AUDibLE: Acoustic-based UAV Detection with Low-Power Embedded Machine Learning

Monika Prakash[§], Uma Maheswara[§], Sumit Sharma[§], Willian T. Lunardi[§], Kavin Thangadorai[§], Markus Schuss[‡], Carlo Alberto Boano[‡], and Michael Baddeley[§]

[§] Technology Innovation Institute, UAE
{monika.prakash, uma.maheswara, sumit.sharma, willian.lunardi, kavin.thangadorai, michael.baddeley}@tii.ae

[‡] Graz University of Technology, Austria
{markus.schuss, cboano}@tugraz.at

Abstract*—Small unmanned aerial vehicles (UAVs) pose increasing security risks to industrial infrastructure. Acoustic sensing offers a practical, low-cost alternative to radar, computer vision, and RF-based detection; however, such systems rely on power-intensive processing that limits deployment in remote locations or at scale. AUDibLE addresses this challenge by performing UAV detection *locally* on a resource-constrained embedded device: specifically, a low-power, low-cost acoustic sensing system based on an off-the-shelf nRF54L15 microcontroller and low-cost micro-electromechanical systems microphones. To support UAV detection on such constrained hardware, we employ feature extraction based on Mel-frequency cepstral coefficients combined with lightweight machine learning classifiers, including support vector machines, XGBoost, and LightGBM. We show that model performance can degrade under embedded deployment because of differences between training and sensing conditions, particularly the microphone frequency response. We therefore introduce a microphone-aware equalization filter that aligns the spectral characteristics of training data with the response of the sensing-node microphone. Retraining with equalized data restores feature distribution alignment and significantly improves detection robustness during on-device inference. Our experimental evaluation demonstrates reliable UAV detection across recordings from 32 UAV platforms, as well as real-world flight experiments. We further show that tree-based models, specifically LightGBM, significantly reduce inference time, CPU utilization, and Flash memory requirements. These results demonstrate that such models are computationally feasible for real-time inference on resource-constrained embedded devices and can provide a solution for UAV acoustic perimeter sensing in scenarios where low-power, low-cost hardware is required.

***Index Terms*—UAV detection, Acoustic sensing, Embedded systems, Machine learning, Edge computing, Mel-frequency cepstral coefficients, Support vector machines, XGBoost, LightGBM.**

I. Introduction

The proliferation of small unmanned aerial vehicles (UAVs) has created new security threats at sensitive sites and critical infrastructure. Unauthorized airborne intrusions are particularly dangerous for large-scale installations such as solar farms, airports, industrial facilities, and isolated energy assets. Since UAVs may operate at low altitudes and exhibit high maneuverability, early detection remains a significant challenge.

Existing UAV detection approaches mainly rely on RF monitoring, radar, and vision-based sensing. RF-based systems detect communication links between drones and controllers; however, such methods may be ineffective against autonomous or low-emission platforms. Radar systems can offer long-range coverage, but often involve high cost, increased power consumption, and reduced sensitivity to small drones with weak radar reflections. Vision-based systems provide high-resolution identification, but depend on favorable lighting and weather conditions; moreover, they require (relatively) computationally intensive processing that makes them unsuitable for battery-powered systems.

Acoustic sensing represents an alternative approach. Rotor-driven UAVs inherently generate characteristic acoustic patterns arising from blade rotation, motor harmonics, and aerodynamic effects. Prior studies have demonstrated that spectral features such as Mel-frequency cepstral coefficients (MFCCs), combined with classical machine learning or deep neural networks, can achieve strong drone detection performance under controlled or augmented conditions [1]–[4]. Recent works have also introduced larger multiclass UAV audio datasets and explored fusion with vision-based systems [5]–[8].

Real-world deployment challenges. Despite these advances, most acoustic UAV detection approaches are developed and evaluated under *offline* processing settings [3], [4], which limits their applicability in practical large-scale deployments. In real-world perimeter monitoring scenarios, sensing systems must operate continuously over wide areas using low-power devices with limited communication capabilities, making continuous transmission of raw audio data impractical. This necessitates performing UAV detection directly on resource-constrained embedded edge nodes.

Embedded deployment challenges. However, enabling reliable on-device detection introduces additional challenges. Embedded platforms impose strict constraints on computational resources, memory footprint, and model complexity, which limit the applicability of conventional high-capacity models. In addition, differences in sensing hardware, particularly the frequency response of low-cost embedded microphones, introduce discrepancies between training data and deployment conditions. These factors can lead to shifts in feature distributions, making the direct transfer of trained models to field-deployed edge nodes unreliable.

