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

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

ACSAC '22, December 5–9, 2022, Austin, TX, USA
© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *Annual Computer Security Applications Conference (ACSAC '22), December 5–9, 2022, Austin, TX, USA*, 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].

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

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

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 eld of drone detection, researchers were trying to compare di erent combinations of feature extraction schemes and di erent classi cation 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 e ect of di erent 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 in uence 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 Coe cients (LPCC) and MFCC. Then, they applied SVM with di erent kernels to compare the performance of LPCC and MFCC with 13 features. The results showed that the detection performance of MFCC surpassed LPCC with di erent kernels. This work reached the conclusion that machine learning is an e cient and accurate tool in the eld 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 Classi cation

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 in uence on the detection and classi cation performance. However, to be more convincing, other classi cation 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 arti cial dataset with the aim of improving the performance of drone detection and classi cation. The result suggested that the bene ts of using GAN to augment datasets outweighed the drawbacks in drone detection and drone classi cation. Utebayeva et al. [38] were the rst group to use Long Short-Term Memory (LSTM) to classify drone sound. They extracted the sound features by MFCC, but the details of the con guration 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 the LSTM model and trained with the background class to solve the open set classi cation. Based on their previous work, Kolamunna et al. [22] developed DronePrint based on LSTM for drone detection and classi cation. The authors analyzed the drone sound characteristic and discussed the in uence of peak normalization in the time domain and the rescaling of the feature vector in MFCC. They used 40 lters 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 classi cation. Furthermore, DronePrint showed resistance to the Doppler e ect 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 classi cation.

2.3 Other Drone Sound Related Studies

Ibrahim et al. [19] were the rst to detect the payload of commercial drones according to the sound features. They extracted 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, LDA, LSVM, and quadratic SVM outperformed other methods with an accuracy of 98.9%. They found that a longer frame length was e ective 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 di erent 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 ies 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 classi cation, however, none have addressed the issue of authentication on ying drones. In this paper, we use MFCC-related features as the acoustic ngerprint for drone authentication. Previous work has already shown that MFCC is an e ective 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 con guration details of MFCC were not clear enough in many of the above-mentioned works. Most of the works did not provide speci c details on the number of used lters [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 e ectiveness of DMFCC and DDMFCC [18] in drone sound analysis. In our work, we discuss the e ect of the following four settings on drone authentication: (1) the frame length of segmented audio, (2) the number of used lters, (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 ngerprint, 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 di erent 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 rst 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 o ce 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 nd out which con guration for feature extraction is appropriate. Then we used DS2 and DS2N to verify the con guration 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 rst 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 di erent con gurations in this part, we use accuracy as a performance indicator:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

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 classi cation [3, 4, 22].

The next step is to split each audio le 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 nd the appropriate frame length, we calculate the authentication accuracy in di erent 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 lters 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 xed, 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 in uence the authentication accuracy: (1) the number of lters, (2) the number of used features, and (3) the use of high-level features. The number of lters determines the number of total generated features. In all experiments, all the rst 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 in uence. We divide the feature vector by its L_2 norm. The formula is shown below:

$$E'_{Feature} = \frac{E_{Feature}}{\|E_{Feature}\|}.$$

Figure 2: The in uence of frame length on accuracy.

where $E_{Feature}$ is the feature vector and $E'_{Feature}$ 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 lters remain unknown for most of the work. However, according to our experiment, adopting these numbers leads to poor accuracy. To nd 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 lter-varying experiment without AWGN is conducted in DS1, and the lter-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 lters 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 in uence of the number of lters and the number of used features. We name this process lter-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 xed number of lters. 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 lter-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 lter-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 lters 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 e ect of DMFCC and DDMFCC is unclear, because six machine learning algorithms achieve a similar accuracy using the same con gurations with an increasing number of lters. 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 e ect of MFCC and DMFCC, we apply the same lter-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 e ect under the in uence of noise. The results are shown in Figure 4. Compared to the lter-varying experiment without AWGN, the accuracy of using a small number of lters and features is greatly reduced. In this case, we can see the importance of using a large number of lters and features. Furthermore, the addition of AWGN has a great in uence on DT, which has a signi cant decrease in drone authentication accuracy. In addition, LDA shows strong resistance to noise and remains the highest accuracy in all con gurations.

These results show that as the number of lters and used features increases, the accuracy of each algorithm almost continuously increases. Their accuracy reaches a bottleneck of about 200 used lters. Although DMFCC and DDMFCC are e ective in speaker recognition [2], their implementation does not have a positive e ect 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 lters 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 rst experiment in this section will con rm that the drone with replaced propellers has a unique acoustic ngerprint, i.e., it can be regarded as a new drone. Then we explore the in uence of AWGN with di erent 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:

$$\%A428B8>= = \frac{)\%}{)\%\ ,\ \%} \bullet \quad (2)$$

$$'420;; = \frac{)\%}{)\%\ ,\ \#} \bullet \quad (3)$$

$$1 = \frac{1}{2}\ ^{1}\ \frac{1}{'420;;}\ ,\ \frac{1}{\%A428B8>=}\ ^{o} \bullet \quad (4)$$

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 justi es the feasibility of using an acoustic ngerprint 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 ngerprint. 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 ve methods is greater than 80%, and for QDA, it is greater than 95%. The results show that QDA outperformed the other methods signi cantly 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 in uence 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 di erent 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 signi cantly low while maintaining a relatively high accuracy. QDA and LSVM are two

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