Drone Authentication via Acoustic Fingerprint

Yufeng Diao
University of Glasgow
School of Engineering
Glasgow, UK
y.diao.1@research.gla.ac.uk

Yichi Zhang
University of Glasgow
School of Engineering
Glasgow, UK
2510899z@student.gla.ac.uk

Guodong Zhao
University of Glasgow
School of Engineering
Glasgow, UK
guodong.zhao@glasgow.ac.uk

Mohamed Khamis
University of Glasgow
School of Computing Science
Glasgow, UK
mohamed.khamis@glasgow.ac.uk

ABSTRACT

As drones become widely used in different applications, drone authentication becomes increasingly important due to various security risks, e.g., drone impersonation attacks. In this paper, we propose an idea of drone authentication based on Mel-frequency cepstral coefficient (MFCC) using an acoustic fingerprint that is physically embedded in each drone. We also point out that the uniqueness of the drone's sound comes from the combination of bodies (motors) and propellers. In the experiment with 8 drones, we compare the authentication *accuracy* of different feature extraction settings. Three kinds of different sound features are used: MFCC, delta MFCC (DMFCC), and delta-delta MFCC (DDMFCC). We choose the feature extraction settings and the sound features according to the best authentication result. In the experiment with 24 drones, we compare the closed set authentication performance of eight machine learning methods in terms of *recall* under the influence of additive white Gaussian noise (AWGN) with different levels of signal-to-noise ratio (SNR). Furthermore, we conduct an open set drone authentication experiment. Our results show that Quadratic Discriminant Analysis (QDA) outperforms other methods in terms of the highest average *recall* (94.19%) in the authentication of registered drones and the third highest average *recall* (82.35%) in the authentication of unregistered drones.

CCS CONCEPTS

• **Security and privacy** → **Authentication**; • **Computing methodologies** → **Machine learning**.

KEYWORDS

Authentication, Drones, MFCC, Acoustic fingerprinting, Machine learning

ACSAC '22, December 5–9, 2022, Austin, TX, USA

ACM ISBN 978-1-4503-9759-9/22/12…$15.00
https://doi.org/10.1145/3564625.3564653

ACM Reference Format:
Yufeng Diao, Yichi Zhang, Guodong Zhao, and Mohamed Khamis. 2022. Drone Authentication via Acoustic Fingerprint. In *Annual Computer Security Applications Conference (ACSAC '22), December 5–9, 2022, Austin, TX, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3564625.3564653

1 INTRODUCTION

In the past decade, advances in both software and hardware have made drones smaller, cheaper, and easier to fly without special training. As a result, drones have been widely used in different sectors, such as agriculture [6, 31], film production [10], rescue operations [7], etc. However, this raises new security concerns from different aspects, e.g., a drone may approach an airport and interfere with aviation safety; it may fly over private homes and lead to privacy issues; it may be used for terrorist attacks and drug smuggling [26].

Current solutions focused on drone detection and drone classification can be divided into four main categories [36]: radio frequency (RF) analyzers [14], acoustic sensors (microphones), optical sensors (cameras) [24], and radar [17]. Although these methods are used to detect and classify drones, they are not capable of authenticating drones, i.e., verifying their identity to distinguish them from impersonating drones. Authentication is crucial to prevent drones from accessing resources and areas they are not authorized to use/enter. For example, when delivery drones approach customers, they should be verified and then got permission to land or drop parcels. When delivery drones return to the warehouse, the warehouse should also verify that they are legitimate drones, rather than malicious drones. By doing so, drone authentication can prevent drone impersonation attacks [30].

In authentication, many drones have software-level digital certificates to indicate the individual identity of each drone [1, 11]. However, this is vulnerable to cyber attacks, such as impersonation. When dealing with cyber attacks, a valid solution is to utilize the physical attributes of drones, such as position [12] and RF [40], as additional authentication. Furthermore, using physical characteristics (such as fingerprints in the case of humans) that are deeply embedded in each drone has the potential to improve authentication by adding an inherence factor to authentication [9].

The source code repository for this paper is located at: https://github.com/Eidos1970/Drone-Authentication-via-Acoustic-Fingerprint

In this paper, we investigate drone authentication methods using acoustic signals from flying drones. In particular, our goal is to identify each individual drone from a group of drones of the same model/manufacturer. Our experiments demonstrate the feasibility of drone authentication using the noise generated by drone flight. Specifically, audio fingerprints can be built from Mel-frequency cepstral coefficient (MFCC), delta MFCC (DMFCC), and delta-delta MFCC (DDMFCC). We compare eight widely used machine learning methods in authentication tasks: (1) Linear Discriminant Analysis (LDA), (2) Quadratic Discriminant Analysis (QDA), (3) Linear kernel Support Vector Machine (LSVM), (4) Radial Basis Function kernel Support Vector Machine (RBF-SVM), (5) K-Nearest Neighbor (KNN), (6) Decision Tree (DT), (7) radio frequency (RF), and (8) Gaussian Naïve Bayes (GNB).

Feature extraction is important to improve the authentication performance of machine learning methods. We conducted a series of experiments with different feature extraction settings, such as: (1) the frame length of segmented audio, (2) the number of used filters, (3) the number of used features, and (4) the use of high-level features. Here, the high-level features refer to DMFCC and DDMFCC, which are generated from MFCC. The results shed light on the optimal MFCC feature extraction setting for drone authentication.

Through our experiment, we also found that the combination of drone bodies (motors) and propellers leads to unique sound features. For example, a nonzero offset of the rotor causes the noise generated by the motor. Meanwhile, the manufacturing imperfections of the different propellers lead to different wind noises. Therefore, given the same model of drones, the sound features of each individual drone could be determined by the combination of drone bodies (motors) and propellers.

The main contributions of this work are summarized as follows:

- To the best of our knowledge, this is the first work that uses acoustic fingerprints to authenticate flying drones.
- We report the results of experiments in which we used different parameter settings to extract MFCC, DMFCC, and DDMFCC. Our feature extraction setting could be used as a reference for future studies.
- We investigate the performance of eight machine learning methods in drone authentication. Furthermore, we also applied AWGN with different levels of SNR to explore the noise resistance ability of eight machine learning methods. These results could be used as a baseline for future research.
- We verify that the acoustic fingerprint could be determined by the combination of the drone body (motors) and the propellers.
- We verify that our proposed authentication settings and methods could not only solve the closed set problem, but also authenticate the drone that has never been seen by the algorithm before (open set problem). The results show that QDA outperforms other methods in terms of the highest average *recall* (94.19%) in registered drones and the third highest average *recall* (82.35%) in unregistered drones.

The rest of this paper is organized as follows. In Section 2, we summarize related work and outline how we extend state of the art. Section 3 describes how we collect audio data and how we combine the drone body and propellers as a "new" drone. Section 4 presents how to set up a series of experiments to find the appropriate feature extraction setting. Section 5 presents the result of the drone authentication experiment and attacks on authentication systems. In Section 6, we discuss the limitations of our research and possible future extensions. Finally, we summarize our work in Section 7.

2 RELATED WORK

Many existing works use microphones to detect the presence of drones (drone detection) and classify the type of drone (drone classification). The former accounts for most of the research in this area. The study of individual drone recognition in terms of sound characteristics is still an under-investigated field. Our work builds on prior work on (1) drone detection, (2) drone classification, and (3) other studies related to drone sound features.

