	1.56	3.16	2.62	1.49
	detection	0.04	0.03	0.01	0.52

IV. Conclusion

This paper proposes a comparison between several tree-based machine learning models to detect GPS spoofing attacks in order to select a suitable model for UAS. The implemented models are RF, Gradient Boost, XGBoost, and LightGBM. The results show that all models can detect GPS spoofing attacks effectively in less than 22 milliseconds. Yet, the performance of XGBoost surpasses the other models in terms of accuracy, detection time, and memory size. This model is three times faster and occupies two times less memory than the other models in the detection phase, which makes it ideal for UAS considering the Size, Weight, and Power (SWaP) constraints.

Acknowledgment

The authors acknowledge the support of the National Science Foundation (NSF) through the Award Number: 2006674.

References

[1] M. R. Manesh and N. Kaabouch, "Cyber-attacks on unmanned aerial system networks: Detection, countermeasure, and future research directions," *Computers & Security*, vol. 85, pp. 386–401, 2019.

[2] F. Anthony, "Navstar gps space segment/user segment l1c interfaces," accessed on: Nov. 4, 2021. [Online]. Available: https://www.gps.gov/technical/

[3] M. M. Yamin, M. Ullah, H. Ullah, and B. Katt, "Weaponized ai for cyber attacks," *Journal of Information Security and Applications*, vol. 57, p. 102722, 2021.

[4] S. Wang, J. Wang, C. Su, and X. Ma, "Intelligent detection algorithm against uavs' gps spoofing attack," in *2020 IEEE 26th International Conference on Parallel and Distributed Systems (ICPADS)*. IEEE, 2020, pp. 382–389.

[5] S. Semanjski, A. Muls, I. Semanjski, and W. De Wilde, "Use and validation of supervised machine learning approach for detection of gnss signal spoofing," in *2019 International Conference on Localization and GNSS (ICL-GNSS)*. IEEE, 2019, pp. 1–6.

[6] N. Kaabouch, M. R. Manesh, and J. R. Kenney, "Detection of spoofing and meaconing for geolocation positioning system signals," Jul. 16 2020, uS Patent App. 16/367,961.

[7] A. Shafique, A. Mehmood, and M. Elhadef, "Detecting signal spoofing attack in uavs using machine learning models," *IEEE Access*, vol. 9, pp. 93 803–93 815, 2021.

[8] F. Gallardo and A. P. Yuste, "Scer spoofing attacks on the galileo open service and machine learning techniques for end-user protection," *IEEE Access*, vol. 8, pp. 85 515–85 532, 2020.

[9] G. Liu, R. Zhang, C. Wang, and L. Liu, "Synchronization-free gps spoofing detection with crowdsourced air traffic control data," in *2019 20th IEEE International Conference on Mobile Data Management (MDM)*. IEEE, 2019, pp. 260–268.

[10] C. J. Wullems, "A spoofing detection method for civilian l1 gps and the e1-b galileo safety of life service," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 4, pp. 2849–2864, 2012.

[11] E. Shafiee, M. Mosavi, and M. Moazedi, "Detection of spoofing attack using machine learning based on multi-layer neural network in single-frequency gps receivers," *The Journal of Navigation*, vol. 71, no. 1, pp. 169–188, 2018.

[12] K. D. Wesson, B. L. Evans, and T. E. Humphreys, "A combined symmetric difference and power monitoring gnss anti-spoofing technique," in *IEEE Global Conference on Signal and Information Processing*. IEEE, 2013, pp. 217–220.

[13] J. J. Spilker Jr, P. Axelrad, B. W. Parkinson, and P. Enge, *Global positioning system: theory and applications, volume I*. American Institute of Aeronautics and Astronautics, 1996.

[14] M. L. Psiaki and T. E. Humphreys, "Gnss spoofing and detection," *Proceedings of the IEEE*, vol. 104, no. 6, pp. 1258–1270, 2016.

[15] A. Rustamov, N. Gogoi, A. Minetto, and F. Dovis, "Assessment of the vulnerability to spoofing attacks of gnss receivers integrated in consumer devices," in *2020 International Conference on Localization and GNSS (ICL-GNSS)*. IEEE, 2020, pp. 1–6.

[16] A. Jovanovic, C. Botteron, and P.-A. Fariné, "Multi-test detection and protection algorithm against spoofing attacks on gnss receivers," in *Proceedings of IEEE/ION PLANS 2014*, 2014, pp. 1258–1271.

[17] Z. Haider and S. Khalid, "Survey on effective gps spoofing countermeasures," in *2016 Sixth International Conference on Innovative Computing Technology (INTECH)*. IEEE, 2016, pp. 573–577.

[18] M. R. Manesh, J. Kenney, W. C. Hu, V. K. Devabhaktuni, and N. Kaabouch, "Detection of gps spoofing attacks on unmanned aerial systems," in *2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2019, pp. 1–6.

[19] A. Cavaleri, B. Motella, M. Pini, and M. Fantino, "Detection of spoofed gps signals at code and carrier tracking level," in *2010 5th ESA Workshop on Satellite Navigation Technologies and European Workshop on GNSS Signals and Signal Processing (NAVITEC)*. IEEE, 2010, pp. 1–6.

[20] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[21] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in neural information processing systems*, vol. 30, pp. 3146–3154, 2017.