Acoustic Detection of Drones through Real-time Audio Attribute Prediction

Sayan Mandal*, Lei Chen†, Vishwa Alaparthy‡, and Mary Cummings§
Humans and Autonomy Lab, 304 Research Drive, Durham, NC, 27708, USA

With the rise in popularity of drones, their use in anti-social activities has also proliferated. Nationwide police increasingly report the appearance of drones in unauthorized settings such as public gatherings and also in the delivery of contraband to prisons. Detection and classification of drones in such environments is very challenging from both visual and acoustic perspective. Visual detection of drones is challenging due to their small size. There may be cases where views are obstructed, lighting conditions are poor, the field of view is narrow, etc. In contrast, acoustic-based detection methods are omnidirectional, however, they are prone to errors due to possible noise in the signal. This paper presents a method of predicting the presence (detection and classification) of a drone using a single microphone and other inexpensive computational devices. A Support Vector Machine classified the spectral and temporal features of pre-segments generated using a sliding window for the audio signal. Additionally, spectral subtraction was used to reconstruct the magnitude spectrum of drone sounds to reduce false alarms. To increase the accuracy of predictions, an added confidence script is proposed based on a queue-and-dump approach to make the system more robust. The proposed system was tested in real time in a realistic environment with various drone models and flight characteristics. Performance is satisfactory in a quiet setting but the system generates excessive false alarms when exposed to lawn equipment.

I. Nomenclature

x	=	sinusoidal waveform of audio signal
a	=	amplitude of sound wave
ϕ	=	phase offset
f	=	frequency of sound wave
t	=	continuous time
n	=	discrete time sample
k	=	sampling frame
N	=	sample length
X	=	fourier transform of audio signal
ω_o	=	angular frequency
M	=	mel-frequency cepstrum

II. Introduction

DRONES have exploded in popularity in both commercial and hobbyist settings, and as a result, managers of outdoor public spaces are increasingly faced with the duality of needing to allow a few drones permission to operate in support of events, while simultaneously preventing the incursion of interloping drones. Other more sensitive public facilities like prisons are also facing an increasing presence of drones, which threatens public safety when contraband like guns and cellphones are dropped into prison yards [1][2][3]. So, in this context, the necessity of detecting and restricting the illegal use of drones emerges.

*Ph.D. Student, Electrical and Computer Engineering, Duke University
†Research Scientist, Humans and Autonomy Laboratory, Duke University
‡Post Doc., Electrical and Computer Engineering, Duke University
§Professor, Electrical and Computer Engineering, Duke University, AIAA Fellow

Detection and classification of drones from the environment has always been a very challenging task in visual, radar-based and acoustic contexts [4] due to complex attributes (geometry, frequency, and sound). Visual detection methods can be complicated by obstructions, poor lighting conditions, and weather. Radar or frequency-based detection methods are expensive and require significant human support.

Acoustic detection is a passive approach in analyzing patterns of wave energy produced by an oscillating body that distinguishes pressure fluctuations between two signals [5]. This type of sensing has many advantages. Since acoustic signals are typically omnidirectional, absolute spherical sensing coverage can be more easily achieved with a smaller sensor footprint. By analyzing the sound pressure levels and spectral peak frequencies of a drone over time, the presence of a drone can be determined [6][7]. A major downside to this pattern-matching approach is that the drone acoustic signature has to be known *a priori* to determine if the real-time spectrogram matches with the captured signal [4].

Because many public venues and facilities do not have budgets to support the acquisition and operation of complex drone detection systems that are very costly [8], we chose to explore the use of acoustic detection methods since the sensors are substantially cheaper. Often such venues simply just want to know whether a drone is present and thus advanced, expensive systems are not needed. We initially explored how to detect and classify drones using a single microphone due to lower cost and system complexity reasons, but eventually intend to link multiple microphones in a network to provide more precise localization. Thus, we present a method of predicting the presence through detection and classification of a drone using a single microphone with inexpensive devices.

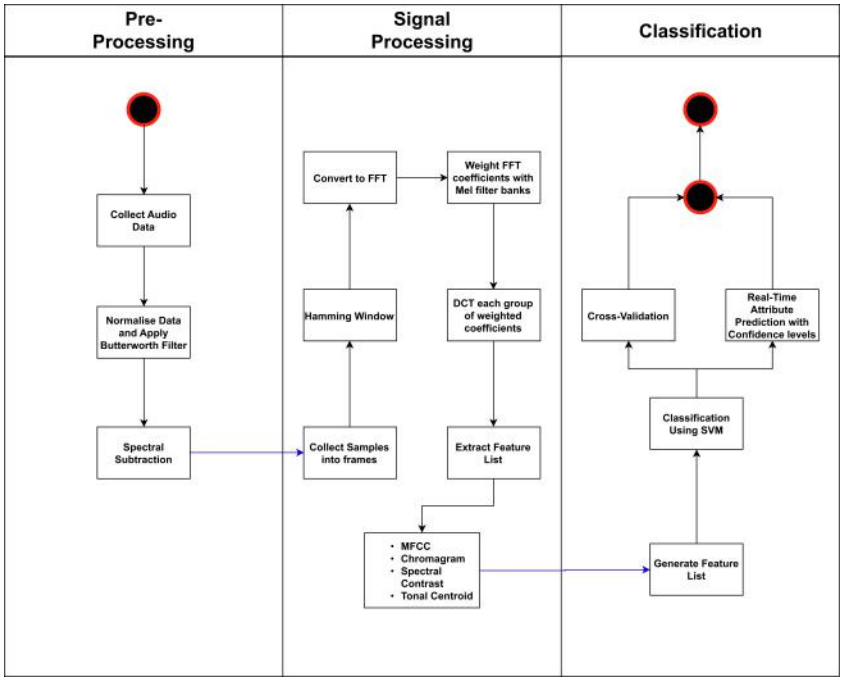

Fig. 1 Flow chart for acoustic detection and classification.

III. Audio Processing

An omnidirectional condenser microphone was used to gather drone audio signatures both in outdoor and indoor environments. The microphone was connected to a low-cost micro-computer i.e., a Raspberry Pi 3B+ that was configured to record and process signals. The Raspberry Pi and microphone can be powered either by solar panels as seen in Figure 2a, or through local outlets as in Figure 2b. This system, along with a mobile phone application detailed in a subsequent section constitute the holistic alerting systems which we call PRIS, Prison Reconnaissance Information System. A 3DR Iris+ drone was used to generate the initial audio data set of drone sounds used in this research since the frequency band for a typical drone lies under 1500Hz with its harmonics within a short range of interval 130 - 180Hz [9][10]. The Iris+ weighs about 1.28kg with average outdoor ambient sound as 39.4079dB and indoor as 33.0261dB [11].