2.1 Drone Detection

Kim et al. [21] designed software that can detect and monitor drones in real-time, every 0.743 seconds of data, based on Fast Fourier Transform (FFT). They used Plotted Image Machine Learning (PIL) to achieve 83% accuracy in detecting the drone without a propeller in a noisy indoor environment. In addition, KNN was applied to achieve 61% accuracy in the detection of flying drones. Although this work used different methods to detect the presence of drones, the accuracy needed to be further improved. The work of Bernardini et al. [8] used another approach to achieve outstanding detection accuracy. They implemented short-term (20 ms) and long-term (200 ms) audio analyses and used RBF-SVM to achieve an accuracy of over 96.4% in distinguishing drone sound and environmental sound. In the short-term analysis, 13 MFCC features were extracted, which contain distinctive information for detection. However, this work did not address detection with unseen data. The unseen data problem was solved by Jeon et al. [20], who were the first to use Gaussian Mixture Model (GMM), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN) in drone detection, considering the application in a real-time detection system. The collected drone audio was augmented using the noise dataset. They used 40 Mel filters to extract MFCC features and used 20 features for GMM and 40 features for CNN and RNN. Furthermore, a 40 ms time window was applied to extract MFCC features for GMM, while a 240 ms time window was applied for CNN and RNN. They achieved drone detection in the unseen data type and found that RNN has the best performance with *F1-score* of 0.6984.

Seo et al. [34] segmented the audio data into a 20 ms frame length with 50% overlap and used Short-time Fourier Transform (STFT) to extract the sound features. They corrupted the audio data with AWGN in different SNR and applied CNN to test the influence of noise on detection. Their results showed that lower SNR led to lower accuracy. While this work used AWGN to explore the influence of noise, noise in the real environment should be more complex. To address drone detection and location in a real outdoor environment, Sedunov et al. [33] developed a Drone Acoustic Detection System (DADS) to detect drone presence and track drone location. The sound signals were captured in 4 seconds with 50% overlap to generate the spectrogram. A novel algorithm was developed to detect drones and distinguish drone sound from other similar

sounds. This algorithm was based on detecting and tracking the number of harmonics in the spectrogram.

In the field of drone detection, researchers were trying to compare different combinations of feature extraction schemes and different classification schemes. Yang et al. [39] designed the experiments in a real outdoor environment. They extracted audio features using MFCC and STFT. Then they applied Support Vector Machine (SVM) and CNN to compare the effect of different combinations of features and methods. The result showed that the STFT-SVM combination had the best drone detection performance. However, in some cases, the drone used in the experiment was pulled by a string to confront the intense wind, which could influence the features of the drone sound. In addition, some audio was recorded from the drone held by a walking pilot. Anwar et al. [5] recorded drone audio in a real noise environment and extracted the sound features using Linear Predictive Cepstral Coefficients (LPCC) and MFCC. Then, they applied SVM with different kernels to compare the performance of LPCC and MFCC with 13 features. The results showed that the detection performance of MFCC surpassed LPCC with different kernels. This work reached the conclusion that machine learning is an efficient and accurate tool in the field of drone detection. Uddin et al. [37] unmixed the recorded signals and then extracted the sound features through MFCC, power spectral density (PSD), and Root Mean Square (RMS) of PSD. They applied SVM and KNN to detect the presence of drones. The author pointed out that the combination of RMS values of PSD and KNN outperformed other combinations in terms of accuracy.

2.2 Drone Classification

Siriphun et al. [35] divided the audio into 4 to 5 seconds of each sample and applied FFT to extract sound features. Then they applied RF to detect and classify drones. The result showed that the drone model had a great influence on the detection and classification performance. However, to be more convincing, other classification methods must be tested. Al-Emadi et al. [4] used CNN, RNN, and Convolutional Recurrent Neural Network (CRNN) to detect and classify drones based on a spectrogram extracted from 1 second of audio. Additionally, public noise datasets were applied to drone audio to mimic real scenes. On the basis of that, Al-Emadi et al. [3] used Generative Adversarial Network (GAN) to generate an artificial dataset with the aim of improving the performance of drone detection and classification. The result suggested that the benefits of using GAN to augment datasets outweighed the drawbacks in drone detection and drone classification. Utebayeva et al. [38] were the first group to use Long Short-Term Memory (LSTM) to classify drone sound. They extracted the sound features by MFCC, but the details of the configuration were ambiguous. Kolamunna et al. [23] extracted MFCC features with a frame length of 25 ms and an overlap of 15 ms. They stacked 20 frames as input to their LSTM model and trained with the background class to solve the open set classification. Based on their previous work, Kolamunna et al. [22] developed *DronePrint* based on LSTM for drone detection and classification. The authors analyzed the drone sound characteristic and discussed the influence of peak normalization in the time domain and the rescaling of the feature vector in MFCC. They used 40 filters to extract 40 MFCC features from the audio, which was split into 200 ms segments without overlap. Then, 10 time steps (frames) are entered into the two-stacked LSTM model for drone detection and classification. Furthermore, *DronePrint* showed resistance to the Doppler effect due to the data augmentation step. The results reported in this work suggested that *DronePrint* can achieve an accuracy of 95% in known sound signals and an accuracy of 86% in unknown sound signals in drone detection. In addition, it achieved an overall accuracy of 92% in drone classification.

2.3 Other Drone Sound Related Studies

Ibrahim et al. [19] were the first to detect the payload of commercial drones according to the sound features. They extracted 40 MFCC features in audio with frame lengths of 0.25 and 1 s. Ten widely used machine learning methods were applied. In addition, the result showed that cubic SVM outperformed other methods and can achieve 98.4% payload detection accuracy in just 0.25 s frame length of recorded audio. Furthermore, with a frame length of 1 s, QDA, LSVM, and quadratic SVM outperformed other methods with an accuracy of 98.9%. They found that a longer frame length was effective in detecting drone payloads.

Ramesh et al. [30] used the noise generated by motor rotation to identify individual drones. They showed that the motor noise (without propellers) was unique for each drone. In addition, they extracted the cepstral features of drone audio and applied SVM to authenticate different drones in the same model without propellers. Fifty-four motors and 11 drones of the same model and make were used in their work. Furthermore, they achieved an accuracy of 99.48% in drone authentication without propellers. Their work is inspiring, but whether this method can still be applied when a drone flies with propeller noise is an open question.

2.4 Our Contribution Compared to Prior Work

Many studies mentioned above have yielded outstanding results in drone detection and classification, however, none have addressed the issue of authentication on flying drones. In this paper, we use MFCC-related features as the acoustic fingerprint for drone authentication. Previous work has already shown that MFCC is an effective method for extracting features from drone audio [5, 8, 19, 20, 22, 23, 37–39]. This motivated us to use MFCC in our work. We also use DMFCC and DDMFCC in our work, since they are popular in speaker recognition [2]. However, the configuration details of MFCC were not clear enough in many of the above-mentioned works. Most of the works did not provide specific details on the number of used filters [5, 8, 19, 22, 23, 37–39]. Furthermore, some authors did not mention the segmented frame length [5, 37–39] and the frame overlap ratio of segmented frames [5, 19, 37–39]. Some works did not mention the number of used features [38, 39]. In addition, none of them discussed the effectiveness of DMFCC and DDMFCC [18] in drone sound analysis. In our work, we discuss the effect of the following four settings on drone authentication: (1) the frame length of segmented audio, (2) the number of used filters, (3) the number of used features, and (4) the use of high-level features. By default, each frame has a 50% overlap with adjacent frames. Our work not only performs drone authentication via acoustic fingerprint, but also can provide a reference to set feature extraction parameters for further research.

