MSG-SET-183 Specialists' Meeting on "Drone Detectability: Modelling the Relevant Signature"

Machine Learning Empowered Radio Frequency Signal Classification For UAS Detection

Michael Nilsen

Research Assistant, Virginia Modeling, Analysis and Simulation Center, Old Dominion University

DSP Engineer, Johns Hopkins University Applied Physics Lab, Laurel, MD US

Sachin Shetty

Associate Director - Virginia Modeling, Analysis and Simulation Center,

Associate Professor, Department of Computational Modeling and Simulation Engineering

Old Dominion University, Norfolk, VA USA

Kimberly Gold

Engineer

Naval Surface Warfare Center

Crane, Indiana USA

Charles Kamhoua

Senior Electronics Engineer

Network Sciences Division

Army Research Lab

Adelphi, MD USA

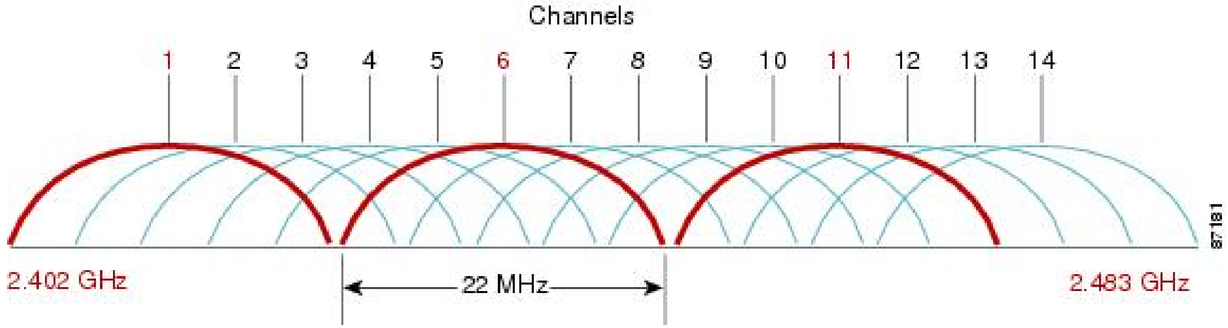
Overview

- **Project Goal**
  - Develop an automated RF classification capability in GNU Radio that can automatically monitor, detect and classify UAS signals across multiple bands
- **Benefits**
  - Detect and classify UAS signals to create a library of known UAS signals.
  - Craft a waveform and launch a protocol attack on the adversarial UAS signal.
  - Signal energy detection, automatic modulation recognition and classification will be useful for C-UAS efforts.

UAS Signal Detection

Bands

- **Band 1: 433 MHz ISM Band: 433.05 - 434.7**
- **Band 2: 2.4 GHz – WiFi g/b/n: ≈ 2.4 – 2.5 GHz**
- **Band 3: 5.8 GHz - WiFi a/h/j/n**

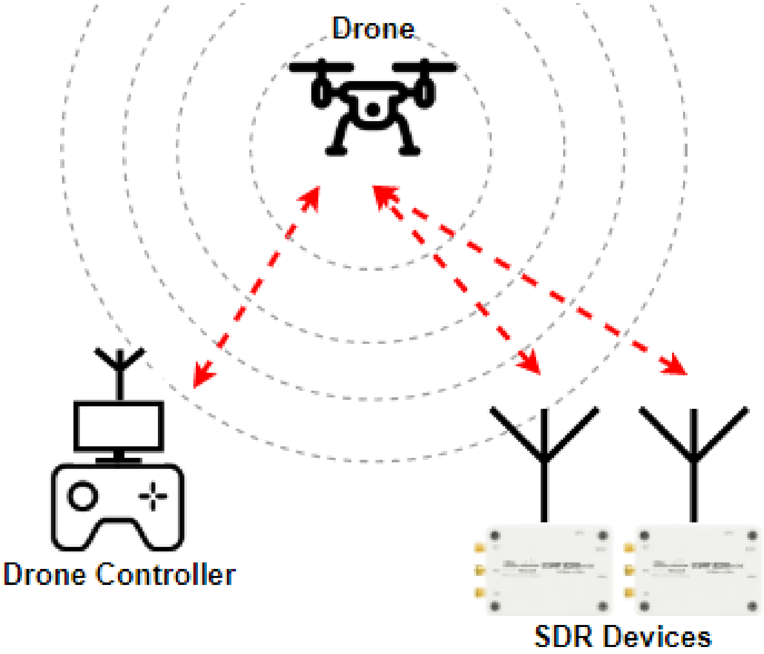


Signal

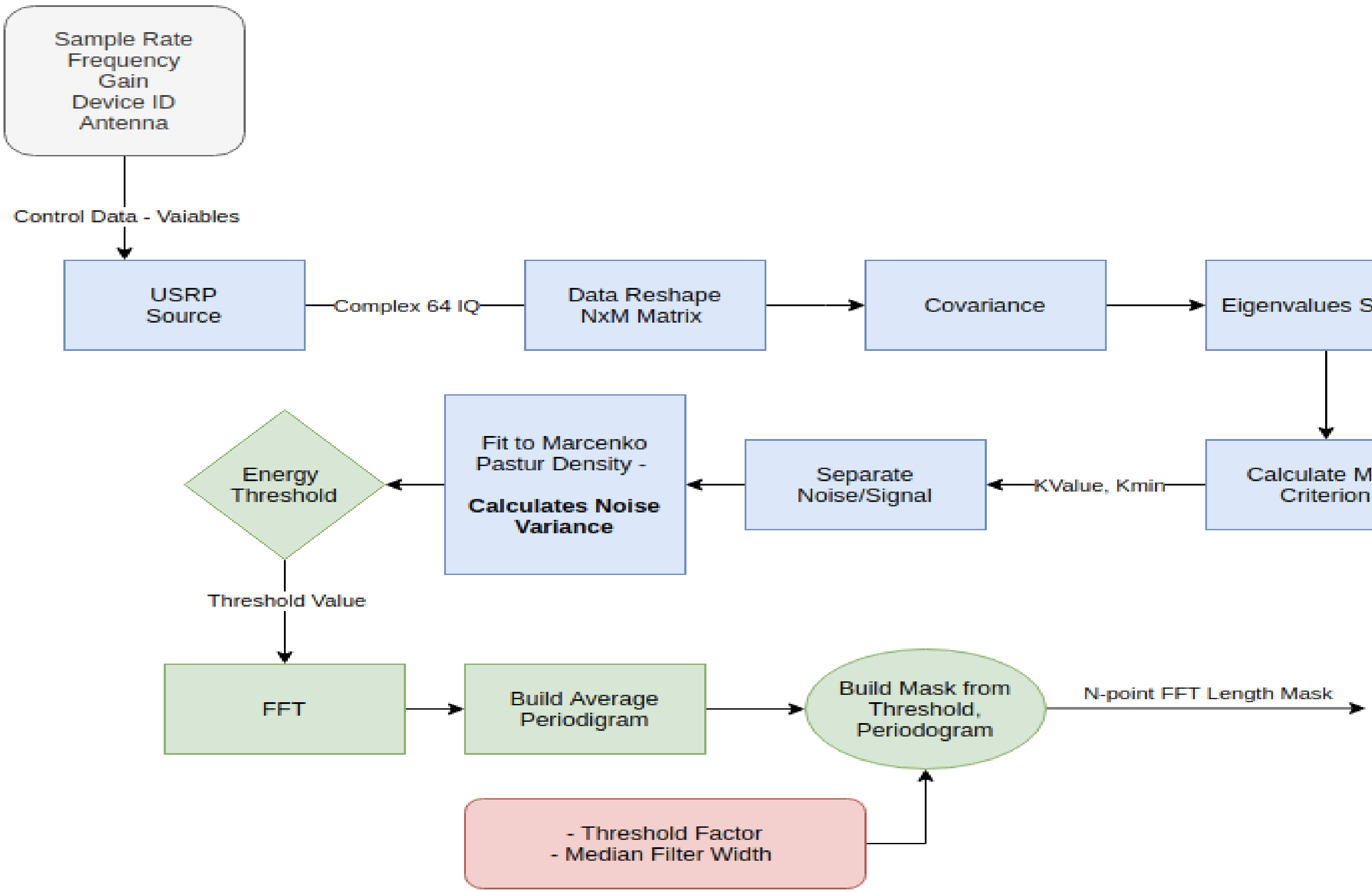
- **OFDM (orthogonal frequency division multiplexing): Drone specific transmit frequency channel will be selected**

Background

- **Radio Frequency Detection**
  - **Adaptive Signal Energy Detection**
  - **Modulation Scheme Recognition**
  - **Machine Learning Model**
  - **Multi-UAS Classification Approach**
    - **Frequency Band Allocation**
    - **ML Feature Data Processing**
    - **Signal Energy Transient Analysis**



