A Review of Automatic Classification of Drones Using Radar
Key Considerations, Performance Evaluation and Prospects

Ahmad, Bashar I.; Rogers, Colin; Harman, Stephen; Dale, Holly; Jahangir, Mohammed; Antoniou, Michael; Baker, Chris; Newman, Mike; Fioranelli, Francesco

DOI
10.1109/MAES.2023.3335003

Publication date
2024

Document Version
Final published version

Published in
IEEE Aerospace and Electronic Systems Magazine

DOI. No. 10.1109/MAES.2023.3335003

A Review of Automatic Classification of Drones Using Radar: Key Considerations, Performance Evaluation, and Prospects

Bashar I. Ahmad**, Thales Land and Air Systems, Cambridge CB24 9NG, U.K., and also University of Cambridge, Cambridge, U.K.**
Colin Rogers, **Thales Land and Air Systems, Cambridge CB24 9NG, U.K.**
Stephen Harman, **Thales UK, Reading RG2 6GF, U.K.**
Holly Dale**,** ***Mohammed Jahangir*****,** ***Michael Antoniou*****, and** ***Chris Baker*****, University of Birmingham, Birmingham B15 2TT, U.K.**
Mike Newman**, Thales UK, Reading RG2 6GF, U.K.**
Francesco Fioranelli**, TU Delft, 2600 GA Delft, The Netherlands**

INTRODUCTION

Over the last few years, there has been a significant surge in the use of unmanned air systems (UASs) or aerial vehicles (UAVs), not only in military applications, but also in the civilian domain given the numerous benefits they bring such as to agriculture, e-commerce, filming, inspection-maintenance, to name a few. This is primarily driven by the wide availability of commercial off-the-shelf miniature UASs. They are relatively cheap, can be easily operated and are becoming more sophisticated, capitalizing on advances in sensing systems, wireless communications, automation, and artificial intelligence (AI). However, the potential security and safety threats UAVs pose, for example, to manned aviation, privacy, and sensitive infrastructure or assets, are widely recognized.

Therefore, there is a growing demand for reliable noncooperative drone surveillance for either of the following:

1) *Counter UAS (C-UAS):* Detect and mitigate the unauthorized use of drones by malicious or novice operators such as in exclusion zones around airports [1] or military bases.
2) *Unmanned Air Traffic Management (UTM):* Harness the full potential of UAVs via enabling their safe, widespread, utilization, and integration into the airspace along with manned aviation, for instance, the Single European Sky ATM Research SESAR programme 2004–2020 [2].

Noncooperative C-UAS and UTM solutions often comprise of multiple sensors such as radar, electro-optical cameras, acoustic, and radio frequency (direction finders) to deliver consistent situational awareness in complex and dynamically changing environments [3], [4], for example, surrounding airports. Nevertheless, only radar offers 24-hour, all weather, surveillance at long ranges and for wide areas. In this article, we consider (ground-based) radar which can be part of a multisensor system governed by a suitable concept of operations (CONOPS). For instance, radar cues a high-resolution camera to confirm the identify of a target of interest, such as a drone.

Here, we treat the specific problem of automatic target classification (ATC) or recognition (ATR) of UASs from

Authors' current addresses: Bashar I. Ahmad is with Thales Land and Air Systems, Cambridge CB24 9NG, U.K., and also with the Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, U.K. (e-mail: b.i.ahmad@gmail.com). Colin Rogers is with Thales Land and Air Systems, Cambridge CB24 9NG, U.K. (e-mail: colin.rogers@aveillant.com). Stephen Harman and Mike Newman are with Thales UK, Reading RG2 6GF, U.K. (e-mail: stephen.HARMAN@uk.thalesgroup.com; mike.newman@uk.thalesgroup.com). Holly Dale, Mohammed Jahangir, Michael Antoniou, and Chris Baker are with the Microwave Integrated Systems Laboratory, School of Engineering, University of Birmingham, Birmingham B15 2TT, U.K. (e-mail: had449@student.bham.ac.uk; c.j.baker.1@ bham.ac.uk). Francesco Fioranelli is with TU Delft, Faculty of Electrical Engineering, Mathematics and Computer Science, 2600 GA Delft, The Netherlands (e-mail: f.fioranelli@tudelft.nl).
Manuscript received 25 April 2023; accepted 16 November 2023, and ready for publication 20 November 2023.
Review handled by Michael Brandfass.

Image licensed by Ingram Publishing

radar data, in particular the proliferating sub-50 kg drones, see the NATO taxonomy in Table 1. This encompasses improvised, commercial, and military grade rotary or fixed wing drones. Spanning the "small" to "nano" labels of Class I, the sub-50 kg platforms are thence referred to as miniature UAS (mUASs) or UAVs (mUAVs) for brevity. They pose unique challenges to radar as highlighted in the next section. Tactical and Medium/High Altitude Long Endurance (M/HALE) UAVs can be regarded to resemble traditional targets such as airplanes and jets in terms of radar cross section (RCS), speed, and altitude.

Different classification tasks can be formulated depending on the sought target categories. For example, the objective might be to distinguish drone from non-drone targets, the UAS type (e.g., rotor or fixed wing), size (small, medium, and large), carrying a payload or not and others. In this article and for simplicity, we predominantly focus on the radar ability to automatically discriminate between drone and non-drone objects. ATC enabler, considerations, performance metrics that are relevant to common operational requirements, and other related capabilities (e.g., ATR with drone subclasses, global classifiers, detecting malicious intent, simulators, digital twins and others) are also discussed. Additionally, here we use real measurements from the Thales Gamekeeper radar for illustrations. It is an L-band staring, otherwise known as ubiquitous or holographic [5], radar with 64-element receiver array designed for detecting, tracking in 3D and classifying mUASs within a 7.5 km range, 90° azimuth coverage, and with an ≈ 0.27 s update period in its current configuration.

The remainder of this article is organized as follows. In the section "Drone Surveillance Radar and ATC," we outline the key problems and considerations of classifying mUASs with radar. ATC performance indicators are introduced in the section "Classification Performance Evaluation" and example results are shown in the section "Example Results From Real Data." Opportunities to achieve enhanced (or more detailed) target recognition results are highlighted in the section "Additional Capabilities and Prospects."

DRONE SURVEILLANCE RADAR AND ATC

WHY ARE DRONES DIFFICULT TARGETS FOR RADAR?

Class I (miniature) drones are particularly challenging targets to detect and track with radar because they can simultaneously have all (or most) of the following attributes:

- *Small (low observable):* mUASs can have low RCSs, which can be $\ll 0.01\ \mathrm{m}^2$ as with nano or micro drones, and discriminative features of their radar

Table 1.

NATO Drone Platforms Designations and Taxonomy

UAS Class	Maximum Take-Off Weight (kg)	Label
I(a)	< 0.2	Nano
I(b)	$0.2 - 2$	Micro
I(c)	$2 - 20$	Mini
I(d)	$20 - 150$	Small
II	$150 - 600$	Tactical
III	> 600	M/HALE

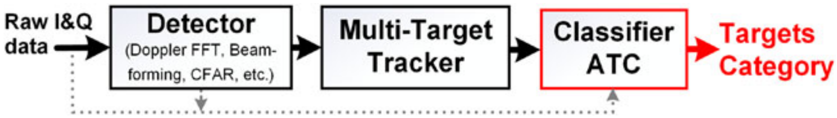

Figure 1.
Standard processing chain of a drone surveillance radar.

signatures (e.g., micro-Doppler components) are a further one to two orders lower [6]. This necessitates the sensor having a high sensitivity and thereby detecting a potentially large number of targets of similarly low RCSs such as birds.

- *Slow:*Drones can have markedly low speeds, for example, less than 10 m/s which renders them virtually undetectable to conventional radar systems (e.g., primary airport radars); rotary-wing mUAVs can also hover. Consequently, their body return and/or any distinguishable features in their radar signatures (e.g., from their on-board rotors, if any) can be easily obfuscated by stationary or slow-moving clutter within the same resolution (e.g., range, azimuth, and elevation) cell making them difficult to consistently detect and track.
- *Low:* mUASs can fly at low altitudes (e.g., between 30 to 150 m above ground and remain not easily visible to the naked eye). At such heights, especially in urban-industrial environments, we can have: a) occlusions to the radar line-of-sight from terrain or buildings and electromagnetic RF interference; b) multipath effects, impacting targets height measurements; c) interference from ground targets (e.g., cars) and man-made equipment with moving or rotating parts such as air-conditioning units, generators or roof fans; and d) the potential presence of large number of birds of various types and sizes.
- *Agile:* While (semi-)autonomous mUAVs often follow optimized smooth paths (e.g., between predefined waypoints), they can be highly maneuverable, and can undertake sharp maneuvers such as abrupt turns and accelerations. Manually operated drones, for example racing or first person view mUAVs, can fly erratically with frequent maneuvers. The radar multitarget tracker (MTT) has to be able to handle a wide range of kinematic behaviors. This is generally constrained by the tracker employing motion models with fixed, fine-tuned, parameters for describing the expected level of variability in the targets movements. This encompasses models specifically developed for manoeuvring targets [7]. Alternative techniques, such as interacting multiple model [8] and adapted MTT [9], can be either hard to correctly configure or prohibitively computationally demanding when simultaneously tracking large number of objects, majority of which are non-drone targets such as cars, pedestrians, birds, etc.

