IET Radar, Sonar & Navigation

WILEY

The Institution of Engineering and Technology

ORIGINAL RESEARCH OPEN ACCESS

Autonomous Radar Ornithology: Benchmarking Automated Bird Detection and Tracking

Amy L. Jones[1] | David L. McCann[2] | Lucy J. Mitchell[1,3] | Jason McIlvenny[1] | Benjamin J. Williamson[1]

[1]Environmental Research Institute, University of Highlands and Islands NWH, Thurso, UK | [2]National Oceanography Centre, Liverpool, UK | [3]Department of Biology, TEREC/EcoBird, Ghent University, Ghent, Belgium

Correspondence: Amy L. Jones (amy.jones@uhi.ac.uk)

Received: 16 December 2025 | **Revised:** 16 March 2026 | **Accepted:** 30 March 2026

Keywords: marine navigation | marine radar | multi-target tracking | object detection | radar cross-sections | renewable energy sources | sea clutter

ABSTRACT

Automated radar-based tracking offers a scalable solution for avian monitoring, particularly in remote or data-intensive settings where manual annotation is impractical. This study evaluates the performance of the GlobAl Nearest Neighbour targEt Tracker (GANNET), a low-cost and customisable algorithm for detecting and tracking bird-like targets in X-band marine radar imagery. Using data from the Fall of Warness tidal test site in Orkney, Scotland, we compared manual and automated outputs across more than 55,000 filtered scan-level detections and over 34,000 consolidated trajectories derived from good-quality imagery. Trajectories were defined using a minimum track-length threshold of six detections, with a mean of 22 detections per track. GANNET processed data over 10 times faster than manual annotation and showed higher sensitivity, particularly in medium- and low-quality imagery where manual performance deteriorated. These results highlight GANNET's value for retrospective analysis of archived radar datasets and its potential application in environmental impact monitoring for birds in marine renewable energy contexts. Targeted improvements in clutter suppression, motion modelling and validation against a cooperative target would further strengthen its suitability for operational deployment. GANNET therefore represents a promising foundation for next-generation radar ornithology and offshore biodiversity assessment.

1 | Introduction

Radar has been used to study bird movements since the 1960s [1], offering continuous observation across large spatial scales and under conditions where visual monitoring is limited [2–4]. Increasing automation in radar processing has improved the efficiency and repeatability of detecting and tracking birds [5–8]. However, without concurrent ground-truthing, distinguishing birds from clutter remains difficult [9], reinforcing the importance of robust target classification based on features, such as size, speed, shape and flight behaviour [2, 10]. Approaches range from manual classification to fully automated machine-learning pipelines [11–14], with lightweight convolutional neural network (CNN) architectures further demonstrating promise for real-time classification in marine radar systems [15].

Accurate interpretation of radar data requires distinguishing three related but analytically distinct processes: detection, tracking and classification (examples in Figure 1). Detection involves identifying radar returns that exceed a defined noise-plus-clutter threshold. Tracking links sequential detections into coherent spatiotemporal trajectories. Classification assigns a physical target class to each trajectory by evaluating radar cross-section, kinematic characteristics and movement patterns [3, 4, 16].

Effective target detection, tracking and classification is particularly important in offshore environments, where radar is increasingly used to monitor interactions between seabirds and marine renewable energy (MRE) infrastructure [17–19]. Radar

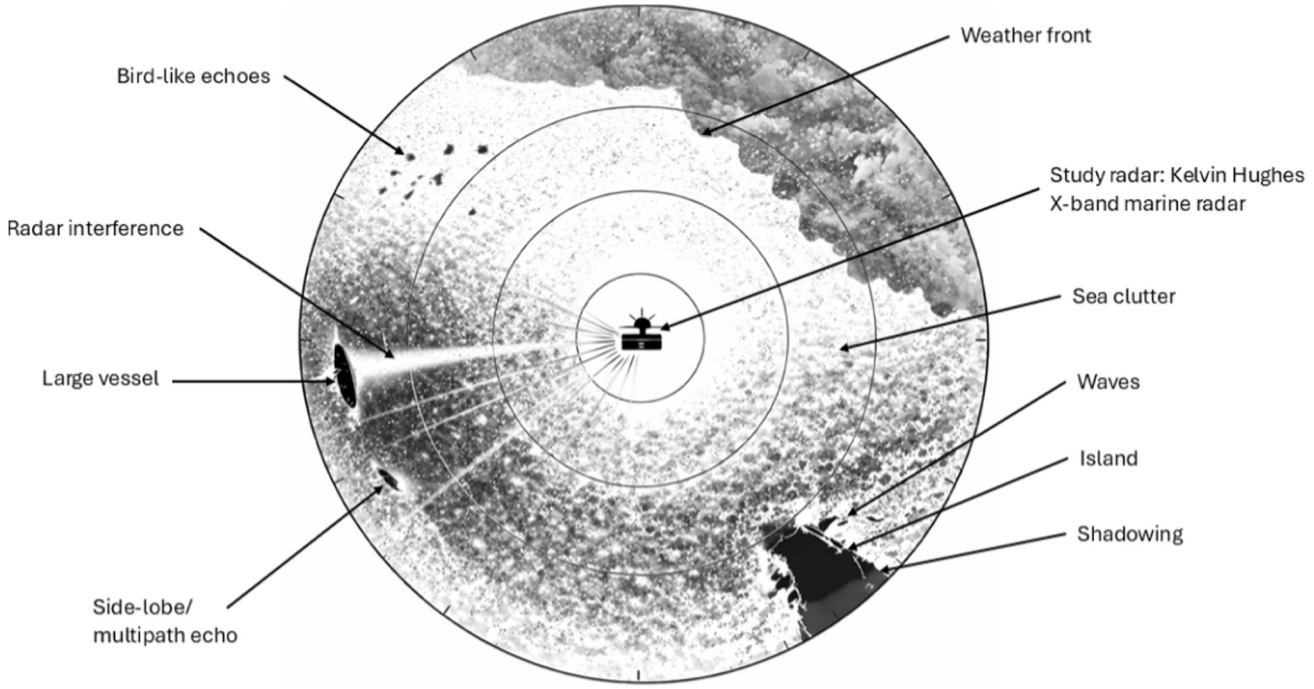

FIGURE 1 | Example schematic representation of a marine X-band radar image illustrating characteristic echo phenomenology in a coastal setting. Annotation shows bird-like echoes embedded within heterogeneous sea clutter and meteorological backscatter, alongside strong vessel returns, side-lobe artefacts, external radar interference and land-induced shadowing and waves.

offers target detection in conditions where traditional visual surveys are limited and can support mitigation measures such as warning and shutdown systems [20, 21]. Remote sensing is therefore an important complement to field observation in dynamic coastal settings [22–24]. However, using radar effectively in these environments depends on the ability to differentiate biological targets from clutter.

Target detection in marine radar is complicated by environmental clutter—rain, wave-breaking and multipath reflections—along with occlusion, weak echoes from small-bodied birds and technical constraints such as beamwidth and resolution [4, 10, 13]. Dynamic sea clutter in particular can obscure bird-sized echoes and degrade classifier performance. Radar setup, including grazing angle and coverage area, strongly influences detection capability [10]. As a result, classification frameworks must accommodate variable clutter, heterogeneous backgrounds and targets with low or inconsistent radar cross-sections.

A range of classification approaches have been tested using marine radar data. Random forests using radar cross section (RCS) features [4], gradient-boosting models [25] and support vector machines [10, 12] have all shown promise for distinguishing bird-like echoes from clutter. Despite methodological differences, these studies consistently demonstrate the feasibility of identifying bird-like targets and highlight the need for continued refinement in clutter handling and generalisation to complex backgrounds.

Alongside advances in machine learning, decision tree techniques remain valuable where computational resources or training data are limited [10, 25], and where model interpretability is a priority. For this reason, we evaluate GANNET, the GlobAl Nearest NEighbour Tracker developed by the National Oceanography Centre [13]. Following the global nearest neighbour (GNN) tracking framework presented in Konstantinova et al. [26], GANNET uses decision-tree-style logic and kinematic constraints to detect and track small airborne targets in marine X-band radar imagery. Although purpose-built avian radar systems such as BirdScan and ROBIN MAX offer sophisticated capabilities, their cost and logistical requirements limit deployment. GANNET provides a low-cost customisable alternative that leverages existing marine radar installations, making it suitable for large-scale and historical datasets [13].

Although not a full classifier, elements of GANNET's processing pipeline—thresholding, clutter-map constant false alarm rate (CM-CFAR) filtering, connected-component labelling (CCL) and GNN-based kinematic constraints—are used to provide an initial classification of targets. CM-CFAR is an adaptive thresholding technique that maintains a constant false alarm probability by estimating local background clutter statistics and adjusting the detection threshold accordingly (e.g., refs. [16, 27]). In the clutter-map variant, a temporally averaged representation of persistent returns (e.g., sea clutter and static features) is maintained and subtracted or used to scale detection thresholds, thereby suppressing stationary or quasi-stationary backscatter while preserving transient moving targets.

