DronePrint: Acoustic Signatures for Open-set Drone Detection and Identification with Online Data

HARINI KOLAMUNNA, The University of Sydney
THILINI DAHANAYAKA, The University of Sydney
JUNYE LI, University of New South Wales
SURANGA SENEVIRATNE, The University of Sydney
KANCHANA THILAKARATNE, The University of Sydney
ALBERT Y. ZOMAYA, The University of Sydney
ARUNA SENEVIRATNE, University of New South Wales

With the ubiquitous availability of drones, they are adopted benignly in multiple applications such as cinematography, surveying, and legal goods delivery. Nonetheless, they are also being used for reconnaissance, invading personal or secure spaces, harming targeted individuals, smuggling drugs and contraband, or creating public disturbances. These malicious or improper use of drones can pose significant privacy and security threats in both civilian and military settings. Therefore, it is vital to identify drones in different environments to assist the decisions on whether or not to contain unknown drones. While there are several methods proposed for detecting the presence of a drone, they have limitations when it comes to low visibility, limited access, or hostile environments. In this paper, we propose *DronePrint* that uses drone acoustic signatures to detect the presence of a drone and identify the make and the model of the drone. We address the shortage of drone acoustic data by relying on audio components of online videos. In drone detection, we achieved 96% accuracy in a closed-set scenario, and 86% accuracy in a more challenging open-set scenario. Our proposed method of cascaded drone identification, where a drone is identified for its 'make' followed by the 'model' of the drone achieves 90% overall accuracy. In this work, we cover 13 commonly used commercial and consumer drone models, which is to the best of understanding is the most comprehensive such study to date. Finally, we demonstrate the robustness of DronePrint to drone hardware modifications, Doppler effect, varying SNR conditions, and in realistic open-set acoustic scenes.

CCS Concepts: • **Computing methodologies** → **Feature selection**; **Neural networks**.

Additional Key Words and Phrases: Drones, Acoustic fingerprinting, LSTM, Drone Audio Dataset

ACM Reference Format:
Harini Kolamunna, Thilini Dahanayaka, Junye Li, Suranga Seneviratne, Kanchana Thilakaratne, Albert Y. Zomaya, and Aruna Seneviratne. 2021. DronePrint: Acoustic Signatures for Open-set Drone Detection and Identification with Online Data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 1, Article 20 (March 2021), 31 pages. https://doi.org/10.1145/3448115

Authors' addresses: Harini Kolamunna, The University of Sydney, harini.kolamunna@sydney.edu.au; Thilini Dahanayaka, The University of Sydney; Junye Li, University of New South Wales; Suranga Seneviratne, The University of Sydney; Kanchana Thilakaratne, The University of Sydney; Albert Y. Zomaya, The University of Sydney; Aruna Seneviratne, University of New South Wales.

2474-9567/2021/3-ART20 $15.00
https://doi.org/10.1145/3448115

1 INTRODUCTION

Drones are becoming ubiquitous and are no longer limited to military use. They are finding many commercial and commodity applications in the likes of last-mile goods delivery, agriculture, surveying, emergency response, and cinematography in addition to their recreational uses [2]. The global drone market is expected to reach 13 billion US dollars by the end of 2025, which is a staggering 200% projected growth over a period of 5 years [78]. While a limited number of countries have entirely banned the sale and the use of drones, in many countries purchasing a drone is quite straight forward, and drone usage is permitted given a set of guidelines are followed by the pilots [35].

Such wide availability of drones also gives rise to potential misuse of drones in illegal activities as well as to cause public disturbances. In last few years, we have seen reports of increasing volumes of such activities in the likes of using drones for drug smuggling, invade private spaces, or even intentionally flying into no-fly areas to create havoc. For instance, in late 2018 and early 2019, departures at the London Gatwick and Heathrow airports were temporarily suspended for several hours after reports of possible drone sightings over the airfields [51, 52]. In 2015, a drone carrying crystal meth crashed into a parking lot in Tijuana, a Mexican city close

to the US-Mexican border [50]. Multiple incidents reported unauthorized drones recording public events and occasionally crashing on to the spectators or participants creating injuries or damaging landmarks [14, 73, 80]. Also, there have been attempts to weaponize commercial drones. For example, in 2018 in Caracas, Venezuela, two DJI M600 drones carrying explosives detonated in the proximity of an outdoor setting where the Venezuelan president was delivering a speech [43]. In many of those instances, the authorities failed to provide a timely response to control the unauthorized drone, finding the perpetrating pilot, or in some cases even to establish the fact that a drone was actually present. The drone-related incidents also resonate with the growing concerns raised by the general public and drone pilots [9, 79, 84, 85, 88], calling for technological and regulatory actions to offer protection from unauthorized drones.

As identified by Lykou et al. [44], unauthorized drones launched near critical sites require rapid identification and response. The authors also mentioned the intention of authorities to develop an identification system for civilian drones to assist airspace segregation and compliance management. Such scenarios motivate a defensive system that could be deployed within or around critical facilities, capable of obstruction-robust, close proximity drone detection (i.e. detecting the presence of a drone in proximity), identification (i.e. identifying the exact make and model of the drone), and neutralization (i.e. carefully take down or divert away unauthorized drones) to counteract both intentional and unintentional threats posed by them.

In this work, we focus on *drone detection* and *drone identification*. *Drone detection* is important in detecting an unauthorized entry to a no-fly zone. The additional step of *drone identification* plays an important role in multiple scenarios. For example, if the specific drone model can be identified, one may find specific ways to *neutralize* the drone by using known vulnerabilities in communication or firmware. Also, in case of managing a fleet of drones, where we know the sound profiles of our drones, the *drone identification* process allows detecting something out of the ordinary. Moreover, there are instances where it is allowed to fly drones over populated areas, subject to some conditions on weight and size. In such instances, *drone identification* can assist the process of distinguishing between complying and non-complying drones.

Multiple studies demonstrated the feasibility of detecting and identifying drones using various forms of data such as video [59, 77], RF [24, 25], thermal imaging [77], radar [12, 87], and acoustics [58, 66]. Many of these technologies are still nascent, and they have their own advantages and disadvantages. Our focus in this paper is to build an acoustics-based drone detection and identification system. Acoustics offer some unique advantages compared to other methods such as the feasibility to detect drones under low visibility and the ability to detect drones at lower altitudes compared to radar-based methods [57, 72].

While there are some existing studies in acoustics-based drone detection and identification, they are either; *i) drone decision systems making a binary decision whether or not a drone is present using acoustics* [3, 7, 67, 75] or *ii) focus on identifying a limited number of drones* [1]. Only a few studies looked into the feasibility of open-set drone detection and identification (i.e. detecting and identifying drone sounds when the input can be any sound) [58]. Moreover, the feasibility of such methods was shown only in limited settings such as working only with drones without hardware modifications/defects. This is understandably due to the limitations of collecting large volumes of data covering different drone models. To this end, in this paper, we propose an acoustics-based open-set drone detection and identification system, DronePrint, that can identify over 13 drone models by sourcing real-world data as well as data collected from online platforms. More specifically, we make the following contributions.

- We explain the identification of prominent and distinguishable features of drone sound. For this, we study the sound generation mechanism of drones and characterize the frequency spectrum of drone sounds. We show that the majority of the signal energy remains in the range of 0-8kHz and there are equally spaced prominent energy peaks at the range of 0-2kHz where the first peak is in the range of 150Hz-400Hz. We show that these are the prominent features of the drone sounds that can also be used to distinguish them from other common mechanical sounds. Thus, we suggest placing more emphasis on lower frequencies in feature extraction, such as Mel-frequency cepstral coefficients (MFCC) in the range of 0-8kHz.
- To address the lack of drone acoustics samples, mainly driven by the practical limitations of getting access to all available drone models and the restrictions of experimentally collecting data, we propose to source data from online drone videos and extract the audio component. We show that such data can be combined with real-world collected data without any performance impact. The models that are trained and validated with the online extracted data can be effectively used in real world testing. To the best of our knowledge, ours is the first study to show this feasibility.
- We nullify the effect of differences in signal strengths with the distances to the drone emanating the sound and the most common hardware modification of low-noise propellers with the use of time-domain amplitude normalization. This avoids the necessity of training the model with different sound signal strengths. Also, in the feature space, we normalize each feature vector to unit length by dividing by its L_2 norm to minimize the effect of instantaneous noise such as birds' sounds.
- We propose an LSTM-based open-set drone detection and identification system, DronePrint, that can first classify a given sound belongs to a drone or not and subsequently identify the make and the exact model of the drone. We used two techniques to address the open-set problem. First, we use a *background class* where we train a binary classifier that differentiates drone sounds from non-drone sounds. To avoid the training time data requirement of non-drone sounds, we propose classwise-OpenMax, which is a modified version of the OpenMax method [5], initially proposed for open-set image classification.
- We compare our methods with previous work and show that both the proposed methods outperform them by over 15% in accuracy in both open-set drone detection with unseen signals and subsequent drone make and model identification.
- Finally, we evaluate the robustness of DronePrint's detection and identification models under different operational conditions such as when using low-noise propellers, with Doppler effect, and under varying signal-to-noise-ratio (SNR) conditions. Also, we evaluated DronePrint's drone detection performance for an extended period through emulation by replaying various natural and artificial sounds. We show that DronePrint still detects and identifies drones with different hardware modifications and under the Doppler effect without the need for re-training. We show that by augmenting the training set with a portion of noise incorporated samples, DronePrint's detection and identification rates remain minimally affected even at SNR conditions as low as 1dB. Next, we demonstrate that DronePrint provides 81% open-set accuracy in realistic scenarios of ever-changing various natural and machinery sounds for an extended time period.

The rest of the paper is organized as below. In Section 2, we present the related work. Section 3 presents our dataset and explain the drone sound generation mechanism and frequency characteristics. Section 4 presents the DronePrint pipeline. Section 5 presents the results of DronePrint detection and identification models followed by the analysis of models' robustness in Section 6. In Section 7, we discuss the limitations and possible extensions of our work. Section 8 concludes the paper.

2 RELATED WORK

Our focus in this paper is on drone detection and identification. As we describe below, most of the existing related work involved only drone detection, and only a few works have focused on the drone identification, and that was also in limited settings and did not consider the more practical requirement of open-set drone detection [72]. We present related work under four broad categories; *i) radar-based methods*, *ii) computer vision-based methods*, *iii) radio-frequency (RF) signature-based methods*, and *iv) acoustics-based methods*.

Radar-based methods use reflected radar signatures to detect and classify drones. Zhang and Li [87], considered the problem of detecting drones when the received radar signal is superimposed by micro-Doppler features of multiple drones. By leveraging the Cadence Frequency Spectrum (CFS) features and a k-means classifier, authors tried to distinguish the presence of various combinations of three drone types; quadcopter, helicopter, and hexacopter, in a closed world setting. Authors achieved over 90% accuracy in all the scenarios they tested. Ritchie et al. [62] used micro-Doppler radar centroid and radar cross-section with a naive Bayes classifier to classify drones with different payload weights and achieved above 95% accuracy. Church et al. [12] evaluated Lidar for drone detection, achieving about 90% of detection rate up to 200m range within 30 deg Field-of-View. Other comparable works include [36, 46, 86]. Radar-based methods have the advantage of not requiring line-of-sight and also can assist drone localization in addition to detection. However, they may have limited use when it comes to drones with smaller surface area or low altitudes [57].

Vision-based methods rely on cameras for data acquisition and process the images in order to identify drones. For example, Unlu et al. [77] employed 2-dimensional Generic Fourier Descriptor (GFD) features to train a neural network that can differentiate drones from birds. The authors achieved an accuracy of around 85% and showed that their method could outperform CNNs when less data is available. Peng et al. [59] considered the problem of detecting drones in a given image (i.e. coming up with the bounding boxes surrounding the drones). To address the problem of lack of labelled drone data in standard object detection datasets, the authors used the Physically Based Rendering Toolkit (PBRT)[1] to synthesize photo-realistic drone images. Then, the authors used the synthesized data to fine-tune a pre-trained Faster R-CNN [61] and showed that an Average Precision (AP) of ~80% could be achieved on real drone images. Other similar studies include [41, 71]. Vision-based methods can achieve high accuracy with high-resolution cameras under strict line-of-sight. Nonetheless, such hardware is expensive, and vision-based methods may fail drastically in adverse weather conditions such as rain, mist, or fog [72].

RF signature-based approaches sense the wireless transmission between the controller and the drone (i.e. control frames, data frames, or both), and detect and identify drones accordingly. Ezuma et al. [24, 25] implemented a drone detection and identification system using the energy-transient statistical features of the RF signal such as mean, skewness, and variance measured in both indoor and outdoor settings. At higher SNRs, the authors achieved 95% accuracy for drone detection and 90% accuracy for identification of 17 drones, using a Markov model and a kNN classifier, respectively. However, when the SNR is lower than 12dB, the identification performance deteriorated significantly by about 10% per dB loss in SNR. Nyugen et al. [53] leveraged the unique RF patterns caused by physical body shift and vibration of drones to detect and identify drone movements and drone types. Two independent detection algorithms were employed for drone body vibration and body shifting detection

[1] https://www.pbrt.org/

using STFT features followed by an evidence-based identifier to achieve above 90% detection accuracy even at long distances. For identification, authors achieved above 80% accuracy with eight drone types in a closed-set setting. When evaluated over an unknown dataset, the model was able to identify only 64% of the unknown drone classes. Shi et al. [68] conducted a similar study.

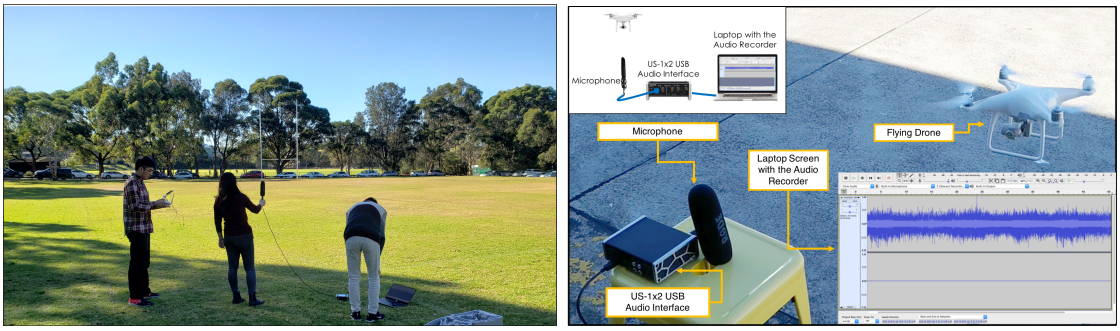
RF signature-based drone detection and identification methods require active communication between the drone and its controller, making them inept for autonomous drones. Moreover, they might suffer from long distance and signal interference from other RF transmissions that is common especially in the unlicensed frequency bands used by commercial and recreational drones.

