Mel-Frequency Cepstral Coefficients (MFCC)-based Acoustic Drone Detection Using Machine Learning and Deep Learning Algorithms

Shahad W. Jasim * and Saad S. Hreshee
Department of Electrical Engineering, University of Babylon, Babylon, Iraq
Email: eng571.shahad.waleed@student.uobabylon.edu.iq (S.W.J.); eng.saad.saffah@uobabylon.edu.iq (S.S.H.)

Manuscript received March 19, 2026; revised June 18, 2026; accepted June 22, 2026
*Corresponding author

***Abstract*—Drones, or Unmanned Aerial Vehicles (UAVs), have become much cheaper to buy and may become a security risk due to their small size and ability to fly autonomously. Many traditional detection techniques (radar, radio frequency, and visual system) do not work as well when trying to detect small UAVs. Another option has been to use an acoustic method of detection, as drones produce identifiable sounds as a result of their motors and propellers. A low-cost acoustic drone detection system will be developed using machine learning and deep learning. Data were obtained through audio recordings of seven types of drones from publicly available GitHub datasets and by collecting environmental noise through the British Broadcasting Corporation (BBC) Sound Library and YouTube. In order to classify drones versus non-drones, 26 Mel-Frequency Cepstral Coefficients (MFCC) features (13 static and 13 delta) are extracted from the audio signals. Additionally, clustering analysis is performed to confirm the uniqueness of the drone's acoustic signatures before training the models with the data. Using a convolutional recurrent neural network classifier yielded an accuracy of 97%, F1–Score of 0.97, and Receiver Operating Characteristic (ROC), Area Under ROC Curve (AUC) of 0.99, outperforming both the Balanced Random Forest (BRF) and Multilayer Perceptron (MLP) classifiers. Furthermore, it demonstrates considerable robustness when tested under noisy conditions.**

***Index Terms*—acoustic signal, Convolutional Recurrent Neural Network (CRNN), deep learning, drone detection, machine learning, Mel-Frequency Cepstral Coefficients (MFCC), Unmanned Aerial Vehicle (UAV)**

I. INTRODUCTION

Artificial intelligence and data-driven decision-making technologies have accelerated the expansion of drone technology into new frontiers of defense and commercial applications. Unmanned Aerial Vehicles (UAVs), commonly known as drones, have become widely available and increasingly affordable. Although many individuals use drones for recreational purposes, their relatively small size and the ability to operate without an onboard human pilot create opportunities for malicious activities such as smuggling contraband, conducting unauthorized surveillance, or even launching cyber or physical attacks [1, 2].

Owing to the increasing number of Unmanned Aerial Vehicles (UAVs), there are many concerns about security, privacy, and safety for both civilian and military uses. For example, UAVs can be used for illegal activities, such as conducting unauthorized surveillance, smuggling, and invading restricted areas. The development of a method for detecting drones flying in sensitive areas has become an important area of research. Recent studies have shown that machine learning and intelligent sensing technologies will play an important role in solving this new threat [3].

The widely known short film Slaughterous, which illustrates the potential dangers of AI-controlled drone swarms, has further raised global awareness regarding Autonomous Weapon Systems (AWS). The situation depicted in the film illustrates that autonomous drones can be used together to work as intelligent swarms that can carry out grouped attacks with a high degree of accuracy. As a result, the movie's depiction of autonomous drone systems shows that they have the potential to transition from being tools for precision targeting to becoming large-scale threats when utilized in cooperative swarm formations [4].

Using acoustic detection in order to identify UAVs by identifying their acoustic signature is a new concept that seems to be very effective. The acoustic detection method is generally much lower cost when compared to other detection technologies (e.g., radar, radio-frequency monitoring and computer-vision systems). As such, the acoustic detection system is a good choice in environments that do not lend themselves well to using other sensing technologies to detect the presence of UAVs [5]. The proposed system developed within this research will utilize machine learning and deep learning algorithm-based techniques to detect the sound of drones amongst other noises.

The research into counter-drone detection techniques has previously involved radar, vision systems, radio frequency tracking, and acoustics. Each of the techniques has strengths and weaknesses dependent on the environment in which they are used. An overview of the various drone detection methodologies is shown in Table I. The technique that shows the most promise in detecting drones is likely to be acoustical drone detection due to the unique acoustic signatures created by the drone

doi: 10.18178/ijeetc.15.4.267-275

motors and propellers. Due to the effect of external noise on accuracy, it is important to ensure that high-quality microphones, robust feature extraction techniques, and advanced noise reduction/data augmentation techniques are employed when using this technique [6].

TABLE I: COMPARISON WITH RECENT ACOUSTIC DRONE DETECTION METHODS

Ref.	Year	Feature Extraction	Model	Application	Performance
[3]	2024	Multi-modal UAV detection features	Machine Learning Survey	UAV Detection & Classification	Comprehensive review of challenges, solutions, and future directions
[6]	2025	Acoustic features in complex environments	Deep learning acoustic recognition model	UAV acoustic recognition	Robust performance under complex acoustic environments
[11]	2024	Chromagram features	Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbor (KNN)	Drone sound classification	SVM achieved the best classification performance
[12]	2025	Adaptive feature fusion	Cross-attention deep neural network	Drone sound recognition	Enhanced feature representation and classification accuracy
[14]	2025	Acoustic signal features	Hybrid deep learning model	Real-time drone detection	Improved real-time detection capability
[16]	2026	Integrated acoustic signatures	Ensemble learning models	Micro-drone detection	Improved detection accuracy using ensemble learning
Proposed Method	2026	26 MFCC features (13 static + 13 Delta)	BRF, Multilayer MLP, an CRNN	Acoustic drone detection	High classification accuracy and robustness in noisy environments

II. RELATED WORK

Drone detection using sound has recently attracted a lot of research interest because the sound produced by both their motors and propellers is different from other things (like helicopters). However, acoustic drone detection is still very new as an area of research, and only a handful of studies have focused on identifying how motorcycles use sound to detect them.