Our contributions. In this work, we design and implement AUDibLE, a low-power acoustic UAV detection system using resource-constrained embedded edge devices. The proposed pipeline employs MFCCs as primary acoustic features and evaluates multiple lightweight machine learning classifiers, including support vector machines (SVMs), XGBoost, and LightGBM, for real-time on-device inference. To address deployment challenges associated with microphone characteristics and domain differences between training data and embedded hardware, we introduce a microphone-aware equalization strategy that improves detection robustness during field operation. The resulting system enables scalable acoustic perimeter sensing suitable for large-area infrastructure protection, offering a cost-effective and energy-efficient complementary sensing layer to existing detection technologies.

The remainder of this paper is organized as follows. Section II reviews related work. Section III presents the system architecture. Section IV describes the datasets used and discusses deployment challenges. Section V outlines the proposed methodology, while Section VI details the system design. Section VII presents the experimental results. Finally, Section VIII concludes the paper.

II. Related Work

Drone detection has been extensively researched across multiple sensing modalities. RF-based methods analyze communication signals emitted by UAV control links to detect and identify drones [9], [10]. Radar-based systems exploit micro-Doppler signatures to distinguish UAVs from birds and other airborne objects [11], [12]. Vision-based techniques use optical sensors combined with convolutional neural networks for UAV detection and tracking [13], [14]. Despite their effectiveness, these approaches can require specialized sensing hardware or computationally intensive processing and may be sensitive to environmental conditions. Acoustic sensing therefore offers a complementary and potentially low-power alternative.

Acoustic UAV detection is commonly formulated as a classification problem using spectral features, particularly MFCCs [15]. Jeon et al. [2] demonstrated that MFCC-based classification can effectively distinguish drone sounds from urban environmental noise using Gaussian Mixture Models (GMMs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs). Similarly, Shi et al. [1] investigated Hidden Markov Models (HMM) with varying MFCC dimensionalities under noisy conditions. Ibrahim et al. [16] used MFCC features with an SVM to infer payload variations from drone acoustic emissions, while Dumitrescu et al. [17] explored microphone-array platforms for combined detection and localization. Deep learning approaches have also been explored. Casabianca et al. [3] evaluated CNN, RNN, and CRNN architectures with late fusion, while Al-Emadi et al. [4] used generative adversarial networks for dataset augmentation. Multimodal approaches combine RF and vision [7], as well as audio and vision [8].

Open-source UAV audio datasets, including the AuDroK database [5] and multiclass UAV sound collections [6] support benchmarking across multiple drone types and environments but are primarily used for *offline* evaluation rather than embedded deployment.

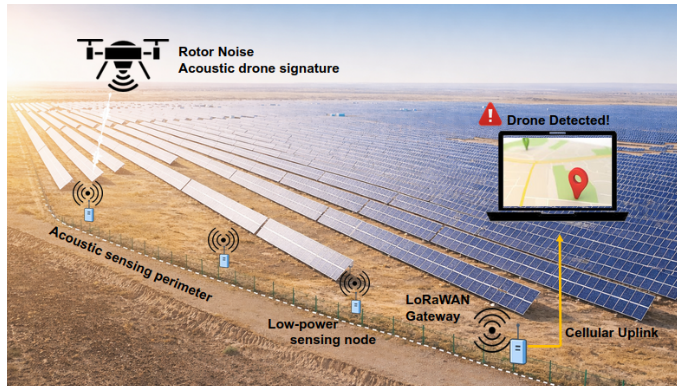


Fig. 1. Use case example of AUDibLE for perimeter monitoring using low-power acoustic sensing nodes.

Despite significant progress in acoustic drone classification, most existing work focuses on *offline* evaluation. While some studies have considered hardware-based implementations [17], [18], comparatively little attention has been given to real-time deployment on low-power embedded platforms, where resource constraints and microphone characteristics can affect detection performance. In particular, transferring models trained on public datasets to low-cost embedded sensing hardware remains insufficiently explored. This work addresses this gap by evaluating MFCC-based SVM, XGBoost, and LightGBM classifiers on resource-constrained sensing nodes and introducing microphone-aware equalization to improve robustness when transferring models trained on open-source datasets to embedded hardware.

III. System Architecture

This section describes the system architecture of `AUDibLE`, a low-power acoustic UAV detection system for perimeter monitoring of large-scale critical infrastructure. It targets scenarios such as solar farms, industrial facilities, and remote energy sites, where continuous wide-area surveillance is required under strict energy and communication constraints. As illustrated in Fig. 1, acoustic sensing nodes can be placed around the monitored area to provide perimeter coverage.

`AUDibLE` follows an edge-based sensing approach in which acoustic analysis and UAV detection are performed locally on each sensing node. Instead of transmitting raw audio, each node generates UAV detection alerts only when activity is detected. This event-driven communication model reduces bandwidth usage and supports energy-efficient operation. Detection alerts are transmitted over LoRaWAN to a gateway and then forwarded to a monitoring station via a cellular uplink for remote alerting and system monitoring. This design avoids continuous audio streaming and enables scalable, energy-efficient deployment in bandwidth-constrained environments.

