Journal of *Marine Science and Engineering*

MDPI

Article

A Computer-Vision Biological Early Warning System for Marine Pollution Detection Using *Aurelia aurita* as a Biosensor: Per-Animal Anomaly Detection of Diesel Exposure

Aleksandr Grekov *, Kirill Paraev, Iuliia Baiandina, Aleksei Baiandin and Elena Vyshkvarkova

A. O. Kovalevsky Institute of Biology of the Southern Seas Russian Academy of Sciences (IBSS), Sevastopol 299011, Russia; paraev57@gmail.com (K.P.); baiandina_yus@ibss-ras.ru (I.B.); baiandin_as@ibss-ras.ru (A.B.); aveiro_7@mail.ru (E.V.)

* Correspondence: i@angrekov.ru

Abstract

Marine pollution monitoring increasingly relies on Biological Early Warning Systems (BEWSs), which use living organisms as continuous, integrative sentinels of water quality. The moon jellyfish *Aurelia aurita* is a sensitive but under-exploited candidate for this role. We present a computer-vision BEWS pipeline that is unsupervised at inference time and operates without labelled pollution-response data, converting side-view aquarium video of single *A. aurita* medusae into a binary pollution alarm. Per-frame YOLO bounding-box detections are reduced to a continuous bell-area signal and a centroid trajectory, from which eleven pulsation, kinematic, and detection-quality features are extracted on 60 s sliding windows. A per-animal baseline is fitted on a clean-water baseline (recommended $\geq$15 min), and a two-layer detector—fast outlier detection on the mean absolute z-score with a k-of-N rule, plus one-sided CUSUM (cumulative sum) accumulation—flags any sustained deviation. Validation on six adult medusae exposed to diesel-WAF detected all six animals (95% CI 54–100%) and produced no false alarms in 203 clean-window opportunities (exact 95% upper bound 1.8%; rule-of-three estimate $\approx$1.5%). First-alarm latencies ranged from 1.0 to 23.7 min, and the observed responses were described as three descriptive patterns in this pilot dataset: sharp step-change, slow drift, and mixed. The deployed anomaly scoring step contains no neural-network weights, runs in under 300 lines of Python, and is designed for field-portable use in settings where a stationary side-view camera can be positioned alongside an aquarium, although field validation remains required. Per-animal anomaly detection accommodates the strong inter-individual variability of the diesel-WAF response that limits supervised clean-versus-polluted classification at this sample size.

Keywords: Biological Early Warning System; *Aurelia aurita*; computer vision; YOLO; anomaly detection; CUSUM; bell pulsation frequency; ecotoxicology; marine pollution; biosensor

Academic Editor: Azizur Rahman

Received: 21 May 2026
Revised: 25 June 2026
Accepted: 26 June 2026
Published: 28 June 2026

1. Introduction

Marine and coastal water bodies are increasingly exposed to a diverse array of anthropogenic pollutants, including petroleum hydrocarbons and polycyclic aromatic hydrocarbons (PAHs), heavy metals, synthetic pesticides, and microplastics. These contaminants pose significant threats to aquatic ecosystem health and, through bioaccumulation pathways, to human populations that depend on coastal resources. Conventional physicochemi-

cal monitoring methods—routine water sampling, laboratory chemical analysis, and sensor networks measuring parameters such as pH, dissolved oxygen, and conductivity—provide essential baseline data but are fundamentally limited in their ability to assess the integrated toxicological impact of complex, multi-component pollution events in real time.

Biological Early Warning Systems (BEWSs) serve as a biological complement to physicochemical monitoring by enabling continuous assessment of water quality changes through measurable organismal responses [1,2]. Existing BEWSs have predominantly relied on taxa with well-defined and readily quantifiable endpoints: valve movement and cardiac activity in bivalves [3,4], ventilatory activity in fish [5], and behavioural or physiological parameters in standard test organisms such as algae, Daphnia, and gammarids [6]. Machine-learning approaches have further been applied to improve the interpretation of warning signals in long-term BEWS datasets, including *Daphnia magna* monitoring [7]. Pelagic gelatinous organisms, by contrast, remain poorly represented in this field, despite their continuous exposure to dissolved and suspended contaminants throughout the water column and the availability of quantifiable behavioural endpoints that can be recorded non-invasively [8].

The limited application of jellyfish in BEWSs does not reflect a lack of biological relevance. It is more likely attributable to practical and methodological constraints. Medusae are mechanically fragile and are less amenable to long-term laboratory maintenance than bivalves, fish, or crustacean model species. Furthermore, standardized behavioural assays for jellyfish have been developed primarily for ephyrae rather than adult medusae [9–12]. Jellyfish have also been investigated more frequently in the context of bloom dynamics and ecosystem disturbance than as bioindicator organisms in contamination monitoring. An additional constraint, until recently, has been the absence of automated tools for continuous, non-invasive recording of medusa locomotor activity. Computer-vision methods now enable quantification of bell pulsation frequency and translational movement from video footage, without physical attachment to or manipulation of the experimental organism.

Among potential sentinel organisms, the moon jellyfish *Aurelia aurita* (Linnaeus, 1758) presents a compelling but under-exploited option for BEWS applications [13,14]. As a scyphozoan medusa abundant across wide latitudinal and coastal ranges, *A. aurita* occupies a central position in pelagic food webs and is highly sensitive to changes in water chemistry. Earlier work documented accumulation of trace metals (Al, As, Cd, Cu, Zn) in jellyfish tissues at levels far above ambient concentrations, with elemental signatures reflecting spatial gradients in coastal contamination [15,16]. Jellyfish have also been proposed as a global-scale bioindicator for plastic pollution [17], with microplastic ingestion documented in 97% of *Pelagia noctiluca* individuals in the North Atlantic [18] and 42% of jellyfish across three species in a tropical estuary [19]. Acute copper and silver toxicity have been documented in *A. aurita* polyps, with both metals impairing budding, strobilation, and survival at coastal-relevant concentrations [20], challenging the historical view of jellyfish as pollution-tolerant.

The most attractive feature of *A. aurita* for vision-based BEWSs is its characteristic rhythmic bell pulsation: the contractile cycle driven by a distributed nervous system provides a readily observable, continuously measurable physiological signal that has been shown to alter significantly in response to a range of chemical stressors. The pulsation frequency Fp has been systematically validated as a sensitive ecotoxicological endpoint. Faimali et al. [9] standardized an *A. aurita* ephyra bioassay against cadmium nitrate and SDS, confirming that Fp is easily measurable, dose-dependent, and complementary to acute immobilization. Costa et al. [10] subsequently demonstrated dose-dependent Fp alterations in *A. aurita* ephyrae under exposure to neurotoxic compounds (eserine and chlorpyrifos), with EC50 values indicating greater sensitivity than for many other marine invertebrates used in ecotoxicology. Sensitivity to microplastic of *A. aurita* ephyrae

pollution was established by Costa et al. [11] for polyethylene and by Di Giannantonio et al. [12] for PVDF and PLA, and to harmful microalgae by Yang et al. [21]. Most recently, Mercado et al. [22] evaluated food-additive toxicity using *A. aurita* ephyrae alongside other model invertebrates, finding compound- and species-specific Fp responses that varied non-linearly with concentration. The response to petroleum-related compounds is particularly relevant to the present study: Echols et al. [23] showed that water-accommodated fractions (WAF) of Macondo crude oil produced no acute mortality in *A. aurita* ephyrae at environmentally realistic loadings, whereas chemically enhanced WAF (CEWAF, prepared with Corexit 9500 dispersant) elicited muscle-contraction abnormalities and reduced pulsation. The neurobehavioural signal in that study cannot be attributed unambiguously to the petroleum fraction alone, as the dispersant itself contributes to toxicity. The present pilot uses neat diesel without an added dispersant, so the petroleum-fraction effect can be evaluated independently of dispersant-mediated bioavailability. Life stage and exposure concentration both modulate sensitivity strongly: *A. aurita* polyps were unaffected by chronic exposure to atrazine and chlorpyrifos at guideline concentrations [24], whereas combined exposure to microplastics and tetracycline produced synergistic apoptosis and irreversible oxidative damage [25]. The bulk of the published Aurelia ecotoxicology has used the ephyra stage, which combines high standardization with a documented sensitivity that often exceeds that of post-metamorphic stages [9–12]. Direct comparison of ephyra-based EC_{50} values with behavioral responses of medusa-stage animals, including the present pilot, is therefore informative to the direction of effect but not to threshold concentration.

The locomotion and bell kinematics of *A. aurita* have been characterized in detail from a biomechanical perspective. McHenry and Jed [26] measured swimming kinematics across ontogenetic size classes, finding that *A. aurita* decreases pulse frequency with growth while maintaining kinematic similarity in thrust generation. Their work established that pulse frequency is the primary determinant of swimming speed and is strongly linked to body size—an observation with direct consequences for monitoring system design, since the same physiological signal can be measured directly from a side-view bounding box. Bajcar et al. [27] used computer-aided video analysis to measure the kinematic properties of bell motion, reporting that bell relaxation is longer than contraction, and that subumbrellar cavity volume changes by up to 50% over the pulsation cycle. A direct precedent for using AI-based tracking to measure *A. aurita* pulsation is provided by Boyd et al. [28], who developed the Jellyfish Tracking and Analysis (JeTA) pipeline using YOLOv5m to identify, track, and quantify pulsation of free-swimming medusae, and observed a significant increase in pulsation rate in response to Artemia extract. The principle of using bounding-box dimensions as a physiological proxy has also been applied successfully outside marine biology, e.g., in livestock behaviour monitoring with YOLOv5 plus DeepSORT [29], confirming the soundness of the approach.

The application of deep learning to jellyfish detection has accelerated considerably over the past decade. The most directly comparable system is Jellytoring [30], which achieved an F1 of 95.2% on real-time *A. aurita* detection and explicitly identified the potential of such pipelines as the basis for a jellyfish-based early warning system. The YOLO family of architectures has been the dominant approach in jellyfish-specific detection studies: JF-YOLO improved bloom detection over a YOLOv4 baseline [31], an enhanced YOLOv4-tiny with CBAM attention reached 95.0% accuracy at 223 FPS [32], an optimized YOLOv3 with underwater image preprocessing improved detection accuracy across seven jellyfish species including Aurelia [33], an improved YOLOv5-nano with GAM and CoordConv reached 89.1% mAP@0.5 with model-assisted labelling [34], and YOLOv8 has been shown to outperform classical CNNs for jellyfish classification [35]. For underwater video specifically, dedicated architectures such as UWV-Yolox [36] and Arctic deep-sea pipelines [37] address

blur, contrast, and species-discrimination challenges. Tracking frameworks for biological video are also mature, with Lopez-Marcano et al. [38] tracking 120 of 169 individual fish in field video using a combination of MOSSE, Seq-NMS, and SiamMask. Collectively, these results establish that YOLO-based detection of Aurelia is a mature and reliable basis on which to build downstream behavioural analysis.