Figure 1 shows a detailed overview of our approach to signal detection and classification. It is divided into

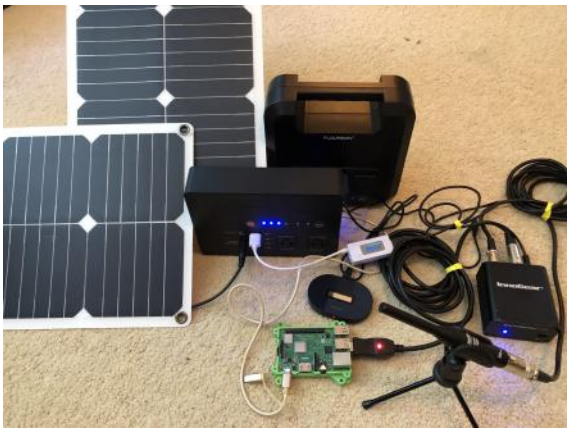

Fig. 2 (a) The acoustic signal detection system that includes a microphone, a Rasberry pi with timer, and optionally solar panels, (b) Appearance of the system when installed inside a waterproof box powered by conventional AC.

pre-processing for noise reduction, signal processing for feature extraction and classification using machine learning, described in a later section. We consider audio signals in the form of $x(t)$ (eq. 1), Discrete Fourier Transforms (DFT) (eq. 2) & and Inverse DFT (eq. 3) for most of our processing.

$$x(t) = a(\sin(2\pi f t + \phi)) \tag{1}$$

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-jk\omega_0 n}, \quad k = 0, 1, ..., N-1 \tag{2}$$

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{jk\omega_0 n}, \quad n = 0, 1, ..., N-1 \tag{3}$$

A. Pre-Filtering

The Butterworth filter is a maximally flat magnitude filter that is used when the signal needs to have a flat frequency response at the passband [12]. Since the audio signals of drones and the background noise are convoluted in most cases, it is beneficial for us to identify the passband of drone sounds and apply the Butterworth filter to them [11]. For this reason, we identified the range of frequencies in which drone audio signals are dominant. We then used these as the cut-off frequencies for the Butterworth filter and the passband subsequently rolls down the frequencies beyond the cut-off point on both sides to zero. An increased filter order results in an increased gain rate of change, which in turn pushes the stopband response to approach the ideal stopband characteristics. However, with an increase in the order, there is an increase in filter coefficients that need to be estimated, representing a tradeoff. Consequently, the accuracy of estimates declines and so does the proximity of the actual stop band response with the ideal stop band [12][13][14]. So an optimal order is required which works with the input signal and the cut-off frequencies to efficiently filter the acoustic signal. After trial and error, we found that the filter design is optimal if a ninth order Butterworth filter is used.

B. Noise Suppressor Design

Because filters such as the Butterworth filter only attenuate certain frequencies from the original signal, not all of the noise is filtered out of the signal. Hence, an additional two-step noise suppressor was designed using spectral subtraction and power spectrum density analysis.

Spectral subtraction is primarily used in the domain of speech enhancement where the main motive is to increase signal quality by reducing distortions caused by ambient noises [15]. Although dual or multi-channel enhancement performs better than single-channel enhancement techniques [15][16], we only have a single-channel since we use only one omnidirectional microphone for audio recordings. Single-channel enhancement is thus used to estimate the spectral

amplitude and phase of noise [15]. This can be done by estimating the power of noise and applying a subtraction algorithm to remove noise estimates from the original signal. Most implementations of spectral subtraction determine pauses in speech to identify the noise, but since drone sounds are continuous, i.e., without definitive pauses, estimating noise is difficult. Thus, we use a separate recording of noise taken just before running the detection algorithm to estimate the noise floor. The recorded noise is divided into windowed segments using a Hanning Window function. A smoothed periodogram is used to compute the short-time power spectrum of each windowed segment [17]. The noise floor power spectrum is then calculated as the weighted minimum of all the short-time power spectrum estimates of windowed segments. We repeated the above process on drone sounds to determine the power spectrum of short-time windowed segments of signals and use the noise floor power spectrum estimate to calculate Segmental Signal-to-Noise Ratio (SNRseg).

Since drone sounds have characteristics that closely match the ambient noise, we need not necessarily filter out all the noise rather we try to reduce the noise power spectrum in the main detection recording (windowed) to a preset minimum threshold [16]. This can be done by subtracting the power spectrum of the windowed signal by an over subtraction parameter which is calculated using a function of SNRseg. The parameters of the function are set such that over-estimates do not effect drone signal quality by suppressing low Signal-to-Noise Ratio (SNR) components of the drone sounds [18]. The subtracted signals are reconstructed using inverse DFT for feature extraction.

IV. Feature Extraction for Machine Learning

In order to apply a machine learning algorithm to classify the data, we needed to extract features from the drone data. For this purpose, we used Mel-frequency cepstral coefficients (MFCCs), spectrogram, chromagram, spectral contrast and tonal centroids (tonnetz) as features.

The most prominent features among these are the Mel-frequency cepstral coefficients and Mel-based chromagrams. Mel-frequency cepstrum (MFC) is a form of power spectrum derived from the log power spectrum of the signal in mel-scale (eq 4) which can map a perceived frequency to its actual frequency [19][20]. MFCCs are derived from the MFC by identifying amplitudes after applying Discrete Cosine Transforms (DCT) on its log powers.

$$M(f) = 1125 \ln (1 + \frac{f}{700}) \tag{4}$$

MFCCs are useful features for speech recognition. They are used as a common feature for competitive baseline benchmarking classification algorithms [21]. Additionally, using filterbanks to retain time-series attributes of the audio features makes MFCCs more robust for detection [22][23]. A major chunk of our feature set (about 70%) was made up of MFC coefficients which helps in correctly classifying the audio signals.

Apart from MFCCs, we used a Mel-based spectrogram to represent a time-varying spectrum of frequencies through a 1D array to extract frequency magnitudes as features. Since most drones have an identifiable pitch but different qualitative characteristics (timbre), we used a chromagram to capture the harmonics of sound which has proved to be ideal for audio matching by correlating harmonic progression, and is robust to the dynamics of source [24].

