Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
April 23, 2026
A Regional Conference of the Society for Industrial and Systems Engineering

CNN-Based Acoustic Detection of Small UAS: Prototype and Field Evaluation

Jordan Follenweider, Zachary Hanna, Loro Lado, Cameron Mock, Justin Won, and John Caddell

Department of Systems Engineering, United States Military Academy, West Point, New York 10996

Corresponding author's Email: john.caddell@westpoint.edu

Author Note: The authors of this paper conducted a year-long research project through the Department of Systems Engineering at the United States Military Academy under the advisement of LTC J.D. Caddell (Assistant Professor). The views expressed in this paper are those of the authors and do not reflect those of the United States Army, the United States Military Academy, or the Department of Systems Engineering.

Abstract: Current Counter-Unmanned Aerial Systems (C-UAS) primarily rely on radar, electro-optical/infrared (EO/IR), and radio frequency detection. Each of these methods has significant limitations against small, low-altitude, or autonomously operated drones. Acoustic sensing offers a passive, low-cost complement but remains underutilized. This paper presents an acoustic detection prototype that pairs a microphone-based sensor with a Convolutional Neural Network (CNN) classifier trained on drone audio signatures. Field testing demonstrated reliable detection at close range (≤25 meters), with intermittent detection extending to 75 meters. Classification performance degraded at longer ranges due to signal attenuation and environmental noise. These results suggest acoustic sensing can meaningfully supplement existing C-UAS architectures. Future work will expand training data diversity and evaluate performance across varied environmental conditions.

Keywords: Counter-Unnamed Aerial System (C-UAS), Machine Learning, Acoustic Sensing, Convolutional Neural Networks.

1. Introduction

The proliferation of small, inexpensive unmanned aerial systems (UAS) has outpaced the detection capabilities of conventional military defense systems. Recent conflicts, particularly the war in Ukraine, demonstrate the growing use of low-altitude drones for reconnaissance, strike missions, and coordinated swarm attacks, often at scales that overwhelm existing countermeasures (Farrell, 2022).

Current C-UAS detection relies on radar, EO/IR, and radio frequency (RF). Radar, especially the low-frequency based systems, provides reliable range and altitude measurements but struggles to detect small drones. These radars are not effective in detecting smaller UAS because they often have minimal radar cross-sections (De Cubber et al., 2025). EO/IR sensors support identification but degrade under poor weather and visibility. RF detection fails entirely when drones operate autonomously or use fiber-optic control links (Khawaja et al., 2025). Integrating these systems together may increase detection range and extend the decision-making window against incoming threats (Majors & O'Neil, 2021). However, integration is constrained by factors such as cost, sensor synchronization complexity, and power requirements.

Acoustic sensing offers a passive, low-cost approach that detects the distinctive sound signatures produced by drone motors and propellers (Tejera-Berengue et al., 2024). Although acoustic methods have shorter detection ranges and are sensitive to environmental noise, recent work demonstrates that machine learning techniques such as Mel-frequency cepstral coefficient (MFCC) feature extraction paired with CNNs can significantly improve classification accuracy (Al-Emadi et al., 2021). This has driven increasing interest in integrating acoustic sensing into multi-sensor C-UAS architectures.

This paper presents the development and field evaluation of an acoustic drone detection prototype that uses a microphone-based sensor paired with a CNN classifier. We describe the system design, training methodology, and results from range-dependent field tests conducted at the United States Military Academy.

2. Stakeholder, Requirements, and Functional Analysis.

Stakeholder engagements informed the system requirements and operational context for this project. The team consulted three primary stakeholders: PAE Fires (Program of Analysis and Experimentation, Fires), who provided operational framing for C-UAS needs; Dmytro Bielievtsov, lead engineer of the Zvook acoustic detection system, who shared operational insight from acoustic C-UAS employment in Ukraine; and leadership from 1st Battalion, 12th Infantry Regiment, who described how acoustic sensors could improve close-fight operations based on their recent experience conducting C-UAS operations along the United States-Mexico border.