The four pollutant classes most relevant to operational marine BEWS (petroleum hydrocarbons and PAHs, heavy metals, pesticides, and microplastics) each have documented effects on jellyfish physiology or on comparable aquatic invertebrates. PAHs are reviewed in Honda and Suzuki [39], who documented endocrine disruption and tissue-specific toxicity, and noted the emerging concern of PAH–microplastic interactions as a vector for enhanced bioaccumulation. For *A. aurita* ephyrae specifically, chemically dispersed Macondo crude oil fractions produced visible muscle-contraction abnormalities [23]. Heavy metals affect aquatic invertebrates through metabolic dysfunction, oxidative stress, and neurotoxicity, with synergistic interactions amplifying toxicity in mixed exposures [40]. Behavioural endpoints are particularly sensitive sub-lethal indicators, as demonstrated by video-based detection of locomotor decreases in marine shrimp at metal concentrations of ~1 ppm within 3 h [41]. Pesticide contamination from agricultural runoff is a major concern: a meta-analysis of pesticide effects on aquatic vertebrate swimming behaviour reported reductions of 35% in swim speed and 72% in overall activity at environmentally relevant concentrations [42], and the organophosphate chlorpyrifos has been shown to alter *A. aurita* Fp dose-dependently at coastal-relevant concentrations [10]. Microplastics impair feeding and swimming in *A. aurita* ephyrae [11] and act as vectors for co-occurring contaminants such as PAHs and heavy metals [43]. Chronic polystyrene microbead exposure of *Rhopilema esculentum* produced growth inhibition and increased oxygen consumption after 15 days, with implications for bell-contraction energetics [44].

Extracting meaningful pollution signals from continuous time series of biological movement metrics requires robust computational methods. Classical and deep machine-learning approaches have both been applied: SVM and ANN have been shown to be more resilient to data imbalance than deep methods on real-world water-quality time series [45], although deep methods generally outperform classical ones in feature-learning accuracy [46,47]. Unsupervised methods are particularly attractive for BEWS because they do not require labelled contamination events for training: Grekov et al. [48] demonstrated that Isolation Forest, elliptic envelope, and local outlier factor could all achieve F1 = 1.0 on bivalve activity data with appropriate hyperparameter tuning, with Isolation Forest providing the fastest detection. Frameworks combining regression-, feature-, and rule-based methods are recommended to maximize detection of different anomaly types while minimizing false positives [49].

In this study we exploit these foundations to develop and validate an end-to-end, unsupervised-at-inference-time (operating without labelled pollution data) BEWS pipeline for marine pollution detection that uses *A. aurita* as the sentinel organism. The pipeline takes raw aquarium video as input, derives a continuous bell-area signal and centroid trajectory from YOLO bounding-box detections, extracts eleven pulsation, kinematic, and detection-quality features on 60 s sliding windows, and applies a per-animal two-layer anomaly detector that combines fast mean-$|z|$ outlier detection with one-sided CUSUM drift accumulation. The principal contributions of this work are: (i) empirical evidence that pulsation-frequency direction of effect is animal-specific under diesel exposure, motivating per-animal anomaly detection rather than supervised classification; (ii) a compact, ML-weight-free detection algorithm with calibrated thresholds derived entirely from clean-water windows; and (iii) proof-of-concept validation on six adult medusae exposed to nominal diesel-WAF, with all six animals detected and no false alarms observed across

203 clean-window opportunities, interpreted together with the confidence intervals reported in Section 3.1. The remainder of the paper is organized as follows. Section 2 describes the experimental design, the detection pipeline, the anomaly detector, and the threshold-calibration and validation protocols. Section 3 presents the validation results across the six pilot animals. Section 4 discusses the implications, limitations, and future directions, and Section 5 concludes the study.

2. Materials and Methods

This section describes the end-to-end pipeline that transforms raw aquarium video of *A. aurita* into a binary pollution alarm. The pipeline has five stages: (i) controlled video acquisition in clean and polluted water with logged injection timestamps; (ii) per-frame detection and tracking with a YOLO model trained on a custom dataset; (iii) signal preprocessing that reconstructs a continuous bell-area time series and removes slow drift; (iv) extraction of a compact set of pulsation, kinematic, and detection-quality features on overlapping temporal windows; and (v) an unsupervised two-layer anomaly detector operating on per-animal baselines. The first four stages are conventional and are described in Sections 2.1–2.4. The fifth stage, which is the main methodological contribution of this work, is described in Sections 2.5–2.7. Section 2.8 summarizes the implementation.

2.1. Experimental Design and Dataset

A. aurita were collected from the coastal zone of Sevastopol (Black Sea) on 1 October 2025 and 24 October 2025 at a water temperature of approximately 19–21 °C. A total of 25 individuals were collected; 22 survived the initial inspection. Twelve individuals were selected for experimental recordings on the basis of biological and technical suitability for side-view video analysis: intact bell morphology, absence of visible mechanical damage, spontaneous pulsation, active swimming, and a bell size compatible with reliable detection within the camera field of view.

Recordings were performed in filtered natural Black Sea water under controlled laboratory conditions. The water temperature was approximately 19–21 °C both at collection and during experimental recording. Camera position and illumination were kept constant within each recording session. Illumination was provided by overhead fluorescent room lamps and aquarium lighting (IEK, Moscow, Russia and Barbus, Moscow, Russia), without direct sunlight. Light intensity was not quantified in lux during this pilot study. Salinity corresponded to natural Black Sea water, approximately 18‰. Dissolved oxygen was not continuously logged during the recordings; therefore, any potential effects of short-term variation in oxygenation are acknowledged as a limitation in Section 4.5.

Before recording, each individual was transferred to a 6 L experimental aquarium containing filtered Black Sea water and allowed to acclimate for approximately 1 h. This acclimation period was intended to allow recovery from handling and transfer before video recording. Recording was started only after the focal animal had resumed regular bell pulsation, showed stable swimming, and could be reliably tracked in side view.

The final analytical dataset consisted of six technically usable recordings. For these six individuals, bell diameter ranged from 1.5 to 4.6 cm, with a mean of 3.1 ± 1.1 cm (mean ± SD, n = 6). Per-animal bell diameters are reported in Supplementary Materials, Table S1. Recordings were excluded only for technical reasons, such as insufficient baseline, poor focus, the medusa leaving the camera frame, or major tracking interruptions. Exclusion was not based on the observed biological response to diesel-WAF exposure or on the anomaly detection result. The specific reason for excluding each recording is provided in Supplementary Materials, Table S1.

Pollutant preparation. The exposure medium was prepared as a water-accommodated fraction (WAF) of commercial diesel fuel. Diesel fuel (8 mL) was added to 200 mL of natural seawater and briefly mixed to promote contact between the fuel and seawater, corresponding to a 1:25 v/v diesel-to-seawater loading ratio. This preparation is based on the general WAF approach used in petroleum-toxicity studies, including previous work on *Aurelia aurita* ephyrae exposed to oil WAFs [23], but was adapted to the smaller experimental volume and proof-of-concept design of the present pilot study. The mixture was then left to equilibrate for 1–12 h, after which the aqueous phase was collected and used for the exposure experiments. The WAF used in this pilot study was not analytically characterized by gas chromatography–mass spectrometry (GC–MS) or by total petroleum hydrocarbon (TPH) analysis. Accordingly, the exposure is described as a nominal diesel-WAF exposure prepared under the specified conditions, rather than as exposure to an analytically verified hydrocarbon concentration. The previously estimated range of 100–350 mg L^{-1} is retained only as an approximate preparation-based estimate and should not be interpreted as a measured hydrocarbon concentration in the exposure aquaria.

Recording protocol and selection of usable records. Each animal was recorded continuously in side view at 1920 × 1080 resolution and 30 frames per second, with the camera held stationary throughout. A clean-water baseline $\geq$15 min (preferably $\geq$30 min) was targeted (in this pilot three animals (ExpID 2, 6, 5) had shorter clean phases of 4, 7.5 and 8 min, respectively (Table S1), followed by exposure to the WAF preparation and continued recording for at least 30 min. Twelve animals were recorded under this protocol across the experimental series of October–November 2025. Six recordings (ExpID 1–6) met the technical requirements for the YOLO-based pipeline used in this study: sufficient duration of the clean-water baseline, the focal animal remaining in focus and within the camera frame during the recording, and the absence of large tracking dropouts that would compromise feature extraction. The remaining six recordings were excluded from the present analysis on technical grounds, including insufficient recording or baseline duration, loss of focus, the animal moving out of the frame for substantial parts of the recording, large tracking dropouts, or operator disturbance affecting the baseline.

The technical-quality criteria were defined a priori and applied at the acquisition/tracking stage, before and independently of anomaly detection analysis. No recording was included or excluded on the basis of the animal's biological response to diesel-WAF exposure or on the basis of the detector output. The specific technical reason for each excluded recording is provided in Supplementary Materials, Table S1.

The dataset consists of recordings of six adult medusae (ExpID 1–6) in a controlled laboratory aquarium with a stationary side-view camera (Nikon Corporation, Tokyo, Japan) at 1920 × 1080 resolution and 30 frames per second. Each experiment follows a paired within-subject protocol: a given individual is first recorded in clean seawater to establish a behavioural baseline, then the pollutant is introduced into the same aquarium while recording continues, and the post-injection behaviour of the same animal is followed for at least 30 min. Diesel fuel was chosen as the first-phase pollutant because its active fractions (petroleum hydrocarbons and polycyclic aromatic hydrocarbons) are among the most widespread contaminants of coastal waters and are known to disrupt jellyfish neuromuscular function at sub-lethal concentrations [23,39]. The pilot dataset comprises 203 clean-water 60 s windows and 225 diesel-phase windows, distributed across the six individuals.

Two protocol details are critical and are enforced in all recordings used for the analysis below. First, the clean-water baseline is acquired continuously for at least 15 min, preferably 30 or more, because per-animal z-score normalization (Section 2.5.2) relies on a stable estimate of the baseline mean and standard deviation. In our pilot data, short baselines

(e.g., 4 min for one individual, for details see Table S1) produced inflated z-scores and inconsistent behaviour of the detector. Second, the absolute time of pollutant injection was logged manually by the operator to the nearest second from the synchronized video timestamp. This is essential because the response to a toxicant is time-varying: in one of the pilot animals (ExpID 3) we observed a stimulation phase lasting approximately 17 min post-injection, followed by a depression phase, which is invisible without a precise injection timestamp. The injection timestamp is attached to every downstream feature window as the field t_inj_s, which is zero at injection, negative before, and positive after.

Video frames during which the camera was physically disturbed, for example, while pouring pollutant into the aquarium or correcting the focal animal's position, cause false centroid displacements and must be excluded. These intervals are logged manually, and any 60 s feature window overlapping a disturbance is marked with an artefact flag and dropped from both calibration and scoring.

The YOLO v11 detector was trained on a custom-labelled image set. Frames were sampled from the experimental videos using an in-house Python v3.11 script that saved one frame every 30 s, yielding 478 candidate images. Annotation was performed in Roboflow: each visible jellyfish was outlined with a polygon mask using the platform's built-in tools, and frames in which no jellyfish could be identified were discarded. The annotated dataset was split at the video level into training, validation, and test subsets in a 70/10/20 ratio. All frames extracted from a given source video were assigned to only one subset; therefore, no frames from the same recording were shared across training, validation, and test sets. This video-level split was used to avoid temporal leakage and to provide a more conservative estimate of detector performance. A representative annotation is shown in Figure 1.