A drone surveillance radar processing chain normally consists of the standard three sequential operations for: 1) detection, 2) tracking and 3) classification of targets within the field of coverage. Each is carried out within a separate software-firmware module, which can share information. An example block diagram is shown in Figure 1. Merging two or more of these tasks is known to substantially improve the radar performance against mUASs, for instance track-before-detect [10], joint tracking-classification [11], and even recognize-before-detect. The latter however regularly refers to a rule-based filtering of detections to prevent overloading the multitarget tracker or the operator. Next, we focus on ATC.

AUTOMATIC DISCRIMINATION: DRONE VERSUS CONFUSER TARGETS

Several radar systems, including in multistatic configurations, have emerged to address the formidable challenges presented by mUASs [3], [5], [12]. Within a relatively large field of coverage (e.g., spanning a few kilometers in range), they have to contend with a large number of potential *confuser* targets, such as birds, which cohabit the same aerospace and exhibit similar characteristics to mUASs, for instance their RCSs, altitude, and speeds. A reliable automatic target classifier is thus fundamentally important in drone surveillance radars to distinguish between mUASs, which are usually rare, and confuser targets (e.g., birds), which can be abundant in (semi-) rural or urban environments. The surveillance system resources (e.g., secondary optical sensors) and/or operator attention should be dedicated to scrutinizing targets that can pose a threat (e.g., mUAVs). Otherwise, they can be overwhelmed by the large number of tracked targets. ATC is also crucial for automation to reduce the overall C-UAS/UTM system operating cost by circumventing human-intensive CONOPS.

To demonstrate the sheer number of targets drone surveillance radars often handle, in Table 2 we list the average numbers of trajectories formed on targets per hour by the Thales Gamekeeper sensor within a range of 7.5 km; 90° azimuth coverage. They are attained from 24-hour

Table 2.

Average Number of Tracks Per Hour From a Staring Radar, Within 7.5 Km Range and 90° Azimuth at Different Sites		
Site	Description	Average Number of Tracks (per hour)
A	Dense urban environment with the radar overlooking a major U.K. city	6803.2
B	Mixed semiurban and semirural environments with small-medium villages/towns distributed in coverage	9204.5
C	Semirural environment with major roads in coverage	9863.3
D	National airport within a semirural environment and surrounded by small-medium villages/towns	6532.3
E	International airport with mixture of dense urban and industrial areas within coverage	9394.4
F	Mixed urban and semirural environment with major roads and villages/towns within coverage	11183.9

continuous recordings at various sites in the U.K. and France. From the table, it can be seen that several thousand tracks per hour are processed on average by the radar; this can exceed 15,000 per hour at certain times in mixed urban and semirural areas. On average several hundred tracks can be considered at any point in time. Figure 2 displays all the trajectories reported by the radar MTT during a 30 min period at semiurban/rural environment (Site B). Although no ground-truth is available for all targets within the radar coverage, analysis of the characteristics of the tracks in Table 2 and Figure 2 (e.g., height, location, speed, etc.) confirm that they are not spurious. Accordingly, ATC module has to correctly classify the vast majority of tracks as non-drone and avoid triggering false alarms (FAs) whose prominence is emphasized in the section "Classification Performance Evaluation."

CLASSIFICATION ALGORITHMS AND ENABLERS

Automatic classification or recognition of noncooperative targets with radar, including UASs, is a well-established research field with a plethora of existing techniques [13]. Conventional approaches are commonly based on hand-crafted rules applied to selected target features (e.g., RCS, height, velocity, etc.) and/or employ classical spectral analysis tools (e.g., cepstrum). Recent ATC algorithms on the other hand are increasingly data-driven and leverage advances in machine learning (ML), such as deep neural networks (DNNs), to achieve impressive classification performance, see [14], [15], [16], [17], [18], [19], [20], [21].

While progress is being made to better understand the behavior of DNNs [22], the main difficulty in developing

Figure 2.
Trajectories of tracked targets within a range of 7.5 km at Site B during a typical 30 minute radar recording.

generalizable machine-learned classifiers, usually within a supervised learning framework, is the availability of extensive and sufficiently representative labeled training datasets. This is due to the great diversity of potential targets to be recognized relative to the available real radar measurements. For example, the wide range of possible mUAS sizes, designs, speeds, heights, trajectory profiles-maneuvers, rolls-pitch-yaw angles with respect to the radar (i.e., incident angles), rotor speeds which may depend on the ambient wind, clutter characteristics, etc. Obtaining ground-truth of bird targets can involve further complications, for example, due to the difficulties and cost of conducting controlled trials with birds instrumented with transceivers or employing a specialized targets labeling solution with electro-optical sensors, secondary radars, etc. Using synthetic data (e.g., to complement the available real sensor data, see the section "Simulators and Digital Twins") can be critical to mitigate overfitting effects and ensure that data-driven classifiers deliver robust performance when applied under real operational conditions (e.g., low SNR and previously unseen drone data).

Given the drastic measures that might need to be taken when an unauthorized or malicious mUAS is declared (e.g., closure of the airspace near civilian airports that can severely disrupt the aviation traffic), it can be highly desired for the radar ATR module to report the certainty level in its classifications/predictions such as confidence scores for all considered target categories. For instance, a DNN micro-Doppler classifier can have a *softmax* output layer to produce these confidence scores, in lieu of the target final label [14], [18], [19], [20]. This permits the multisensor counter UAS or UTM system to not only adopt different risk management strategies for different targets, but also a more informed data fusion and CONOPS.

Drone classification often relies on target discriminative features that can be grouped into three categories: micro-Doppler, kinematics, and long-term behavior. They are described next along with their limitations.

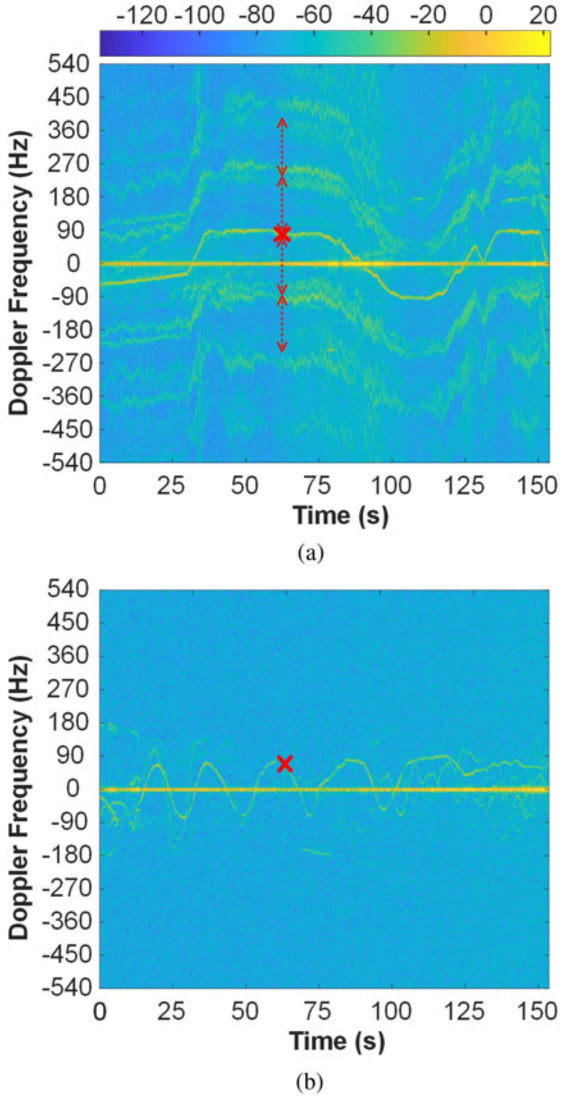

Figure 3.
Doppler spectrograms from real radar data of a quadcopter drone (a) and bird (b) at ranges $\approx 1.5 - 2$ km. Red crosses mark the target body Doppler frequency at one time step; arrows indicate the UAS micro-Doppler returns.