Three consistent themes emerged from these engagements: the sensor must be lightweight and portable, it must identify drones based on acoustic signatures, and it must maintain detection capability in noisy environments. Figure 1 presents the functional hierarchy derived from these requirements. The three major system functions are *detect drones*, *operate in adverse conditions*, and *communicate status*. Further analysis of these functions and their subfunctions revealed the benefits of a modular system design that enables incremental improvement following the establishment of a baseline capability.

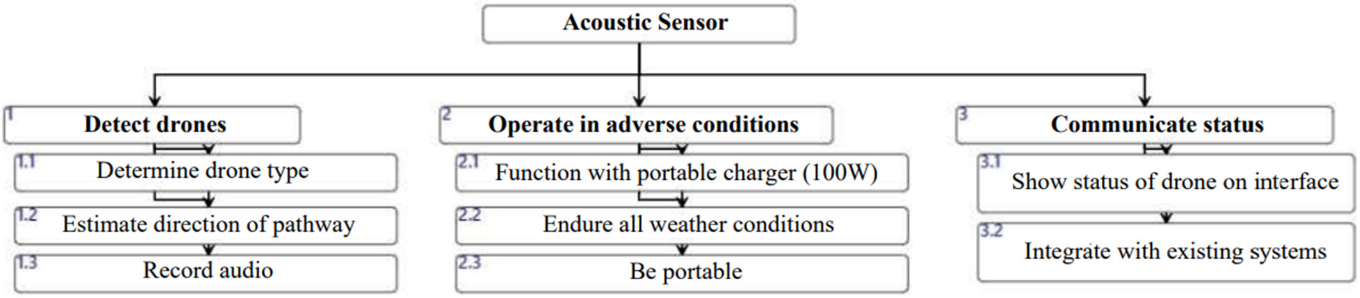

Figure 1. Functional Hierarchy

3. Prototype Development

The prototype was developed iteratively, in accordance with the Spiral Development Method (Driscoll at el., 2023), beginning with acoustic data collection hardware and progressing through model training and interface development. The activity diagram in Figure 2 illustrates the compartmentalization of the development. Raw acoustic waveform data is first converted into MFCCs, then split into training (60%), validation (15%), and testing (25%) datasets. A CNN is trained on these spectrograms to classify audio as drone or no-drone.