Adaptive Energy Detection



Data Stream to Eigenvalues

Complex64
float32 I, float32 Q

L=16384

81ms Burst Capture

$$\rightarrow y = \begin{pmatrix} y_{1,1} & \cdots & y_{1,N} \\ \vdots & \ddots & \vdots \\ y_{L,1} & \cdots & y_{L,N} \end{pmatrix} \longrightarrow \begin{array}{c} \mathcal{H}_0 : y_i(n) = w_i(n) \\ \mathcal{H}_1 : y_i(n) = x_i(n) + w_i(n) \end{array} \longrightarrow \begin{array}{c} \Sigma_W = E\{w(n)w^H(n)\} = \sigma_W^2 . I_L \; ; \; -\infty < n < +\infty \\ \Sigma_x = E\{x(n)x^H(n)\} \\ \Sigma_y = E\{y(n)y^H(n)\} \end{array}$$

$$\Sigma_y = \Sigma_x + \Sigma_W = \Sigma_x + \sigma_W^2 . I_L \longrightarrow \begin{array}{l} \lambda_{yi} = \lambda_{yi} + \sigma_W^2 \,, \forall i = 1,2 \ldots K \\ \lambda_{yi} = \sigma_W^2 \,, \forall i = K+1, K+2, \ldots L \end{array} \longrightarrow \hat{\Sigma}_y = \frac{1}{N} y y^H$$

- **Complex IQ Stream is converted to a sample covariance matrix in order to calculate eigenvalues**
  - **Signal noise (captured) is independent of signal**
  - **Eigenvalues are used to determine power threshold of noise/signal**

MDL Criterion

$$K = \arg\min_{K} \left(-(L-K)N \log \frac{\varphi(K)}{\theta(K)} + \frac{1}{2}K(2L - K) \log N \right); 0 \le K \le L-1$$

Where $\varphi(K)$ and $\theta(K)$ are given by:

$$\varphi(K) = \prod_{i=K+1}^{L} \lambda_i^{\frac{1}{L-K}}$$

$$\theta(K) = \frac{1}{L-K} \sum_{i=M+1}^{L} \lambda_i$$

$$\sigma_{W1}^2 = \frac{\lambda_L}{(1-\sqrt{p})^2}$$

$$\sigma_{W2}^2 = \frac{\lambda_{\hat{K}+1}}{(1+p)^2}$$

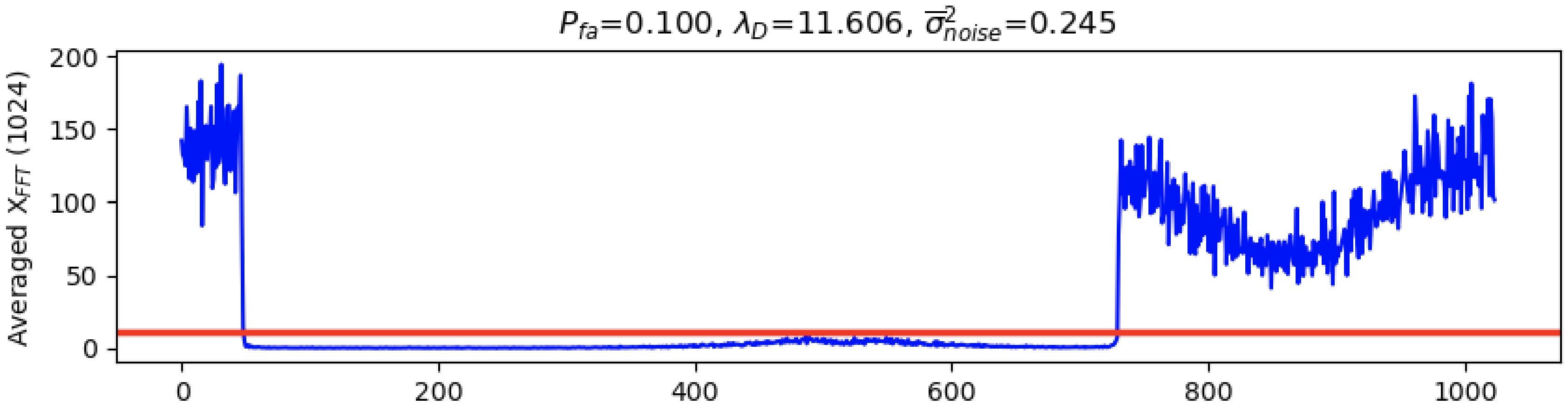
- **MDL (Minimum Description Length) is used to determine K, index of eigenvalues separating noise and signal**
  - **This function returns both index and noise variance**

Marcenko Pasture Density

$$mp(p, \sigma_w) = dF^W(Z)$$
$$= \frac{\sqrt{(z - \sigma_W^2(1-\sqrt{p})^2)(\sigma_W^2(1+\sqrt{p})^2 - z)}}{2\pi\sigma_W^2 zp} dz \longrightarrow \lambda_D = \widehat{\sigma_W^2}\left(Q^{-1}(P_{fa})\sqrt{2N} + N\right)$$

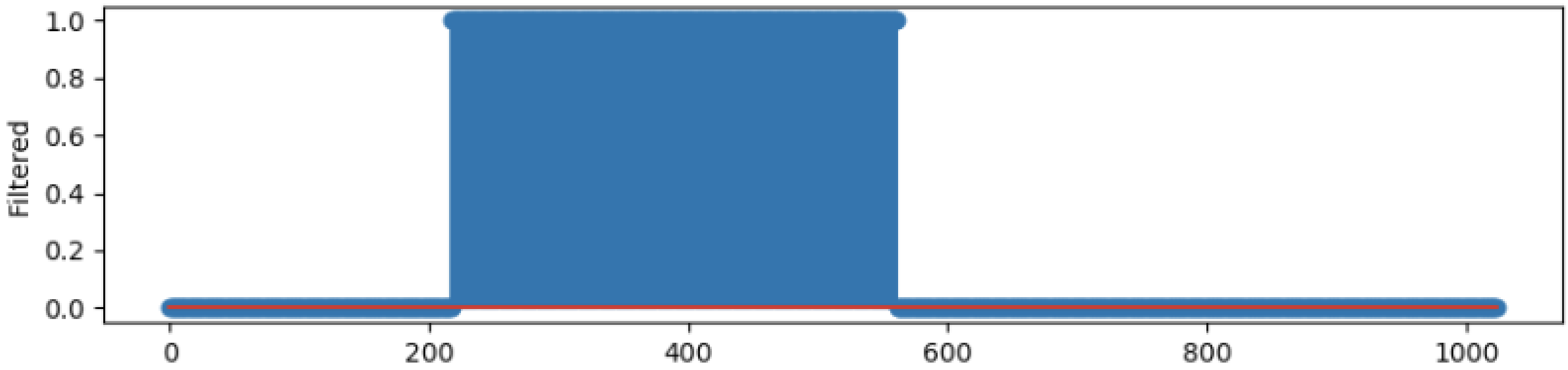
Where

$$\sigma_W^2(1-\sqrt{p})^2 \le z \le \sigma_W^2(1+\sqrt{p})^2$$

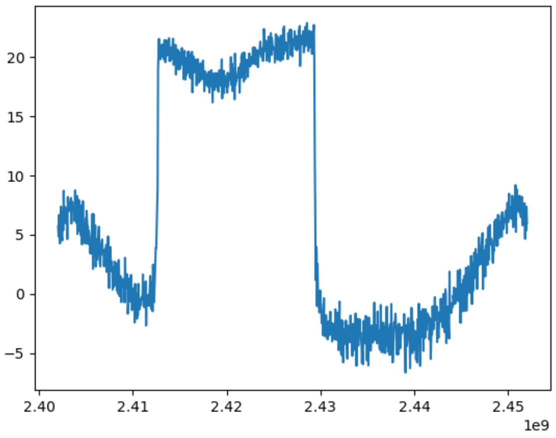


- **Returns noise/signal threshold used to build mask of periodogram**

Masking Function



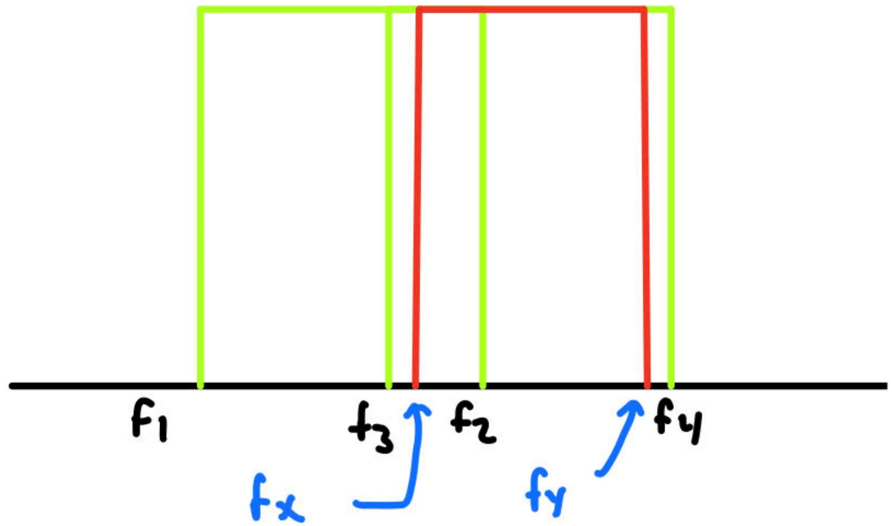
- **Mask is built with FFT bins where the value is greater than calculated threshold value**
  - **Returns a binary array of length L = NFFT**
  - **Mask span is converted to frequency values and interpolated based on a priori knowledge**