MICRO-DOPPLER SIGNATURES

The motion of rotors or propellers on-board a mUAS produce spectral lines in the radar Doppler spectrum, with approximate harmonic structure around the target body Doppler frequency. These are dubbed micro-Doppler components and their characteristics depend on the number of blades, blade length, frequency of rotation, radar wavelength, and respective incident angle [23]. Example Doppler spectrograms from a bird and DJI Inspire 2 quadcopter drone are depicted in Figure 3, where both targets are approximately $1.5 - 2$km from the L-band Gamekeeper radar. Unlike the bird, the mUAV Doppler spectrogram in Figure 3(a) has visible micro-Doppler components symmetrically distributed around the target body. Nevertheless, bird wings motion can result in intricate micro-Doppler-type spectral features for appropriately short radar wavelengths [24]; they are notably distinctive from those originating from mUAS blades rotating up to several thousand times a minute (i.e., drone's propellers move at a very different speed to a bird's wings). On detecting micro-Doppler signatures for ATC, convolutional neural networks (CNNs) have shown great promise [14], [15], [16], [18], [19], [21]. Their input can be Doppler spectrograms from multiple radar frames/scans (e.g., magnitude spectrogram as in Figure 3), complex time series data or covariance matrices. The former can be treated as an image and micro-Doppler harmonic structure becomes the sought pattern. Popular and specialized CNN architectures such as GoogLeNet, AlexNet, CRNet, SPDNet or an aggregation of various models can be utilized. Recurrent neural networks and other hybrid DNNs have also proven effective for micro-Doppler-based drone recognition [14]. While micro-Doppler is a strong classification cue, especially against prevalent confuser targets such as birds, rotors radar signatures can be -15 to -20 dB lower than that of the mUAS body. The corresponding signal-to-noise (SNR) ratios decay rapidly with range and can be reduced further due to the blades characteristics (e.g., small size, fairing and constructions from a low reflective material such as carbon or plastic) as well as deliberate concealment (e.g., with blade guards). Consequently, the dectectability of micro-Doppler from drone propellers can be restricted to specific scenarios such as mUASs at short ranges and/or with certain rotor blades physical properties. Furthermore, there are several (false positive) micro-Doppler sources in industrial/urban settings such as air-conditioning units, generators or roof fans, etc.

TARGET BODY KINEMATICS AND CHARACTERISTICS

Features extracted from the kinematic movements and characteristics of the target main body can facilitate distinguishing between drone and non-drone targets, for instance velocity, acceleration, jerk, height above ground, 2D/3D trajectory curvature, torsion, body Doppler stability and spread over time, RCS and others [16], [17]. They are typically derived from the MTT output or even associated detections/plots; hence their quality is dependent on the detection-tracking accuracy. The classifier employs a statistical measure of each kinematic feature (e.g., mean, L-moments, median, standard deviation, smoothness metric, quantiles, etc.) computed from the time series of its consecutive instantaneous values for a given trajectory. Contrary to micro-Doppler signature, the range of possible values of the kinematic features for the mUAS and bird targets can largely overlap. Thereby, relying on them alone can lead to a relatively low classification performance.

LONG-TERM BEHAVIOR AND PATTERNS OF LIFE

Drones can follow distinctive trajectory shapes dictated by a mission planning software to optimize use of resources (e.g., time the platform is airborne under limited battery life) such as waypoints-driven paths, hippodromes, hexagon, and others [25]. These are generally not characteristic of birds behavior. Conversely, frequent swooping maneuvers are more likely to be displayed by birds. Revealing such distinctive long-term kinematic patterns can allow identifying mUASs targets. They are however not always present and demand a persistent tracking of low observable and agile targets over extended durations, which is difficult to maintain at long ranges and in high clutter-noise environments. Additionally, a substantial delay is incurred before a distinguishable motion pattern (if any) materializes and this degrades the ATC timeliness which we discuss in the section "Timeliness: Classification Time Delay (CTD)." For some target types (e.g., cars or birds), revealing high activity areas and times (e.g., on major roads) can be salient pattern-of-life information that can be exploited by the classifier; this can be viewed as contextual data.

Subsequently, it is imperative that combining micro-Doppler, kinematic and long-term behavior features (if available) can boost the ATC overall performance [16], [17].

SUMMARY OF ATC CONSIDERATIONS

Linked to the challenges of detecting/tracking mUASs in the section "Why are Drones Difficult Targets for Radar?," in summary the major considerations for formulating drone surveillance radar ATC approaches are:

- *Large number of confuser targets* that can potentially trigger FAs; this can render stringent FA specifications (e.g., a maximum of one FA every 24 h or even every several days in urban-rural settings) unachievable in practice with radar alone.

- *Fleeting discriminative target features*, such as micro-Doppler from a drone's rotors, that are intermittently observed due diversity in target type, flight profiles, behaviors, clutter, multipath, respective incident angle, etc. This can lead to fluctuations in the classification results over time. For illustration, an example of confidence scores in the UAS class from three ML classifiers are depicted in Figure 4. This is from real Gamekeeper radar data of the DJI Inspire 2 drone, whose Doppler spectrogram is shown in Figure 3(a). The ATC methods are: **i)** multistage with machine-learned decision tree (DT) [17] trained on real sensor data and using micro-Doppler as well as kinematic features at each radar frame with a Simple Moving Average (SMA) at its output, **ii)** AlexNet-based CNN in [18] with nonoverlapping Doppler spectrograms of length ≈ 5.5s (updates every ≈ 5.5s) and trained on real radar data, **iii)** low-latency simple CNN model [20], trained exclusively on synthetic data to use as its input the Doppler spectrum from one update/time-step. Both multistage and low-latency CNN classifiers update every ≈ 0.27s. The noticeable changes in the classification results over time is visible in Figure 4 for the three classifiers which all perform reasonably well in terms of declaring this tracked target a UAV (e.g., with a 50% decision threshold). This variability in their outputs can be attributed to

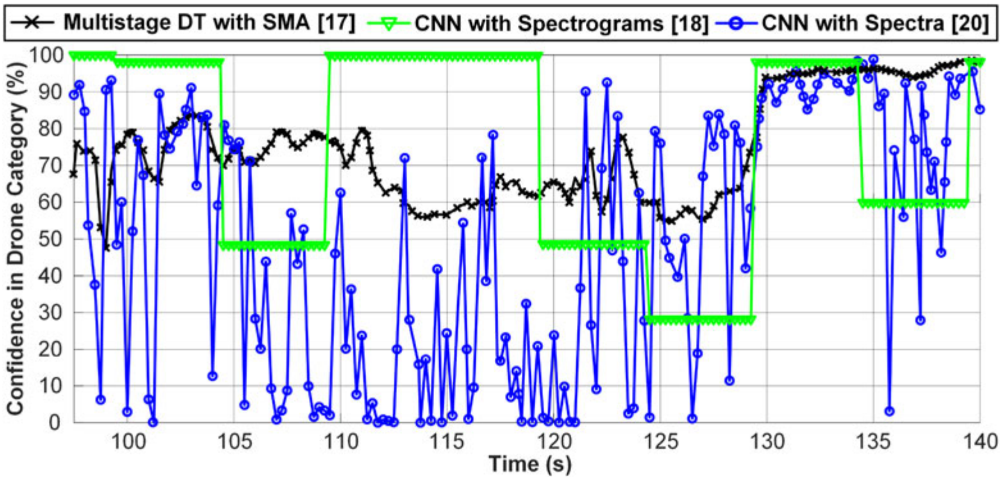

Figure 4.
Confidence scores in UAS category from three classifiers for the quadcopter whose spectrogram is shown in Fig. 3(a); decision tree (DT), simple moving average (SMA), and convolutional neural network (CNN).

changes in the radar measurements quality over time: see spectrogram in Figure 3(a).

- *Limited available data and generalizability of the classifier*, especially for data-driven (machine-learned) ATC algorithms expected to tackle previously unseen drone types or targets whose data are not in the training datasets.
- *Capturing classification certainty* to enable multi-sensor C-UAS or UTM solutions to apply effective CONOPS at the command and control (C2) system level.
- *Computational efficiency and swarms*, to process large number of targets with a track-level classifier. An initial coarse plot-level classification is sometimes used by radars for pruning detections fed to the tracker. This can be exacerbated by drones ability to fly in swarms, where 3281 is the Guinness world record for most UAVs airborne simultaneously [26]. With large number of confuser targets, a much smaller swarm could risk overwhelming the radar sensor processing chain.

