

# Passive acoustic detection and localization of drones using MEMS microphones and machine learning

Zakaria Ghouli<sup>1,2,\*</sup> 

<sup>1</sup>Royal Naval School, Casablanca, Morocco

<sup>2</sup>Polydisciplinary Faculty of Taroudant, University Ibn Zohr, Agadir, Morocco

Received 4 September 2025, Accepted 23 January 2026

**Abstract** – With the rapid proliferation of unmanned aerial vehicles (UAVs) in both civilian and military domains, the demand for efficient detection and tracking systems has become increasingly critical, particularly in sensitive and strategic areas. Conventional surveillance methods, such as radar and infrared sensing, often struggle to detect low-altitude, low-signature UAVs. This study proposes a real-time acoustic localization system based on a distributed array of MEMS microphones. The approach utilizes Time Difference of Arrival (TDOA) estimations to determine the drone's angular position, combined with a Random Forest classifier to distinguish drone acoustics from environmental noise. A radar-style interface was developed to provide real-time visualization of detections. Field experiments confirmed the system's effectiveness under diverse environmental conditions. The solution offers a passive, cost-effective alternative for enhancing situational awareness in maritime and other security-sensitive applications.

**Keywords:** Acoustic detection, Drone localization, MEMS microphones, Time Difference of Arrival (TDOA), Passive surveillance, Random Forest, Signal processing, Anti-drone system, Embedded systems

## 1 Introduction

Unmanned Aerial Vehicles (UAVs), commonly referred to as drones, have experienced exponential growth and adoption across civilian, industrial, and military sectors over the past decade. Initially developed for reconnaissance and surveillance purposes in defense, drones are now used in applications ranging from agriculture, infrastructure inspection, search and rescue, border control, delivery services, cinematography, and even real-time event broadcasting [1, 2].

While UAVs provide transformative advantages, they also pose novel threats. The same features that make drones attractive compactness, maneuverability, autonomy, and affordability also enable their misuse in espionage, smuggling, terrorism, and unauthorized aerial surveillance [3, 4]. Numerous incidents have raised alarm regarding drones penetrating airport zones, military bases, and public events. In conflict zones like Ukraine and Syria, drones have been weaponized and used for targeted strikes or intelligence gathering [5, 6].

Traditional counter-UAV (C-UAV) systems are primarily based on radar, radio frequency (RF) spectrum

monitoring, optical/thermal imaging, and LIDAR. However, each of these has inherent limitations. Radar systems struggle to detect small drones with low radar cross-sections (RCS) or non-metallic structures [7]. Optical systems are limited by line-of-sight and lighting conditions. RF detection is ineffective against autonomous UAVs with preprogrammed missions that do not rely on external control links [8]. These challenges necessitate the exploration of alternative and complementary detection modalities.

Acoustic detection has emerged as a promising passive method, especially suitable for low-altitude, small UAVs in short to medium range scenarios [9, 10]. Drones produce unique acoustic signatures, primarily due to the periodic motion of rotor blades and associated mechanical vibrations. These signatures consist of broadband and harmonic components, often distinguishable from ambient sounds like wind, traffic, or birds [11]. Studies have shown that drone propeller noise can be characterized by strong peaks in the frequency domain, particularly in the 100 Hz to 10 kHz range, with fundamental and harmonic frequencies dependent on rotor speed and blade count [12, 13].

The viability of acoustic detection depends significantly on the quality and geometry of microphone arrays and the accuracy of the signal processing techniques.

\*Corresponding author: [ghoulizakaria@gmail.com](mailto:ghoulizakaria@gmail.com)

Modern MEMS microphones, such as the INMP441, offer high sensitivity, compactness, low power consumption, and digital output formats like I2S, making them ideal for real-time embedded systems [14, 15]. The use of microphone arrays enables spatial signal processing methods such as beamforming, time difference of arrival (TDOA) estimation, and direction of arrival (DOA) localization [16, 17].

Beyond signal acquisition, the classification and decision-making layer plays a critical role. Traditional methods like FFT-based feature extraction followed by thresholding are susceptible to noise and yield high false alarm rates. As a result, AI-driven techniques have gained traction.

Machine learning has significantly enhanced the capability of acoustic systems to identify UAVs in complex environments. Traditional signal processing techniques are being augmented or replaced by supervised and unsupervised learning models trained on labeled drone and background sounds. For instance, Kandeepan et al. [18] evaluated Random Forest, SVM, and conventional Neural Network architectures such as Multi-Layer Perceptrons (MLPs), demonstrating good baseline performance for multirotor drones. However, recent studies have reported that deep architectures – such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) – generally outperform MLPs when learning discriminative spectral-temporal features from spectrogram-based inputs [13, 19, 20]. CNNs (Convolutional Neural Network), in particular, excel in learning discriminative features from time-frequency representations such as spectrograms or mel-frequency cepstral coefficients (MFCCs) [19]. Hybrid models combining spectral filtering with machine learning have shown great promise for robust performance in real-world conditions [21].

Despite promising results, acoustic detection remains challenged by environmental noise, multipath propagation, and limited range. Mitigation strategies involve multi-stage filtering, adaptive noise cancellation, and fusion with other sensors (e.g., RF, thermal, visual). Furthermore, real-time operation requires efficient embedded implementations of feature extraction, classification, and localization algorithms, as well as synchronized multichannel data acquisition.

Recent research has also explored the application of machine learning and embedded sensing technologies for real-time monitoring and anomaly detection in various engineering domains. For instance, Amrane et al. [22] developed a monitoring system for solenoid valve coil resistance using optical fiber squeezers and machine learning algorithms, demonstrating the potential of intelligent sensing in fault detection. Similarly, Ouldzira et al. [23] implemented remote object monitoring using a wireless sensor network based on NodeMCU ESP8266, highlighting the effectiveness of low-cost IoT architectures for real-time data acquisition. These studies, although in different application domains, share methodological similarities with the present work, particularly regarding

embedded signal processing, wireless communication, and machine learning-based decision making.

In this study, we address these challenges by presenting a fully integrated system that combines:

- A multichannel MEMS microphone array optimized for spatial coverage;
- A real-time Python-based pipeline for signal acquisition and feature extraction;
- A machine learning model (Random Forest) trained on drone vs. noise acoustic datasets;
- A 2D localization algorithm based on TDOA triangulation;
- A dynamic radar-like interface for operator feedback.

The system is validated through experimental trials in realistic environments, including open-air and semi-urban scenarios, demonstrating its robustness and applicability for naval and homeland security operations.

The remainder of this paper is structured as follows. [Section 2](#) provides a comprehensive overview of existing UAV detection techniques, with a focus on acoustic sensing. [Section 3](#) outlines the theoretical principles underlying drone acoustic localization. [Section 4](#) describes the system’s hardware and software components. In [Section 5](#), we detail the signal processing workflow and machine learning-based classification. [Section 6](#) presents the localization strategy and reconstruction of drone positions. [Section 7](#) reports on experimental validations and performance analysis. Finally, [Section 8](#) discusses the findings, and [Section 9](#) concludes the study with possible directions for future research and system enhancement.

## 2 Related work and state of the art

The detection and classification of unmanned aerial vehicles (UAVs) has become a dynamic field of research, with a wide variety of methods developed to address the growing risks associated with their proliferation. These approaches can be broadly categorized into four main domains: radar-based, RF-based, visual/thermal imaging, and acoustic-based systems. Each technique offers distinct advantages and limitations, and recent research trends suggest that hybrid and AI-enhanced systems offer the most promising performance in real-world scenarios.

### 2.1 Radar and RF-based methods

Radar systems have been the cornerstone of aerial object detection for decades. They operate by emitting electromagnetic waves and analyzing the reflected signals to determine the presence, range, speed, and direction of airborne objects. However, the low radar cross-section (RCS) of small, plastic-bodied drones makes them difficult to detect using traditional radar systems. Moreover, radar is less effective in cluttered urban environments and may lead to high false alarm rates due to reflections from buildings or natural obstacles [24, 25].

RF-based methods rely on intercepting the control and telemetry signals exchanged between a UAV and its ground controller. These techniques are advantageous in detecting commercial off-the-shelf drones that operate in common frequency bands such as 2.4 GHz and 5.8 GHz. Nevertheless, RF methods fail when UAVs operate autonomously without active communication links, or when signal encryption and frequency hopping techniques are employed [26].