The sensing node is built around an `nRF54L15` microcontroller with a 128 MHz ARM Cortex-M33 processor [19], and is optimized for low-power operation. It is powered by a 3.7 V 3400 mAh Li-ion battery. The node is equipped with

TABLE I
KEY HARDWARE PARAMETERS OF THE SENSING NODE

Parameter	Value
Microcontroller	nRF54L15 (ARM Cortex-M33)
Microphone	Digital PDM microphone
Signal-to-noise ratio (SNR)	63 dB
Max. sound pressure level (SPL)	120 dB
Sampling rate	8 kHz
Active current	10 mA
Sleep current	3 μA
Battery capacity	3400 mAh (Li-ion)

TABLE II
DATASETS USED FOR ACOUSTIC UAV DETECTION

ID	Description
DS0	In-house UAV and background recordings collected using the proposed sensing node
DS1	Public UAV recordings with payload variations collected under controlled conditions [21]
DS2	Outdoor UAV and background recordings in real-world environments [22]
DS3	UAV recordings captured at varying distances and noise conditions [23]
DS4	Multi-class recordings of consumer UAV platforms [24]

a digital PDM microphone (MP34DT05-A, STMicroelectronics) [20], providing a signal-to-noise ratio (SNR) of 63 dB and a maximum sound pressure level (SPL) of 120 dB for acoustic sensing. Table I summarizes the key hardware parameters that influence sensing performance, energy consumption, and the design of the acoustic processing pipeline. The acoustic detection pipeline is described in Section V.

IV. DATASETS AND DEPLOYMENT CHALLENGES

This section describes the datasets used for training and evaluation and the cross-domain mismatch between training data and sensing-node recordings.

A. Datasets

To train and evaluate the proposed acoustic UAV detection system, we use a combination of in-house and publicly available UAV audio datasets, summarized in Table II. The datasets contain UAV and background recordings across indoor and outdoor environments, varying distances, multiple UAV platforms, and environmental conditions including wind and human activity.

B. Cross-Domain Mismatch and Deployment Impact

UAV audio datasets are captured using different microphones, environments, and acquisition setups, resulting in variations in their spectral characteristics. In particular, publicly available datasets are typically recorded using high-quality microphones, whereas the proposed sensing node uses a low-cost micro-electromechanical system (MEMS) microphone. This introduces frequency-dependent coloration, altering the spectral representation of the audio signal. Such variations lead to inconsistencies in Mel-frequency cepstral coefficient (MFCC) feature distributions between training datasets and sensing-node data.

To quantify the impact of this variability, we compare the distribution of MFCC features extracted from training datasets with those obtained from the sensing node. Fig. 2 shows the distribution of selected MFCC feature dimensions across both domains. A clear shift in feature distributions is observed, indicating that sensing-node data deviates significantly from the training data. This mismatch alters the feature space and can destabilize classifier decision boundaries, leading to degraded performance when models trained on external datasets are deployed on embedded hardware. This domain mismatch motivates the microphone-aware equalization approach introduced in Section VI-B.

V. METHODOLOGY

Fig. 3 presents the acoustic UAV detection pipeline of AUDibLE, consisting of two stages: offline model preparation and on-device inference.

Step A: Offline model preparation. In the offline stage, UAV and background recordings from the datasets in Section IV are resampled, segmented, and microphone-equalized to reduce the spectral mismatch between training data and sensing-node recordings. MFCC features are then extracted to train lightweight machine learning classifiers, including SVM and LightGBM. The equalization, feature extraction, and model design are described in Section VI. The trained models are exported in compact representations for deployment on the sensing node.

Step B: On-device inference. During runtime, the sensing node converts PDM microphone data to PCM samples at 8 kHz, selected to reduce processing and memory overhead while preserving low-frequency UAV rotor harmonics. The audio stream is segmented into overlapping frames, and MFCC features and their temporal derivatives are computed in real time. Mean and standard deviation pooling over a 1 s window produces a fixed-length feature vector for binary classification. A positive UAV detection triggers an alert transmitted via LoRaWAN for remote monitoring (Section III).

VI. DESIGN

We now describe the design choices underlying the proposed acoustic detection pipeline, detailing the configuration of MFCC features, the proposed microphone-aware equalization approach, and the selection of lightweight ML models for embedded deployment.

A. MFCC Feature Design and Parameter Selection

MFCCs were selected to capture the spectral characteristics of UAV rotor noise while providing a compact representation suitable for embedded processing. Their configuration was determined using the UAV and background datasets described in Section IV, balancing detection performance and computational efficiency.

