12th INTERNATIONAL SCIENTIFIC CONFERENCE ON DEFENSIVE TECHNOLOGIES
OTEH 2026
Serbia, Tara, 21 - 24 October 2026

SENSOR FUSION FOR DRONE CLASSIFICATION IN RADIO FREQUENCY DOMAIN

DARKO MIKANOVIĆ
Military Technical Institute, Belgrade, darko.mikanovic@mod.gov.rs

BOBAN SAZDIĆ-JOTIĆ
Military Technical Institute, Belgrade, boban.sazdic.jotic@vs.rs

VASILIJA JOKSIMOVIĆ
Military Technical Institute, Belgrade, vasilija.joksimovic@mod.gov.rs

Abstract: *The increasing usage of Unmanned Aerial Systems in critical areas requires robust Counter Unmanned Aerial Systems solutions to maintain security and operational integrity. This paper investigates the impact of sequential sensor data fusion on the classification accuracy of Unmanned Aerial Systems signals within the Radio Frequency domain for real-time inference. Given the hardware constraint of a single Radio Frequency receiver, we propose a tailored fusion architecture that integrates sequential data acquired over time. Utilizing Radio Frequency spectrograms as the primary input, various sensor fusion methods were evaluated in combination with simple Convolutional Neural Networks. To ensure compatibility with our deep learning models, particularly under the constraints of Radio Frequency sensors that are often affected by low Signal-to-Noise Ratio and spectral interference in dense environments, the proposed fusion methods produce post-processed spectrograms that retain the original input dimensions. This design enables seamless integration into pre-trained deep learning pipelines without necessitating architectural modifications, while also addressing conditions that may otherwise contribute to elevated false-positive rates. The proposed methods were benchmarked against multiple pre-recorded validation datasets specifically designed to simulate real-world scenarios. The results demonstrate that specific temporal sensor fusion approaches significantly enhance prediction accuracy and overall decision-making robustness, providing a more stable decision boundary compared to baseline models without data fusion. Ultimately, the proposed approach offers a computationally efficient technique to improve inference reliability without additional overhead hardware or extensive model fine-tuning. By effectively mitigating stochastic noise and signal fading, this method reduces false alarm rates by up to 15%, suggesting that incorporating time-domain sensor fusion as a preprocessing layer is a scalable, cost-effective solution for real-time drone classification in complex Radio Frequency environments.*

Keywords: *classification, drone, convolutional neural network, fusion, radio frequency, sensors.*

1. INTRODUCTION

The development of Unmanned Aerial Systems (UAS) or colloquial “drones” and the technology they use is progressing at an incredible pace, while the reduction in their price and the simplification of their use contribute to their increasing availability and accessibility. As a result, there has been a need for protection against malicious drones that seek to damage critical structures in some way. This is how Counter Unmanned Aerial Systems (CUAS) emerged, which, with the help of sensor systems and Artificial Intelligence (AI) algorithms, detect and identify drones and react appropriately in order to neutralize the threat [1].

Most AI algorithms developed for this application rely primarily on individual sensor inputs, such as images, whereas real-time scenarios in which sensors continuously generate large volumes of image data are typically not sufficiently addressed [2]. When used in real-time, detection algorithms often give too many decisions per unit of time, and the end-user does not benefit from that. Moreover, such decisions are imperfect because of the detection algorithms themselves, which have a certain accuracy over trained datasets that are generally not large and diverse enough due to unavailability of adequate training data and the very nature of the problem. This is strictly connected with the characteristics of Radio Frequency (RF) environment often affected by low Signal-to-Noise Ratio (SNR), spectral interference and propagation of electromagnetic waves [3], [4], [5].

The research subject includes the introduction of automatic data fusion into a system with a single sensor for detection and identification of drones in the RF domain. Multiple early and late fusion techniques were tested through experimental research (spectrogram merging, hard and soft fusion), and the corresponding results are presented offering the tailored approach for this specific problem. Each of the fusion approaches are explained in the Methodology chapter of the paper. The main goal is to increase the reliability and accuracy of drone detection and identification CUAS systems that

Corresponding author: D. Mikanović;
E-mail: darko.mikanovic@mod.gov.rs
DOI: 10.5937/OTEH26057M

leverage already trained Convolutional Neural Network (CNN) models for predictions.