However, chromagrams and MFCCs consider average spectral characteristics, which might not represent the signal as a cumulative whole. Since drone signals do not have a large separation between the peak and valley in the spectrum, the averaged function might smooth the spectral characteristics. Spectral contrast helps maintain relativity among the spectral characteristics in each frame of the signal, and this is useful for retaining the spectral information in each sub-band [23]. Additionally, we used tonnetz to detect changes in the harmonics of the audio signals by computing Euclidean distance between them [25]. Since drone sounds are monotonic and harmonic changes might be less, tonnetz was a low priority feature.

V. Machine Learning and Classification

A. Support Vector Machine

We elected to use a discriminative statistical machine learning technique called Support Vector Machines (SVM) for classification. SVM is a lightweight, stable and computationally inexpensive alternative for neural networks [26]. SVMs essentially map non-linear input vectors to a high-dimensional feature space called hyperplanes and use a decision surface and margins to partition the input based on labels [27]. Since calculating non-linear mapping hyperplanes can be computationally intensive, SVMs use kernel functions to reflect the inner products of input vectors without explicitly

representing the separation space [28]. There is sparseness in the drone sound dataset even though the dataset is large, thus we determined that separating the data with hyperplanes would be an optimal approach.

B. Model Building

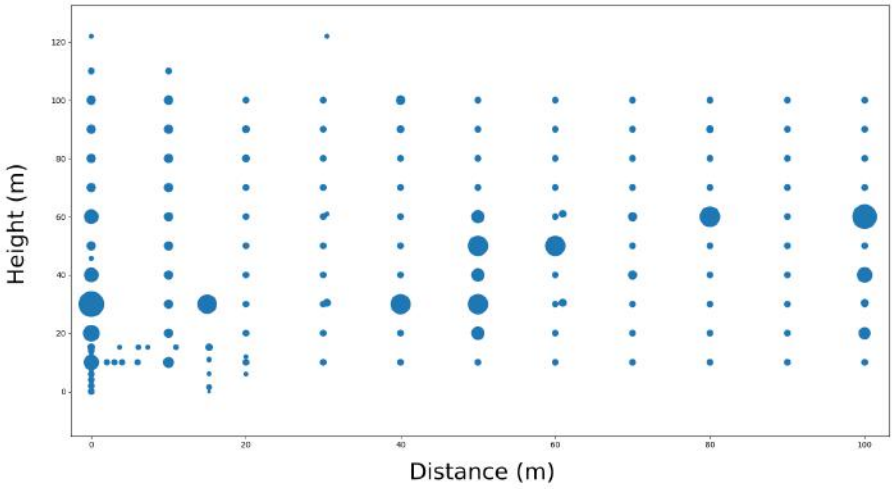

Fig. 3 Distribution of the training data at different locations where the size of the bubbles represent the length of the recordings.

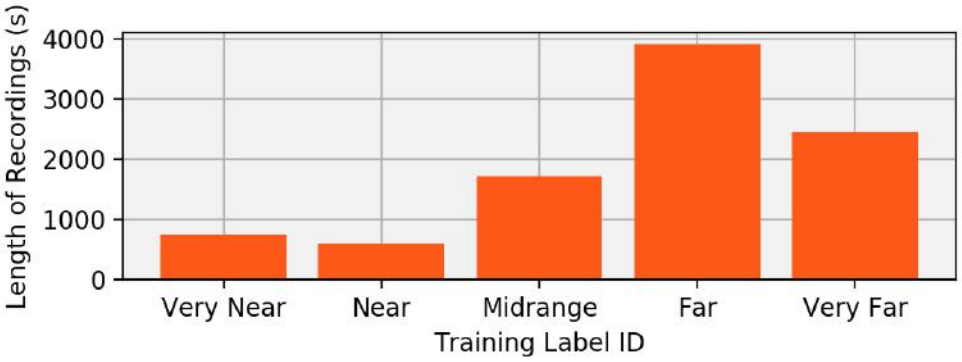

Fig. 4 Data sample distribution for different label IDs used in the initial training algorithm.

Two drones were used to develop a library of drone sounds to train the SVM, a DJI Inspire 2 (79.8dB noise level) along with a 3DR Iris+. With the omnidirectional microphone ($20Hz - 20kHz, uncertainty \pm 3dB$) at the epicenter of 150 coordinate points varying in distance and altitude (Fig. 3), 14932 recordings of one second each were generated and used to train the SVM. The recordings were labeled with heights and distances in a vertical 2D plane, with the microphone at the origin (Fig. 3). Each of the collected sounds were labeled according to radial distances with multiple classes for cross-validation so as to avoid overfitting [29]. The library of drone sounds resulting from these trials can be found at [30].

The dataset was stored temporarily as wav files, then parsed using a sliding window approach to generate pre-segments to generate audio frames for further analysis [22]. Given that some features of audio are more distinct than others [31] and the time series model (Fig. 5) is extremely difficult to classify [21][31], we selected features based on audio filter

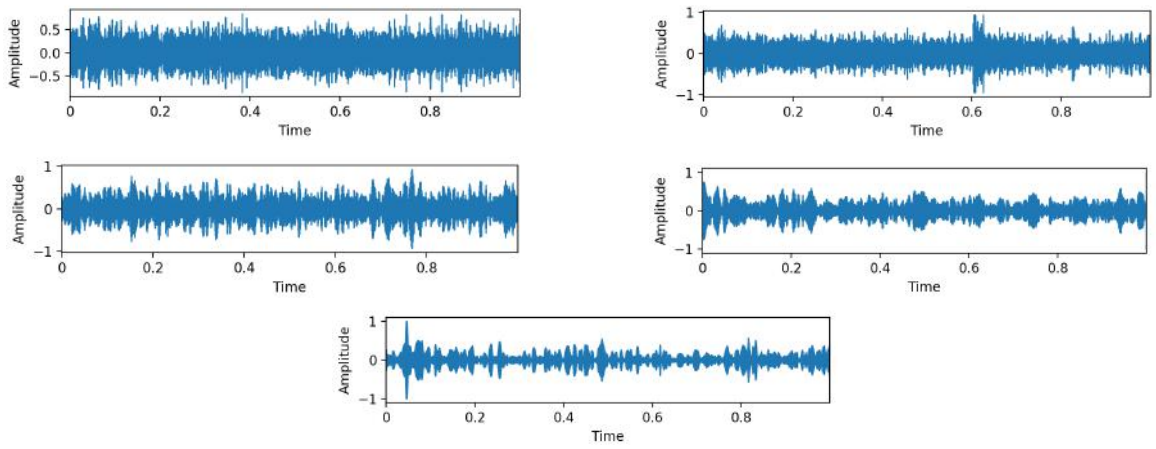

Fig. 5 Amplitude vs. Time plots for drone sounds at (a)very near, (b)near, (c)mid range, (d)far and (f)very far labels, from top-left to bottom-right respectively.

banks generated by auto-correlation methods [22] (e.g., Mel scale, coefficients of Mel-frequency Cepstrum, chromagram, spectral-contrast and tonal centroid) [23], as these have shown exemplary characteristics for classification (subsection IV) [32]. We used a nu-SVM classifier in which we controlled the number of support vectors through the parameter 'nu' which subsequently penalizes incorrect classifications [33]. This algorithm uses a 3rd-degree polynomial basis function as a one-vs-rest classifier to categorize the labels according to their corresponding features. This process is described in (Fig. 1).