Acoustics-based methods generate signatures from drone sounds using feature representations such as Linear Predictive Cepstral Coefficients (LPCC), Mel-Frequency Cepstrum Coefficients (MFCC), Short-Time Fourier Transform (STFT) coefficients, or spectrograms. Then the signatures are matched using simple similarity measures or by training machine learning models. For instance, Sedunov et al. [66] showed that there were two unique frequency components on the audio spectrum generated by two groups of drone motors (4300 and 4700 RPM) and that feature alone yields a 92% detection accuracy.

Drone detection - Uddin et al. [75] used investigated Independent Component Analysis (ICA) and MFCC features of drone acoustic signals to build an SVM classifier that achieved 98% accuracy in simulation setting where data was sourced from online databases. Anwar et al. [3] compared MFCC and LPCC features with SVM classifier for drone detection from environmental sounds of birds, aeroplane and thunderstorm, and concluded that MFCC with SVM outperforms LPCC with SVM by about 20%, achieving overall 96.7% detection accuracy. It is unclear the number and the make of the drones used in this study. Shi et al. [67] implemented an SVM-based drone detection using MFCC that achieved above 85% accuracy. However, the detection algorithm was not evaluated against other unknown sounds, rather only with different levels of environmental noise. Jeon et al. [34] compared Gaussian Mixture Models (GMM), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN), for drone sound detection and showed that the RNN models achieved the best detection performance with an F-Score of 0.8009 for the known sounds. When evaluated with unknown drone sounds, the RNN model maintained an F-score of 0.6984. Similarly, Cabrera-Ponce et al. [7] used microphone arrays mounted on a patrolling drone to detect nearby intruding drones by treating the spectrograms of the acoustics signals as images and training a CNN on Inception v.3 architecture [13]. The authors did not evaluate the system with unknown drone sounds, and only used one drone in the training.

Drone identification - Ibrahim et al. [29] designed a lightweight model that uses the acoustic feature of MFCC with SVM to detect drone with/without payload at, and a more computational-intensive model to classify the payload weight classes. In a more comprehensive study, Pathier [58] used MFCC with SVM to identify the drones, each with unique manufacturing defects in their motors. Overall, 54 motors and 11 drones of the same make were used for evaluation, achieving above 99% accuracy in an experiment carried inside a sound booth with stationary drones. The authors also evaluated the model by marking a fraction of the drones as 'other' and reached an average accuracy of 90.8% for the known 'other' drones. Al-Emadi et al. [1] tested a simple two-class drone classification with CNN, RNN, and CRNN (Convolutional-Recurrent Neural Network). The authors found that the CRNN, which also had a significantly lower complexity and was easier to train, produced the best performance of 92% accuracy.

Acoustics-based methods can work without line-of-sight and also can facilitate other related application scenarios such as localization and payload detection. They have limitations in terms of range and how they might perform under noise. However, acoustics-based methods have some advantages over other methods, such as no Line-of-Sight requirement in contrast to visual methods, which are critical in countering drones from within a critical facility such as an airport [44]. Therefore, acoustic-based detection and classification system is going to be part of holistic drone detection and classification frameworks that deploy multiple methods. However,



(a) Data collection in the field (b) Experimental setup

Fig. 1. Experimental data collection

acoustics-based classification of drones is not yet been thoroughly studied, as relevant studies mostly suffer from insufficient data in terms of drone classes and the amount of data available.

In contrast to existing work in acoustics-based drone detection and identification, we present an acoustic-based open-set drone detection and subsequent identification system that can identify over 13 drone models. We demonstrate that the expansion to the real-world collected data can be done by scraping from online videos without any performance impact. Kolamunna et al. [38] introduced the usage of online sourced data. However, in this work, we demonstrate a system solely trained with available online data, and we propose feature normalization methods to address distance effects and instantaneous noise. Also, we conduct multiple experiments to demonstrate the robustness of DronePrint.

3 DATASETS & DRONE SOUND CHARACTERISTICS

We categorize the data we collected into three datasets; 1) **DS1:** Drone acoustic samples that are experimentally collected and online scraped (from YouTube videos and other online sources), 2) **DS2:** Mechanical and natural background noise samples, 3) **DS3:** Drone and noise samples that are unknown to DronePrint for open-set evaluation. In this section, we describe how we built these datasets, followed by acoustic analysis of drone sounds. We make available our datasets to the research community through our `GitHub` page [22].

3.1 Data Collection

DS1: We experimentally collected data in a park in three sessions over two days when there is not much other activity happening, using the setup illustrated in Figure 1. In particular, we captured the acoustic signals emanating from the drone sampled at 44.1kHz, using a high-quality directional microphone (RODE NTG4), when the drones were flying at around 20m above the ground and within a 50m radius. The directional microphone has a *Supercardioid* polar pattern, which picks the front and sides and rejects 150 degrees to the rear. However, since we are holding the microphone up-right at 1m above the ground level, we do not miss any drone sound of interest coming from any of the directions. We collected 3-4 minutes of data at each session for five classes of drones; *Parrot: Bebop 2* [56], *DJI: Spark* [21], *DJI: Mavic Pro* [19], *DJI: Matrice 100* [18], and *DJI: Phantom 4 Pro* [20].

Since it is not possible to obtain all popular drone models for experimental data collection, we also used data from online sources. First, we focused on YouTube videos of people flying the drones for drones reviewing and noise level testing purposes. We downloaded the audio component of these YouTube videos using the *4K*

Table 1. Details of the DS1

Data type	Source	No. of traces (Duration)		
		Training (150s)	Validation (50s)	Testing (50s)
Parrot: Bebop 2	YouTube (*Training & Validation*); Experimental (*Testing*)	8 (10s-40s)	2 (15s & 35s)	3 (10s-20s)
DJI: Spark		9 (5s-27s)	4 (10s-20s)	
DJI: Mavic Pro		8 (10s-25s)	3 (10s-20s)	
DJI: Phantom Pro 4		9 (6s-25s)	2 (20s & 30s)	
DJI: Matrice 100		4 (5s-70s)	1 (50s)	
Autel: EVO	YouTube	5 (5s-100s)	3 (10s-30s)	3 (12s-20s)
DJI: Inspire 2		3 (40s-60s)	3 (11s-23s)	3 (8s-24s)
DJI: Mavic Air		7 (8s-45s)	4 (9s-16s)	3 (5s-25s)
Feima Robotics: J.ME		3 (all 50s)	2 (25s each)	2 (10s & 40s)
Parrot: Anafi		10 (5s-70s)	2 (25s each)	3 (10s-25s)
Parrot: Bebop	Online dataset [1]	150 (all 1s)	50 (all 1s)	50 (all 1s)
Parrot: Membo				
MikroKopter: MK-Quadro	Online dataset [70]	3 (all 50s)	1 (50s)	1 (50s)

Table 2. Details of the DS2

Data type	Source	No. of traces (Duration)		
		Training (270s)	Validation (90s)	Testing (90s)
Busy Road Noise	YouTube	3 (all 90s)	1 (90s)	1 (90s)
Airplane				
Hair Dryer				
Raining				
Birds				
Human Talking				
Calm Environment	Experimental	3 (all 90s)	1 (90s)	1 (90s)

Video Downloader software.[2] In addition to collecting extra data samples for the five drone classes we covered in our experimental data collection, we were able to collect data for five additional drone classes, as shown in Table 1. We manually removed the signal with audible background noise from the audio file to get audio samples with minimal amount of background noise. Second, we used two publicly available data sets for three more drone classes; *Parrot: Membo* and *Parrot: Bebop* [1], and *MikroKopter: MK-Quadro* [70]. These are experimentally collected audio data for research purposes that were released publicly.

Finally, we split **DS1** into three parts; 60% (150s) for *training* and 20% (50s) each for *validation* and *testing* sets. To ensure the generalizability of the subsequent classifier we build, we ensured that the data used in training, validation, and testing are from different sources.

DS2: We built DS2 to capture various possible background sounds that are to be used in training and testing of DronePrint. We captured the acoustic signal of *calm environment* that has no audible noise in an indoor environment, by using the above data capturing setup in five different sessions. In addition, we scraped audio

[2]https://www.4kdownload.com/

Table 3. Details of the DS3

Data type	Source	No. of traces (Duration)
		Testing (30s)
Drones		
AIrHogs: DR1FPV Race	Soundsnap 3	1 (30s)
Blade: NanoQX		
ZeroZero: Falcon		
DJI: Inspire1		
DJI: MavicMini		
DJI: Phantom3		
Syma: X5SW		
Non-drone mechanical sounds		
Helicopter	YouTube	1 (30s)
Motorcycle		
Lawn Mover		
Air Blower		
Car		
Natural background noise		
Slight Breeze	YouTube	1 (30s)
Bees Humming		

signals from YouTube videos for *busy road noise*, *air plane sounds*, *hair dryer*, *rain*, *bird sounds*, and *human speech*. Here also we split these audios into three parts; 60% (270s) for *training* and 20% (90s) each for *validation* and *testing* sets. We collected five YouTube traces (90s each) for each class of sound. Similar to DS1, we ensured that the data used in training, validation and testing are from different sources. As such, we used three of them for extracting the *training* samples, and one each for extracting *validation* and *testing* samples. Details of DS2 are shown in Table 2.

DS3: We built DS3 for open-set evaluation consisting of both drone and non-drone sounds. We scraped acoustic signals for seven drone classes that are not included in DS1 from soundsnap, which is an online database for various sounds.[3] Also, we scraped acoustic signals from YouTube for five other non-drone mechanical sounds and two natural sounds. Details of DS3 are shown in Table 3.

3.2 Drone Sound Characteristics

Next, we analyze acoustic patterns generated by drones, their sources, and signal characteristics that indicate the possibility of fingerprinting drones based on acoustic signatures.

Sources of the drone sounds. According to the principle of aerodynamics, drones fly by pushing the air close to the propellers downwards as the propeller blades spin at very high speeds. The audio noise generated from drones is primarily due to its propellers, motors, and the mechanical vibration of the body. Among these sources, propeller noise is significant due to the air drifting with their high-speed motion. This rapid contact of blades and air results in a chopping or buzzing sound, which is heard as noise in the human audible range. The motor

[3]https://www.soundsnap.com

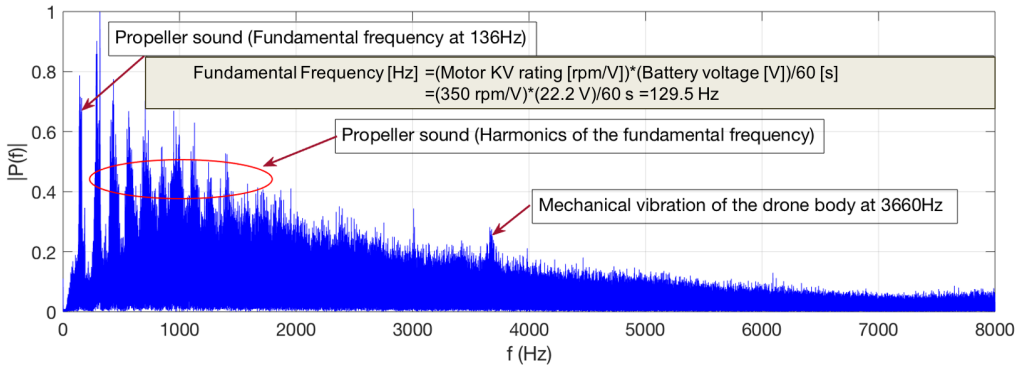


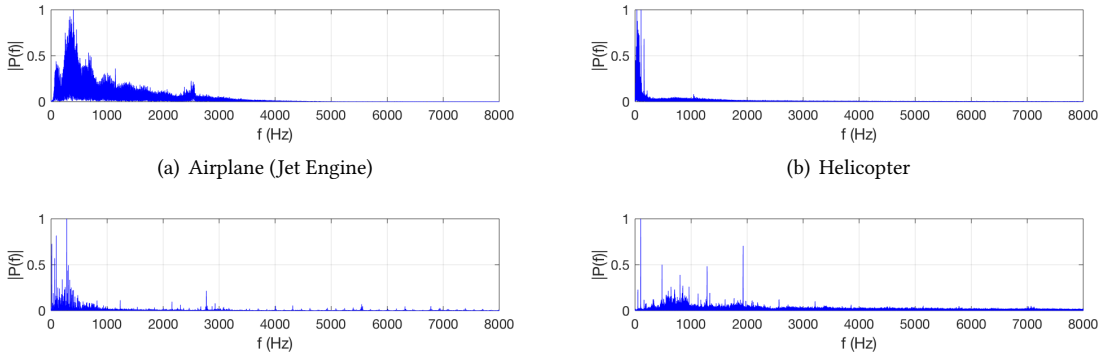
Fig. 2. Power spectrum of a drone (DJI Inspire 1) sound.

contributes to the noise in terms of a whining tone. Nonetheless, its intensity is significantly less compared to the propeller noise. The propellers and motors transmit vibrations through the arms into the drone frame. This results in generating the noise due to the mechanical vibration of the drone body [37].

Frequency characteristics of the drone sounds. The spinning of air with multiple propellers results in a fundamental frequency and its higher-order harmonics in the range of 0-2kHz [74]. This fundamental frequency is defined by the KV rating of the stepper motors of these drone propellers. KV rating is the ratio of the motor's unloaded revolutions per minute (rpm) to the battery voltage. The KV rating multiplied by the drone battery voltage provides the rpm value, and when this value is divided by 60, that results the fundamental frequency in Hz. For most of the commercial drones, the fundamental frequency is at the range of 150-400Hz depending on their motor specifications [37]. The drone body is relatively a small area with many fixed pointsand it vibrates at a frequency around 3000-4000Hz, and the intensity of this noise may not be prominent for smaller drones [37]. Different models of drones have different KV ratings for the stepper motors that alter the fundamental frequency and its harmonics. Also, different drone sizes and physical shapes differentiate the body vibration frequency. These changes are perceived as distinguishable sounds in different drone models.