There are several advantages of using optimal sensor fusion techniques. The reliability and accuracy of drone detection and identification are increased with very little overhead and using the available prediction models as-is. The number of decisions made per unit of time is brought closer to the optimum for the end-user. An optimal sensor fusion is also a good starting point for more advanced data fusion approaches that include other types of fusion, such as post-prediction late fusion, that can increase the accuracy and reliability of detection systems even more.

The limitation of the sensor fusion approach is that it cannot improve the predictions accuracy by much on its own without other data fusion layers, though incorporating sensor fusion along with advanced feature fusion or late fusion methods tend to bring the best results.

The main contribution of this work consists of several facts: improvement of the classification accuracy in CNN models with temporal data fusion approaches, examination of the classification accuracy in 2,4GHz frequency band, testing the possibility of detecting multiple drones operating at the same time, and novel benchmark for examining the classification accuracy of drone classification in RF domain.

The rest of the paper is organized as follows: chapter Literature presents the recent research in this area specifically solving this problem, chapter Methodology presents the system model and explains to the reader how the experiments are performed, chapter Results presents the obtained results with discussion, and finally chapter Conclusion concludes the study.

2. LITERATURE

2.1 RF Domain and Spectrogram Analysis for UAS Classification

The rapid proliferation of UAS within critical airspace underscores the immediate necessity for dependable CUAS technologies [6]. Traditional drone monitoring techniques—primarily dependent on visual recognition, electro-optical sensors (cameras), and acoustic arrays—exhibit severe performance drop-offs when subjected to high ambient noise, physical occlusions, and adverse weather conditions [4], [7], [8]. To bypass these vulnerabilities, RF sensing has emerged as an indispensable modality, offering an all-weather, long-range, non-line-of-sight strategy by intercepting telemetry and control packets exchanged between a drone and its ground control station [9]. Recent trends favor Deep Learning (DL) and AI pipelines over hand-crafted statistical extractions to interpret these signal environments [7], [10], [11]. Specifically, by applying the Short-Time Fourier Transform (STFT), researchers transform one-dimensional raw time-series data or In-phase and Quadrature (I/Q) samples into two-dimensional visual spectrogram layouts [3], [5], [12], [13]. This processing stabilizes signal variance, preserves underlying temporal-spectral fluctuations, and maps the physical channel behaviors directly into joint time-frequency domains optimal for CNN feature backbones [3], [5], [11].

2.2 Multi-Sensor and Cross-Modality Data Fusion Architecture

While 2D DL models yield high accuracy profiles under pristine configurations, real-world RF landscapes remain prone to multipath scattering, intermittent fading, and deliberate or stochastic spectral noise, inflating false-positive metrics [3]. To reinforce system robustness, cross-modality data fusion frameworks have been widely engineered. For instance, multi-sensory CUAS architecture combining heterogeneous sensors—such as radars, electro-optical sensors (cameras), acoustic arrays, and RF sensors—leverage complementary physical properties to yield substantial improvements in drone verification accuracy, spatial classification, and real-time tracking efficiency [7], [10], [14], [15], [16]. Within specific dual-stream frameworks, DL models have been successfully fused with parallel convolutional channels to combine raw acoustic waveforms with RF signatures, illustrating high tracking efficacy across heterogeneous clusters [4], [9]. Furthermore, multi-modal network schemes are frequently supported by advanced algorithmic backbones, such as contrastive learning Siamese topologies or Kalman filters, to reliably identify unauthorized payloads and stealth drone threats under highly complex electromagnetic interference [3], [11], [17].

