RF-based data acquisition and detection of drones using Machine Learning.

[*]A project survey for the course of "Wireless Networks & Mobile Application" by Professor Claudio Enrico Palazzi

[*]The project Dataset and code is available publicly available at: https://github.com/SRiazRaza/WNMA

Syed Riaz Raza
Dipartimento di Matematica "Tullio Levi-Civita" (DM)
Universita degli studi di Padova
Padova, Italy
riazraza0@gmail.com
riazraza.me

***Abstract*—Drones have emerged as versatile tools with applications across various domains, yet they present considerable risks to public safety and security. In response to these threats, numerous anti-drone solutions have been developed; however, most rely on the drone's prior detection by the defender, posing a significant challenge. This survey introduces a novel drone detection system that harnesses radio frequency wireless signals collected through different platforms for autonomous drone detection and identification. Initially, two distinct technical approaches were proposed [1]: active tracking, wherein the system emits a radio signal and analyzes its reflection, and passive listening, wherein the system captures, extracts, and analyzes wireless signals emitted by the drone. Afterwards, the approach was refined to solely focus on passive listening, culminating in the successful implementation of preliminary experiments for ML model.**

The experiments involved the utilization of 2.4GHz of Tx/Rx for data acquisition. To further enhance the system's accuracy and robustness, a machine learning model, was integrated to introduce a factor of noise into the analysis. This allowed for an evaluation of the feasibility and effectiveness of the proposed approaches.

The preliminary results underscore the potential of the developed system, demonstrating its proficiency in detecting and identifying drones autonomously using SDR based platforms. By presenting the findings, this research paves the way for future investigations and advancements in drone detection technologies. The combination of passive listening and machine learning showcases a promising direction for bolstering security measures against the ever-evolving landscape of drone threats.

***Index Terms*—UAVs, Drone Detection; Wireless Technology; RF; Machine Learning; SDR**

I. INTRODUCTION

Unmanned Aerial Vehicles (UAVs), also known as Unmanned Aerial Systems (UAS), have emerged as a revolutionary technology with a wide range of applications, encompassing recreational, civilian, and professional domains. The term "drone" has become prevalent among hobbyists and the general public, referring to these autonomous or remotely controlled aircraft. Despite their growing popularity and diverse applications, the distinction between UAS and UAV remains blurry, as they are often used interchangeably in literature and common language.

Drones have found diverse applications across various industries, ranging from aerial photography and videography to mapping and surveying, search and rescue operations, precision agriculture, and scientific research [2]. Notably, Amazon has put forth an ambitious proposal to revolutionize package delivery using drones in their "Amazon Prime Air" project [4].

Their vision for the future involves the use of drones to transport packages through the airspace, potentially transforming the logistics industry and offering a novel approach to efficient and rapid delivery services. These multifaceted applications of drones showcase the significant impact they have on modern technology and their potential to shape various sectors in the foreseeable future. This survey will explore the advancements in drone applications and contribute to a deeper understanding of their capabilities and potential privacy penetration in these domains.

The proliferation of UAVs in recent years has been fueled by advancements in embedded sensor technology, making these aerial devices more accessible and affordable for civilians. The societal impact of UAVs has been substantial, with applications spanning educational and commercial sectors. However, alongside their beneficial applications, the use of drones has also raised concerns about potential negative effects and challenges in their detection and monitoring.

One of the primary challenges faced by researchers and hobbyists is the reliable detection of UAV signals. These flying devices can operate in various directions, necessitating monitoring equipment capable of tracking multiple flight paths simultaneously. Moreover, distinguishing UAVs from other airborne objects, such as kites or birds, poses an additional difficulty, especially when the drone is operating at a distance from the detection module.

Another constraint affecting UAVs is their dependence on power supply, leading to limitations in battery life and communication range. As a result, consumer-grade drones are often

confined to low altitudes during operation. Additionally, obstacles in the surrounding environment can hinder the seamless utilization of drones.

In light of the rapid expansion of UAV applications, this survey explores the challenges associated with UAV detection and monitoring. It delves into the various obstacles faced by researchers and hobbyists in effectively distinguishing UAV signals from other flying objects and explores potential solutions to overcome these hurdles. By addressing these challenges, I aim to contribute to the development of efficient and reliable UAV detection systems, enabling the responsible and safe integration of drones into our daily lives.