An 80/20 pooled train-test split was used for parameter tuning, while leave-one-domain-out (LODO) validation assessed

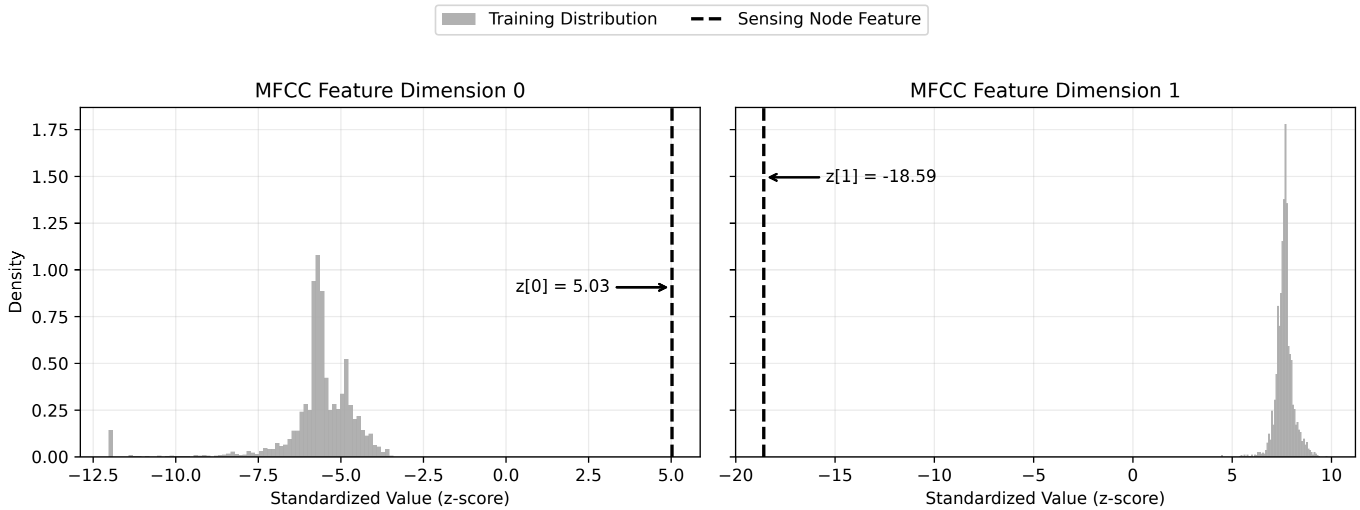


Fig. 2. Distribution of MFCC features for training datasets and sensing-node data, illustrating cross-domain mismatch.

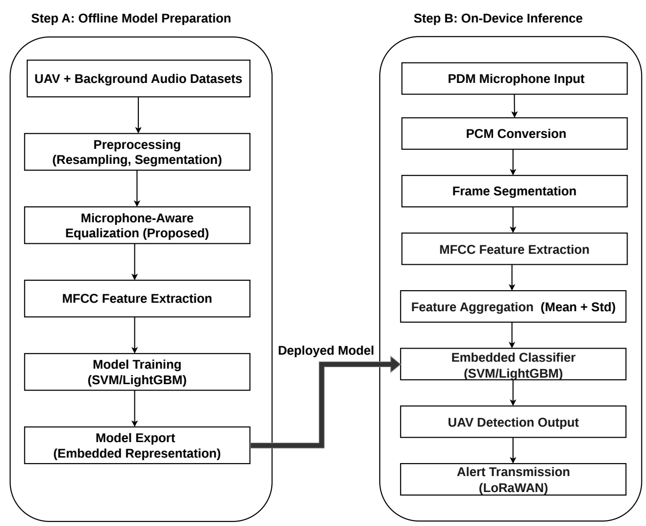


Fig. 3. Acoustic UAV detection pipeline of `AUDibLE`.

TABLE III
MFCC PARAMETER SEARCH RANGES

Parameter	Values Evaluated
Mel filters	64, 80, 128
MFCC coefficients	20, 22, 36
Maximum frequency	1500 Hz, 1800 Hz
Frame duration	25 ms, 35 ms
Hop duration	10 ms, 20 ms

TABLE IV
MFCC PARAMETER SELECTION RESULTS

Mel filters	MFCC	$f_{\max}$	Frame/Hop	F1
80	22	1500 Hz	25/10 ms	0.8939
80	22	1800 Hz	35/20 ms	0.9066
64	20	1500 Hz	25/10 ms	0.8998
64	20	1800 Hz	25/10 ms	0.8959
64	**20**	**1800 Hz**	**35/20 ms**	**0.9171**
128	36	1800 Hz	35/20 ms	0.8978

robustness across recording conditions. The baseline configuration of 80 Mel filters, 22 MFCC coefficients, 1500 Hz maximum frequency, and 25/10 ms frame/hop durations achieved an F1-score of 0.8939. A grid search was then performed over the parameter ranges in Table III, spanning different spectral and temporal resolutions while keeping classifier hyperparameters fixed.