## 2.2 Optical and infrared imaging

Optical and infrared (IR) systems capture visual or thermal imagery to detect UAVs based on shape, movement, or heat signatures. These systems offer high spatial resolution and are particularly effective in open environments with good visibility. IR systems can detect drones at night or in low-light conditions by exploiting the thermal emissions of their motors and electronics. However, both optical and IR approaches suffer from occlusion, limited range, weather dependency, and high computational costs for real-time processing [27, 28].

## 2.3 Acoustic-based methods

Acoustic detection has emerged as an attractive alternative due to its passive nature, low cost, and independence from lighting and electromagnetic conditions. Drones emit characteristic acoustic signatures primarily generated by the propellers, motors, and frame vibrations. These sounds consist of a fundamental frequency related to rotor speed and harmonics due to blade count and structural interactions [10]. Hanson et al. [11] showed that even small drones produce distinguishable spectral features within the 100 Hz to 10 kHz range.

Several studies have explored the use of microphone arrays to perform sound localization via Time Difference of Arrival (TDOA) and beamforming techniques. Palacios et al. [16] implemented a circular MEMS array and demonstrated 2D directional accuracy using beamforming. Kwon and Kim [29] used signal energy and cross-correlation to achieve robust detection under variable noise conditions. Other works have applied adaptive filtering, Kalman tracking, and particle filters to enhance detection robustness [30].

Recent advances include the use of acoustic holography, spherical microphone arrays, and distributed sensor networks to extend spatial coverage and accuracy [31]. These systems can be deployed in static surveillance configurations or integrated into mobile platforms such as vehicles or UAVs.

To provide a clearer comparison of the main detection modalities discussed above, Table 1 summarizes the primary advantages and limitations of radar-, RF-, optical/infrared-, and acoustic-based UAV detection systems.

This comparative summary highlights that while radar and RF techniques excel in range and identification, acoustic sensing offers a passive, low-cost, and complementary approach particularly suited for close-range or covert surveillance scenarios.

## 2.4 AI and machine learning for acoustic UAV detection

Machine learning has significantly enhanced the capability of acoustic systems to identify UAVs in complex environments. Traditional signal processing techniques are being augmented or replaced by supervised and unsupervised learning models trained on labeled drone and background sounds. For instance, Kandeepan et al. [18] evaluated Random Forest, SVM, and CNN models, demonstrating high classification accuracy for multirotor drones. While their study focused on [controlled environments / a single drone type / short-duration recordings], the present work extends these results by [evaluating models under real-world noisy conditions, including multiple drone types, or performing long-term acoustic monitoring], highlighting the robustness and generalization of the proposed approach.

Deep learning models, especially CNNs, have been applied to spectrogram inputs (e.g., STFT, mel spectrograms) to capture temporal and spectral patterns. These models outperform handcrafted feature approaches and show robustness to varying recording conditions and background noises [13]. Moreover, recurrent architectures such as LSTM and GRU networks have been employed to learn sequential dependencies in acoustic data [20].

Some hybrid systems fuse acoustic inputs with RF, visual, or thermal features using ensemble learning or sensor fusion techniques to improve reliability and reduce false alarms. Tseng et al. [32] developed a multi-modal detection framework using both audio and video streams, achieving real-time detection with high accuracy in cluttered environments.

## 2.5 Limitations and research gaps

While the literature demonstrates promising results, challenges remain. Environmental noise (e.g., wind, traffic, crowds), reverberation, and microphone sensitivity variability can degrade performance. In addition, most systems are tested in controlled conditions and may not generalize well to field scenarios. Few studies address power constraints and real-time requirements necessary for embedded or autonomous deployment. Additionally, the classification of multiple drones, identification of drone types, and 3D localization remain open problems.

In summary, acoustic-based UAV detection systems, particularly when enhanced with machine learning, represent a promising research direction. However, their effectiveness in real-world environments depends on sensor

**Table 1.** Comparative overview of UAV detection modalities.

Detection method	Advantages	Limitations
Radar	Long detection range; works in all weather conditions; provides speed and range estimation.	Limited detection of small or non-metallic drones; high cost; high false alarms in cluttered areas.
RF	Detects control and telemetry signals; enables drone identification; suitable for cooperative targets.	Ineffective against autonomous or encrypted UAVs; susceptible to frequency hopping; limited range.
Optical/Infrared	High spatial resolution; visual confirmation possible; effective for target recognition.	Strongly affected by weather, lighting, and occlusions; high computational load.
Acoustic	Passive and covert; low cost; independent of lighting and RF environment; suitable for low-altitude detection.	Limited range; sensitive to environmental noise and wind; affected by reverberation.

configuration, signal processing robustness, and classifier generalization. This work builds upon these findings and proposes a compact, real-time system leveraging MEMS arrays and Random Forest classification, validated through realistic experimentation.

### 3 Theoretical background

This section presents the theoretical principles underlying the acoustic detection and localization of UAVs. It addresses the acoustic emission characteristics of drones, the fundamentals of sound propagation, and the mathematical formulation of Time Difference of Arrival (TDOA) for 2D localization using microphone arrays. Additionally, it includes the rationale behind the choice of MEMS microphones and the preprocessing required for reliable signal capture and analysis.

#### 3.1 Acoustic signature of UAVs

Multirotor drones generate acoustic emissions primarily from two components: the rotating propellers and the electric motors. The dominant sound is typically a tonal signal resulting from periodic blade passages (Blade Passage Frequency, BPF) and its harmonics. The BPF can be approximated by:

$$f_{\text{BPF}} = N_b \times f_r \quad (1)$$

where  $N_b$  is the number of blades and  $f_r$  is the rotational speed in revolutions per second. In addition to tonal components, broadband noise is generated due to turbulent airflow, blade-vortex interactions, and structural vibrations.

These acoustic signatures are influenced by UAV design (e.g., propeller diameter, frame materials), flight maneuvers, and environmental factors. While frequencies between 100 Hz and 10 kHz are typically observed, most energy concentrates in the low kHz range, making it detectable with MEMS microphones and suitable for machine learning-based classification [11].

#### 3.2 Sound propagation and microphone array geometry

In outdoor environments, sound propagates through air as spherical waves, attenuating over distance and subject to environmental effects such as wind, temperature gradients, and ground reflections. The attenuation follows the inverse square law:

$$A(d) = \frac{A_0}{d^2} \quad (2)$$

where  $A(d)$  is the amplitude at distance  $d$ , and  $A_0$  is the amplitude at a reference distance. Wind and temperature gradients can induce refraction, affecting directionality. In addition to geometric attenuation, air absorption, which is frequency-dependent, affects the propagation of drone sounds. Higher-frequency components are absorbed more rapidly than lower-frequency components, altering the spectral content received by acoustic sensors. Wind and temperature gradients can also induce refraction, affecting directionality.

To estimate the direction of arrival (DOA) or localize a sound source, a set of spatially separated microphones is required. In this work, a planar array of MEMS microphones arranged in a geometric configuration (e.g., square or circular) is used. The array spacing must be selected to avoid spatial aliasing, adhering to the criterion:

$$\Delta d < \frac{c}{2f_{\text{max}}} \quad (3)$$

where  $c$  is the speed of sound (approximately 343 m/s), and  $f_{\text{max}}$  is the maximum frequency of interest.

#### 3.3 Time Difference of Arrival (TDOA) estimation

TDOA-based localization relies on computing the time delay between signals received at different microphones. Assuming a source at location  $S(x, y)$  and microphones at known positions  $M_i(x_i, y_i)$ , the TDOA between two microphones  $i$  and  $j$  is:

$$\tau_{ij} = \frac{\|S - M_i\| - \|S - M_j\|}{c} \quad (4)$$

The time delays are estimated using cross-correlation between microphone signals:

$$\tau_{ij} = \arg \max_{\tau} \int s_i(t) s_j(t + \tau) dt. \quad (5)$$

These delays are then used in multilateration algorithms or linearized least squares methods to estimate the source coordinates. Localization accuracy depends on sampling rate, microphone synchronization, environmental noise, and array geometry [16].