Apart from their constructive applications, civilian drones have increasingly become a cause for concern due to their problematic uses. Instances of drone-related incidents have raised alarms among both the public and authorities. For instance, a significant incident occurred on March 29, 2016, near Los Angeles International Airport (LAX) when a Lufthansa jet narrowly avoided a collision with a drone, with just 200 feet separating the two [5]. Furthermore, drones have been known to interfere with firefighting efforts during forest fires in certain regions [6]. Notably, a drone crash disrupted a U.S. Open tennis match, instilling fear in spectators, and one player expressed apprehension, stating, "It was a little bit scary, I have to say because with all the things happening now in the world, I imagine maybe it's a bomb" [7].

In addition to aviation safety concerns, drones have also raised security risks. One such incident involved a drone crashing at the White House, sparking fears about the vulnerability of government buildings and facilities [8]. Moreover, drones have been associated with instances of stalking and privacy violations, with accusations of their misuse to intrude on individuals' personal lives [9].

As the prevalence of civilian drones continues to grow, addressing the challenges posed by their problematic use becomes increasingly imperative. This thesis aims to delve into these concerns and explore potential solutions to promote the responsible and safe integration of drones into our society. By mitigating the risks associated with drone operations, we can facilitate the positive contributions of this technology while safeguarding public safety and privacy.

II. BACKGROUND

A. Drone Detection Methods

Various methods for detecting drones have been explored in the literature, including audio, video, thermal, radar, and radio frequency detection.

- One notable approach discussed in [10] involves using an array of microphones to capture ambient drone sounds, focusing on the distinct high-frequency noise generated by brushless motors.
- In [11], a passive radar technique was implemented for drone signal detection. The radio frequency (RF) detection method, as outlined in [1], has proven effective for long-range drone detection due to the communication protocols used by drones, making it challenging for drones to evade complete RF detection.
- On the other hand, video detection faces difficulties in accurately distinguishing drones from birds, particularly when birds are gliding [10].
- Thermal detection relies on the detection of infrared radiation emitted by warmer objects, making it suitable for drones with propulsion engines, but less reliable for plastic quadcopters with electric motors that emit less heat [12]. Radar, commonly useful for detecting larger aircraft, struggles with the detection of small drones like quadcopters, as they were not primarily designed for this purpose [11].
- The effectiveness of radio frequency detection, as outlined in [13], depends on the power of the transmitter and the sensitivity of the receiver. As the drone technology landscape continues to evolve, comprehensive and reliable detection mechanisms are crucial to address safety and security concerns associated with drone usage.

B. Working of Drones

The drone comprises two essential components: the remote control and the aircraft, and they establish communication through a radio frequency link. The figure below illustrates the architecture and components commonly found in most drones. The standard frequencies utilized for communication between the remote control and the drone are 433 MHz or 933 MHz for telemetry, 2.4GHz for control, and 5.8 GHz for FPV systems / OSD (On-Screen Display).

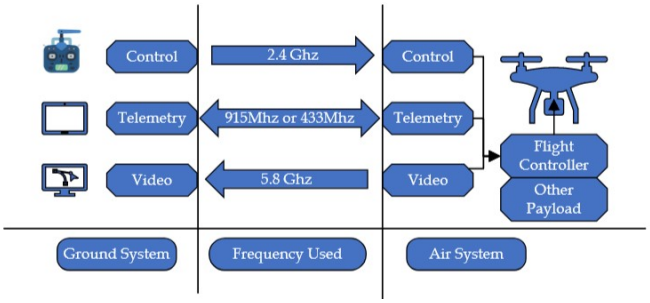

Fig. 1. Working components of drones and frequency used.

In our RF classification, our focus will be on the 2.4GHz spectrum communication between the ground (remote controller) and the air (aircraft). The remote controller facilitates direct control of the drone by enabling control, data, and video transmission, allowing seamless communication between the two components. Most controllers utilize the 2.4 GHz spectrum, employing a proprietary frequency hopping spread spectrum (FHSS) modulation technique. This FHSS modulation is commonly used in military applications and select electronic devices, such as cell phones, due to its ability to maximize robustness and control distance.

The pairing of the transmitter and receiver enhances the resilience of these signal modulation types against interference, making them challenging to detect. In military applications,

this technology can be used for low probability of detection (LPD), ensuring that the enemy remains unaware of the signal's presence. However, this characteristic also makes it more difficult to detect drones employing FHSS controllers.