CLASSIFICATION PERFORMANCE EVALUATION

There are widely used detection and multitarget tracking key performance indicators (KPIs) for noncooperative surveillance systems, for example, the single integrated air picture (SIAP) [27]. On the other hand, radar ATC efficacy is often assessed in terms of standard object recognition metrics from the ML field, such as accuracy, true positive (TP) and false positive (FP) rates, confusion matrix, F_1 score, and receiver operating characteristic (ROC), if relevant. While these are informative, especially for developing and refining the classification algorithm, in this section we revisit the definitions of some of the traditional KPIs and propose additional metrics that are important to the end-users in various C-UAS/UTM applications.

EVALUATION FRAMEWORK

For simplicity, we note that the classification KPIs are presented here for a binary discrimination problem, namely miniature drone/truth versus non-drone/non-truth (i.e., birds, cars, pedestrians, etc.) targets. This also considers a framework underlined by the following essential stipulations from the ATC considerations outlined in the section "Summary of ATC Considerations":

- *ATC specific metrics are exclusively studied*; they are disentangled from issues that arise from the detection and tracking steps which have own KPIs (e.g., SIAP). For instance, true positive classifications are obtained from time steps where a track is formed and associated with the truth/drone target. This permits a more objective examination of the classifier behavior and benchmarking its performance. This is despite the dependence of the ATC on the detection-tracking results.
- *Classification confidence and hard positives/negatives are reported*. The former, whose values can range from 0 to 1 (or 0 to 100%), is the certainty level in the classifier predictions for each of the nominal target categories. A decision rule is then applied to determine the target category, for instance the most probable class, i.e., maximum *a Posteriori* (MAP), or the confidence score in the UAS class exceeding a threshold value. Thus, the notion of "*Hard*" TP (HTP) and FP (HFP) is

introduced to signify their post-decision nature. Confidence scores can be attained from a time average (e.g., SMA) or other method for combining the classifier results over time.

- *All metrics are calculated per radar recording/dataset* (e.g., during a live drone trial) and combined measures, such as average, across multiple ones can be obtained. An example is the 30 min recording in Figure 2. This ensures that a poor performance at a specific site or time (e.g., due to high number of confuser targets) is not overlooked, especially in relation to FAs, see Table 2 for the variability between locations. Even at one site, the surveyed scene characteristics can substantially change with time. Conversely, *traditional* ATC performance assessments (e.g., with accuracy and confusion matrix) are usually computed from all of the available test data (i.e., from the aggregate of all datasets).
- *FPs from individual non-truth tracks in a recording are first calculated and subsequently combined* (e.g., via average, standard deviation, etc.) to represent FP metrics. This facilitates defining KPIs such as timeliness and FAs. Truth information is regularly unavailable for non-drone targets such as birds. This is unlike TP metrics pertaining to cooperative UAS targets during flight trials (e.g., from on-board GPS). True positive KPIs are attained from data of all tracks associated with the truth, within the corresponding radar recording.

For context, consider a fixed site protection scenario, where a C-UAS radar at a civilian airport is tasked with reporting the presence of mUASs in the vicinity of (or within) the restricted flying zone covering the protected aerodrome [1]. Notation of the discussed metrics are listed in Table 3. We make one further assumption here to streamline the notation. At any kth time step (i.e., radar frame or update at time instant t_k), only one track can associate with the jth truth/drone target. This can be easily generalized to multiple associations at t_k by appropriately defining the set $\mathbb{A}_j$ in Table 3.

TP CONFIDENCE, DEVIATIONS, AND HARD TRUE POSITIVES

The confidence $c^+_{j,k}$ estimated by the classifier is the one for the UAS/truth category when a track plot at time instant t_k (i.e., kth radar frame or update interval) is associated with the jth drone/truth target. Its values can range from 0 to 1 (or 0 to 100%) and represent the ATC certainty level in the target being a UAV. A decision on the target class can then be taken leading to the Hard TP (HTP) $C^+_{j,k} \in \{0, 1\}$ at t_k. This can be based on $c^+_{j,k}$ being larger than the value for all other target classes, or its own value exceeding a certain threshold γ, for instance, $C^+_{j,k} = 1$ if $c^+_{j,k} > \gamma = 0.9$ and zero otherwise.

The mean TP Confidence (TPC) for the jth UAS/truth is

$$\mathrm{TPC}_j = \frac{\sum_{k \in \mathbb{A}_j} c^+_{j,k}}{N^+_j} \tag{1}$$

from the set $\mathbb{A}_j$ of all time steps where a track-plot is associated with the jth truth and $N^+_j = |\mathbb{A}_j|$. The average across all J drone targets in the radar recording (e.g., during a live drone trial) is given by: $\mathrm{TPC} = \sum_j \mathrm{TPC}_j / J$. An increase in TPC implies a better ability to recognize drone targets and with higher confidence. A related metric that can measure the consistency of the TP confidences is the standard deviation

$$\mathrm{DevTPC}_j = \sqrt{\sum_{k \in \mathbb{A}_j} \left(c^+_{j,k} - \mathrm{TPC}_j\right)^2 / N^+_j} \tag{2}$$

with average $\mathrm{DevTPC} = \sum_j \mathrm{DevTPC}_j / J$. A decrease in its value signifies an improvement in the ability of the classifier to recognize drones over a period of time.

Another metric to capture the variability in the TPC can be explored. Hard True Positive Classification Probability (HTPCP) for the $j = 1, 2, \ldots, J$ drone/truth targets is

$$\begin{aligned} \mathrm{HTPCP}_j &= \sum_{k \in \mathbb{A}_j} C^+_{j,k} / N_j \\ \mathrm{HTPCP} &= \sum_j \mathrm{HTPCP}_j / J \end{aligned} \tag{3}$$

where HTPCP is the *traditional* true positive or recall KPI.

UAS DECLARATION PROBABILITY (UDP)

The likelihood that the jth UAS/truth target is correctly classified as a drone for at least one time step. This is termed UAS target declaration probability (UDP) and is given by

$$\mathrm{UDP}_j = \begin{cases} 1, & \text{if } C^+_{j,k} = 1 \text{ for any } k \in \mathbb{A}_j \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

and $\mathrm{UDP} = \sum_j \mathrm{UDP}_j / J$ for $j = 1, 2, \ldots J$. In other words, if a drone is within the radar system field of view and it has been detected-tracked, this is the likelihood of it being identified as a UAS. Higher UDP indicates a higher ATC reliability; it should be considered with the timeliness metric detailed next. A more demanding UDP rule can be adopted such as HTPs are required to be maintained for a fixed duration.

Table 3.

Notation for the Proposed Key Performance Metrics	
T	**Duration (in Hours) of the Processed Radar Recording**
J	Total number of detected-tracked drone/truth targets
N_T	Total number of formed tracks with unique IDs in the radar recording
$\mathbb{A}_j$	Set of time steps (i.e., radar updates/frames), where there is a track-plot associated with the jth truth/drone target
N_j^+	Total number of time steps in set $\mathbb{A}_j$ and $N_j^+ = \lvert\mathbb{A}_j\rvert$
$\mathbb{T}_{j,i}$	Set of time steps where the ith track is associated with a UAS
$c_{j,k}^+$	Confidence $c_{j,k}^+ \in [0,1]$ in the truth/UAS class at the kth step (at time instant t_k) when a track is associated with the jth drone target
$C_{j,k}^+$	Hard true positive, $C_{j,k}^+ \in \{0,1\}$, that the target is a UAS/truth when a track plot at t_k is associated with the jth UAS/truth target
γ	Threshold value of a hard TP, $C_{j,k}^+ = 1$ if $c_{j,k}^+ > \gamma$ and 0 otherwise
$\mathbb{H}_j^+$	Set of tracks (of unique IDs) that are associated with the jth drone/truth with $C_{j,k}^+ = 1$ for at least one time step
$t_{j,i}^0$	Time instant (in seconds) of the start of the ith track, which associates with the jth truth/drone target for at least one time step
$\tau_{j,i}^+$	Time instant (in seconds) of the first/earliest HTP declared for the ith track associated with the jth drone/truth target
I	Total number of tracks (with unique IDs) not associated with any drone/truth target
$c_{i,k}^-$	Confidence $c_{i,k}^- \in [0,1]$ in the UAS/truth category for the ith track plot at t_k not associated with the drone/truth target
$C_{i,k}^-$	Hard false positive, $C_{i,k}^- \in \{0,1\}$, that the target is a UAS/truth at t_k for the ith track plot not associated with the drone/truth target
$\hat{\gamma}$	Threshold value for a hard FP, $C_{i,k}^- = 1$ if $c_{i,k}^- > \hat{\gamma}$ and 0 otherwise
$\mathbb{B}_i$	Set of time steps (i.e., radar update intervals) where ith track is not associated with any of the truth/ drone targets
N_i^-	Total number of time steps in set $\mathbb{B}_i$ and $N_i^- = \lvert\mathbb{B}_i\rvert$
M_u	Number of successive hard FP for a track not associated with a drone/truth track that produces a unique FA as per (9)