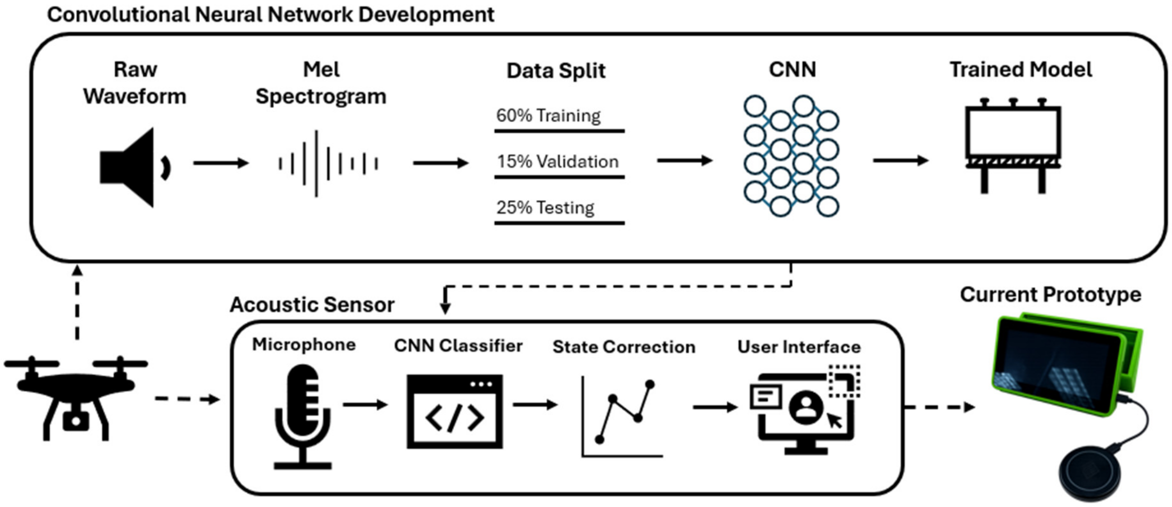

Figure 2. Activity Diagram of Acoustic Sensor Prototype

CNNs can learn spatial patterns within structured input data, making them well-suited for acoustic classification (Wang, 2023). This is important because MFCCs represent sound as a time-frequency pattern, from which the CNN learns the key features of a drone's acoustic signature. The prototype CNN architecture is based on the model proposed by Al-Emadi et al. (2021) and consists of two convolutional layers followed by a linear layer and a hidden fully connected layer. In their work, Al-Emadi et al. (2021) used this CNN architecture specifically for audio-based drone detection and identification on a dataset but never tested it in an operational environment. The model was trained on one-second audio clips captured by the initial sensor prototype, matching the clip length used by Al-Emadi et al. (2021) to enable real-time detection; their experiments confirmed that one-second segments were sufficient for accurate classification compared to longer durations.

During training, the model was evaluated against the validation dataset every 100 steps. If validation accuracy failed to improve over multiple consecutive checks, indicating the model had stopped learning, training was halted via early stopping and the best-performing model was evaluated against the held-out test set.

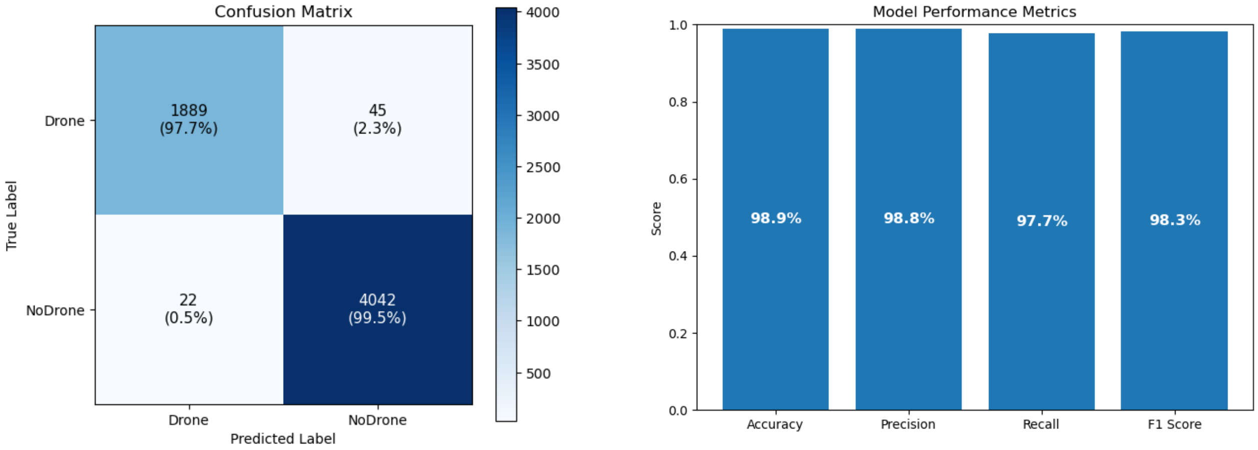

Figure 3. Confusion Matrix and Model Performance Metrics

The test set contained 5,998 clips at a 1:2 ratio of drone to no-drone audio. As shown in the confusion matrix (Figure 3), the model achieved high rates of true positives and true negatives with low false positive and false negative rates. Accuracy (98.9%), precision (98.8%), recall (97.7%), and F1 scores (98.3%) were consistently high, comparable to the results reported by Al-Emadi et al. (2021). However, these metrics reflect performance on held-out test data and do not capture real-world field conditions, which are evaluated in Section 4

The prototype was built on a Raspberry Pi 5 paired with a 7-inch display, selected for its compact form factor to approximate a field-deployable system. Audio sensing was provided by the ReSpeaker, a four-microphone omnidirectional array, chosen for its directional detection capabilities and low cost. The system was powered by a 100W portable charger, with a micro-SD card for storage and USB interfaces for peripherals and the microphone array. To house the microphone, processor, and display in a soldier-portable form factor, the team designed and 3D-printed a custom enclosure that protects internal components and supports rapid setup during field deployment.