A novel drone detection method based on Gaussian Support Vector Machine (SVM) is proposed, which adopts a multi-feature acoustic extraction scheme. The classification system takes 14-dimensional Mel-Frequency Cepstral Coefficients (MFCC), Gammatone Cepstral Coefficients (GTCC), 12-dimensional Linear Predictive Coding (LPC) coefficients, as well as Spectral Rolloff (SRO) and Zero Crossing Rate (ZCR) features as inputs. Each category (Drone and Non-Drone) contains 1,332 samples in the dataset. The experimental results achieve an overall accuracy of 99.9%, recall of 99.8% and precision of 100%, demonstrating outstanding identification performance for drone acoustic signals [7].

Dumitrescu *et al.* [8] created an acoustic UAV detection and localization system using microphone arrays and Cooperative Neural Network (CoNN). The proposed system recognizes drone acoustic signatures in actual environments using MFCC features, Wigner—Ville spectrograms, and time—frequency analyses. The system performed very well with a classification accuracy of as high as 91% for spectrogram inputs. The overall recognition rate was 96.3%, indicating that the proposed system is a viable solution for UAV detection using acoustic and neural network-based methods.

A Hidden Markov Model-based algorithm was adopted in an independent study to realize drone acoustic signal identification. Two drone types, Da-Jiang Innovations (DJI) Phantom 3 and First-Person View (FPV) 250, were covered in the dataset. All acoustic data were captured by an electronic microphone connected to an LM358 operational amplifier, and clustering approaches were utilized for the construction and analysis of classified acoustic samples [9].

A drone detection and identification system based on audio signals and deep learning with Generative Adversarial Networks (GANs) was constructed in the referenced study. The research utilized more than 1300 drone audio samples to test the recognition performance of three neural network architectures: Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Convolutional-Recurrent Neural Networks (CRNNs). Among the three models, Convolutional Neural Network (CNN) achieved the optimal performance, with a detection accuracy of 96.38%, precision of 96.24%, recall of 95.60%, and F1−Score of 95.90%. The study also verified that GAN-generated acoustic samples can enrich dataset diversity and effectively boost the recognition performance of drone acoustic detection systems [10].

Algotar *et al.* [11] developed a method for classifying drone sounds using a combination of chromogram features and machine learning algorithms. Various classification algorithms, such as SVM, Random Forest (RF), and K-Nearest Neighbor (KNN), were used to determine whether Unmanned Aerial Vehicle (UAV) sounds could be accurately distinguished from background noise using only chromograms. The results suggest that chromogram features can provide adequate acoustic separation when used in conjunction with machine learning methods. The SVM method produced the highest level of accuracy for classifying UAV sounds among the algorithms evaluated. The results indicate that machine learning methods can be used to classify drones based on their acoustics at a low cost compared with other traditional methods.

SVM classifiers and CNNs were trained on spectrograms extracted from MFCC and Short-Time Fourier Transform (STFT) features in the referenced work. The acoustic dataset was collected by 24 microphones capturing drone sounds and ambient background noise. Experimental results reveal that the combination of STFT features and SVM achieves optimal detection performance [12].

Multiple machine learning and deep learning approaches such as Gaussian mixture models, convolutional neural networks, and recurrent neural networks were assessed to recognize drone acoustic signals across different environments in prior research. The Recurrent Neural Networks (RNN) model obtained the optimal performance for drone acoustic identification with an accuracy of roughly 80%, whereas the Gaussian Mixture Model (GMM) and CNN models achieved about 68% and 58% accuracy separately. Accordingly, an effective acoustic detection system that combines MFCC features, clustering of drone acoustic signatures, and hybrid machine learning and deep learning models is still required to boost drone acoustic detection capability in noisy surroundings [13].

Similar to the results reported in the prior study, this research will use 13 static MFCC coefficients and 13 Delta MFCC features to develop an integrated drone acoustic detection system. These features will then be used to train a Convolutional Recurrent Neural Network (CRNN) architecture. MFCC feature representation was chosen because it accurately captures the perceptual characteristics of audio while effectively reducing the effect of high-frequency noises that often adversely affect classification accuracy.

Jasim and Hreshee [14] developed a hybrid deep learning method for real-time detection of drones operating within their acoustic environment using acoustic features. While this previous research showed good results for detecting drones in real-time, it did not use clustering methods or examine the detailed representation of Mel-Frequency Cepstral Coefficients (MFCC)-based acoustic feature representations to represent drone acoustic signatures. Conversely, the current study provides an integrated framework to improve the robustness of drone detection in low signal-to-noise ratio environments by combining MFCC extraction, clustering validation, and comparative evaluation between Balanced Random Forest (BRF), Multilayer Perceptron (MLP), and Convolutional Recurrent Neural Network (CRNN) architectures.