2.3 Single RF sensor Preprocessing Paradigms and the Temporal RF Fusion Gap

Despite the high detection performance guaranteed by multi-sensor deployment, the physical implementation of distributed, multi-modal hardware remains heavily constrained. Deploying physical radar arrays alongside synchronized electro-optical sensors (cameras) and acoustic arrays introduces severe system configuration complexity, excessive computational workloads, and aggressive calibration costs [9]. Consequently, modern methodologies are actively pivoted toward data reconstruction and virtual feature fusion frameworks capable of extracting hidden or multi-feature temporal streams inside a single-sensor hardware pipeline [12]. Synthesizing sequential temporal segments or multi-scale data profiles prior to or within intermediate convolutional layers provides a highly stable decision boundary, maximizing the model's macro-precision and robustness without scaling hardware boundaries [5], [12].

However, to the best of our knowledge, analysis of existing CUAS literature reveals a stark architectural dichotomy. Current state-of-the-art data fusion solutions are overwhelmingly dependent on either multi-modal hardware configurations (fusing RF with radar [1], [4], [7], [8], [10], [14], [15], [16], [17], vision [9], [18], or acoustics [4]) or concurrent multi-channel array nodes [3]. There remains a severe scarcity of scientific research evaluating sequential, time-domain data fusion

architectures engineered specifically for single-receiver RF infrastructures. Most single-sensor designs still rely on isolated, single-frame spectrogram inferences that lack temporal context and fail during deep signal fading or transient interference bursts. This critical research gap establishes a clear demand for lightweight, computationally efficient preprocessing frameworks capable of dynamically aggregating temporal RF states into standardized dimensions, optimizing classification accuracy without demanding overhead hardware or resource-intensive model fine-tuning.

3. METHODOLOGY

3.1 Fusion Architecture

The proposed system (Figure 1) processes RF spectrograms produced by a single receiver through two independently pre-trained CNN classifiers, one for counting drones (Model 1) and the other for identifying a single drone (Model 2). These models were also used in paper [13], where they are described in detail. Model 1's classes are: 0 drones, 1 drone, 2 drones and 3 drones; model 2's classes are: Lightbridge (0), OcuSync20 (1) and OcuSync40 (2). Model 2's predictions are relevant only when model 1 outputs 1 drone as the prediction. Combining both models' predictions, the proposed data fusion approach has 6 possible decisions to make:

- No drones (model 1: 0),
- Lightbridge (model 1: 1, model 2: 0),
- OcuSync20 (model 1: 1, model 2: 1),
- OcuSync40 (model 1: 1, model 2: 2),
- Two drones (model 1: 2),
- Three drones (model 1: 3).

Spectrograms are recorded in advance and grouped into four different scenarios which are used in this research for fusion evaluation and they will be explained in the Experimental setup section and presented in the Results and discussion chapter. Two families of data fusion are considered: early (sensor) fusion, which merges several successive spectrograms into a single input prior to classification, and late (hard and soft) fusion, which combines the classifiers' outputs over several consecutive predictions. Sensor and late fusion can also be combined, and every technique is configurable in terms of how many spectrograms or predictions are aggregated. Each fusion technique will be described in the following sections.