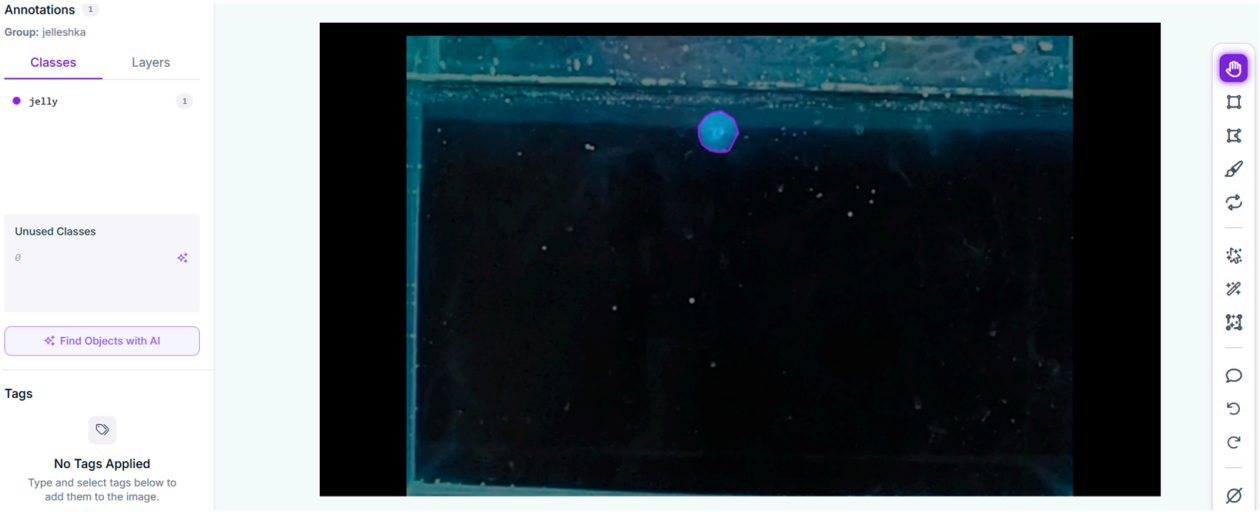

Figure 1. Example of image annotation in Roboflow's dataset labelling interface. The left panel displays the annotation classes, where the object class "jelly" is selected. The highlighted circular region in the central part of the image indicates the annotated jellyfish region of interest marked by the user.

The annotated dataset is available on the Roboflow platform.

2.2. Detection and Tracking

Per-frame detection is performed by a YOLO object detector trained on a custom-labelled dataset of *A. aurita* images. The detector used in this work is YOLOv11 (Ultralytics, Frederick, MD, USA, and London, UK, released 2024), although the pipeline is architecturally agnostic and earlier YOLO versions can be substituted without changes downstream of the detector. For each frame, the model emits a single axis-aligned bounding box char-

acterized by pixel coordinates (x1, y1, x2, y2), a centroid (cx, cy), the pixel area of the box, and a detection confidence score in [0,1]. These quantities, together with the frame index and timestamp, constitute the entire input to all downstream processing. No additional image-level information is propagated beyond the detector. This design choice keeps the pipeline lightweight, portable, and storage-efficient in field deployments, the raw video need not be retained after detection, and is analogous to the JeTA pipeline of Boyd et al. [28], which demonstrated that bounding-box dimensions alone are sufficient to detect biologically meaningful changes in pulsation frequency. Single-target tracking is used throughout: each aquarium is monitored for one focal jellyfish at a time, so no data-association step is required and consecutive detections are treated as a contiguous trajectory (Figure 2).

A B C

Figure 2. Annotated YOLO frames: (**A**)—typical detection, (**B**)—jellyfish merged with the rib of the aquarium (failure mode), (**C**)—the reflection has higher confidence than the jellyfish itself (failure mode), YOLO detector validation. Detector performance was evaluated using the video-level test split described in Section 2.1, so that frames from the same recording were not shared between training and testing. This reduced the risk of overestimating performance due to visually similar consecutive frames. On this test set, the YOLO detector achieved precision = 0.89, recall = 0.90, mAP@0.5 = 0.92, and mAP@0.5:0.95 = 0.72. The test-set confusion matrix contained 42 true positives, 7 false positives, and 3 false negatives. False positives corresponded mainly to background structures or low-contrast regions detected as jellyfish-like objects, whereas false negatives occurred primarily during partial bell occlusion or in low-contrast frames. The main tracking-related failure modes were missed detections during strong bell occlusion and bounding-box instability when the animal approached the edge of the frame. The YOLO detector confusion matrix and confidence-score distribution on the video-level held-out test split are provided in Figure 3.

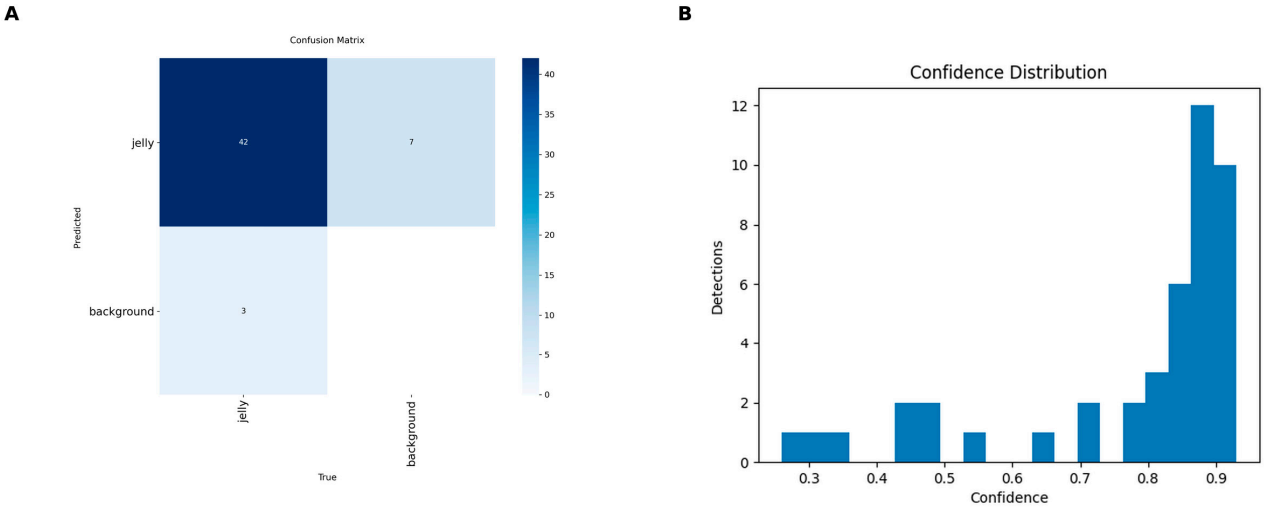

Figure 3. YOLO detector validation on the video-level held-out test split. (**A**): Confusion matrix. (**B**): Confidence-score distribution.

Owing to the small video-level test set, these values constitute preliminary detector validation on this dataset, not evidence of robustness under operational field conditions.

2.3. Signal Preprocessing

Two signal-quality issues are handled before feature extraction. First, YOLO occasionally fails to detect the medusa; across the pilot data, the detection rate varies from 95.8% to 100%. Frames in which all bounding-box coordinates are missing are treated as detection gaps and linearly interpolated in the area and centroid signals, up to a maximum permitted gap length of 1 s. Windows whose total dropout fraction exceeds 20% are flagged as unreliable and excluded from subsequent analysis.

Second, the raw bell-area signal contains a slow drift of tens of seconds' duration that reflects the animal swimming closer to or further from the camera, rather than any physiological pulsation. This drift is estimated by a Savitzky–Golay low-pass filter (window 15 s, polynomial order 2) and subtracted, yielding a zero-mean AC signal that contains only the contractile cycle. A third-order Butterworth band-pass filter (0.1–2.0 Hz, applied forward and backward to avoid phase distortion) further restricts the signal to the physiologically plausible range of *A. aurita* pulsation [26]. Contraction events are detected as prominent negative peaks in the band-passed signal, corresponding to minima of bell area, i.e., the moments of peak contraction, subject to a minimum inter-event spacing of 0.4 s and a prominence threshold of one-half of the signal's standard deviation. The effect of these steps is illustrated in Figure 4.

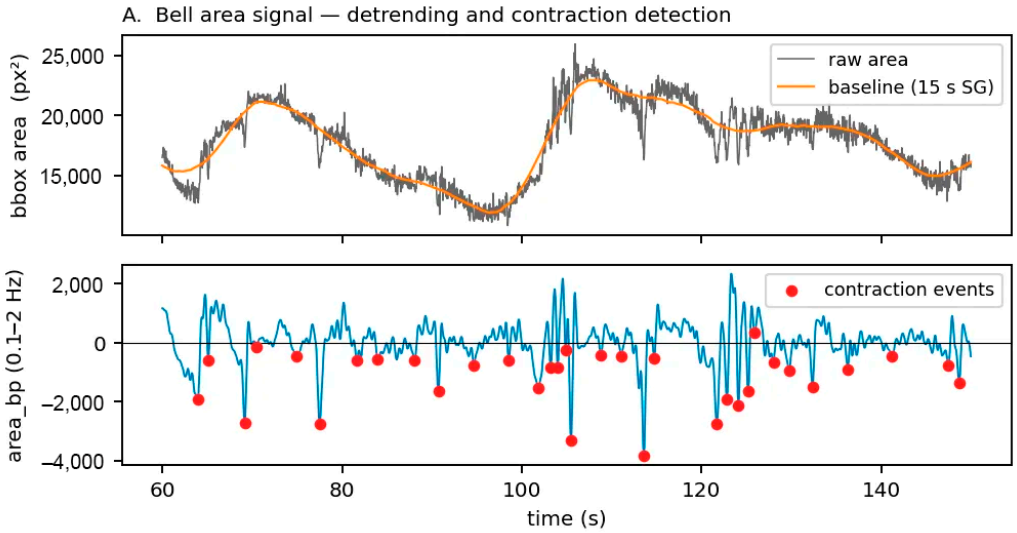

Figure 4. Preprocessing of the bell-area signal on a 90 s excerpt. (**Top**): Raw bounding-box area (grey) and 15 s Savitzky–Golay baseline (orange) that captures non-physiological drift caused by the animal swimming toward or away from the camera. (**Bottom**): Band-passed signal (0.1–2.0 Hz) with detected contraction events marked in red.

2.4. Feature Extraction on Overlapping Temporal Windows

Classification operates on 60 s windows with 50% overlap (stride 30 s). The 60 s length is a compromise between temporal resolution and statistical robustness: at a typical Fp of 0.5–1.3 Hz, a 60 s window contains 30–80 contraction events, which is sufficient to estimate both the mean and the coefficient of variation in the inter-pulse interval. Eleven features are computed per window (Table 1): four pulsation features from the band-passed area signal, five kinematic features from the centroid trajectory, and two detection-quality features. The first nine enter the anomaly detector. The last two are used to gate windows before scoring.

Table 1. Feature catalogue. Features marked "direction-consistent" had the same sign of effect under diesel exposure in every animal examined in the pilot. Other features had animal-specific directions of effect but contributed consistently to the aggregated magnitude metric mean $|z|$ used by the detector.

Group	Feature	Definition	Role in Detector
Pulsation	F_p	1/median(IPI) from contraction events	Primary ecotoxicological endpoint [17]; direction of effect is animal-specific; contributes to mean $|z|$
Pulsation	IPI CV	std(IPI)/mean(IPI)	Rhythm regularity; contributes to mean $|z|$
Pulsation	Relative amplitude	$(P_{95}(\text{area}) - P_5(\text{area}))/P_{95}(\text{area})$	Contraction depth
Pulsation	Spectral entropy	Normalized Shannon entropy of Welch PSD of area_bp	Rhythm complexity; rises under arrhythmia
Kinematic	Mean speed	Mean frame-to-frame centroid displacement/dt	Locomotor activity; direction of effect is animal-specific
Kinematic	Speed CV	std(speed)/mean(speed)	Burstiness of swimming
Kinematic	Tortuosity	Total path length/net displacement	Direction-consistent (rises under diesel in all pilot animals)
Kinematic	Vertical bias	Mean vertical velocity of centroid	Surface-/bottom-seeking tendency
Kinematic	Immobility	Fraction of frames with speed < bell_diam/5 s	Direction-consistent (rises under diesel in all pilot animals)
Quality	Mean YOLO confidence	Mean of conf over detected frames	Detection reliability; excluded from z-scoring but monitored
Quality	Dropout fraction	Fraction of frames with no valid detection	Tracking continuity; windows with >20% dropout are dropped.