To perform detection and tracking under these conditions, GANNET uses a five-stage pipeline: (i) threshold-based detection, (ii) connected-component labelling for identification, (iii) a simple CM-CFAR adaptive clutter suppression using sectorial counts of identified targets and a threshold, (iv) data association using the Munkres algorithm and (v) track updating through an Extended Kalman Filter (EKF) [13].

The Munkres algorithm is a combinatorial optimisation method for solving the linear assignment problem, enabling optimal pairing between new detections and existing track predictions by minimising a global cost function (e.g., distance or kinematic residual) [28]. Here the Munkres algorithm is used to minimise the global cost of the difference between all detected target state vectors and all live tracks for each radar scan. This ensures globally consistent data association across multiple simultaneous targets, reducing track fragmentation and mis-association in cluttered environments. These components collectively support the detection and tracking of small moving targets under variable clutter conditions.

Here, we evaluate the performance of GANNET as a small-target tracker in X-band marine radar data and identify key areas for methodological refinement. Using high-volume radar datasets, we compare GANNET outputs with manual annotation to examine thresholds, identification performance and kinematic tracking behaviour. We also outline future steps for validating GANNET using cooperative targets, such as reflector-equipped drones, to support its development as a robust tool for avian monitoring in marine environments.

2 | Methods

2.1 | Study Site

As part of the FLOWBEC-4D multi-sensor project, which was primarily aimed at identifying the physical conditions influencing the behaviour of fish, their predators and benthic communities in high-energy tidal environments [29, 30], a noncoherent uncalibrated Kelvin Hughes X-band marine radar (9.4 GHz, 3.8 cm wavelength) was deployed by the UK National Oceanography Centre. The system was installed overlooking the Fall of Warness tidal energy test site, Orkney, Scotland (McCann and Bell [13]; Figure 2). Shore-based visual seabird surveys conducted at this site [31] document both species-level observations and the practical challenges associated with monitoring in such energetic tidal conditions, including constraints imposed by weather and visibility. The use of these radar data for automated detection and tracking of bird-like targets therefore represents a secondary application of an existing observational dataset, with potential to complement visual survey approaches in this dynamic environment.

The radar system comprised a 2.4 m horizontally polarised slotted-waveguide ('T-bar') antenna mounted approximately 7 m above mean sea level. The antenna rotated at approximately 46 rpm with a 0.8° horizontal and 20° vertical 3 dB beamwidth. The radar was operated in short-pulse mode (pulse width $\tau = 0.05\ \mu s$), giving a pulse-limited range resolution of $\Delta R = c\tau/2 \approx 7.5$ m, where c is the speed of light. Radar backscatter was digitised using an OceanWaveS GmbH WaMoS II Ocean State Monitoring system, which samples the analogue radar video output synchronously with antenna rotation at a rate of 30 MHz. This corresponds to a temporal sampling interval of 33 ns and an effective range sample spacing of 4.7 m, which oversamples the pulse-limited range resolution without increasing the fundamental resolving capability. One radar image was acquired per antenna rotation, with a maximum instrumented range of 4 km. Azimuth resolution was defined by the antenna 3 dB beamwidth (0.8°), whereas the azimuthal sampling interval was 1/3° per pixel. Images were stored in stacks of 256 scans (≈ 5.5 min).

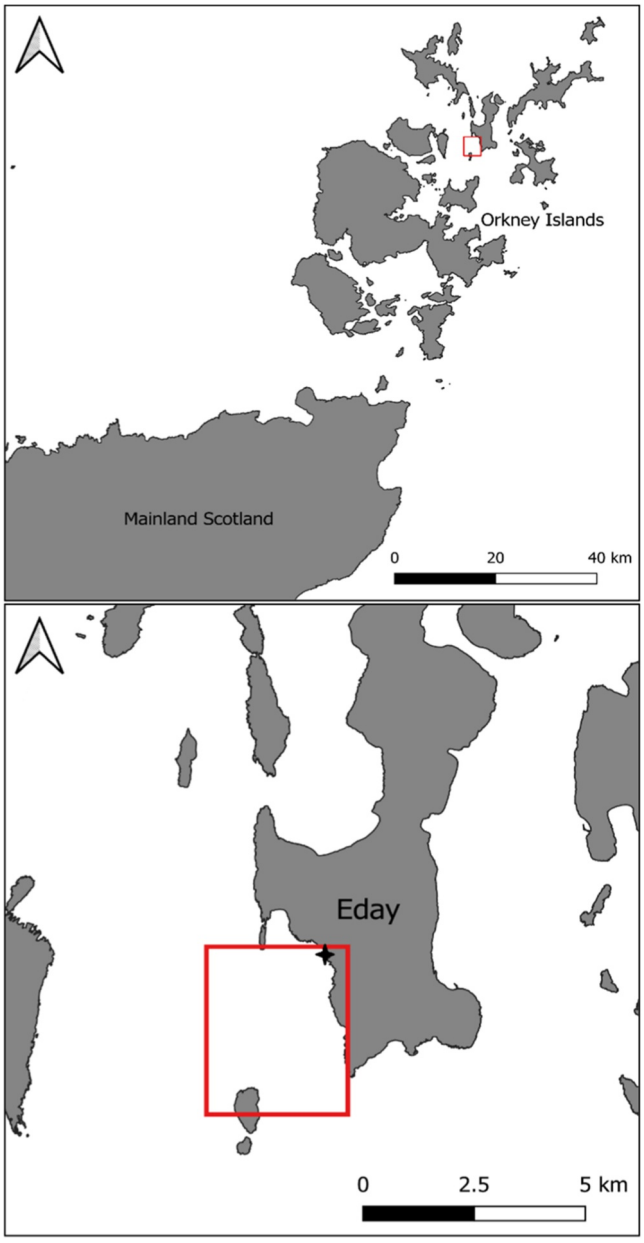

FIGURE 2 | Maps showing the location of the European Marine Energy Centre (EMEC) Fall of Warness tidal energy test site, situated off the west coast of Eday, Orkney, Scotland. Top: wider Orkney archipelago map illustrating the regional setting of EMEC tidal test site (red rectangle) relative to the Scottish mainland. Bottom: zoomed-in view of Eday, with the designated tidal test site boundary indicated. EMEC tidal energy test site highlighted by red rectangle, position of X-band marine radar indicated by black cross.

The Fall of Warness is a high-energy tidal channel (peak flow $> 3\ m\ s^{-1}$) hosting multiple tidal-stream energy demonstration platforms [32]. Ecological surveys conducted by Scottish Natural

Heritage [33] identified numerous ecologically important seabird and marine mammal species in the area, including Atlantic puffins (*Fratercula arctica*), black guillemots (*Cepphus grille*), common guillemots (*Uria aalge*), European shags (*Phalacrocorax aristotelis*), northern gannets (*Morus bassanus*), razorbills (*Alca torda*), grey seals (*Halichoerus grypus*) and harbour porpoises (*Phocoena Phocoena*). Spring migration of these bird species typically begins in mid-March, with peak breeding activity from May–August. Analyses were therefore restricted to May–August to maximise the likelihood of observing bird-like radar targets.

2.2 | Data Preparation

Radar image stacks collected between May and August of 2012 and 2013 were extracted from the WaMoS II system. Each stack consisted of 256 consecutive radar rotations, corresponding to approximately 5.5 min of data, given a rotation period of 1.3 s per full antenna sweep.

For each radar image stack $s(S)$, static land areas were excluded prior to statistical analysis using a binary land mask derived from georeferenced shoreline boundaries, ensuring that threshold statistics were computed using marine pixels only. The mean image intensity, denoted MII_S, was calculated over sea-surface pixels only. The distribution of pixel intensities within this masked stack was then characterised, and the interquartile range, IQR_S, was computed to derive a stack-specific adaptive detection threshold, T_S. Restricting the statistical analysis to the marine domain avoids bias from persistent high-amplitude land backscatter. These thresholds were determined iteratively from each stack's mean pixel return, enabling objective quality categorisation into three distinct image classes:

- Good: $<$ 25% of pixels exceeded the noise plus clutter threshold
- Medium: 25%–50%
- Bad: $>$ 50%

The adaptive noise plus clutter threshold T_s was then defined as follows:

$$T_S = MII_S + \alpha IQR_S$$

where α is a dimensionless scaling coefficient. In this study, $\alpha = 1$. Image quality classification was then classified based on the proportion of pixels exceeding this threshold.

$$\text{Quality}\,(S) = \begin{cases} \text{Good}, Ps < 0.25, \\ \text{Medium}, .25 \geq Ps \geq 0.50, \\ \text{Bad}, Ps > 0.50. \end{cases}$$

Only good-quality images were used for primary manual and automated analyses. Medium-quality images were retained to assess algorithmic robustness under moderate clutter conditions. Bad-quality images were excluded due to extensive clutter preventing reliable annotation. Figure 3 illustrates representative examples of the different image-quality categories and their associated clutter characteristics. All images were processed through a standardised workflow to facilitate direct comparison of manual and automated detections and tracks (Figure 4).