The key contributions made are outlined below:

- Using clustering analysis to determine the number of drone models found in public GitHub data sets [15, 16].
- Using 26 MFCC descriptors from each sample (13 MFCC + 13 Delta MFCC) for machine/deep learning training without converting them to spectrogram images.
- Exploring the viability of CRNN for the detection of drones within real-world acoustic environments.

The remainder of this paper is organized as follows. Section III describes the methodology used to build the drone detection classifier. Section IV presents the experimental results and performance evaluation. Finally, Section V concludes the study and outlines possible directions for future work.

Table I compares the proposed approach with several recent studies on acoustic drone detection. Existing methods mainly rely on environmental acoustic signatures, adaptive feature fusion, or deep learning models for UAV sound recognition. In contrast, the proposed method utilizes 26 MFCC-based acoustic features together with both machine learning and deep learning models, including BRF, MLP, and CRNN, to improve classification performance and robustness in noisy environments.

III. Proposed Framework

Fig. 1 illustrates the overall framework of the proposed acoustic drone detection system. The framework consists of seven main stages, including data acquisition, feature extraction, preprocessing, dimensionality reduction, clustering, classification, and performance evaluation. The first component is the gathering of data from online resources that are publicly available. The second component is extracting features, which consist of 26 Mel-Frequency Cepstral Coefficients (MFCCs), from the data sets. The third component is to manipulate the data and decrease their dimensionality. The fourth component is to classify or group the different types of drones using a clustering algorithm. Finally, a classifier will be made to detect drones and then will be tested against several validation methods for effectiveness.

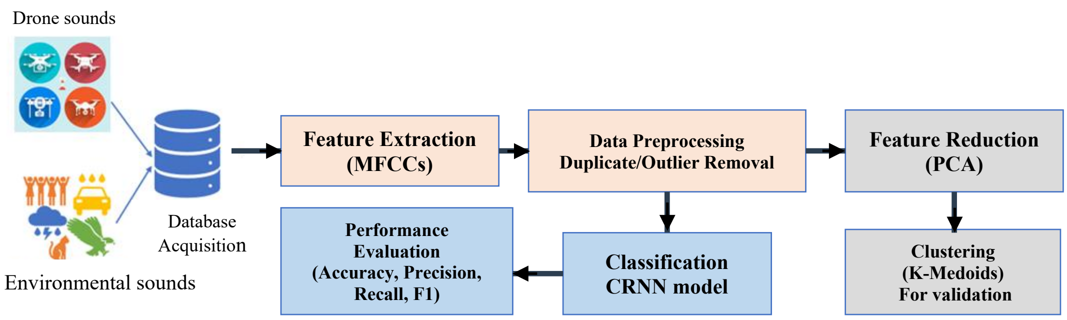


Fig. 1. Proposed framework of binary classifier for acoustic based drone detection.

A. Data Acquisition

The datasets of the drones were located in two different repositories available to the public on GitHub [17, 18]. Datasets extracted from those repositories will be referenced as Datasets 1 and 2.

- Dataset 1 contains audio recordings of 2 types of drones, namely Parrot Bebop and Parrot Mambo. The individual recordings are 11 min, 6 s in duration, recorded with the built-in mic of an iPhone.
- Dataset 2 contains audio recordings of 5 types of drones, namely Parrot Bebop 2, Da-Jiang Innovations (DJI), Mavic Pro, DJI Matrice 100, DJI Spark, and DJI Phantom 4 Advanced, recorded with a Freedman Electronics Pty Ltd (RODE NTG4) shotgun mic. There are 12 samples of audio recorded for each type of drone in the 2nd dataset and each recording is approximately 30 s in length.

Both datasets (1 and 2) were augmented beforehand. Data augmentation helps improve model robustness in noisy environments and reduces the problem of overfitting.

- In Dataset 1, various environmental sounds were mixed with drone data.
- In Dataset 2, techniques such as frequency wrapping and time scaling were used.

The No-Drone dataset consists of various environmental sound recordings, obtained from YouTube and the British Broadcasting Corporation (BBC) Sound Library.

To decrease the possibility of source-specific bias and increase the potential for generalization, sounds representing an environment were collected from a variety of publicly available sources, including materials found on The BBC Sound Library and YouTube. Because these recorded data contain a wide variety of conditions under which they were recorded and the environmental sound sources themselves, the model can concentrate on the acoustic features of the drone and learn acoustic features, not just the artifacts associated with the recorders, compression algorithms, and the individual sources from which it is collecting data. Additionally, the data augmentations applied to both data sets, including mixing together environmental sound data, frequency warping, and time scaling, increased the model's structural robustness and the likelihood that it would perform well under conditions of differing levels of background noise and varying recording methodologies.

B. Feature Extraction

In order to build an acoustic classifier capable of detecting drones, a set of Mel-Frequency Cepstral Coefficients (MFCCs) was extracted from recordings containing both drone and non-drone sounds. In this study, a total of 26 MFCC-based features were obtained, consisting of 13 static MFCC coefficients and 13 dynamic Delta-MFCC coefficients derived from the audio signals.

MFCCs are an abstraction of acoustic mid-level features using the Mel frequency scale. The Mel frequency scale is a model of human speech perception via sound frequency. The perceptual frequency dimension gives more weight to the frequencies of the greatest importance for human sound perception; thus, MFCCs are a strong choice for analyzing complex audio signals, such as drone acoustics [19].