2.5. Detection Architecture

2.5.1. Motivation for an Unsupervised, Per-Animal Design

The initial design of this study called for a supervised classifier trained on labelled clean-vs-polluted windows, as is standard in early BEWS work [7,45]. Pilot-data analysis on the six individuals exposed to diesel revealed three obstacles that rule out a pure supervised approach.

(i) Pulsation frequency is not a universal marker of pollution. Clean-water baseline Fp differs substantially between individuals (from 0.77 Hz for one animal to 1.01 Hz for another, a 30% range), and the direction of the diesel-induced shift is not the same across individuals. In the extreme, one animal showed an effect size of Cliff's $\delta = +0.94$ (strong depression) while another showed $\delta = -0.07$ (no effect), with a third showing a biphasic pattern of stimulation followed by depression. Figure 5 (top) makes this point quantitatively: baseline Fp distributions differ, and Figure 5 (bottom) shows that when trajectories are aligned to the moment of injection, animals follow qualitatively different paths in Fp space.

(ii) Biphasic responses place the same class on opposite sides of the feature space. In one well-characterized animal, post-injection behaviour for the first 17 min was an upward shift in Fp and locomotion (stimulation), and after minute 17 it was a downward shift (depression). A supervised classifier attempting to learn a single hyperplane separating "diesel" from "clean" must accommodate both regions simultaneously, and in practice collapses to majority-class prediction. Leave-one-video-out evaluation of a logistic-regression classifier on this animal gave 34% accuracy, well below the 50% majority-class baseline.

(iii) BEWS field deployment requires open-world detection. Unknown pollutants in natural waters will produce behavioural signatures different from any that have been labelled. A detector that flags "anything sufficiently unlike clean water for this animal" is inherently more robust than one that learns "here is what diesel looks like" [47,48].

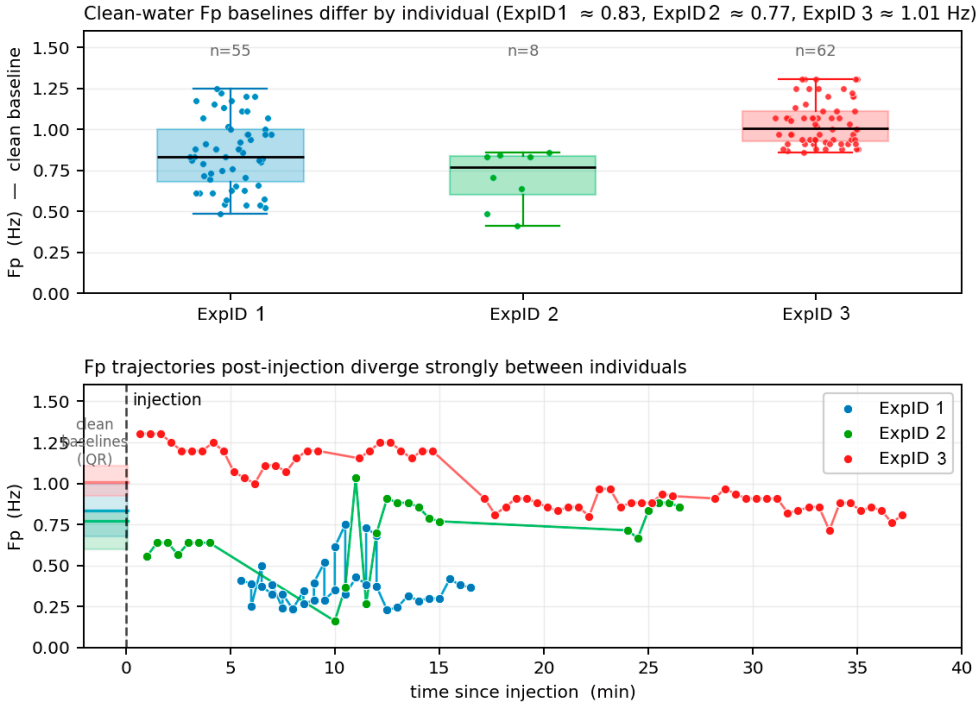

Figure 5. Individual variability in the pulsation response to diesel. (**Top**): Clean-water Fp baselines for three animals of the same species differ by approximately 30%, establishing that absolute Fp thresholds are not transferable between individuals. (**Bottom**): When diesel-phase Fp trajectories are aligned to the moment of injection, animals follow qualitatively different paths—sustained depression (blue), transient depression followed by return to baseline (green), and biphasic stimulation-then-depression (red). This heterogeneity is the primary reason supervised classification of "clean vs. polluted" is not viable on this dataset and motivates a per-animal anomaly detector.

A fourth observation suggests the form of an unsupervised alternative: although the direction of per-feature effects varies across individuals, the magnitude of the multivariate departure from the clean distribution is consistently elevated under diesel exposure. Specifically, the mean absolute z-score averaged across the nine behavioural features (mean$|z|$) is elevated in every animal under diesel relative to its own clean baseline. This observation motivates a detector based on the magnitude of deviation from a per-animal baseline, which is the construct developed in the remainder of this section.

The anomaly detector was developed as part of this pilot study. The selected behavioural features, the aggregation of feature-wise deviations into mean$|z|$, the two-layer structure, the three-window rule in Layer 1, and the 1.3× factor applied to the CUSUM threshold were chosen during exploratory analysis of the six pilot recordings. The detector is unsupervised in the sense that it does not require labelled pollution-response data; during analysis, each animal is compared only with its own clean-water baseline. However, because the detector design was developed using the pilot dataset, its performance and transferability require confirmation on independent data. This limitation is discussed in Section 4.5.

2.5.2. Per-Animal Baseline Calibration

For each animal, the first 15 or more min of clean-water recording constitute a calibration set from which a baseline object is fitted. This baseline consists of the mean μi and standard deviation σi of each feature i over the clean windows. During monitoring, every new window is transformed to a vector z of z-scores $zi = (xi - \mu i)/\sigma i$, and the scalar mean$|z| = \text{mean}(|zi|)$ is computed and passed to the two detection layers described below. This per-animal normalization neutralizes the inter-individual variability docu-

mented in Section 2.5.1: a diesel-induced 30% drop in Fp that is indistinguishable from noise in absolute Hz becomes a z-score of -2 or -3 relative to the same animal's own clean distribution.

The 15 min lower bound on the calibration period is dictated by two considerations. Empirically, within-clean variability is non-trivial: in one pilot animal, Fp within a single 22 min clean recording ranged from 0.5 to 1.3 Hz over 43 consecutive windows. The 4 min baseline of ExpID 2, the shortest clean-water baseline in the pilot dataset, illustrates this limitation: a short clean-water segment may sample only part of the underlying behavioural variability and thereby underestimate σ, producing inflated z-scores later. Theoretically, the standard-error scaling of the mean as $1/\sqrt{n}$ implies that at least 30 effectively independent windows are needed to estimate μ within a quarter of its own standard deviation. With a 50% overlap between consecutive 60 s windows, the autocorrelation between adjacent windows reduces the effective sample size by approximately a factor of two; we therefore adopt 30 overlapping windows (corresponding to 15 min of clean recording at 30 s stride) as a practical minimum, acknowledging that this is a conservative lower bound and that longer baselines further improve the precision of the per-animal μ and σ. The ExpID 2/5/6 calibration set comprised all available clean windows (8/16/15), fewer than the recommended ~30.

2.5.3. Two Detection Layers

The mean $|z|$ stream feeds two parallel detection layers (Figure 6). They target two distinct anomaly regimes, discrete step changes and sustained distributional drifts, and the union of their alarms constitutes the detector output. An earlier prototype included a third layer based on Isolation Forest novelty in the full z-vector space [48], intended to detect unusual feature combinations that would not show up as a large mean $|z|$. Empirical evaluation on five individuals showed that this layer produced zero unique true positives, every window it flagged was already flagged by Layer 1, because the toxicological response to diesel is monotonic along the dominant axis already captured by mean $|z|$. The Isolation Forest layer was therefore removed. The simplified two-layer architecture has the same sensitivity at lower implementation complexity and is reported here.

Layer 1—fast outlier detection. Layer 1 thresholds mean $|z|$ against a calibrated value Tz. A step-change anomaly, in which one or more features shift abruptly (as observed within 30 s of diesel injection in two of the six pilot animals), produces a mean $|z|$ above Tz within a single window. To suppress isolated outliers caused by tracking glitches, the layer requires the threshold to be exceeded for k consecutive windows before raising an alarm. In the present implementation k = 3 windows: with a 60 s window and 30 s stride, three consecutive triggered windows span 120 s from the start of the first window to the end of the third (60 s + 2 × 30 s = 120 s). At the validation threshold Tz = 1.559, Layer 1 identifies time windows in which mean $|z|$ is unusually high. Because neighbouring windows overlap by 50%, consecutive mean $|z|$ values are positively autocorrelated and are not statistically independent: a high value in one window can also affect the adjacent window. For this reason, we did not calculate a theoretical false-alarm probability for the k = 3 rule. Instead, we evaluated the complete k = 3 rule directly on the clean-water data. No Layer-1 alarms were produced in the pooled clean-water windows (0/203 clean-window opportunities).

Layer 2—drift detection via CUSUM. Layer 1 does not reliably detect a distributional shift in which no individual window is extreme but the median of the signal is noticeably elevated. In three of the six pilot animals, the diesel response takes exactly this form: mean $|z|$ shifts from a clean median around 0.7 to a diesel median around 1.0, without

any window exceeding Tz. To detect this pattern, Layer 2 applies a one-sided cumulative sum [50] to the mean $|z|$ stream:

$$S_i = \max(0, S_{i-1} + (\text{mean}\,|z|_i - \text{target})) \quad (1)$$

where target is the median mean $|z|$ observed on pooled clean-water calibration data.