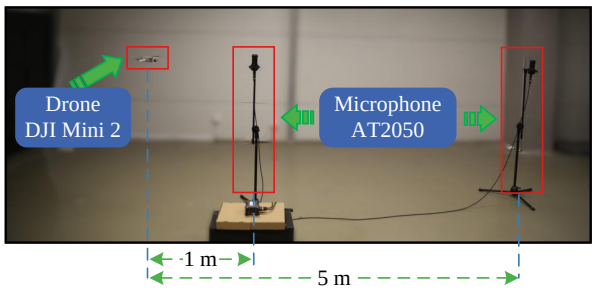

Figure 1: Recording room setup.

3 DATASETS

In our experiment, 8 DJI Mini 2 [13] and two sets of spare propellers were used. We collected audio data from 8 original drones and 16 reassembled drones. Here, reassembled drones mean using the body of the original drones but with different spare propeller sets. Each set includes four propellers. As shown in Table 1, we labeled the bodies of the 8 drones from "A" to "H". The propeller sets that come with drones from "A" to "H" were labeled as a1-a4, b1-b4, c1-c4, etc. Here, the letters ("a" to "h") represent the original drone body of the propellers. In addition, the numbers (1 to 4) represent the numerical label of the propellers.

The original 8 drones were marked as drones No. 1 - No. 8. Next, we labeled the two sets of spare propellers as x1-x4 and y1-y4, respectively. We used the first set of spare propellers x1-x4 to replace the original propellers a1-a4 in drone No. 1. After that, we labeled this reassembled drone (body "A" with the propeller set x1-x4) as drone No.9. We repeated this for drones No. 2 - No. 8 to obtain drones No. 10 - No. 16. Then, we used the second set of spare propellers y1-y4 to obtain drones No. 17 - No. 24 in the same way. Based on the above setup, we collected audio clips from 8 original drones and 16 reassembled drones ("new" drones).

The size of the recording room is approximately 5 m in width, 8 m in length, and 3 m in height. As a preliminary study of drone authentication, we recorded drone audio only when the drone was hovering. We used two multi-pattern condenser microphones (model: AT2050) to record the audio with a sampling rate of 44.1 kHz and a bit depth of 16, where the microphone is 1 m and 5 m away from the drone, respectively. Figure 1 shows the setup of the recording room.

The audio recording was stored in WAV format as a mono channel. We recorded similar lengths of audio clips for each drone (No. 1 - No. 24) to avoid bias when training our algorithms. For each drone, we recorded about 10 minutes (600 seconds) at 1 and 5 meters at the same time, respectively. The entire audio data was collected over 15 days to reduce bias caused by weather or temperature. Since the recording room is close to another office area, the recorded audio contained some noise. Table 1 shows the details of the audio data collected and the combinations of drone bodies and propellers.

The audio of the drones under each label (No.) was divided into two parts: 70% for training (training set) and 30% for testing (test set). Furthermore, the audio recorded for each drone at 1 m and 5 m is maintained with the same timestamp after division. Based on collected drone audio, we created four datasets.

Table 1: Collected Drone Audio

Drone No.	Combination	1m (s)	5m (s)	Total (s)
1	"A" & a1-a4	609.94	609.94	1219.88
2	"B" & b1-b4	605.00	605.00	1210.00
3	"C" & c1-c4	612.01	612.01	1224.02
4	"D" & d1-d4	606.00	606.00	1212.01
5	"E" & e1-e4	605.93	605.93	1211.87
6	"F" & f1-f4	605.93	605.93	1211.87
7	"G" & g1-g4	606.00	606.00	1212.01
8	"H" & h1-h4	607.07	607.07	1214.14
9	"A" & x1-x4	609.94	609.94	1219.88
10	"B" & x1-x4	604.93	604.93	1209.87
11	"C" & x1-x4	615.08	615.08	1230.16
12	"D" & x1-x4	608.07	608.07	1216.14
13	"E" & x1-x4	603.93	603.93	1207.87
14	"F" & x1-x4	607.00	607.00	1214.01
15	"G" & x1-x4	610.07	610.07	1220.15
16	"H" & x1-x4	608.00	608.00	1216.01
17	"A" & y1-y4	626.02	626.02	1252.05
18	"B" & y1-y4	605.93	605.93	1211.87
19	"C" & y1-y4	607.07	607.07	1214.14
20	"D" & y1-y4	627.96	627.96	1255.92
21	"E" & y1-y4	604.93	604.93	1209.87
22	"F" & y1-y4	605.00	605.00	1210.00
23	"G" & y1-y4	605.00	605.00	1210.00
24	"H" & y1-y4	635.96	635.96	1271.93
Total	-	14642.89	14642.89	29285.79

- **DS1:** this dataset contained drone audio from No. 1 - No. 8.
- **DS2:** this dataset contained drone audio from No. 1 - No. 24.
- **DS1N:** we added AWGN to **DS1** with 0 dB SNR to create **DS1N**. The lengths of the corresponding drone audio in **DS1N** and **DS1** are equal to each other.
- **DS2N:** we added AWGN to **DS2** with 93 levels of SNR ranging from -8.00 dB to 15.00 dB with a step of 0.25 dB to create **DS2N**. The size of **DS2N** is 93 times larger than the size of **DS2**. In other words, each level of SNR creates a new subset in **DS2N**, whose size is equal to DS2.

We tested on **DS1** and **DS1N** to find out which configuration for feature extraction is appropriate. Then we used **DS2** and **DS2N** to verify the configuration we chose. In addition to drone audio, we also recorded about 60 minutes of real indoor noise from our recording room. This noise audio was used with **DS2** for security studies in the drone authentication experiment.

4 AUTHENTICATION METHODS

In this section, we first discuss the method for pre-processing the data. Then we investigate the proper setting of the MFCC parameters for authentication. To evaluate the performance of different configurations in this part, we use *accuracy* as a performance indicator:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (1)$$

where *TP* is true positive, *TN* is true negative, *FP* is false positive, and *FN* is false negative.

In our work, eight machine learning methods are based on the code of *scikit-learn* [28]. In addition, we use *python_speech_features* [27] to extract MFCC, DMFCC, and DDMFCC.

4.1 Data Preprocessing

Before we extract the MFCC, DMFCC, and DDMFCC features from the audio, we must properly pre-process the original audio.

First, we notice that it is not necessary to extract features from the whole frequency domain. This is because the energy contained in the high-frequency domain is too small for drone sound. We calculate the average energy distribution of all the collected audio. The results show that almost 95% of the energy is located within the frequency range of 0-8 kHz. Thus, we focus on extracting the features within the range of 0-8 kHz. This range is also commonly used by other works in drone detection and classification [3, 4, 22].

The next step is to split each audio file into small frames for feature extraction. Based on previous works, the frame lengths commonly used in previous work ranged from 20 to 1000 ms [8, 19, 20, 22, 23]. In general, the short frame length aims to capture the instant features in the audio, while the long frame length can show the general features.

To find the appropriate frame length, we calculate the authentication *accuracy* in different frame lengths ranging from 20 to 2520 ms with an increment of 50 ms. In addition, each frame has a 50% overlap with adjacent frames by default. Here, we train and test the eight machine learning methods on drone audio in **DS1**. In this experiment, 50 filters and 49 features (from 2 to 50) of MFCC are used.

Figure 2 shows the variation of the authentication *accuracy* with respect to the frame length. The results suggest that the sound features in a very short frame length (20 ms) are not enough to distinguish individual drones. A longer frame length can enhance the sound features of each drone. However, since the total length of drone audio is fixed, a longer frame length leads to a smaller size of the extracted feature data, which could decrease the performance of the machine learning algorithm. Here, we choose the frame length of 1000 ms to have a good balance between *accuracy* and the size of the extracted feature data.