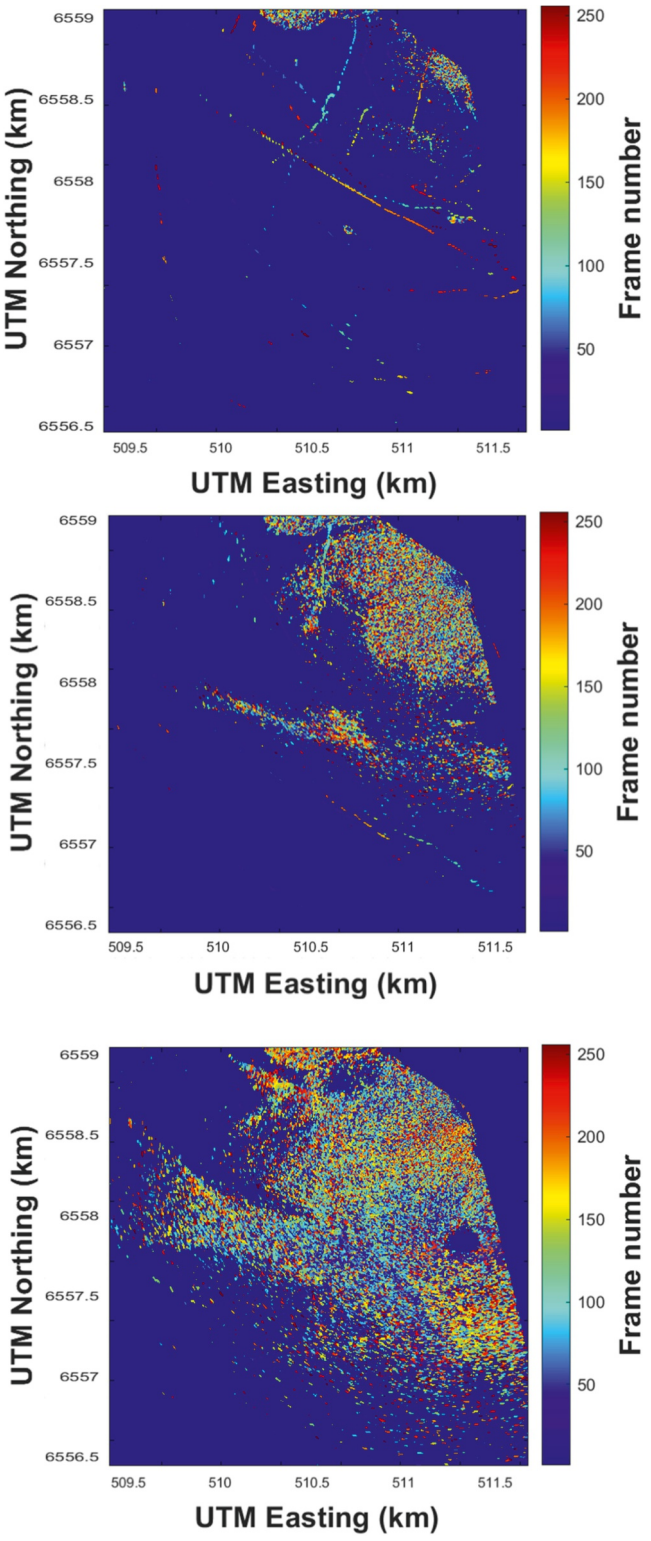

FIGURE 3 | Visual examples of the radar image quality categories. Top: good-quality image with fewer than 25% of pixels exceeding the adaptive noise plus clutter threshold. Middle: medium-quality image (25%–50% exceedance). Bottom: bad-quality image, where more than 50% of pixels exceed the threshold.

2.3 | Manual Analysis

Each good-quality image stack was converted into two types of images:

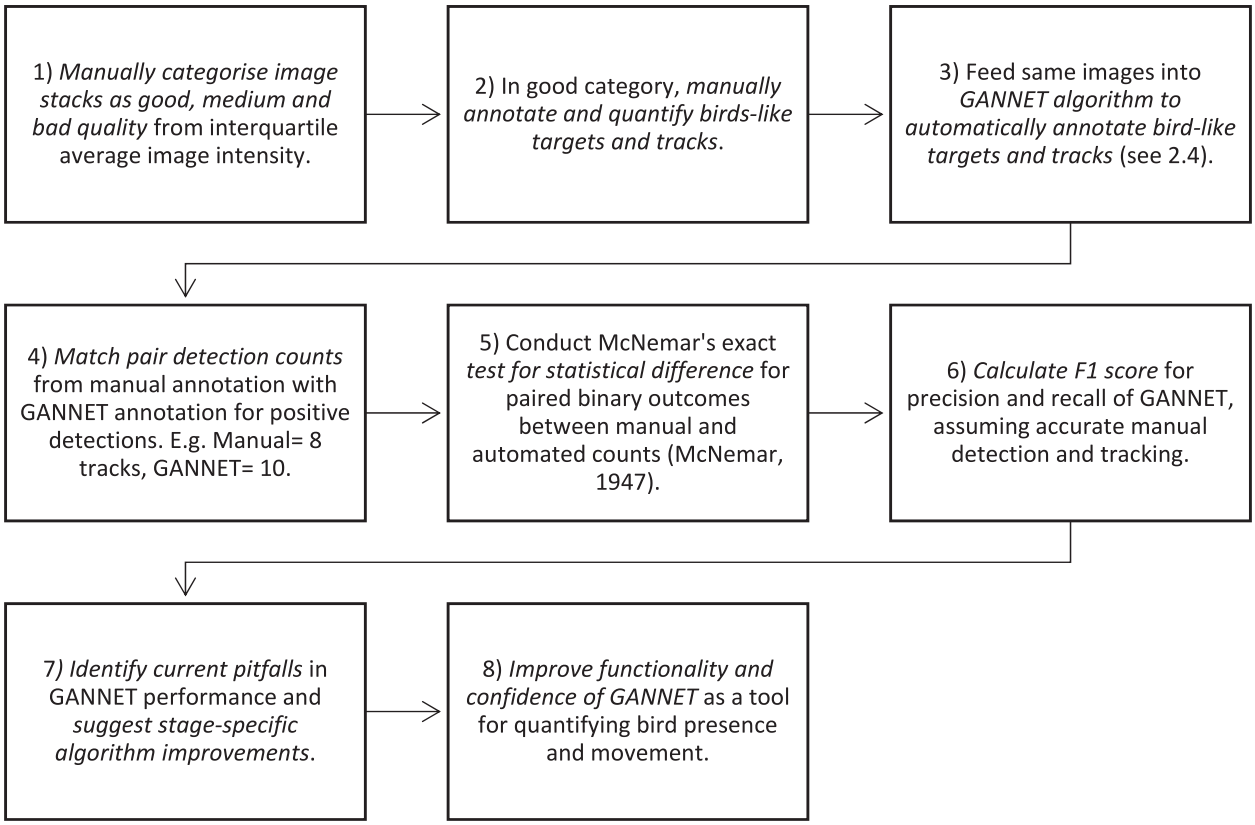

FIGURE 4 | Workflow for evaluating and refining the GANNET algorithm for automated bird detection and tracking.

1. Mean images—highlighting static features (e.g., land, infrastructure).
2. Maximum images—emphasising dynamic, transient returns (e.g., vessels, bird-like echoes).

Consecutive radar rotations were then combined into 'rainbow plots' by temporally stacking frames and colour-coding each according to acquisition order, preserving the original pixel intensities. Rainbow plots enabled visual assessment of target direction, speed, and motion tortuosity over the approximately 5 min period.

Objects with a small spatial footprint (1–4 pixels; ≤ 3 m^2), high apparent velocity (5–20 m s^{-1}), and tortuous movement ($9t = 0.1$–$0.$) were classified as bird-like. Tortuosity (t) was calculated as follows:

$$\text{Tortuosity} = \frac{D}{L}$$

where D is the straight-line distance between the start and end points of the track and L is the total path length (including curves).

Each rainbow plot was labelled for (1) presence/absence of bird-like detections, (2) presence/absence of bird-like tracks and (3) the number of bird-like tracks, which served as ground truth for later comparison with GANNET outputs.

2.4 | Automated Analysis

The same 'good' image stacks were then analysed using GANNET (automated processing outlined in Figure 5). Initially, pre-processing used a CM-FAR filter to suppress homogeneous clutter such as sea surface reflection or atmospheric interference (e.g., fog or light rain) while preserving potential targets.