Figure 2 shows the power spectrum of the *DJI Inspire 1* drone. As can be seen from the figure, the fundamental frequency of the propellers generated sound is at 136 Hz. Also, the higher peaks after the fundamental frequency show the harmonic response. The small peak in the frequency plot at approximately 3660Hz is due to the mechanical vibration of the drone body. This *DJI Inspire 1* uses the DJI 3510H motor having the KV rating of 350rpm/V, and has the drone battery voltage of 22.2V [17, 23]. These results in the fundamental frequency at 129.5Hz and its harmonics. However, the frequency response can be slightly deviated specifically due to the Doppler effect, which is a result of the drone's movements and the wind. According to the frequency spectrum in Figure 2, the fundamental frequency of 136Hz has a 5% difference to the calculated fundamental frequency (129.5Hz). This is a result of the Doppler effect as the drone is flying towards the observer.

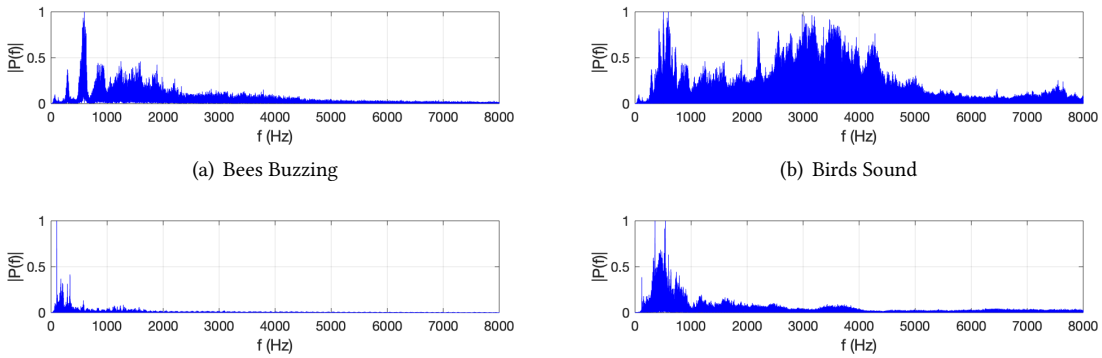
Figure 3 shows the power spectrums of a few other mechanical sounds, and Figure 4 shows the power spectrums of a few most common natural sounds. All of these sound signals have prominent energy components in the frequency range where the drone sound is also prominent (0-8kHz). However, we observe that there are finer level frequency characteristics that all the drone sounds generally have yet not so common among other mechanical and natural sounds. These prominent features of drones include equally spaced energy peaks representing the harmonics of a fundamental frequency in the range of 0-2kHz (as the operating rpm of the motors attached to propellers are different from drones) and a considerable amount of energy in the range of



(a) Airplane (Jet Engine) (b) Helicopter

(c) Washing Machine (d) Hair Dryer

Fig. 3. Power spectrum of different mechanical sound.



(a) Bees Buzzing (b) Birds Sound

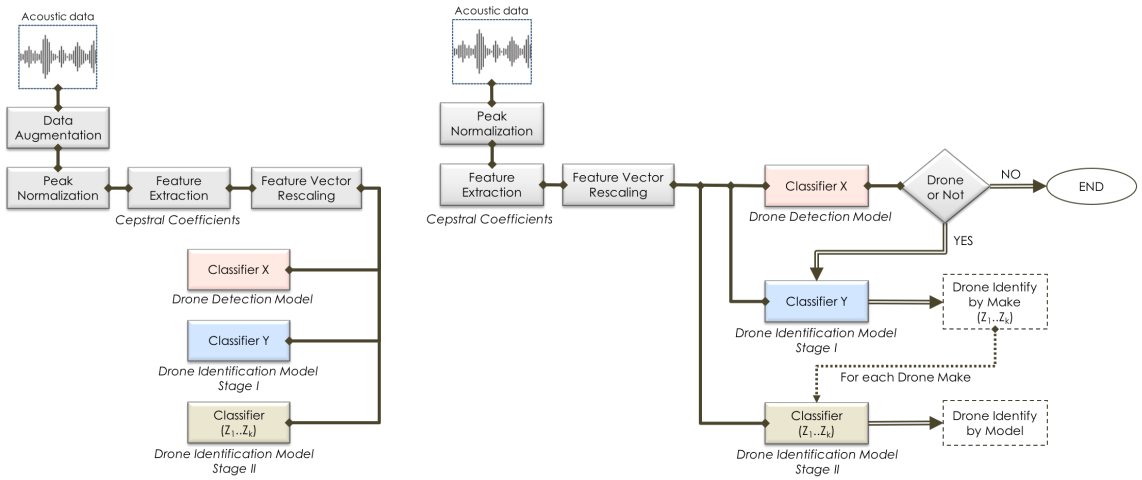
(c) Calm Environment (d) Human Talking

Fig. 4. Power spectrum of different natural sound.

0-8kHz. None of these other mechanical sounds considered here have these features as the operating rpm of the motors are different from drones. Also, when compared to the buzzing sound of bees, which is one of the closest sound to the drone sound, the prominent differences are present in the lower frequencies 200Hz-500Hz [32]. However, classifying sound by characterizing the power spectrum can lead to ambiguities and may not guarantee the best performance. On the other hand, features obtained from Mel-Frequency Ceptral Coeffienents (MFCCs) that models the human auditory perception have proven performance in audio classification. Therefore, in DronePrint we extract the MFCC features and train neural networks to artificially create the environment of how a human classifies these sounds.

4 DRONEPRINT SYSTEM

In this section, we present the DronePrint system architecture. The schematic overview of the training and testing pipeline of DronePrint is illustrated in Figure 5. After pre-processing data, we train classifiers for drone detection, drone make and model identification. We describe each primary step involved in the remainder of this section.



(a) Training (b) Testing

Fig. 5. System Architecture

4.1 Data Augmentation

Even in an environment with minimal noise, the microphone capturing acoustic signals for the same drone type can vary with different operational conditions, e.g., different manoeuvres; different distances to the microphone; and hardware modifications such as inserting low noise propellers and carrying additional weight. These differences are reflected in the power spectrum with frequency alterations and/or different power levels.

To incorporate the variations of the frequency due to the differences in the drones' operation conditions in the DronePrint training, we augment the *training* and *validation* samples in DronePrint using *frequency warping*, a technique that has shown good results in comparable audio applications [11, 31, 33, 60]. Each acoustic sample of the *training* and *validation* dataset is re-sampled such that $f_i = \alpha f$. We selected the α value ranging from 0.8 to 1.2 with 0.02 steps. Therefore, each sample goes through the re-sampling process 21 times, which gives 21 fold increment to the dataset. In general, data augmentation is done with both amplitude scaling (incorporate the different power levels) and frequency warping techniques [11]. However, in DronePrint, we do not perform the common practice of augmenting the data with amplitude scaling. Instead, we effectively nullify the amplitude variations by performing time-domain peak-normalization detailed in Section 4.2.

4.2 Peak Normalization

The various power levels in the samples can be addressed by augmenting the training samples with amplitude scaling. However, including the whole span of the amplitude differences is impossible and will prohibitively increase the augmented dataset size. We eliminate the need of having training samples with different time-domain amplitudes by nullifying this amplitude dependency in the signals with *peak normalization*. We normalize all the samples to the maximum amplitude sample within the selected time frame. To the best of our knowledge, none of the similar types of work has replaced the need for augmenting data using amplitude scaling with time-domain amplitude-normalization. We select a frame size of 200ms with no overlap. We assume that the external factors such as distance do not change much within such a short time. Also, we consider no overlap because frame overlap is majorly important in applications such as speech recognition. In these applications,

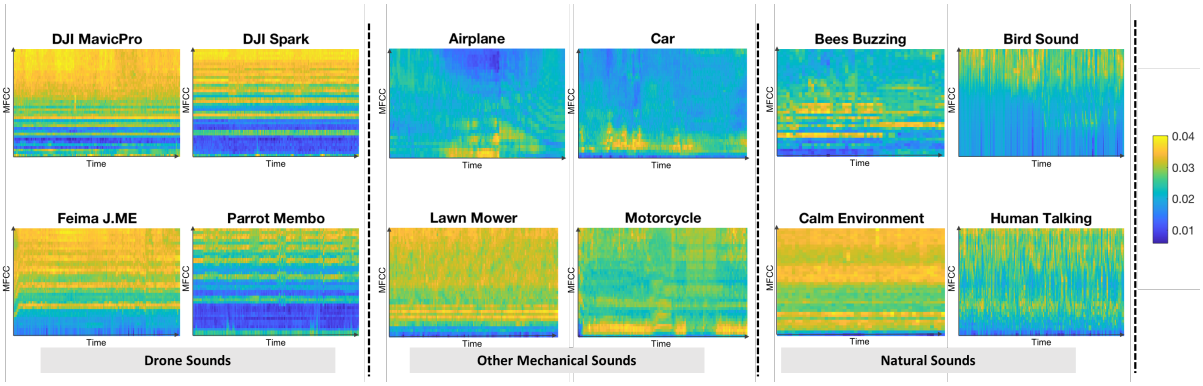


Fig. 6. Example feature matrices of different sounds

frame shift determines the particularity of information about speech dynamics where the lower the frame shift, the more details we can retrieve about speech dynamics. In our application of drone sound analysis, we do not need to analyze the dynamics of the sound, and therefore, uses non-overlapping windows. We perform peak normalization for each frame. We describe the impact of peak normalization in the features in Section 4.4.

4.3 Feature Extraction

For each frame, we calculated Mel-Frequency Cepstral Coefficients (MFCC) that are commonly used in audio analysis. As detailed in Section 3.2, the majority of the signal energy for drones remain in the range of 0-8kHz, and therefore, we focused on 0-8kHz. Specifically, we used the commonly used Mel-scale in audio analysis to capture the comparatively higher energy in lower frequencies compared to higher frequencies in the range and calculated 40 MFCC features.

In Figure 6, we show the calculated MFCC for some drones, other mechanical sounds, and some common sounds. The prominent differences of drones with other sound are i) periodic high energy levels in specific MFCC coefficients (as highlighted by horizontal lines), and ii) higher values in higher MFCC coefficients. The horizontal lines represent the propeller sound that consists of the fundamental frequency and its harmonics (cf. Section 3.2). Next, the higher values in higher coefficients are results of both the harmonics and also the high energy in higher frequencies that are generated by the mechanical vibration of the drone body. As shown in Figure 3 and 4, many of the considered sounds (e.g., airplane, bees buzzing, bird sound, human talking) have prominent energy in the frequency range of 0-8kHz. However, the calculated MFCCs of drones have the above distinguishable patterns to them.

4.4 Feature Vector Re-scaling

Figure 7 shows two signal frames that were recorded at a distance of 10m and 50m. In one case, there was minimal background noise, whereas in the other case there is an instantaneous noise (bird sound). As shown in Figure 7, when there is minimal noise, peak normalization (cf. Section 4.2) brings the time-domain signal frames into similar/comparable range. To observe the impact of time-domain peak normalization in the feature space, we calculate the vector of element-wise distances in the two feature vectors of the same class. Each element of the distances vector is calculated by taking the absolute difference between the elements at a particular index of the two feature vectors and represented as a percentage to the element in the first feature vector. Then we calculate the average and the standard deviation of the elements of the distances vector and shown under each

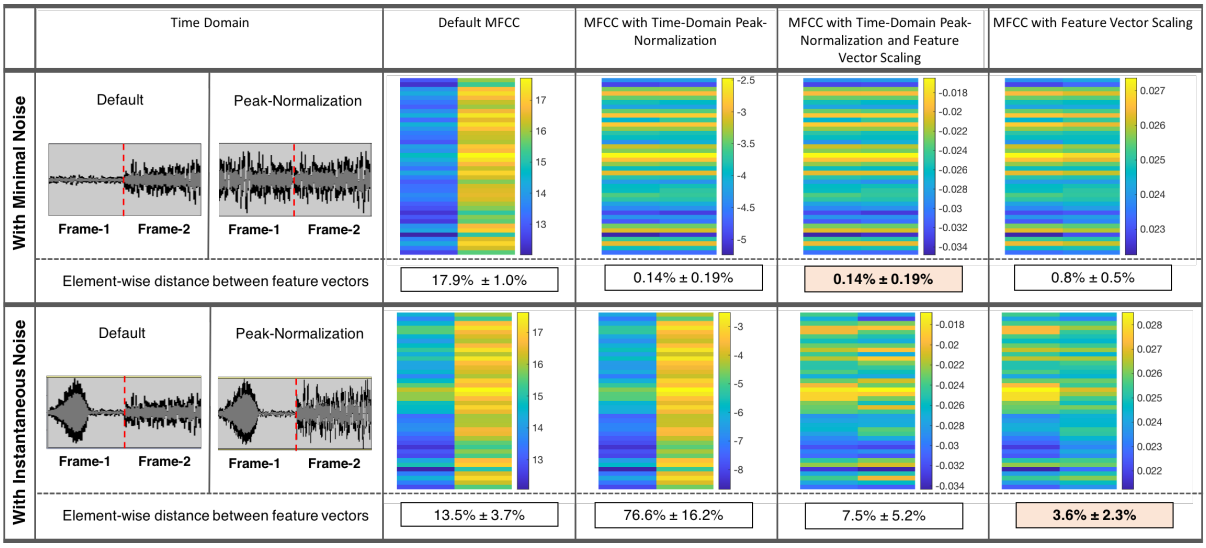


Fig. 7. Feature vector re-scaling.

feature space in Figure 7. The average element-wise difference indicates how sparse the two feature vectors are. Higher the average element-wise difference is, more sparse the two feature vectors are. MFCC calculations with time-domain peak normalization have only 0.14% ± 0.19% distance between the two feature vectors where the default MFCC has 17.9% ± 1% distance. Therefore, the time-domain peak normalization reduces the difference between the two feature vectors.

However, time-domain peak normalization technique will still be problematic when there are instantaneous and high peak noise such as birds sound is present, since the peak normalization of the frame with high noise can result in even more significant difference in the feature vectors. We show an example case in Figure 7 where there is instantaneous noise. In this case, MFCC with time-domain peak normalization provides much a larger difference as 76.6% ± 16.2% where the default MFCC calculations provide 13.5% ± 3.7%. This is because that the time-domain peak normalization makes the drone sounds in the two frames more sparse, and therefore, the drone sound features are less prominent in the feature vector. As a solution, we re-scale the feature vector where we represent the feature vectors as unit vectors. Unit vector scaling re-scales the cepstral coefficients feature vector v by dividing by its Euclidean norm, i.e., L_2 norm; $v^{'} = v/||v||$. Though feature vector re-scaling is generally used for efficient training of machine learning models, to the best of our knowledge it is not used for a targeted objective like this in any of the acoustic-based classification models. We observe that when instantaneous noise is present, the least element-wise distance between feature vectors is given by MFCC with unit vector scaling and without time-domain peak normalization. However, this is different with minimal noise where the least distance is given with the time-domain peak normalization. To explain this, next we describe the MFCC calculation procedure and how the time-domain peak normalization gives adverse effect with instantaneous peaks.