Channel Overlap

- Goal is to provide the distributed USRPs with information on which channel to tune to
- Function is needed to convert binary mask to start frequency, stop frequency, and channel span
- Masking calculation occurs with some error due to wide band signal captures from UAS devices
- Channel overlap function calculate the % overlap with an default channel on the observed band (2.4/5.8GHz)

Channel Overlap



channel $x = [f_1, f_2]$

channel $y = [f_3, f_4]$

unknown channel $u = [f_x, f_y]$

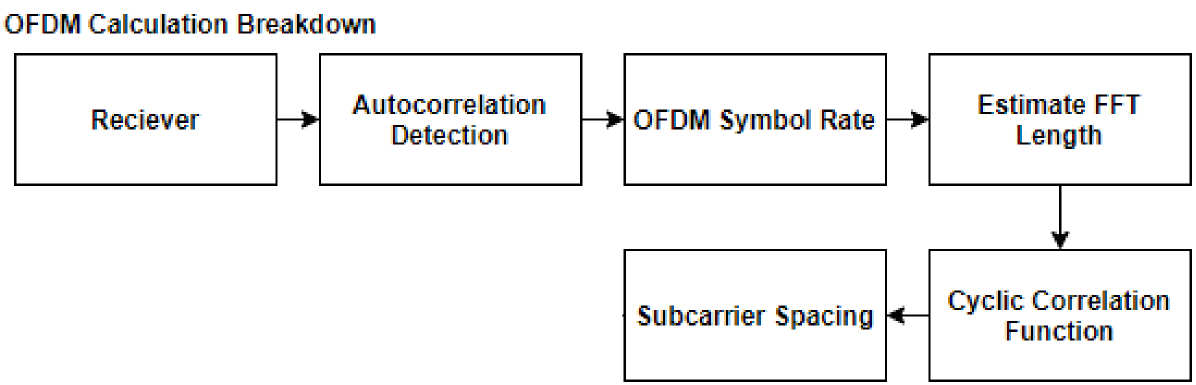
Feature Extraction

OFDM Parameters

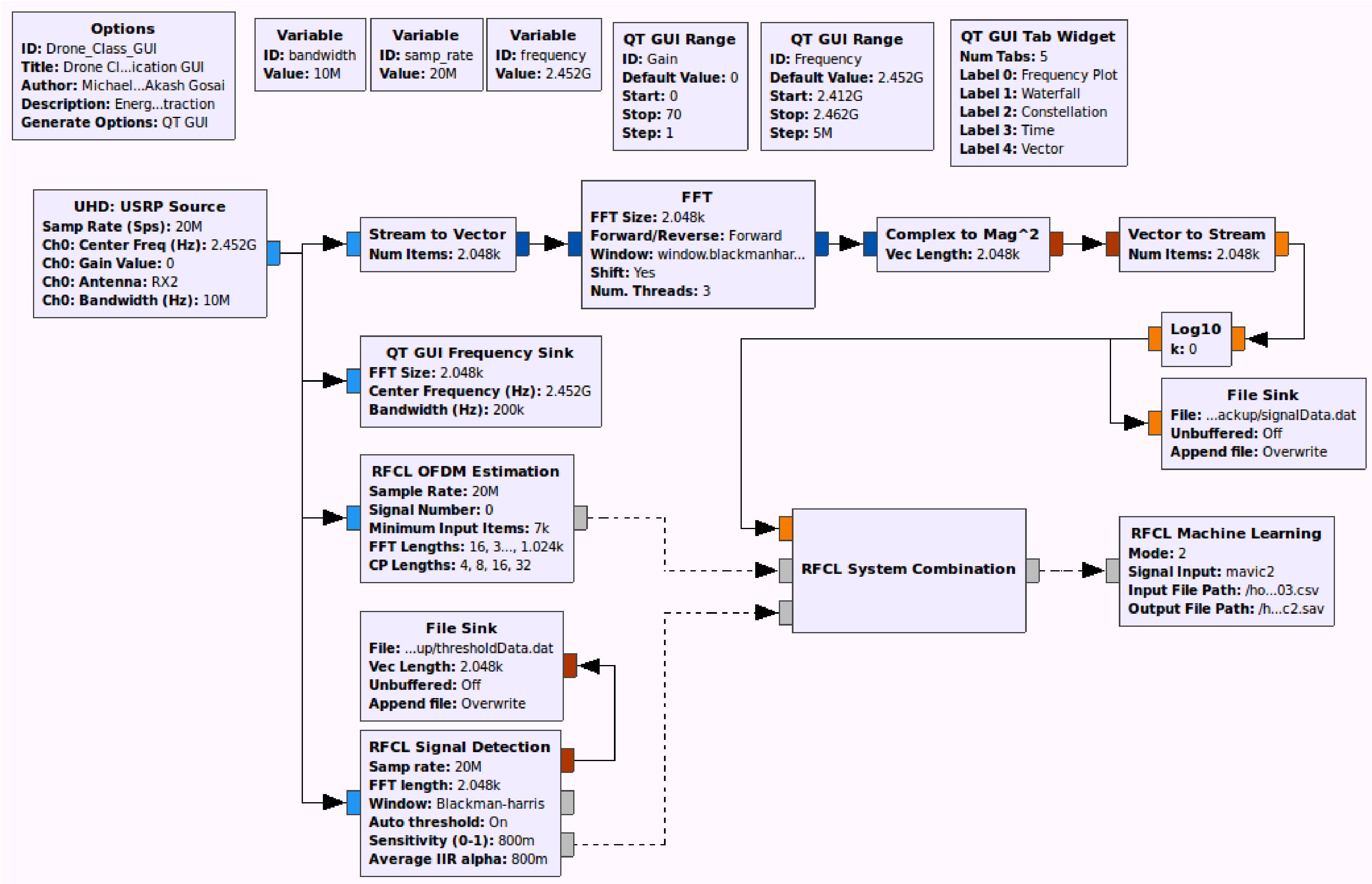
- **Cyclic Prefix Length :** Circular extension added to OFDM symbols in order to eliminate inter symbol interference
- **FFT Size:** Fast Fourier transformation of the digital signal to frequency domain using different bin sizes 64, 128, 256, 512, 1024…
- **Subcarrier Spacing:** spacing of the subcarriers of the OFDM symbol = 1/Ts, Ts (Symbol Time)
- **OFDM Symbol Duration:** Length of symbol time

Modulation Scheme Recognition

- **Blind OFDM can be detected by calculating the autocorrelation of the cyclic prefix of the received signal**
- **FFT lengths are estimated to determine the number of signal subcarriers**
- **The breakdown below outlines each step in the recognition process.**