Detections exceeding the adaptive MII-based threshold T_s were labelled as discrete objects and passed to GANNET's tracking module, which employed a recursive Kalman filter to associate detections over successive scans. Adaptive MII-based threshold T_s based on stack-specific mean image intensity and interquartile range were used to accommodate temporal variability in background backscatter and sea clutter, which precludes the use of a single global intensity threshold under dynamic marine conditions. The adaptive threshold T_S was derived on a per-stack basis and applied uniformly across the marine domain. Spatial heterogeneity in clutter (e.g., localised weather cells, wave breaking or wake structures) was mitigated during the preceding CM-CFAR stage, which performs local background normalisation prior to global threshold application. Consequently, temporal variability in background intensity was accommodated through stack-specific threshold adjustment, whereas spatial variability was addressed through CM-CFAR processing rather than through multiple region-specific thresholds. Here, k denotes the discrete time (scan) index, K is the Kalman gain, H is the measurement matrix, $P(k)$ is the state error covariance at time step k, R is the measurement noise covariance matrix and superscript T denotes matrix transposition.

$$K = \begin{pmatrix} HP(k)H^T + R \\ HP(k) \end{pmatrix}^T$$

Within the GANNET tracking framework, the EKF and GNN algorithms perform distinct but sequential functions. For each

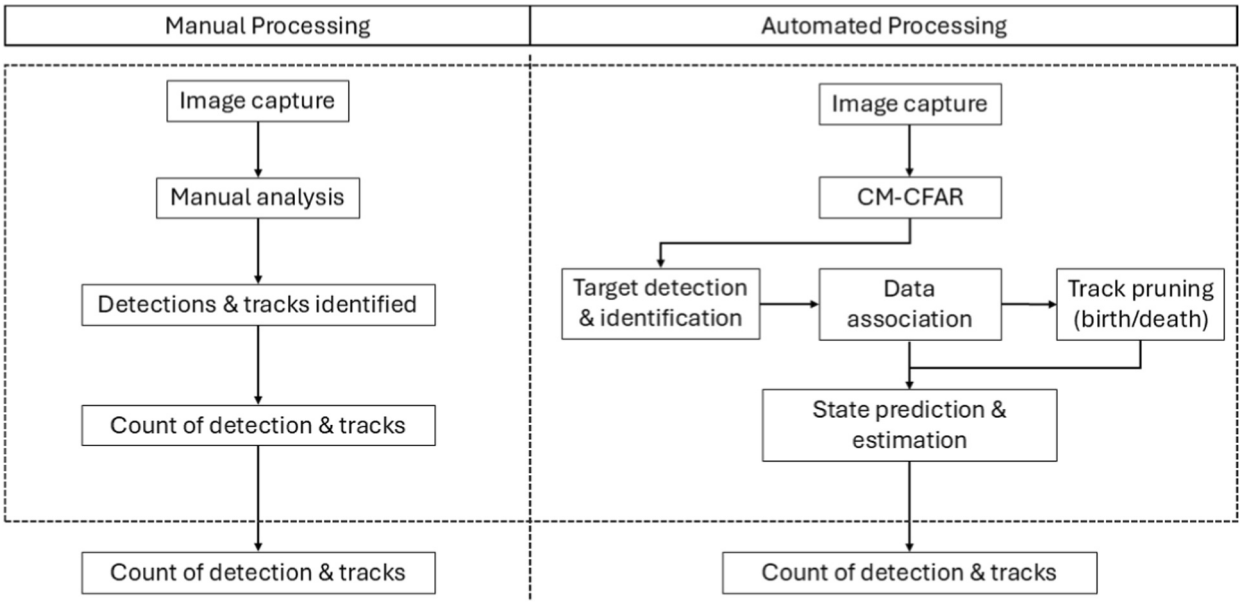

FIGURE 5 | Comparison of manual and automated radar-processing workflows used to derive detection and track counts. In the manual pathway (left), radar imagery is visually interpreted to identify detections and tracks prior to aggregation. In the automated pathway (right), imagery is processed using CM-CFAR, followed by target detection and identification, data association, track pruning (birth/death management) and state prediction and estimation [13]. Both approaches produce a count of detections and tracks, enabling direct comparison between manual and automated outputs.

scan k, existing track states are first propagated forward in time using the EKF prediction step, producing predicted state estimates and associated error covariances. These predicted states define spatial and kinematic gating regions. The GNN algorithm then performs measurement-to-track assignment by solving a global cost-minimisation problem (Munkres algorithm) using the predicted states [26]. Following assignment, the EKF performs the measurement update step to refine track state estimates. Track management procedures (birth, confirmation and pruning) are subsequently applied. The EKF therefore provides state estimation, whereas the GNN performs association; only the latter assigns detections to tracks.

The GNN algorithm is widely used single-hypothesis multi-target tracking method that associates detections with existing tracks by minimising a global assignment cost subject to spatial and kinematic gating constraints. In GANNET, the GNN framework links individual pixel detections into coherent tracks based on spatial proximity, velocity and similarity in radar cross-section and return intensity [26]. This approach is computationally efficient and well-suited to tracking transient targets in cluttered radar environments [13]. The residual vector Y between the measured state Y of each target j and the predicted state X of each track i for the current scan, where H is the measurement vector, is given by the following equation:

$$Y_{ij} = Y_j - HX_i$$

The residual covariance matrix S between the process error covariance Q and the measurement covariance R is then calculated as follows:

$$S = HQH^T + R$$

With the residual vector and covariance matrix obtained, the norm of the residual vector d_{ij}^2 is found using the following equation:

$$d_{ij}^2 = Y_{ij}T\, S^{-1} Y_{ij}$$

GANNET initiated new tracks using the gating procedure, where track birth denotes the establishment of a new tracking hypothesis when an incoming measurement cannot be matched to any existing track. Conversely, track death denotes the termination of a track following six consecutive missed target associations (approximately 7.02 s). Track births and deaths were thus managed dynamically within the gating procedure to reflect transient target behaviour (see Figure 6).

Automated tracks were stored with associated metadata and visualised as annotated outputs, where each track was colour-coded by average track speed (Figure 7). Tracking performance was evaluated by statistically comparing automated track counts and characteristics with manual annotations.

2.5 | Computational Environment

All automated analyses using GANNET were performed on a standard office-grade workstation (Intel Core i5-1235U CPU, 16 GB RAM) using MATLAB R2023b. Reported processing times reflect this baseline configuration, whereas substantially faster performance is achievable on higher-performance hardware.

2.6 | Statistical and Post-Hoc Analysis

An exact McNemar's binomial test was used to assess significance between discordant paired binary outcomes derived from the

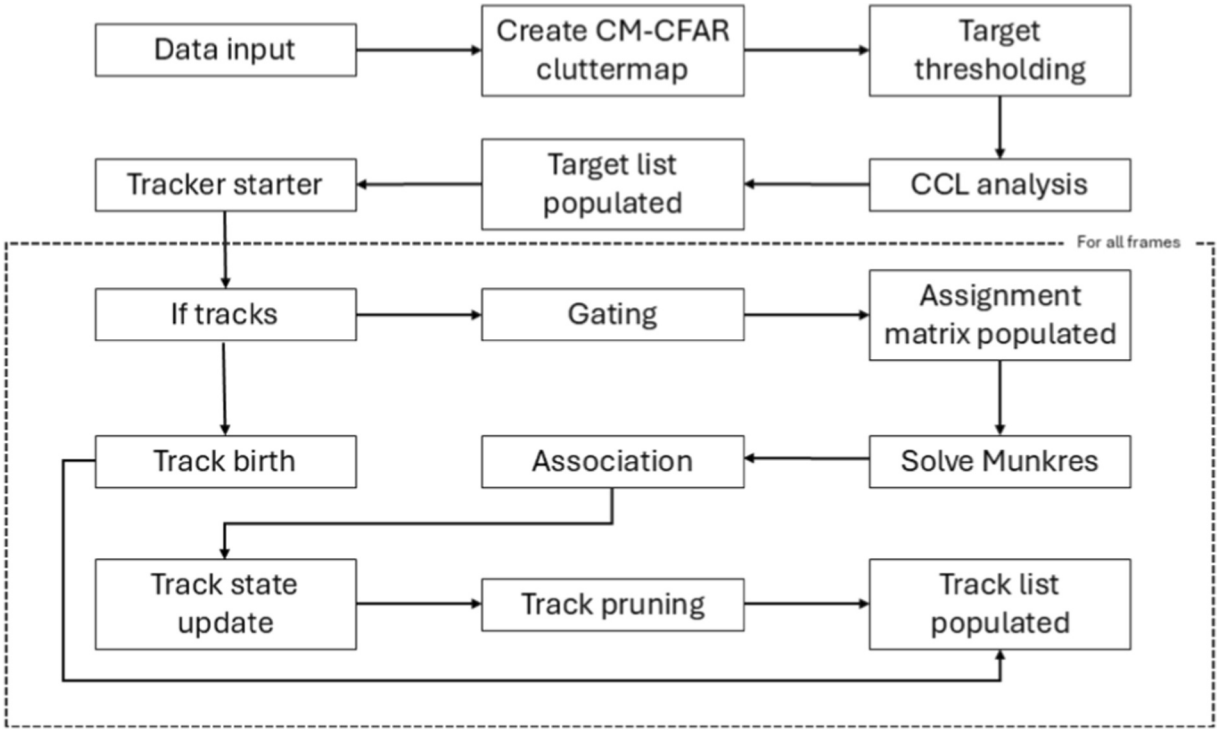

FIGURE 6 | Processing pipeline for GANNET detection and multi-target tracking [13]. Radar data are ingested to generate a CM-CFAR clutter map, followed by adaptive thresholding and connected-component labelling (CCL) to populate the target list. The tracker is initialised and, for each frame, performs gating, assignment matrix construction and optimal measurement-to-track association using the Munkres algorithm [28]. Track birth, state update and pruning steps manage track life cycles, resulting in an updated track list for subsequent frames.

same radar image, comparing manual and automated methods for detection and quantification (i.e., two classifiers applied to the same dataset [34]). The exact form of the test was selected to accommodate small monthly sample sizes.