Assume that the two signals in two different frames are $X_1(t)$ and $X_2(t)$ where the first signal is recorded when the drone is far away than the second signal. With minimal noise conditions, the first signal's time-domain amplitude is multiplied by a constant k to get the second signal. i.e. $X_2(t) = k * X_1(t)$. As the first step in MFCC calculation, the Fast Fourier Transform (FFT) is applied to the time-domain signal and calculate the frequency spectrum. Since FFT is linear, the frequency domain representations, $Y_1(f)$ and $Y_2(f)$ can be expressed

as $Y_2(f) = k * Y_1(f)$. In the second step, Mel filterbanks are applied to the frequency spectra and sum the energy in each filter. Assume these calculated filterbank energy vectors are $Y_1[E]$ and $Y_2[E]$. The linear relationship will still be preserved between the energy vectors such that $Y_2[E] = k * Y_1[E]$. The final step of the MFCC calculation is taking the logarithm of the filterbank energies and take the Discrete Cosine Transform (DCT) to get the feature vectors. Assume the feature vectors for the two signals are $Z_1[C]$ and $Z_2[C]$. Although the DCT is a linear transformation, because of the logarithmic function, the final step of MFCC calculation is not linear as shown in (1). Since $Z_1[C] \neq k * Z_2[C]$, the unit vector scaling will not nullify the effect of k, and the distance between the vectors is higher with higher k values.

$$Z_2[C] = DCT(log_{10}(Y_2[E])) = DCT(log_{10}(k * Y_1[E])) = DCT(log_{10}(k) + log_{10}(Y_1[E])) \neq k * Z_1[C] \quad (1)$$

Since there is no prominent noise energy in the minimal-noise scenario, the time-domain peak normalization makes the signal in two frames comparable (minimizes the k values), and hence, the distance between the two feature vectors is minimal. Therefore, performing both time-domain peak normalization and unit vector scaling when there is minimal or constant and persistent noise provides the minimal distance between the feature vectors. However, in the example of instantaneous and prominent noise, the time-domain peak normalization makes the k value higher, and hence, makes the two feature vectors more distant. Therefore, when there is instantaneous and prominent noise, the sole operation of unit vector scaling provides the minimal distance between the feature vectors.

In machine learning algorithms, highly diverse data within a particular class affects the performance and decreases the ability to make accurate predictions. However, zero diversity among the samples make them duplicated and duplicated samples in training and validation makes the variance of parameter estimates improperly small. Therefore, our goal here is to minimize the diversity in the samples, while not producing duplicate samples with zero diversity.

4.5 Classifier

4.5.1 LSTM Model. We used a Long Short-Term Memory (LSTM) network as our classifier since it has shown promising results in audio signal classification tasks with both speech [27] and non-speech [10] data as well as other time-series data [49]. The LSTM architecture we used as illustrated in Figure 8 consists of 10 time steps, two stacked LSTM layers. We used a hidden state size of 32.

We trained the model using the Adam optimizer with a learning rate of 0.0001 for 500 epochs. The validation dataset is used to evaluate the performance of the trained model on that validation dataset at each epoch. Furthermore, we select the trained model having the lowest validation loss. At the inference time, we average three predictions of three consecutive windows that cover a total duration of 6 seconds. Then we select the class having the highest predicted probability. All three classifiers, X, Y, and Z have the same architecture and followed the same training procedure.

4.5.2 Drone Detection in Open-set Scenarios. Our drone detection classifier (*Classifier X*) detects whether the input acoustic signal is from a 'Drone' or 'Non-Drone'. We used two techniques to design this classifier; *Background Class* and *Modified OpenMax* where we further improved the performance of OpenMax method initially proposed for image classification in [5]. Under background class, we train a binary classifier that differentiates drone sounds from non-drone sounds. We used samples from other mechanical sounds such as *vehicle sounds in a busy road*, *Airplane*, *Hair dryer*; and common non-mechanical sounds such as *human voice*, *Raining sound*, and *Birds sound* as training samples for the background class.

However, as the performance of the background class method is restricted by the availability of samples from non-drone signals at training time, we next explored OpenMax method proposed by Bendale and Boult,

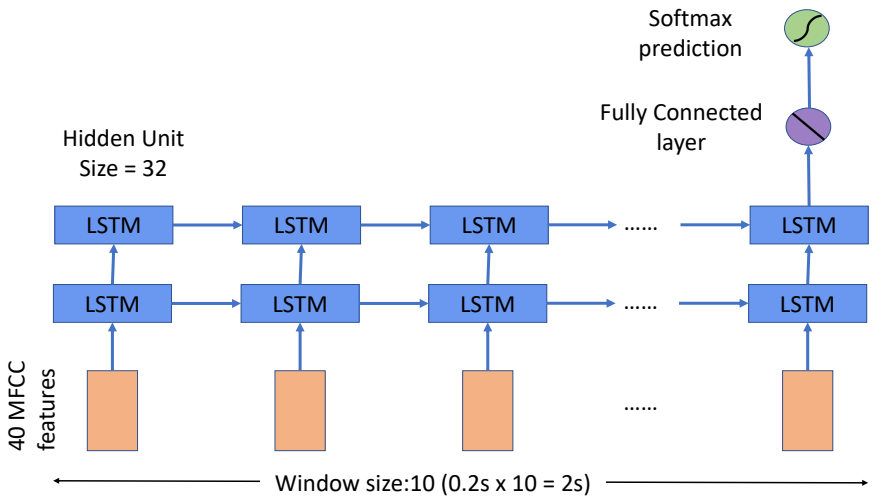


Fig. 8. LSTM model.

which does not require samples from non-drone signals at the training time [5]. OpenMax is based on the intuition that the values from the penultimate layer (layer before softmax) of a deep neural network are not just independent per-class score estimates, but provide a distribution of how classes are related to each other. Accordingly, OpenMax uses the concept of *meta-recognition* on the scores from the penultimate layer to estimate if a given input is far from known training data. Since the classification task we consider here is binary, and OpenMax does not use open-set data at training time, we used a simple closed-set classifier that classifies 13 drone classes as the underlying deep learning model used by OpenMax.

At inference time, for each sample, OpenMax will output a vector of $n + 1$ elements which represent the probabilities of the sample belonging to one of the $n + 1$ classes (n known drone classes and the non-drone class). At the final step of the original OpenMax model, if the model is confident enough about the prediction; i.e. if the probability of the predicted class is higher than a predetermined threshold, the prediction from the model is accepted as it is. If the model is not confident about the prediction and if the probability of the predicted class is less than the predetermined threshold, the sample is considered to be from an unknown class.

In order to further improve the performance of the OpenMax method, we assign a separate threshold value for each known class. The intuition behind this choice is that as the distribution of the output from OpenMax varies for different drone classes, it is possible for samples for some classes to have relatively high probability values corresponding to the most probable class while samples from other classes have relatively lower probability values and hence, having class-wise thresholds that suits the distribution of the probability value for the most probable class for each class would be more effective. At the inference time, we sum the prediction probabilities of all the drone classes and consider as the prediction probability of being a drone.

4.5.3 Drone Identification. If a sound is detected as a drone by Classifier X, it is then passed to drone identification classifiers to identify its make and the model subsequently. First, we use the *Classifier Y* to identify the drone by its make, e.g., *DJI, Parrot, Autel, Femia, MicroKopter*. Next, we use *Class-wise Classifiers* Z_1,...,Z_k for each class of drone make identified in *Classifier Y*. These classifiers identify the drone model. All these classifiers in the drone identification are multi-class classifiers.

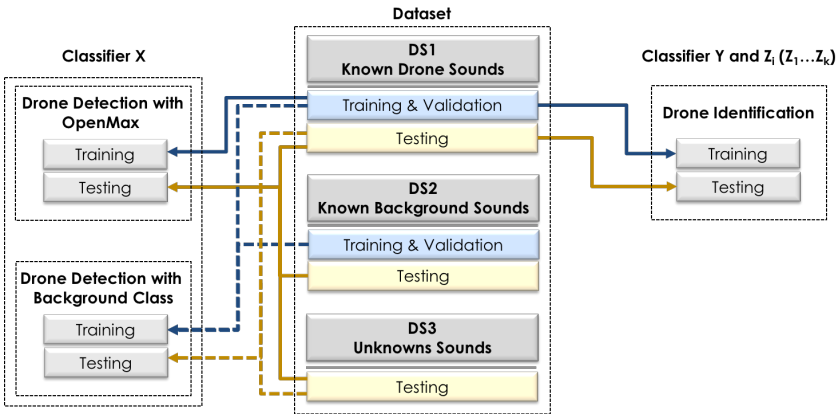


Fig. 9. Usage of datasets

In Figure 9 we show a summarized view of how we use our three datasets to build various classifiers in the DronePrint pipeline. For instance, the drone detection classifier with the background class is trained and validated with *training* and *validation* data of *known drone sounds* and *background sounds*, and tested with *testing* data of all three datasets. The drone detection classifier with OpenMax is trained and validated with *training* and *validation* data of *known drone sounds*, and tested with the *testing* samples of *known drone sounds* and *unknown sounds*. Finally, drone identification classifier is trained and validated and tested with *training*, *validation* and *testing* data of *known drone sounds*.

5 RESULTS AND ANALYSIS

In this section, we present the performance results of DronePrint's drone detection and identification models. We also compare the performance of DronePrint with other similar work.

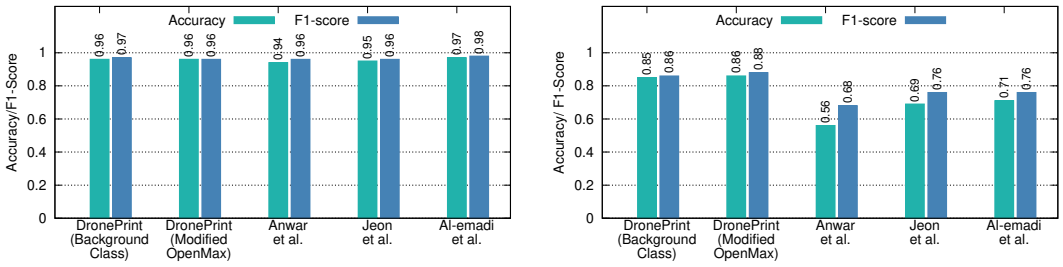
5.1 Evaluation

Drone Detection: We compare DronePrint detection models with three other similar work that use MFCC for the features and use a background class to enable open-set detection. The implementation details of these models are shown in Table 4. Machine Learning features used in DronePrint is different from other methods as we implement MFCC with time-domain peak normalization and feature vector re-scaling where other methods implement default MFCC. We re-implemented all the models using the given details. We performed two sets of testing, 1) *known* classes (closed-set scenario), and 2) *unknown* classes (open-set scenario). The closed-set scenario is tested with *testing* samples of *known* (DS1), and *background* (DS2) datasets in background class method, and *testing* samples of *known* (DS1) in OpenMax method. The open-set scenario is tested with *testing* samples of *unknown* (DS3) in both methods of drone detection.

Drone Identification: As detailed in Section 4.5, DronePrint performs the drone identification in two stages; drone make identification followed by the drone model identification. The 13 drones in our dataset are from five different drone makes, *Autel*, *DJI*, *Femia*, *MikroKopter*, and *Parrot*. As shown in Table 1, *DJI* and *Parrot* have six and four drones, respectively, and other classes have only one drone. Hence, the augmented dataset (cf. Section 4.1) is not balanced for the five classes of drone makes. Therefore, to get a balanced dataset in this stage, we augmented the *training* and *validation* samples of each drone type differently. Specifically, each sample was scaled differently along in the time axis; 8 times for the 6 drones in *DJI*; 12 times for the 4 drones in *Parrot*; and 48 times

Table 4. Comparison with related work.

Details	DronePrint	Anwar et al. [3]	Jeon et al. [34]	Al-emadi et al. [1]
Features	MFCC with time-domain peak normalization and feature vector scaling	MFCC	MFCC	MFCC
	0-8kHz (40 Filter-banks)	0-20kHz (13 Filter-banks)	0-1.5kHz (40 Filter-banks)	0-8kHz (40 Filter-banks)
Frames	200ms with no overlap	Not Specified (used 200ms with no overlap)	240ms with no overlap	100ms with 10% overlap
Classifier	LSTM with 2 Layers	SVM with Cubic Kernal	LSTM with 2 Layers	CRNN
	10 time-steps	-	1 time-step	101 time-steps



(a) Known Classes (b) Unknown Classes

Fig. 10. Performance of Drone Detection Model

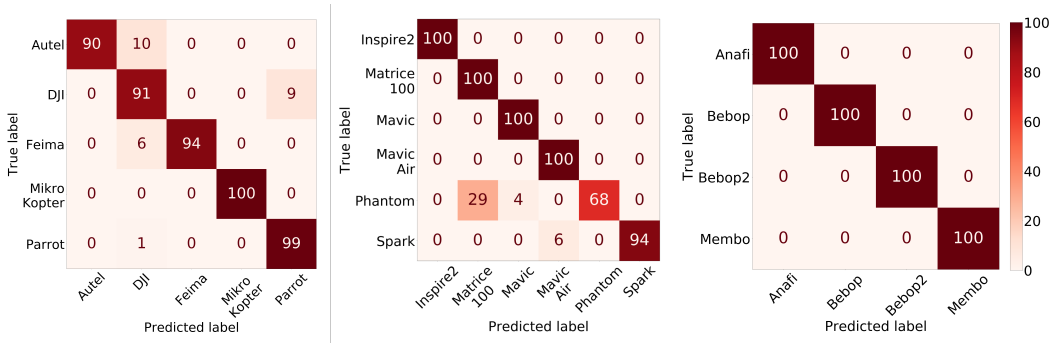
for *Autel*, *Feima*, and *MikroKopter*. Also, only a part of the *testing* data is used from each of the *DJI* and *Parrot* drones to have a balanced testing dataset.

Next, in the drone model identification, the drone make classes that have two or more drone models are classified. Therefore, the signals identified as *DJI* or *Parrot* are passed to this classification. Two separate models are trained with the six drone models in *DJI*, and four drone models in *Parrot* with the data augmentation detailed in Section 4.1.