Additionally, FHSS controllers offer effectiveness against jamming as they transmit information across the entire width of the spectrum in which the drone is hopping. Both the remote controller and the aircraft facilitate the transmission and reception of control, video, and other types of signals between them. Table 1 provides a list of information associated with control signals for some available drones used by hobbyists, students, or researchers.Table 2 shows some information associated with the communication signal for the video transmission between the remote controller and the aircraft.

Brand	Frequency	Modulation	Technology
DJI Phantom	2.4Ghz/5.8Ghz	FHSS/DSSS	FASST/Lightbridge
Futaba	2.4Ghz	FHSS/DSSS	FASST
Spektrum	2.4Ghz	FHSS/DSSS	DSMX
Parrot AR2	2.4Ghz/5.8Ghz	OFDM	Wi-Fi
JR	2.4Ghz	FHSS/DSSS	DMSS
Hitec	2.4Ghz	FHSS/DSSS	AFHSS
Graupner	2.4Ghz	FHSS/DSSS	HOTT
Yuneec	2.4Ghz	DSSS	ZigBee

TABLE I
POPULAR CONTROLS(RC) USED BY DRONES, FREQUENCIES THEY OPERATE ON AND THE USED MODULATION TECHNIQUE AND TECHNOLOGY.

Brand	Frequency	Modulation
DJI Phantom	2.4Ghz	OFDM
Immersion	2.4Ghz	FM
Yuneec	5.8Ghz	OFDM
Connex	5.8Ghz	OFDM
Immersion	5.8Ghz	FM
Boscam	5.8Ghz	FM

TABLE II
POPULAR VIDEO TRANSMISSION USED BY DRONES, FREQUENCIES THEY OPERATE ON AND THE USED MODULATION TECHNIQUE AND TECHNOLOGY .

C. Experimented Approach

The approach is inspired by the concept discussed in [1] Fig. 2 (b). Our goal is to utilize RF detection knowledge for the detection, feature extraction, and classification of drone signals using machine learning. In this survey, we aim to extract drone signal parameters, energy, and other statistics from received signals. To enhance data flexibility and account for noise and interference, we introduced additional classes like distance and experimented with data collection during motor rotation and rest to simulate vibrations.

Once the features are extracted from the incoming signals of interest, we will proceed with further processing. This involves labeling and training the collected features since we plan to use supervised learning in machine learning to classify the drone signals. Python will be utilized for data training and processing.

After completing the training in Python, a testing file will be generated and it can be loaded into the GNU radio block for real-time testing and analysis. The use of GNU radio block requires an SDR, and for 2.4 GHz frequency, HackRF One is required. Although beyond the scope of this survey, it could be considered for future work. The hardware used to collect the data for Because of lack of funds, the experimented approach didn't include HackRF One, which would have been an invaluable asset to this survey, speacially the hack-rf sweep function, SDR and GNU could have saved the time spent to preprocess the raw data, and HackRF One is robust frequency range, we could have tested

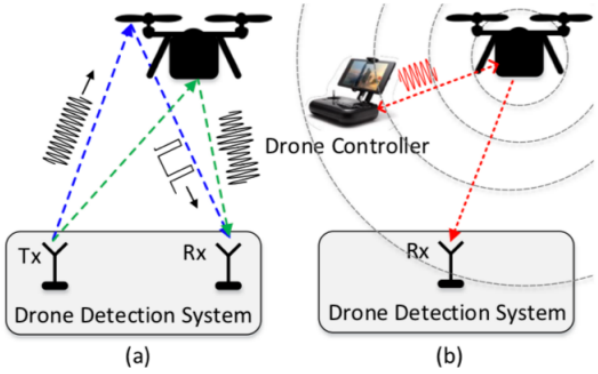

Fig. 2. The overview of drone detection system: (a) active and (b) passive approaches

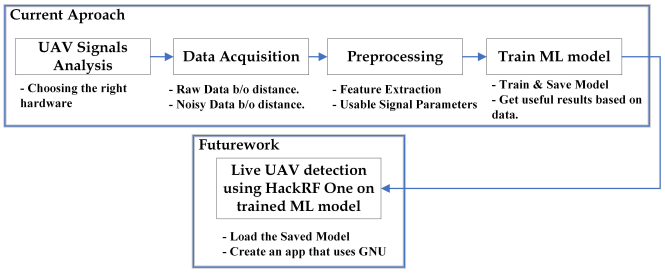

Fig. 3. Process diagram of experimented approach

D. Hardware

The signal analysis hardware utilized in this study consists of a Flysky FS-i6X transmitter, featuring 6 channels and an RF range of 2.408 - 2.475GHz. For reception, a Flysky FS-iA10B receiver is employed, equipped with 10 PWM channels, operating at a wireless frequency of 2.4GHz, and using the AFHDS 2A wireless protocol, which additionally supports RSSI and various Data ports such as PWM, PPM, i.bus, and s.bus. The drone used is a DIY open-sourced drone originating from China.