Post-hoc Fisher's exact tests were then employed to identify months exhibiting the greatest disparity in detection and tracking performance between methods [35]. A Bonferroni correction was applied to adjust for multiple comparisons and to limit the family-wise false-positive error rate to below 5% [36].

Finally, an F1 score, defined as the harmonic mean of precision and recall, was calculated to summarise classification performance, with manually annotated detections treated as ground truth (verified accuracy $>$ 95% [37]).

The analysis pipeline comprised five steps:

1. Binary classification of detections (presence/absence),
2. Binary classification of tracks,
3. Enumeration of tracks,
4. Exact McNemar test on detection outcomes,
5. Exact McNemar test on tracking outcomes.

Each step was applied to both the manual and automated GANNET image analyses to evaluate agreement and performance under varying conditions. The full workflow for this study is outlined in Figure 8.

3 | Results

3.1 | Detection and Tracking Summary

Across the two monitoring periods, 55,383 scan-level bird-like detections were retained from good-quality radar imagery following preprocessing and quality filtering. These detections represent individual target positions identified during single antenna rotations. The detections were linked into 34,617 unique trajectories after data association and application of a minimum track-length threshold of six detections; single-scan detections were excluded.

When calculated consistently at the scan level, the mean trajectory length was 22 detections per track (seasonally variable), comparable to the 23 detections per track reported by McCann and Bell [13]. Detection totals reflect individual scan-level target positions, whereas trajectory counts represent multi-scan tracks; the two totals therefore do not directly represent detections per track.

Image quality fluctuated over time due to environmental variability (e.g., rain, wind and wave breaking), influencing the visibility and continuity of target tracks. High-clutter frames exhibited elevated MII values and reduced spatial heterogeneity, whereas low-clutter frames allowed clear target separation (Figure 9). Examination of MII and clutter characteristics revealed transitions in noise and clutter regimes that aligned with the predefined good, medium, and bad quality categories.

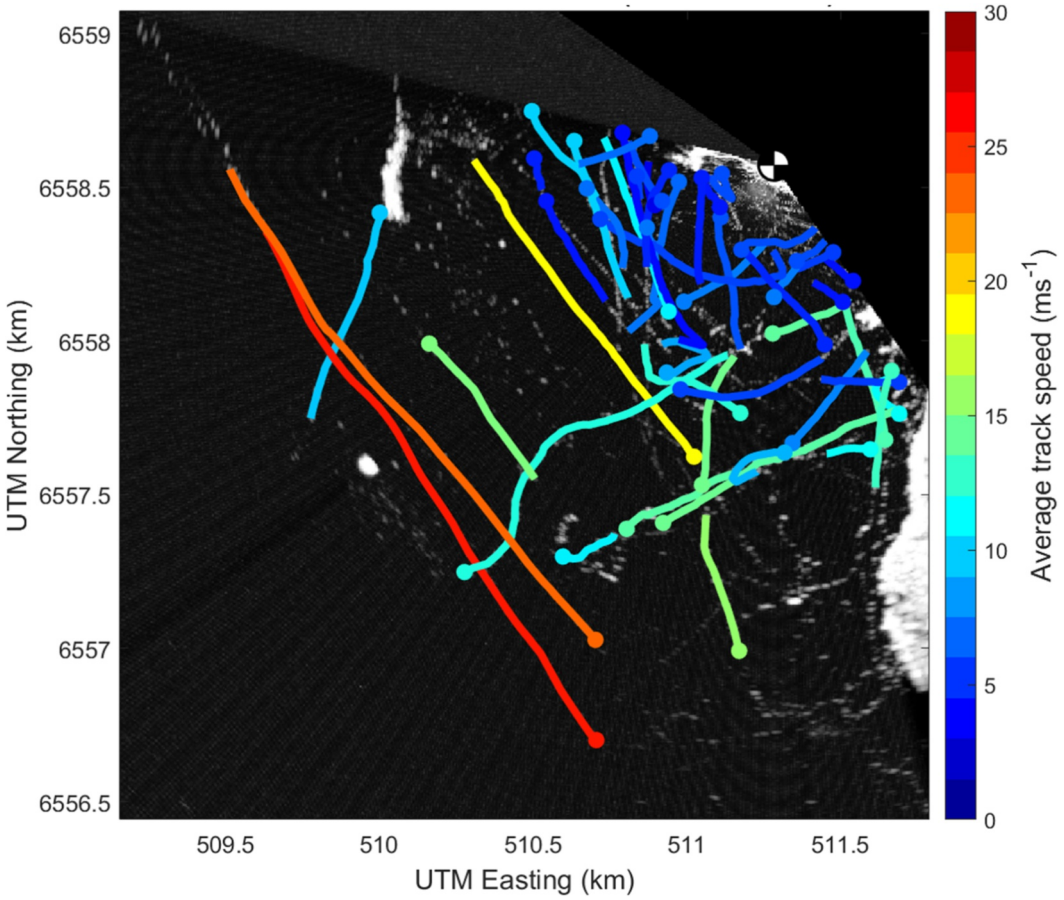

FIGURE 7 | Example GANNET output showing bird-like detection and tracks. Track are coloured according to average speed along track from blue (slow) to red (fast). Monochrome chequered dot in top right shows radar. White speckles and features are unclassified radar echoes. The top right black section is the blanked land. The remaining dark grey area is the sea.

3.2 | Statistical Performance Analysis

Manual annotations were treated as ground truth for all statistical comparisons, and automated performance metrics therefore quantify agreement with these annotations rather than absolute ecological truth. Manual annotation of good-quality imagery required 196 h, whereas GANNET processed the same dataset in 19 h, achieving a ten-fold increase in processing speed.

Paired exact McNemar testing showed 79% agreement between manual and automated detection classifications ($p = 0.0156$), indicating significant differences in detection outcomes. Tracking classifications also differed significantly ($p = 0.0313$), suggesting the two methods varied in their ability to maintain consistent trajectories.

Post-hoc Fisher's exact tests showed that discrepancies were concentrated in June–July 2013, during periods of persistent homogeneous sea clutter. In these months, manual annotation showed increased variability during these periods, whereas automated detections remained more consistent in their response to clutter, resulting in systematic differences between methods.

Both methods showed strong agreement for negative classifications (100% for detections; 95% for tracks). Relative to the manually annotated reference dataset, GANNET exhibited higher sensitivity, whereas manual annotation showed a higher rate of missed detections (approximately 5% of targets subsequently identified by the algorithm and confirmed upon review).

In medium- and low-quality imagery, agreement between automated and manual annotations declined, reflecting increased difficulty of target identification under elevated clutter conditions. Although manual annotation became increasingly time-consuming and inconsistent in these conditions, GANNET maintained more stable detection behaviour until the mean image intensity exceeded ~50%, beyond which clutter-driven false positives increased.

3.3 | Precision, Recall and F1 Scores

Across the full dataset, GANNET achieved a recall of 65% (0.654), precision of 54% (0.538) and an F1 score of 0.59, indicating moderate sensitivity with a tendency toward false positives in cluttered conditions (Table 1).

Here, true positives (TP) represent targets correctly detected by GANNET and confirmed by manual annotation, false positives (FP) represent detections identified by GANNET that were not present in the manual reference dataset and false negatives (FN) represent manually annotated targets that were not detected by GANNET.