4.2 Feature Extraction

In this experiment, three factors influence the authentication *accuracy*: (1) the number of filters, (2) the number of used features, and (3) the use of high-level features. The number of filters determines the number of total generated features. In all experiments, all the first extracted features for MFCC, DMFCC, and DDMFCC are discarded, which contain unwanted low-frequency information generated from the environment. In addition, since the audio may contain some instantaneous noise, we apply the feature vector rescaling [22] method on the extracted feature vector to reduce its influence. We divide the feature vector by its L_2 norm. The formula is shown below:

$$v'_{\text{Feature}} = \frac{v_{\text{Feature}}}{\|v_{\text{Feature}}\|},$$

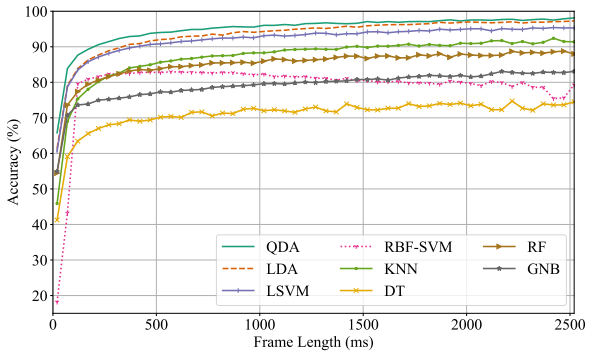

Figure 2: The influence of frame length on *accuracy*.

where v_{Feature} is the feature vector and v'_{Feature} is the feature vector after rescaling.

Previous work [5, 8, 19, 20, 22, 23, 37] suggests that the common number of used features is between 13 and 40, while the applied filters remain unknown for most of the work. However, according to our experiment, adopting these numbers leads to poor *accuracy*. To find the proper setup of three factors, we use eight machine learning methods to obtain the authentication *accuracy* with a frame length of 1000 ms, which is based on the result of data pre-processing. In this part, the filter-varying experiment without AWGN is conducted in **DS1**, and the filter-varying experiment with AWGN is conducted in **DS1** and **DS1N**.

4.2.1 Filter-varying Experiment: Using MFCC only. First, we only use MFCC features. We increase the number of filters from 26 to 271 with an increment of 5 while keeping using one-third, two-thirds, and all the generated features, respectively, to explore the influence of the number of filters and the number of used features. We name this process "filter-varying process". When the number of used features is not divisible, we round that number down. Figures 3 (a), (b), and (c) illustrate that using more features leads to better authentication *accuracy*, with a fixed number of filters. Furthermore, the number of features commonly used in previous work does not perform well in drone authentication.

4.2.2 Filter-varying Experiment: Using MFCC and DMFCC. Second, we combine MFCC and DMFCC together and repeat the filter-varying process. The results are shown in Figures 3 (d), (e), and (f), respectively. Compared to using only MFCC, using MFCC and DMFCC has a similar *accuracy* in most algorithms, except RBF-SVM and KNN. Using one-third of the features, these two methods have a dramatic decrease in *accuracy*. Furthermore, using two-thirds or all the features, RBF-SVM returns to a level of *accuracy* similar to before, while KNN remains at a low *accuracy*.

4.2.3 Filter-varying Experiment: Using MFCC, DMFCC and DDMFCC. Third, we combine MFCC, DMFCC and DDMFCC together and repeat the filter-varying process. The results are shown in Figures 3 (g), (h), and (i), which are similar to the last experiment. However, the *accuracy* of RBF-SVM and KNN is further reduced.

The results of the three experiments show that LDA, QDA, and LSVM always perform well compared to the other methods. While KNN has relatively good *accuracy* only using MFCC.

These three experiments also show that adding more features may not improve performance. Although using a smaller number of features can lead to poorer *accuracy*, when the number of filters is large enough, using a third of the features can also give good results, but slightly lower than using all features. This means that in the range of 0-8 kHz, the high frequency also contains valuable information for drone authentication. However, the effect of DMFCC and DDMFCC is unclear, because six machine learning algorithms achieve a similar *accuracy* using the same configurations with an increasing number of filters. Only the *accuracy* of RBF-SVM and KNN is reduced by introducing DMFCC and DDMFCC.

4.2.4 Filter-varying Experiment with AWGN. To explore the effect of MFCC and DMFCC, we apply the same filter-varying experiment in **DS1N**. We use the models trained on the training set of **DS1** and test them in the test set of **DS1N**. This experiment aims to explore whether the use of DMFCC and DDMFCC has a special effect under the influence of noise. The results are shown in Figure 4. Compared to the filter-varying experiment without AWGN, the *accuracy* of using a small number of filters and features is greatly reduced. In this case, we can see the importance of using a large number of filters and features. Furthermore, the addition of AWGN has a great influence on DT, which has a significant decrease in drone authentication *accuracy*. In addition, LDA shows strong resistance to noise and remains the highest *accuracy* in all configurations.

These results show that as the number of filters and used features increases, the *accuracy* of each algorithm almost continuously increases. Their *accuracy* reaches a bottleneck of about 200 used filters. Although DMFCC and DDMFCC are effective in speaker recognition [2], their implementation does not have a positive effect on drone authentication. On the contrary, DMFCC and DDMFCC reduce the *accuracy* of RBF-SVM and KNN.

Based on the results of the above experiments, we segment drone audio with a frame length of 1000 ms, and each frame has a 50% overlap with adjacent frames. For feature extraction, 201 filters are applied to extract MFCC features from 2 to 201 and use them for training and testing. DMFCC and DDMFCC are unused.

5 AUTHENTICATION EXPERIMENT RESULTS

We conducted 3 authentication experiments:

- Authentication of 24 drones without AWGN.
- Authentication of 24 drones with varying AWGN.
- Security study.

Since the experiment in Section 4 has already proven the feasibility of drone authentication, the first experiment in this section will confirm that the drone with replaced propellers has a unique acoustic fingerprint, i.e., it can be regarded as a new drone. Then we explore the influence of AWGN with different SNR on our drone authentication methods. Finally, we show the results of our scheme against unregistered drones.

To evaluate the performance of the eight machine learning methods in the authentication of 24 drones, in addition to *accuracy*, we also use the following three performance indicators:

$$Precision = \frac{TP}{TP + FP}, \tag{2}$$

$$Recall = \frac{TP}{TP + FN}, \tag{3}$$

$$F1 = \frac{1}{2} \times \left(\frac{1}{Recall} + \frac{1}{Precision}\right), \tag{4}$$

where *TP* is true positive, *FP* is false positive, and *FN* is false negative.

Table 2: Authentication Results of 24 Drones without AWGN

Method	*Accuracy* (%)	*Precision* (%)	*Recall* (%)	*F1* (%)
QDA	96.20	96.32	96.20	96.20
LDA	92.43	92.47	92.42	92.39
LSVM	93.68	93.71	93.67	93.64
RBF-SVM	66.05	72.67	66.06	65.31
KNN	90.49	90.74	90.49	90.49
DT	62.83	63.24	62.83	62.87
RF	83.73	83.96	83.73	83.61
GNB	67.15	67.79	67.16	66.73

5.1 Authentication of 24 Drones without AWGN

In this authentication experiment, eight machine learning methods are used to authenticate 24 drones without applying AWGN. This result justifies the feasibility of using an acoustic fingerprint for drone authentication and builds a baseline for further noise analysis. Furthermore, this experiment shows that a new combination of drone body and propellers can be regarded as a new drone in terms of acoustic fingerprint. We used all drone audio in **DS2** to train and test the model for each method. We calculate the evaluation metric that includes *accuracy*, *precision*, *recall*, and *F1-score* for all methods. The result is shown in Table 2. The general values of *precision*, *recall*, and *F1-score* in the table are calculated using the unweighted average values of *precision* and *recall* of each drone label.