Performance Metrics: We use the two most commonly used performance evaluation metrics in classification models, *Accuracy* and *F1-score* [8]. In the multi-class classification model in drone identification, we treat all classes equally where we have an equal number of per class samples in training and also in testing. Therefore, we compute the macro-averages of *Accuracy* and *F1-score*. The *Accuracy* per each class is calculated independently and then take the average to compute the macro-averages of *Accuracy*. Therefore, *Accuracy* in the multi-class classification is calculated as the sum of the main diagonal elements in the confusion matrix divided by the sum of all the elements. *F1-score* is the harmonic mean of the precision and recall varying between the best score at 1 and worst score at 0. Similar to average accuracy, we compute the macro-average of *F1-score* using the macro-averages of precision and recall.

5.2 Drone Detection

Figure 10(a) shows that all models equally performs for *known classes*, i.e. closed-set scenarios. *Accuracy* for all the models is in the range of 94% to %97 and *F1-score* is in the range of 0.96 to 0.98. However, a drone detection model is realistic when it can perform better not only with previously seen classes but also with unseen



(a) Classifier Y (b) Classifier Z: DJI (c) Classifier Z: Parrot

Fig. 11. Confusion Matrices of Drone Identification

classes. Otherwise, the models' real-world use is limited. Therefore, we evaluate the open-set classification performance of the models. The two proposed DronePrint models well outperform the related work models in open-set identification as shown in Figure 10(b). Modified OpenMax method provides marginal improvement over Background class method with *Accuracy* of 86% vs 85% respectively and *F1-score* of 0.88 and 0.86 respectively. All the models detect the presence of an unknown drone sounds correctly where the *recall* is high in value. However, the unknown non-drone sounds are also detected as drone in other methods, and therefore, having lower *Accuracy* and *F1-score*. We observe that the lowest performance is provided by the SVM based method whereas all the other methods implement LSTM based methods. The detection performance for the known classes is comparable with the reported results in the respected work [1, 3, 34]. However, the performance for the unknown classes are reported only in [34], and those results are also comparable to the results we get in these tests. Overall in the drone detection, DronePrint with modified OpenMax achieves 92% *Accuracy* and 0.94 *F1-score*; and DronePrint with Background class achieves 91% *Accuracy* and 0.93 *F1-score*. Also, our experimental results show that the performance of the original method of OpenMax is improved with our modifications in DronePrint. Specifically, the *accuracy* is improved by 5% and the *F1-score* is improved by 0.06.

5.3 Drone Identification

Figure 11(a) shows the confusion matrix of the *Classifier Y*, i.e. drone make identification. The percentage of the instances in the class are shown in each row of the confusion matrix. The overall *Accuracy* is 95%, and *F1-score* is 0.95. For each class, more than 90% are in the diagonal in the confusion matrix. We also tested feature concatenation with combining time-domain (*Zero Crossing Rate*, *Energy*, and *Entropy of Energy*) and spectral features (*Spectral Centroid*, *Spectral Spread*, *Spectral Entropy*, *Spectral Flux*, and *Spectral Rolloff*). However, despite the marginal improvements in certain classes, overall accuracy was reduced to 82% and 74% respectively for temporal and spectral features concatenation.

Figure 11(b) and 11(c) depict the results of individual drone identification with *Classifier Z* for two manufacturers. For *DJI*, most of the drone models are classified with 100% accuracy, except for *Phantom* and *Spark* drone models. For *DJI*, the overall *Accuracy* is 95%, and *F1-score* is 0.96. On the other hand, all models were correctly identified for *Parrot* as shown in Figure 11(c).

In the cascaded multi-stage drone identification, the performance of classifying a drone as the correct make in *Classifier Y* is multiplied by the performance of classifying a drone as the correct model in *Classifier Z* to get the final performance of the particular drone model identification. As such, the calculated overall *Accuracy* is 90%,

and the *F1-score* is 0.91 in this multi-stage drone identification. Also, these results show that the DronePrint model that is trained and validated with online sourced data performs well when tested with the experimentally collected data. This is exhibited in the cases of *Parrot: Bebop 2*, *DJI: Spark*, *DJI: Mavic Pro*, *DJI: Matrice 100*, and *DJI: Phantom 4 Pro*. The online sourced data would have recorded with different microphones that would have different frequency characteristics; also, these audios could have been through multiple audio processing steps. However, our results indicate that drone-specific acoustic signatures are still preserved even in the processed online data collected from highly assumptive environments. Also, we compared DronePrint identification model with existing work. Only the work in [1] classifies the commercially available drones using the acoustic signals. However, they classify only two drone models. For comparison, we built the model in [1] and trained it as a 13 class classifier. The results of this classification give the overall *Accuracy* of 76% and *F1-score* of 0.76. This is ∼15% lower than DronePrint in both performance metrics. The reported results in [1] for the two drone model classification are 92% *Accuracy* and 0.92 *F1-score*. However, when the number of classes in the classifier is increased to 13 drone classes, the performance is reduced drastically.

6 ROBUSTNESS OF DRONEPRINT

Next, we conduct several experiments to demonstrate the robustness of DronePrint. We evaluate the model in different operational conditions; i.e. i) with the most common hardware modifications, ii) with Doppler effect, and iii) under different SNR conditions. Moreover, we evaluated DronePrint's drone detection performance for an extended period through an emulation by replaying various natural and artificial sounds. During these experiments, we use DronePrint with the background class for drone detection and the DronePrint classifiers Y and Z for drone identification.

6.1 Effect of Hardware Modifications

6.1.1 Scenarios. There are many instances where the drone owners modify the drone hardware. For example, one of the most popular modification is the use of low-noise propellers. As explained in Section 3.2, a major part of the drone sound is due to the propellers, and there are many ongoing projects aiming at introducing low-noise propellers both in academia and industry [45, 82]. The overall shape, including the tip of the propellers, are modified in these low-noise propellers, and these modifications reduce the acoustic signal strength. For example, the commercially available low-noise propellers of *DJI Phantom 4 Adv. Pro* reduces the sound strength by 8dB [45].

As such, it is necessary to assess the robustness of DronePrint to such modifications. Also, there are other common modifications such as custom-built propellers, adding propeller guards, and upgrading the camera and the gimbal, which can potentially change the acoustic signature. We next consider these most common hardware modifications, and evaluate whether DronePrint can still detect and identify drones without re-training.

6.1.2 Experiment. We extracted the audio component of YouTube videos that experiment with *DJI Phantom 4 Pro* for four scenarios; i) commercially available low-noise propellers, ii) custom-made low-noise propellers, iii) propeller guards, and iv) upgraded camera and the gimbal. We compare all these modifications with the signal with no modifications (i.e., the original drone). We list all the YouTube files used in these experiments in Appendix B. We used 12 seconds of sound traces for each scenario. Then, we tested these signals with our DronePrint detection and identification classifiers.

6.1.3 Results. We show the frequency spectrum, feature space and the prediction for each frame in Figure 12. As can be seen from the figure, the frequency spectrums for the low-noise propellers are distinguishable from the spectrums of original propellers where the original propellers have higher energy levels. However, the feature space does not contain these differences as a result of the time-domain peak normalization and feature vector

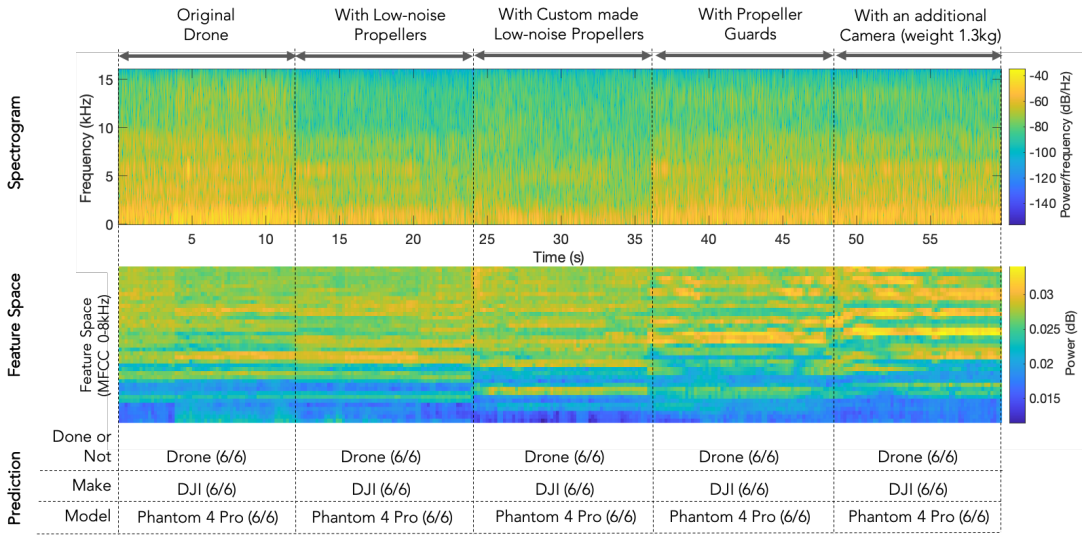


Fig. 12. Effect of Hardware Modifications

scaling (cf. Section 4.4). Therefore, DronePrint can correctly detect the drone as well as identify its make and model. Since the effect of using low-noise propellers is the same as the drone is flying at longer distances, the results are also applicable for the scenarios of drone operating at different distances. Figure 12 also shows that the additional weight and the changes in the drone in the scenarios of adding propeller guards and mounting a new camera, results in slightly higher values in the higher coefficients in the feature space. Nonetheless, DronePrint still can detect and identify the drones with 100% accuracy.

6.2 Doppler Effect

6.2.1 Scenario. Doppler effect in sound signals occurs when the sound source or the observer moves relatively towards or away from each other. This effect shifts the signal frequency up or down. The scaling of the frequency with the Doppler effect is shown in Equation (2), where the velocities of sound V_{sound}, wind V_w, observer V_{obs}, and source V_{src} are given [4]. As shown in Equation (2), the signal frequency is not changed as long as the observer or the sound source is not moving. However, the movements of the observer or the source scale the frequency response by a factor. Doppler effect is applicable to DronePrint when a drone is flying away or vice versa from the microphone when the audio signal is being recorded.

$$f^{'} = f_0 . \frac{V_{sound} \pm V_w \pm V_{obs}}{V_{sound} \pm V_w \pm V_{src}} \tag{2}$$

6.2.2 Experiment. We recorded the sound signal emanated from a *DJI Phantom 4 Pro* drone when it is flying towards the observer, i.e., microphone, passes the microphone and finally flying away. The drone was flying at its maximum speed in the *sports mode* (20m/s) up to a distance of 60 meters. This resulted in a 6s length audio signal. Next, we passed these traces through the detection and identification classifiers.

6.2.3 Results. We show the result of this experiment in Figure 13. DronePrint was able to make correct predictions for both drone detection and drone identification due to our data augmentation step (cf. Section 4.1).

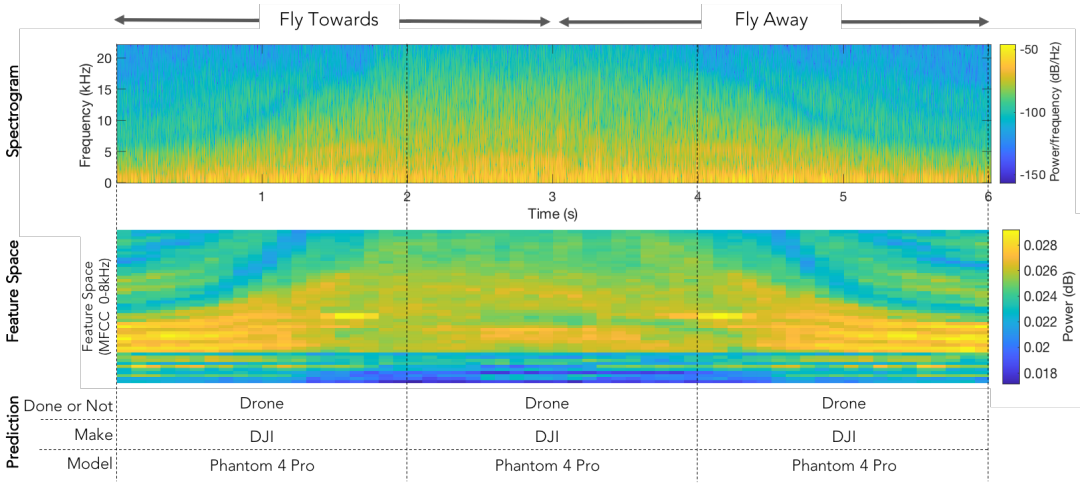


Fig. 13. Doppler Effect

Assume the sound source is moving and in a fresh breeze environment. The maximum and minimum possible frequency changes that can happen are given in Equation 3. The drone is flying at the speed of 20 m/s and we assume the maximum velocities for fresh breeze (10.7 m/s) [69]. By substituting these values in (3) as $V_{sound} = 343m/s$, $V_w = 10.7m/s$, and $V_{src} = 20m/s$, the frequency can increase by 6.4% (when flying towards) and decrease by 5.6% (when flying away). As detailed in Section 4.1, we augmented the *training* and *validation* samples with frequency warping that scaled along in the time axis by values ranging from 0.8 to 1.2. This is ±20% of frequency scaling, which includes the frequency variations in this particular scenario. Moreover, our training samples consist of the drone free flying sound signals with the Doppler effect. Therefore, DronePrint can correctly detect and identify even at the presence of Doppler Effect.

$$f'_{Max} = f_0 \cdot \frac{V_{sound} - V_w}{V_{sound} - V_w - V_{src}}; \qquad f'_{Min} = f_0 \cdot \frac{V_{sound} - V_w}{V_{sound} - V_w + V_{src}}; \tag{3}$$

6.3 Effect of Noise

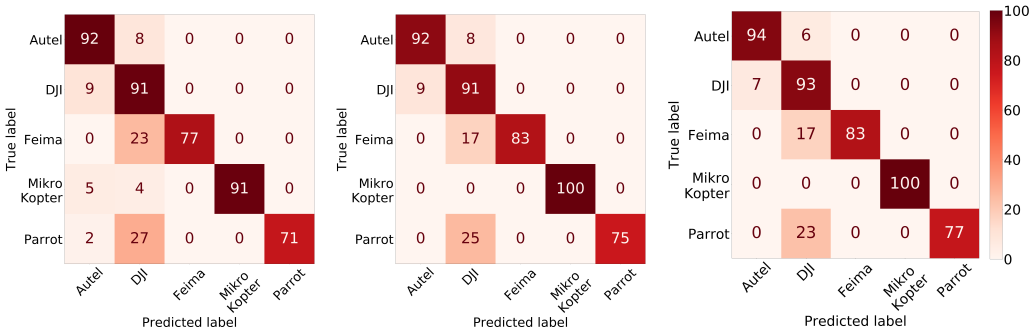
6.3.1 Scenario. In many realistic drone detection and identification scenarios, background noise will be present with different relative strengths compared to the desired signal. This is formally measured by the signal-to-noise ratio (SNR), which is the ratio of signal power to the noise as shown in Equation 4. A ratio higher than 1:1 (i.e. greater than 0 dB) indicates the signal strength is higher than the noise signal. We evaluate the performance of DronePrint when it is operating in environments where there is continuous background noise.

$$SNR_{dB} = 10log_{10}\frac{P_{signal}}{P_{noise}} \tag{4}$$

6.3.2 Experiments. Earlier we trained and validated DronePrint under minimal background noise. To improve the performance at lower SNRs, here we re-trained each classifier of DronePrint by augmenting the *training* and *validation* samples with noise incorporated samples. We used 80% of the samples with minimal noise, i.e., the earlier training set, and the other 20% is the noise integrated samples. We prepared these noise integrated samples as follows.