The trained CNN model was deployed on the prototype hardware as shown in Figure 2. The omnidirectional microphone continuously records ambient audio, which is passed to the CNN classifier to identify potential drone acoustic signatures. To improve detection stability, the raw classifier output is processed through a state correction module that applies exponential smoothing and a hysteresis threshold. Exponential smoothing dampens transient fluctuations in prediction confidence, while the hysteresis threshold prevents the system from rapidly toggling between drone-detected and no-drone states. The corrected detection state is then displayed to the user through a graphical interface.

4. Testing Implementation

System performance was evaluated through stationary field tests designed to measure detection reliability as a function of range. The test drone was a DJI Mini 4k, although not being used in a conflict, it was selected because its low-noise motors produce a relatively weak acoustic signature, representing a challenging detection target. The drone was positioned at fixed distances of 10, 25, 50, 75, and 100 meters from the sensor, with horizontal distance and altitude held equal at each position (e.g., 50x50 denotes 50m horizontal distance at 50m altitude). Three one-minute drone-present recordings were collected at each distance. An eight-minute baseline recording with no drone present was also collected to characterize ambient noise conditions. The team selected one-minute intervals because the key performance indicators, time to first detection and sustained detection capability, are both observable within that duration. This interval also allowed efficient iteration between test distances while preserving drone battery life. The eight-minute no-drone baseline provided sufficient duration to observe false positive behavior across varying ambient conditions while optimizing available test times. Trial order was randomized to reduce environmental bias.

Testing was conducted at the approved drone range along the Hudson River at the United States Military Academy, an open field environment with relatively low ambient noise relative to other areas that could potentially be used for testing. During each trial, the prototype recorded raw audio data and stored real-time classifier predictions for subsequent analysis. Dynamic flight trials were planned but not executed because of lack of time and weather issues. However, it remains a priority for future testing.

5. Evaluation and Results

Figure 4 presents the classifier output probabilities across all test distances for Model v3, the highest performing model. At 10x10 meters, the model consistently produced probability values near 1.0, and the smoothing algorithm maintained continuous detection windows. As distance increased, prediction confidence became more variable, with output fluctuating between drone and no-drone classifications. At 100x100 meters, detections became sporadic and the smoothing algorithm could not sustain a stable detection state. The no-drone baseline trial maintained probability values near zero, with occasional false positives attributed to ambient sounds with spectral similarity to drone rotor harmonics. However, the state correction algorithm successfully filtered these intermittent false positives from the final detection state classification.

The percentage of time the system maintained a corrected detection state declined with range, with one exception. At 10x10 meters, the system maintained detection for 91.67% of the trial duration. Performance degraded to 76.73% at 25x25 meters and 31.41% at 50x50 meters. At 75x75 meters, however, the corrected state of detecting drone as present rose to 89.31%, departing from the expected monotonic decline with range. Visual inspection of the individual trials in Figure 4 shows considerable variability across the three 50x50 trials, suggesting this result may reflect favorable environmental conditions rather than a true performance characteristic. During testing, intermittent construction noise from nearby jackhammers, passing river traffic, and nearby aircraft may have contributed to variability in detection performance across trials, and these acoustics present one of the major limitations in our experiment. At 100×100 meters, the system failed to detect the drone during all three trials.