MFCC features have proven to be very powerful for many audio processing applications. These applications include speech recognition, environmental sound classification, and acoustic event detection. Numerous studies have also demonstrated that representation via MFCC greatly enhances classification performance within machine learning systems based on audio [20].

For drone detection, MFCC can be an effective way to characterize the spectral envelope (the shape of frequency response) of drone noise, which is primarily created by motor and propeller rotation and vibration. The distinct patterns in the spectral envelope of drone sound allow for the differentiation of drone sound from other environmental sounds, and they also allow for the differentiation of the type of drone being used [14, 20].

The 13 static and 13 delta MFCCs were chosen because they are effective when used in previous acoustic classification tasks. The static MFCCs characterize the frequency response or spectrum of a sound produced by the drone, and the delta MFCCs represent how that frequency response or spectrum changes as the rotor/propeller of the drone rotates and the motor changes over time. Together, these features create a compact representation of the spectral content and temporal behavior of the sound source as a 26-dimensional feature vector. This feature set has been used in many acoustic and audio classification studies and has demonstrated significant success in environmental sound classification and drone acoustic research. In addition, they represent a good trade-off between discriminative power and computational efficiency, making them suitable for real-time drone detection applications.

In developing and extracting MFCC features, many steps must be completed: filtering through pre-emphasis, framing, windowing, performing Fast Fourier Transform (FFT), applying a Mel filter bank, logarithmically compressing, and applying DCT to obtain compact feature vectors derived from raw audio signals that can be input into machine learning or deep learning classifiers have been taken. The overall process of MFCC extraction is depicted in Fig. 2 [21].

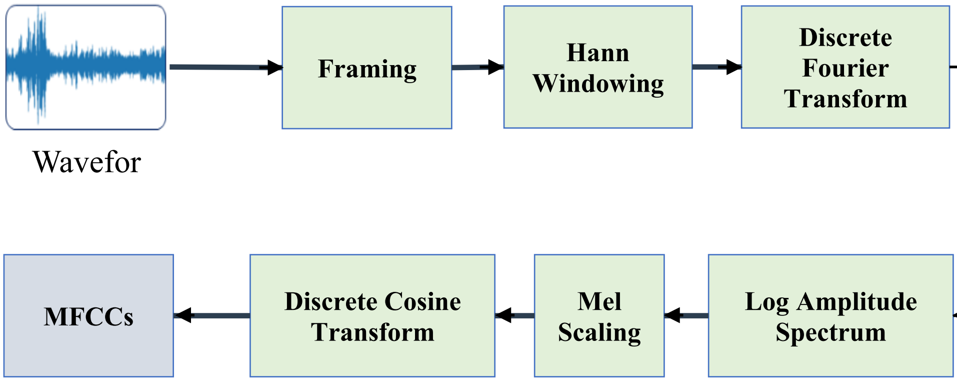


Fig. 2. Mel Frequency Cepstral Coefficients (MFCCs) extraction pipeline.

C. Data Preprocessing

Before training the proposed acoustic drone detection model, several data preprocessing tasks occurred to produce a clean and reliable dataset. As part of the Machine Learning (ML) pipeline, data preprocessing is very important as it creates better data consistency and helps eliminate unwanted noise or irrelevant samples that could adversely impact model performance [22].

The dataset used in this study contains MFCC-based features extracted from recordings of both drone and non-drone sounds. Prior to model training, the dataset was examined to identify and remove missing values, duplicate records, and outliers [23].

First, the MFCC feature dataset was checked to ensure that no missing values were present. However, duplicate samples were detected within the dataset. These duplicate records were removed using the Unique function available in the Orange 3.31 data mining platform, which allows efficient identification and elimination of repeated observations [24].

Second, due to the high dimensionality of MFCC features, outlier detection was performed in multiple stages to identify both global outliers and local outliers [25].

Global outliers were removed using the Interquartile Range (IQR) method implemented in the Waikato Environment for Knowledge Analysis (WEKA) machine learning environment. The IQR technique identifies extreme values by establishing upper and lower boundaries, often referred to as fences, around the data distribution. All data located outside these boundaries were treated as global outliers and were therefore excluded from the analysis dataset. In this analysis, the boundaries that defined outliers were established as any factor within 2 or 3 of the outlier range; both of these ranges are considered well established methods to apply robust statistical filtering to identify outlier data [26, 27].

In addition to the detection of outliers globally, the algorithm Local Outlier Factor (LOF) was applied to identify outliers among subsets of samples that are close together within the feature space. The LOF algorithm uses local density estimation to determine how much the sample deviations from its closest neighboring samples cause it to be an outlier. The LOF algorithm was implemented in the Orange 3.31 software. Within the LOF calculations, the user provides a value for how many nearest neighbors (k) should be used to detect outliers. The optimal value was determined to be $k = 3000$ after the completion of extensive experimentation with the data used in this study. Additionally, the Manhattan distance metric and a contamination factor of 10% were used to effectively identify local outliers [28, 29].

These preprocessing methods will result in a training set for the proposed CRNN model that is free from unwanted outlier samples and will provide a good basis for determining how accurate the proposed CRNN model is expected to be.