Table 5. Performance of Drone Detection Model with Noise Incorporation (unknown noise).

Metric	Known Classes			Unknown Classes		
	SNR=1dB	SNR=3dB	SNR=10dB	SNR=1dB	SNR=3dB	SNR=10dB
Accuracy [%]	94	94	95	83	85	86
F1-Score	0.96	0.96	0.96	0.85	0.87	0.88



(a) SNR = 1dB (b) SNR = 3dB (c) SNR = 10dB

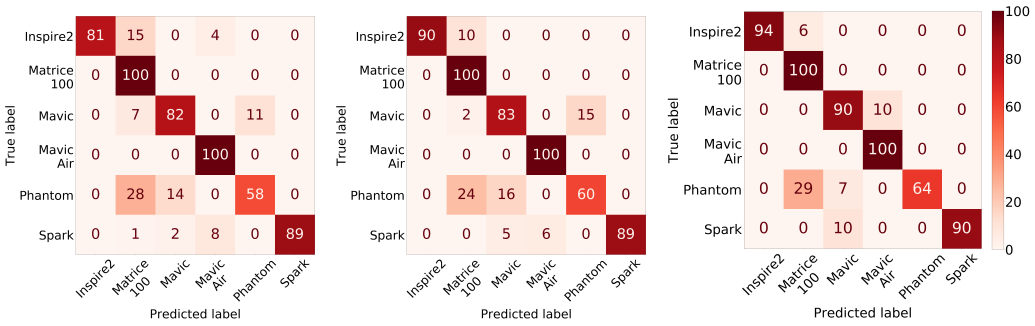
Fig. 14. Confusion matrices of Classifier Y drone make identification with noise

For the *Classifiers X* and *Classifiers Z*, we picked 37s of each training signal in Table 1. As *Classifiers Y* has different number of drone models in each class, we use different duration from each drone. Such that, we used 15s from each *DJI* drone, 20s from each *Parrot* drone and 85s each from all other drones. Then, we mix and render each signal with the seven background sounds of the same length listed in Table 2. We used the *Audacity* software to mix and render the signals, which appears as one combined audio file.[4] Here, we used three instances from each signal where the average power of the noise is 1dB, 3dB and 10dB lower than the testing signal's average power, where the signal is 1.25, 2, and 10 times stronger than the noise, respectively.

Next, for testing, we consider two scenarios, i) drone sound with an unknown noise, i.e., background sound that was not incorporated in the training, and ii) drone sound with a known noise, i.e., background sounds listed in Table 2. For each scenario, we consider three situations where there is continuous background noise in different SNRs, (1dB, 3dB and 10dB). We use the lawn-mower sound (audio file in Table 3) as the unknown noise and human talking sound as the known noise (testing audio file in Table 2). Sound of the lawn-mower is a common sound in outdoor environments and also having a closer feature profile to the drones as detailed in Section 4, but detected as a non-drone sound by DronePrint detection model in Section 5. As earlier, each testing signal is mixed and rendered with this noise signal using *Audacity* software to appear as one combined audio file.

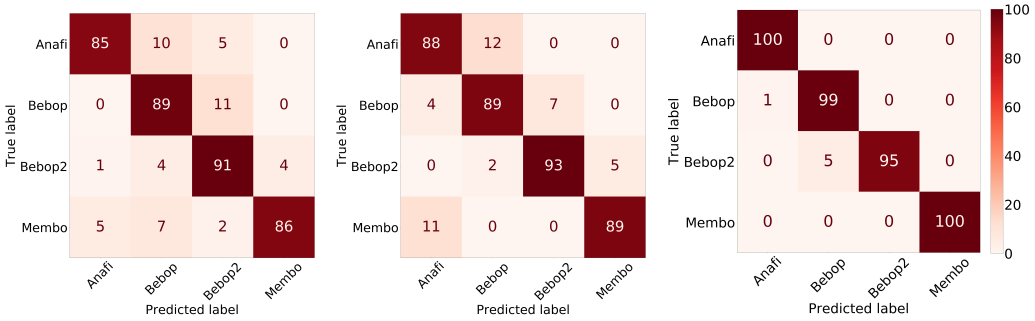
6.3.3 Results. With the scenario of with human talking noise, drone detection and identification models provide the same performances regardless of the SNR value. The *Accuracy* and *F1-Score* of the drone detection for *Known Classes* are 96% and 0.96 respectively; for *Unknown Classes* the values are 86% and 0.87 respectively. For the drone identification, the *Accuracy* and *F1-Score* of the *Classifier Y* are 95% and 0.95; *Classifier Z:DJI* are 95% and

[4]https://manual.audacityteam.org/man/mixing.html



(a) SNR = 1dB (b) SNR = 3dB (c) SNR = 10dB

Fig. 15. Confusion matrices of Classifier Z:DJI drone model identification with noise



(a) SNR = 1dB (b) SNR = 3dB (c) SNR = 10dB

Fig. 16. Confusion matrices of Classifier Z: Parrot drone model identification with noise

0.96; *Classifier Z:Parrot* are 99% and 0.99. These performances are similar to the minimal noise testing results (cf. Section 5).

Next, we evaluate the performance with unknown noise. Table 5 shows the results of drone detection. As expected, all the performance metrics of the detection model for both the known and unknown classes have improved when the SNR increases. Also, when the SNR is 10dB, the performance metrics have equal performances to the minimal noise testing results (cf. Figure 10). We observe a similar behavior for drone make and model identification as shown in Figure 14, Figure 15, and Figure 16. For instance, The prediction accuracy of *Parrot* drone model identification improves from 88% to 99% when SNR increases from 1dB to 10dB. Some of the drone makes (*Autel*, *DJI*, and *MikroKopter*), and models (*DJI Inspire2*, *DJI Matrice100*, *DJI MavicPro*, and *DJI Spark*) are less affected by lower SNRs and are correctly identified even at 1dB of SNR. This is well below the industry standard for acoustic-based systems that require at least 10dB SNR for reliable operation [30, 76]. Note that these performances of DronePrint are achieved via augmenting the training dataset with the samples of noise incorporation.

6.4 Increased Number of Classes

6.4.1 Scenario. DronePrint provides an overall drone identification *Accuracy* of 90% as detailed in Section 5 (with 13 classes of drones). However, it is important to evaluate the performance of Droneprint in a realistic scenario of the classes of drone makes (*Classifier Y*) and the classes of drone models by each make (*Classifier* Z_i) are increased.

6.4.2 Experiment. We consider three additional drone models (*Autel: EVO II*, *Yuneec: Mantis-G*, and *Yuneec: Mantis-Q*) than the 13 drones detailed in Table 1. Specifically, we added a new drone to the existing drone make class (*Autel*), and another a new drone to the make class (*Yuneec*). Identical to the data collection process detailed in Section 3, we downloaded the audio component of these YouTube videos. We re-trained the *Classifier Y* by incorporating all three new drone models and trained new classifiers drone model classifiers (*Classifier* Z_i) 1) for *Autel* by incorporating *Autel: EVO* and *Autel: EVO II*; 2) for *Yuneec* by incorporating *Yuneec: Mantis-G*, and *Yuneec: Mantis-Q*.

6.4.3 Results. The performance of *Classifier Y* with six drone make classes is not deviated from the earlier results in Section 5 with only five drone make classes. The *Accuracy* is slightly improved with the six classes from 94% to 95%. Results for the *F1-Score* has the same value of 0.94 in both cases. This shows that DronePrint is minimally affected by the addition of a new class. Also, the results for *Classifier Z* for *Autel* and *Yuneec* shows 100% *Accuracy* and *F1-score* of 1.00. We calculate the overall *Accuracy* of the cascaded multi-stage drone identification as in Section 5.

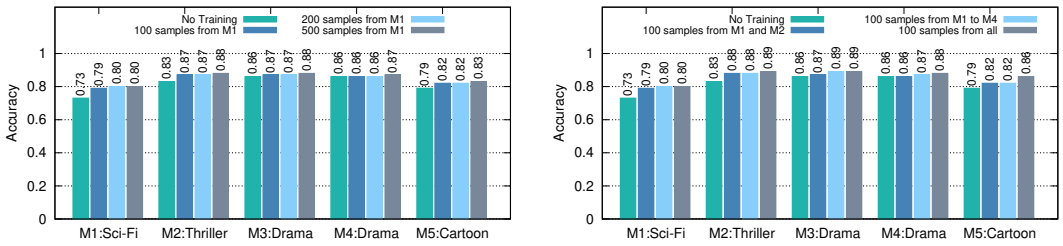
In this case, we have the *Classifier Z* for four drone make classes (*Autel*, *Yuneec*, *DJI*, and *Parrot*). Note that the *Classifier Z* for *DJI* and *PArrot* are not re-trained, and therefore, we use the same confusion matrices shown in Section 5 in the calculations. As such, the calculated overall *Accuracy* is 91%, and the *F1-score* is 0.92, which is almost the same as in the results in Section 5, in fact, the overall performance is slightly improved. The addition of new drone classes has minimal impact on the overall *Accuracy* and the *F1-score* as well.

6.5 Open-set Drone Detection

6.5.1 Scenario. Although DronePrint provides improved performance from existing drone detection models in the open-set scenarios (cf. Section 5), it is important to evaluate the performance in realistic scenarios of ever-changing various natural and machinery sounds for an extended time period.

6.5.2 Experiment. We considered the realistic scenarios of changing sound conditions with the sound signal of movies that contains various natural and artificial sounds. We used different types of movies to create different acoustic environments, i) *Science-Friction* and *Thriller* types that can have more machinery sounds (Movies; 'Taking Earth' and 'Kidnap'), ii) *Drama* type having more human voice and day-to-day indoor and outdoor sounds (Movies; 'Five Feet Apart' and 'McKenna'), and iii) *Cartoons* containing more music and various sound effects ('Masha and the Bear'). The length of these movies are 100 minutes each, and none of these have drone sounds. First, we tested the DronePrint detection model in an open-set scenario where no training samples are taken from any of the movies. Then we conducted experiments by adding training samples from the movies and calculated the accuracy improvements. We conducted two separate sets of experiments; i) we added 100, 200, and 500 number of samples; 20s, 40s, and 100s of length respectively, as *training* samples labeled as non-drone and selected at a random time instance from one of the movies (M1); ii) we added 100 samples (20s in length) selected at a random time instance as *training* samples that labeled as non-drone from M1 and M2; from M1 to M4; and from all.

6.5.3 Results. As shown in Figure 17, when no training samples are taken from any of the movies, the average *Accuracy* is 81% (ranging between 73% and 87% for different movies). This means close to 80% of time DronePrint



(a) Effect of the Number of Training Samples

(b) Effect of the Number of Movies Included in Training

Fig. 17. Accuracy in open-set Testing

correctly discards non-drone sounds over a total of 100 minutes (i.e. total of 1000 predictions based on 6 second windows)

Then, when different numbers of random *training* samples are added, we observe that the accuracy is improved by ∼5% in M1 and by ∼1% to ∼3% in other movies. Next, when 100 number of random *training* samples are added from different sets of movies, we observe ∼4% to ∼5% accuracy improvement in each of the movies when the *training* samples are added from the movie itself and ∼1% to ∼3% accuracy improvement when the *training* samples are added from another movie. This analysis shows that DronePrint can correctly detect non-drone with 81% accuracy in more realistic open-set scenarios and by adding a only few hundreds of new training samples new random sounds to the background class, the accuracy can be further improved.

7 DISCUSSION

Using experimentally and online sourced data, in this paper we showed that it is possible to build a drone detection and identification system that can accurately identify most of the typical commercial and consumer drones with high accuracy. With controlled experiments, we also showed that DronePrint is resilient to background noise, minor drone alterations, and Doppler effect. We next discuss the limitations of our work and potential improvements that can be explored as future research directions.

Microphone arrays and reflectors - We built DronePrint using data collected from a single, stationary microphone. Use of microphone phased arrays [66] may allow acquiring more fine-granular acoustic characteristics, especially when drones are constantly moving rather than hovering. Such spatially dispersed groups of microphones enable other related use cases such as drone localization, angle of arrival estimation, and drone identification in the presence of multiple drones [47]. Moreover, microphone arrays can also be helpful in extended applications such as drone maneuver classification, payload classification, and fault detection. Finally, in our experiment, we did not use a parabolic reflector together with the directional microphone. Adding properly designed parabolic reflectors will help to improve the range the microphones can cover [66].

Multi-drone scenario While we built and tested with DronePrint with single-drone scenarios, it is possible to extend the system to cover multi-drone cases. This requires additional techniques such as Independent Component Analysis (ICA), sound source separation, that allow identification and classification of each drone sound source. Several such frameworks have been proposed for speech processing applications [28, 48]. The adaptation of such frameworks in drone-related works focuses mainly on localization [39, 66]. However, they could be combined with DronePrint to perform multi-drone identification and classification.

Instantaneous background noise One of the major challenges in an acoustic-based classification model is the background noise. In general, the signal is not recognizable when the SNR ≤0dB where the noise power is greater than the signal. However, in the case of instantaneous noise, even when the noise power is larger than the signal power for an instance, we show that the feature vector re-scaling provides features that are close to the signal with minimal noise, and therefore, allows the model to classify the signal more accurately. With the feature vector re-scaling, we show the distance between the feature vectors (with the instantaneous noise and with minimal noise) is only 3.6%. However, further reductions of this vector distance would further improve the performances. Observing irregular and high power peaks in the waveform are being used in different disciplines to detect abnormalities [40]. Such approach can be taken to detect and then eliminate such instantaneous peaks at the pre-processing stage to better handle these scenarios.

Other hardware modifications and active attacks - In Section 6.1 we showed that DronePrint is resilient to some of the common hardware modifications due to the use time domain peak normalization. However, there are other hardware modifications such as motor replacement and addition of sound/vibration dampening materials that will also have an effect on a drone's acoustic signature. This is an area that can be further explored through experimentation and potentially can be solved by data augmentation once typical changes were identified.