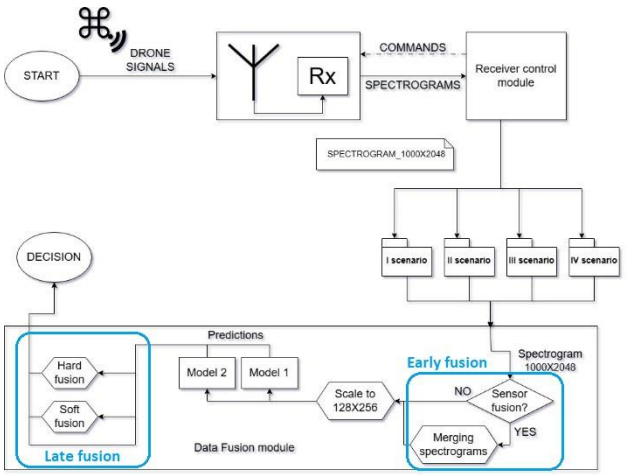


Figure 1. Methodology

3.2 Sensor (Early) Fusion

Since only a single RF sensor is available, sensor fusion here is limited to merging successive spectrograms from that sensor into one spectrogram matching the models' input size. Three merging methods were implemented: vertical concatenation of successive spectrograms followed by rescaling, which compresses drone signal patterns along the time axis while preserving their width — the primary feature distinguishing drone types (Figure 2); pixel-wise logical OR of the aligned spectrograms, which preserves the original dimensions but accumulates noise and interference from every merged frame (Figure 3); and pixel-wise logical XOR, intended to reduce this noise accumulation relative to OR at the cost of a more fragmented resulting signal (Figure 4). Vertical concatenation showed the best results among these three methods in qualitative tests in the 2,4GHz frequency band use case, where the accuracy is increased only slightly when compared to late fusion approaches.

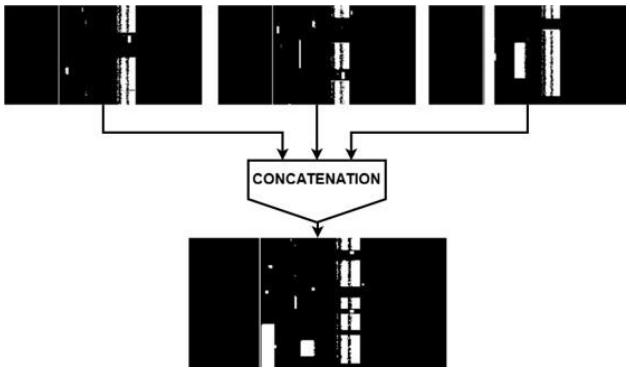


Figure 2. Sensor Fusion – Concatenation

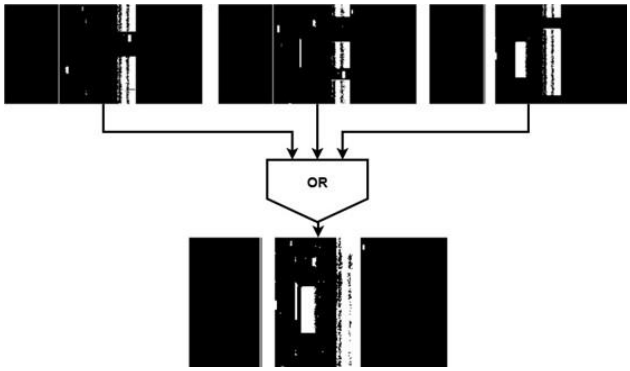


Figure 3. Sensor Fusion – OR operation

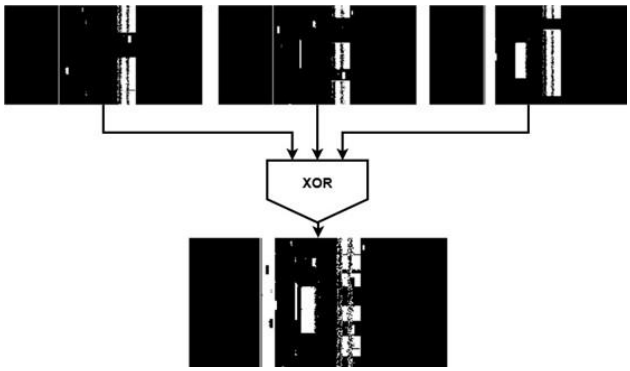


Figure 4. Sensor Fusion – XOR operation

3.3 Score-Level Fusion: Soft Fusion

Soft fusion combines the probability vectors produced by a classifier over several consecutive spectrograms by summing them element-wise and selecting the class with the highest cumulative probability. The soft fusion additionally folds the previous decision window's summed vector into the current sum, scaled by an empirically obtained factor k = 0.25. This makes fusion outcomes more resistant to short-lived state changes, at the cost of slightly slower adaptation in environments where drones frequently appear or disappear, and is therefore best suited to relatively stable operating environments.