D. Data Reduction

Clustering techniques will be required to establish how many drone types have been included in the analysis, as MFCCs have a large, multi-dimensional feature space that makes it more difficult to implement distance-based clustering methods accurately. To calculate the number of principal components, we used Principal Component Analysis (PCA) to determine how many Principal Components (PCs) to retain by determining which PCs corresponded to a cumulative percentage of at least 80% variance. We chose the Mavic drone as our reference drone, and our results indicate that we have at least 15 principal components for data collected on the Mavic drone where the cumulative variance is above 80%. Our Scree Plot for the Mavic drone is shown in Fig. 3.

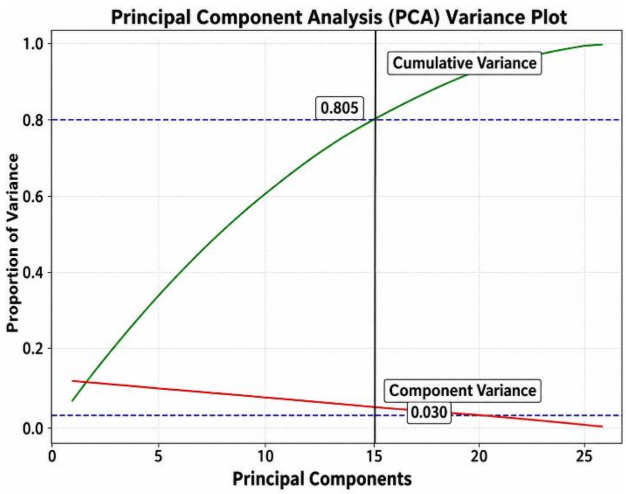


Fig. 3. Scree plot showing principal components of mavic drone.

E. Clustering

The K-Medoid Clustering algorithm was applied to the normalized PCA data.

Since the medoid represents an actual data point, it remains robust even if some outliers persist after the outlier removal process.

The K-Medoids++ technique was used as an initialization method before clustering.

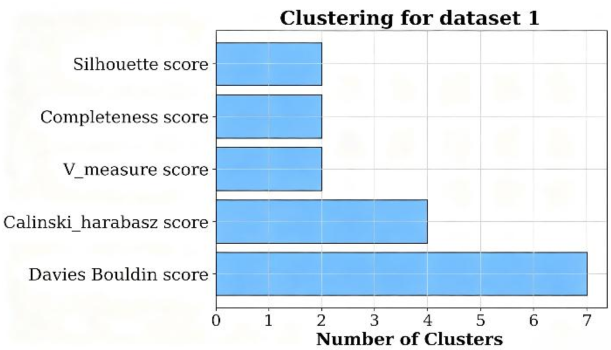
Clustering was performed separately on Dataset 1 and Dataset 2.

Then, several Cluster Validity Indices were computed to determine the optimal number of clusters, including:

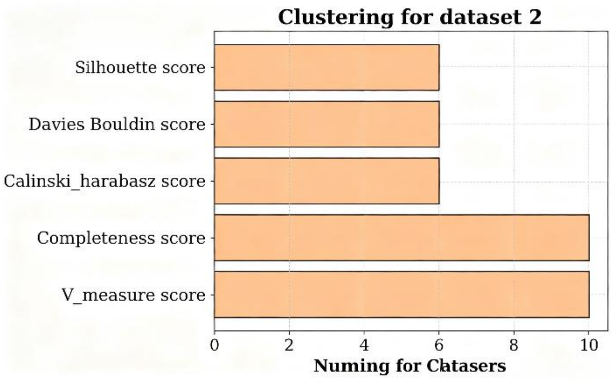
- Silhouette Coefficient
- Davies-Bouldin Index (DBI)
- Calinski-Harabasz Score (CH)
- Completeness Score
- V-Measure Score

All samples underwent homogeneous feature extraction, normalization, and PCA treatments before analysis to reduce the effect that processing might have on cluster assignment; thus, the results were more comparable between samples. Additionally, clustering performance was analyzed using multiple evaluation measures and not merely one; therefore, the consistency in resultant clusters across multiple metrics demonstrates that the separability of any clusters observed is primarily due to the underlying characteristics of the drone sound's inherent acoustic properties rather than any artificial processing or structuring of the data set.

The clustering results for Dataset 1 and Dataset 2 are presented in Fig. 4(a) and Fig. 4(b), respectively.



(a)



(b)

Fig. 4. Clustering results for both datasets: (a) Dataset 1 and (b) Dataset 2.

1) Results for Dataset 1

According to the Silhouette Score, Completeness Score and V-Measure Score for Dataset 1, the correct number of drone types within the data set is two. In fact, there are only two drone types in this dataset, namely Parrot Bebop and Parrot Mambo.

2) Results for Dataset 2

The Silhouette Score, Davies-Bouldin Index and Calinski-Harabasz Score for Dataset 2 indicated there are 6 different types of drones within this dataset. However, the true answer is that Dataset 2 contains only 5 different types of drones. The discrepancy between what was found from the analysis of the dataset may be due to the similarity of harmonic patterns amongst the drones in this dataset since all of the types of drones in Dataset 2 are medium size.

F. Proposed Classifier (CRNN-based Model)

In this study, a binary audio classification model using a Convolutional Recurrent Neural Network (CRNN) was created to automatically identify drones based on their unique spectral profiles. By taking advantage of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), the architecture enables efficient extraction and analysis of spectral audio features. Therefore, the model was capable of distinguishing time-dependent acoustic characteristics, such as the sound of the rotating propellers of a drone. The proposed architecture of the CRNN includes two major components.