According to Table 2, the *accuracy* of five methods is greater than 80%, and for QDA, it is greater than 95%. The results show that QDA outperformed the other methods significantly in all performance metrics. All evaluation metrics for QDA are greater than 96%. On the contrary, RBF-SVM, DT, and GNB perform poorly, with all performance metrics below 73%, 64%, and 68%, respectively.

5.2 Authentication of 24 Drones with Varying AWGN

To test the influence of noise, we continue to use the models trained in **DS2**, but we test them on the test set of **DS2N**. The drone audio in the test set of **DS2N** is similar to the drone sound in **DS2** but with the corruption of AWGN with different levels of SNR.

Figure 5 shows the result of this experiment. Generally, with the increase of SNR, *accuracy* of all methods increases. KNN shows an outstanding noise resistance when SNR is significantly low while maintaining a relatively high *accuracy*. QDA and LSVM are two

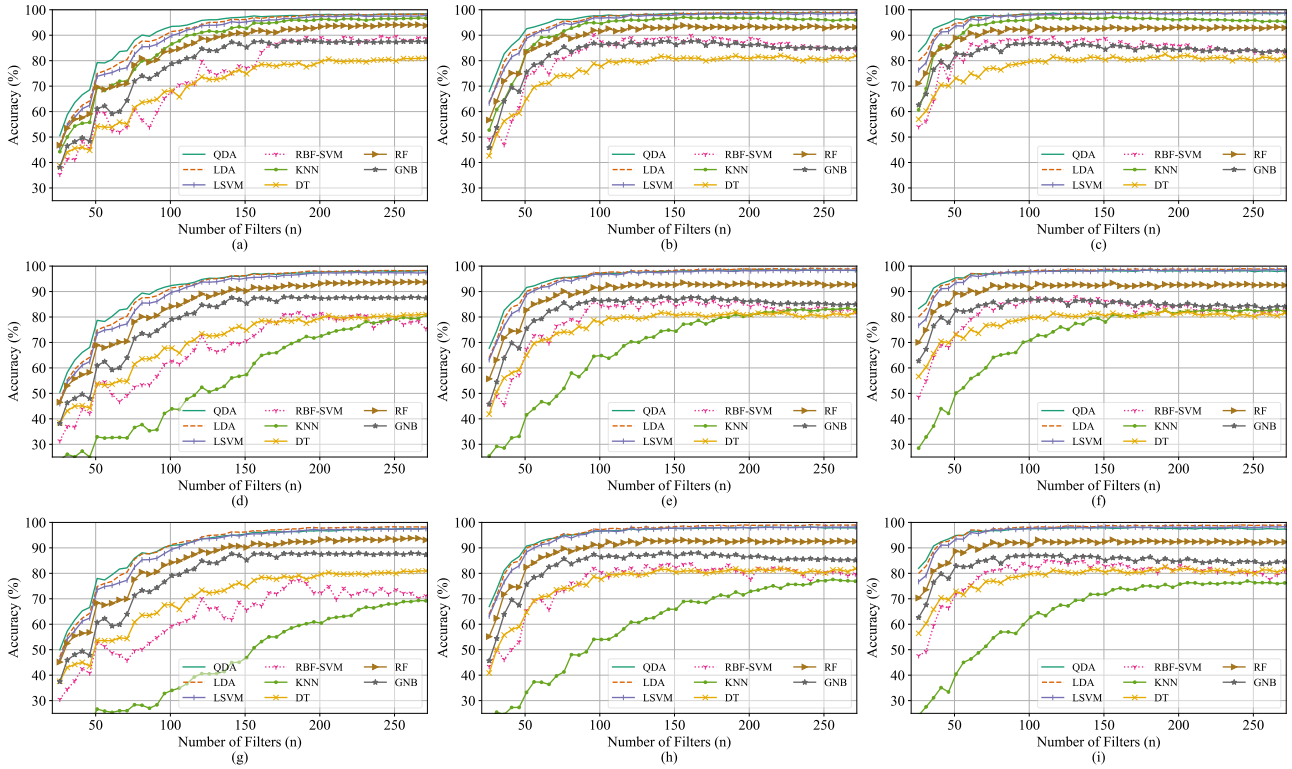

Figure 3: Number of filters versus *Accuracy* without AWGN on DS1 test set. (a) Increasing number of filters using one-third of the MFCC features. (b) Increasing number of filters using two-thirds of the MFCC features. (c) Increasing number of filters using all the MFCC features. (d) Increasing number of filters using one-third of the MFCC and DMFCC features. (e) Increasing number of filters using two-thirds of the MFCC and DMFCC features. (f) Increasing number of filters using all the MFCC and DMFCC features. (g) Increasing number of filters using one-third of the MFCC, DMFCC, and DDMFCC features. (h) Increasing number of filters using two-thirds of the MFCC, DMFCC, and DDMFCC features. (i) Increasing number of filters using all the MFCC, DMFCC, and DDMFCC features.

powerful drone authentication methods at high SNR, but they are very sensitive to low SNR.

When SNR is less than 0 dB, as SNR decreases, *accuracy* of all methods decreases fast. Furthermore, KNN becomes the most effective method when SNR is less than -3 dB. When SNR is around 0 dB, QDA, LDA, LSVM, and KNN outperform other methods with *accuracy* of 80.55%, 74.57%, and 76.65%, respectively. When SNR exceeds 4 dB, the *accuracy* increasing rate of all methods starts to slow down, and there is almost no further growth after 10 dB. It can be assumed that most of the methods perform well at 2 dB SNR or more, where the *accuracy* of QDA, LDA, LSVM, and KNN is higher than 80%.

5.3 Security Study

The previous authentication experiments show that our proposed authentication method can perform a good closed set classification. However, in reality, it is more about authentication among registered and unregistered drones, which is an open set problem. To evaluate the authentication performance under possible malicious drone attacks, we designed this drone attack study.

5.3.1 Threat Model. We assume that the attacker has access to a drone of the same model and is able to navigate it to a target location they are not authorized to access. The attacker tries to use the unregistered drone to pass our authentication procedure. In a real scenario, this authentication may occur, for example, when a drone needs to enter a warehouse to pick up a shipment that needs to be delivered [16, 32].

5.3.2 Design. To solve the open set authentication problem, we decided to build a background class as the unregistered drone type. We recorded about 60 minutes of real indoor noise from our recording room and combined it with 8 types of drone audio (background drones) chosen from **DS2** to train a background class. Meanwhile, we chose 8 drones as registered drones and 8 drones as attack drones (unregistered drones). The labels of these three types of drones do not overlap each other. During the training process, the audio of the attack drones is not present in the training set. After training, we used the drone audio of 8 registered drones and 8 attack drones, which came from the test set of **DS2**, to show the authentication performance.