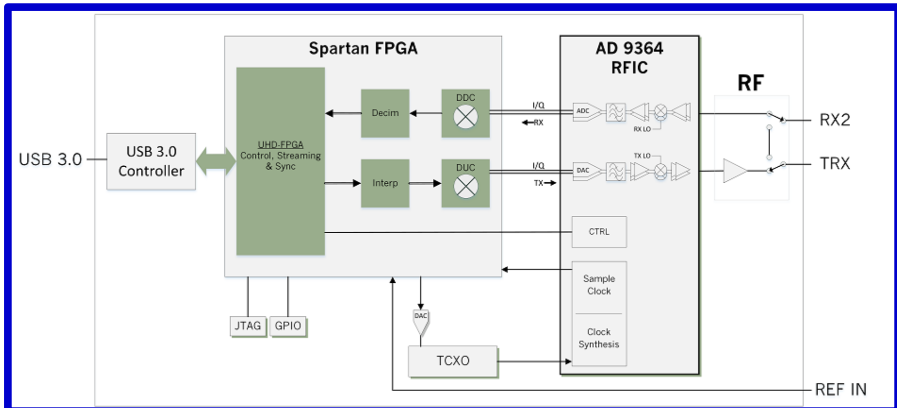


RF Classification Library



SDR Implementation

Ettus B205-1 Block Diagram



Complex64 IQ

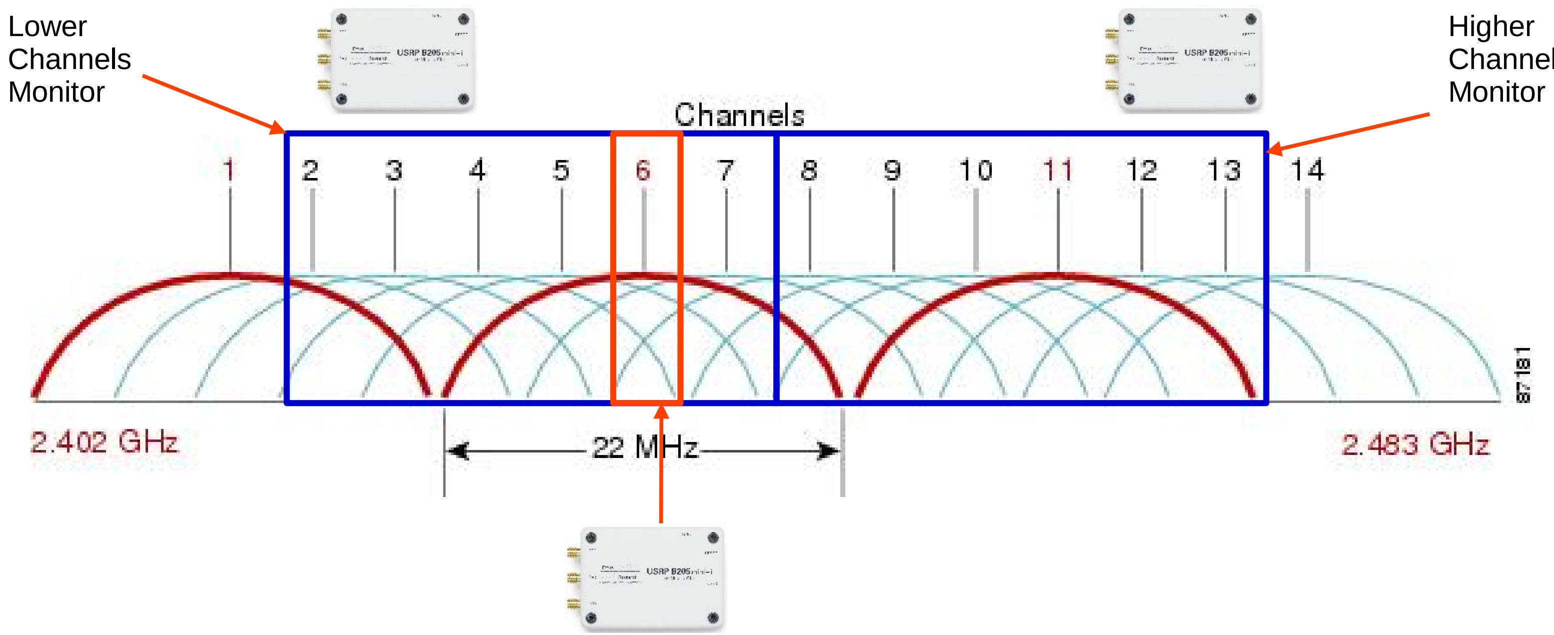
Control

Host PC

GNU Radio

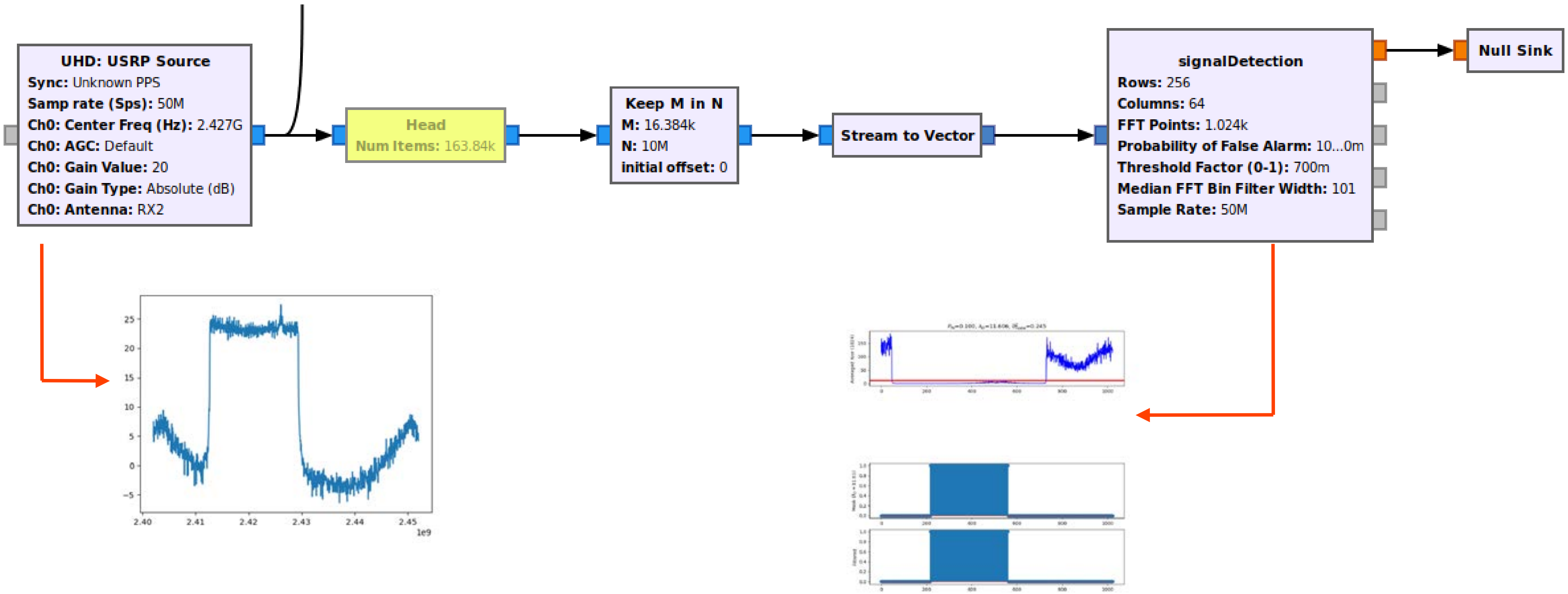
- Signal Processing
- Energy Detection
- OFDM Synchronization
- OFDM Demodulation
- Offline Data Collection
- Machine Learning Model