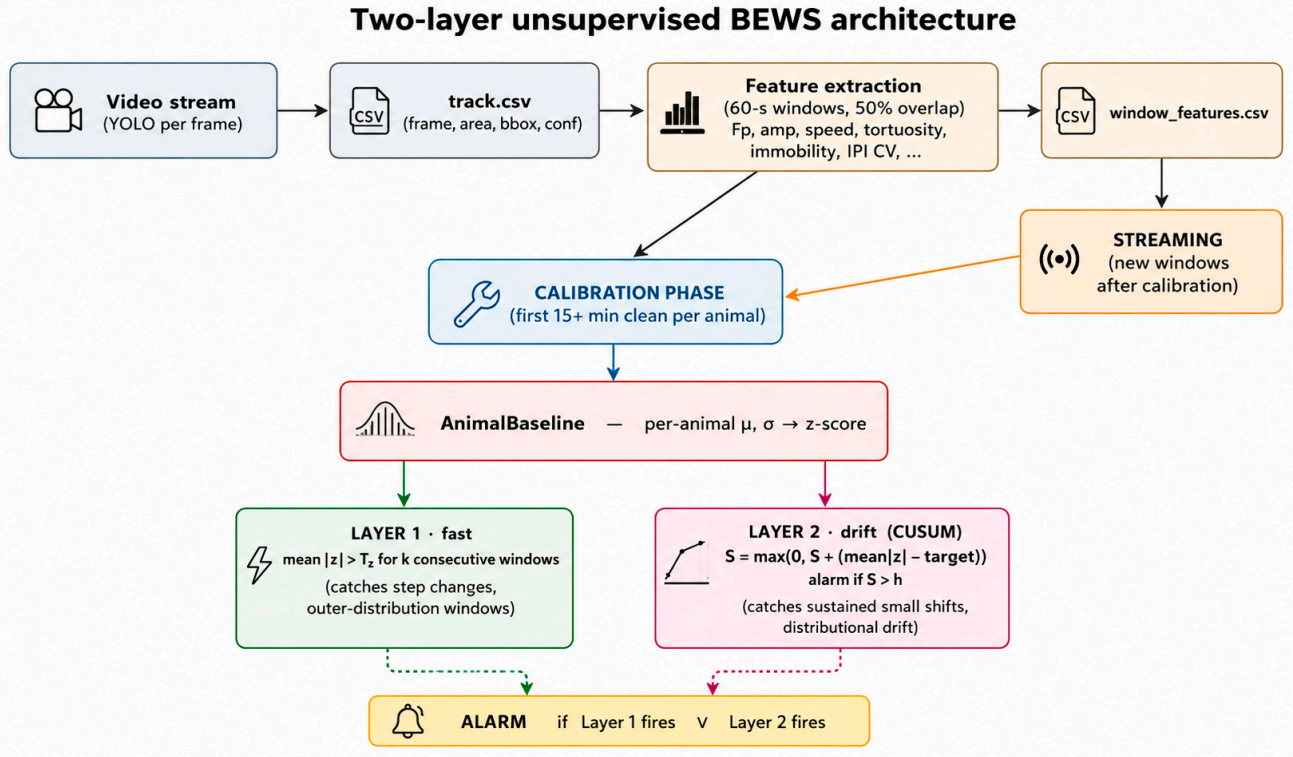

Figure 6. Two-layer unsupervised BEWS architecture. A calibration phase (clean-water recording; recommended ≥15 min) fits a per-animal AnimalBaseline object consisting of feature-wise μ and σ for z-score normalization. Incoming windows are scored against this baseline and evaluated by two parallel layers: Layer 1 thresholds mean $|z|$ (with a k-consecutive rule) for step changes and outlier windows; Layer 2 applies a one-sided CUSUM to detect sustained small distributional shifts. An alarm fires when either layer's condition is met.

The CUSUM reference value was set to the median of the clean-water mean $|z|$ distribution because mean $|z|$ is bounded below by zero and can be right-skewed when occasional clean windows contain larger behavioural deviations or tracking noise. The median therefore provides a robust estimate of the in-control central tendency. The reference value affects the in-control average run length (ARL_0): a lower reference increases sensitivity but can shorten ARL_0, whereas a higher reference is more conservative but may delay detection. In the present dataset, the mean and median references were nearly identical. A robustness check using the mean clean-water reference value (target = 0.714) instead of the median reference value (target = 0.715) produced the same outcome: 6/6 animals detected and 0/203 clean-window false alarms. The CUSUM Si accumulates whenever mean $|z|$ exceeds this reference, resets to zero otherwise, and triggers an alarm when it exceeds a threshold h calibrated (Section 2.6) to be higher than any CUSUM value reached during any calibration-animal clean period. A one-sided positive CUSUM is appropriate because we expect pollution exposure to increase, not decrease, the animal's departure from its clean baseline; no two-sided construction is needed. The reset after an alarm permits the detector to re-fire if the exposure continues. No k-consecutive rule is needed on Layer 2 because the CUSUM construction already implements temporal accumulation.

2.6. Threshold Calibration

The two thresholds Tz and h, together with the CUSUM target, are determined entirely from clean-water windows and are frozen before any test data is scored. mean$|z|$ is computed for each clean window of each calibration animal, Tz is set to the 99th percentile of the pooled clean-window distribution, and target is set to the median of the same pooled distribution. The CUSUM threshold h is determined by running the CUSUM recursion on each calibration animal's clean stream and recording the maximum value reached. h is then set to 1.3 times this maximum (safety factor), which guarantees that no calibration animal produces a false alarm on its clean period and leaves approximately 30% headroom for inter-animal variation. The safety factor was chosen to balance two competing concerns: too low a factor produces false alarms on new animals whose clean baselines are noisier than the calibration distribution; too high a factor increases detection latency on real pollution events.

Two threshold sets are reported, corresponding to two different calibration regimes. The validation thresholds are obtained by fitting on three calibration animals (ExpIDs 1, 2, 3) and reserving three animals (ExpIDs 4, 5, 6) for held-out testing. They underpin the validation results in Section 3. The production thresholds are obtained by fitting on all six animals and are intended for field deployment, where any new animal will be scored against the most stable threshold estimate available from the pooled pilot data. Both sets are reported in Table 2. The production thresholds are slightly more conservative because the additional animals (4, 5, 6) reveal a wider range of clean-baseline variability than the original three.

Table 2. Calibrated thresholds under the validation regime (3 calibration/3 held-out animals) and the production regime (all 6 animals used for calibration). Differences are within ±5% for Tz and target and approximately +11% for h. The latter is driven primarily by ExpID 4, whose clean-water mean$|z|$ reaches a slightly higher peak than any of the original three calibration animals.

Threshold Set	Calibration Animals	Tz	Target	h
Validation	ExpIDs 1–3 (n = 125 windows)	1.559	0.715	5.498
Production	ExpIDs 1–6 (n = 203 windows)	1.591	0.711	6.117

2.7. Validation Protocol

Validation uses a held-out-animals protocol in which a subset of individuals is reserved entirely from threshold calibration. In the present analysis, three of the six animals (ExpIDs 4, 5, 6) form the held-out test set and the remaining three (ExpIDs 1, 2, 3) form the calibration set. The test animals' own per-animal baselines (μ, σ) are still fitted on their own clean windows; this is permitted and indeed required by the per-animal design, and the validation thresholds (Tz = 1.559, target = 0.715, h = 5.498) are then applied to the test animals' diesel-phase windows. Because calibration uses only the calibration animals' clean data and each animal's baseline is fitted only on its own clean data, there is no leakage of pollution-response information between calibration and test. This protocol mirrors the operational deployment scenario exactly: a new animal placed in a monitoring aquarium contributes its own 15 min clean calibration and is then monitored against globally calibrated thresholds.

Three metrics are reported per animal. The false-positive rate on clean windows is the fraction of clean test windows that raise an alarm, and should be zero or very close to zero. The detection outcome on diesel is a binary yes-or-no at the animal level: the detector is credited with detection if any alarm fires during the diesel phase. The detection latency is the time between pollutant injection and the first alarm, in minutes. This is the

operationally most important metric for a BEWS, because a longer latency means more pollutant enters a receiving water body before action is taken.

2.8. Implementation and Reproducibility

The pipeline is implemented in Python 3 using NumPy, SciPy, and pandas, peak detection and filtering use scipy.signal. The detector itself has no machine-learning dependencies. The implementation is split between two modules: pipeline.py (approximately 160 lines) ingests a single track.csv produced by the YOLO detector and emits a window_features.csv table of per-window features, and bews.py (approximately 130 lines) defines the AnimalBaseline, BEWSDetector, and calibrate_thresholds constructs. The total code footprint for feature extraction and detection is well under 300 lines, which is significant for field deployment: the whole pipeline can be audited by a domain scientist without ML expertise, and it contains no neural-network weights that would have to be shipped and maintained. On a standard laptop, the feature extraction runs at roughly 600× real time and the detection step runs at well over 10,000× real time, so the binding performance constraint for real-time deployment is the YOLO detector itself, which runs on a GPU. All features, z-scores, layer outputs, and alarm decisions are written to CSV to allow downstream statistical analysis without re-running the pipeline. Source code, the demonstration notebook, and the experimental database for the six pilot animals are openly available in the project repository (see Data Availability Statement).

3. Results

3.1. Detector Performance Across Individuals

Given n = 6, all-animal detection should be read as an encouraging proof-of-concept outcome, not robust validation. The validation protocol described in Section 2.7 was applied to the six pilot animals (Table 3, Figures 7 and 8). All six animals exposed to diesel were correctly detected, and no false alarms were raised on any of the 203 clean-water windows aggregated across all individuals. Because of the small number of animals, these point estimates have wide confidence intervals: animal-level detection of 6/6 corresponds to a 95% Clopper–Pearson confidence interval of 54–100%, whereas 0/203 clean-window false alarms corresponds to an exact 95% upper bound of 1.8% and a rule-of-three estimate of approximately 1.5%. The held-out subset (ExpIDs 4, 5, 6), which contributed neither to threshold calibration nor to the pooled clean-baseline distribution, was detected with first-alarm latencies of 1.9, 21.0, and 7.0 min, respectively. For the held-out animals alone, detection of 3/3 corresponds to a 95% confidence interval of 29–100%, and 0/78 clean-window false alarms corresponds to an exact 95% upper bound of 4.6%. Layer-resolved scoring also produced no clean-water false alarms: Layer 1 produced 0 alarms and Layer 2 produced 0 alarms across the pooled clean-window set.

To test whether threshold calibration was driven by the shortest clean-water baseline, we repeated the calibration after excluding ExpID 2, which had the shortest baseline in the pilot dataset. This ExpID-2 leave-out calibration used 148 clean-water windows and produced Tz = 1.667, CUSUM target = 0.709, and h = 6.147. These values were close to those obtained with the pooled clean-window calibration using all six animals (Tz = 1.591, target = 0.711, h = 6.117; 203 clean-water windows). Applying the ExpID-2 leave-out thresholds to the full six-animal dataset retained complete detection of exposed animals (6/6) and produced no clean-water false alarms (0/203).

Table 3. Detector performance on six animals using the validation thresholds (Tz = 1.559, target = 0.715, h = 5.498). ExpIDs 1–3 contributed their clean windows to threshold calibration. ExpIDs 4–6 was held out entirely. All six animals were detected with zero false alarms on clean windows. Layer labels in "True detection" indicate which detection layers fired; the latency column is the time of the first alarm relative to pollutant injection. The "Clean windows" column reports the number of 60 s windows that passed quality gating (dropout < 20%).

ExpID	Role	Clean Windows	Diesel Windows	False Alarms on Clean	True Detection	Latency (min)
1	calibration	55	36	0/55	yes (Layer 2)	8.0
2	calibration	8	24	0/8	yes (Layer 1 + 2)	1.0
3	calibration	62	64	0/62	yes (Layer 2)	23.7
4	held-out test	47	24	0/47	yes (Layer 1 + 2)	1.9
5	held-out test	16	42	0/16	yes (Layer 1 + 2)	21.0
6	held-out test	15	35	0/15	yes (Layer 2)	7.0

Note. The clean-water baseline duration for ExpID 2 (4 min, 8 windows) is shorter than the 15 min minimum recommended in Section 2.5.2. ExpID 2 was the earliest recording in the pilot series; analysis of these initial data revealed the need for longer baselines, and subsequent experiments (ExpIDs 3–6) were therefore conducted with extended clean-water recordings. ExpID 2 is retained in the calibration set because it nonetheless exhibited a sharp step-change response with the shortest detection latency observed (1.0 min), and excluding it would reduce the calibration sample size without changing the reported detection outcome.