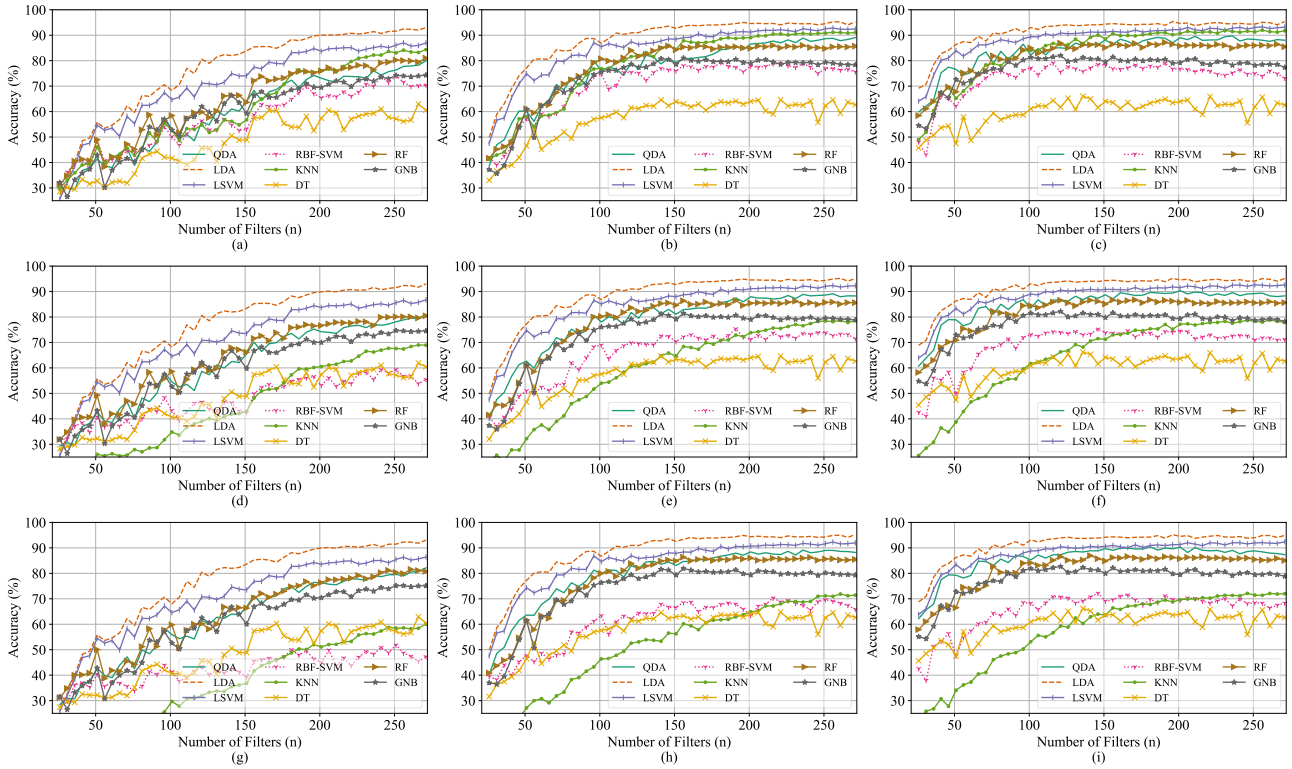

Figure 4: Number of filters versus *Accuracy* with 0 dB SNR on DS1N test set. (a) Increasing number of filters using one-third of the MFCC features. (b) Increasing number of filters using two-thirds of the MFCC features. (c) Increasing number of filters using all the MFCC features. (d) Increasing number of filters using one-third of the MFCC and DMFCC features. (e) Increasing number of filters using two-thirds of the MFCC and DMFCC features. (f) Increasing number of filters using all the MFCC and DMFCC features. (g) Increasing number of filters using one-third of the MFCC, DMFCC, and DDMFCC features. (h) Increasing number of filters using two-thirds of the MFCC, DMFCC, and DDMFCC features. (i) Increasing number of filters using all the MFCC,DMFCC, and DDMFCC features.

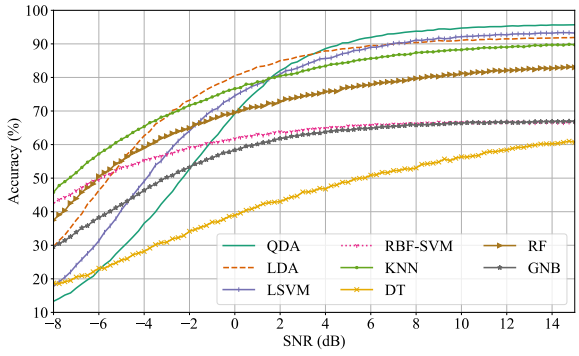

Figure 5: The influence of AWGN with different SNR.

5.3.3 Procedure. We conducted the experiment 10 times with different combinations of registered drones, background drones, and attack drones. We manually selected drones No. 1 - No. 8 as registered drones, drones No.9 - No. 16 as background drones, and drones No. 17 - No. 24 as attack drones in the first experiment. Then we randomly chose 8 drones as registered drones, background drones, and attack drones, respectively, for attack experiments No. 2 - No. 10. Table 3 shows the setups of the 10 experiments.

5.4 Security Study Results

As an authentication issue, we are more concerned about (1) whether registered drones can be properly authenticated as their identities and (2) whether unregistered drones can be authenticated as unregistered by the system. In that case, we used the average *recall* value of all registered drones and all attack drones, respectively, as a performance metric.

5.4.1 Authentication on Registered Drones. Table 4 shows the authentication *recall* on registered drones. Although this is an open set problem, QDA (94.19%), LDA (89.64%), LSVM (91.11%), and KNN (92.71%) still perform better than others on average *recall*. On the contrary, RBF-SVM has very poor performance and only achieves an average *recall* of 13.05%, which means that most registered drones will be classified as unregistered drones.

Table 3: Attack Experiment Setup

Experiment No.	Registered Drone No.	Background Drone No.	Attack Drone No.
1	1, 2, 3, 4, 5, 6, 7, 8	9, 10, 11, 12, 13, 14, 15, 16	17, 18, 19, 20, 21, 22, 23, 24
2	1, 2, 4, 7, 10, 16, 17, 20	3, 5, 6, 14, 15, 18, 19, 21	8, 9, 11, 12, 13, 22, 23, 24
3	3, 5, 9, 14, 16, 17, 19, 24	2, 6, 8, 11, 12, 15, 18, 21	1, 4, 7, 10, 13, 20, 22, 23
4	3, 12, 13, 16, 20, 21, 22, 24	6, 7, 8, 9, 14, 15, 17, 18	1, 2, 4, 5, 10, 11, 19, 23
5	2, 5, 6, 8, 14, 17, 18, 20	1, 4, 7, 9, 10, 12, 19, 23	3, 11, 13, 15, 16, 21, 22, 24
6	1, 10, 11, 15, 17, 18, 23, 24	3, 5, 8, 9, 12, 13, 19, 20	2, 4, 6, 7, 14, 16, 21, 22
7	2, 3, 4, 5, 11, 12, 15, 24	1, 6, 7, 10, 16, 18, 21, 22	8, 9, 13, 14, 17, 19, 20, 23
8	4, 5, 7, 8, 10, 17, 18, 21	1, 3, 11, 13, 15, 19, 23, 24	2, 6, 9, 12, 14, 16, 20, 22
9	4, 5, 6, 11, 12, 13, 14, 23	1, 2, 3, 17, 19, 20, 22, 24	7, 8, 9, 10, 15, 16, 18, 21
10	5, 9, 10, 12, 17, 18, 21, 23	1, 6, 7, 13, 15, 16, 20, 22	2, 3, 4, 8, 11, 14, 19, 24

5.4.2 Authentication on Unregistered Drones. Table 5 shows the authentication *recall* on unregistered drones. Compared to the authentication result on the registered drone, the average *recall* for all methods is reduced, except RBF-SVM (98.10%). However, combining the result of the previous experiment, it suggests that RBF-SVM simply classifies almost all drone sounds as unregistered drones to achieve this significant *recall*. Although RF has a relatively poor authentication performance for registered drones (75.30%), it is the best for unregistered drone authentication (84.68%). Meanwhile, QDA has the best overall authentication performance, ranking first for registered drone authentication (94.19%) and third for unregistered drone authentication (82.35%).