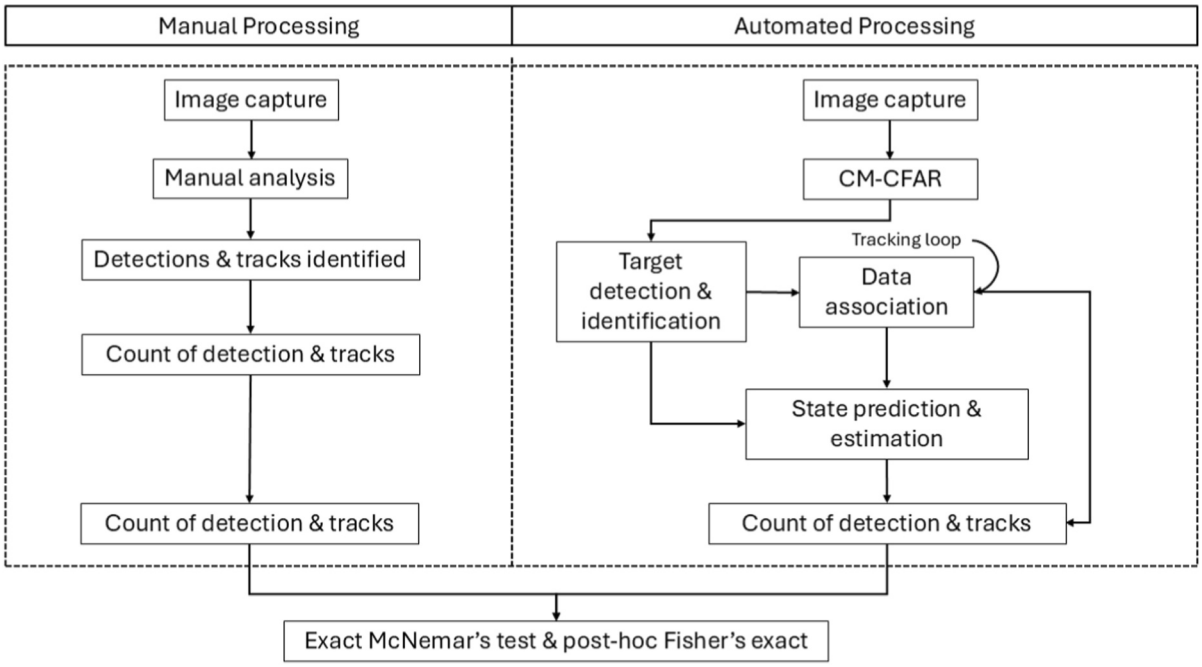

FIGURE 8 | Comparative analysis workflow used to assess agreement between manual interpretation and the automated GANNET tracking framework. For each radar image stack, manual processing (left) involved visual identification and enumeration of detections and tracks, whereas automated processing (right) comprised CM-CFAR thresholding, target detection and iterative multi-target tracking (state prediction, data association, and track management) prior to enumeration. Paired binary detection and tracking outcomes from the same radar images were compared using exact McNemar's binomial tests, with Bonferroni-adjusted Fisher's exact tests applied post-hoc to identify months with the greatest method-specific disparities.

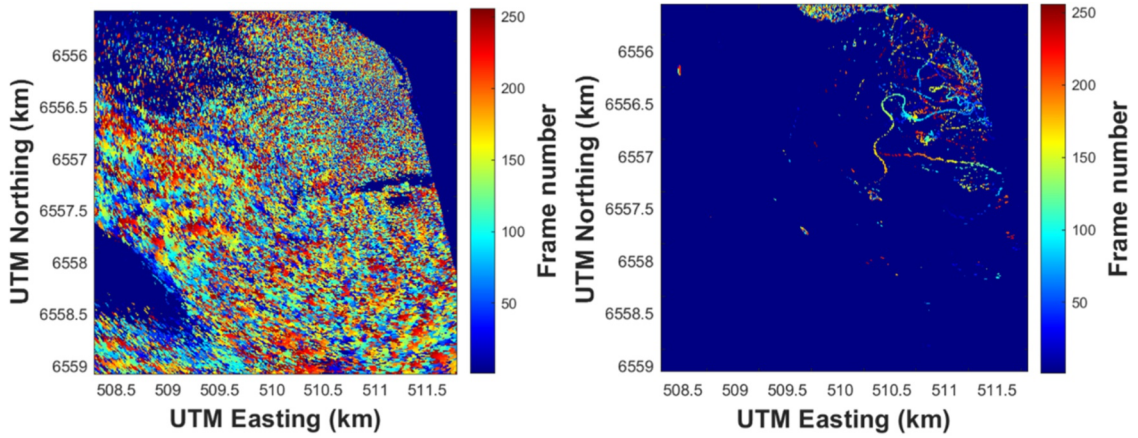

FIGURE 9 | Rainbow plots (overlays of maximum and mean images). (Left) High mean image intensity with extensive sea clutter obscuring individual targets. (Right) Low mean image intensity with low clutter, permitting clear separation of bird-like echoes and stable tracking.

TABLE 1 | Performance metrics for GANNET across the full dataset and under good-quality imaging conditions.

Data subset	Recall	Precision	F1 score	Interpretation
Full dataset (all conditions)	0.654 (65%)	0.538 (54%)	0.59	Moderate sensitivity with a tendency toward false positives. Reflecting clutter-limited performance under variable marine conditions.
Good-quality imagery	0.80 (80%)	0.70 (70%)	0.75	Performance comparable to or exceeding reported values for manual radar-based wildlife detection in cluttered marine environments.

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{399}{399 + 342} = 0.538$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{399}{399 + 211} = 0.654$$

$$\text{F1 Score} = 2\text{x}\,\frac{\text{Precision x Recall}}{\text{Precision} + \text{Recall}} = 2\text{x}\frac{0.538\text{ x }0.654}{0.538 + 0.654} = 0.590$$

In good-quality imagery, GANNET achieved recall and precision values of approximately 80% and 70%, respectively (Table 1). These findings are comparable to the range of values reported for automated radar-based wildlife detection systems operating in cluttered marine environments (e.g., [6, 10]). Performance requirements for precision, recall and F1 score are inherently application dependent. For large-scale radar monitoring and screening applications, recall is typically prioritised to minimise missed targets, with moderate false-positive rates considered acceptable and amenable to subsequent filtering or expert review. Within this operational context, the achieved performance represents a practical and effective compromise between detection sensitivity, robustness to sea clutter and substantial reductions in manual annotation effort. Small-bodied or weak-RCS birds were under-detected, particularly under cluttered conditions, likely due to target signatures approaching the radar noise plus clutter floor.

4 | Discussion

4.1 | GANNET Performance

In applications where GANNET is intended to support regulatory monitoring and large-scale screening (e.g., environmental monitoring at MRE sites), recall is often more critical than precision. Missing true targets carries greater risk than generating a limited number of false positives, particularly where automated detections can be subsequently filtered or reviewed. Within this context, GANNET demonstrates clear advantages over manual annotation, most notably in its ability to consistently detect and track bird-like targets across large and heterogeneous datasets.

Across the full dataset, recall and precision values of 65% and 54%, respectively, indicate relatively high sensitivity even under lower-quality imaging conditions. Although performance declines under increased sea clutter, these results are consistent with previous studies demonstrating the value of automated radar-based tracking in environments where visibility is constrained, data volumes are large and human annotation capacity is limited [6, 10].

Despite being implemented in MATLAB and executed on modest computational hardware, GANNET processed data approximately ten-times faster than manual annotation. This highlights its scalability and potential for substantial performance gains when deployed on high-performance computing systems. Manual annotation cannot be accelerated by improved hardware, making automation essential for large-scale and time-sensitive monitoring. GANNET's ability to rapidly process historical datasets is particularly valuable for MRE applications, where long-term radar archives support baseline assessments and adaptive environmental management [17–19].

As described in the introduction, environmental clutter strongly affects X-band radar performance. In our dataset, precision declined from approximately 70% in low-clutter frames (MII < 30) to $< 45\%$ when MII exceeded 50, whereas recall decreased from approximately 80% to $< 55\%$. This indicates nonlinear increases in clutter-driven false positives and missed detections, consistent with known X-band clutter regimes involving wave-breaking, precipitation and multi-path interference.

Sea clutter could, in principle, be reduced through geometric mitigation such as physical screening or elevation masking of the radar line of sight. However, in practice this approach is limited for seabird monitoring, as sea clutter originates from the ocean surface at low grazing angles that overlap with the airspace occupied by many birds flying close to the sea surface. Any fixed physical screen or imposed elevation mask capable of blocking sea-surface returns would therefore also attenuate or exclude low-altitude bird targets. In addition, clutter energy can enter the receiver through both the antenna main lobe and sidelobes, meaning that physical screening alone is unlikely to suppress clutter across all ranges and azimuths. Consequently, practical mitigation is more effectively achieved through deployment geometry (e.g., antenna height, siting relative to dominant wave directions and sidelobe control) combined with adaptive signal processing, rather than through physical screening alone.

These performance declines have ecological implications. Smaller-bodied or low-RCS species (e.g., puffins, auklets) are more likely to be masked during high-clutter periods, leading to systematic underestimation of their activity. Conversely, clutter-driven false positives can inflate apparent movement rates, potentially biasing flight-intensity metrics used in MRE risk assessments. GANNET also showed sensitivity to flight kinematics. Tracks associated with tortuous or erratic movements—typical of foraging seabirds—fragmented more frequently, reducing track continuity by 20%–35% compared with straighter trajectories. Rare behaviours such as stooping (rapid, steep diving attacks on prey) or escape manoeuvres were often under-sampled, potentially biasing behavioural interpretations.

Linking these performance patterns to known ecological behaviours helps identify where GANNET is robust and where methodological assumptions constrain interpretation. These insights highlight priority areas for improvement, including enhanced clutter suppression, behaviour-informed motion models and explicit uncertainty scoring.

In terms of validation context, no formal systematic human ornithological observations were conducted concurrently with the radar deployment, as the original FLOWBEC project was not designed specifically for radar-based bird monitoring. However, radar operators made qualitative visual observations during deployment periods and noted correspondence between real-time radar returns and birds observed in the airspace.