Table IV summarizes selected configurations from the parameter sweep. The best configuration used 64 Mel filters, 20 MFCC coefficients, a maximum frequency of 1800 Hz, and 35/20 ms frame/hop durations, achieving an F1-score of 0.9171. Increasing the feature dimensionality to 128 Mel filters and 36 MFCC coefficients did not improve performance, supporting the compact configuration selected for embedded deployment.

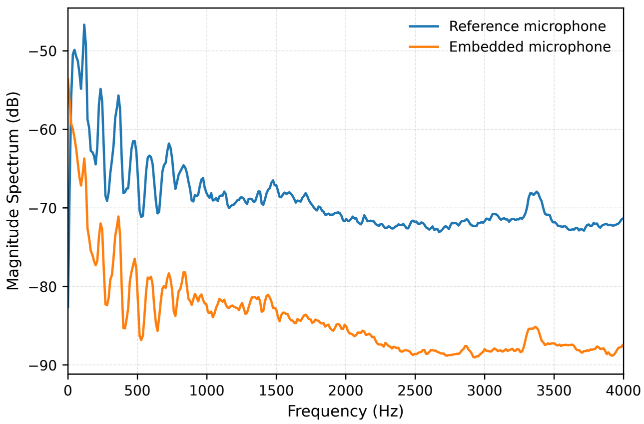
B. Microphone-Aware Equalization

To address the cross-domain mismatch described in Section IV, we designed a microphone-aware equalization filter using paired recordings captured simultaneously with a reference microphone and the sensing-node microphone. Each recording is resampled to 8 kHz, and its average magnitude spectrum is computed using short-time Fourier transform (STFT) analysis.

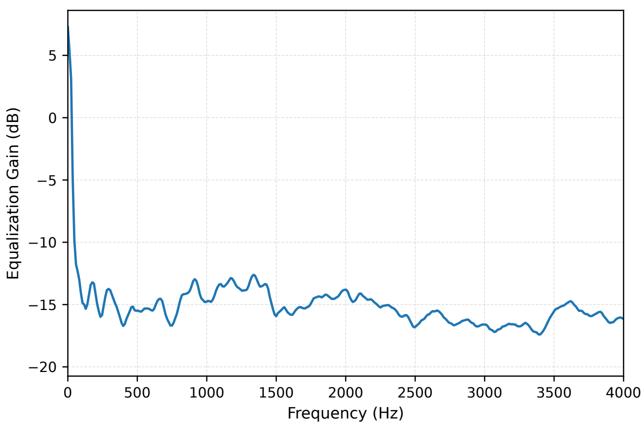
The microphone response mismatch is estimated as the ratio between the average spectra of the sensing-node and reference microphones:

$$H(f) = \frac{S_{\text{node}}(f)}{S_{\text{ref}}(f)}, \tag{1}$$

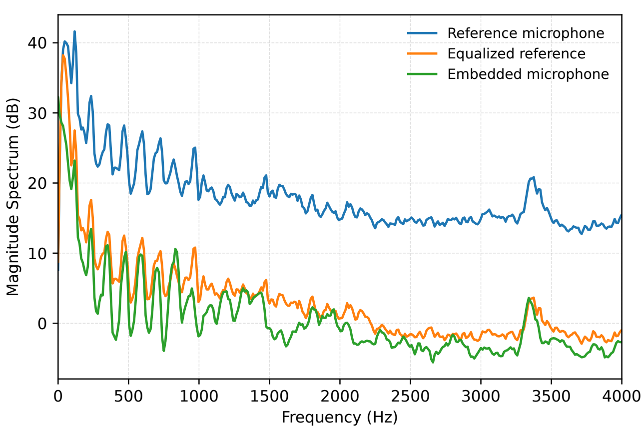
where $S_{\text{node}}(f)$ and $S_{\text{ref}}(f)$ denote the mean magnitude spectra of the sensing-node and reference microphones, respectively. The resulting frequency response is smoothed in the logarithmic domain and used to design a 257-tap linear-phase finite impulse response (FIR) filter, providing sufficient frequency resolution with moderate offline preprocessing complexity. The filter is applied to the training data to compensate for microphone-induced spectral distortion.



(a) Reference vs sensing-node spectra



(b) Estimated equalization gain



(c) Post-equalization spectral alignment

Fig. 4. Design and validation of the microphone-aware equalization filter.

Fig. 4a compares the average spectra of paired reference and sensing-node recordings, revealing frequency-dependent attenuation introduced by the embedded microphone. Fig. 4b shows the equalization gain estimated from their spectral ratio. After filtering, Fig. 4c shows reduced spectral mismatch between the reference and sensing-node recordings.