The SVM's prediction was distributed into one of five categories - 'Very Near', 'Near', 'Mid-range'. 'Far' and 'Very Far', which was designed to roughly capture distances from the epicenter (Fig. 4 & 5). The near and far samples were significantly larger than the other categories, which could lead to bias in the data (Fig. 4).

C. Prediction Analysis

In order to show the relevance of the basic SVM model predictions, the model was fit to the training data consisting of 12200 data points. Accuracy was measured on the remaining validation data which includes 2732 data points. The validation confusion map, shown in Figure 6 indicates that the SVM model was able to accurately classify the data into its corresponding labels with an aggregate accuracy of 96.86 %.

Even though the accuracy of classification was high on validation data, we observed during the real-time implementation of the algorithm that the SVM could classify 'very near' and 'near' correctly, but everything else was classified as 'very far'. This could be due to the fact that the validation data was a subset of the audio dataset and therefore the sound characteristics for different classes were similar. Because of this, we decided that 'very near' and 'near' labels meant a drone was identified by the classifier and the remaining labels indicated no drones were in the vicinity of the detector.

After the above modifications were made, a set of initial tests were conducted in the laboratory and at an external test site with various models of drones to measure the performance of the detector. Table 1 summarizes the accuracy of the detector in different settings. We noticed that in a quiet environment such as laboratory, the detector classified about 80% of the labels correctly for two different models of drones, a Holystone Predator and Parrot Bepop 2. We saw similar classification results with DJI Inspire 2 flown in a field with only wind as the background noise. Although performance was satisfactory in a quiet setting, it dropped to about 60% when some background noise (generator) was introduced in the environment.

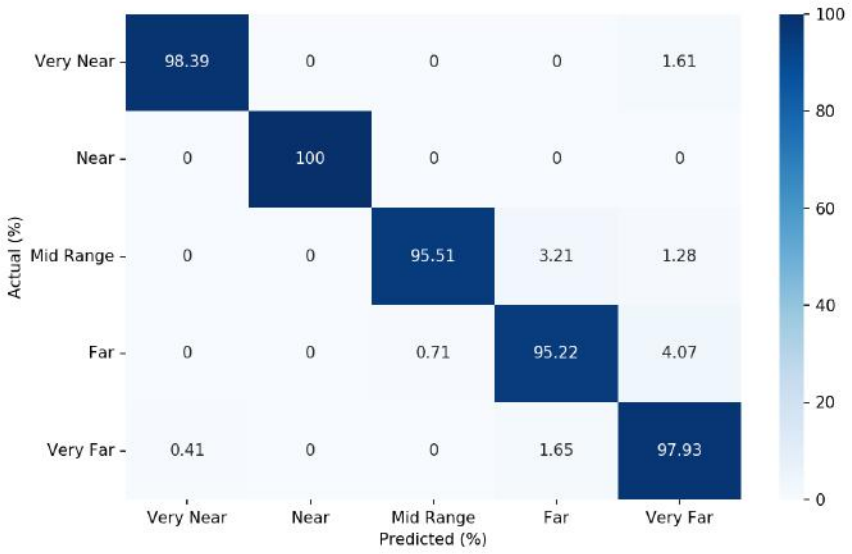

Fig. 6 SVM validation confusion map on 2732 test data.

Environment	Drone Model	Accuracy %
Quiet Lab	Holystone Predator	82.7 %
Quiet Field	DJI Inspire 2	80.4 %
Quiet Lab	Parrot Bepop 2	79.3 %
Field with generator	Parrot Bepop 2	65.2 %
Field with generator	DJI Inspire 2	64 %

Table 1 Initial test results.

D. Alerting and Confidence Estimations

The system was designed to communicate with an app on one or more mobile devices via the microphone and Raspberry Pi placed in a field setting. This app, the Mobile Alerting Interface (MAI) in (Fig. 7) was developed to alert users(in this case prison officials), of a potential interloping drone [34]. Given the probabilistic nature of such an alert and the potential for human bias [35], MAI includes a confidence estimation with each alert.

When a "Drone Detected" decision is made via the microphone and Raspberry Pi, a push alert is sent to a mobile device as depicted in Figure 7a. The user then selects this notification and is shown an interactive map that indicates the estimated location of the detected drone as well as the confidence level of this prediction (Fig. 7b). (Fig. 7).

A simple and robust confidence algorithm was developed to generate the confidence level of each detection based on past detections. The predictions from the raspberry pi are pushed through a queue to calculate a confidence level. If more than half of the predictions within the past 30 seconds are "Drone Detected", then the final prediction will be "Drone Detected" and a push alarm will be generated. When the queue is filled, the oldest predictions drop out of the list via a first-in, first-out approach.