As with most radar-based wildlife studies, targets classified as bird-like may represent either single individuals or small aggregations, particularly at longer ranges or under cluttered conditions. Although this limits species attribution and precise individual counts, marine radar provides continuous omnidirectional surveillance over kilometre-scale ranges and extended

time periods independent of daylight. This capability far exceeds the spatial and temporal coverage achievable by human observers at exposed marine sites. The advantage is especially important for seabird monitoring, as many species undertake nocturnal or crepuscular movements that are poorly captured by visual surveys alone. Radar- and observer-based methods therefore offer complementary, rather than directly comparable, capabilities. Human observers remain essential for species-level identification and fine-scale behavioural interpretation at close range (< 2 km), as demonstrated in previous studies combining shore-based seabird surveys with remote-sensing techniques at the Fall of Warness [31]. In contrast, radar-based automation is particularly well suited to large-scale screening, long-term trend analysis and monitoring in challenging marine environments where visual observation is impractical or incomplete [13].

GANNET does not provide wingbeat analysis due to the temporal limitations of the rotating marine radar configuration, which produces an effective image update interval of 1.3 s. At this sampling rate, typical avian wingbeat frequencies cannot be reliably resolved and would be subject to severe sampling-induced distortion.

From an applied monitoring perspective, automated radar-based approaches, such as GANNET, offer distinct advantages over human observer surveys for long-term deployment. Radar-based tracking provides objective repeatable measurements that are not influenced by observer experience, visibility, fatigue or subjectivity and can be sustained over months to years at relatively low marginal cost once infrastructure is in place. In contrast, human observations are labour-intensive, costly to maintain at scale and inherently limited in duration and spatial extent.

As discussed in the introduction, commercial avian radar systems offer advanced classification but at high cost. In contrast, GANNET can be integrated into existing marine radar infrastructure, providing a practical and scalable alternative for small-target tracking.

4.2 | GANNET Limitations and Future Developments

A key limitation of the current GANNET implementation is its limited ability to suppress temporally nonstationary (and frequently non-Gaussian) sea-clutter conditions. CM-CFAR effectively attenuates homogeneous background noise and clutter but is less capable of handling transient and spatially variable artefacts generated by wave-breaking, precipitation cells, and antenna sidelobes. These features drive many of the false positives observed in high-clutter regimes, consistent with long-standing challenges in radar ornithology [2, 10].

Tracking performance is constrained by the use of a generalised EKF motion model. Incorporating species-or behaviour-specific kinematic constraints—such as turning limits, acceleration profiles or directional persistence—could improve track stability and reduce fragmentation [38, 39]. Rare or highly dynamic behaviours remain difficult to detect because they deviate from the motion and persistence assumptions embedded within the tracking framework, increasing the likelihood of track fragmentation or rejection during association. Expanding coverage of these behaviours, ideally through cooperative target validation, will be important for improving behavioural inference.

The use of tortuosity as a screening metric introduces an inherent trade-off between excluding nonbiological artefacts and retaining valid biological tracks. In this study, tortuosity was applied as a bounded filter (0.1–0.9) to remove near-linear tracks likely associated with clutter streaks or interference, as well as highly erratic tracks indicative of fragmentation or noise. Importantly, visually straight tracks may still fall within this range, as small deviations accumulated over long path lengths can yield moderate tortuosity values. Nevertheless, genuinely straight bird flight paths, such as sustained commuting or migratory movements, may be underrepresented by this criterion. This limitation highlights the need for more flexible, context-dependent trajectory metrics, such as behaviour-specific motion models or adaptive tortuosity thresholds, to better accommodate the full spectrum of avian flight behaviours.

Operationally, the current rendering and gridding pipeline (the conversion of raw polar radar returns into gridded Cartesian image representations for visualisation and analysis) also introduces computational overhead that constrains real-time deployment. Streamlining these stages through parallelisation or more efficient data structures could support responsive applications, such as turbine-curtailment systems [40]. Addressing these limitations will require both methodological refinement and broader validation. High-clutter and low signal to noise ratio (SNR) testing using reflector-equipped drones would provide essential ground-truth benchmarks for assessing algorithmic improvements. Enhanced feature-extraction methods, including shape- and motion-based descriptors, improved CM-CFAR tuning and better-balanced classifier thresholds, offer practical routes to strengthening detection sensitivity.

Machine learning approaches also present opportunities for more robust generalisation. Deep learning architectures such as CNNs and recurrent models (such as long short-term memories and recurrent neural networks) have shown strong performance in complex radar environments and could complement GANNET's rule-based framework [7, 41]. These methods are particularly suited to modelling nonlinear background patterns, learning rare or high-variability behaviours, and distinguishing biological motion from environmental noise and clutter.

Sensor-fusion represents another promising direction, where combining radar with optical or thermal sensors could increase detection confidence, reduce false positives and provide behavioural context that radar alone cannot capture. Cross-sensor confirmation, particularly during high clutter periods, would help mitigate both false positives and false negatives. Complementing this, systematic benchmarking of GANNET against both proprietary and open-source tracking systems across controlled gradients of clutter and SNR would clarify its operational limits and support the development of adaptive filtering strategies. Together, these advancements outline a clear trajectory toward improving GANNET's reliability, ecological fidelity and suitability for real-world conservation and mitigation applications.

5 | Conclusion

By quantifying GANNET's performance across diverse environmental conditions, we define the operational space in which GANNET can be used with confidence and highlight the methodological developments required to enhance future capability.

GANNET's compatibility with existing marine radar installations positions it as a practical tool for analysing large historical datasets and expanding the spatial and temporal scope of avian monitoring. With targeted refinement and continued validation, it has the potential to underpin next-generation approaches to radar-based wildlife surveillance in dynamic marine environments.

Author Contributions

Amy L. Jones: conceptualization, data curation, formal analysis, investigation, methodology, project administration, validation, visualization, writing – original draft, writing – review and editing. **David L. McCann:** conceptualization, data curation, formal analysis, methodology, resources, software, supervision, writing – review and editing. **Lucy J. Mitchell:** supervision, writing – review and editing. **Jason McIlvenny:** supervision, writing – review and editing. **Benjamin J. Williamson:** funding acquisition, project administration, supervision, writing – review and editing.

Acknowledgements

Many thanks to the University of the Highlands and Islands PhD Studentship award for funding this research.

Funding

This research was funded by the University of the Highlands and Islands PhD Studentship Award.

Ethics Statement

This study received full ethical approval from the University of the Highlands and Islands Research Ethics Committee.

Conflicts of Interest

This research forms part of the requirements for the degree of Doctor of Philosophy (PhD) of Amy Jones at Environmental Research Institute, University of the Highlands and Islands North, West and Hebrides.

Data Availability Statement

Datasets analysed in the current study are available from the corresponding author on reasonable request.

References

1. E. Eastwood, *Radar Ornithology* (Methuen & Co. Ltd, 1967).

2. I. M. Rosa, A. T. Marques, G. Palminha, et al., "Classification Success of Six Machine Learning Algorithms in Radar Ornithology," *Ibis* 158, no. 1 (2015): 28–42, https://doi.org/10.1111/ibi.12333.

3. S. S. Urmy and J. D. Warren, "Quantitative Ornithology With a Commercial Marine Radar: Standard-Target Calibration, Target Detection and Tracking, and Measurement of Echoes From Individuals and Flocks," *Methods in Ecology and Evolution* 8, no. 1 (2017): 860–869, https://doi.org/10.1111/2041-210X.12699.

4. S. S. Urmy and J. D. Warren, "Evaluating the Target-Tracking Performance of Scanning Avian Radars by Augmenting Data With Simulated Echoes," *Methods in Ecology and Evolution* 11, no. 4 (2020): 559–569, https://doi.org/10.1111/2041-210x.13365.

5. M. W. Klope, R. C. Beason, P. O. Box, T. J. Nohara, and M. J. Begier, "Role of Near-Miss Bird Strikes in Assessing Hazards," *Human-Wildlife Interactions* 3, no. 2 (2009): 208–215, http://www.jstor.org/stable/24875708.

6. A. M. Dokter, M. J. Baptist, B. J. Ens, K. L. Krijgsveld, and E. E. Van Loon, "Bird Radar Validation in the Field by Time-Referencing Line-Transect Surveys," *PLoS One* 8, no. 9 (2013): e74129, https://doi.org/10.1371/journal.pone.0074129.

7. C. Chilson, K. Avery, A. McGovern, E. Bridge, D. Sheldon, and J. Kelly, "Automated Detection of Bird Roosts Using NEXRAD Radar Data and Convolutional Neural Networks," *Remote Sensing in Ecology and Conservation* 5, no. 1 (2019): 20–32, https://doi.org/10.1002/rse2.92.

8. I. Schekler, T. Nave, I. Shimshoni, and N. Sapir, "Automatic Detection of Migrating Soaring Bird Flocks Using Weather Radars by Deep Learning," *Methods in Ecology and Evolution* 14, no. 8 (2023): 2084–2094, https://doi.org/10.1111/2041-210x.14161.

9. M. Peng, G. Li, X. Zhou, et al., "A Registration-Error-Resistant Swath Reconstruction Method of ZY1-02D Satellite Hyperspectral Data Using SRE-ResNet," *Remote Sensing* 14, no. 22 (2022): 5890, https://doi.org/10.3390/rs14225890.