TABLE V
CLASSIFIER PERFORMANCE USING OFFLINE MFCC IMPLEMENTATION

Model	**Accuracy**	**Precision**	**Recall**	**F1**
SVM	0.9594	0.9483	0.9717	0.9599
XGBoost	0.9735	0.9653	0.9823	0.9737
LightGBM	0.9700	0.9586	0.9823	0.9703

TABLE VI
CLASSIFIER ROBUSTNESS WITH EMBEDDED MFCC IMPLEMENTATION

Model	**Accuracy**	**Precision**	**Recall**	**F1**
SVM	0.8917	0.8750	0.8305	0.8522
XGBoost	0.9427	0.9310	0.9153	0.9231
LightGBM	0.9618	0.9492	0.9492	0.9492

C. Model Selection for Embedded Deployment

To enable real-time UAV detection on resource-constrained devices, we evaluate SVM, XGBoost, and LightGBM using the MFCC representation described above. An SVM with a radial basis function (RBF) kernel is used as a baseline due to its strong acoustic classification performance, although support-vector evaluation can increase computational and memory requirements.

XGBoost and LightGBM are considered as tree-based alternatives offering efficient non-linear classification. In particular, LightGBM uses compact gradient-boosted trees that enable efficient inference on embedded hardware. All models use identical feature representations and dataset splits for fair comparison, with detection performance and computational efficiency evaluated in Section VII.

VII. RESULTS

This section evaluates `AUDibLE` under playback, real-world UAV flights, and embedded deployment conditions. We compare offline and embedded classifier performance, evaluate microphone-aware equalization across diverse UAVs, and report memory and latency measurements.

A. Classifier Comparison

We compare SVM, XGBoost, and LightGBM to evaluate classifier performance and robustness under embedded deployment conditions. All classifiers are trained using the same MFCC feature representation and identical dataset splits. Table V reports reference performance when the same offline MFCC implementation is used for training and testing. All classifiers perform strongly, with tree-based models achieving slightly higher accuracy and F1-score than SVM.

To evaluate deployment robustness, classifiers trained using the reference MFCC pipeline are also tested using features generated by the lightweight embedded implementation, where minor numerical differences may arise.

As shown in Table VI, all classifiers experience some degradation, but LightGBM maintains the highest F1-score under embedded feature extraction. Based on these results, subsequent deployment evaluations focus on SVM as the kernel-based baseline and LightGBM as the best-performing tree-based model.

TABLE VII
UAV PLATFORMS USED IN PLAYBACK EVALUATION (32 MODELS)

Manufacturer	Model	Env.
Self-built	David Tricopter	Out
Self-built	PhenoBee	Out
Autel	Evo 2 Pro	Out
DJI	Avata	Out
DJI	FPV	Out
DJI	Matrice 200	Out
DJI	Matrice 200 V2	Out
DJI	Matrice 600p	Out
DJI	Mavic Air 2	Out
DJI	Mavic Mini 1	Out
DJI	Mini 2	Out
DJI	Mini 3	Out
DJI	Mini 3 Pro	Out
DJI	Mavic 2 Pro	Out
DJI	Neo	Out
DJI	Mavic 2s	Out
DJI	Phantom 2	Out
DJI	Phantom 4	Out
DJI	Tello	In
DJI	RoboMaster TT	In
Hasakee	Q11	In
Holystone	HS210	In
Hover	X1	Out
Syma	X5SW	In
Syma	X5UW	In
Syma	X8SW	In
Syma	X20	In
Syma	X20P	In
Syma	X26	In
Swellpro	Splash 3+	Out
Yuneec	Typhoon H+	Out
UDI RC	U46	Out

B. Effect of Microphone-Aware Equalization

To evaluate the impact of the proposed microphone-aware equalization filter, the filter described in Section VI was applied during training data preparation, and the SVM classifier was retrained using the equalized data together with recordings acquired from the sensing node.

Playback evaluation used recordings from 32 UAV platforms in the drone-visualization dataset (DS4) [24], which was excluded from training and reserved for evaluating generalization to unseen UAVs. Models were trained on DS0–DS3 using a pooled 80/20 train-test split. Table VII summarizes the evaluated UAVs, spanning different manufacturers, rotor configurations, sizes, and operating environments.

The recordings were normalized to a fixed playback amplitude and extended to 30 s duration to simulate continuous UAV operation. After applying microphone-aware equalization, the retrained SVM classifier detected 31 out of 32 UAV recordings, corresponding to a detection rate of 96.9%. The only missed case was the DJI Mini 3 Pro, which produced comparatively weak acoustic signatures during playback.

These results demonstrate that compensating for microphone-induced spectral distortion significantly improves the transferability of models trained on heterogeneous datasets and enables reliable detection of previously unseen UAV platforms on the embedded sensing node.