Initially, the receiver was connected to an ArduPilot (RCS) capable of handling multiple data protocols and offering application support for built-in Python scripts. This setup was

used to gain a deeper understanding of the signals and the communication between the transmitter and receiver.

Subsequently, for data acquisition purposes, the receiver was connected to a Raspberry Pi to collect data. The ArduPilot and Raspberry Pi worked independently. To introduce noise into the data, measurements were taken both while the motors were running and while the system was at rest. Moreover, distance measurements in multiples of 5 (ranging from 5 to 50) were incorporated to analyze signal strength and other relevant parameters.

III. DATA SET AND PROCESSING

A. Raw Data Set

The raw data was gathered in two distinct types. The first type, which contained PWM values from channel 1 to channel 6 of the transmitter. To validate these values, we used ArduPilot applications. The second dataset "Table III" focused on Received Signal Strength Indications (RSSIs), obtained freely through the energy detection modules of wireless devices. We captured RSSI data on the 2.4 GHz frequency band using the spectrum analysis function of Flysky FS-iA10B, and the public dataset is accessible in Github Repo.

The raw RSSI dataset was divided into sets of ten CSV files, each corresponding to a specific distance from the drone in multiples of 5, ranging from 5 meters to 50 meters. Additionally, another set of RSSI data was collected with the drone's motors at rest, introducing increased noise. This step was undertaken to generate positive and negative examples for training our Machine Learning (ML) model [14].

Each representative distance in multiples of 5 had a collection of 60 seconds, resulting in N = 48,766 samples within the frequency range of 2.400 GHz to 2.490 GHz. FFT bin width (frequency resolution), was set at 600000 hz.

Parameter	Value
Date	date
Time	time
(hz) Low	2.40Ghz
(hz) High	2.49Ghz
(hz) Bin Width	600000
Num Samples	48766
db*10	float value

TABLE III
THE STRUCTURE OF RAW DATASET THAT FOCUSED ON RSSI PARAMETERSS

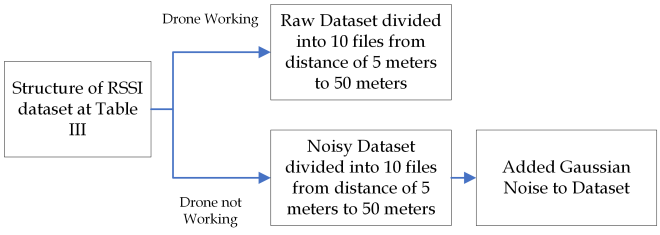

Fig. 4. Dataset structure and division

B. Prepossessing

To enhance the effectiveness of our ML model, we follow a preprocessing step before feeding the raw RSSI samples. This involves handling missing bins, gaining insights from the data, and carefully selecting essential parameters for training the ML model.

To improve data structure, we transformed the dataset, where each row represents a complete sweep (containing 180 bins) indexed by datetime timestamp. The columns, corresponding to each bin, are labeled with the starting frequency in hertz, resulting in a data frame with 180 rows.

For each frequency, we generated a list of dB values using specific equations:

1. Calculate the number of frequency bins:

$$m = \text{num\_bins}$$

2. Find the minimum and maximum frequencies in the dataset:

$$f_{\text{min}} = \min(\text{frequency\_values})$$

$$f_{\text{max}} = \max(\text{frequency\_values})$$

3. Calculate the width of each frequency bin:

$$\Delta f = \frac{f_{\text{max}} - f_{\text{min}}}{m}$$

4. Set the column names of the DataFrame to represent the frequency values of each bin using sigma notation:

$$\text{db\_column} = [f_{\text{min}} + \sum_{x=0}^{m-1} (\Delta f \cdot x)]$$

.

By carrying out this preprocessing, we ensure that the data is properly organized and ready for training our ML model effectively.