10. R. C. Beason, T. J. Nohara, and P. Weber, "Beware the Boojum: Caveats and Strengths of Avian Radar," *Human-Wildlife Interactions* 7, no. 1 (2013): 16–46, https://www.jstor.org/stable/24874115.

11. R. Mead, J. Paxton, and R. S. Sojda, "Identifying Biological Echoes in Radar Scans Using Machine Learning," *International Environmental Modelling and Software Society* 4th (2008): Biennial Meeting, https://scholarsarchive.byu.edu/iemssconference/2008/all/165/.

12. S. Zaugg, G. Saporta, E. Van Loon, H. Schmaljohann, and F. Liechti, "Automatic Identification of Bird Targets With Radar via Patterns Produced by Wing Flapping," *Journal of Royal Society Interface* 5, no. 26 (2008): 1041–1052, https://doi.org/10.1098/rsif.2007.1349.

13. D. L. McCann and P. S. Bell, "Visualising the Aspect-Dependent Radar Cross Section of Seabirds Over a Tidal Energy Test Site Using a Commercial Marine Radar System," *International Journal of Marine Energy* 17 (2017): 56–63, https://doi.org/10.1016/j.ijome.2017.01.002.

14. M. U. Haque, J. J. Dabrowski, R. M. Rogers, and H. Parry, "Detection of Animal Movement From Weather Radar Using Self-Supervised Learning," *Neural Information Processing* (2024), https://doi.org/10.48550/arXiv.2408.04424.

15. X. Zhang, W. Hu, W. Qu, and S. Maybank, "Multiple Object Tracking via Species-Based Particle Swarm Optimization," *IEEE Transactions on Circuits and Systems for Video Technology* 20, no. 11 (2010): 1590–1602, https://doi.org/10.1109/tcsvt.2010.2087455.

16. M. I. Skolnik, *Radar Handbook*. 3rd ed. (McGraw-Hill, 2008).

17. M. Desholm, A. D. Fox, P. D. L. Beasley, and J. Kahlert, "Remote Techniques for Counting and Estimating the Number of Bird–Wind Turbine Collisions at Sea: A Review," *Ibis* 148, no. 1 (2006): 76–89, https://doi.org/10.1111/j.1474-919x.2006.00509.x.

18. A. T. Marques, H. Batalha, S. Rodrigues, et al., "Understanding Bird Collisions at Wind Farms: An Updated Review on the Causes and Possible Mitigation Strategies," *Biological Conservation* 179 (2014): 40–52, https://doi.org/10.1016/j.biocon.2014.08.017.

19. B. J. Williamson, S. Fraser, P. Blondel, P. S. Bell, J. J. Waggitt, and B. E. Scott, "Integrating a Multibeam and a Multifrequency Echosounder on the FLOWBEC Seabed Platform to Track Fish and Seabird Behavior Around Tidal Turbine Structures," in *4th Annual Marine Energy*

Technology Symposium (METS), (2016), https://tethys.pnnl.gov/sites/default/files/publications/Williamson-et-al-2016-METS.pdf.

20. A. L. Drewitt and R. H. W. Langston, "Assessing the Impacts of Wind Farms on Birds," *Ibis* 148, no. 1 (2006): 29–42, https://doi.org/10.1111/j.1474-919x.2006.00516.x.

21. R. May, O. Reitan, K. Bevanger, S. Lorentsen, and T. Nygård, "Mitigating Wind-Turbine Induced Avian Mortality: Sensory, Aerodynamic and Cognitive Constraints and Options," *Renewable and Sustainable Energy Reviews* 242 (2014): 170–181, https://doi.org/10.1016/j.rser.2014.10.002.

22. B. J. Williamson, S. Fraser, P. Blondel, P. S. Bell, J. J. Waggitt, and B. E. Scott, "Multisensor Acoustic Tracking of Fish and Seabird Behavior Around Tidal Turbine Structures in Scotland," *IEEE Journal of Oceanic Engineering* 42, no. 4 (2017): 948–965, https://doi.org/10.1109/joe.2016.2637179.

23. N. Largey, A. S. C. P. Cook, C. B. Thaxter, et al., "Methods to Quantify Avian Airspace Use in Relation to Wind Energy Development," *Ibis* 163, no. 3 (2021): 747–764, https://doi.org/10.1111/ibi.12913.

24. M. Atencio, M. A. De La Colina, and B. Mahler, "Behavioural Responses Are Associated With Mortality and Mobility After Reintroduction in the Endangered Yellow Cardinal," *Ibis* 165, no. 4 (2023): 1145–1155, https://doi.org/10.1111/ibi.13191.

25. J. Bhatia, A. Dayal, A. Jha, et al., "Classification of Targets Using Statistical Features From Range FFT of mmWave FMCW Radars," *Electronics* 10, no. 16 (2021): e1965, https://doi.org/10.3390/electronics10161965.

26. P. Konstantinova, A. Udvarev, and T. Semerdjiev, "A Study of a Target Tracking Algorithm Using Global Nearest Neighbor Approach," in *CompSysTech '03: Proceedings of the 4th International Conference on Computer Systems and Technologies*, (2003), 290–295, https://doi.org/10.1145/973620.973668.

27. H. Rohling, *New Procedure for Disturbance-Adaptive Target Recognition in a CFAR Processor* (Radar Technology 5th Symposium, 1983), https://scixplorer.org/abs/1983OrNav.....15R/abstract.

28. J. Munkres, "Algorithms for the Assignment and Transportation Problems," *Journal of the Society for Industrial and Applied Mathematics* 5, no. 1 (1957): 32–38, https://doi.org/10.1137/0105003.

29. P. S. Bell, D. L. McCann, B. E. Scott, et al., "Flow and Benthic Ecology 4D- FLOWBEC- An Overview," in *Proceedings of the 2nd International Conference on Environmental Interactions of Marine Renewable Energy Technologies (EIMR2014)*, 28 April – 02 May 2014, (2014).

30. B. J. Williamson, B. E. Scott, J. Waggitt, et al., "FLOWBEC Seabed Frame to Understand Underwater Interactions Between Seabirds, Prey, Hydrodynamics and Tidal and Wave Energy Structures," in *Proceedings of the 2nd International Conference on Environmental Interactions of Marine Renewable Energy Technologies (EIMR2014)*, 28 April – 02 May 2014, (2014).

31. J. J. Waggitt, P. W. Cazenave, R. Torres, B. J. Williamson, and B. E. Scott, "Predictable Hydrodynamic Conditions Explain Temporal Variations in the Density of Benthic Foraging Seabirds in a Tidal Stream Environment," *ICES Journal of Marine Science* 73, no. 10 (2016): 2677–2686, https://doi.org/10.1093/icesjms/fsw100.

32. S. P. Neill, A. Vögler, A. J. Goward-Brown, et al., "The Wave and Tidal Resource of Scotland," *Renewable Energy* 114, no. A (2017): 3–17, https://doi.org/10.1016/j.renene.2017.03.027.

33. A. Robbins, "Analysis of Bird and Marine Mammal Data for the Fall of Warness Tidal Test Site, Orkney," *Scottish Natural Heritage Commissioned Report* (2011).

34. Q. McNemar, "Note on the Sampling Error of the Difference Between Correlated Proportions or Percentages," *Psychometrika* 12, no. 2 (1947): 153–157, https://doi.org/10.1007/BF02295996.

35. R. A. Fisher, "On the Interpretation of χ^2 From Contingency Tables, and the Calculation of P," *Journal of the Royal Statistical Society* 85, no. 1 (1922): 87–94, https://doi.org/10.2307/2340521.

36. O. J. Dunn, "Multiple Comparisons Among Means," *Journal of the American Statistical Association* 56, no. 293 (1961): 52–64, https://doi.org/10.1080/01621459.1961.10482090.

37. D. M. W. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation," *International Journal of Machine Learning Technology* 2, no. 1 (2011): 37–63, https://doi.org/10.48550/arXiv.2010.16061.

38. P. Zhang, S. Liu, and Q. Xu, "Identifying Doppler Velocity Contamination Caused by Migrating Birds," *Journal of Atmospheric and Oceanic Technology* 22, no. 8 (2005): 1105–1113, https://doi.org/10.1175/jtech1757.1.

39. L. Dinevich and Y. Leshem, "Algorithmic System for Identifying Bird Radio-Echo and Plotting Radar Ornithological Charts," *Ring* 29, no. 1–2 (2007): 3–39, https://doi.org/10.2478/v10050-008-0040-z.

40. P. B. Garcia-Rosa and J. O. G. Tande, "Mitigation Measures for Preventing Collision of Birds With Wind Turbines," *Journal of Physics: Conference Series* 2626, no. 2023 (2023): 012072, https://doi.org/10.1088/1742-6596/2626/1/012072.

41. R. Dadial and B. Thakur, "Advancing Wildlife Conservation Through Drone-Based Multi Classification of Species With Fused CNN-LSTM Deep Learning Model," *10th International Conference on Advanced Computing and Communication Systems* (2024): 1317–1321, https://doi.org/10.1109/IACCS60874.2024.10717116.