In the queue and dump approach, we used 8 predicted data samples for the previous 30 seconds to generate three levels of confidence (low, medium and high). A confidence level 'high' of a certain label indicates that more than 85% of the recent 8 predictions in the queue were of that label. Similarly a confidence level 'medium' indicates that about 60 - 85% of the recent 8 predictions were of that label and a confidence level of 'low' meant that less than 60% of the predictions in the queue of 8 were of the same label. Both 'high' and 'medium' confidence levels for the detection warranted that the user should be notified about the presence of a drone and thus a notification alert was sent to the MAI via the API. If the confidence interval was low, no detection is communicated through MAI. As compared to the single prediction approach, this queue and dump evaluation made the final "Detected" decision more robust by taking all

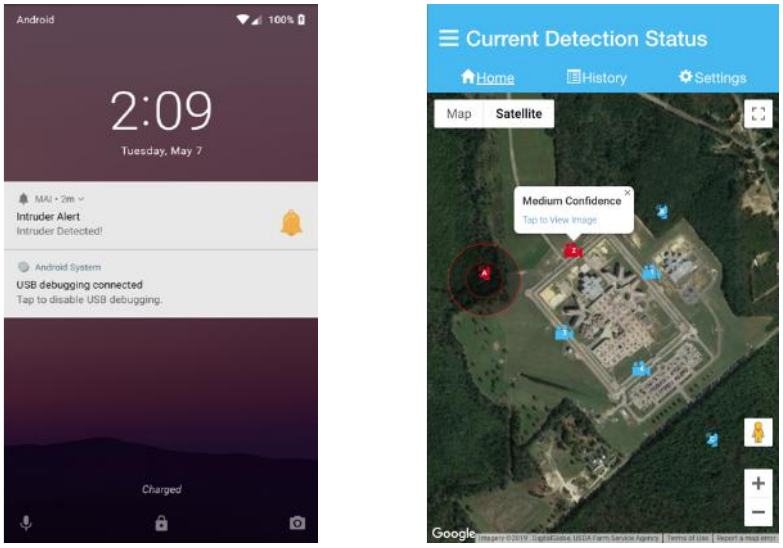

Fig. 7 Mobile Application for prediction output (a) push notification and (b) geo-coordinates of detection [34].

prediction points in a span of 30 seconds into consideration.

VI. Field Tests

A. Hardware Setup

Field tests were conducted for the proposed system with various drone models and flight characteristics to determine performance outside the laboratory. The system as described in subsection III was installed at a local state prison, the Dan River Work Farm (NC). A hotspot was used to facilitate the transfer of the data to the server as well as a push alert to the user. Two drone models were used for testing including a DJI Phantom 4 Pro (76.8dB noise level) and a DJI Mavic Pro with standard propellers (79db noise level). For each drone model, two sets of tests were conducted including hovering static tests and dynamic flight tests from a half-mile away to the location of the microphone. In the dynamic testing, each drone was flown along a continuous path at a relatively constant speed to simulate a drone quickly attempting to approach a drop point.

B. Static Testing Results

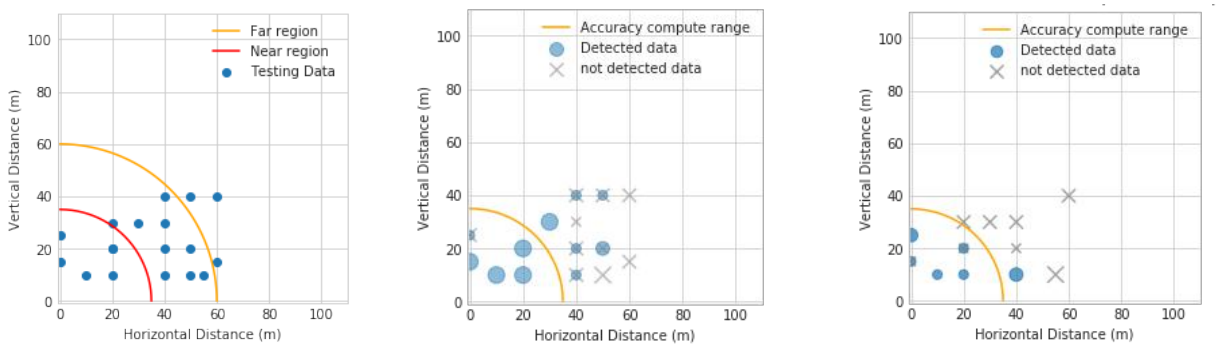

Fig. 8 Dan River Work Farm static test plan and results (a) Static test coordinate points (b) Phantom 4 Pro static testing results and (c) Mavic Pro static testing results. Size of markers represent the number of predictions.

In the static hovering testing, the drones hovered at 17 pre-designated coordinate points as depicted in Fig. 8. These testing coordinate points were chosen in a random distribution in the radial distance range of 0 to 72 meters with

variations in both vertical and horizontal distance. At each coordinate point, the testing drone hovered for 30 seconds and the prediction result via the confidence interval algorithm was collected. The red curve in Fig. 8 (a) indicates the "near" region, which was set at 35m, while the orange curve indicates the "far" region at 60m, which is reflected in the training set. Detection accuracy was calculated for the DJI Phantom and the Mavic as shown in Fig. 8 (b) & (c). Due to the time and resource limitations, all the distance points in the training set were not included in the testing set. Thus the reported accuracies are a reflection of the testing sets.

Detection accuracy of 86.7% was reached for the Phantom 4 Pro model when the drone hovered within the radial distance of 35m. The detection accuracy decreased as the radial distance of the drone from the microphone increased, with a detection accuracy of 62% when the drone hovered within the radial distance of 60m (Fig. 9). The detection accuracy of the Mavic model was slightly lower with an accuracy of 75% achieved at 35m (Fig. 9). The detection accuracy again decreased for Mavic as the detection distance increased. This is not surprising given the fact that no Mavic sounds were used in the training set. Indeed, even though accuracy was slightly less for the Mavic, the detection algorithm performed well given the lack of Mavic training data, suggesting that such approaches may be able to generalize to similar platforms.

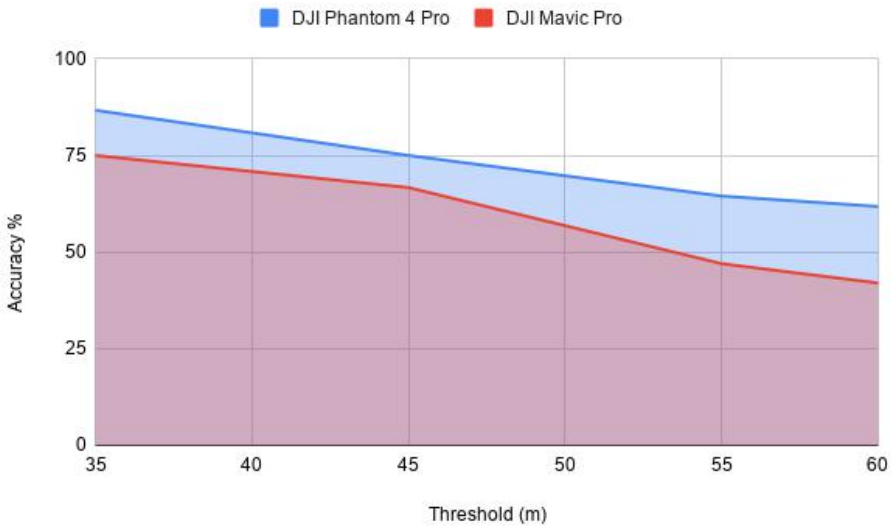

Fig. 9 Distance vs Accuracy comparison graph for DJI Phantom 4 Pro and DJI Mavic Pro static drone test.