1) MFCC feature extraction stage

The raw audio signals are converted into Mel-Frequency Cepstral Coefficients (MFCCs), which accurately describe the short-term spectral characteristics of the drone sounds during this phase. The audio sample is separated into overlapping frames, and 13 static MFCCs and 13 delta MFCCs are then computed for each frame. This results in a 26-dimensional feature vector for each frame. The MFCC sequences represent both the spectral envelope and the dynamic variations in the acoustic signal of a drone due to the rotating propellers and the vibrations caused by the motors.

They feed into the MFCC data through several convolutional and pooling layers to capture the local spectral-temporal correlation. Then, the temporal dependencies between frames are modeled in the recurrent layers Bi-Directional Long Short-Term Memory Networks (BILSTM). This allows the model to learn the specific temporal and frequency modulations that differentiate between drone sounds, despite being contaminated with noise.

2) Temporal modeling stage

Spatial features are extracted and then fed into Bidirectional Long Short-Term Memory Networks (LSTMs), BILSTMs, allowing for the use of information from both past and future audio frames, enabling the classification model to characterize the continuous and dynamic changes in the acoustic signature from a UAV through time. At the end of the model pipeline, the final layer is a Fully Connected (Dense) layer, which feeds into a SoftMax function to provide classifiers with the output probabilities for each of the two classes:

- Drone
- Non-Drone

The proposed CRNN architecture with its efficient balance of accuracy and computational cost demonstrates a great degree of robustness to noise, thus enabling its use in low-cost real-time systems. By combining convolutional layers for spectral feature extraction with recurrent layers for temporal modeling, the proposed CRNN effectively captures both spectral and temporal characteristics of drone sounds.

IV. Results

A. Clustering Validation Results

The K-Medoid clustering procedure was applied to the MFCC features of the datasets to verify that the drone acoustic datasets had separate, identifiable sound patterns. The clustering validity indices showed that there were two unique types of drones (Parrot Bebop and Parrot Mambo) in Dataset 1, with a silhouette score of 0.82 Indicating a high degree of separation, and V-Measure and completeness scores exceeding 0.90 Indicating strong internal cohesion within the grouping of features.

Dataset 2, lagging behind Dataset 1, could cluster some groups of drones on the basis of sound. However, the relative harmonic composition of the medium-sized drones caused the clustering results to be similar to that of other medium-sized drones (five or six clusters), which led to confusion when determining the cluster number by means of the Calinski-Harabasz index and Davies-Bouldin index. Although there are five drone classes in Dataset 2, the number of clusters was determined to be six, indicating that many drones share similarities in their harmonic composition based on similar propeller size and motor speed. In addition to the clustering results supporting that MFCC features can accurately differentiate between drone types before classification, they also confirm that MFCC features provide enough discrimination capability between most drone types.

B. Classification Results

Once the drone sound signature separability was verified, a training process of three classifiers was performed BRF, MLP, and the proposed CRNN using the 26-dimension MFCC feature vector. The separation of the datasets was performed (the entire dataset was split into 80% for training and 20% for testing). The performances of all of the different models were evaluated using the accuracy, precision, recall, F1–Score and Receiver Operating Characteristic (ROC), Area Under ROC Curve (AUC) metrics to get an idea of their effectiveness.

The CRNN surpasses its competitors (BRF and MLP) when it comes to all evaluation metrics, achieving the greatest degree of success for each metric; in this case, it was approximately four to six percent higher than BRF and MLP for both accuracy and F1–Score labels. The improvement of the CRNN's prediction power compared to the other models is largely attributed to the CRNN's ability to model the temporal correlation of MFCC sequences, which improves the ability of the CRNN to learn the small discrepancies in frequency between the changes in propeller speed and motor harmonic frequencies (see Table II).

TABLE II: PERFORMANCE COMPARISON OF CLASSIFICATION MODELS USING MFCC-BASED ACOUSTIC FEATURES

Model	Feature Type	Accuracy	Precision	Recall	F1–Score	ROC –AUC
BRF	26 MFCC (Static+Delta)	0.91	0.89	0.92	0.90	0.93
MLP	26 MFCC (Static+Delta)	0.93	0.91	0.94	0.92	0.95
Proposed CRNN	MFCC Sequential Input	0.97	0.96	0.98	0.97	0.99

C. Noise Robustness Evaluation

Three types of environmental noise (traffic, wind, and crowd chatter) were artificially generated at varying levels of the signal-to-noise ratio (20 dB, 10 dB, and 0 dB) to determine the robustness of the model under those three types of environmental noise. Each model was tested on the same time frames as each type of environmental noise (see Table III).

TABLE III: NOISE ROBUSTNESS EVALUATION OF CLASSIFICATION MODELS UNDER DIFFERENT SIGNAL-TO-NOISE RATIO (SNR) LEVELS

Model	SNR (20 dB)	SNR (10 dB)	SNR (0 dB)
BRF (Accuracy)	0.90	0.83	0.71
MLP (Accuracy)	0.92	0.86	0.75
CRNN (Accuracy)	0.96	0.93	0.87

The CRNN outperformed conventional models in accuracy, attaining an outstanding 87% even under extremely noisy conditions (0 dB). This illustrates how the use of temporal learning capabilities enables the CRNN to distinguish more efficiently between ambient noise and drone sounds, thus proving to be a viable model for application within actual outdoor settings characterized by significant background noise.