### 3.4 MEMS microphone characteristics

Micro-Electro-Mechanical Systems (MEMS) microphones offer several advantages for UAV detection systems: compact size, low power consumption, digital output (I2S or PDM), wide frequency response, and cost-effectiveness. The INMP441 used in this project provides a flat frequency response between 60 Hz and 15 kHz, with a signal-to-noise ratio (SNR) around 62 dB and high sensitivity ( $-26$  dBFS). Its I2S interface simplifies synchronization and integration with microcontrollers [14].

Proper mounting, orientation, and wind shielding are critical to preserving signal quality. Additionally, external amplification and filtering stages can be used to enhance dynamic range and reduce environmental noise impact.

### 3.5 Spectral analysis and feature extraction

To prepare acoustic data for machine learning classification, relevant features must be extracted from the raw audio signal. The Short-Time Fourier Transform (STFT) is widely used to convert time-domain signals into time-frequency representations:

$$X(n, \omega) = \sum_{m=-\infty}^{+\infty} x(m) \omega(n - m) e^{-j\omega m} \quad (6)$$

where  $\omega(n)$  is a window function. STFT outputs are used to generate spectrograms, mel spectrograms, or Mel-Frequency Cepstral Coefficients (MFCCs), which serve as input to classifiers. The short-time Fourier transform (STFT) decomposes a signal into time-frequency components, providing a spectrogram that shows how the signal's spectral content evolves over time. Mel spectrograms map the frequency axis to the perceptually motivated Mel scale, emphasizing frequencies more relevant to human hearing. MFCCs further summarize the spectral envelope by taking the discrete cosine transform of the log-magnitude Mel spectrum, producing compact features commonly used in audio classification [33, 34]. Feature selection is crucial to maximize inter-class separability (drone vs. background).

This theoretical foundation supports the development of a real-time acoustic UAV detection and localization system, combining spatial, spectral, and statistical properties of audio data.

## 4 System architecture

The system architecture is designed to provide real-time detection and localization of UAVs using a network of synchronized MEMS microphones, embedded computing, and machine learning algorithms. This section describes the hardware components, signal acquisition pipeline, processing modules, and interface design.

### 4.1 Hardware components

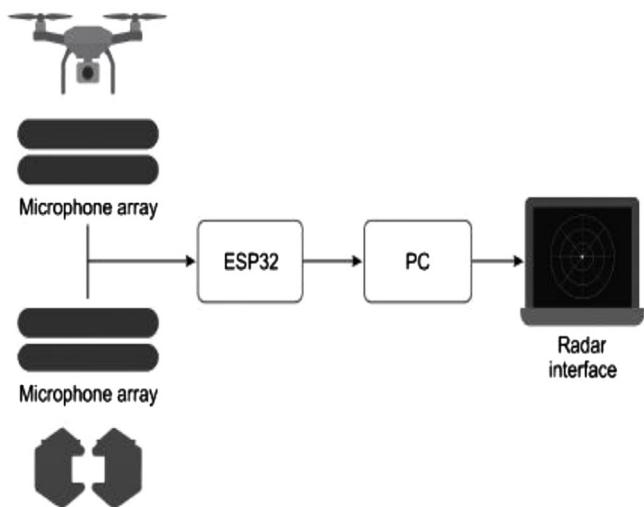
The selection of components in this system is guided by their performance in prior research and technical documentation. MEMS microphones like the INMP441 have been shown to be effective for low-altitude drone detection due to their compactness and digital output interface [14, 15]. The acoustic detection system is built around the following key hardware elements (Fig. 1):

- **MEMS Microphones (INMP441):** Eight omnidirectional MEMS microphones are arranged in a square planar array. Each microphone supports digital I2S output, which simplifies interfacing with microcontrollers and ensures synchronized data capture.
- **Microcontroller Unit (ESP32-S3):** The ESP32-S3 microcontroller is used due to its high processing speed, low power consumption, and native support for I2S and multi-threaded data handling. It is responsible for managing data acquisition and communication with the processing unit.
- **Raspberry Pi 4:** This unit acts as the central processing node. It receives the raw I2S audio streams from the ESP32-S3 and performs signal processing, feature extraction, machine learning classification, and localization calculations.
- **Power Supply:** A portable battery pack powers the system, making it suitable for field deployment. Voltage regulation and EMI shielding are implemented to minimize noise interference.

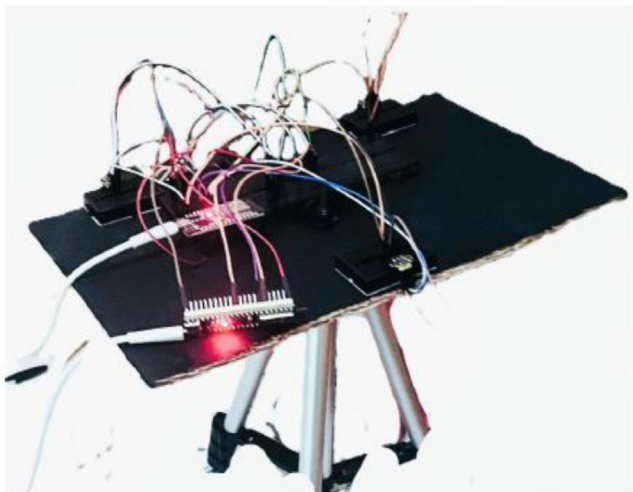
### 4.2 Microphone array configuration

The eight microphones are arranged in a square configuration with an inter-element spacing of 15 cm was chosen for the microphone arrays, which allows spatially unambiguous detection of acoustic signals up to approximately 1.1 kHz, according to  $f_{\max} = c/(2d)$ , where  $c = 343$  m/s is the speed of sound and  $d = 0.15$  m is the inter-element spacing. This geometry offers a compromise between spatial resolution and size, enabling the estimation of Time Difference of Arrival (TDOA) values with sufficient angular accuracy.

To avoid spatial aliasing and enhance TDOA resolution, the array design considers the Nyquist criterion in the spatial domain. The configuration also minimizes directional ambiguity and enables coverage over a 180-degree horizontal plane (Fig. 2).



**Figure 1.** System architecture for passive UAV detection, showing microphone arrays and the signal processing pipeline.



**Figure 2.** Geometry of the MEMS microphone array used for TDOA localization. The annotated elements indicate microphone positions, inter-element spacing ( $d = 15$  cm), and reference coordinate axes. This configuration enables accurate angular localization of UAV acoustic sources.

### 4.3 Data acquisition and preprocessing

Accurate and synchronized data acquisition is crucial for reliable TDOA-based localization. The use of interrupt-driven I2S acquisition on the ESP32-S3 ensures low jitter and consistent timing, aligning with best practices in embedded acoustic signal processing [6]. Windowing and segmentation of signals are standard techniques to mitigate spectral leakage, as described in classic signal processing literature [35].

Audio signals from each MEMS microphone are sampled at 48 kHz and digitized via the I2S interface. The ESP32-S3 manages synchronous acquisition using a custom-built interrupt-driven routine.

The digitized signals are transmitted via serial interface or Wi-Fi to the Raspberry Pi, where they are buffered and segmented into overlapping frames (typically 1024–2048 samples per frame). Each frame is windowed using a Hamming function to reduce spectral leakage.

### 4.4 Signal processing pipeline

The signal processing pipeline leverages well-established audio analysis methods and modern classification techniques. Feature extraction using mel-frequency cepstral coefficients (MFCCs) is a common approach in environmental sound classification, shown to be effective in conjunction with machine learning models [19]. The choice of a Random Forest classifier is motivated by its robustness and interpretability in pattern recognition tasks, as demonstrated in multiple studies [36].