SDR Implementation



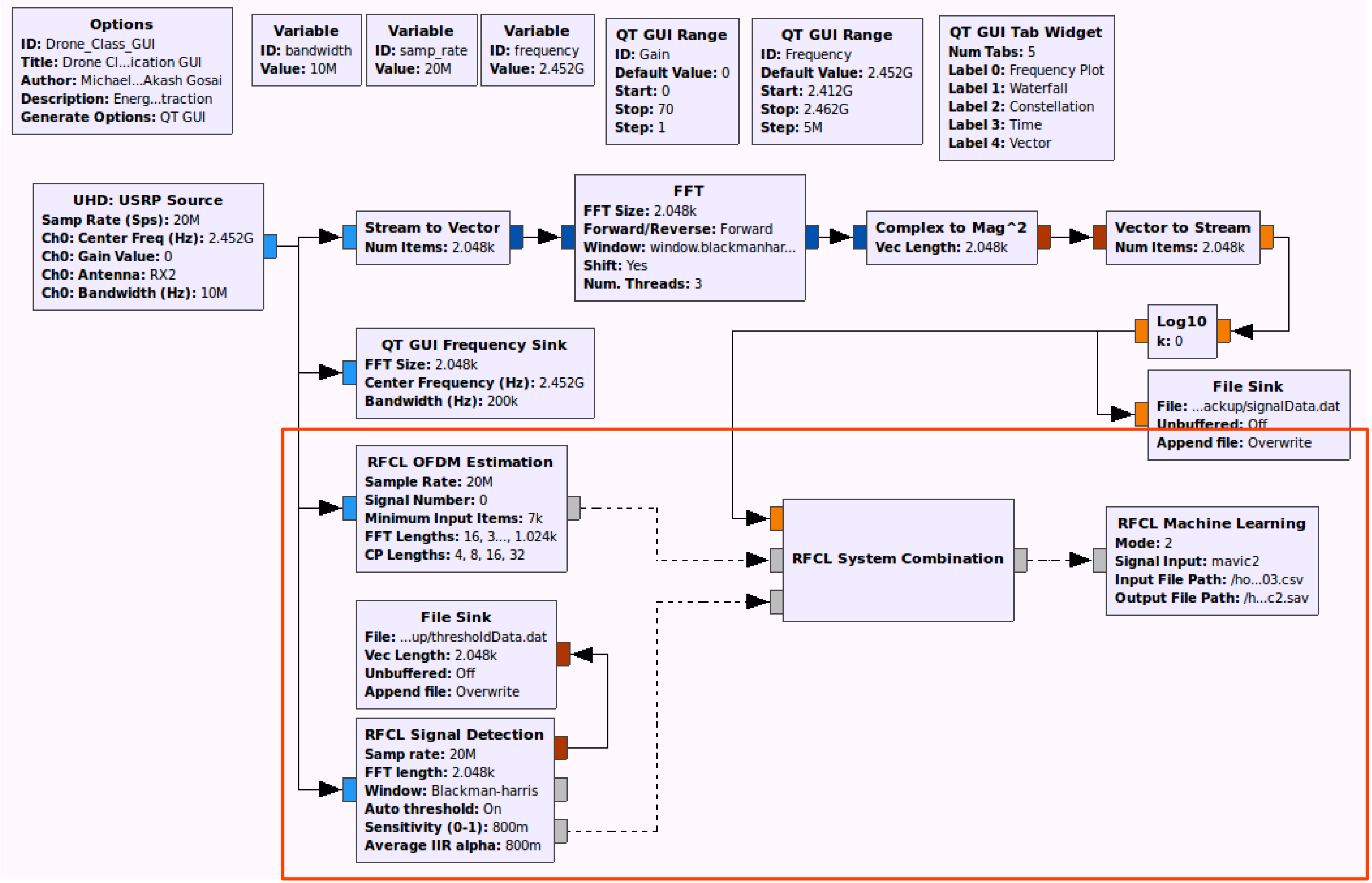
- Multiple B205 SDR split up the spectrum available by continuously monitoring and running an algorithm to find channels of potential interest
- A third B205 SDR (or B210) is then able to tune to a specific channel and determine the second phase drone identification algorithm. But each device can potentially split is 61.4 bandwidth into multiple streams.

Phase 1 GNU Radio



- Phase 1 algorithm is responsible for determining a channel that is occupied or potentially occupied with a drone. Phase 2 is then responsible for determining if the channel is occupied and then to classify the drone on the channel.
- Replaces current energy detection algorithm

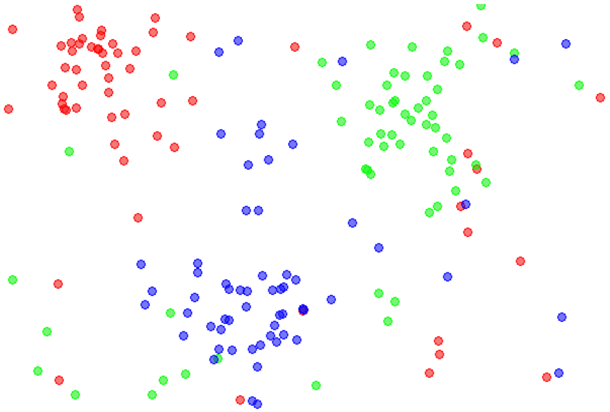
Phase 2



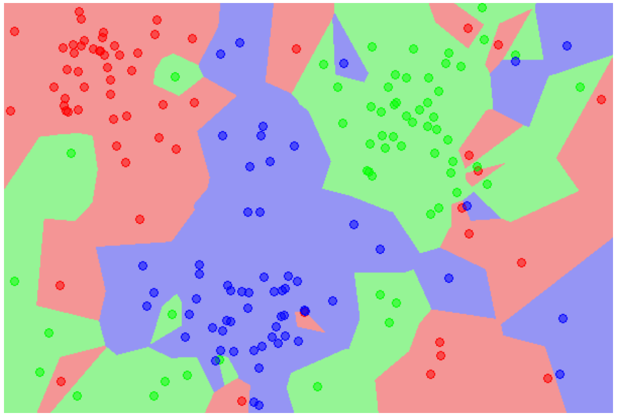
Options
ID: Drone_Class_GUI
Title: Drone Cl...ication GUI
Author: Michael...Akash Gosai
Description: Energ...traction
Generate Options: QT GUI
Variable
ID: bandwidth
Value: 10M
Variable
ID: samp_rate
Value: 20M
Variable
ID: frequency
Value: 2.452G
QT GUI Range
ID: Gain
Default Value: 0
Start: 0
Stop: 70
Step: 1
QT GUI Range
ID: Frequency
Default Value: 2.452G
Start: 2.412G
Stop: 2.462G
Step: 5M
QT GUI Tab Widget
Num Tabs: 5
Label 0: Frequency Plot
Label 1: Waterfall
Label 2: Constellation
Label 3: Time
Label 4: Vector
UHD: USRP Source
Samp Rate (Sps): 20M
Ch0: Center Freq (Hz): 2.452G
Ch0: Gain Value: 0
Ch0: Antenna: RX2
Ch0: Bandwidth (Hz): 10M
Stream to Vector
Num Items: 2.048k
FFT
FFT Size: 2.048k
Forward/Reverse: Forward
Window: window.blackmanhar...
Shift: Yes
Num. Threads: 3
Complex to Mag^2
Vec Length: 2.048k
Vector to Stream
Num Items: 2.048k
Log10
k: 0
QT GUI Frequency Sink
FFT Size: 2.048k
Center Frequency (Hz): 2.452G
Bandwidth (Hz): 200k
File Sink
File: ...ackup/signalData.dat
Unbuffered: Off
Append file: Overwrite
RFCL OFDM Estimation
Sample Rate: 20M
Signal Number: 0
Minimum Input Items: 7k
FFT Lengths: 16, 3..., 1.024k
CP Lengths: 4, 8, 16, 32
RFCL System Combination
RFCL Machine Learning
Mode: 2
Signal Input: mavic2
Input File Path: /ho...03.csv
Output File Path: /h...c2.sav
File Sink
File: ...up/thresholdData.dat
Vec Length: 2.048k
Unbuffered: Off
Append file: Overwrite
RFCL Signal Detection
Samp rate: 20M
FFT length: 2.048k
Window: Blackman-harris
Auto threshold: On
Sensitivity (0-1): 800m
Average IIR alpha: 800m

K-Nearest Neighbors Algorithm

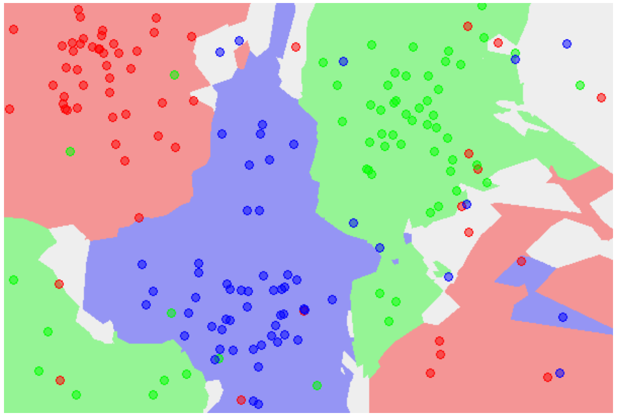
- In KNN, the input consists of the k closest training examples in the feature space.
- Parameter selection and data collection are the two time consuming parts of the machine learning process



Dataset [4]



1NN Classification Map [4]