Also, an active attacker can take deliberate measures (e.g., carefully designed propeller modifications) to change the acoustic signature of a drone with the objective of evading the classifier. Equally, the attacker can adapt Active Noise Control (ANC) and masking techniques to reduce the sound emitted by drones [63]. In the current scope of the paper, we do not consider such active attacks. Such attacks and defences usually is an arms race between the attackers and defenders. Some of the evasion attempts can potentially be mitigated using techniques such as defensive distillation [54] that are proposed to make deep neural networks more robust for adversarial examples [26] through smoother decision boundaries.

Other data sources - Building comprehensive drone acoustic datasets is a challenging task because of the difficulties in getting access to all common drone models and flying restrictions (e.g. no-fly zones and pilot licensing requirements in some countries and states). Therefore, it is vital to explore the possibility of using other sources to build such datasets. In this work, we showed the viability of sourcing drone acoustic data from online platforms. However, identifying and exploring more sources can contribute towards building better and more representative datasets required to train deep learning methods. One possible direction to explore is whether drone simulators[5] can be used effectively for this task. Furthermore, the feasibility of using novel audio data augmentation techniques such as SpecAugment [55], which performs augmentation using the spectrogram of an audio source needs to be studied.

Open-set methods - Open-set classification is still an evolving area in machine learning [6]. In this work, we showed that the OpenMax [5] method gives a slightly better performance compared to the background class method; a state-of-the-art open-set classification method that does not require open-set training samples. Nonetheless, the background class method may not work under all conditions because it has not seen all possible open-set data during training time. Thus, further experiments can be done with other comparable open-set audio classification methods [15, 42, 64, 81]. For example, the approach proposed by Wilkinghoff and Kurth [81] which involves a combination of a CNN based closed-set classification algorithm and an outlier detection algorithm based on deep convolutional autoencoders (DCAE) to recognize unknown samples, has shown promising results in acoustics scene classification.

Hybrid methods - Finally, in this work, we considered only acoustics data for drone identification. However, multi-sensor data such as vision, thermal, radar, and RF, can be combined for more precise classification and mostly likely the real-world drone detection systems will rely on heterogeneous data sources. Some work attempted

[5]https://www.dji.com/au/simulator

combining multiple methods for drone detection [65, 71]. However, a major drawback in multi-sensory drone detection applications is that the information from the different sensors is not fused to produce a result. Instead, the alert signals are used independently from each system component to provide multiple early warnings that are later confirmed by a human operator [16]. The system can be fully automated by leveraging recent advances in data fusion techniques (e.g. DeepSense [83]) without a considerable trade-off in classification capability.

8 CONCLUSION

In this paper, we showed that drones have unique acoustic signatures that are associated with their sound generation mechanism. We developed a prediction pipeline called DronePrint to detect the presence of the drone in an open-set scenario and followed by the identification of the make and model of the drone through cascaded identification stages. In our experiment settings, DronePrint drone detection could achieve 95% accuracy with known signal and 86% accuracy with unknown signals. The overall accuracy we achieved in DronePrint identification is 92%.

To address the challenges in building large datasets of drone sounds, we proposed to source data from online sources. We demonstrated that online-sourced data can be readily used and even such data is heavily processed, the acoustic signatures are preserved. DronePrint was trained using only the data collected from online sources, and it was able to make accurate predictions on experimentally collected data as well as data collected from other online sources that were not used during the training time. This is an important result because due to various reasons, it is not practical to collect data from all drone types in all possible flying scenarios.

We analyzed how the most common hardware modifications, Doppler effect, persistence and/or instantaneous background noise, and the addition of more drone classes can affect the performance of DronePrint. We showed that because of our data augmentation, peak normalization, and feature scaling steps DronePrint is resilient to such changes. Finally, we evaluated the DronePrint's drone detection classifier over 100 minutes of movie audio data having various real-world sounds and showed that it achieves close to 80% accuracy. We showed the careful selection of signals to train the additional background class results in further accuracy improvements.

ACKNOWLEDGMENTS

This project was financially supported by NSW Defence Innovation Network and NSW State Government (DINPP-2018-04-16).

REFERENCES

[1] Sara Al-emadi, Abdulla Al-ali, Amr Mohammad, and Abdulaziz Al-ali. 2019. Audio Based Drone Detection and Identification using Deep Learning. In *IWCMC'19*. IEEE, 459–464.

[2] Riham Altawy and Amr M Youssef. 2016. Security, privacy, and safety aspects of civilian drones: A survey. *ACM Transactions on Cyber-Physical Systems* 1, 2 (2016), 1–25.

[3] Muhammad Zohaib Anwar, Zeeshan Kaleem, and Abbas Jamalipour. 2019. Machine Learning Inspired Sound-Based Amateur Drone Detection for Public Safety Applications. *IEEE Transactions on Vehicular Technology* 68, 3 (2019), 2526–2534.

[4] S Batra. 2018. *41 Years (1978-2018) JEE Advanced (IIT-JEE) + 17 yrs (2002-2018).* Disha.

[5] Abhijit Bendale and Terrance E Boult. [n.d.]. Towards Open Set Deep Networks 1 Introduction. ([n. d.]).

[6] Terrance E Boult, Steve Cruz, Akshay Raj Dhamija, M Gunther, James Henrydoss, and Walter J Scheirer. 2019. Learning and the unknown: Surveying steps toward open world recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 9801–9807.

[7] Aldrich A Cabrera-Ponce, J Martinez-Carranza, and Caleb Rascon. 2020. Detection of nearby UAVs using a multi-microphone array on board a UAV. *International Journal of Micro Air Vehicles* 12 (2020), 1756829320925748. https://doi.org/10.1177/1756829320925748 arXiv:https://doi.org/10.1177/1756829320925748

[8] Gürol Canbek, Seref Sagiroglu, Tugba Taskaya Temizel, and Nazife Baykal. 2017. Binary Classification Performance Measures/Metrics: A comprehensive visualized roadmap to gain new insights. 821–826. https://doi.org/10.1109/UBMK.2017.8093539

[9] Victoria Chang, Pramod Chundury, and Marshini Chetty. 2017. Spiders in the Sky: User Perceptions of Drones, Privacy, and Security. 6765–6776. https://doi.org/10.1145/3025453.3025632

[10] Jagmohan Chauhan, Yining Hu, Suranga Seneviratne, Archan Misra, Aruna Seneviratne, and Youngki Lee. 2017. BreathPrint: Breathing acoustics-based user authentication. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. 278–291.

[11] Jagmohan Chauhan, Jathushan Rajasegaran, Suranga Seneviratne, Archan Misra, Aruna Seneviratne, and Youngki Lee. 2018. Performance Characterization of Deep Learning Models for Breathing-Based Authentication on Resource-Constrained Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 4 (Dec. 2018).

[12] Philip Church, Christopher Grebe, Justin Matheson, and Brett Owens. 2018. Aerial and surface security applications using lidar. In *Laser Radar Technology and Applications XXIII*, Monte D. Turner and Gary W. Kamerman (Eds.), Vol. 10636. International Society for Optics and Photonics, SPIE, 27 – 38. https://doi.org/10.1117/12.2304348

[13] Google Cloud. 2020. Advanced Guide to Inception v3 on Cloud TPU. https://cloud.google.com/tpu/docs/inception-v3-advanced

[14] Gael Fashingbauer Cooper. 2017. Watch a drone crash into Seattle's Space Needle. https://www.cnet.com/news/drone-crash-seattle-space-needle/

[15] Akshay Raj Dhamija, G Manuel, and Terrance E Boult. 2018. Reducing Network Agnostophobia. NeurIPS (2018).

[16] Eleni Diamantidou, Antonios Lalas, Konstantinos Votis, and Dimitrios Tzovaras. 2019. Multimodal Deep Learning Framework for Enhanced Accuracy of UAV Detection. In *Computer Vision Systems*, Dimitrios Tzovaras, Dimitrios Giakoumis, Markus Vincze, and Antonis Argyros (Eds.). Springer International Publishing, Cham, 768–777.

[17] DJI. [n.d.]. INSPIRE 1Specs. https://www.dji.com/au/inspire-1/info

[18] DJI. [n.d.]. Matrice 100. https://www.dji.com/au/matrice100

[19] DJI. [n.d.]. Mavic Pro. https://www.dji.com/au/mavic

[20] DJI. [n.d.]. Phantom 4 Pro. https://www.dji.com/au/phantom-4-pro

[21] DJI. [n.d.]. Spark. https://www.dji.com/au/spark

[22] DronePrint. [n.d.]. https://github.com/DronePrint/DronePrint.

[23] Sphere Drones. [n.d.]. DJI Inspire 1 - 3510 Motor (CW). https://shop.spheredrones.com.au/products/dji-inspire-1-3510-motor-cw

[24] Martins Ezuma, Faith Erden, Chethan Kumar Anjinappa, Ozgur Ozdemir, and Ismail Guvenc. 2019. Micro-UAV Detection and Classification from RF Fingerprints Using Machine Learning Techniques. In *2019 IEEE Aerospace Conference*. 1–13.

[25] Martin Ezuma, Faith Erden, Chethan Kumar Anjinappa, Ozgur Ozdemir, and Ismail Guvenc. 2020. Detection and Classification of UAVs Using RF Fingerprints in the Presence of Wi-Fi and Bluetooth Interference. *IEEE Open Journal of the Communications Society* 1 (2020), 60–76.

[26] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. 2014. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572* (2014).

[27] Alex Graves, Abdel rahman Mohamed, and Geoffrey Hinton. 2013. Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. 6645–6649.

[28] François Grondin and François Michaud. 2019. Lightweight and optimized sound source localization and tracking methods for open and closed microphone array configurations. *Robotics and Autonomous Systems* 113 (2019), 63 – 80. https://doi.org/10.1016/j.robot.2019.01.002

[29] Omar Adel Ibrahim, Savio Sciancalepore, and Roberto Di Pietro. 2020. Noise2Weight: On Detecting Payload Weight from Drones Acoustic Emissions. arXiv:2005.01347 [eess.AS]

[30] ISCE. 2020. What's so sacrosanct about 10 dB Signal to noise ratio? https://www.isce.org.uk/articles/whats-so-sacrosanct-about-10-db-signal-to-noise-ratio/

[31] Md Tamzeed Islam, Bashima Islam, and Shahriar Nirjon. 2017. SoundSifter: Mitigating Overhearing of Continuous Listening Devices. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '17)*. Association for Computing Machinery, 29–41.

[32] Raya Islam and Dr. Alexander Stimpson. 2017. Small UAV Noise Analysis.

[33] navdeep jaitly and e geoffrey hinton. 2013. Vocal Tract Length Perturbation (VTLP) improves speech recognition. (2013).

[34] Sungho Jeon, Jong-Woo Shin, Young-Jun Lee, Woong-Hee Kim, YoungHyoun Kwon, and Hae-Yong Yang. 2017. Empirical study of drone sound detection in real-life environment with deep neural networks. In *2017 25th European Signal Processing Conference (EUSIPCO)*. 1858–1862.

[35] Therese Jones. 2017. *International commercial drone regulation and drone delivery services.* Technical Report. RAND.

[36] Byungkwan Kim, Hyunseong Kang, and Seong-Ook Park. 2017. Drone Classification Using Convolutional Neural Networks With Merged Doppler Images. *IEEE Geoscience and Remote Sensing Letters* 14 (2017), 38–42.

[37] Nicola Kloet, Simon Watkins, and Reece Clothier. 2017. Acoustic signature measurement of small multi-rotor unmanned aircraft systems. *International Journal of Micro Air Vehicles* 9, 1 (2017), 3–14.

[38] Harini Kolamunna, Junye Li, Thilini Dahanayaka, Suranga Seneviratne, Kanchana Thilakaratne, Albert Y. Zomaya, and Aruna Seneviratne. 2020. POSTER: AcousticPrint: Acoustic Signature based Open Set Drone Identification. https://wisec2020.ins.jku.at/proceedings/wisec20-5.pdf.