Fig. 5 displays the heatmap of the raw dataset, recording frequency bins at distances ranging from 5 meters to 50 meters. The decibel (db) values are calculated based on the mean of 180 bins in each sweep row, representing positive examples. In contrast, Fig. 6 focuses on negative examples by recording data when the drone motors are at rest, capturing only frequency communication and background noise without motor vibrations. The difference between both figures is not significant, and the mean Signal-to-Noise Ratio (SNR) was calculated to be very low.

For effective drone detection using machine learning, it is desirable to have a higher SNR as it enhances model accuracy, feature extraction, detection range, and resistance to noise and interference. However, achieving a high SNR may require careful data acquisition and consideration of various factors, such as recording data in a soundless room with no background noise. To artificially increase the SNR, we added Gaussian noise with a standard deviation of 0.2 Fig. 7 . This approach aims to improve the overall detection performance.

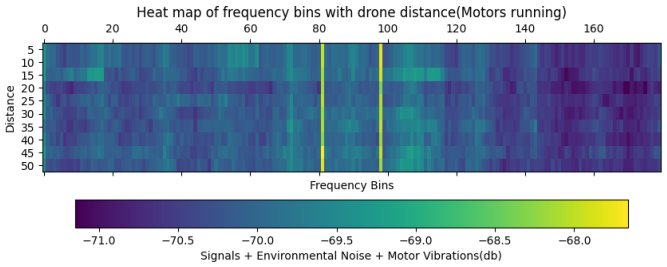

Fig. 5. Heat map of frequency bins and drone distance(Motors running)

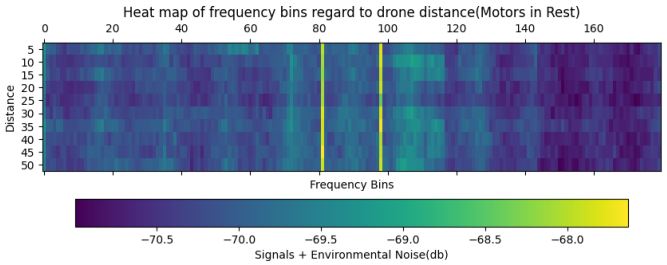

Fig. 6. Heat map of frequency bins and drone distance during motors in rest, no vibrations

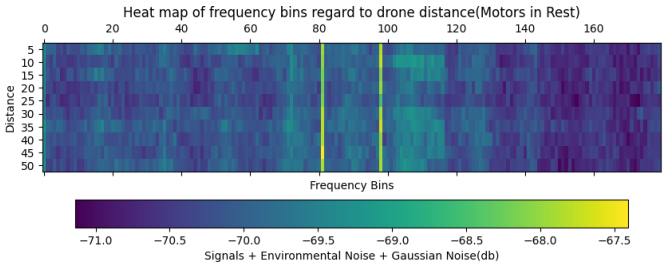

Fig. 7. Heat map of frequency bins and drone distance when motors in rest with addition of Gaussian Noise(0.2)

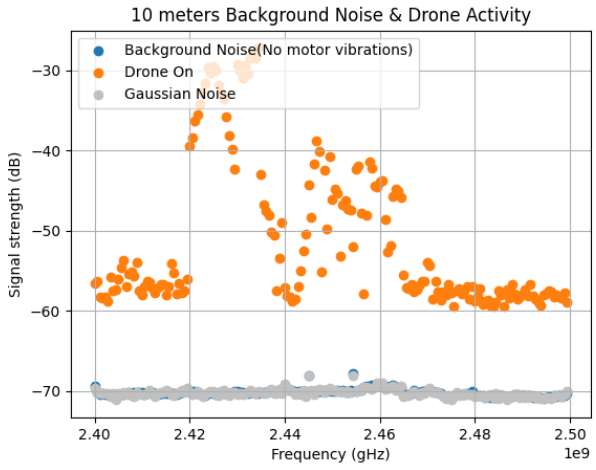

Fig. 8. The 10-meter signal strength comparison of data using Mean Function in Fig. 5,6,7