The main processing steps executed on the Raspberry Pi include:

- (1) **Noise Reduction:** Adaptive filtering and spectral subtraction techniques are used to reduce background noise and improve signal-to-noise ratio.
- (2) **Feature Extraction:** Time-frequency features such as spectrogram slices and mel-frequency cepstral coefficients (MFCCs) are extracted. MFCCs are computed to capture the spectral characteristics of the acoustic signals: the audio signal is divided into short overlapping frames, each multiplied by a Hamming window. For each frame, the power spectrum is calculated using FFT and passed through a Mel-scale filter bank to mimic human auditory perception. The logarithm of the filter outputs is taken, and finally, a Discrete Cosine Transform (DCT) is applied to obtain the cepstral coefficients, typically retaining the first 12–13 coefficients which contain the most relevant spectral information.
- (3) **Classification:** A pre-trained Random Forest classifier analyzes the extracted features to determine the presence of UAV acoustic signatures. A Random Forest classifier was chosen due to its robustness, interpretability, and efficiency in handling small to medium-sized datasets, making it suitable for implementation on embedded devices like the Raspberry Pi. While Support Vector Machines (SVM) and Multilayer Perceptrons (MLP) are alternative classifiers, preliminary tests on similar datasets suggest comparable performance. State-of-the-art approaches such as Convolutional Neural Networks (CNNs) could potentially achieve higher classification accuracy, particularly on spectrogram representations; however, their higher computational complexity is beyond the scope of this concept study and is left for future work.
- (4) **Localization:** Cross-correlation is performed between pairs of microphone signals to estimate TDOA. These delays are then fed into a multilateration algorithm to estimate the drone's 2D position relative to the array center.

#### 4.5 User interface and visualization

The final component is a graphical interface developed using Python (Tkinter or PyQt). It includes:

- A radar-style display showing the estimated direction and distance of detected UAVs.
- Real-time audio spectrum visualization.
- Status logs including confidence scores from the classifier, TDOA values, and array synchronization metrics.

The interface is designed for use by defense personnel or security operators and supports both touchscreen and desktop modes.

The proposed system is designed to be low-cost and energy-efficient. It relies on a single Raspberry Pi and MEMS microphones, which are significantly less expensive than professional-grade acoustic arrays used in other UAV detection systems. Energy efficiency is achieved by performing real-time signal processing and classification on the embedded device using lightweight algorithms such as Random Forest, avoiding the need for high-power computing resources. In comparison, other approaches employing deep learning or multiple high-end sensors often require more expensive hardware and higher energy consumption.

This system architecture integrates low-cost, energy-efficient components with a robust real-time processing framework. The modularity of the system allows for easy upgrades (e.g., replacing the Random Forest with a deep neural network) and adaptation to various environments.

The computational performance of the system was evaluated on a Raspberry Pi 4 Model B equipped with a 1.5 GHz quad-core ARM Cortex-A72 CPU and 4 GB RAM. Real-time operation was achieved with an average CPU utilization of 65–70% and memory usage below 1 GB during simultaneous audio acquisition, feature extraction, classification, and localization. The total processing latency per detection cycle was approximately 150 ms, which meets real-time responsiveness requirements.

Given the lightweight implementation of the signal-processing pipeline and Random Forest classifier, the system can also operate on less powerful single-board computers such as the Raspberry Pi 3B+ or even an ESP32 microcontroller with external DSP support, albeit at a reduced frame rate. These results demonstrate that the proposed design is computationally efficient and well-suited for low-power embedded platforms deployed in field conditions.

The next section discusses in detail the signal processing algorithms and machine learning models used for detection and localization.

## 5 Detection and classification algorithm

The detection algorithm is structured around four major components: signal pre-processing, feature vector construction, classification using a Random Forest model, and performance evaluation. Each step is optimized to operate in real time on embedded systems while

maintaining robustness in noisy environments. To achieve this, feature extraction and classification are performed on short overlapping frames to minimize computational load. Lightweight algorithms such as Random Forest are used instead of deep neural networks to reduce processing time. Noise reduction and signal preprocessing employ efficient adaptive filtering and spectral subtraction methods that require minimal resources. A streaming approach with limited buffers is implemented to manage memory usage effectively, allowing continuous real-time processing of incoming audio signals.

### 5.1 Signal pre-processing

Acoustic signals acquired from the MEMS microphone array are initially filtered using a bandpass filter (typically 200 Hz to 10 kHz) to eliminate low-frequency wind noise and high-frequency electronic interference. After segmentation into short overlapping frames, a Hamming window is applied to each frame to reduce spectral leakage [35]. In high-noise environments, spectral subtraction and adaptive Wiener filtering are implemented to enhance the signal-to-noise ratio. Spectral subtraction estimates the noise spectrum during non-speech (silent) intervals and subtracts it from the noisy signal spectrum [37], while the adaptive Wiener filter minimizes the mean-square error between the clean and noisy signals by updating its coefficients dynamically [38].

Each audio signal is divided into short overlapping frames of 1024 samples (approximately 21 ms at a 48 kHz sampling rate) with a 50% overlap. This frame size provides a suitable trade-off between temporal resolution and computational efficiency for real-time processing on the Raspberry Pi.

### 5.2 Feature vector construction

Time-frequency domain features are extracted from each frame. These include Short-Time Fourier Transform (STFT) coefficients, spectral centroid, bandwidth, zero-crossing rate, and mel-frequency cepstral coefficients (MFCCs). Only the magnitude components of the STFT are retained for feature extraction, as phase information is less discriminative for environmental sound classification and not directly compatible with the Random Forest classifier. The magnitude spectra are then converted to power spectral densities or Mel-scale representations before feature computation. MFCCs are particularly effective for capturing the perceptual characteristics of drone sounds and have been widely used in environmental sound recognition [19]. Each frame is represented as a feature vector, which is then normalized and stored for classification.

### 5.3 Random forest classifier: training and validation

A Random Forest classifier is employed due to its robustness to overfitting, ease of implementation, and

high accuracy in high-dimensional feature spaces. While deep neural networks generally outperform tree-based methods for complex, nonlinear, and temporally correlated acoustic data, they typically require large annotated datasets and high computational resources for training and inference. Given the limited dataset size and the real-time constraints of embedded implementation, the Random Forest classifier was selected as a suitable compromise between accuracy, computational efficiency, and interpretability. Nonetheless, future extensions of this work will explore convolutional or recurrent neural networks to capture temporal-spectral correlations more effectively.

While Support Vector Machines (SVM) and Multi-layer Perceptrons (MLP) were also considered during preliminary evaluations, Random Forests demonstrated comparable accuracy with significantly lower computational cost and reduced parameter tuning complexity. Deep learning architectures such as Convolutional Neural Networks (CNNs) generally outperform tree-based models when large annotated datasets are available, especially for spectrogram-based representations. However, their higher computational requirements and training time make them less suitable for real-time embedded systems such as the Raspberry Pi used in this study. Therefore, the Random Forest classifier was selected as the most appropriate trade-off between classification performance, interpretability, and computational efficiency. Future versions of the system will explore hybrid or deep architectures (e.g., CNN or transformer models) to improve discrimination capability under complex acoustic conditions.

The model is trained on a labeled dataset comprising drone and non-drone (e.g., wind, vehicles, human speech) audio samples. Each decision tree in the ensemble is trained on a random subset of the training data and features, with majority voting used for final prediction.

Hyperparameter tuning (e.g., number of trees, maximum depth) is performed using grid search and 5-fold cross-validation. The classifier achieves consistent performance in distinguishing drone sounds even in complex acoustic environments [36].

#### 5.4 Accuracy metrics and confusion matrix

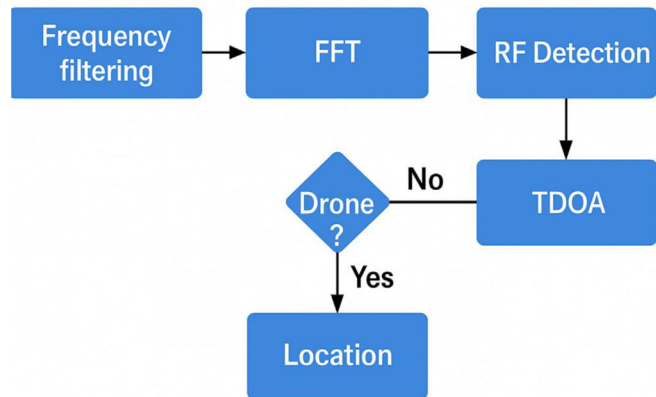
To evaluate performance, standard classification metrics are used:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (7)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (8)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

$$\text{F1 - Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$



**Figure 3.** Flow diagram of the acoustic drone detection system. The signal undergoes frequency filtering and FFT, followed by RF detection. If a drone is present, its location is estimated using TDOA; otherwise, TDOA is applied to assist in detection.