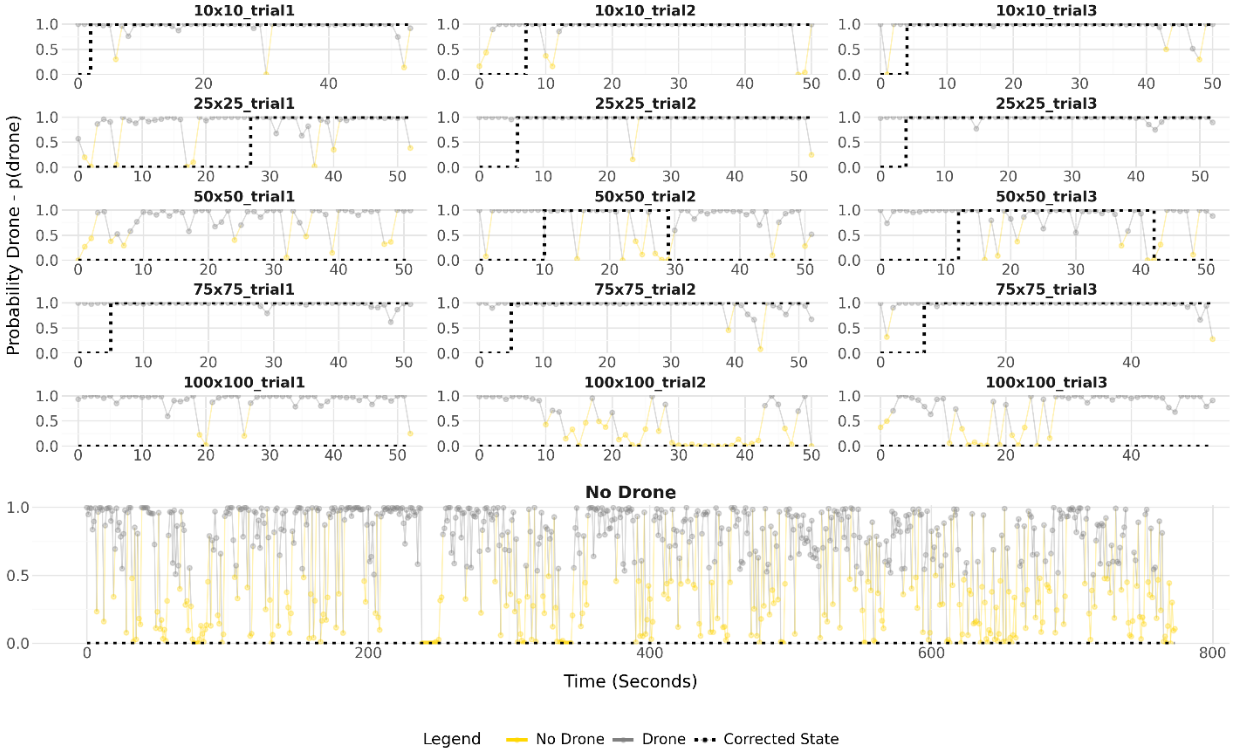

Figure 4. Results of Model v3

6. Discussions

The system performs most reliably when drone motor harmonics are clearly distinguishable from environmental noise, which occurs consistently at short ranges. As distance increases, signal weakening reduces the signal-to-noise ratio, producing greater variability in classifier output and occasional misclassifications. Because the classifier responds primarily to the dominant acoustic signal in the input, environmental sounds that exceed drone rotor harmonics in amplitude cause the model to classify based on the louder source rather than the drone signature. The intermittent construction noise observed during testing further demonstrates the system's sensitivity to environmental interference, particularly at ranges where the drone signal is already weak. The transition from Model v3 to Model v4 introduced additional no-drone recordings intended to improve environmental robustness. However, this expansion imbalanced the training dataset toward no-drone samples, and Model v4 significantly underperformed relative to Model v3. This result highlights that dataset composition, not only dataset size, is critical to classifier performance. Future training efforts must carefully maintain class balance while expanding the diversity of both drone and no-drone recordings. From an operational standpoint, the system's portability, low power consumption, and passive sensing make it well-suited for forward operating environments where minimizing electromagnetic emissions is critical. Additionally, because the test drone had a relatively weak acoustic signature, detection performance would likely improve against larger or louder platforms.