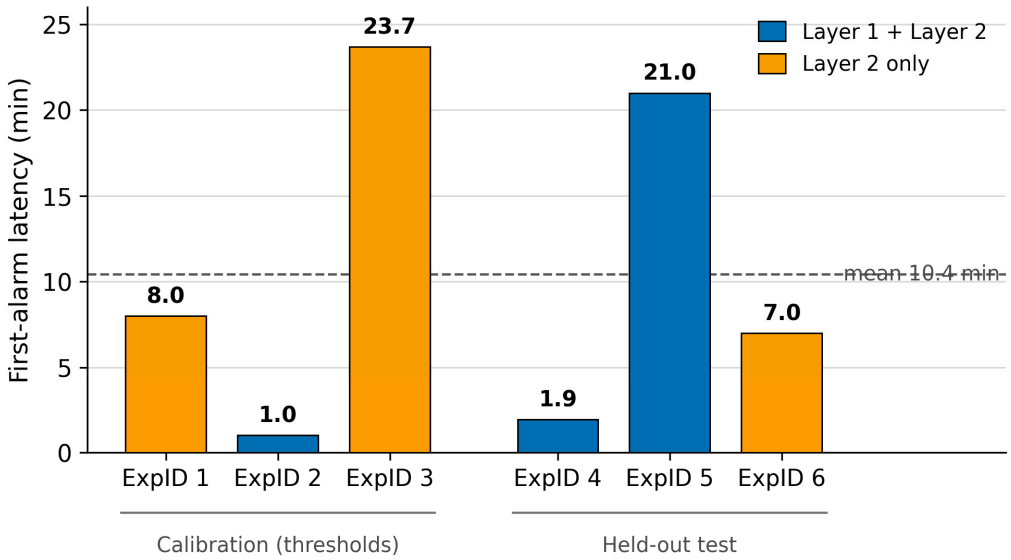

Figure 7. Per-animal first-alarm latency (time of the first alarm relative to pollutant injection) for the six pilot animals, as a graphical companion to Table 3. Bars are coloured by the detection layer(s) that fired (Layer 1 + Layer 2, or Layer 2 only). ExpID 1–3 contributed clean windows to threshold calibration; ExpID 4–6 were held out entirely. The dashed line marks the mean latency (10.4 min). Latencies range from 1.0 to 23.7 min across the six animals; given the small sample, these values are illustrative of the pilot dataset rather than estimates of operational detection time.

Across the full set of six animals, detection latencies range from 1.0 to 23.7 min. The contribution of each detection layer is animal-dependent: for ExpIDs 2 and 4, both Layer 1 (fast outlier on mean $|z|$) and Layer 2 (CUSUM drift) fire, with Layer 1 firing first; for ExpIDs 1, 3, and 6, only Layer 2 fires; and for ExpID 5, both layers fire. The combined two-layer design therefore exploits both the ability of Layer 1 to react sharply to abrupt step changes and the ability of Layer 2 to integrate small but persistent shifts that no single window would flag.

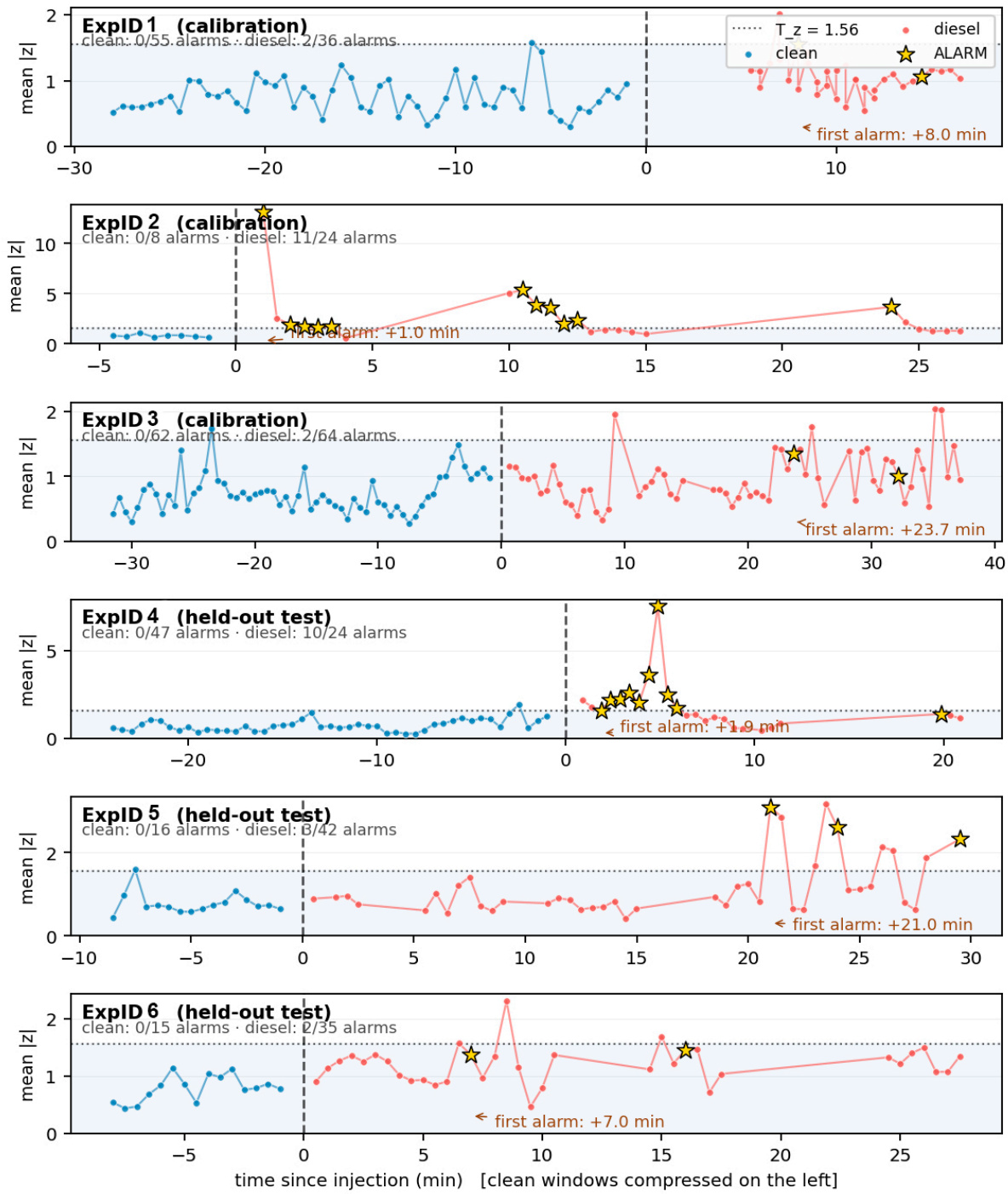

Figure 8. Validation across six individuals. Each panel shows mean $|z|$ (Layer 1 score) on a time axis centred at the injection moment (vertical dashed line); clean-baseline windows are shown on the left in compressed form. The dotted horizontal line is the validation threshold Tz = 1.56. Gold stars mark windows at which the full two-layer detector raises an alarm. No animal shows false alarms on clean windows. ExpIDs 2 and 4 show sharp step-change responses; ExpIDs 1, 3, and 5 show distributional drifts detected by Layer 2 (CUSUM); ExpID 6 is intermediate.

3.2. Descriptive Response Patterns and Detection-Layer Activation

The per-animal alarm patterns cluster into three descriptive response patterns observed in this six-animal pilot that are recognizably different from one another and that engage the two detection layers in different ways.

Sharp step-change responders (ExpIDs 2, 4; latencies 1.0 and 1.9 min). Within the first one or two windows after injection, mean $|z|$ jumps to 2–7 and Layer 1 fires immediately.

Layer 2 also fires shortly afterwards, providing redundant confirmation. These are the animals for which a single-window threshold on mean|z| would already be sufficient. The k-of-N rule on Layer 1 prevents a single tracking glitch from triggering an alarm but does not slow detection meaningfully for this response pattern.

Slow-drift responders (ExpIDs 1, 3, 5; latencies 8.0, 23.7, and 21.0 min). mean|z| stays near or below Tz for an extended period after injection but consistently above the CUSUM target. Layer 1 alone would miss these animals. Layer 2 accumulates the small-but-persistent shift and triggers between 8 and 24 min post-injection. ExpID 3 is the most extreme case and corresponds to the biphasic response described in Section 2.5.1: a stimulation phase of ~17 min, in which mean|z| rises modestly, is followed by a depression phase that pushes the CUSUM over h at minute 23.7. Without precise per-second injection logging this pattern would have been invisible.

Mixed responders (ExpID 6; latency 7.0 min). The pattern is intermediate between the two extremes: mean|z| crosses Tz sporadically but not for k = 3 consecutive windows, while the CUSUM accumulates faster than in the slow-drift cases. Layer 2 fires first. Layer 1 does not fire at all on this animal during the recorded window.

3.3. Cross-Validation of Threshold Calibration Regimes

The two threshold sets reported in Table 2, derived under the validation regime (3-of-6 animals for calibration) and the production regime (all 6 animals for calibration) agree closely. The 99th-percentile threshold Tz shifts from 1.559 to 1.591 (+2%), the CUSUM target shifts from 0.715 to 0.711 (−0.6%), and the CUSUM threshold h shifts from 5.498 to 6.117 (+11%). The larger shift in h is driven primarily by ExpID 4, whose clean-water mean|z| stream reaches a slightly higher peak than any of the original three calibration animals. This is the expected effect of adding more animals to the calibration pool: the estimate of the worst-case clean drift becomes more conservative as the pool grows. The validation regime is reported throughout this paper because it permits quantitative held-out evaluation. The production thresholds are reported for use in field deployments where the largest available calibration pool should be exploited.

A useful diagnostic is the ratio of detector latency to calibration baseline length. For the three slow-drift responders (ExpIDs 1, 3, 5), the latency is 0.36×, 0.76×, and 1.31× the duration of their respective calibration baselines, suggesting that increasing baseline length beyond 30 min may not materially shorten detection latency for this descriptive response pattern, because Layer 2 latency is dominated mainly by the magnitude of the post-injection drift rather than by the precision of the baseline μ/σ estimate. For the sharp step-change responders, latency is essentially independent of baseline length once the 15 min minimum is satisfied.

A one-at-a-time sensitivity analysis of the main detector parameters is provided in Supplementary Materials, Table S2, and confirms that the main animal-level detection outcome was stable across the tested parameter ranges. Analysis performed within the same small pilot dataset characterizes internal robustness and does not replace independent external validation.

4. Discussion

4.1. The Case for Per-Animal Anomaly Detection

The central methodological finding of this study is that supervised classification of clean-vs-polluted windows is not viable for *A. aurita* under diesel exposure at the available sample size, and that an unsupervised, per-animal anomaly detector is both more interpretable and more transferable. Three observations support this conclusion. First, baseline Fp varies by ~30% across individuals of the same species under

nominally identical conditions, consistent with the kinematic scaling demonstrated by McHenry and Jed [26] and with the species-specific dose–response patterns reported by Mercado et al. [22]. Second, the direction of the diesel-induced shift in Fp and locomotion is itself animal-specific. One animal showed a strong depression (Cliff's $\delta = +0.94$) while another showed essentially no effect ($\delta = -0.07$). Third, biphasic responses—stimulation followed by depression—place the same exposure class on opposite sides of the feature space, which is in principle unlearnable by a single-hyperplane classifier and which collapses logistic regression to majority-class prediction (34% LOVO accuracy in our pilot). The signal that survives this heterogeneity is the magnitude of the multivariate departure from the per-animal clean distribution, captured here by the scalar mean $|z|$. This result echoes a broader theme in the BEWS literature: organism, individual, and time-of-day variability in baseline behaviour means that absolute thresholds derived in one context generalize poorly. Jeong et al. [7] addressed this through gradient boosting on a very large Daphnia dataset. Grekov et al. [48] addressed it through unsupervised novelty detection on bivalves. The per-animal mean $|z|$ construct used here is in the same spirit, but exploits the within-subject paired design that is natural in single-aquarium jellyfish monitoring.