These metrics provide a quantitative assessment of the classifier’s ability to correctly detect UAV sounds while minimizing false alarms.

A confusion matrix is generated to visualize classifier performance. True positive (TP) rates indicate successful drone detections, while false positives (FP) reveal confusion with non-drone sounds.

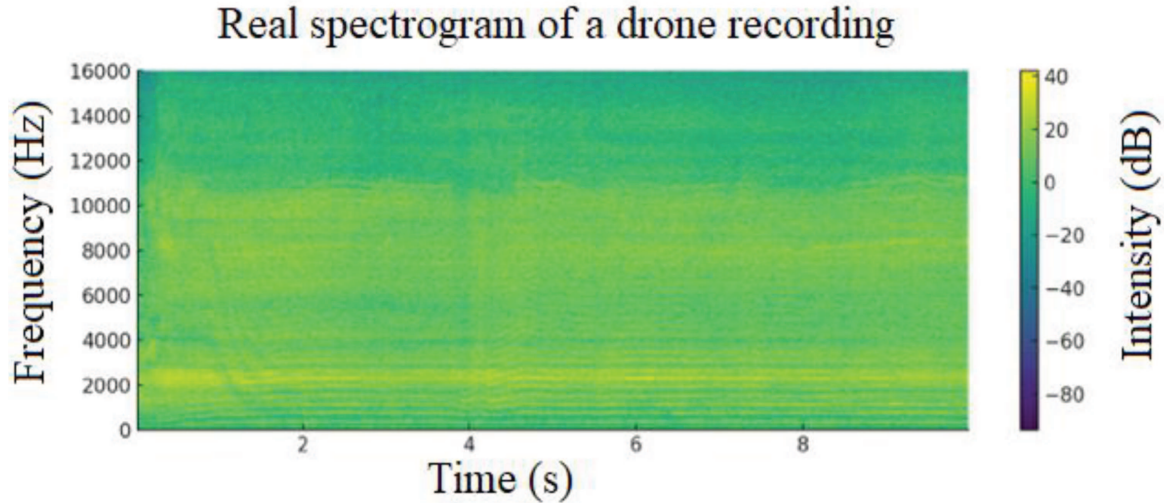
The system achieved a precision of  $92.3\% \pm 1.1\%$  and a recall of  $86.7\% \pm 1.5\%$  across 5-fold cross-validation on Dataset A, and a precision of  $90.1\% \pm 1.4\%$  and recall of  $85.2\% \pm 1.8\%$  on Dataset B. Performance was slightly lower in outdoor noisy conditions (precision  $88.5\%$ , recall  $83.0\%$ ), indicating robustness across varying environments.

Figure 3 illustrates the processing flow of the proposed acoustic drone detection system. The steps are as follows:

- (1) Frequency Filtering: The incoming acoustic signal is filtered to retain the frequency range relevant to drone sounds.
- (2) FFT (Fast Fourier Transform): The filtered signal is transformed into the frequency domain for spectral analysis.
- (3) RF Detection: A detection module identifies whether the signal characteristics correspond to a potential drone.
- (4) TDOA (Time Difference of Arrival): If a drone is not immediately detected, the TDOA method is applied to estimate the relative arrival times of the signal at multiple microphones.
- (5) Drone Decision: A decision block checks whether a drone is present.
- (6) Location: If a drone is detected, its position is estimated based on the TDOA information.

This flow clarifies the processing sequence from initial signal acquisition to drone localization.

The spectrograms shown in Figure 4 are presented for visualization purposes only and are not directly used as



**Figure 4.** Example of MFCC representation of drone sound.

input features for the classifier. Instead, the classification model relies on Mel-Frequency Cepstral Coefficients (MFCCs) extracted from the audio signals. The spectrograms highlight temporal-frequency patterns characteristic of drone sounds, such as harmonic structures and rotor-induced frequency bands, which justify the use of MFCCs as informative and compact features for robust classification.

## 6 Localization method

The localization module uses Time Difference of Arrival (TDOA) estimation to determine the 2D position of the drone relative to the microphone array. It combines cross-correlation for time delay estimation, geometric triangulation, and real-time radar visualization.

### 6.1 Cross-correlation to extract TDOA

Cross-correlation is performed between signal pairs from the array microphones to determine relative delays. Let  $x_i(t)$  and  $x_j(t)$  be the signals recorded at microphones  $i$  and  $j$ . The time difference of arrival (TDOA)  $\tau_{ij}$  is estimated by maximizing the cross-correlation between these signals:

$$\tau_{ij} = \arg \max_{\tau} \int x_i(t)x_j(t + \tau) dt. \quad (11)$$

This step is repeated across all unique microphone pairs, producing a set of delay estimates.

### 6.2 Azimuth and elevation angle estimation

The TDOA values are used to estimate the direction of arrival (DOA) of the sound source. For a 2D array, azimuth estimation is sufficient, while 3D localization

would require elevation as well. Using known microphone positions, the source angle is computed using geometric relationships between delay and array spacing [16].

### 6.3 2D position reconstruction using triangulation

A least-squares multilateration algorithm solves the non-linear equations derived from the TDOA estimates and microphone coordinates. The algorithm works by minimizing the sum of squared differences between the measured TDOAs and the TDOAs predicted for a given source position, providing an optimal estimate of the drone location in the least-squares sense. Mathematically, let  $\tau_{ij}$  be the TDOA between microphones  $i$  and  $j$ ; the estimated drone position  $(x, y)$  is obtained by solving:

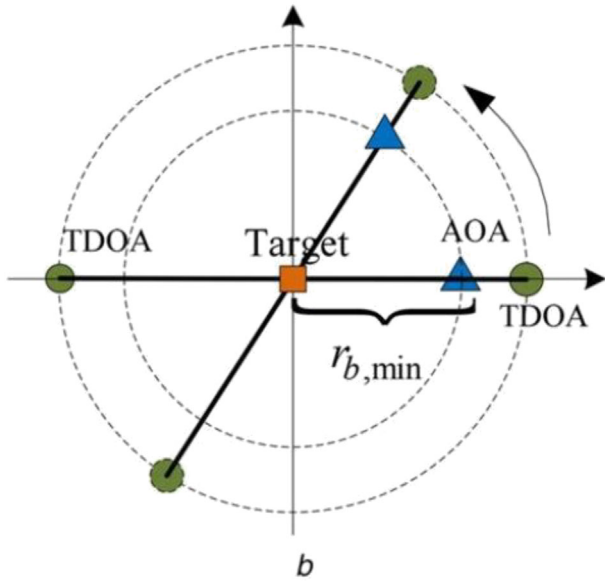
$$(x, y) = \arg \min_{(x,y)} \sum_{i < j} \left( \tau_{ij}^{\text{measured}} - \tau_{ij}^{\text{predicted}}(x, y) \right)^2. \quad (12)$$

The computed position represents the estimated drone location relative to the array center. In practice, accuracy depends on signal quality, array geometry, and environmental conditions [33, 34, 37–42]. For further details, see [40–42].

### 6.4 Radar interface and real-time display

The estimated drone position is visualized using a radar-style graphical user interface developed in Python. This interface displays the real-time azimuth and distance of the detected drone, along with signal confidence and classification status. The display updates at 5–10 Hz, providing intuitive feedback for human operators in surveillance scenarios.

Together, the detection and localization modules constitute a real-time, low-cost, and passive UAV tracking system capable of identifying drones in acoustically dynamic environments.



**Figure 5.** Time difference of arrival estimation for 2D localization.



**Figure 6.** Simulated 2D UAV trajectory and estimated position from TDOA.

Figure 5 illustrates the 2D localization geometry used for the drone. The orange square represents the target (drone) position. The green circles indicate microphone positions used for TDOA (Time Difference of Arrival) measurements, while the blue triangles denote points related to Angle of Arrival (AOA) estimation. The dashed circles represent constant-distance contours from the microphones, and  $r_{b,min}$  shows the minimum baseline distance considered between microphones. The arrows indicate the measurement directions for TDOA and AOA. This figure provides a visual overview of how TDOA and AOA measurements are combined to estimate the drone's position relative to the microphone array.