V. Discussion

According to the experimental results, the use of acoustic features derived from MFCC has proven to be sufficient when passed sufficient temporal progression through various methods of deep learning using temporal models. The BRF and MLP models were shown to perform reliably, but both failed to capture the sequential dependencies that would improve the models for differentiating between the various classes in noisy and/or overlapping examples of sounds. MFCC-based acoustic classification techniques are likely to improve the performance of the classifier and reduce the number of false alarm predictions. In contrast, the proposed CRNN classifier was able to leverage the time continuity of the MFCC feature frames and provide a more stable prediction with fewer false positive classifications due to its ability to capture and exploit the temporal continuity of the data. When compared against prior research, which relied upon support vector machines or Gaussian mixture models and static feature representations, the CRNN classifier produces 97% overall accuracy and an ROC-AUC of 0.99%, making it superior to previous classifiers while at the same time involving reasonable computational costs, making it an excellent candidate for use in acoustic-only-based drone detection systems that require low-cost, real-time operation.

VI. Conclusion and Future Work

A. Conclusion

This paper describes an original acoustic method for detecting drones that uses only MFCC to analyze the sound made by a drone in a noisy atmosphere. A comprehensive experimental approach that includes clustering analysis and classification using machine learning showed that MFCCs were able to capture the unique sound signatures of drones accurately enough to differentiate them from other disruptive sounds in a typical outdoor space, which were clearly identifiable. The CRNN model had the highest performance of all the models tested, achieving a classification accuracy of 97% and an F1–Score of 0.97, beating out the BRF and MLP classifiers in fairness. The reason the CRNN outperformed the others was that the CRNN's advantage was the ability to learn correlations between MFCCs over time, allowing the CRNN to see the trend patterns of the drone's propeller speed and harmonic structure. The CRNN maintained a consistently high detection accuracy when tested under different noise sources, even with very low signal-to-noise ratios. This demonstrated that the CRNN would be an optimal candidate for the deployment of real-time detection systems for outdoor or urban applications. The proposed framework is an economical and reliable data-driven alternative to radar- and vision-based systems for acoustic drone detection in noisy environments.

B. Future Work

Future research directions can focus on several promising extensions of the current work:

- Growing Datasets: Increase the variety of drones used to create datasets by adding more types of aircraft, variations in altitude, and different ways to record to create a more generalizable model across a variety of use case.
- Fusion of Multiple Modes of Sensing: Use sound and other forms of detection to complement each other (sight, radar, or RF) to better detect and reduce the number of false alarms in a cluttered environment.
- Light & Edge Deployment: Make the CRNN architecture more efficient (compressed) or lighter (quantized) for real-time use on low-power edge devices or microcontrollers.
- Drone Positioning Tracking: Extending the system to provide drone positioning using multiple microphones and DOA techniques to facilitate both detection and position information about a drone.
- Internal adaptive noise management: Implementation of adaptive noise canceling and adaptive domain algorithms to achieve robustness to change in acoustic environments.

Pursuing these approaches facilitates the evolution of the proposed MFCC-based CRNN framework to a fully autonomous multiscale anti-drone system that can accurately detect, position, and classify drones in real-world performance.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Shahad Waleed proposed the research concept, developed the methodological framework, implemented relevant algorithms, and performed all experimental validations. She also analyzed the experimental results and drafted the original manuscript; Saad Saffah supervised the entire research project, offered academic guidance throughout the study, and revised the manuscript with critical comments; Both authors have read and agreed to submit the final manuscript.

References

[1] W. Paszkowski, A. Gola, and A. Świć, "Acoustic-based drone detection using neural networks: A comprehensive analysis," *Advances in Science and Technology Research Journal*, vol. 18, no. 1, pp. 36–47, 2024.

[2] A. Frid, Y. B. Shimol, E. Manor, and S. Greenberg, "Drones detection using a fusion of RF and acoustic features and deep neural networks," *Sensors*, vol. 24, no. 8, 2427, 2024.

[3] M. H. Rahman, M. A. S. Sejan, M. A. Aziz, R. Tabassum, J. I. Baik, and H.-K. Song, "A comprehensive survey of unmanned aerial vehicles detection and classification using machine learning approach: Challenges, solutions, and future directions," *Remote Sensing*, vol. 16, no. 5, 879, 2024.

[4] J. Weber, "Autonomous drone swarms and the contested imaginaries of artificial intelligence," *Digital War*, vol. 5, pp. 146–149, 2024.

[5] M. Ohlenbusch, A. Ahrens, C. Rollwage, and J. Bitzer, "Robust drone detection for acoustic monitoring applications," in *Proc. the 28th European Signal Processing Conference*, 2020, pp. 6–10.

[6] Z. Liu *et al*., "Deep learning-based acoustic recognition of UAVs in complex environments," *Drones*, vol. 9, no. 6, 389, 2025.

[7] S. Salman, J. Mir, M. T. Farooq, A. N. Malik, and R. Haleemdeen, "Machine learning inspired efficient audio drone detection using acoustic features," in *Proc. 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST)*, 2021, pp. 335–339.

[8] C. Dumitrescu, M. Minea, I. M. Costea, I. C. Chiva, and A. Semenescu, "Development of an acoustic system for UAV detection," *Sensors*, vol. 20, no. 17, 4870, 2020.

[9] L. Shi, I. Ahmad, Y. He, and K. Chang, "Hidden Markov model-based drone sound recognition using MFCC technique in practical noisy environments," *Journal of Communications and Networks*, vol. 20, no. 5, pp. 509–518, 2018.