6 DISCUSSION AND FUTURE WORK

We achieved drone authentication by applying eight machine learning methods and a fine-tuned feature extraction strategy. Based on the investigation of 8 drones, we found the MFCC settings that work best for drone authentication. Then we showed that the new combination of the drone body and propellers could be regarded as a new drone in terms of acoustic fingerprint. Adding AWGN with varying SNR, we also showed that our proposed methods are resistant to AWGN with SNR greater than 2 dB. Furthermore, the security study showed that our methods could address the problem of open set authentication and detect unregistered drones through their acoustic fingerprints. Next, we discuss the limitations of our research and possible improvements that could be investigated as a direction for future research.

6.1 Complex Outdoor Environment

Our drone audio was collected in the quiet room. Although this room is close to other offices, noise is relatively low. To explore the performance of drone authentication at different levels of noise, we add AWGN with different SNR to our collected drone audio. However, the noise pattern of AWGN is quite monotone compared to the outdoor environment. In real outdoor scenarios, there is some noise similar to drone sounds, such as motorcycle, helicopter, and airplane noise. In addition, a real environment may contain more instantaneous noise, such as bird sounds, building construction noise, and the sound of a car horn. The influence of these real environmental noises could be investigated in the future.

6.2 Repeated Disassembly and Assembly

In order to test different combinations of drone bodies and propellers, we frequently disassembled and assembled drones. During disassembly and assembly, we notice that the wear and tear on the propellers and screws of the drone can also slightly change the sound features. In particular, one screw hole of the "F" drone body slipped during the experiment. This makes it possible to confuse the sound characteristics of drone No. 14 and drone No. 22, which share the same body. This means that we can use this feature to detect the wear and tear of a drone or to check if a registered drone has been modified by someone unauthorized. Meanwhile, the same drone may not pass authentication after a long period of work and wear and tear. This can be a limitation in some cases, but it can also allow improving safety by preventing damaged drones from entering certain areas. An interesting direction for future work is to automatically estimate if it is safe for a drone to proceed into an area based on its acoustic signature.

6.3 Model and Number of Drones

By using DJI Mini 2, our proposed methods achieve a good authentication result. However, the feasibility of our methods for different drone models is uncertain. More drone models could be used in future studies. Meanwhile, using more spare propellers to create more "new" drones is a valid way to further prove our proposed idea.

6.4 Anonymity and Untraceability

Anonymity and untraceability are critical in an authentication system to protect personal privacy. Since only the drone identity is necessary for an acoustic drone authentication system, the owner's identity should remain anonymous. However, by utilizing the drone sound, the identity of drone owners could be traced and revealed. In that case, anonymity and untracebility for acoustic drone authentication should be developed in the future.

6.5 Further Threat Models

In our security study, we assumed that the attacker would use a drone of the same model and try to fool the authentication system. Although this is a natural and reasonable assumption, it can be

Table 4: *Recall* of Authentication on Registered Drones

Experiment No.	QDA (%)	LDA (%)	LSVM (%)	RBF-SVM (%)	KNN (%)	DT (%)	RF (%)	GNB (%)
1	97.16	95.24	97.28	35.41	95.53	75.24	89.30	83.80
2	95.66	91.69	92.49	9.92	93.78	70.12	77.84	83.84
3	94.23	87.78	90.76	2.68	93.05	64.50	71.89	80.60
4	94.30	90.32	91.34	10.71	91.91	65.09	74.86	85.12
5	94.56	88.21	89.19	10.94	93.67	69.93	76.02	85.16
6	92.53	86.85	88.74	6.14	90.44	57.52	67.96	72.48
7	95.56	94.54	94.85	23.84	95.99	72.64	83.30	87.67
8	92.97	86.14	88.64	7.06	90.00	58.48	69.31	69.70
9	94.29	92.24	92.42	18.53	95.52	69.98	82.34	88.72
10	90.60	83.39	85.40	5.31	87.28	54.80	60.21	66.05
Average	94.19	89.64	91.11	13.05	92.71	65.83	75.30	80.31

Table 5: *Recall* of Authentication on Unregistered Drones

Experiment No.	QDA (%)	LDA (%)	LSVM (%)	RBF-SVM (%)	KNN (%)	DT (%)	RF (%)	GNB (%)
1	99.78	98.51	99.69	100.00	97.96	87.45	99.88	79.21
2	77.93	60.90	57.64	99.43	46.57	56.23	73.85	20.64
3	73.85	63.79	68.06	99.71	42.60	59.74	85.33	16.52
4	90.95	84.81	84.50	100.00	73.82	73.98	94.05	44.09
5	75.89	60.10	67.33	99.95	52.77	60.60	72.40	16.23
6	87.64	67.85	72.71	96.74	65.12	74.06	90.16	60.47
7	77.86	63.19	60.98	98.76	55.03	60.11	80.46	16.72
8	84.50	58.08	56.49	95.58	55.02	56.06	91.33	41.40
9	65.93	53.98	45.34	96.94	29.71	38.56	70.34	8.64
10	89.19	53.75	51.62	93.85	50.55	57.99	88.95	35.52
Average	82.35	66.50	66.44	98.10	56.92	62.48	84.68	33.94

extended further. Feng et al. [15] noted that voice-based authentication systems are inherently under threat of impersonation attacks and replay attacks. In addition, Pradhan et al. [29] addressed the importance of defending the replay attack and designed a voice-liveness detection system, named *REplay-resilient VOice Legitimacy Tester* (REVOLT), to prevent replay attacks on wearable devices. Furthermore, Lu et al. [25] proposed a user authentication system, called *VocalLock*, using acoustic features to face replay attacks on smartphones.

Since we used drone acoustic features as the basis for drone authentication, the system has the potential to be attacked by replaying drone audio. The further threat model could be the combined attack model, in which an attacker tries to use both the impersonation attack and the replay attack. Additionally, this attacker can navigate drones to a target location and place the replay device(s) near the authentication system (microphone).

7 CONCLUSION

This was the first work to demonstrate that flying drones can be authenticated via their acoustic fingerprint. To address the feature extraction configuration, we compared the authentication *accuracy* of different parameter settings for the extraction of MFCC, DMFCC, and DDMFCC. QDA, LDA, and LSVM outperformed other methods in terms of *accuracy* in closed set 8 drone authentication without AWGN. Furthermore, LDA outperformed other methods under the influence of AWGN with 0 dB SNR.

Based on that, we decided to use (1) a frame length of 1000 ms, (2) 50% frame overlap, (3) 201 filters, and (4) only 2 to 201 MFCC features as the feature extraction setting. Then, we conducted the closed set experiment, authentication of 24 drones, to show that the new combination of the drone body and propellers could be regarded as a new drone. QDA, LDA, LSVM, and KNN outperformed other methods in terms of *accuracy*, *precision*, *recall*, and *F1-score*, which are all greater than 90%, in closed set 24 drone authentication without AWGN. Meanwhile, applying AWGN, KNN became the most effective method when SNR is less than -3 dB.

Furthermore, the security study result showed that QDA outperformed other methods in terms of the highest average *recall* (94.19%) in registered drones and the third highest average *recall* (82.35%) in unregistered drones.

REFERENCES

[1] Federal Aviation Administration. 2021. UAS Remote Identification Overview. Retrieved April 29, 2022 from https://www.faa.gov/uas/getting_started/remote_id/

[2] Khan Suhail Ahmad, Anil S. Thosar, Jagannath H. Nirmal, and Vinay S. Pande. 2015. A unique approach in text independent speaker recognition using MFCC feature sets and probabilistic neural network. In *2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR)*. 1–6.