[39] J. Lauzon, F. Grondin, D. Létourneau, A. L. Desbiens, and F. Michaud. 2017. Localization of RW-UAVs using particle filtering over distributed microphone arrays. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2479–2484.
[40] C. Lazaro, J.P. Marques, G. Marchesan, and G. Cardoso. 2018. Waveform asymmetry of instantaneous current signal based symmetrical fault detection during power swing. *Electric Power Systems Research* 155 (2018), 340 – 349.
[41] Dongkyu Rroyr Lee, Woong Gyu La, and Hwangnam Kim. 2018. Drone Detection and Identification System using Artificial Intelligence. In *9th International Conference on Information and Communication Technology Convergence*. Institute of Electrical and Electronics Engineers Inc., 1131–1133. https://doi.org/10.1109/ICTC.2018.8539442 9th International Conference on Information and Communication Technology Convergence, ICTC 2018 ; Conference date: 17-10-2018 Through 19-10-2018.
[42] Bernhard Lehner, Khaled Koutini, Christopher Schwarzlmüller, Thomas Gallien, and Gerhard Widmer. 2019. Acoustic scene classification with reject option based on resnets. In *Proceedings of the Detection and Classification of Acoustic Scenes and Events Workshop (DCASE), New York, NY, USA*. 25–26.
[43] Natasha Lomas. 2018. Analysis backs claim drones were used to attack Venezuela's president. https://techcrunch.com/2018/08/08/analysis-backs-claim-drones-were-used-to-attack-venezuelas-president/
[44] Georgia Lykou, Dimitrios Moustakas, and Dimitris Gritzalis. 2020. Defending Airports from UAS: A Survey on Cyber-Attacks and Counter-Drone Sensing Technologies. *Sensors* 20, 12 (2020), 3537. https://doi.org/10.3390/s20123537
[45] Abdirahman Mohamud and Ashwin Ashok. 2018. Drone Noise Reduction through Audio Waveguiding. In *Proceedings of the 4th ACM Workshop on Micro Aerial Vehicle Networks, Systems, and Applications* (Munich, Germany) *(DroNet'18)*. Association for Computing Machinery, New York, NY, USA, 92–94. https://doi.org/10.1145/3213526.3213543
[46] Pavlo Molchanov, Karen O. Egiazarian, Jaakko Astola, R. Harmanny, and Jos De Wit. 2013. Classification of small UAVs and birds by micro-Doppler signatures. *2013 European Radar Conference* (2013), 172–175.
[47] M. Y. Mustafa, G. Polanco, Q. Gao, Y. Xu, A. Mustafa, and Q. Z. Al-Hamdan. 2014. Application of microphone arrays for the detection of acoustic noise in porous panel shields. In *2014 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom)*. 549–553.
[48] F. Nesta and M. Omologo. 2012. Generalized State Coherence Transform for Multidimensional TDOA Estimation of Multiple Sources. *IEEE Transactions on Audio, Speech, and Language Processing* 20, 1 (2012), 246–260.
[49] Natalia Neverova, Christian Wolf, Griffin Lacey, Lex Fridman, Deepak Chandra, Brandon Barbello, and Graham Taylor. 2016. Learning human identity from motion patterns. *IEEE Access* 4 (2016), 1810–1820.
[50] BBC News. [n.d.]. Drug delivery drone crashes in Mexico. https://www.bbc.com/news/technology-30932395.
[51] BBC News. [n.d.]. Gatwick airport: How can a drone cause so much chaos? https://www.bbc.com/news/technology-46632892.
[52] BBC News. [n.d.]. Heathrow airport: Drone sighting halts departures. https://www.bbc.com/news/uk-46803713.
[53] Phuc Nguyen, Hoang Truong, Mahesh Ravindranathan, Anh Nguyen, Richard Han, and Tam Vu. 2017. Matthan: Drone Presence Detection by Identifying Physical Signatures in the Drone's RF Communication. In *MobiSys '17*. 211–224.
[54] Nicolas Papernot, Patrick McDaniel, Xi Wu, Somesh Jha, and Ananthram Swami. 2016. Distillation as a defense to adversarial perturbations against deep neural networks. In *2016 IEEE Symposium on Security and Privacy (SP)*. IEEE, 582–597.
[55] Daniel S Park, William Chan, Yu Zhang, Chung-Cheng Chiu, Barret Zoph, Ekin D Cubuk, and Quoc V Le. 2019. Specaugment: A simple data augmentation method for automatic speech recognition. *arXiv preprint arXiv:1904.08779* (2019).
[56] Parrot. [n.d.]. SUPPORT - PARROT BEBOP 2. https://support.parrot.com/us/support/products/parrot-bebop-2-fpv/support-produit
[57] Jarez S Patel, Francesco Fioranelli, and David Anderson. 2018. Review of radar classification and RCS characterisation techniques for small UAVs ordrones. *IET Radar, Sonar & Navigation* 12, 9 (2018), 911–919.
[58] Thomas Pathier. 2019. SoundUAV : Towards Delivery Drone Authentication via Acoustic Noise Fingerprinting. In *DroNet'19*. 27–32.
[59] Junkai Peng, Changwen Zheng, Tianyu Cui, Ye Cheng, and Lingyu Si. 2018. Using Images Rendered by PBRT to Train Faster R-CNN for UAV Detection. https://doi.org/10.24132/CSRN.2018.2802.3
[60] Ilyes Rebai, Yessine BenAyed, Walid Mahdi, and Jean-Pierre Lorré. 2017. Improving speech recognition using data augmentation and acoustic model fusion. *Procedia Computer Science* 112 (2017), 316 – 322.
[61] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*. 91–99.
[62] Matthew Ritchie, Francesco Fioranelli, Hervé Borrion, and Hugh Griffiths. 2017. Multistatic micro-Doppler radar feature extraction for classification of unloaded/loaded micro-drones. *IET Radar, Sonar Navigation* 11, 1 (2017), 116–124.
[63] N. Roy, M. Gowda, and R. Choudhury. 2015. Ripple: Communicating through Physical Vibration. In *NSDI*.
[64] Fatemeh Saki, Yinyi Guo, Cheng-Yu Hung, Lae-Hoon Kim, Manyu Deshpande, Sunkuk Moon, Eunjeong Koh, and Erik Visser. 2019. Open-set evolving acoustic scene classification system. (2019).
[65] Stamatios Samaras, Eleni Diamantidou, Dimitrios Ataloglou, Nikos Sakellariou, Anastasios Vafeiadis, Vasilis Magoulianitis, Antonios Lalas, Anastasios Dimou, Dimitrios Zarpalas, Konstantinos Votis, Petros Daras, and Dimitrios Tzovaras. 2019. Deep Learning on Multi Sensor Data for Counter UAV Applications—A Systematic Review. *Sensors* 19 (11 2019), 4837. https://doi.org/10.3390/s19224837
[66] Alexander Sedunov, Darren Haddad, Hady Salloum, Alexander Sutin, Nikolay Sedunov, and Alexander Yakubovskiy. 2019. Stevens Drone Detection Acoustic System and Experiments in Acoustics UAV Tracking. In *2019 IEEE International Symposium on Technologies*

for Homeland Security (HST). 1–7.
[67] Zhiguo Shi, Xianyu Chang, Chaoqun Yang, Zexian Wu, and Junfeng Wu. 2020. An Acoustic-Based Surveillance System for Amateur Drones Detection and Localization. *IEEE Transactions on Vehicular Technology* 69, 3 (2020), 2731–2739.
[68] Zhiyuan Shi, Minmin Huang, Caidan Zhao, Lianfen Huang, Xiaojiang Du, and Yifeng Zhao. 2017. Detection of LSSUAV using hash fingerprint based SVDD. In *2017 IEEE International Conference on Communications (ICC)*. 1–5.
[69] Royal Meteorological Society. [n.d.]. The Beaufort Scale How is wind speed measured? https://www.rmets.org/resource/beaufort-scale
[70] Martin Strauss, Pol Mordel, Victor Miguet, and Antoine Deleforge. 2018. DREGON: Dataset and Methods for UAV-Embedded Sound Source Localization. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2018)*. IEEE, Madrid, Spain, 5735–5742. https://doi.org/10.1109/IROS.2018.8593581
[71] Fredrik Svanstrom, Cristofer Englund, and Fernando Alonso-Fernandez. 2020. Real-Time Drone Detection and Tracking With Visible, Thermal and Acoustic Sensors. arXiv:2007.07396 [cs.CV]
[72] Bilal Taha and Abdulhadi Shoufan. 2019. Machine Learning-Based Drone Detection and Classification: State-of-the-Art in Research. *IEEE Access* 7 (2019), 138669–138682.
[73] Sarah Taillier. 2014. Triathlete injured as drone filming race falls to ground. https://www.abc.net.au/news/2014-04-07/triathlete-injured-as-drone-filming-race-drops-to-ground/5371658
[74] Charles E. Tinney and Jayant Sirohi. 2018. Multirotor Drone Noise at Static Thrust. *AIAA Journal* 56 (2018), 2816 – 2826.
[75] Zahoor Uddin, Muhammad Altaf, Muhammad Bilal, Lewis Nkenyereye, and Ali Kashif Bashir. 2020. Amateur Drones Detection: A machine learning approach utilizing the acoustic signals in the presence of strong interference. *Computer Communications* 154 (Mar 2020), 236–245. https://doi.org/10.1016/j.comcom.2020.02.065
[76] Cornell University. 2020. A General Guide for Deriving Abundance Estimates from Hydroacoustic Data. http://www.acousticsunpacked.org/AcousticBackground/Signal-to-noiseRatio.html
[77] Eren Unlu, Emmanuel Zenou, and Nicolas Riviere. 2018. Using Shape Descriptors for UAV Detection. *Electronic Imaging* 2018 (01 2018), 1–5. https://doi.org/10.2352/ISSN.2470-1173.2018.09.SRV-128
[78] I. Wagner. [n.d.]. Commercial UAVs - Statistics & Facts. arXiv:https://www.statista.com/topics/3601/commercial-uavs/
[79] Yang Wang, Huichuan Xia, Yaxing Yao, and Yun Huang. 01 Jul. 2016. Flying Eyes and Hidden Controllers: A Qualitative Study of People's Privacy Perceptions of Civilian Drones in The US. *Proceedings on Privacy Enhancing Technologies* 2016, 3 (01 Jul. 2016), 172 – 190. https://doi.org/10.1515/popets-2016-0022
[80] Martin Weil. 2013. Drone crashes into Virginia bull run crowd. https://www.washingtonpost.com/local/drone-crashes-into-virginia-bull-run-crowd/2013/08/26/424e0b9e-0e00-11e3-85b6-d27422650fd5_story.html
[81] Kevin Wilkinghoff and Frank Kurth. 2019. Open-set acoustic scene classification with deep convolutional autoencoders. (2019).
[82] Yuhang Wu, Yan-ting Ai, Wang Ze, Tian Jing, Xiang Song, and Yingtao Chen. 2019. A Novel Aerodynamic Noise Reduction Method Based on Improving Spanwise Blade Shape for Electric Propeller Aircraft. *International Journal of Aerospace Engineering* 2019 (2019), 3750451. https://doi.org/10.1155/2019/3750451
[83] Shuochao Yao, Shaohan Hu, Yiran Zhao, Aston Zhang, and Tarek Abdelzaher. 2017. Deepsense: A unified deep learning framework for time-series mobile sensing data processing. In *Proceedings of the 26th International Conference on World Wide Web*. 351–360.
[84] Yaxing Yao, Huichuan Xia, Yun Huang, and Yang Wang. 2017. Free to Fly in Public Spaces: Drone Controllers' Privacy Perceptions and Practices. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6789–6793. https://doi.org/10.1145/3025453.3026049
[85] Yaxing Yao, Huichuan Xia, Yun Huang, and Yang Wang. 2017. Privacy Mechanisms for Drones: Perceptions of Drone Controllers and Bystanders. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6777–6788. https://doi.org/10.1145/3025453.3025907
[86] Pengfei Zhang, Le Yang, Gao Chen, and Gang Li. 2017. Classification of drones based on micro-Doppler signatures with dual-band radar sensors. *2017 Progress in Electromagnetics Research Symposium - Fall (PIERS - FALL)* (2017), 638–643.
[87] Wenyu Zhang and Gang Li. 2018. Detection of multiple micro-drones via cadence velocity diagram analysis. *Electronics Letters* 54, 7 (2018), 441–443.
[88] Adam Zwickle, Hillary B. Farber, and Joseph A. Hamm. 2019. Comparing public concern and support for drone regulation to the current legal framework. *Behavioral Sciences & the Law* 37, 1 (2019), 109–124. https://doi.org/10.1002/bsl.2357 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/bsl.2357

A NULLIFYING THE EFFECT OF DISTANCE TO THE DRONE

In this appendix we show the results of additional experiments we conducted to show the performance of time-domain peak-normalization and feature vector re-scaling methods when the distance between the microphone and drone gradually increases. We experimentally collected data when the drone is hovering at distances of 10m,

20m, 50m, and 100m. Figure 18 shows the time domain signal and the feature space consisting of default MFCC calculations, MFCC with time-domain peak-normalization, and MFCC with feature vector re-scaling. We observe that the default MFCC calculations have significant value reductions in the feature space as the distance between the microphone and drone is increased. However, we observe these differences are minimal with the distance increments for the cases of time-domain peak-normalization and feature vector re-scaling. Note that even at the same distance, feature vectors have slight variations due to the small changes in the sound profile because of the drone's slight movements and trying to stay still with the varying environmental conditions. However, these experiments show that, apart from these slight changes in the feature vectors, the distance has no prominent changes in the feature space.

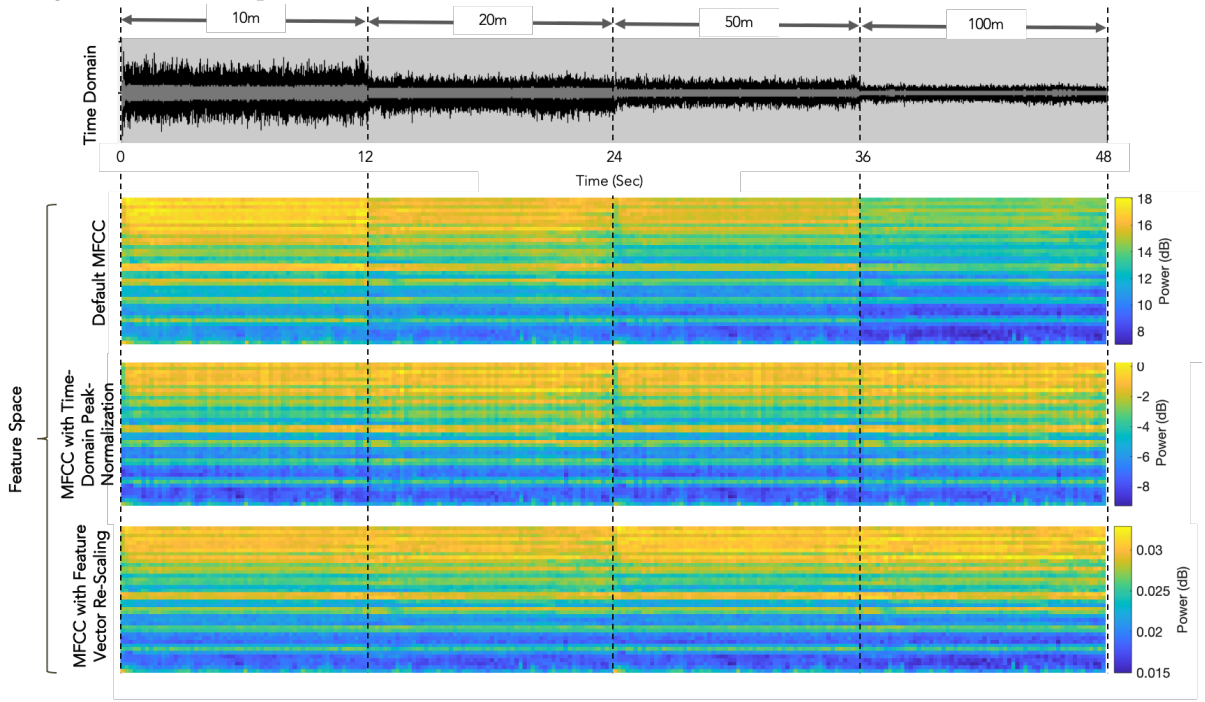


Fig. 18. Effect of time-domain peak-normalization.

B ONLINE VIDEOS WITH DRONE HARDWARE MODIFICATIONS

In Table 6 we list the online videos we used as sources for hardware modification experiments in Section 6.1.

Table 6. YouTube videos we used in the hardware modification experiments.

Scenario	File
Original	https://www.youtube.com/watch?v=u-hR7Nsy0lk
Commercially available low-noise propellers	
Custom-made low-noise propellers	https://www.youtube.com/watch?v=Idn0Y8WoHMI
With propeller guards	https://www.youtube.com/watch?v=olNSKLvM90Y
Upgraded camera and the gimbal	https://www.youtube.com/watch?v=gu2v-PPCg2g