TIMELINESS: CLASSIFICATION TIME DELAY (CTD)

This conveys the delay incurred prior to successfully declaring, for the first time, a detected-tracked UAS/truth target as a drone. This is an ATC timeliness measure. For the ith track which starts at time instant $t_{j,i}^0$ and associates with the jth drone/truth target, its CTD (in seconds) is

$$\mathrm{CTD}_{j,i} = \tau_{j,i}^+ - t_{j,i}^0 \tag{5}$$

where $\tau_{j,i}^+ = \min\{t_k : C_{j,k}^+ = 1, k \in \mathbb{T}_{j,i}\}$ is the first time step this track is declared as a drone target. A demonstration is shown in Figure 5, where we have a track break caused by a temporary loss of the target (e.g., due to a sharp manoeuvre, loss of detections of the low observable target, a high clutter region, etc.). Time instants $t_{j,i}^0$ and $\tau_{j,i}^+$ for the two tracks formed on the drone are marked by arrows. Track 1 and 2 last for five and four time steps (or updates), respectively. In this case and for a radar update period of 0.27 s, $\mathrm{CTD}_{j,1} = 0.27\mathrm{s}$ and $\mathrm{CTD}_{j,2} = 0.54\mathrm{s}$; $\mathrm{HTPCP} = 5/9$, and $\mathrm{UDP} = 1$.

The average and minimum CTD from all tracks (with unique IDs) that are associated with the jth UAS/truth target and each have at least one HTP, i.e., set $\mathbb{H}_j^+$, are

$$\mathrm{CTD}_j = \sum_{i\in\mathbb{H}_j^+} \mathrm{CTD}_{j,i}/\lvert\mathbb{H}_j^+\rvert$$
$$\mathrm{CTD}_{j,\min} = \min\{\mathrm{CTD}_{j,i}, i \in \mathbb{H}_j^+\}. \tag{6}$$

Averages can be computed for all J targets as per $\mathrm{CTD} = \sum_j \mathrm{CTD}_j/J$ and $\mathrm{CTD}_{\min} = \sum_j \mathrm{CTD}_{j,\min}/J$. If one

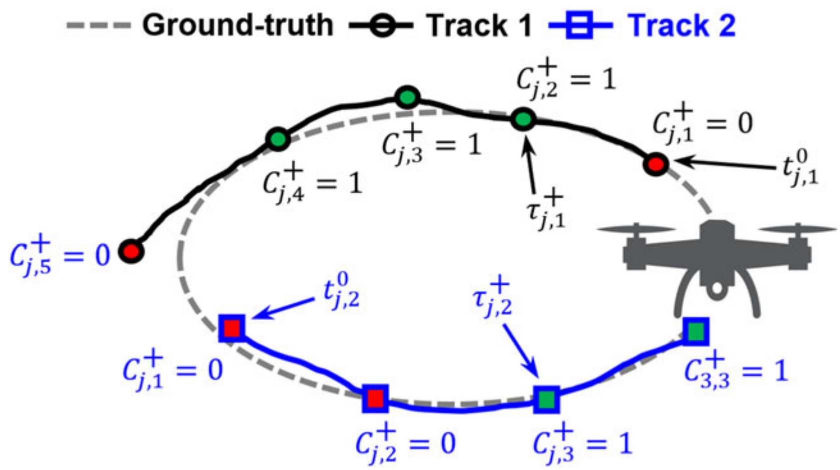

Figure 5.
Example of a UAS trajectory with two formed tracks on a mUAS due to a temporary loss of the target (i.e., a track break/death, for instance induced by the target going through zero-Doppler and/or undertaking a sharp maneuver). Hard TPs are displayed (green is for 1 and red for 0). Time instants of tracks start and first declared hard true positive are indicated by arrows.

continuous track is formed on the UAS, i.e., no breaks, then $|\mathbb{H}_j^+| = 1$.

ATC timeliness can be critical where a minimal CTD is often sought. Early recognition of the mUAS provides C-UAS system operators with sufficient time to take necessary action to address any threat the drone might present, for example, air traffic management (ATM) can divert flights. If a new trajectory with a unique ID is created by the MTT for the same drone target (e.g., following a track break), it will be treated as a new track that may require scrutiny/interrogation by the surveillance system. Consequently, the average CTD can be a more suitable timeliness metric with a radar MTT that has a high rate of track number changes (R) and/or low longest track segment (both are SIAP measures of MTT continuity). CTD can be in reference to the time instant $tD0_j$ the mUAS was detectable in principle in lieu of $t_{j,i}^0$ in (5). This mixes detection and classification metrics and it is usually difficult to specify $tD0_j$ in complex environments.

FP CONFIDENCE, DEVIATIONS, AND HARD FP

For the ith track, which is not associated with a drone/truth target during the time steps in set $\mathbb{B}_i$ and $N_i^- = |\mathbb{B}_i|$,

$$\mathrm{FPC}_i = \frac{\sum_{k\in\mathbb{B}_i} c_{i,k}^-}{N_i^-},$$

$$\mathrm{DevFPC}_i = \sqrt{\frac{\sum_{k\in\mathbb{B}_i}\left(c_{i,k}^- - \mathrm{FPC}_i\right)^2}{N_i^-}} \tag{7}$$

are the mean false positive confidence (FPC) and its deviations, respectively. Hard false positive (HFP) follows from a decision scheme with $C_{i,k}^- \in \{0,1\}$ at t_k, for instance, $C_{i,k}^- = 1$ if $c_{i,k}^- > \hat{\gamma}$ and 0 otherwise. While average from all non-truth trajectories can be obtained, reporting the L (e.g., $L = 10$) tracks with the highest mean FPC can be highly beneficial to understanding the ATC false positive behavior.

A decrease in the mean FPC implies an improvement in the ability to classify a non-drone as a non-drone, hence potentially reducing hard FPs and FAs (see the section "Summary of ATC Considerations"). Lower DevFPC (for selected tracks or average across all non-drone tracks) suggests a more consistent classification of non-drone targets. The *traditional* (hard) false positive metric, referred to here by false positive classification probability (HFPCP), within the studied radar recording is

$$\mathrm{HFPCP} = \sum_i \sum_{k\in\mathbb{B}_i} C_{i,k}^-/N_i^-.$$

FALSE ALARM RATE

This is the number of tracks that trigger a FA and can require the operator and/or secondary sensor attention (e.g., camera to confirm the target class). We measure this as a rate, per hour and/or per R_{Track} tracks. FAs are based on hard false positives, i.e., post deciding the target class (e.g., with a thresholding or MAP criterion). The ith track triggers a unique FA (UFA), i.e., $\mathrm{UFA}_i = 1$, if the ATC makes a M_U successive hard false classifications for this non-drone trajectory, at least in one occasion such that $M_U \geq 1$ is an integer. We can formulate this for the ith track not associated with the UAS target at the $k \in \mathbb{B}_i$ time steps as follows:

$$\mathrm{UFA}_{i,k} = \begin{cases} 1, & \text{if } k\in\mathbb{B}_i \text{ and } \sum_{\min(k,k-M_U+1)}^{k} C_{i,k}^- \geq M_U \\ 0, & \text{otherwise} \end{cases}$$

$$\mathrm{UFA}_i = \begin{cases} 1, & \text{if } \sum_k \mathrm{UFA}_{i,k} > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

It is thus unique false alarms since one track can cause at most one FA.

The FA rate (FAR) per hour is

$$\text{FAR} = \sum_{i=1}^{I} \text{UFA}_i / T$$

and FAR ratio (FARR) per R_{Track} tracks (e.g., $R_{\text{Track}} = 100$)

$$\text{FARR} = R_{\text{Track}} \sum_{i=1}^{I} \text{UFA}_i / N_T.$$

These two performance metric determine: a) the user confidence in the radar classifications since FAs can be extremely distracting for operators; b) the load on the C-UAS solution secondary sensors (e.g., cameras, RF direction finders, etc.); and c) potential for automation to reduce human-intensive CONOPS. Minimizing the FA rate is among the main challenges for long-range drone surveillance radars [14], [17], [20]. It is noted that less strict UFA_i scheme can be employed in (9), e.g., M_U non-successive hard false positives.