[3] Sara Al-Emadi, Abdulla Al-Ali, and Abdulaziz Al-Ali. 2021. Audio-Based Drone Detection and Identification Using Deep Learning Techniques with Dataset Enhancement through Generative Adversarial Networks. *Sensors* 21, 15 (2021).

[4] Sara Al-Emadi, Abdulla Al-Ali, Amr Mohammad, and Abdulaziz Al-Ali. 2019. Audio Based Drone Detection and Identification using Deep Learning. In *2019 15th International Wireless Communications Mobile Computing Conference (IWCMC)*. 459–464.

[5] 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.

[6] Guest Author. 2020. Drones find rising role in agriculture. Retrieved March 24, 2022 from https://www.beefcentral.com/ag-tech/drones-and-automated-vehicles/drones-find-rising-role-in-precision-agriculture/

[7] V Baiocchi, D Dominici, and M Mormile. 2013. UAV application in post-seismic environment. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Archives)* 1 (2013).

[8] Andrea Bernardini, Federica Mangiatordi, Emiliano Pallotti, and Licia Capodiferro. 2017. Drone detection by acoustic signature identification. *Electronic Imaging* 2017, 10 (2017), 60–64.

[9] John Brainard, Ari Juels, Ronald L. Rivest, Michael Szydlo, and Moti Yung. 2006. Fourth-Factor Authentication: Somebody You Know. In *Proceedings of the 13th ACM Conference on Computer and Communications Security (CCS)*. Association for Computing Machinery, 168–178.

[10] BRWC. 2019. How Drones Are Being Used In The Film Industry. Retrieved March 24, 2022 from https://battleroyalewithcheese.com/2019/08/how-drones-are-being-used-in-the-film-industry/

[11] Eduardo Castelló Ferrer. 2018. The blockchain: a new framework for robotic swarm systems. In *Proceedings of the future technologies conference*. Springer, 1037–1058.

[12] Carlos Felipe Emygdio de Melo, Tulio Dapper e Silva, Felipe Boeira, Jorgito Matiuzzi Stocchero, Alexey Vinel, Mikael Asplund, and Edison Pignaton de Freitas. 2021. UAVouch: A Secure Identity and Location Validation Scheme for UAV-Networks. *IEEE Access* 9 (2021), 82930–82946.

[13] DJI. 2022. DJI Mini 2. Retrieved May 1, 2022 from https://www.dji.com/uk/mini-2?from=store-product-page

[14] Martins Ezuma, Fatih 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*. IEEE, 1–13.

[15] Huan Feng, Kassem Fawaz, and Kang G. Shin. 2017. Continuous Authentication for Voice Assistants. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking (MobiCom)*. Association for Computing Machinery, 343–355.

[16] Eitan Frachtenberg. 2019. Practical drone delivery. *Computer* 52, 12 (2019), 53–57.

[17] Lars Fuhrmann, Oliver Biallawons, Jens Klare, Reinhard Panhuber, Robert Klenke, and J Ender. 2017. Micro-Doppler analysis and classification of UAVs at Ka band. In *2017 18th International Radar Symposium (IRS)*. IEEE, 1–9.

[18] Md Afzal Hossan, Sheeraz Memon, and Mark A Gregory. 2010. A novel approach for MFCC feature extraction. In *2010 4th International Conference on Signal Processing and Communication Systems (ICSPCS)*. IEEE, 1–5.

[19] Omar Adel Ibrahim, Savio Sciancalepore, and Roberto Di Pietro. 2022. Noise2Weight: On detecting payload weight from drones acoustic emissions. *Future Generation Computer Systems (FGCS)* 134 (2022), 319–333.

[20] 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)*. IEEE, 1858–1862.

[21] Juhyun Kim, Cheonbok Park, Jinwoo Ahn, Youlim Ko, Junghyun Park, and John C Gallagher. 2017. Real-time UAV sound detection and analysis system. In *2017 IEEE Sensors Applications Symposium (SAS)*. IEEE, 1–5.

[22] 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 (IMWUT)*. 5, 1 (mar 2021), 31 pages.

[23] Harini Kolamunna, Junye Li, Thilini Dahanayaka, Suranga Seneviratne, Kanchana Thilakaratne, Albert Y. Zomaya, and Aruna Seneviratne. 2020. AcousticPrint: Acoustic Signature Based Open Set Drone Identification. In *Proceedings of the 13th ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec)*. Association for Computing Machinery, 349–350.

[24] Dongkyu Lee, Woong Gyu La, and Hwangnam Kim. 2018. Drone detection and identification system using artificial intelligence. In *2018 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 1131–1133.

[25] Li Lu, Jiadi Yu, Yingying Chen, and Yan Wang. 2020. VocalLock: Sensing Vocal Tract for Passphrase-Independent User Authentication Leveraging Acoustic Signals on Smartphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*. 4, 2 (jun 2020), 24 pages.

[26] 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).

[27] James Lyons, Darren Yow-Bang Wang, Gianluca, Hanan Shteingart, Erik Mavrinac, Yash Gaurkar, Watcharapol Watcharawisetkul, Sam Birch, Lu Zhihe, Josef Hölzl, Janis Lesinskis, Henrik Almér, Chris Lord, and Adam Stark. 2020. *jameslyons/python_speech_features: release v0.6.1.*

[28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[29] Swadhin Pradhan, Wei Sun, Ghufran Baig, and Lili Qiu. 2019. Combating Replay Attacks Against Voice Assistants. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*. 3, 3 (sep 2019), 26 pages.

[30] Soundarya Ramesh, Thomas Pathier, and Jun Han. 2019. SoundUAV: Towards Delivery Drone Authentication via Acoustic Noise Fingerprinting. In *Proceedings of the 5th Workshop on Micro Aerial Vehicle Networks, Systems, and Applications*. Association for Computing Machinery, 27–32.

[31] Catur Aries Rokhmana. 2015. The Potential of UAV-based Remote Sensing for Supporting Precision Agriculture in Indonesia. *Procedia Environmental Sciences* 24 (2015), 245–253. The 1st International Symposium on LAPAN-IPB Satellite (LISAT) for Food Security and Environmental Monitoring.

[32] Judy Scott and Carlton Scott. 2017. Drone delivery models for healthcare. In *Proceedings of the 50th Hawaii international conference on system sciences (HICSS)*.

[33] 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)*. IEEE, 1–7.

[34] Yoojeong Seo, Beomhui Jang, and Sungbin Im. 2018. Drone detection using convolutional neural networks with acoustic stft features. In *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 1–6.

[35] Nappaphol Siriphun, Shigeru Kashihara, Doudou Fall, and Assadarat Khurat. 2018. Distinguishing drone types based on acoustic wave by IoT device. In *2018 22nd International Computer Science and Engineering Conference (ICSEC)*. IEEE, 1–4.

[36] 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.

[37] 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 (2020), 236–245.

[38] Dana Utebayeva, Akhan Almagambetov, Manal Alduraibi, Yelmurat Temirgaliyev, Lyazzat Ilipbayeva, and Sungat Marxuly. 2020. Multi-label UAV sound classification using Stacked Bidirectional LSTM. In *2020 Fourth IEEE International Conference on Robotic Computing (IRC)*. IEEE, 453–458.

[39] Bowon Yang, Eric T Matson, Anthony H Smith, J Eric Dietz, and John C Gallagher. 2019. UAV detection system with multiple acoustic nodes using machine learning models. In *2019 Third IEEE International Conference on Robotic Computing (IRC)*. IEEE, 493–498.

[40] Abbas Yazdinejad, Reza M Parizi, Ali Dehghantanha, and Hadis Karimipour. 2021. Federated learning for drone authentication. *Ad Hoc Networks* 120 (2021), 102574.