4.2. Comparison with Other Behaviour-Based BEWSs

The performance reported here—detection of all six exposed animals with no observed false alarms across 203 clean-water windows, with uncertainty quantified in Section 3.1, and detection latencies of 1–24 min—is broadly comparable to recent results obtained on bivalve and Daphnia platforms with much larger datasets. Grekov et al. [48] reported F1 = 1.0 on bivalve activity data using Isolation Forest, elliptic envelope, and local outlier factor, with Isolation Forest providing the fastest anomaly detection. Jeong et al. [7] reported precision and recall improvements of 29.5% and 43.4% over a traditional alarm index on eight years of Daphnia BEWS data using LightGBM. Both of these systems benefit from supervised or semi-supervised access to historical labelled events that are simply not available for jellyfish under controlled diesel exposure. The fact that an unsupervised detector with no learned weights and a baseline of only 8–62 clean windows per animal detected all six animals without observed clean-window false alarms in this pilot dataset is therefore an encouraging result. It suggests that *A. aurita* bell pulsation is a sufficiently rich behavioural signal to permit reliable detection without large training corpora, provided that the per-animal calibration step is carried out properly.

Detection latency deserves separate comment. The two sharp-responder animals (ExpIDs 2, 4) were detected within 1.0–1.9 min, which is competitive with reported bivalve and fish-ventilation BEWSs [3,5] and substantially faster than typical chemical-sensor response times for petroleum hydrocarbons in coastal waters. The slower latencies (8.0–23.7 min) of the drift responders reflect a genuine biological time constant: the animal's integrated departure from baseline takes ~10–25 min to cross the CUSUM threshold. This is the bound that any monitoring system using these animals must respect. It cannot be reduced by algorithmic refinement alone, because the underlying mean $|z|$ signal does not provide single-window evidence of exposure. In an operational context, this latency is still well within the time scale on which most pollution events develop, and is materially shorter than the typical hours-to-days turnaround of laboratory chemical analysis.

4.3. Biological Plausibility and Choice of Pollutant

Diesel was chosen as the first pollutant because its active fractions (PAHs and lighter aliphatic hydrocarbons) are among the most operationally relevant marine contaminants. The closest published precedent for an Aurelia neuromuscular response to petroleum is the work by Echols et al. [23], who reported muscle-contraction abnormalities in *A.*

aurita ephyrae exposed to chemically dispersed Macondo crude oil. We note that their preparation included Corexit 9500 dispersant, whereas the present study uses a water-accommodated fraction (WAF) of diesel without a chemical dispersant; the comparison is therefore at the level of qualitative effect (perturbation of contraction kinematics) rather than at the level of effect threshold. The two direction-consistent features in our pilot, increased tortuosity and increased immobility under diesel, are consistent with the broader observation that pesticide and metal exposure reduce locomotor speed and activity in aquatic invertebrates by 35–72% at environmentally relevant concentrations [40]. The within-species, animal-specific direction of effect on Fp observed here is consistent with the non-monotonic, compound-specific Fp responses reported by Mercado et al. [22] for food additives in *Aurelia* ephyrae and with the dose-dependent Fp depression reported by Faimali et al. [9] for cadmium and SDS in the same stage. In other words, the heterogeneity that limits supervised clean-versus-polluted classification is unlikely to be only a measurement artefact, but may reflect a real biological property of the response; per-animal anomaly detection is therefore an appropriate approach for this pilot dataset.

4.4. Sources of Inter-Individual Heterogeneity

The three response patterns described in Section 3.2 (fast step-change (ExpIDs 2, 4), slow drift (ExpIDs 1, 3, 5) and mixed (ExpID 6)) were not predicted in advance, and the present data do not allow us to assign individual animals to a specific mechanism. Several published sources of variability in *A. aurita* pulsation rate were, however, not controlled in this pilot, and they are plausible contributors to the differences we recorded.

Body size and baseline pulsation rate. McHenry and Jed [26] reported that pulsation frequency in A. aurita decreases with increasing bell diameter. The ~30% spread of clean-water baseline Fp across our six animals (0.77–1.01 Hz, Section 4.1) is consistent with this allometric scaling alone for animals of unequal size, and bell diameter was not standardized in the pilot. Crucially, however, body-size effects on absolute pulsation frequency do not propagate directly to the detector output. Because every feature is converted to a z-score against the same animal's own clean-water baseline (Section 2.5.2), the detector operates on each individual's departure from its own baseline and does not compare absolute Fp values between animals of different size. A size-driven difference in baseline Fp is therefore removed by per-animal normalization and can neither raise the clean-water false-alarm rate nor determine whether a diesel exposure is detected. Body size may still modulate the magnitude or direction of the response to diesel-WAF exposure, but this cannot be analyzed by size class in the present six-animal pilot. A separate size-class analysis is therefore deferred to the larger follow-up study, in which bell diameter will be treated as a covariate.

Environmental history and baseline stability. Dillon [51] reported that pulsation rate in *A. aurita* ephyrae responds to acute changes in temperature and salinity, and that the response to a salinity decrease is not stable over time: in his experiments, ephyrae transferred from 30 ‰ to 20 or 25 ‰ pulsed significantly faster than controls after 24 h, and returned to baseline rates only after 48 h. Temperature and salinity were nominally constant in our aquarium, but this earlier work shows that *A. aurita* baseline Fp depends on the recent environmental conditions experienced by the animal—one reason the first 15 min of a clean recording is not always equivalent to the longer-term baseline (Section 2.5.2).

Non-uniform exposure during WAF mixing. The WAF preparation was introduced into a small experimental aquarium that was not actively stirred during recording. Complete mixing of the introduced volume with the resident water in a 6 L aquarium under purely passive conditions is not instantaneous; concentration gradients can persist for

at least the first few minutes after introduction. The position of the focal animal in the aquarium at the moment of WAF introduction was not controlled, so individuals located close to the point of WAF addition and individuals further away were exposed to different hydrocarbon concentrations during the first few minutes—the same window in which the fast step-change responses were recorded.

The biphasic profile recorded for ExpID 3 (a stimulation phase of approximately 17 min followed by depression) is qualitatively similar to biphasic profiles previously reported in *A. aurita* ephyrae exposed to the cholinesterase inhibitor chlorpyrifos [10] and to certain food additives [22]. We note the similarity for context only; we do not interpret it mechanistically, as diesel is not a known cholinesterase inhibitor and only one of six animals showed the pattern. Whether the ExpID 3 profile reflects individual susceptibility, a particular exposure trajectory, or a specific mode of action of diesel cannot be resolved from the present dataset. A controlled follow-up, with standardization of bell diameter, life-cycle stage, acclimation duration, exposure dose and aquarium hydrodynamics, is needed to attribute response patterns to individual covariates with confidence.

4.5. Limitations

Several limitations of the present study should be acknowledged. First, the pilot dataset comprises six individuals of a single species exposed to a single nominal diesel-WAF preparation under controlled laboratory conditions. The detection of all six exposed medusae should therefore be interpreted as proof-of-concept evidence rather than as a precise estimate of operational performance. This uncertainty is reflected in the confidence intervals reported in Section 3.1: animal-level detection of 6/6 corresponds to a 95% Clopper–Pearson interval of 54–100%, and 0/203 clean-window false alarms corresponds to an exact 95% upper bound of 1.8% (rule-of-three estimate ≈1.5%). When the held-out subset is considered separately, the corresponding confidence intervals become wider still because of the smaller number of animals and clean-water windows. Generalization to additional pollutants (heavy metals, pesticides, microplastics) and to the lower, more chronic concentrations characteristic of background coastal contamination remains to be demonstrated. The pollutant classes most likely to produce a detectable behavioural response are those with documented *A. aurita* effects on Fp or locomotion: chlorpyrifos and other organophosphates [10,42], polyethylene microplastics [11], dispersed petroleum fractions [23], harmful microalgae [21], and copper or zinc [20]. However, the present detector settings should not yet be considered broadly transferable. Although the detector does not require labelled pollution-response data at inference time, its design was developed using the six pilot recordings. The selected behavioural features, the mean $|z|$ summary metric, the two-layer structure, the $k = 3$ consecutive-window rule, and the 1.3× CUSUM threshold factor were all chosen during exploratory analysis of this pilot dataset. Therefore, the detector should be regarded as a proof-of-concept implementation. Before these settings can be considered broadly applicable, they need to be tested on new recordings from additional animals, other pollutants, and field conditions.

A related limitation concerns specificity. The detector raises an alarm whenever the behaviour of the focal animal departs in a sustained way from its own clean-water baseline, regardless of the cause of that departure. It is therefore a sensitive but non-specific anomaly detector. Strong non-pollution perturbations, including abrupt changes in temperature, salinity, illumination, or mechanical and human disturbance, could in principle also exceed the alarm thresholds. This is particularly relevant for *A. aurita*, because pulsation rate is known to respond to acute temperature and salinity change [51]. In operational use, the biological alarm should therefore be interpreted together with co-located physicochemical measurements, including temperature, salinity, oxygen, light, turbidity, and other rele-

vant environmental variables. Future negative-control experiments should include sham clean-seawater injections and controlled temperature, salinity, and light step-changes to quantify specificity.

The present dataset was also affected by substantial technical attrition. Twelve medusae were recorded, but only six recordings were technically usable for the full YOLO-based pipeline, corresponding to approximately 50% attrition. The exclusion criteria were technical, response-blind, and applied at the acquisition/tracking stage rather than on the basis of detector output. Nevertheless, attrition remains a practical limitation for field deployment. According to Supplementary Materials, Table S1, four of the six excluded recordings (X1, X3, X5, and X6) failed before diesel-WAF introduction, whereas two recordings (X2 and X4) failed after diesel-WAF introduction. This pattern does not indicate enrichment of the analyzed subset for diesel-responsive animals, but it shows that robust acquisition, stable focus, and reliable tracking are essential prerequisites for routine use.

Second, the 15 min lower bound on the calibration baseline is a hard constraint of the per-animal design and may be impractical in some field deployments. This bound was itself an outcome of the present pilot: the earliest recording in the series (ExpID 2, 4 min baseline) demonstrated that short baselines produce inflated z-scores and unstable detector behaviour, and subsequent experiments were therefore conducted with extended clean-water recordings. Within-clean Fp variability of 0.5–1.3 Hz over a 22 min window in one of our pilot animals further supports this requirement, since shorter baselines run a real risk of sampling only one tail of the underlying distribution. ExpID 2 was nonetheless detected with the shortest latency in the pilot set (1.0 min), which is attributable to the sharpness of its post-injection response rather than to the short baseline itself; this result should therefore not be generalized to slow-drift response patterns, which require a longer clean-water baseline for reliable σ estimation. A useful direction for future work is to investigate adaptive baseline updating that can shorten the cold-start period without compromising the σ estimate. Short baselines (ExpID 2/5/6) can inflate z-scores (as noted for ExpID 2), but detection is still correct (Table S1).

The hydrocarbon content of the prepared diesel-WAF was not measured analytically. The present study was designed as a pilot proof-of-concept evaluation of a computer-vision BEWS pipeline, rather than as a quantitative toxicity assay. Within this pilot design, the WAF concentration was not verified by GC–MS or TPH analysis. Accordingly, the term "diesel-WAF exposure" is used here to denote exposure to a nominal preparation produced under the described conditions, rather than exposure to a measured concentration of dissolved or dispersed petroleum hydrocarbons. Because the WAF was prepared over a 1–12 h equilibration interval and introduced into the experimental aquarium without analytical confirmation of hydrocarbon content, the exact hydrocarbon concentration in the exposure medium remains uncertain. Future studies will include standardized WAF preparation and analytical verification of exposure concentrations.

Third, single-target tracking is assumed throughout. Multi-individual aquaria would require a data-association step (e.g., DeepSORT-style tracking, as demonstrated for fish in Lopez-Marcano et al.'s work [38]) and a per-individual baseline pool. The detection logic at the per-window level is unaffected, but the bookkeeping is non-trivial and is reserved for future work.