C. Dynamic Testing Results

For dynamic testing, a series of flight profiles were tested, all starting 100 meters away from a single overfly point. The DJI Phantom and Mavic flew approximately 35 mph for all the flight profiles. The dynamic flight tests assessed the sensitivity of the drone detection system under more real world conditions where drones are quickly moving through space, typically only hovering for brief periods of time to drop contraband. As with the static tests, these tests measured the number of push notifications generated in the MAI, which only indicate if the drone is either "near" or "very near".

There were four types of dynamic flight profile with multiple runs, which included 1) a direct flight with a 10s hover at a specific distance, 2) a descending flight profile, 3) an ascending flight profile, and 4) a complex flight profile that included a step change in altitude as the vehicle approached the drop point. The simplest, most direct profile is depicted in (Fig. 10). In this profile, the drone approached with a constant vertical distance of 30 meters. Over the three test runs with hover points a 50m, 40m, and 25m (30m altitude), a push alert was triggered at horizontal distances of 35-40m (Fig. 10). In the case of the descending flight profile (Fig. 11), no clear conclusion could be made about predictions since there were no push notifications in any of the three runs, but a prediction was reported in the landing process of run 3. For the ascending flight profile, push notifications were reported at 50m in one run and at 8m for the two following runs (Fig. 12). Finally, for the complex flight profile, (Fig. 13), all three runs reported push notifications in the landing process while in run 1, a push notification was reported at 86m (50m altitude and 70m horizontal distance) and during run 3, a push notification was reported at 35m (30m altitude, 20m in horizontal distance).

These flight test results show that the detector performance did not align with the prediction analysis drawn from initial lab testing. This behaviour was expected since as laboratory tests typically differ from actual flight tests in the environment. Although the performance results were inconsistent among the tests, we saw that the detector could correctly identify different types of drones in various environments.

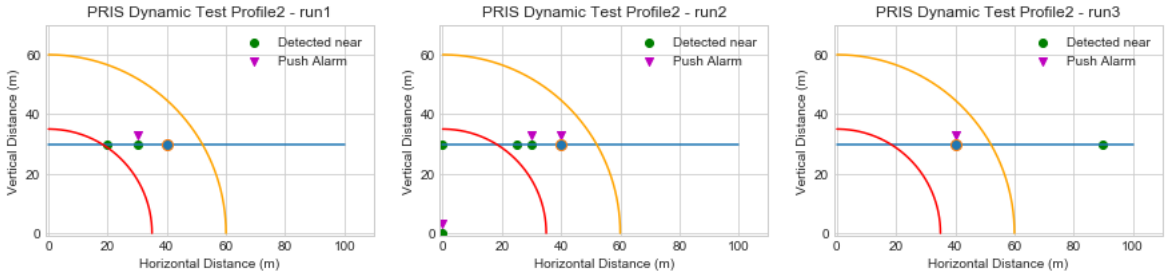

Fig. 10 Dan River Work Farm dynamic test profile replay for constant altitude flight.

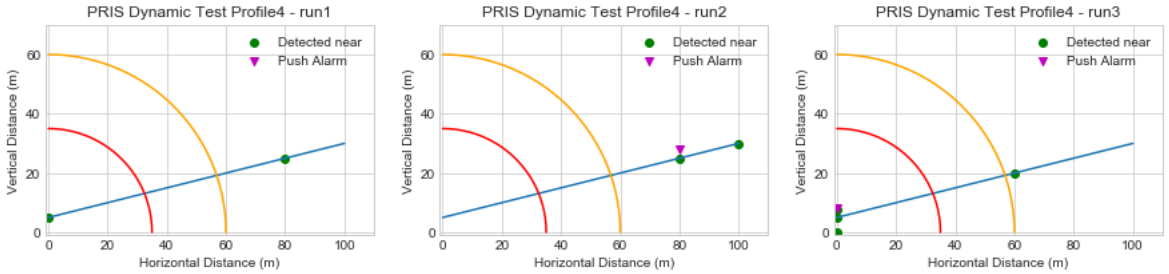

Fig. 11 Dan River Work Farm dynamic test profile replay for descending flight.

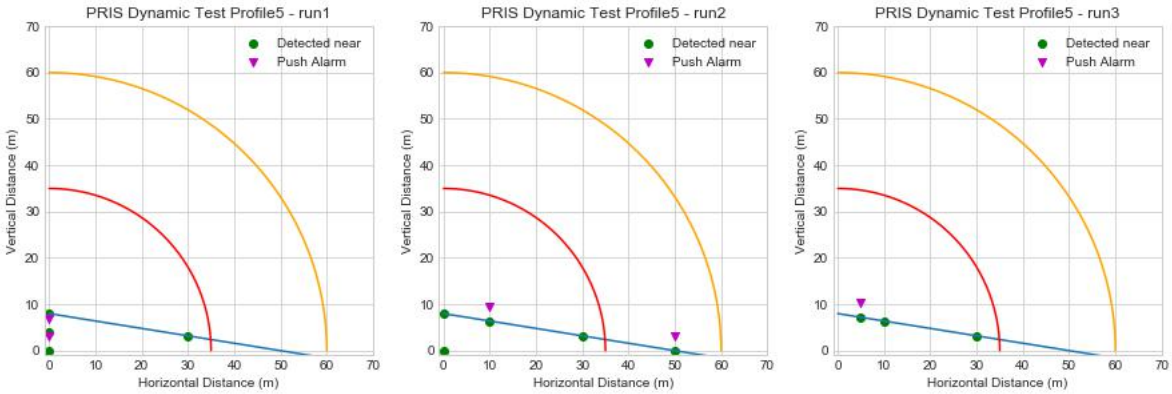

Fig. 12 Dan River Work Farm dynamic test profile replay for ascending flight.