TABLE VIII
REAL-WORLD EVALUATION RESULTS AND FALSE-POSITIVE ANALYSIS

Test Condition	SVM	LightGBM
UAV Detection Performance		
Drone Flights (40 trials)	40 / 40	40 / 40
False Positives (Background Noise Only)		
Ambient (birds, wind)	0 / 10	0 / 10
Human activity (speech, movement)	0 / 10	0 / 10
Transient events (clap, door bell)	0 / 10	0 / 10
Mechanical (motors, mixers, vacuum)	8 / 10	7 / 10

C. Real-World Drone Evaluation

To validate `AUDibLE` under real-world conditions, experiments were conducted using a T-Motor M690B quadcopter. The sensing node was deployed along the perimeter of an outdoor arena of $30\,\text{m} \times 15\,\text{m}$. The drone executed maneuvers including hovering, approach, landing, and rotational motion. Each scenario was repeated for 10 independent 30 s trials. Hovering tests were conducted at approximately 15 m, while the approach scenario involved the drone moving from approximately 15 m to 5 m. Additional tests were also performed at distances up to 20 m. The UAV was detected in all trials at distances between 5 m and 20 m.

The experiments were conducted in an outdoor environment with realistic background acoustic activity, including vehicles, birds, nearby aircraft, human conversation, and construction noise. Across all UAV-present trials, both classifiers detected the UAV, with no false negatives. Detection was typically triggered within the first processing window (approximately 1 s) after the UAV became acoustically distinguishable from the background. To evaluate robustness against environmental sounds, background-only trials included ambient noise, human activity, transient events, mechanical sources, and transportation noise.

Table VIII summarizes the real-world evaluation results. False positives were primarily associated with mechanical sound sources such as electric motors, mixers, and vacuum cleaners, which can resemble UAV rotor acoustics. In contrast, ambient sounds such as bird chirping, human conversation, and general outdoor noise did not trigger false detections. Incorporating such mechanical sounds as hard-negative training examples may further improve robustness.

D. Embedded Deployment Performance

We evaluate deployment feasibility in terms of memory footprint and processing latency on the `nRF54L15` sensing node.

1) Memory Footprint: Table IX reports the incremental memory footprint relative to baseline firmware without UAV detection. SVM requires 534 KB Flash compared with 106 KB for LightGBM, while both require 71.2 KB RAM.

2) Processing Latency: Table X reports execution time per 1 s analysis window. MFCC extraction requires 6.0 ms for both models, while classification requires 264.2 ms for SVM and only 0.2 ms for LightGBM.

Both models satisfy the 1 s real-time processing requirement; however, LightGBM provides substantially better mem-

TABLE IX
INCREMENTAL MEMORY FOOTPRINT OF THE UAV DETECTION MODULE RELATIVE TO THE BASELINE FIRMWARE

Model	Flash (KB)	RAM (KB)
SVM (RBF)	534	71.2
LightGBM	106	71.2

TABLE X
AVERAGE PROCESSING LATENCY PER 1 S AUDIO WINDOW ON THE SENSING NODE

Model	MFCC (ms)	Classification (ms)	Total (ms)	CPU (%)
SVM	6.0	264.2	270.2	27.0
LightGBM	6.0	0.2	6.2	0.6

ory and computational efficiency, making it more suitable for resource-constrained deployment.

VIII. CONCLUSIONS

This paper presented `AUDibLE`, a low-power, low-cost acoustic UAV detection system that performs MFCC-based feature extraction and machine learning inference directly on an `nRF54L15` sensing node. To address microphone-induced domain mismatch between training datasets and embedded sensing hardware, we introduced a microphone-aware equalization strategy that aligns training data with the sensing-node microphone response.

Playback evaluation across 32 UAV platforms achieved a detection rate of 96.9% with the retrained SVM, while real-world flight experiments achieved 100% detection across all trials. LightGBM provided significantly better deployment efficiency than SVM with lower Flash memory usage and inference latency, demonstrating the feasibility of real-time UAV detection on resource-constrained sensing nodes.

False positives were primarily associated with mechanical sounds exhibiting spectral characteristics similar to UAV rotor noise, while typical ambient sounds did not trigger false detections, indicating that the limitation is localized and can be addressed through targeted training. Future work will incorporate hard-negative training using such mechanical sound sources and explore temporal modeling of rotor harmonics to further improve detection robustness.

ACKNOWLEDGMENTS

This work has been partially executed within the SPIDR^2 project ("Secure, Performant, Intelligent, Dependable, and Resilient Wireless Mesh Networks") financed by the Technology Innovation Institute.

REFERENCES

[1] 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.

[2] S. Jeon, J.-W. Shin, Y.-J. Lee, W.-H. Kim, Y. Kwon, and H.-Y. Yang, "Empirical Study of Drone Sound Detection in Real-Life Environment with Deep Neural Networks," in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, 2017, pp. 1858–1862.