STANDARD METRICS AND OTHERS

The standard recall (i.e., HTPCP) in (3) and FP (i.e., HFPCP) in (8), other standard metrics such as accuracy, F_1 score, confusion matrix, and ROC (e.g., for any of the TP versus FP measures to ascertain a suitable decision threshold) are well-understood and can be utilized to evaluate the ATC performance. The mUAS classification problem is however markedly unbalanced, with typically far more non-drone data compared to drone due to confuser targets (e.g., birds). It can be more reasonable to apply different weights to HTPs and HFPs when calculating any measure that mixes them, thereby "*weighted*" accuracy, F_1 score, etc. Otherwise, the true negative and FP data points can dominate the outcome. For this reason, we introduced in this article additional metrics.

Examining ATC results with accuracy and/or F_1 scores alone can hide a high FA problem. This is because only if a small "percent" (e.g., 0.1%) of the large number of non-drone tracks (e.g., 8,000 per hour, see Table 2) can lead to an excessive FAR (per hour) that can deluge the C-UAS system and/or operators. This can occur even if the accuracy and F_1 scores are satisfactory high (e.g., exceeding 95%).

We explored SIAP inspired ATC metrics [28] such as classification spuriousness (i.e., ratio of extant HFPs to all hard positive classifications), continuity (i.e., maintaining correct UAS identification over time), and ambiguity (i.e., over reporting drone presence). They were viewed to rank lower in terms of relevance to requirements of noncooperative drone surveillance systems compared with those detailed above.

SENSITIVITY TO SNR

Classifiers efficacy is generally sensitive to the SNR pertaining to the drone target, especially in terms of the detectability of its micro-Doppler signatures [14], [19], [21]. For example, CNNs' impressive classification accuracy drastically degrades as SNR decreases [21]. A lower SNR can be due to one or more of the following reasons: 1) longer target range, 2) smaller UAS platform (i.e., RCS), and 3) complex environment with higher background noise from clutter and interference.

Therefore, it is paramount to quantify the ATC sensitivity to SNR. This can be achieved by plotting the true-positives-related classification metrics (e.g., HTPCP, TPC, accuracy, and F_1 score) versus the estimated SNR from real radar data (see next section) and/or from synthetically injected noise as in [21]. This ensures a better assessment of the radar "classification range" against different mUAS types/sizes and resilience to environmental factors. Maximizing ranges at which miniature drones can be recognized is vital to implementing effective system level CONOPS and threat mitigation protocols. For instance, a mUAS moving at a speed of 15 m/s toward the airport glide slope and classifying it at 1.8 km away from this prohibited region gives ATM operators 2 min to warn civilian airplanes landing and taking off; this can be increased to 4 min if the ATC range is extended to 3.6 km.

EXAMPLE RESULTS[1] FROM REAL DATA

To demonstrate the ATC metrics in the previous section, we use real measurements from the Gamekeeper staring radar. They were collected during 25 live drone trials (i.e., radar recordings) at various sites, including those in Table 2. Each recording is of a duration of 5 to 16 min and can have up to 6300 target tracks with unique IDs. It has one UAS target (i.e., $J = 1$), whose ground truth information from onboard GPS is available. The overall test data are $\approx$ 4-hour of radar measurements in total and comprises of over 55,000 trajectories to be classified (drone versus non-drone). It contains observations from numerous mUASs such as the DJI Phantom 2 (diameter ≈ 0.4 m and weight ≈ 1.38 kg), DJI Inspire 2 as in Figure 3, DJI Matrice 200 (diameter ≈ 0.9 m and weight ≈ 5.5 kg), DJI Mavic 2 (≈ 0.35 m and weight ≈ 0.9 kg), fixed-wing BlueBear Blackstart (diameter ≈ 1.5 m and weight ≈ 4 kg), Alta X (diameter ≈ 1.4 m and weight ≈ 10 kg) and Octocopter (diameter ≈ 0.9 m and weight ≈ 4.5 kg) at ranges up to 7.5 km; over 50% of drone flights were at

[1]Results here are from experimental algorithms and should *not* be considered in anyway to represent the performance of the Thales Gamekeeper radar, which uses proprietary processing chain inclusive of the ATC.

Table 4.

Average ATC Performance From Real Radar Data		
Metric	**Value**	**Optimal Value**
Accuracy	0.97	1
F_1 Score	0.85	1
Hard TP Classification Probability (HTPCP)†	0.83	1
TP confidence (TPC)	0.76	1
TPC Deviation (DevTPC)	0.17	0
Hard FP Classification Probability (HFPCP)†	0.02	0
False Alarm Rate (FAR) (per hour)	3.4	0
FAR Ratio (FARR) (/100 tracks); $R_{\text{Track}} = 100$	0.03	0
Classification Time Delay (CTD) (sec)	9.15	0
UAS Declaration Probability (UDP)	0.98	1

†*They correspond to the traditional (hard) TP and FP metrics definitions.*

ranges exceeding 3 km. This is a diverse and challenging data with a wide range of SNRs (mean 35.14 dB and standard deviation 7.42 dB).

The machine-learned ATC algorithm [17] with an SMA as in Figure 4 is utilized with a MAP decision criterion for the hard true and false positives. Table 4 lists the averages of selected ATC metrics from all of the 25 radar recordings. One hard FP per track produces a FA with $M_U = 1$ in (9). It can been noticed that the high accuracy and low hard FP (HTPCP) does not give any insights on the potential FA rate, where $3.4/\text{h}$ on average can be regarded as high in some C-UAS scenarios if only a radar sensor is employed.

The average true positive confidence and hard true positive classification probability versus SNR from the real data is depicted in Figure 6. The substantial decay in TPC and HTPCP is visible as SNR declines. Finally,

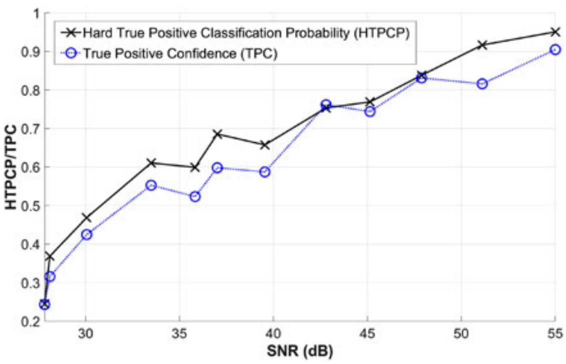

Figure 6.
Hard true positives and TP confidence for the test dataset versus the estimated SNR for the truth (drone) targets.

Figure 7 shows the ROC plot of the average HTPCP (recall) versus FAR (per hour) from all of the testing datasets; this is for a wide range of possible decision thresholds (i.e., rather than choosing the target category with the highest confidence score as with the MAP decision criterion as in Table 4). It exhibits the usual compromise between the classifier's TP and FP performance.

ADDITIONAL CAPABILITIES AND PROSPECTS

We now discuss a few opportunities and technologies that can aid improving the ATC functionality in drone surveillance radars, for example, to keep pace with the continuously evolving drone platforms. This goes beyond general system-level resilience, coverage, and data quality issues [3], [5], [29], for instance, utilizing distributed, multistatic, sensing-processing concepts, and state-of-the-art

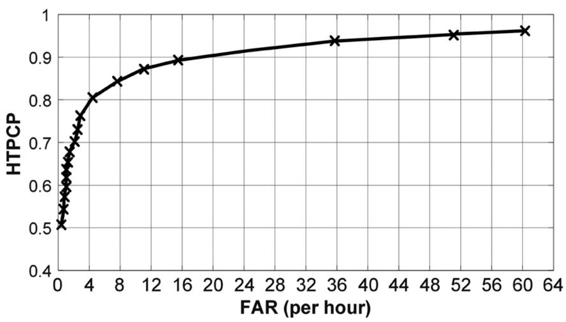

Figure 7.
Hard true positive classification probability versus FA rate per hour from radar measurements.

hardware (e.g., quantum oscillators, antenna designs, GPU-TPU-CPU, etc).