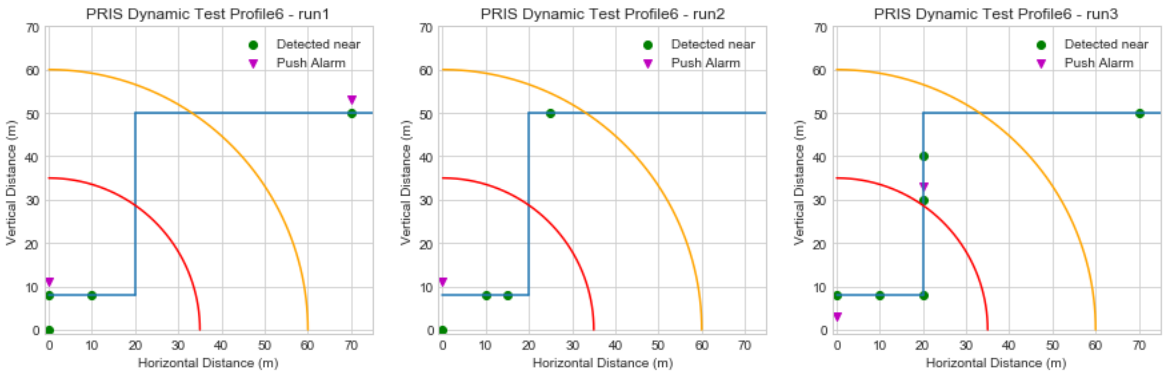

Fig. 13 Dan River Work Farm dynamic test profile replay for complex flight profile.

D. Noise Resilience Test

During the testing at Dan River Work Farm, there were very few false alarms, as the environment is rural with little external disruptions. Thus, we decided to conduct additional testing to determine how PRIS would respond to noises that were similar to that of drones. A series of 10 tests were conducted at the Raleigh Durham airport which included exposing PRIS to aircraft takeoff, landing, and taxiing scenarios. Push notifications were reported with constant "near" predictions for aircraft taxiing while no detection was reported for the other two cases. This indicates that the system is not resilient with aircraft noise present in the background in the same plane. In additional helicopter noise resilience testing, we found out that PRIS continuously generates detections/false alarms when the main propeller was operating within approximately 50m. PRIS was also exposed to two types of maintenance equipment noise, including a leaf blower and a string trimmer which also resulted in continuous false alarms at this same range.

E. Test Summary

Taking the quiet field tests in conjunction with the noise resilience tests suggest that in its current form, PRIS can detect drones sounds somewhat reliably at distances of approximately 40m, which is encouraging given its low cost and ease of operation. If such devices were linked, the range could be greatly expanded. However, PRIS's high number of false alarms suggest significant vulnerability and it is likely that any such deployed system that generated a high number of false alarms would effectively be ignored by the population it was meant to serve. In conversations with prison officials, their feedback was that such systems would likely be well-tolerated if 1 out of 5 alerts was a false alarm, so significantly more work is needed to achieve that threshold.

VII. Conclusion and Future Work

In order to make the current system more robust, we are examining the introduction of a multi-fold detection and evaluation mechanism, as well as expanding our dataset of acoustic signatures. Moreover, we will use a combination of acoustic and radio frequency (RF) analyzers to detect the drone's presence. We originally avoided this approach due to the increased cost of the sensors, but the baseline work indicated that a more robust approach was needed to deal with the false alarms. We are also exploring the use of Convolutional Neural Network's (CNN) in lieu of SVMs due to their efficiency in distinguishing background noises. In addition to these improvements, we are also exploring the development of a human-guided acoustic monitoring and labeling system to characterize the acoustic footprint of a particular location. Using MAI, we are examining whether users can aid in classifying environmental sounds so that this information can be used to retrain the CNN. This would then aid in decreasing false positives and generate a superior data set. Such a human-autonomy collaborative approach would also aid in generalizing this system to different environments.

Acknowledgement

This work was sponsored by the National Science Foundation under the National Robotics Initiative. The field testing was made possible by the efforts of the North Carolina Department of Public Safety, Division of Prisons, including

Scotland Correctional Institution and the Dan River Work Farm. Tigey Jewell-Alibhai flew the drones, and Richard Ruiz and Aneesh Gupta assisted with the flight tests.

References

[1] White, M., "Drone used to smuggle cell phones and saw blades into prison by flying it through 4th-floor window," , 2016. URL https://www.worldwideweirdnews.com/2016/11/24-drone-smuggle-cellphones-saw-blades-into-prison-flying-through-4th-floor-window.html, [Online],[Accessed: 2019-06-06].

[2] Russon, M., "Drones are being used to smuggle drugs into Canadian prisons," , 2013. URL https://www.businessinsider.com/drones-are-being-used-to-smuggle-drugs-into-canadian-prisons-2013-11, [Online],[Accessed: 2019-06-06].

[3] Khaw, C., "Drone crashes while smuggling weed into maximum security prison," , 2014. URL https://www.theverge.com/2014/8/1/5958101/drone-contraband-smuggling-south-carolina, [Online],[Accessed: 2019-06-06].

[4] Güvenç, İ., Ozdemir, O., Yapici, Y., Mehrpouyan, H., and Matolak, D., "Detection, localization, and tracking of unauthorized UAS and jammers," *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*, IEEE, St. Petersburg, 2017, pp. 1–10. doi:10.1109/DASC.2017.8102043.

[5] Harvey, B., and O'Young, S., "Acoustic Detection of a Fixed-Wing UAV," *Drones*, Vol. 2, No. 1, 2018, p. 4. doi:10.3390/drones2010004.

[6] Brandes, T. S., and Benson, R. H., "Sound source imaging of low-flying airborne targets with an acoustic camera array," *Applied Acoustics*, Vol. 68, No. 7, 2007, pp. 752–765. doi:10.1016/j.apacoust.2006.04.009.

[7] Rafaely, B., Peled, Y., Agmon, M., Khaykin, D., and Fisher, E., "Spherical microphone array beamforming," *Speech Processing in Modern Communication: Challenges and Perspectives*, edited by J. B. I. Cohen and E. S. Gannot, Springer, Berlin, 2010, pp. 281–305. doi:10.1007/978-3-642-11130-3_11.

[8] Prisznyák, S., "Drones and Jails," *Scientific Bulletin*, Vol. 23, No. 1, 2018, pp. 43–52. doi:10.2478/bsaft-2018-0006.

[9] Strauss, M., Mordel, P., Miguet, V., and Deleforge, A., "DREGON: Dataset and Methods for UAV-Embedded Sound Source Localization," *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 1–8. doi:10.1109/IROS.2018.8593581.