3.4 Decision-Level Fusion: Hard Fusion

Hard fusion instead combines the discrete class predictions made over several consecutive spectrograms, using at least three predictions to avoid frequent ties. Under the majority voting rule, the final decision is the class occurring most often within the prediction window (e.g., for predictions [0, 1, 0, 2, 2, 0], the majority voting outcome is class 0). Another rule that is implemented is the longest sequence rule, where the final fusion decision is the class that makes the longest consecutive sequence in the prediction window (e.g., for predictions [0, 0, 1, 1, 1, 0, 0], the longest sequence outcome is class 1). Residual ties are resolved through a configurable secondary rule.

3.5 Experimental Setup

To emulate realistic operating conditions, all recordings were made indoors, where sporadic 2,4GHz interference from other wireless devices is present, and drone signals were intentionally attenuated by placing the drones at a distance from the receiving antenna. Manual band switching between 2,4GHz and 5,8GHz and drone activation/deactivation were used to simulate dynamic scenarios. Four scenarios were each recorded once, so that every fusion configuration is evaluated on identical data: (I) a single Lightbridge drone featured in 108 spectrograms; (II) a single OcuSync40 drone featured in 72 spectrograms; (III) two drones (Lightbridge and OcuSync20) in 108 spectrograms; and (IV) three drones (Lightbridge, OcuSync20 and OcuSync40), 252 spectrograms – each scenario including drone-free intervals. Sensor fusion was limited to merging at most four spectrograms, and late fusion windows were capped at nine consecutive predictions, reflecting the real-time constraint of at most one fusion decision every three seconds at an acquisition rate of three spectrograms per second.

4. RESULTS AND DISCUSSION

Table 1 summarizes, for each scenario, the classification accuracy obtained without fusion and the best-performing early and late fusion configurations identified during testing. Fusion decisions are compared against ground truth at the timestep granularity determined by the corresponding fusion window. Early fusion results in scenarios II and III are omitted because no early fusion configuration could achieve the accuracy of models without fusion due to scenarios' nature. N_{sp} is the number of successive spectrograms used for sensor fusion.

Table 1. Classification accuracy by scenario

Scenario	No fusion	Best early fusion	Best late fusion
I	85.19%	89.81% ($N_{sp} = 2$)	**98.15%** (hard, 4 predictions)
II	91.67%	–	**97.22%** (soft, 3 predictions)
III	96.30%	–	96.30% (soft, 3 predictions)
IV	80.95%	84.52% ($N_{sp} = 3$)	**89.29%** (hard, 4 predictions)

The used scenarios and relevant results are illustrated using plots in the upcoming figures. The blue strip is used to mark the ground truth, that is, the true state at the given time in the scenario. Red dots represent models' predictions without any fusion influence. Models' predictions tend to fluctuate frequently when no fusion method is used to compensate. Finally, the golden stars represent fusion's decisions. The stars appear less frequently than the red dots, since fusion algorithms take several predictions from model's to make a single decision. Number of taken predictions is a configurable parameter.

Scenario I, characterized by pronounced interference, benefited most from fusion: sensor fusion via concatenation of two successive spectrograms raised accuracy from 85.19% to 89.81% (Figure 5), while decision-level hard fusion with four predictions using the majority voting rule reached 98.15% (Figure 6).