TARGET RECOGNITION BEYOND DRONE VERSUS NON-DRONE

Estimating the drone physical parameters such as the number of rotors and their rotation/flash rates, from the radar micro-Doppler signatures has a long history. It can offer indicators of the mUAS type (e.g., fixed or rotary wing), specific platform, and any payload [12], [30], [31], [32]. For example, the number of blades and their rotation speeds can be referenced against a database to establish the drone model. Similarly, the heavier the weight a drone needs to carry, the faster its rotors typically need to spin to provide sufficient lift. An abrupt change in the rotation rates can pertain to the drone releasing a payload. This information can aid ATC, where classical spectral analysis tools (e.g., cepstrum) or neural networks can be utilized for estimating the UAS physical parameters [30].

The ML classifier can be explicitly trained to recognize particular drone types or models [14], [18], [33]. This necessitates the availability of datasets per target label, thus higher training data requirements and/or applying a well-defined pipeline for refining a learned ATC algorithm, for instance, with transfer learning [14]. An alternative approach is to label the training data with parent (abstract) classes that the ATR then treats, for example, small fixed-wing, large rotary wing, and small rotary wing for a CNN micro-Doppler classifier as in [18].

COGNITIVE RADAR AND SENSING NETWORKS

The availability of low-cost electronic antennas, high performing, easily programmable, signal/data processing hardware and high-quality digital waveform generators are among those technology advancements that enable embedding intelligence or cognition within modern radar systems. These sensors can in principle be proactive and tailor their resources to multiple mission, for example, to increase performance against certain low observable targets such as drones whose salient radar signatures (e.g., micro-Doppler from rotors) can be otherwise undetectable due to the background noise-clutter [34], [35]. The radar can suitably adjust its transmission, beam-forming, and other parameters. It can specifically adapt its data acquisition (dwell-time) and processing (e.g., complexity of applied tracking and ATC algorithms) in a given target resolution cell in order to increase the SNR and maximize chance of detecting any micro-Doppler signature(s).

Given the prevalence of occlusions to the radar line of sight and persistent clutter in dense urban or other environments with large structures such as wind turbine, a networked or multistatic radar system with multiple spatially distributed transmitters and/or receivers might be required to maintain situational awareness over wide monitored regions. Such solutions have additional system-level challenges to overcome such as synchronization, data fusion, networking topology, etc. This is an active research area, see [3], [12], [29], [31], [36].

SIMULATORS AND DIGITAL TWINS

The main limitation for training generalizable ML classifiers is the availability of extensive radar datasets for all targets of interest. This is compounded by the relatively high cost of conducting controlled drone trials to collect real measurements; more so for birds instrumented with sensors (e.g., GPS tags) to provide ground truth information or utilizing a sophisticated automated labeling system (e.g., using cameras). Hence, there is a pressing need to generate representative synthetic radar data, including to augment the limited available real radar measurements. Conventional full electromagnetic physics-based radar simulators are generally prohibitively complex and time consuming to construct as well as difficult to validate. On the other hand, generative AI technology, such as transformers, generative adversarial networks, variational autoencoders, can expedite the process of simulating realistic radar signatures of various targets for training ATR algorithms. Easy and cheap access to representative simulated radar data can be revolutionary to the drone surveillance radar functionalities in the era of AI and data centric engineering. For example, it can permit adopting advanced fully or model-driven ML/AI for target detection [19] and track-before-detect methods [10] to enhance the radar ability to detect micro-Doppler and track low observable UAS targets (including when hovering).

Leveraging fully digitized and easily programmable processing chain as well as access to elaborate data simulators, digital twins can promote the rapid development, verification-validation, and integration of new ATR algorithms [35]. For example, the Thales Gamekeeper radar has a full digital twin of its processing chain, and the raw I&Q data from each receiver element can be recorded and reprocessed. Refinements to the detection, MTT and ATC modules can then be rapidly validated and deployed on radars in the field.

GLOBAL CLASSIFICATION ARCHITECTURES

Different classifiers can rely on distinct underlying salient characteristics in the target radar signatures (e.g., micro-

Doppler components or kinematic behavior) and over different time-scales as in Figure 4. Simultaneously employing disparate classifiers that can be potentially asynchronous (i.e., update at different rates) and heterogeneous (i.e., have different target categories) can deliver more robust ATC, such as the global classification architecture with classifiers dedicated to kinematic-features-based discrimination and others to micro-Doppler detection in [16]. Applying different versions of a classifier (e.g., several realizations of the same DNN, each configured following a particular initialization of its weights) and combining their results is common in supervised learning; see [37]. This can be extended to utilizing asynchronous and heterogeneous recognition algorithms with the associated fusion mechanism.

EXPLOITING CONTEXTUAL AND PATTERN-OF-LIFE INFORMATION

Given the complexity and diversity of the large surveyed areas with radar, available contextual (e.g., terrain type) or pattern-of-life (e.g., bird migration times) information can be highly effective at improving ATC and addressing the high FA challenge at the sensor level. These can bias/influence the classifier results and the associated decision criteria on the target category. For example, FAs originating from objects near major roads or dense urban areas, where high density vehicle traffic is expected (or even dynamically detected) at selected times can be suppressed or have a higher decision threshold. Similar schemes can be applied for detected large flocks of birds or even the occasional presence of bird species that can trigger FAs such as large gliding birds (e.g., from ornithologist studies or observations).

Therefore, a data fusion approach (e.g., within a Bayesian framework) would be required to capitalize on additional information about the monitored scene or present targets at the radar sensor or even at the C2 system level for a more robust ATC. This nonetheless carries the risk that inaccurate priors can undermine the radar ATC effectiveness and adversaries can exploit them. It is noted that contextual or pattern-of-life data can also be learned from historical radar data as with discriminative behavior features in the section “Drone Surveillance Radar and ATC.”

META-LEVEL INFORMATION INFERENCE AND MALICIOUS INTENT

As drones use is set to proliferate further, it will be critical for surveillance systems to be able to infer “*meta-level*” information on the detected-tracked-classified UAVs, namely their intent (e.g., final destination and future trajectory [38] to unveil, as early as possible, malicious activities) and group interactions-hierarchies in drone swarms (e.g., reveal coordinated mUASs groups and, if relevant, their leaders which can have more on-board capabilities). This can circumvent the system or operator being overwhelmed by swamping tactics. It facilitates timely decision making, automation, and prioritization of potential threats as well as selective deployment of countermeasures (if relevant), thereby minimizing potential collateral damage.

Bayesian meta-level tracking offers a generic framework, for instance to determine, early, if a drone intends to reach a prohibited zone [38], [39] or reveal a swarm hierarchy [40]. It can incorporate results from other threat assessment schemes, see [38] for a recent overview, and/or the ATC results as priors. For example, if the ATC indicates that a drone is carrying a payload, then this is strong indicator of malicious intent. Some of these functionalities can be employed at the C2 level, rather than by the radar sensor.

CONCLUSION

Robust ATC is fundamentally important for drone surveillance radar, especially given the large number of potential confuser targets (e.g., birds) and complex monitored environments. Although mUASs are formidably difficult targets to detect, track, and classify with radar, several sensors (including in multistatic configurations) have emerged over the last few years. They increasingly exploit recent advances in data processing and ML (e.g., DNNs) to deliver a strong ATC performance.

It is however crucial to: a) understand the unique challenges mUAVs pose to radar sensors and ATC enablers, especially that drone platforms are expected to continuously evolve and adapt as adversaries strive to make them harder to detect; b) consider relevant classification metrics when evaluating the efficacy of the classification approach; and c) highlight opportunities for future radar solutions. These aspects are discussed in this article. The objective is to promote a better appreciation of what is achievable in practice now and in the future, i.e., articulate the relationship between the art of the possible and operational effectiveness of automatic classification of drones with radar. An example is the common stringent ATC false alarm rates requirement on a C-UAS solution (e.g., one FA every several days or weeks). This is currently unrealistic to meet with a radar sensor alone while maintaining the ability to detect-track-classify miniature (e.g., micro and nano) drones in complex urban/semirural environments. While this can be fulfilled by a multisensor system within which radar is a critical component, the full potential of the C-UAS/UTM radar technology is yet to be realized.

Although we predominately focused here on ground surface radar and performance evaluation for a binary classification task (i.e., miniature drone versus non-drone), several of the presented arguments seamlessly apply to maritime and airborne radars. Metrics can also be easily extended to multi-class ATC scenarios and other targets (e.g., larger UAVs).

ACKNOWLEDGMENTS

This work was supported by Defence Science at Technology Laboratory (DSTL), U.K., through Defence and Security Accelerator (DASA) under contract DSTLX1000152816. The views in this article are of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.K. Ministry of Defence or Government.