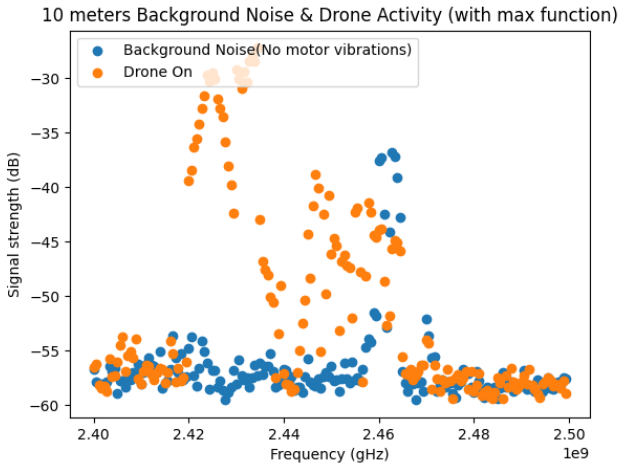

Fig. 9. The 10-meter signal strength comparison of data using Max Function in Fig. 5,6,7

IV. MACHINE LEARNING BASED LEARNING

In this section, we will begin by presenting the Machine Learning models we utilized and elaborate on their training process. Subsequently, we will delve into an evaluation of their performance and conduct a comparative analysis.

We employed three different machine learning models in our analysis: Support Vector Machine (SVM), XGBoost, and Random Forest. For each model, we fine-tuned the parameters to optimize their performance. [15]

1) Support Vector Machine (SVM):
   SVM is a powerful and widely used classification algorithm in Machine Learning. It is particularly effective for binary classification tasks like ours, where we aim to differentiate between drone signals and background noise. SVM works by finding the optimal hyperplane that best separates the two classes, maximizing the margin between them. Its ability to handle high-dimensional data and nonlinear decision boundaries makes it well-suited for signal classification tasks.
2) XGBoost:
   XGBoost, short for Extreme Gradient Boosting, is an ensemble learning method that has gained immense popularity due to its exceptional performance in various classification tasks. It is based on decision trees and operates by iteratively adding weak learners to boost the model's accuracy. XGBoost excels in handling imbalanced datasets, which is crucial for our scenario where the number of positive and negative examples may vary significantly.
3) Random Forest:
   Random Forest is another ensemble learning technique that utilizes multiple decision trees to achieve accurate classification results. By aggregating the predictions of multiple trees, Random Forest can provide robust and stable classifications. Its ability to handle large

datasets and feature importance analysis makes it an excellent choice for signal classification, where different frequency bins may have varying levels of significance.

Each of these models has unique strengths that make them suitable for signal classification tasks. Their combination allows us to explore different approaches and select the most effective model for our specific use case.

A. Experiment Results

Starting with SVM, we achieved an accuracy of 0.56, precision of 0.54, recall of 0.56, and an F1-score of 0.55. The confusion matrix revealed 279 true positives, 224 true negatives, 208 false positives, and 264 false negatives.

Moving on to XGBoost, we obtained an accuracy of 0.53, precision of 0.51, recall of 0.57, and an F1-score of 0.54. The confusion matrix indicated 247 true positives, 256 true negatives, 204 false positives, and 268 false negatives.

For Random Forest, we achieved an accuracy of 0.55, precision of 0.54, recall of 0.56, and an F1-score of 0.55. The confusion matrix showed 273 true positives, 230 true negatives, 206 false positives, and 266 false negatives.

Comparing the three models, we observed that SVM had the highest accuracy, while XGBoost demonstrated slightly higher recall. However, Random Forest showed balanced performance in all metrics. Overall, the models yielded comparable results, with slight variations in their predictive capabilities. TABLE IV is a tabular representation of the evaluation metrics for each model, .

Model	Accuracy	Precision	Recall	F1-score
SVM	0.56	0.54	0.56	0.55
XGBoost	0.53	0.51	0.57	0.54
Random Forest	0.55	0.54	0.56	0.55

TABLE IV
EVALUATION METRICS FOR EACH MODEL

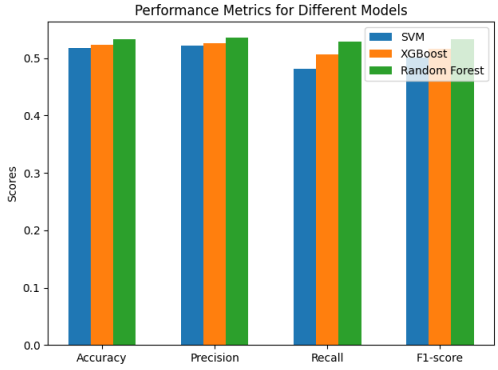

Fig. 10. Comparison of Evaluation metrics for each model

V. ACKNOWLEDGMENT

I would like to extend my heartfelt gratitude and appreciation to my Professor Claudio Enrico Palazzi, for their invaluable support and guidance throughout this project. The encouragement and receptiveness to my ideas played a crucial role in shaping the direction of my survey. His expertise in wireless networks and the comprehensive course they delivered allowed me to gain a profound understanding of the subject matter.