7. Limitations

Four primary limitations affect the current system. First, environmental noise, particularly wind, degrades detection performance by masking drone rotor harmonics. Wind screens or directional filtering could mitigate this effect. Second, natural acoustic attenuation reduces the microphone's ability to capture distinct rotor harmonics at longer ranges, producing increased variability in classification confidence. More sensitive microphone hardware or signal amplification techniques may extend effective detection range. Third, the training dataset remains limited in size and diversity, representing a narrow range of drone types, flight conditions, and environmental backgrounds. Fourth, as demonstrated by the degraded performance of Model v4, expanding the dataset without maintaining class balance between drone and no-drone samples can degrade rather than improve classifier performance. Future iterations must simultaneously increase data diversity and preserve balanced class representation to avoid repeating this outcome.

8. Conclusion

This study demonstrates that acoustic sensing is a viable approach to small UAS detection. The prototype, pairing a microphone-based sensor with a CNN classifier trained on drone audio signatures, achieved reliable detection at 10x10 meters with 91.67% corrected-state time and sustained detection capability out to 75 meters under favorable conditions. Because testing used a DJI Mini 4k, a platform with a relatively weak acoustic signature, these results represent a conservative baseline for detection performance against operationally relevant platforms.

Two key findings emerged from this work. First, the classification model responds primarily to the dominant acoustic signal in the input. When environmental noise exceeds drone rotor harmonics, particularly at longer ranges, it masks the drone's signature and degrades classification accuracy. Second, dataset composition critically affects classifier performance. The transition from Model v3 to Model v4 demonstrated that expanding training data without maintaining class balance degrades rather than improves detection capability.

As a passive, low-cost, and portable sensing modality, acoustic detection addresses gaps in existing C-UAS architectures that rely on active emissions or line-of-sight. Realizing its full operational potential will require expanding and balancing the training dataset across diverse drone types and environmental conditions, developing improved noise filtering or signal isolation techniques to extract drone harmonics from competing sounds, and conducting dynamic flight trials to evaluate performance against maneuvering targets.

9. References

Al-Emadi, S., Al-Ali, A., & Al-Ali, A. (2021). Audio-Based Drone Detection and Identification Using Deep Learning Techniques with Dataset Enhancement through Generative Adversarial Networks. *Sensors, 21*(15), 4953. https://doi.org/10.3390/s21154953.

De Cubber, G., Doroftei, D., Petsioti, P., Koniaris, A., Brewczyński, K., Życzkowski, M., Roman, R., Sima, S., Mohamoud, A., van de Pol, J., Maza, I., Ollero, A., Church, C., & Popa, C. (2025). Standardized Evaluation of Counter-Drone Systems: Methods, Technologies, and Performance Metrics. Drones, 9(5), 354. https://doi.org/10.3390/drones9050354.

Driscoll, P. J., Parnell, G. S., & Henderson, D. L. (2023). *Decision Making in Systems Engineering and Management* (3rd ed.). Wiley.

Farrell, F. (2022). Ex-Navy captain: Ukraine's drone attack in Sevastopol 'has no equivalent in the history of naval warcraft'. *The Kyiv Independent.* https://kyivindependent.com/former-navy-captain-ukraine-drone-attack-in-sevastopol-has-no-equivalent/.

Khawaja, W., Ezuma, M., Semkin, V., Erden, F., Ozdemir, O., & Guvenc, I. (2025). *A survey on detection, classification, and tracking of UAVs using radar and communications systems.* arXiv. https://arxiv.org/abs/2402.05909.

Majors, D. W., & O'Neil, R. P. (2021). *Integration of Radar Sensor Data with Situational Awareness Tools to Respond to an Unmanned Aerial Threat* [Master's Thesis, Naval Postgraduate School]. https://apps.dtic.mil/sti/trecms/pdf/AD1151058.pdf.

Tejera-Berengue, D., Zhu-Zhou, F., Utrilla-Manso, M., Gil-Pita, R., & Rosa-Zurera, M. (2024). Analysis of distance and environmental impact on UAV acoustic detection. *Electronics, 13*(3), 643. https://doi.org/10.3390/electronics13030643

Wang, Y. (2023). *A large-scale UAV audio dataset and audio-based UAV classification using CNN* (Doctoral dissertation). Purdue University.