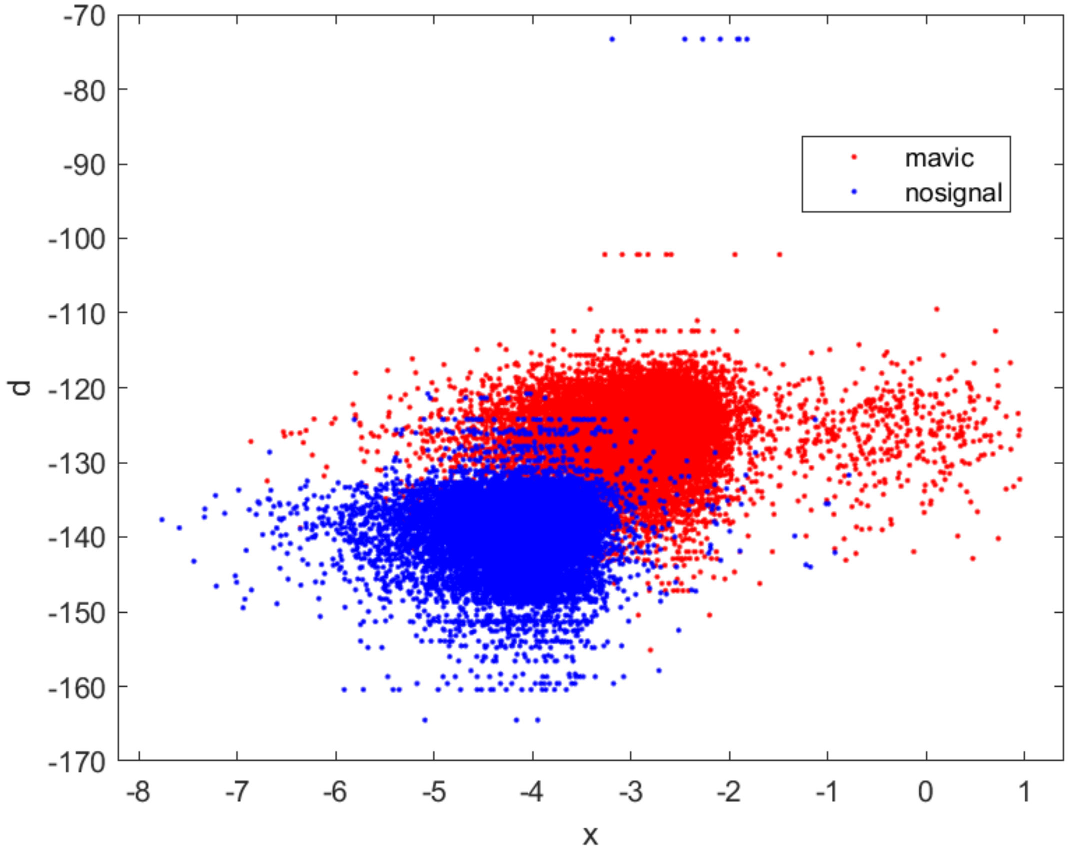


5NN Classification Map [4]

Source: Wikipedia contributors. (2021, February 21). K-nearest neighbors algorithm. In Wikipedia, The Free Encyclopedia.

Machine Learning Model

8NN Classification – Distance: 0m & 5m

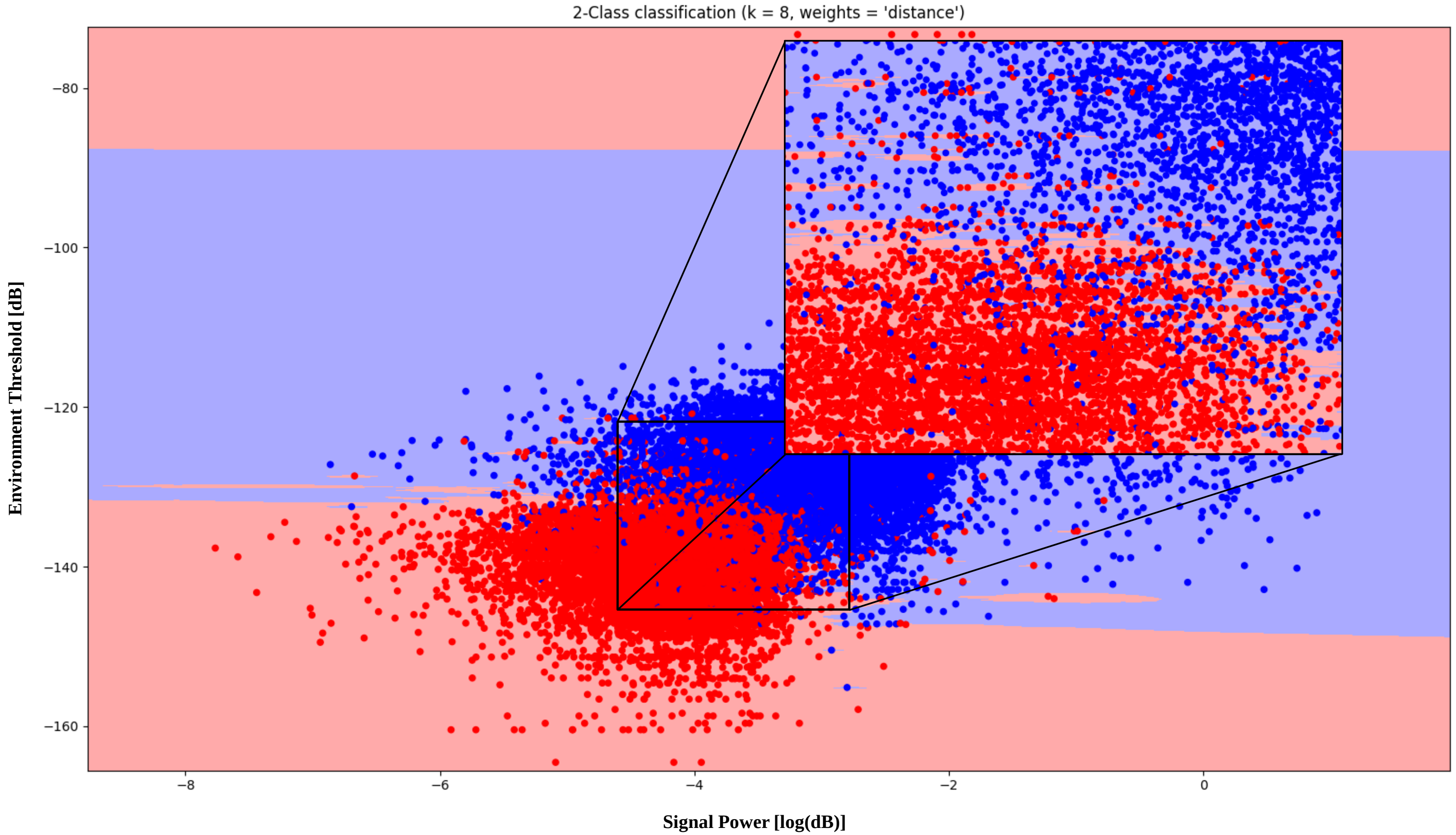


Environment Threshold [dB]

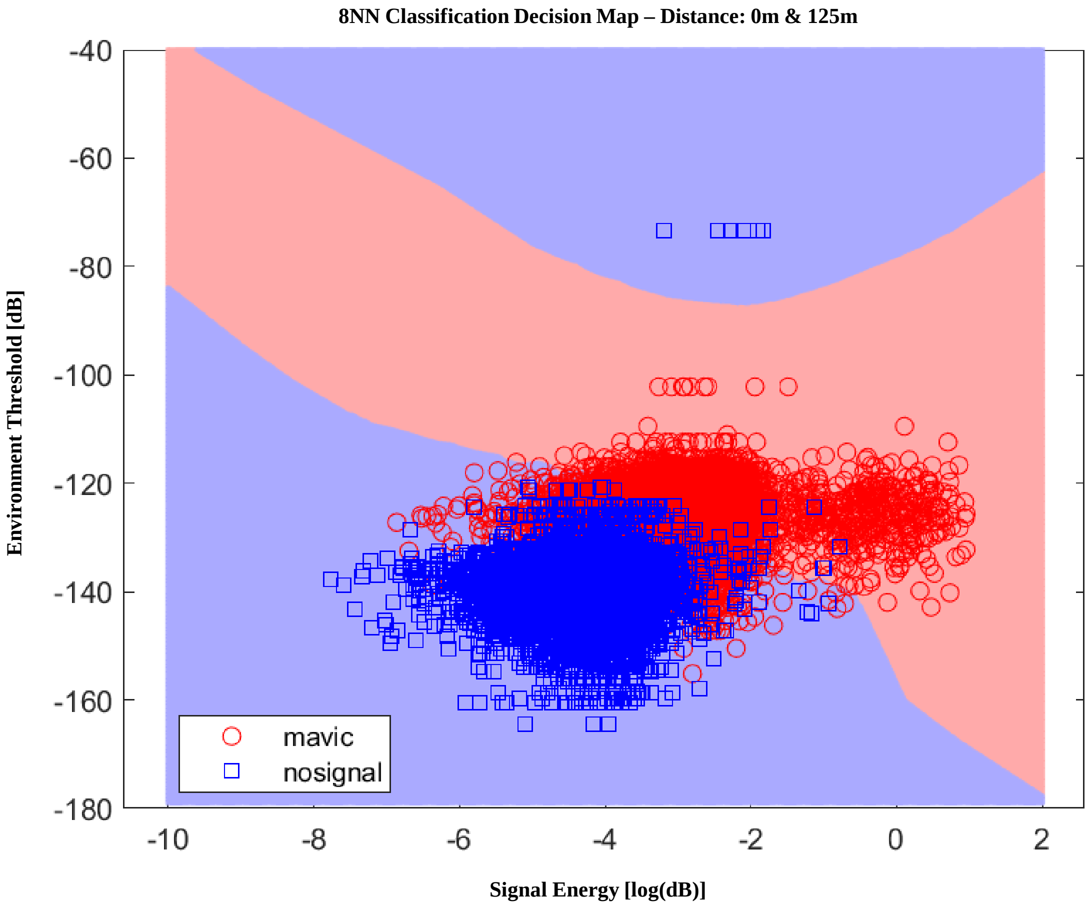
Signal Power [log(dB)]