I am especially grateful for the exposure to cutting-edge research topics from around the world through the informative videos shared during the course. This not only expanded my general knowledge but also inspired me to explore new avenues in my academic journey. His dedication to fostering a conducive learning environment has been instrumental in my growth as a student. Thank you for being an exceptional mentor and contributing immensely to my educational experience.

VI. CONCLUSION & FUTURE-WORK

In conclusion, this project, titled ”RF-based data acquisition and detection of drones using Machine Learning,” involved the collection and analysis of a dataset for the 2.40 - 2.49 GHz frequency range, followed by testing drone detection with machine learning models. Throughout the survey, I gained valuable insights into the impact of environmental noise on data acquisition and the importance of achieving a higher Signal-to-Noise Ratio (SNR) for generating positive examples. A higher SNR proved beneficial in enhancing model accuracy, feature extraction capabilities, detection range, and overall resistance to noise and interference.

During the course of this project, I also recognized certain limitations, including hardware constraints and areas where my knowledge could be further expanded. For future endeavors in this domain, it is highly recommended to utilize Software-Defined Radios (SDRs) with wide frequency ranges to enhance data collection capabilities. Additionally, adopting different data acquisition techniques, such as SDR-based active detection, [1] could potentially yield more robust and accurate results.

Moving forward, the proposed future work involves exploring alternative approaches for signal classification and drone detection. Instead of solely relying on traditional machine learning models, a potential advancement could involve employing Generative Adversarial Networks (GANs) to generate fake data from adversaries, such as noise or connected transmitters. This generated data could then be fed into a Recurrent Neural Network (RNN) for signal classification and drone detection. Such a technique, when deployed with HackRF One, has the potential to create a comprehensive Live Drone Detection system, capable of real-time monitoring and accurate identification of drones.

REFERENCES

[1] P. Nguyen, A. Nguyen, and T. Vun, “Investigating Cost-effective RF-based Detection of Drones,” https://home.cs.colorado.edu/ rhan/Papers/p17-nguyen.pdf.

[2] Z. Liu et al, Rise of mini-drones: Applications and issues, In Proceedings of the 2015 Workshop on Privacy-Aware Mobile Computing, ACM MobiHoc, pages 7–12, 2015.

[3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] Amazon. Amazon prime air. http://www.amazon.com/b?node=8037720011.
[5] J. Serna. Lufthansa jet and drone nearly collide near lax. LA Times, March 19, 2016 in press.
[6] P. McGreevy. Private drones are putting firefighters in 'immediate danger,' california fire official says. LA Times, August 18, 2015.
[7] D. Waldstein. Drone crash interrupts match. New York Times, September 3, 2015
[8] B. Jansen. Drone crash at white house reveals security risks. USA Today, January 26, 2015.
[9] A. Morrow. Couple accuses neighbor of stalking with drone. USA Today, December 13, 2014.
[10] H. Liu, Z. Wei, Y. Chen, J. Pan, L. Lin and Y. Ren "Drone detection based on an audi-assisted camera array."
[11] X. Yang, K. Huo, W. Jiang, J. Zhao and Z. Qiu, "Drone detection based on an audi-assisted camera array."
[12] P Andrasi, T Radisic, M Mustra and J Ivosevic, "Night-time detection of UAVs using thermal infrared camera."
[13] S. Kunze, A. Weinberger and R. Poeschl, "Night-time detection of UAVs using thermal infrared camera."
[14] M. D. L. Angrisani et al., "Clustering-based method for detecting and evaluating I/Q impairments in radio-frequency digital transmitters," IEEE Transactions on Instrumentation and Measurement, vol. 56, no. 6, pp. 2139–2146, 2007.
[15] K. Youssef et al. "Machine Learning Approach to RF Transmitter Identification", IEEE RFID, 2018
[16] S. Riyaz et al. "Deep Learning Convolutional Neural Networks for Radio Identification", IEEE Com. Mag., 2018.