REFERENCES

[1] Civil Aviation Authority, U.K., "Airspace restrictions for unmanned aircraft and drones," Accessed: Apr. 13, 2023. [Online]. Available: https://www.caa.co.uk

[2] SESAR JU, "Single European sky ATM research joint undertaking," Accessed: Apr. 13, 2023. [Online]. Available: http://www.sesarju.eu

[3] C. Clemente, F. Fioranelli, F. Colone, and G. Li, *Radar Countermeasures for Unmanned Aerial Vehicles*. London, U.K.: IET, 2021.

[4] J. Wang, Y. Liu, and H. Song, "Counter-unmanned aircraft system(s) (C-UAS): State of the art, challenges, and future trends," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 36, no. 3, pp. 4–29, Mar. 2021.

[5] G. Oswald and C. Baker, *Radar Countermeasures for Unmanned Aerial Vehicles*. London, U.K.: IET, 2021.

[6] S. Harman, "Analysis of the radar return of micro-UAVs in flight," in *Proc. IEEE Radar Conf.*, 2017, pp. 1159–1164.

[7] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking. Part I. Dynamic models," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1333–1364, Oct. 2003.

[8] R. Visina, Y. Bar-Shalom, and P. Willett, "Multiple-model estimators for tracking sharply maneuvering ground targets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 3, pp. 1404–1414, Jun. 2018.

[9] F. Goodall and B. I. Ahmad, "Adaptation of multi-target tracker using neural networks in drone surveillance radar," in *Proc. IEEE Radar Conf.*, 2023, pp. 1–6.

[10] M. üney, P. Horridge, B. Mulgrew, and S. Maskell, "Coherent long-time integration and Bayesian detection with bernoulli track-before-detect," *IEEE Signal Process. Lett.*, vol. 30, pp. 239–243, 2023.

[11] S. Harman and B. I. Ahmad, "The need for simultaneous tracking and recognition in drone surveillance radar," in *Proc. Int. Radar Symp.*, 2021, pp. 1–10.

[12] M. Ritchie, F. Fioranelli, H. Borrion, and H. Griffiths, "Multistatic micro-doppler radar feature extraction for classification of unloaded/loaded micro-drones," *IET Radar, Sonar Navig.*, vol. 11, no. 1, pp. 116–124, 2017.

[13] D. Blacknell and H. Griffiths, "Radar automatic target recognition ATR and non-cooperative target recognition NCTR," IET, London, U.K., Tech. Rep. 11300, 2021.

[14] B. Taha and A. Shoufan, "Machine learning-based drone detection and classification: State-of-the-art in research," *IEEE Access*, vol. 7, pp. 138669–138682, 2019.

[15] A. Huizing, M. Heiligers, B. Dekker, J. de Wit, L. Cifola, and R. Harmanny, "Deep learning for classification of mini-UAVs using micro-doppler spectrograms in cognitive radar," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 34, no. 11, pp. 46–56, Nov. 2019.

[16] F. Barbaresco, D. Brooks, and C. Adnet, "Machine and deep learning for drone radar recognition by micro-doppler and kinematic criteria," in *Proc. IEEE Radar Conf.*, 2020, pp. 1–6.

[17] M. Jahangir, B. I. Ahmad, and C. J. Baker, "Robust drone classification using two-stage decision trees and results from SESAR SAFIR trials," in *Proc. IEEE Int. Radar Conf.*, 2020, pp. 636–641.

[18] H. Dale, M. Antoniou, C. Baker, M. Jahangir, and A. Catherall, "Convolutional neural networks for drone model classification," in *Proc. Eur. Radar Conf.*, 2022, pp. 361–364.

[19] S. D. Rizvi, S. Ahmad, K. Khan, A. Hasan, and A. Masood, "Deep learning approach for fixed and rotary-wing target detection and classification in radars," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 37, no. 3, pp. 32–42, Mar. 2022.

[20] B. I. Ahmad, J. Grey, M. Newman, and S. Harman, "Low-latency convolutional neural network for classification of previously unseen drone types," in *Proc. Eur. Radar Conf.*, 2022, pp. 1–4.

[21] H. Dale, C. Baker, M. Antoniou, M. Jahangir, G. Atkinson, and S. Harman, "SNR-dependent drone classification using convolutional neural networks," *IET Radar, Sonar Navig.*, vol. 16, no. 1, pp. 22–33, 2022.

[22] G. Alain and Y. Bengio, "Understanding intermediate layers using linear classifier probes," in *Proc. Int. Conf. Learn. Representations*, 2016, pp. 1–13.

[23] V. C. Chen, F. Li, S.-S. Ho, and H. Wechsler, "Micro-Doppler effect in radar: Phenomenon, model, and simulation study," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 1, pp. 2–21, Jan. 2006.

[24] S. Rahman and D. A. Robertson, "Radar micro-Doppler signatures of drones and birds at K-band and w-band," *Sci. Rep.*, vol. 8, no. 1, pp. 1–11, 2018.

[25] C. Conte, G. de Alteriis, R. Schiano Lo Moriello, D. Accardo, and G. Rufino, "Drone trajectory segmentation for real-time and adaptive time-of-flight prediction," *Drones*, vol. 5, no. 3, 2021, Art. no. 62.

[26] Guinness World Record, “3281 drones break dazzling record,” Accessed: Apr. 13, 2023. [Online]. Available: https://www.guinnessworldrecords.com

[27] P. Votruba, R. Nisley, R. Rothrock, and B. Zombro, “Single integrated air picture (SIAP) metrics implementation,” Single Integrated Air Picture System Engineering Task Force, VA, Tech. Rep. 2001-001/002, 2001.

[28] M. Jahangir, B. I. Ahmad, and C. J. Baker, “The application of performance metrics to staring radar for drone surveillance,” in *Proc. 17th Eur. Radar Conf.*, 2021, pp. 382–385.

[29] M. Jahangir et al., “Networked staring radar testbed for urban surveillance: Status and preliminary results,” in *Proc. Int. Conf. Radar Syst.*, 2022, pp. 471–476.

[30] N. Regev, I. Yoffe, and D. Wulich, “Classification of single and multi propelled miniature drones using multilayer perceptron artificial neural network,” in *Proc. Int. Conf. Radar Syst.*, 2017, pp. 1–5.

[31] D. Dhulashia, N. Peters, C. Horne, P. Beasley, and M. Ritchie, “Multi-frequency radar micro-doppler based classification of micro-drone payload weight,” *Front. Signal Process.*, vol. 1, 2021, Art. no. 14.

[32] X. Ren, M. Jahangir, D. White, G. Atkinson, C. Baker, and M. Antoniou, “Estimating physical parameters from multirotor drone spectrograms,” in *Proc. Int. Conf. Radar Syst.*, 2022, pp. 20–25.

[33] L. Lehmann and J. Dall, “Simulation-based approach to classification of airborne drones,” in *Proc. IEEE Radar Conf.*, 2020, pp. 1–6.

[34] S. Brüggenwirth, M. Warnke, S. Wagner, and K. Barth, “Cognitive radar for classification,” *IEEE Aerosp. Electron. Syst. Mag.*, vol. 34, no. 12, pp. 30–38, Dec. 2019.

[35] M. Klein et al., “AI-augmented multi function radar engineering with digital twin: Towards proactivity,” in *Proc. IEEE Radar Conf.*, 2020, pp. 1–6.

[36] R. Palamà, F. Fioranelli, M. Ritchie, M. Inggs, S. Lewis, and H. Griffiths, “Measurements and discrimination of drones and birds with a multi-frequency multistatic radar system,” *IET Radar, Sonar Navigation*, vol. 15, no. 8, pp. 841–852, 2021.

[37] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, “An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes,” *Pattern Recognit.*, vol. 44, no. 8, pp. 1761–1776, 2011.

[38] J. Yun, D. Anderson, and F. Fioranelli, “Estimation of drone intention using trajectory frequency defined in radar’s measurement phase planes,” *IET Radar, Sonar Navig.*, vol. 17, pp. 1327–1341, 2023.

[39] J. Liang, B. I. Ahmad, M. Jahangir, and S. Godsill, “Detection of malicious intent in non-cooperative drone surveillance,” in *Proc. Sensor Signal Process. Defence Conf.*, 2021, pp. 1–5.

[40] Q. Li, B. I. Ahmad, and S. J. Godsill, “Sequential dynamic leadership inference using Bayesian Monte Carlo methods,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 4, pp. 2039–2052, Aug. 2021.