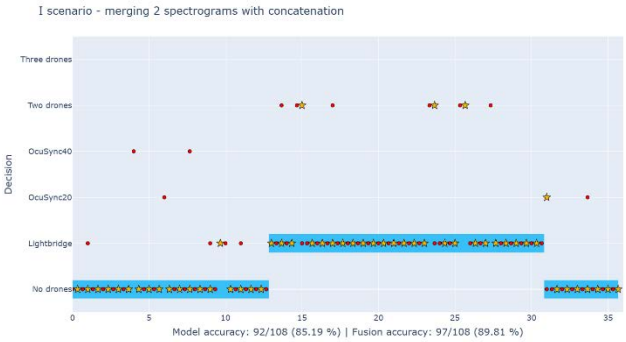


Figure 5. Scenario I – Early fusion

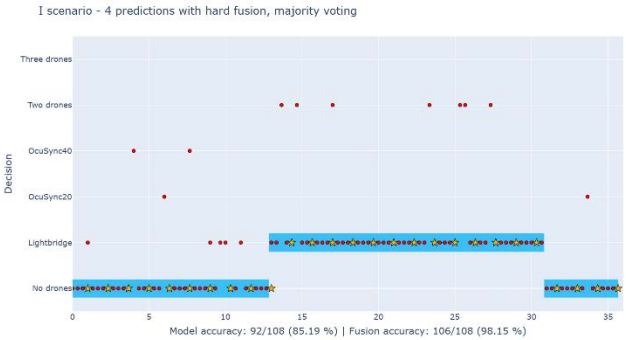


Figure 6. Scenario I – Hard fusion

In the cleaner scenarios II (Figure 7) and III (Figure 8), sensor fusion could not improve the unfused models' accuracy. Soft fusion over three predictions matched or exceeded the unfused baseline in these scenarios (97.22% and 96.30%, respectively) while producing visibly more stable decision sequences than the individual model predictions, which fluctuated at several points despite comparable aggregate accuracy.

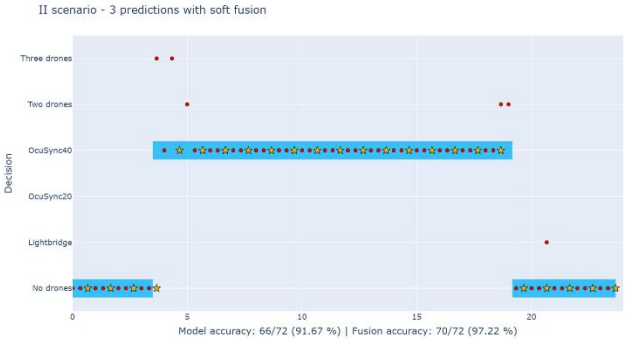


Figure 7. Scenario II – Soft fusion

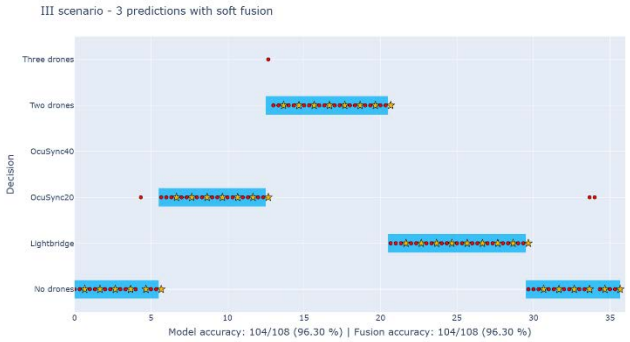


Figure 8. Scenario III – Soft fusion

In the most complex scenario IV (Figures 9 and 10), involving three drones and frequent state changes, hard fusion over four predictions using the longest sequence rule improved accuracy from 80.95% to 89.29%.