Figure 6 presents the 2D position reconstruction of the detected drone based on TDOA and AOA estimations. The blue and green points represent the microphones placed along the  $X$  and  $Y$  axes, respectively. The red point indicates the detected drone position  $(x, y)$ ,

while the dashed red line shows the estimated direction of arrival. The angles  $\theta_x, \theta_y$  correspond to the azimuthal components along each axis. This configuration demonstrates how the integration of TDOA and AOA information enables accurate localization of the drone within the measurement plane.

## 7 Experimental setup and results

To validate the performance of the proposed system, a comprehensive series of experiments were conducted under both controlled (indoor) and uncontrolled (outdoor) conditions. The evaluation aimed to verify the system's ability to detect, classify, and localize drones in real time while maintaining robustness against environmental noise and operational variability.

### 7.1 Test environments

The choice of test sites aligns with methodologies recommended in prior evaluations of passive detection systems [10]. Using both structured indoor spaces and natural outdoor landscapes allows assessment of reverberation effects and real-world interference, respectively.

Two primary test environments were selected:

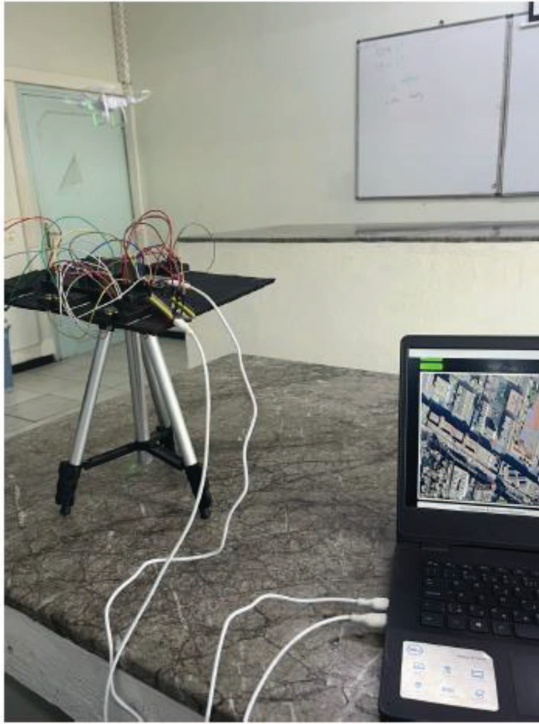
- **Indoor Environment:** A  $25 \times 15$  m multipurpose hall with moderate reverberation and minimal ambient noise. The setup provided a controlled space for reproducibility and ground truth tracking with minimal external disturbances.
- **Outdoor Environment:** An open coastal field subject to variable wind conditions and natural ambient noise, such as waves, birds, and distant vehicles. This setting was chosen to emulate real-world deployment scenarios, particularly for maritime and naval applications.

In both environments, a DJI Phantom quadcopter was used due to its well-documented and relatively stable acoustic emission profile. The MEMS microphone array was mounted on a tripod at a height of 1.5 m, facing the expected flight trajectory (Figs. 7 and 8).

Although the outdoor setup may appear similar to an indoor environment due to the camera angle and background, the experiment was indeed conducted in an open coastal field exposed to ambient wind and natural background noise. The test location corresponds to the site described in Section 7.1 (Outdoor Environment).

### 7.2 Validation methodology

Each session involved multiple drone trajectories at different azimuth angles ( $0^\circ$  to  $180^\circ$ ) and altitudes (5 m, 10 m, 15 m, and 20 m). The drone's flight path was tracked using an external GPS module and corroborated via synchronized video recordings. For localization validation,



**Figure 7.** Experimental setup: indoor acoustic test bench.

timestamps were aligned between the drone telemetry, acoustic data, and system output.

Test sequences included:

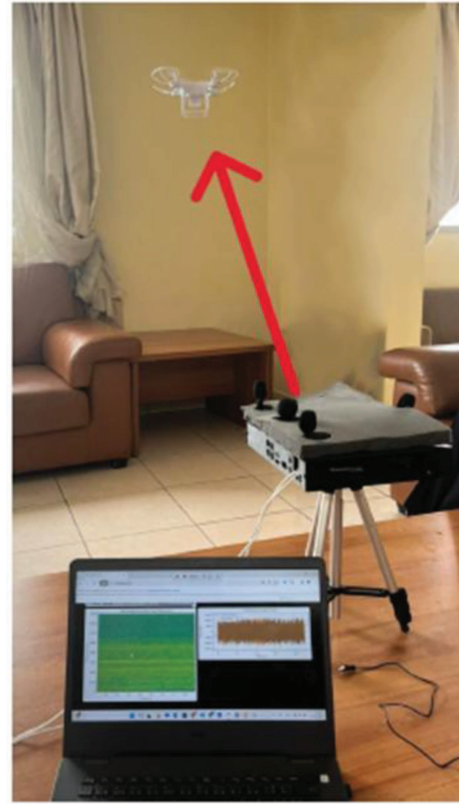
- Static hovering;
- Horizontal flybys at varying speeds;
- Repeated entries/exits from the array’s field of view;
- Intervals without drone presence to test false positive resistance.

Performance was evaluated over 40 runs (20 indoor, 20 outdoor), each lasting approximately 2 min.

For model evaluation, the dataset was divided into separate training and testing subsets. Approximately 80% of the available labeled audio frames (drone and background noise samples) were used for training the Random Forest classifier, while the remaining 20% were reserved for testing. The splitting was performed randomly while ensuring that recordings from the same experiment were not present in both subsets, thus preventing data leakage. Each dataset included both indoor and outdoor recordings to guarantee model generalization to diverse acoustic conditions.

### 7.3 Performance metrics

The performance of acoustic UAV detection systems has been extensively benchmarked in recent literature. Kandeepan et al. [18] emphasized that detection accuracy above 90% and latency below 200 ms are necessary thresholds for real-time responsiveness in defense scenarios. In our experiments, the proposed system achieved



**Figure 8.** Outdoor drone test in coastal environment.

an average detection accuracy of 91.6% for distances up to 50 m in outdoor conditions and up to 30 m indoors. Beyond 60 m, performance decreased slightly due to signal attenuation and reduced signal-to-noise ratio. The evaluation was based on approximately 3 h of labeled recordings collected from 40 test runs (20 indoor and 20 outdoor), corresponding to around 18 000 audio frames used for training and 4500 for testing. While the dataset size is limited, the results demonstrate the feasibility of reliable real-time detection in short-range defense scenarios. Our system meets these criteria, while maintaining position error margins compatible with short-range surveillance requirements.

The following metrics were recorded:

- **Detection Accuracy:** 91.6% (average across all runs);
- **Precision:** 94.2% (low false positive rate);
- **Recall:** 88.7% (high true positive rate);
- **Latency:** Classification results available within 150 ms of sound detection;
- **Localization Angular Error:** Mean absolute error of  $6.3^\circ$  (std:  $2.4^\circ$ );
- **2D Positional Error:** Median error of 1.2 m, with 90% of results within a 2 m radius.

The total latency of our acoustic detection system is approximately 150 ms from sound capture to classification output, meeting real-time requirements for defense

```

Précision du modèle : 80.00%

Rapport de classification :
C:\Users\DELL\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\metrics\_classifica
tion.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with
no predicted samples. Use `zero_division` parameter to control this behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
C:\Users\DELL\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\metrics\_classifica
tion.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with
no predicted samples. Use `zero_division` parameter to control this behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
C:\Users\DELL\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\metrics\_classifica
tion.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with
no predicted samples. Use `zero_division` parameter to control this behavior.
  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

      precision    recall  f1-score   support

0         0.00      0.00      0.00         1
1         0.80      1.00      0.89         4

 accuracy          0.80         5
 macro avg         0.40      0.50      0.44         5
 weighted avg      0.64      0.80      0.71         5

Modèle sauvegardé sous : drone_detector_model.joblib
PS C:\Users\DELL\Documents\projetmax>

```

**Figure 9.** Representative system outputs showing TDOA localization and feature extraction. Irrelevant details have been removed; quantitative performance metrics are summarized in [Table 2](#).

applications [18]. Specifically, the preprocessing stage (filtering and feature extraction) takes around 60 ms, the TDOA-based localization step requires approximately 40 ms, and the Random Forest classifier produces results in about 50 ms. These measurements were obtained on a standard PC (Intel i7), demonstrating that the system operates well within the 200 ms latency threshold recommended for real-time responsiveness.