[3] P. Casabianca and Y. Zhang, "Acoustic-Based UAV Detection Using Late Fusion of Deep Neural Networks," *Drones*, vol. 5, no. 3, 2021. [Online]. Available: https://www.mdpi.com/2504-446X/5/3/54

[4] S. Al-Emadi, A. Al-Ali, and A. Al-Ali, "Audio-Based Drone Detection and Identification Using Deep Learning Techniques with Dataset Enhancement through Generative Adversarial Networks," *Sensors*, vol. 21, no. 15, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/15/4953

[5] S. Kümmritz and L. Paul, "Comprehensive Database of UAV Sounds for Machine Learning," in *Proceedings of the* 10^{th} *Convention of the European Acoustics Association Forum Acusticum*, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:267136465

[6] M. Wang, M. Linn, A. 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.

[7] M. Aledhari, R. Razzak, R. M. Parizi, and G. Srivastava, "Sensor Fusion for Drone Detection," in *Proceedings of the* 93^{rd} *IEEE Vehicular Technology Conference (VTC-Spring)*, 2021.

[8] J. Kim, Y. Kim, H. Shin, Y. Wang, and E. T. Matson, "How Far Can a Drone be Detected? A Drone-to-Drone Detection System Using Sensor Fusion." in *Proceedings of the* 15^{th} *International Conference on Agents and Artificial Intelligence (ICAART)*, 2023, pp. 877–884.

[9] M. Ezuma, F. Erden, C. Kumar Anjinappa, O. Ozdemir, and I. Guvenc, "Detection and Classification of UAVs Using RF Fingerprints in the Presence of Wi-Fi and Bluetooth Interference," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 60–76, 2020.

[10] M. F. Al-Sa'd, A. Al-Ali, A. Mohamed, T. Khattab, and A. Erbad, "RF-Based Drone Detection and Identification Using Deep Learning Approaches: An Initiative Towards a Large Open Source Drone Database," *Future Generation Computer Systems*, vol. 100, pp. 86–97, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X18330760

[11] P. Molchanov, K. Egiazarian, J. Astola, R. I. A. Harmanny, and J. J. M. de Wit, "Classification of Small UAVs and Birds by Micro-Doppler Signatures," in *Proceedings of the European Radar Conference*, 2013, pp. 172–175.

[12] S. Rahman and D. A. Robertson, "Classification of Drones and Birds Using Convolutional Neural Networks Applied to Radar Micro-Doppler Spectrogram Images," *IET Radar, Sonar and Navigation*, vol. 14, no. 5, pp. 653–661, 2019.

[13] A. Rozantsev, V. Lepetit, and P. Fua, "Detecting Flying Objects Using a Single Moving Camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 5, pp. 879–892, 2017.

[14] M. Saqib, S. Daud Khan, N. Sharma, and M. Blumenstein, "A Study on Detecting Drones Using Deep Convolutional Neural Networks," in *Proceedings of the* 14^{th} *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2017, pp. 1–5.

[15] M. Mrabet, M. Sliti, and L. B. Ammar, "Machine learning algorithms applied for drone detection and classification: benefits and challenges," *Frontiers in Communications and Networks*, vol. 5, 2024. [Online]. Available: https://www.frontiersin.org/journals/communications-and-networks/articles/10.3389/frcmn.2024.1440727

[16] O. A. Ibrahim, S. Sciancalepore, and R. Di Pietro, "Noise2Weight: On Detecting Payload Weight from Drones Acoustic Emissions," *Future Gener. Comput. Syst.*, vol. 134, no. C, p. 319–333, 2022. [Online]. Available: https://doi.org/10.1016/j.future.2022.03.041

[17] C. Dumitrescu, M. Minea, I. M. Costea, I. Cosmin Chiva, and A. Semenescu, "Development of an Acoustic System for UAV Detection," *Sensors*, vol. 20, no. 17, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/17/4870

[18] S. Mandal, L. Chen, V. Alaparthy, and M. L. Cummings, "Acoustic Detection of Drones through Real-Time Audio Attribute Prediction," in *Proceedings of the AIAA Scitech Forum*, 2020, p. 0491.

[19] Nordic Semiconductor, "nRF54L15 Product Specification," 2024, https://www.nordicsemi.com/Products/nRF54L15, Accessed: 2026-03-03.

[20] STMicroelectronics, "MP34DT05-A Digital MEMS Microphone Datasheet," 2019, https://www.st.com/resource/en/datasheet/mp34dt05-a.pdf, Accessed: 2026-03-03.

[21] "Drone-Payload Dataset," 2020, available: https://github.com/cri-lab-hbku/Drone-Payload, Accessed: 2026-03-03.

[22] "Drone Audio Dataset: Binary Drone Audio Classification Dataset," 2021, available: https://github.com/saraalemadi/DroneAudioDataset, Accessed: 2026-03-03.

[23] "Drone Detection Audio Dataset," 2020, available: https://github.com/DroneDetectionThesis/Drone-detection-dataset, Accessed: 2026-03-03.

[24] "Drone Visualization Audio Dataset," 2022, available: https://github.com/mackenzie-jane/drone-visualization, Accessed: 2026-03-03.