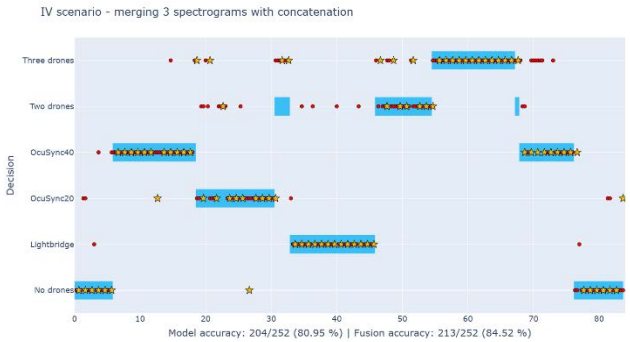


Figure 9. Scenario IV – Early fusion

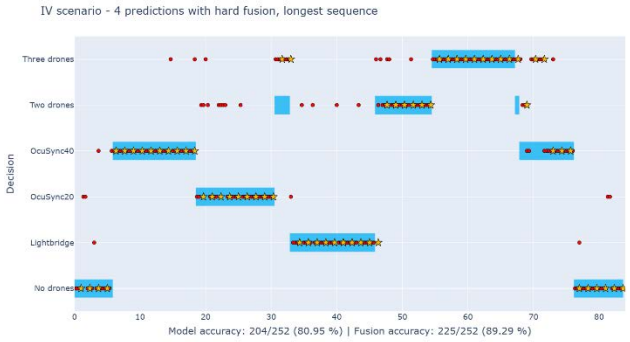


Figure 10. Scenario IV – Hard fusion

Across all scenarios, late fusion consistently outperformed sensor fusion alone, confirming that pairing sensor fusion with a late fusion stage, rather than relying on sensor fusion in isolation, is necessary to obtain the largest accuracy gains. Models used should also have certain amount of accuracy so the fusion algorithms can work out a reliable decision using model's predictions as a baseline.

5. CONCLUSION

This work evaluated sensor and late data fusion for single-receiver RF drone classification with CNN models across four realistic multi-drone scenarios. Sensor fusion alone provided only marginal accuracy gains in the 2,4GHz frequency band. Late fusion, in contrast, improved classification accuracy by up to 13% over the

unfused baseline and produced markedly more stable decisions over time, while also allowing the decision-making frequency to be tuned to the target application.

These findings support time-domain fusion as a lightweight, hardware-free preprocessing layer for improving the precision and reliability of real-time RF-based drone classification. As a result, the predictions that are shown to users in real time became stable and more useful, without fluctuations and excessive output. Future work includes extending the sensor system with additional RF or complementary sensors, fine-tuning the CNN models directly on fused spectrograms, and evaluating further soft and hard fusion strategies.

Acknowledgement

This work was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Contract No. 451-03-34/2026-03/200325).

References

[1] X. Shi, C. Yang, W. Xie, C. Liang, Z. Shi, and J. Chen, "Anti-drone system with multiple surveillance technologies: architecture, implementation, and challenges," *IEEE Communications Magazine*, vol. 56, no. 4, pp. 68–74, 2018, doi: 10.1109/MCOM.2018.1700430.

[2] M. A. Khan, H. Menouar, A. Eldeeb, A. Abu-Dayya, and F. D. Salim, "On the Detection of Unauthorized Drones—Techniques and Future Perspectives: A Review," *IEEE Sens. J.*, vol. 22, no. 12, pp. 11439–11455, Jun. 2022, doi: 10.1109/JSEN.2022.3171293.

[3] H. Zhang, T. Li, Y. Li, J. Li, O. A. Dobre, and Z. Wen, "RF-Based Drone Classification Under Complex Electromagnetic Environments Using Deep Learning," *IEEE Sens. J.*, vol. 23, no. 6, pp. 6099–6108, Mar. 2023, doi: 10.1109/JSEN.2023.3242985.

[4] A. Frid, Y. Ben-Shimol, E. Manor, and S. Greenberg, "Drones Detection Using a Fusion of RF and Acoustic Features and Deep Neural Networks," *Sensors*, vol. 24, no. 8, p. 2427, Apr. 2024, doi: 10.3390/s24082427.

[5] Y. Fu and Z. He, "Radio Frequency Signal-Based Drone Classification with Frequency Domain Gramian Angular Field and Convolutional Neural Network," *Drones*, vol. 8, no. 9, p. 511, Sep. 2024, doi: 10.3390/drones8090511.

[6] B. Sazdić-Jotić, I. Pokrajac, J. Bajčetić, and N. Stefanović, "Review of RF-based drone classification: Techniques, datasets, and challenges," *Vojnotehnicki glasnik*, vol. 72, no. 2, pp. 764–789, Apr. 2024, doi: 10.5937/vojtehg72-49286.