[10] Intaratep, N., Alexander, W. N., Devenport, W. J., Grace, S. M., and Dropkin, A., "Experimental Study of Quadcopter Acoustics and Performance at Static Thrust Conditions," *22nd AIAA/CEAS Aeroacoustics Conference*, 2016. doi:10.2514/6.2016-2873, URL https://arc.aiaa.org/doi/abs/10.2514/6.2016-2873.

[11] Raya Islam, D., Stimpson, A., and Cummings, M., "Small UAV Noise Analysis," Tech. rep., Humans and Autonomy Laboratory, Durham, NC, USA, April 26 2017. URL https://hal.pratt.duke.edu/sites/hal.pratt.duke.edu/files/u24/Small_UAV_Noise_Analysis_rqi.pdf.

[12] Butterworth, S., et al., "On the theory of filter amplifiers," *Wireless Engineer*, Vol. 7, No. 6, 1930, pp. 536–541.

[13] Zumbahlen, H., *Basic Linear Design*, Analog Devices, Incorporated, 2005. URL https://books.google.com/books?id=ZZdzAAAACAAJ.

[14] Bianchi, G., and Sorrentino, R., *Electronic Filter Simulation & Design*, McGraw-Hill Education, 2007. URL https://books.google.com/books?id=5S3LCIxnYCcC.

[15] Ephraim, Y., Lev-ari, H., and Roberts, W., "A Brief Survey of Speech Enhancement," *The Electronic Handbook*, Vol. 2, 2005. doi:10.1201/9780849333910.ch20.

[16] Upadhyay, N., and Karmakar, A., "Speech Enhancement using Spectral Subtraction-type Algorithms: A Comparison and Simulation Study," *Procedia Computer Science*, Vol. 54, 2015, pp. 574 – 584. doi:https://doi.org/10.1016/j.procs.2015.06.066, URL http://www.sciencedirect.com/science/article/pii/S1877050915013903, eleventh International Conference on Communication Networks, ICCN 2015, August 21-23, 2015, Bangalore, India.

[17] Welch, P., "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Transactions on audio and electroacoustics*, Vol. 15, No. 2, 1967, pp. 70–73. doi:10.1109/TAU.1967.1161901.

[18] Martin, R., "Spectral Subtraction Based on Minimum Statistics," *in Proc. Euro. Signal Processing Conf. (EUSIPCO)*, 1994, pp. 1182–1185.

[19] Davis, S., and Mermelstein, P., "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 28, No. 4, 1980, pp. 357–366. doi:10.1109/TASSP.1980.1163420.

[20] Huang, X., Acero, A., and Hon, H.-W., *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*, 1st ed., Prentice Hall PTR, Upper Saddle River, NJ, USA, 2001.

[21] PoojaR, K., Shetty, S., Şuhani, M. F., and JanardanaD, R., "Sound classification using machine learning and neural networks," *Semantic Scholar*, 2018.

[22] Fayak, H., "Speech Processing for Machine Learning: Filter banks, Mel-Frequency Cepstral Coefficients (MFCCs) and What's In-Between," , 2016. URL https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html, [Online],[Accessed: 2019-05-25].

[23] Jiang, D.-N., Lu, L., Zhang, H.-J., Tao, J.-H., and Cai, L.-H., "Music type classification by spectral contrast feature," *Proceedings. IEEE International Conference on Multimedia and Expo (ICME)*, Vol. 1, IEEE, Lausanne, Switzerland, 2002, pp. 113–116. doi:10.1109/ICME.2002.1035731.

[24] Muller, M., Kurth, F., and Clausen, M., "Chroma-based statistical audio features for audio matching," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005.*, 2005, pp. 275–278. doi:10.1109/ASPAA.2005.1540223.

[25] Harte, C., Sandler, M., and Gasser, M., "Detecting Harmonic Change in Musical Audio," *Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia*, ACM, New York, NY, USA, 2006, pp. 21–26. doi:10.1145/1178723.1178727, URL http://doi.acm.org/10.1145/1178723.1178727.

[26] Akande, K. O., Owolabi, T. O., Twaha, S., and Olatunji, S. O., "Performance comparison of SVM and ANN in predicting compressive strength of concrete," *IOSR Journal of Computer Engineering*, Vol. 16, No. 5, 2014, pp. 88–94. doi:10.9790/0661-16518894.

[27] Cortes, C., and Vapnik, V., "Support-vector networks," *Machine Learning*, Vol. 20, No. 3, 1995, pp. 273–297. doi:10.1007/BF00994018, URL https://doi.org/10.1007/BF00994018.

[28] Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3rd ed., Cambridge University Press, New York, NY, USA, 2007.

[29] Cawley, G. C., and Talbot, N. L., "On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation," *J. Mach. Learn. Res.*, Vol. 11, 2010, pp. 2079–2107. URL http://dl.acm.org/citation.cfm?id=1756006.1859921.

[30] Cummings, M., Mandal, S., Wang, C., and Chen, L., "Prison Reconnaissance Information System," Duke Wordpress, 2019. URL https://sites.duke.edu/prisdatabase/, [Online Database],[Accessed: 2019-06-09].

[31] Salamon, J., and Bello, J. P., "Unsupervised feature learning for urban sound classification," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2015)*, IEEE, South Brisbane, Queensland, Australia, 2015, pp. 171–175. doi:10.1109/ICASSP.2015.7177954.

[32] Jeon, S., Shin, J.-W., Lee, Y.-J., Kim, W.-H., Kwon, Y., and Yang, H.-Y., "Empirical study of drone sound detection in real-life environment with deep neural networks," *25th European Signal Processing Conference (EUSIPCO)*, Kos Island, Greece, 2017, pp. 1858–1862. doi:10.23919/EUSIPCO.2017.8081531.

[33] Schölkopf, B., Smola, A. J., Williamson, R. C., and Bartlett, P. L., "New support vector algorithms," *Neural computation*, Vol. 12, No. 5, 2000, pp. 1207–1245. doi:10.1162/089976600300015565.

[34] Wang, C., and Cummings, M., "A Mobile Alerting Interface for Drone and Human Contraband Drops," American Institute of Aeronautics and Astronautics (AIAA) Aviation and Aeronautics Forum and Exposition, Dallas, Texas, 2019. URL https://hal.pratt.duke.edu/sites/hal.pratt.duke.edu/files/u35/final-chungeAIAA2019.pdf.

[35] Tversky, A., and Kahneman, D., "Judgment under Uncertainty: Heuristics and Biases," *Science*, Vol. 185, No. 4157, 1974, pp. 1124–1131. doi:10.1126/science.185.4157.1124, URL https://science.sciencemag.org/content/185/4157/1124.