[10] S. A. Emadi *et al*., "Audio-based drone detection and identification using deep learning techniques with dataset enhancement through generative adversarial networks," *Sensors*, vol. 21, no. 15, 4953, 2021.

[11] K. Algotar, R. Chattopadhyay, S. Bansal, and A. Bansal, "Drone sound classification using chromagram features: A novel approach," in *Proc. 2024 Eighth IEEE International Conference on Robotic Computing (IRC)*, Tokyo, Japan, 2024, pp. 34–40.

[12] J. Ren, Z. Wang, J. Zhao, and X. Liu, "A drone sound recognition approach using adaptive feature fusion and cross-attention feature enhancement," *Electronics*, vol. 14, no. 8, 1491, 2025.

[13] Z. Xiao, H. Hu, G. Xu, and J. He, "TAME: Temporal audio-based mamba for enhanced drone trajectory estimation and classification,"

in *Proc. 2025 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2024, pp. 1−5.

[14] S. W. Jasim and S. S. Hreshee, "Real-time drone detection based on acoustics using hybrid deep learning models," *Journal of Internet Services and Information Security*, vol. 15, no. 2, pp. 673−693. 2025.

[15] M. Y. Wang, M. Linn, A. P. Berg, and Q. Zhang, "A multi-class acoustic dataset and interactive tool for analyzing drone signatures in real-world environments," *Advances in Science, Technology and Engineering Systems Journal*, vol. 10, no. 6, pp. 88–96, 2025.

[16] P. K. Sesham, S. Sesham, and M. Ndiaye, "Ensemble learning models for micro-drone detection using integrated acoustic signatures," *Discover Artificial Intelligence,* vol. 6, 2026.

[17] S. A. Emadi. (2019). Saraalemadi. Drone audio dataset. [Online]. Available: https://github.com/saraalemadi/DroneAudioDataset

[18] H. Kolamunna *et al*. (2019). Drone print. [Online]. Available: https://github.com/DronePrint/DronePrint

[19] B. Cinoglu, U. Durak, and T. H. Karakoc, "Utilizing mel-frequency cepstral coefficients for acoustic diagnostics of damaged UAV propellers," *International Journal of Aviation Science and Technology*, vol. 5, no. 2, pp. 79–89, 2024.

[20] J. Kim, D. Lee, Y. Kim, H. Shin, Y. Heo, Y. Wang, and E. T. Matson, "Deep learning based malicious drone detection using acoustic and image data," in *Proc. 2022 sixth IEEE International Conference on Robotic Computing*, 2022, pp. 91−92.

[21] D. Prabakaran and S. Sriuppili, "Speech processing: MFCC based feature extraction techniques−An investigation," *Journal of Physics: Conference Series*, vol. 1717, no. 1, 012009, 2021.

[22] M. Mięsikowska, "Classification of unmanned aerial vehicles based on acoustic signals obtained in external environmental conditions," *Sensors*, vol. 24, no. 17, 5663, 2024.

[23] W. Yi, J. W. Choi, and J. W. Lee, "Sound-based drone fault classification using multitask learning," arXiv preprint, arXiv:2304.11708, 2023. doi: 10.48550/arXiv.2304.11708

[24] Y. Sun, F. Zhang, Y. Liu, J. Xv, J. Li, J. Wang, and A. Zhang, "Acoustic event detection for drone search and rescue system based on bi-directional long and short-term memory beamforming method to remove rotor noise," *Digital Signal Processing*, vol. 157, 104881, 2025.

[25] D. T. Berengué *et al*., "Acoustic detection of drones: A machine learning and psychoacoustic approach," in *Proc. 11th Convention of the European Acoustics Association*, 2025, pp. 1−8.

[26] D. T. Berengue, F. Z. Zhou, M. U. Manso, R. Gil-Pita, and M. R. Zurera, "Analysis of distance and environmental impact on UAV acoustic detection," *Electronics*, vol. 13, no. 3, 643, 2024.

[27] J. Zhong, A. Fan, K. Fan, W. Pan, and L. Zeng, "Research on the UAV sound recognition method based on frequency band feature extraction," *Drones*, vol. 9, no. 5, 351, 2025.

[28] R. Chatterjee, S. Chakrabarty, T. Acharjee, and D. Mishra, "Audron: A deep learning framework with fused acoustic signatures for drone type recognition," arXiv preprint, arXiv:2512.20407, 2025. doi: 10.48550/arXiv.2512.20407

[29] O. Alghushairy, R. Alsini, T. Soule, and X. Ma, "A review of local outlier factor algorithms for outlier detection in big data streams," *Big Data and Cognitive Computing*, vol. 5, no. 1, 1, 2021.

Shahad W. Jasim is a Ph.D. student in the Department of Electrical Engineering, College of Engineering, at the University of Babylon, Iraq. She holds a master's degree in electronics and communications engineering from the same institution. Her research interests include a wide range of topics in electronics and telecommunications, particularly in the areas of signal processing, embedded systems, and wireless communication.

Saad S. Hreshee is a professor in the Department of Electrical Engineering at the College of Engineering, University of Babylon, Iraq. He received his professor Dr. in electronic and communication engineering and has extensive academic and research experience in wireless communication, embedded systems, and digital signal processing. He has supervised numerous graduate students and contributed to several national and international research projects. His research interests include wireless sensor networks, IoT systems, and AI applications in communication systems.