[7] V. Semenyuk, I. Kurmashev, A. Lupidi, D. Alyoshin, L. Kurmasheva, and A. Cantelli-Forti, "Advances in UAV detection: integrating multi-sensor systems and AI for enhanced accuracy and efficiency," *International Journal of Critical Infrastructure Protection*, vol. 49, p. 100744, Jul. 2025, doi: 10.1016/j.ijcip.2025.100744.

[8] U. Seidaliyeva, L. Ilipbayeva, K. Taissariyeva, N. Smailov, and E. T. Matson, "Advances and Challenges in Drone Detection and Classification Techniques: A State-of-the-Art Review," *Sensors*, vol. 24, no. 1, p. 125, Dec. 2023, doi: 10.3390/s24010125.

[9] H. Lee *et al.*, "CNN-Based UAV Detection and Classification Using Sensor Fusion," *IEEE Access*, vol. 11, pp. 68791–68808, 2023, doi: 10.1109/ACCESS.2023.3293124.

[10] S. Verma, O. P. Verma, H. Gupta, T. K. Sharma, and S. Agarwal, "AI-Driven Efficient and Reliable Anti-UAV Monitoring System Using Sensor Data Fusion: A Short Report," in *Lecture Notes in Networks and Systems*, vol. 1343 LNNS, Springer Science and Business Media Deutschland GmbH, 2025, pp. 613–623. doi: 10.1007/978-981-96-5955-5_52.

[11] T. Li, Z. Hong, Q. Cai, L. Yu, Z. Wen, and R. Yang, "BisSiam: Bispectrum Siamese Network Based Contrastive Learning for UAV Anomaly Detection," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 12109–12124, Dec. 2023, doi: 10.1109/TKDE.2021.3118727.

[12] J. E. Ko, S. Kim, J. H. Sul, and S. M. Kim, "Data Reconstruction Methods in Multi-Feature Fusion CNN Model for Enhanced Human Activity Recognition," *Sensors*, vol. 25, no. 4, p. 1184, Feb. 2025, doi: 10.3390/s25041184.

[13] B. Sazdić-Jotić, I. Pokrajac, J. Bajčetić, B. Bondžulić, and D. Obradović, "Single and multiple drones detection and identification using RF based deep learning algorithm," *Expert Syst. Appl.*, vol. 187, Jan. 2022, doi: 10.1016/j.eswa.2021.115928.

[14] C. A. Kumar and O. Giridhar, "UAV Detection Multi-sensor Data Fusion," *Journal of Research in Science and Engineering*, vol. 6, no. 7, pp. 6–12, Jul. 2024, doi: 10.53469/jrse.2024.06(07).02.

[15] M. Aledhari, R. Razzak, R. M. Parizi, and G. Srivastava, "Sensor Fusion for Drone Detection," in *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, IEEE, Apr. 2021, pp. 1–7. doi: 10.1109/VTC2021-Spring51267.2021.9448699.

[16] J. Dudczyk, R. Czyba, and K. Skrzypczyk, "Multi-Sensory Data Fusion in Terms of UAV Detection in 3D Space," *Sensors*, vol. 22, no. 12, p. 4323, Jun. 2022, doi: 10.3390/s22124323.

[17] V. Mehta, F. Dadboud, M. Bolic, and I. Mantegh, "A Deep Learning Approach for Drone Detection and Classification Using Radar and Camera Sensor Fusion," in *2023 IEEE Sensors Applications Symposium (SAS)*, IEEE, Jul. 2023, pp. 01–06. doi: 10.1109/SAS58821.2023.10254123.

[18] Y. Wan, H. Liang, G. Wu, F. Zhou, B. Crispo, and Q. Wu, "Precise RF-Vision Fusion UAV Positioning and Identification for 6G Spectrum Security," *IEEE Journal on Selected Areas in Communications*, vol. 44, pp. 3949–3964, 2026, doi: 10.1109/JSAC.2026.3671229.