Machine Learning Model

8NN Classification – Distance: 0m & 125m

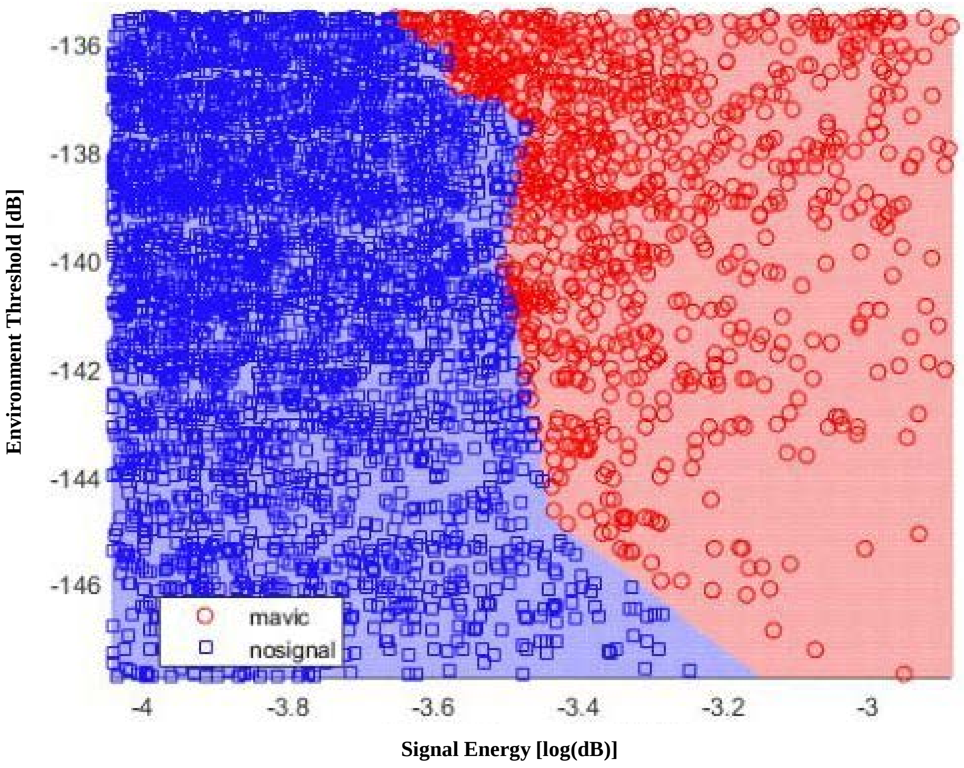


Machine Learning Model



Machine Learning Model

CNN Classification Decision Map (Zoomed) – Distance: 0m to 500m (all)



Testing and Evaluation

- Lab Testing
- Site 1 Testing

Experimental Setup

S/N	Hardware	Software
1	Dell Laptop (i5, 2.3 GHz, 16 Gb RAM)	GNU Radio Companion
2	Ettus Research B210 USRP SDR	
3	2.4/5.8 GHz, 9 dBi Directional Antenna	
4	DJI Mavic 2	
5	DJI Phantom 4 V2	

Lab Testing

- **Data Acquisition** – Capture RF data from drones operating in indoor and outdoor settings for machine learning training
- **Model Building** - Train machine learning model on one drone class
- **Evaluation** - Assess performance based on detection rate.

Lab Test Setup

- **Clear Line of Sight**
  - Tested link conditions of various drone distances
- **Shadowing/Fading**
  - Physical environment obstructions with some line of sight knowledge
- **No Line of Sight**
  - Behind building and out of line of sight

Lab Results (Mavic 2)

Number of Samples = 5000 (Random 20% of 25k)					
Distance (m)	Detection Rate (%) (True Positive + True Negative)/5000	True Positive	False Positive	True Negative	False Negative
5	99.98	2519	0	2480	1
25	99.84	2514	5	2478	3
50	99.34	2501	18	2466	15
100	99.46	2502	17	2471	20
125	98.82	2483	36	2458	23
150	95.34	2311	124	2456	109

Number of Samples = 5000 (Random 20% of 25k)					
Distance (m)	Detection Rate (%)	True Positive	False Positive	True Negative	False Negative
Shadowing/Fading	88.57	2134	243	2294	329

Number of Samples = 5000 (Random 20% of 25k)					
Distance (m)	Detection Rate (%)	True Positive	False Positive	True Negative	False Negative
Unkown Location	81.22	1987	563	2074	376

Site 1 Purpose

- **Data Acquisition** - Capture real data in unique environment
- **Model Building** - Developing models for each drone family within environment
- **Model Comparison** – Assess isolated model performance vs. lab created model. Analyze impact merging dataset into one model.
- **Detection Range** – Test maximum detection range of system

Site 1 Test Setup

- **Clear Line of Sight**
  - Tested link conditions of various drone distances
- **Shadowing/Fading**
  - Physical environment obstructions with some line of sight knowledge

Site 1 Results (Phantom 4 V2)

Distance (m)	Detection Rate (%)	True Positive	False Positive	True Negative	False Negative
100	99.98	10040	0	9956	4
200	99.34	10156	43	9712	89
500	98.77	10012	111	9745	132
600	93.14	9842	435	9456	267
700	91.88	8859	740	9767	632
800	87.20	9069	1254	8371	1306
900	81.09	8271	1967	7947	1815
1000	74.67	7915	2685	7019	2381
Number of Samples = 20000 (Random 20% of 100k)					
Distance (m)	Detection Rate (%)	True Positive	False Positive	True Negative	False Negative
Shadowing/Fading	64.58	6845	3755	6071	3329

- Detection Rate = 1-(FalsePositve (Class1) + FalseNegative (Class2))*100
- This calculation is the accurate detection rate of identifying both class 1 and class 2. Both true cases listed above

Site 1 Results (Phantom 4 V2) Combined Model

Distance (m)	Detection Rate (%)	True Positive	False Positive	True Negative	False Negative
100	99.98	10398	2	9598	2
200	99.56	10155	46	9757	42
500	99.12	10507	93	9317	83
600	96.34	10212	388	9056	344
700	93.04	9676	682	8932	710
800	90.33	9394	948	8672	986
900	86.29	8802	1426	8456	1316
1000	83.44	8845	1755	7843	1557
Number of Samples = 20000 (Random 20% of 100k)					
Distance (m)	Detection Rate (%)	True Positive	False Positive	True Negative	False Negative
Shadowing/Fading	75.88	2557	7133	2267	8043

- Detection Rate = 1-(FalsePositve (Class1) + FalseNegative (Class2))*100
- This calculation is the accurate detection rate of identifying both class 1 and class 2. Both true cases listed above

Analysis

- **Approach extends detection range to roughly 1Km**
- **Largest training data set size with multiple model performance metrics**
- **Combined model resulted in ~3.7% increase in performance vs. isolated model bringing average accuracy over 90%**