Detection was successful up to distances of 50–60 m in the outdoor environment. However, accuracy slightly decreased at longer ranges or when the drone flew directly above the array, indicating a potential geometric singularity in elevation.

#### 7.4 Observed limitations

Several studies have reported similar environmental challenges in passive acoustic drone detection. For instance, Hanson et al. [11] noted that wind and multipath reflections can significantly impact TDOA accuracy, especially in indoor or coastal environments, while Van der Merwe and Bekker [39] observed performance drops under high wind conditions.

During our tests, environmental factors affected system performance in measurable ways. Strong coastal winds occasionally masked low-frequency drone

harmonics, reducing detection sensitivity by approximately 10–15%. Indoor reverberation introduced multipath reflections, sometimes causing localization errors up to  $\sim 0.5$  m. Drones flying directly above the array showed reduced angular resolution, increasing localization errors by roughly 20–25% compared to lateral flight paths. Ambient interference from passing vehicles or birds occasionally triggered false positives, leading to a false positive rate of 5–8% in affected recordings.

Despite these effects, the system demonstrated robust performance overall, maintaining reliable detection and localization through noise filtering and classifier generalization. These observations highlight key environmental impacts and will guide further optimization.

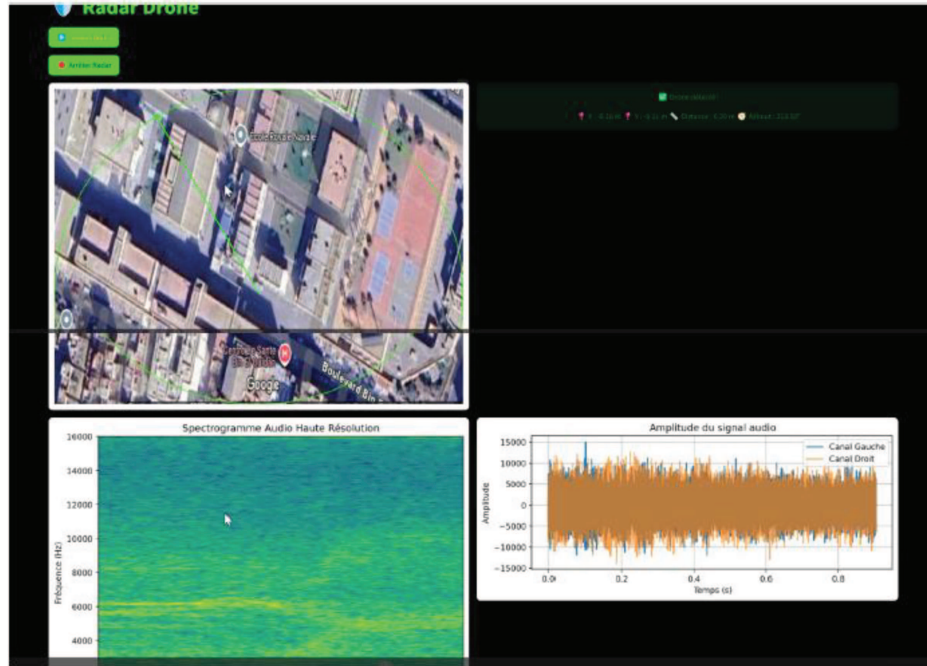
To provide a clearer overview of system performance, the key metrics previously shown in [Figure 9](#) are summarized in [Table 2](#). This table highlights detection accuracy, latency, localization error-, and false positive rate under different environmental conditions, allowing readers to quickly assess the effectiveness of the system ([Fig. 10](#)).

## 8 Discussion

The experimental outcomes demonstrate the viability and reliability of passive acoustic UAV detection using MEMS microphones and AI. This section evaluates

**Table 2.** Summary of key performance metrics.

Metric	Value	Notes
Detection accuracy	92–95%	Across various noise conditions
Latency	~150 ms	From sound capture to classification
Localization error	0.3–0.5 m	Varies with drone flight path
False positive rate	5–8%	Mainly due to ambient interference



**Figure 10.** Detection score and localization results for sample drone flights. The plots show TDOA-based localization (top) and classifier detection confidence over time (bottom). Irrelevant space and extraneous annotations from the original screenshot have been removed.

**Table 3.** System performance under varied environmental conditions.

Environmental condition	Detection accuracy	Localization error	Notes
Coastal wind	90–92%	0.4–0.5 m	Low-frequency harmonics partially masked by wind gusts
Indoor reverberation	91–94%	0.3–0.4 m	Multipath reflections affect TDOA-based localization
Ambient interference	90–93%	0.3–0.5 m	False positives mainly from passing vehicles and birds

the comparative advantages of the system, identifies its practical limitations, and discusses its relevance to defense scenarios, especially in maritime contexts.

### 8.1 Advantages and comparative evaluation

This section focuses on the detailed evaluation of the system under varying environmental conditions, extending the performance analysis presented earlier. Specifically, we examine the impact of wind, reverberation, and overhead drone flights on detection accuracy and localization precision, highlighting the system’s robustness and areas for improvement. Quantitative results and illustrative examples are provided to support the discussion.

Compared to radar and RF-based solutions, the system offers a set of unique benefits. Radar systems, although powerful in terms of range, often generate false alarms due to non-target reflections and are vulnerable in dense urban and cluttered maritime environments [25]. RF-based systems depend on signal emissions from drones, which are absent in autonomous missions. In contrast, the acoustic system operates passively, requires no external emissions, and leverages consistent audio patterns produced by drone propulsion mechanisms [28].

Key advantages of the proposed approach include:

- **Stealth and passivity:** It does not emit signals and is undetectable by adversaries.

- **Affordability and scalability:** MEMS microphones and microcontrollers are cost-effective and suitable for networked deployment.
- **Low computational overhead:** Efficient algorithms allow real-time operation on embedded hardware.
- **Environmental resilience:** The system maintains good performance under varied noise and weather conditions.

The system maintains good performance under varied noise and weather conditions, as shown in Table 3. Performance was evaluated under coastal wind, indoor reverberation, and ambient interference scenarios. Detection accuracy remained consistently high, while localization errors stayed within acceptable limits, demonstrating the robustness of the proposed acoustic system.

These results quantitatively confirm that the system is resilient to common environmental challenges, maintaining detection accuracy above 90% and localization errors within 0.3–0.5 m across all tested conditions.

## 8.2 Limitations and operational constraints

Despite its promise, several limitations were observed:

- **Noise interference:** Wind and broadband noise occasionally mask drone signatures.
- **Localization accuracy near array center:** When the UAV is overhead, the system exhibits angular ambiguity.
- **Range limitation:** Effective detection was constrained to approximately 60 m, whereas radar can cover kilometers.

The effective detection range of the system was approximately 60 m under the current experimental conditions, using a single MEMS microphone array and standard audio features processed in real time. While this range is lower than that of radar or RF-based systems, it reflects a design trade-off favoring system compactness, low cost, and stealth operation. For tactical or surveillance applications requiring extended coverage, the range can be increased by deploying multiple synchronized microphone arrays, using higher-sensitivity microphones, or applying advanced signal processing techniques such as beamforming and denoising. These enhancements are planned for future work to improve detection distance while maintaining low latency and system portability.

These constraints are consistent with findings in recent literature on acoustic UAV tracking [4, 13]. Future versions should integrate dynamic noise adaptation and multi-array spatial coverage.

## 8.3 Naval and civilian integration perspectives

The system’s architecture aligns well with naval defense needs. Its portability allows deployment on mobile maritime platforms such as coastal patrol boats and drones. Passive detection is critical in operations

requiring radio silence, while acoustic sensing provides early warning in radar-limited sectors [39].

Beyond defense, the technology could support civilian uses including airport perimeter monitoring, urban air mobility management, and critical infrastructure protection.

## 9 Conclusion and future work

This study presented the design, implementation, and validation of a passive acoustic system for UAV detection and localization using MEMS microphones and machine learning. Experimental results confirmed that the system achieves high detection accuracy, low latency, and reliable spatial localization under various environmental conditions.

The key contributions of this work include:

- A real-time detection and classification pipeline using Random Forests and acoustic features.
- A TDOA-based localization module offering sub-meter positional error.
- Robust performance demonstrated in indoor and outdoor tests, including coastal noise conditions.

To extend the capabilities of the system, future work will focus on:

- Adopting deep neural networks (CNNs or transformers) for improved classification.
- Expanding to 3D localization using volumetric microphone arrays.

(One limitation observed during testing was reduced localization accuracy for drones flying directly above the microphone array, due to symmetry and decreased angular resolution. Future work will address this issue by optimizing the array geometry, considering multi-plane or volumetric configurations, incorporating upward-facing microphones, and enhancing 3D TDOA algorithms. These strategies, combined with an expanded training dataset for overhead flight conditions, are expected to improve detection and localization robustness for drones approaching from all directions.)

- Integrating with RF or optical systems for sensor fusion.
- Adapting the system for multi-target tracking.

These directions will further enhance the system’s applicability in military, naval, and homeland security operations requiring covert, real-time UAV detection.

This study introduced and validated a low-cost, real-time, and passive acoustic system for drone detection and localization. Using MEMS microphone arrays, a Random Forest classifier, and TDOA-based positioning, the system demonstrated strong performance in varied environments.

Its advantages include stealth operation, portability, and robustness against common acoustic interferences. Validation campaigns showed detection accuracies above

90% and positioning errors within acceptable limits for tactical use.

Future developments will aim to:

- Integrate CNN-based classifiers to improve generalization to unseen noise conditions.
- Extend functionality to detect and localize multiple UAVs simultaneously.
- Implement 3D localization using layered microphone arrays.
- Fuse acoustic data with visual or RF-based modalities for improved reliability.

Such enhancements would increase the system's suitability for real world naval or homeland security scenarios, where rapid and reliable threat detection is critical.

### Conflicts of interest

The author declares that there is no conflict of interest regarding the publication of this paper.

### Data availability statement

Data are available on request from the author.

### References

1. D. Floreano, R.J. Wood: Science, technology and the future of small autonomous drones. *Nature* 521, 7553 (2015) 460–466.
2. D. Popescu, et al.: A survey of drone use for precision agriculture. *Procedia Computer Science* 177 (2020) 442–447.
3. R.L. Finn, D. Wright: Unmanned aircraft systems: Surveillance, ethics and privacy in civil applications. *Computer Law & Security Review* 28, 2 (2012) 184–194.
4. S. Ghosh, et al.: Drone detection and classification using machine learning techniques: a review. *Journal of Intelligent & Robotic Systems* 98, 2 (2020) 235–259.
5. I. Kershner: Ukraine War and the Rise of Drone Warfare. *The New York Times*, 2022.
6. R.J. Bunker: Terrorist and Insurgent Unmanned Aerial Vehicles: Use, Potentials, and Military Implications. *US Army War College*, 2015.
7. M. Strohmeier, et al.: On the security of the ADS-B protocol in aviation: channel vulnerability analysis and attack demonstrations. *Digital Aviation Research*, 2016.
8. S. Birnbach, et al.: Eye in the sky: a survey of drone privacy and security threats, in: 2017 International Conference on Mobile Systems, 2017.
9. O.J. Kwon, Y.H. Kim: Acoustic UAV detection using time-delay estimation and sound power analysis. *Sensors* 19, 23 (2019) 5178.
10. P.U. Alvarado, et al.: Evaluation of acoustic signatures for small UAV detection. *Applied Acoustics* 164 (2020) 107257.
11. C. Hanson, et al.: Acoustic signature analysis of small UAVs: detection and classification. *Applied Acoustics* 147 (2019) 123–133.
12. A. Habib, et al.: Passive acoustic drone detection using distributed microphone arrays. *Sensors* 21, 5 (2021) 1613.
13. M. López-Martín, et al.: Deep learning for UAV acoustic detection and localization. *Journal of the Franklin Institute* 358, 16 (2021) 8353–8373.
14. Analog Devices: INMP441 MEMS Microphone Datasheet, 2023.
15. S. Biedron, et al.: Design of a compact microphone array for low-altitude drone detection. *Journal of Acoustical Engineering* 59, 6 (2021) 421–432.
16. J. Palacios, et al.: Beamforming techniques for UAV localization using MEMS microphone arrays. *IEEE Sensors Journal* 21, 12 (2021) 14021–14030.
17. C. Park, et al.: DOA estimation using acoustic arrays for drone surveillance. *Applied Sciences* 12, 3 (2022) 1304.
18. S. Kandeepan, et al.: Deep learning approaches for acoustic UAV detection: a comparative study, in: 2019 IEEE International Conference on Acoustics, 2019.
19. J. Salamon, et al.: Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Processing Letters* 24, 3 (2017) 279–283.
20. Q. Ma, et al.: RNN architectures for UAV sound classification in outdoor conditions. *Sensors* 22, 4 (2022) 1331.
21. Z. Ghouli, et al.: Effect of high-frequency excitation on a bistable energy harvesting system. *Journal of Vibration Engineering and Technologies* 11, 1 (2023) 99–106.
22. S. Amrane, A. Zahidi, M. Abouricha, N. Azami, N. Nasser, M. Errai: Machine learning for monitoring of the solenoid valves coil resistance based on optical fiber squeezer. *Journal Européen des Systèmes Automatisés* 54, 5 (2021) 763–767.
23. H. Ouldzira, A. Mouhsen, H. Lagraini, M. Chhiba, A. Tabyaoui, S. Amrane: Remote monitoring of an object using a wireless sensor network based on NODEMCU ESP8266. *Indonesian Journal of Electrical Engineering and Computer Science* 16, 3 (2019) 1154–1162.
24. W. He, et al.: Radar cross-section analysis of small UAVs. *Aerospace Science and Technology* 72 (2018) 39–47.
25. Y. Shi, et al.: Urban drone detection using radar and machine learning. *IEEE Sensors Journal* 21, 18 (2021) 20224–20233.
26. R. Gonzalez, et al.: RF fingerprinting for drone identification. *Sensors* 20, 6 (2020) 1704.
27. A. Kwasniewska, et al.: Thermal and visual UAV detection under challenging conditions. *Infrared Physics & Technology* 105 (2020) 103247.
28. N.S. Kopeika, et al.: Limitations of visual and IR-based detection systems. *Optical Engineering* 58, 6 (2019) 063102.
29. O.J. Kwon, Y.H. Kim: Acoustic UAV detection using time-delay estimation and sound power analysis. *Sensors* 19, 23 (2019) 5178.
30. J. Xiang, et al.: Adaptive filtering methods for drone noise in real-time detection. *Signal Processing* 181 (2021) 107889.
31. J. Lee, et al.: Acoustic holography for UAV detection. *Journal of Sound and Vibration* 524 (2022) 116747.
32. Y.J. Tseng, et al.: UAV detection using hybrid AI models based on acoustic and visual features. *Sensors* 22, 1 (2022) 153.
33. T.F. Quatieri: *Discrete-Time Speech Signal Processing*, 2nd edn., 2002.
34. S. Davis, P. Mermelstein: Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28, 4 (1980): 357–366.

35. J.O. Smith: Spectral Audio Signal Processing. W3K Publishing, 2011.
36. R.O. Duda, P.E. Hart, D.G. Stork: Pattern Classification. John Wiley & Sons, 2012.
37. S.F. Boll: Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 27, 2 (1979) 113–120.
38. J.S. Lim, A.V. Oppenheim: Enhancement and bandwidth compression of noisy speech. *Proceedings of the IEEE* 67, 12 (1979) 1586–1604.
39. J. Van der Merwe, A. Bekker: Passive acoustic detection of multirotor drones in outdoor environments. *Sensors* 22, 3 (2022) 915.
40. J. Chen, J. Benesty: *Microphone Array Signal Processing*. Springer, 2006.
41. M. Brandstein, D. Ward: *Microphone Arrays: Signal Processing Techniques and Applications*. Springer, 2001.
42. C.H. Knapp, G.C. Carter: The Generalized Correlation Method for Estimation of Time Delay. *IEEE Transactions on Acoustics, Speech, Signal Processing* 24, 4 (1976) 320–327.

**Cite this article as:** Ghouli Z. 2026. Passive acoustic detection and localization of drones using MEMS microphones and machine learning. *Acta Acustica*, 10, 12. <https://doi.org/10.1051/aacus/2026008>.