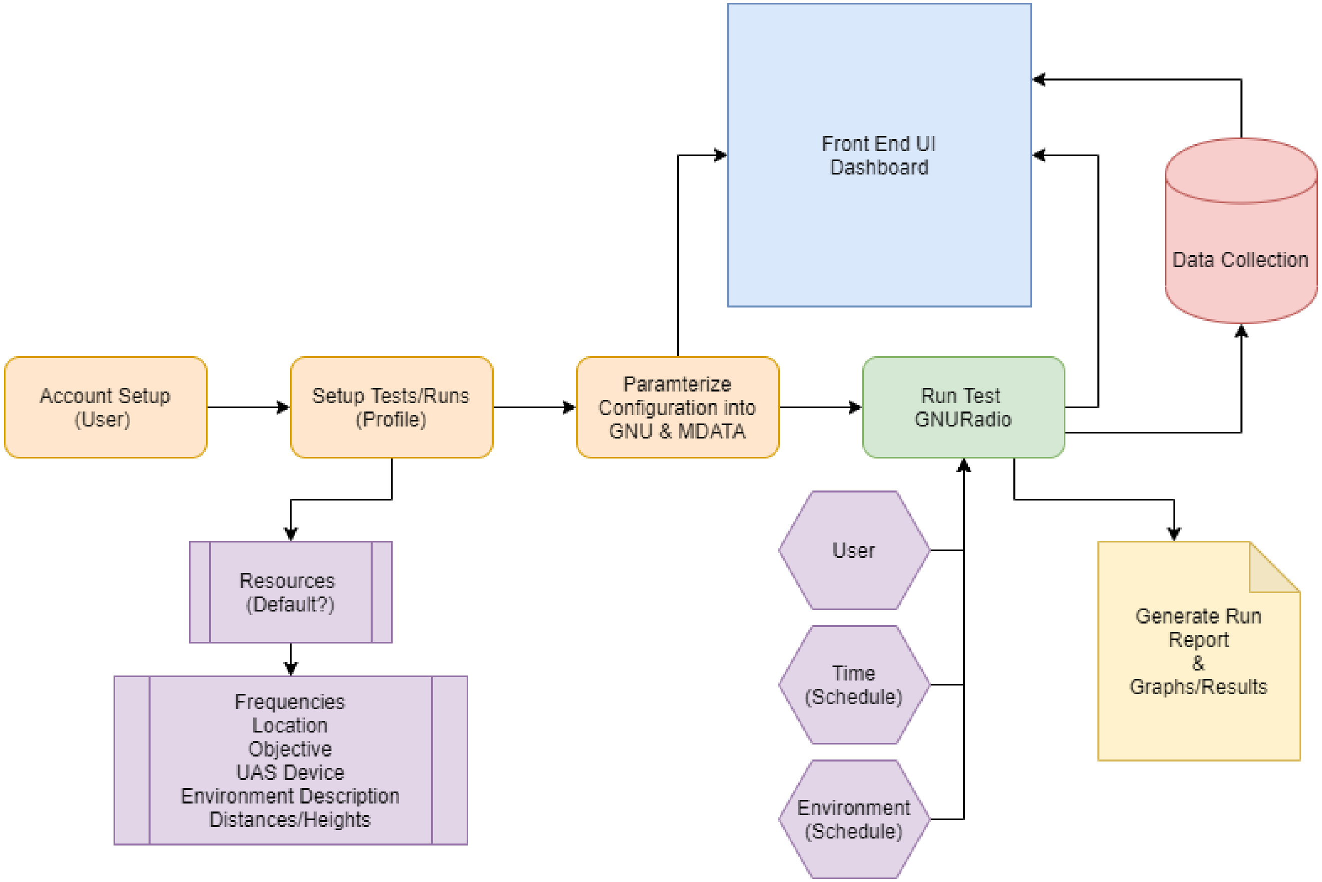
ML Parameter Data Processing

- Processing of seven current ML (Machine Learning) Model parameters specifically looking for drone class specific correlations
- Addition of environmental thresholding max value to capture full signal amplitude of drone
- Include spectral correlation function output in ML model feature input
- Expand classification model to include multi-drone classification steps

System Portal/Workflow

- User – Engineers or Testers responsible for conducting on site experiments
- Capabilities (under development)
  - Provide experiment profile
  - Customize experiment profile specific to system
  - Manage experiment execution
  - Manage Results and Data Files
  - Reuse Models
  - Schedule experiments (Batch)

System Portal/Workflow

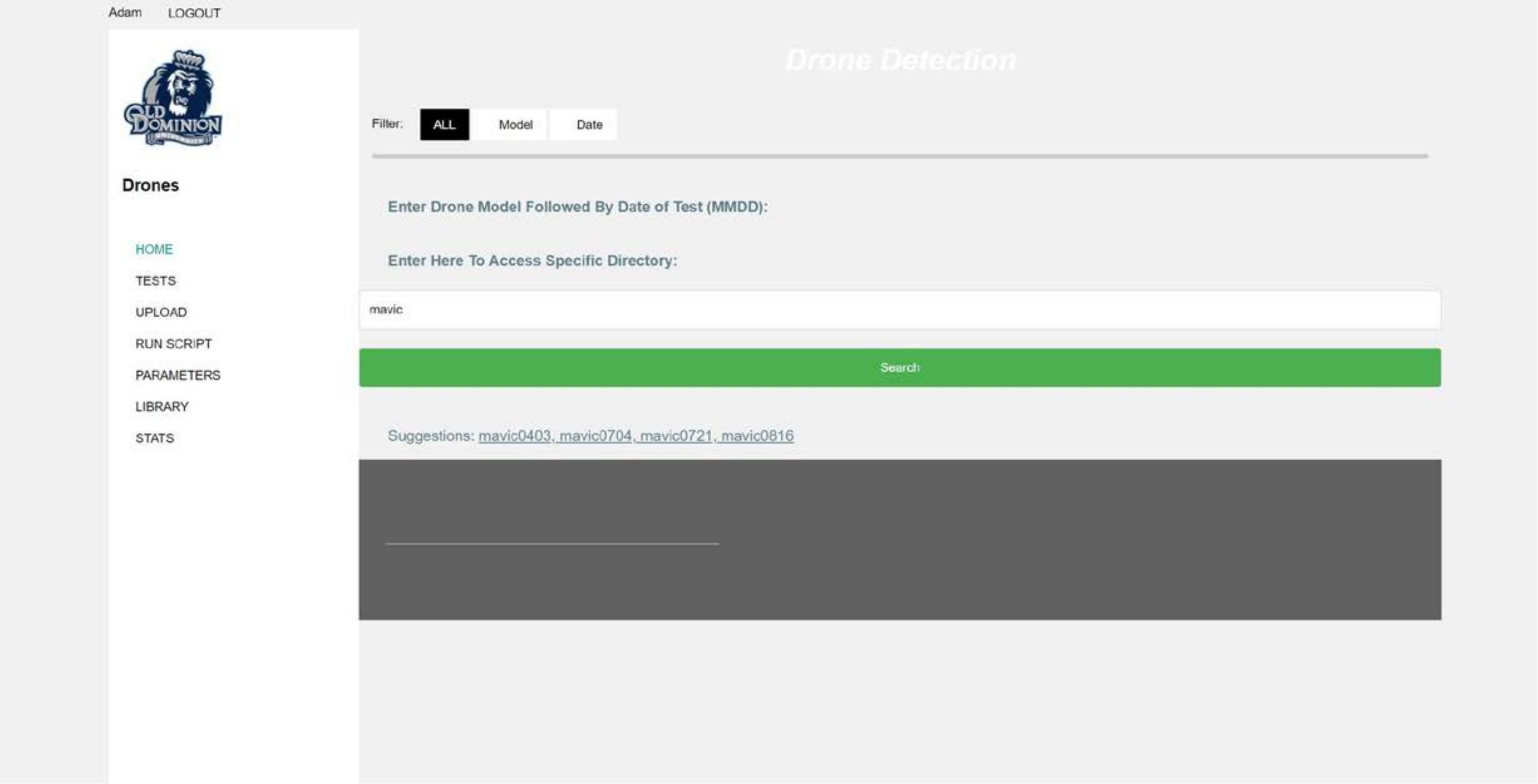


System (Back End) & Front End

- Include all data from trials with the ability to sort by different included system meta data
- Cluster of trials used in certain model creation (ex. trials 1-10 of 50 used for model creation)
- Ability to start model creation or visualization from front end (GUI)
- Automatic data generation, processing, and analysis
- Backend data storage and signal processing to be shown in the front end software

Front End Development

- **Example webpage front end in that is being currently developed**



Conclusion

- Provided a ML based RF classification platform for UAS detection
- Provides an agnostic approach to detecting and identifying presence of UAS in several SNR regimes.

- Aid operators in eliminating non-UAS signals in physical space and focus their detection in physical spaces where UAS signals have been detected.

- Improve the efficiency of detection as prior knowledge of RF fingerprints will aid in ensuring every detection will only focus on unknown RF signal detection.
- Help operators to speed up the detection and identification of UAS devices.

References

[1] M. Hamid, N. Bjorsell, and S. Ben Slimane, “Sample covariance matrix eigenvalues based blind SNR estimation,” IEEE International Instrumentation and Measurement Technology Conference Proceedings, 2014, pp. 718–722.

[2] J. Zhu, Z. Xu, F. Wang, B. Huang, and B. Zhang, “Double Threshold Energy Detection of Cooperative Spectrum Sensing in Cognitive Radio,” International Conference on Cognitive Radio Oriented Wireless Networks and Communications, 2008, pp. 1–5

[3] A. Quadri, M. R. Manesh, and N. Kaabouch, "Performance Comparison of Evolutionary Algorithms for Noise Cancellation in Cognitive Radio Systems," The IEEE Annual Computing and Communication Workshop and Conference,pp. 1-6, 2017.

[4] Wikipedia contributors. (2021, February 21). K-nearest neighbors algorithm. In Wikipedia, The Free Encyclopedia.