Fourth, neural-network weights are deliberately excluded from the deployed detector, but the upstream YOLO detector does of course rely on a trained model. The YOLO detector was trained on a modest custom image set, and although performance was evaluated using a video-level test split, the test set remains small. Errors in the upstream detector can propagate into the downstream behavioural features through missed detections, bounding-box instability, or changes in confidence score. We do not address the question of how

robust the YOLO model is to changes in lighting, water clarity, or background. Existing literature [30,33,36] suggests that with appropriate underwater image preprocessing this is a manageable problem, but it is outside the scope of the present anomaly detection contribution. Field deployment will therefore require stronger evaluation under changing light, turbidity, biofouling, and background conditions. Future field versions of the system should include explicit testing under degraded visibility and, if necessary, image preprocessing steps to stabilize object detection.

Finally, the validation comprises six individuals. This is sufficient for proof-of-concept demonstration under the present held-out protocol, but it is not sufficient for tight quantification of detector sensitivity, specificity, or receiver-operating characteristics at intermediate decision thresholds. Expansion of the pilot dataset to ~20 individuals, which would also begin to make individual-level frame-rate autoencoder approaches tractable, is a natural next step. The zero false-alarm rate reported here should also be understood as a short-term, within-session result obtained under stable laboratory conditions. It does not demonstrate long-term stability over days to weeks of monitoring, when illumination, water clarity, animal condition, background complexity, biofouling, and environmental variability may drift. Establishing long-term stability will require longer deployments, periodic per-animal recalibration, and explicit evaluation of false-alarm behaviour under realistic environmental variability.

4.6. Implications and Future Directions

Several practical challenges must be resolved before the proposed BEWS can be used for routine field monitoring. Long-term video monitoring will require control of biofouling on aquarium walls and optical surfaces, stable illumination, and clear procedures for animal husbandry and periodic replacement. The thermal tolerance and seasonal condition of the sentinel species are also important, especially for cold- and temperate-water taxa whose baseline pulsation may change with temperature. In addition, the detector first needs a clean-water recording for each animal to establish its normal behaviour. In the present implementation, at least 15 min of clean-water recording is required before reliable anomaly detection can begin; during this initial period, the system would not yet be able to issue a reliable alarm. Field deployment will therefore require testing of baseline stability over longer periods, including under changing temperature and salinity conditions, as emphasized by Dillon et al. [51].

Future validation should not rely on the biological alarm alone. Behavioural alarms should be evaluated together with simultaneous physicochemical measurements, such as temperature, salinity, dissolved oxygen, light, turbidity, and other relevant environmental variables. This is necessary because the detector identifies sustained departures from each animal's own baseline, but does not identify the cause of the change. Negative-control experiments are therefore required to quantify specificity. These should include sham clean-seawater injections and controlled step-changes in temperature, salinity, and light, so that pollution-induced behavioural anomalies can be distinguished from behavioural changes caused by non-pollution environmental perturbations.

The combination of biological sensitivity, non-contact measurement, lightweight implementation, and absence of neural-network weights in the deployed anomaly scoring step positions *A. aurita* as a promising organism for next-generation vision-based BEWSs for marine and coastal water quality monitoring. The pipeline reported here can be deployed on commodity hardware (a Raspberry Pi 5 with an AI HAT+ accelerator and a Pi Camera Module 3 was sufficient for our pilot acquisitions) and produces a binary alarm in the same format as legacy bivalve and fish-ventilation systems, allowing potential integration into existing BEWS infrastructure. However, transfer to routine field monitoring should not be

assumed from the present pilot dataset alone. Field deployment will require independent validation under realistic environmental variability, including changes in illumination, turbidity, temperature, salinity, animal condition, and background complexity.

Such hybrid systems are likely to be the operationally most useful endpoint of this line of work. A further technical direction is the incorporation of physics-informed computer-vision approaches, which may help combine image-based tracking with prior knowledge of optical degradation, animal motion, and environmental constraints [52]. For deployment under degraded visibility, maritime dehazing and image-enhancement methods may also be evaluated as preprocessing steps to stabilize upstream object detection [53].

5. Conclusions

We have presented an end-to-end, computer-vision BEWS pipeline that is unsupervised at inference time and operates without labelled pollution-response data, using Aurelia aurita as the sentinel organism. The pipeline ingests YOLO bounding-box detections of a single focal jellyfish, derives a continuous bell-area signal and centroid trajectory, and computes eleven pulsation, kinematic, and detection-quality features on overlapping 60 s windows. A per-animal baseline is fitted on at least 15 min of clean-water video, after which a two-layer detector (fast outlier detection on the mean absolute z-score, plus one-sided CUSUM accumulation) flags any sustained departure from the animal's own clean distribution. In this six-animal proof-of-concept pilot, all six adult medusae exposed to nominal diesel-WAF were detected, and no false alarms were observed in 203 clean-window opportunities. As reported in Section 3.1, these point estimates have wide confidence intervals: 6/6 animal-level detection corresponds to a 95% Clopper–Pearson interval of 54–100%, and 0/203 clean-window false alarms corresponds to an exact 95% upper bound of 1.8% (rule-of-three estimate ≈1.5%). Detection latencies ranged from 1.0 to 23.7 min, and the observed responses were described as three descriptive response patterns in this pilot dataset: sharp step-change, slow drift, and mixed. These patterns engaged the two detection layers in different ways. The deployed anomaly scoring step contains no neural-network weights, is implemented in fewer than 300 lines of Python, and is designed for field-portable use in settings where a stationary camera can be positioned alongside an aquarium, although field validation remains required. Per-animal anomaly detection accommodates the strong inter-individual variability of the diesel-WAF response that limits supervised clean-versus-polluted classification at this sample size. The same per-animal anomaly detection framework may be applicable to additional pollutant classes, including heavy metals, pesticides, microplastics, and dispersed petroleum fractions, but this remains to be tested on independent datasets with analytically verified exposures and pollutant-specific threshold evaluation. The detector's alarm stream is expected to be most useful in operational deployment as a biological complement to physicochemical sensor networks rather than as a stand-alone indicator of pollutant identity.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jmse14131189/s1, Table S1: Inventory of all 12 experimental recordings, including inclusion status, recording date, bell diameter, clean- and diesel-phase durations, exclusion reason, and whether diesel-WAF was introduced before exclusion; Table S2: One-at-a-time sensitivity analysis of the main detector parameters.

Author Contributions: Conceptualization, A.G.; methodology, A.G. and K.P.; formal analysis, A.G. and K.P.; investigation, A.G.; data curation, I.B. and A.B.; writing—original draft preparation, A.G., K.P. and I.B.; writing—review and editing, A.G. and E.V.; visualization, A.G. and K.P.; supervision,

A.G.; project administration, E.V.; funding acquisition, E.V. All authors have read and agreed to the published version of the manuscript.

Funding: The study was supported by a grant from the Russian Science Foundation № 25-19-00551, https://rscf.ru/project/25-19-00551/ (accessed on 25 June 2026).

Institutional Review Board Statement: The work involved invertebrate animals (*Aurelia aurita*), which in many jurisdictions are not subject to ethical review under the relevant national legislation.

Data Availability Statement: The source code (pipeline.py, bews.py), demonstration notebook (bews_demo.ipynb), unit tests, and the experimental metadata for the six pilot animals are openly available in the GitHub repository https://github.com/keepersas/jellyfish (accessed on 27 April 2026) under an MIT licence. The full annotated dataset is publicly hosted on Roboflow https://universe.roboflow.com/bidjo-yvkeu/jellyfish-detection-oyfap/dataset/4 (accessed on 27 April 2026). Raw aquarium video files are available from the corresponding author upon reasonable request.

Acknowledgments: During the preparation of this manuscript, the authors used Claude (Anthropic, model Claude Opus 4.7, 2026) for English language proofreading. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BEWS	Biological Early Warning System
CUSUM	Cumulative Sum (control chart)
CV	Coefficient of variation
F_p	Pulsation frequency
IPI	Inter-pulse interval
LOVO	Leave-one-video-out
mean | z |	Mean absolute z-score across nine behavioural features
PAH	Polycyclic aromatic hydrocarbon
PSD	Power spectral density
ROI	Region of Interest
YOLO	You Only Look Once (object detector family)

References

1. Bae, M.-J.; Park, Y.-S. Biological early warning system based on the responses of aquatic organisms to disturbances: A review. *Sci. Total Environ.* **2014**, *466–467*, 635–649. [CrossRef] [PubMed]
2. Bownik, A.; Wlodkowic, D. Advances in real-time monitoring of water quality using automated analysis of animal behaviour. *Sci. Total Environ.* **2021**, *789*, 147796. [CrossRef] [PubMed]
3. Vereycken, J.E.; Aldridge, D.C. Bivalve molluscs as biosensors of water quality: State of the art and future directions. *Hydrobiologia* **2023**, *850*, 231–256. [CrossRef]
4. Grekov, A.N.; Kuzmin, K.A.; Mishurov, V.Z. Automated Early Warning System for Water Environment Based on Behavioral Reactions of Bivalves. In Proceedings of the 2019 International Russian Automation Conference (RusAutoCon), Sochi, Russia, 8–14 September 2019. [CrossRef]
5. van der Schalie, W.H.; Shedd, T.R.; Knechtges, P.L.; Widder, M.W. Using higher organisms in biological early warning systems for real-time toxicity detection. *Biosens. Bioelectron.* **2001**, *16*, 457–465. [CrossRef] [PubMed]
6. Kizgin, A.; Schmidt, D.; Joss, A.; Hollender, J.; Morgenroth, E.; Kienle, C.; Langer, M. Application of biological early warning systems in wastewater treatment plants: Introducing a promising approach to monitor changing wastewater composition. *J. Environ. Manag.* **2023**, *347*, 119001. [CrossRef] [PubMed]
7. Jeong, H.; Park, S.; Choi, B.; Yu, C.S.; Hong, J.Y.; Jeong, T.Y.; Cho, K.H. Machine learning-based water quality prediction using octennial in-situ Daphnia magna biological early warning system data. *J. Hazard. Mater.* **2024**, *465*, 133196. [CrossRef] [PubMed]

8. Gambardella, C.; Costa, E.; Miroglio, R.; Nugnes, R.; Giuga, M.; Castelli, F.; Faimali, M. Early Warning Systems for Marine Monitoring. In Proceedings of the 2024 IEEE International Workshop on Metrology for the Sea—Learning to Measure Sea Health Parameters (MetroSea), Portorose, Slovenia, 14–16 October 2024; pp. 24–28. [CrossRef]
9. Faimali, M.; Garaventa, F.; Piazza, V.; Costa, E.; Greco, G.; Mazzola, V.; Beltrandi, M.; Bongiovanni, E.; Lavorano, S.; Gnone, G. Ephyra jellyfish as a new model for ecotoxicological bioassays. *Mar. Environ. Res.* **2014**, *93*, 93–101. [CrossRef] [PubMed]
10. Costa, E.; Piazza, V.; Lavorano, S.; Faimali, M.; Garaventa, F.; Gambardella, C. Effect of neurotoxic compounds on ephyrae of *Aurelia aurita* jellyfish. *Hydrobiologia* **2015**, *759*, 75–85. [CrossRef]
11. Costa, E.; Gambardella, C.; Piazza, V.; Vassalli, M.; Sbrana, F.; Lavorano, S.; Garaventa, F.; Faimali, M. Microplastics ingestion in the ephyra stage of *Aurelia* sp. triggers acute and behavioral responses. *Ecotoxicol. Environ. Saf.* **2020**, *189*, 109983. [CrossRef] [PubMed]
12. Di Giannantonio, M.; Gambardella, C.; Miroglio, R.; Costa, E.; Sbrana, F.; Smerieri, M.; Carraro, G.; Utzeri, R.; Faimali, M.; Garaventa, F. Ecotoxicity of polyvinylidene difluoride (PVDF) and polylactic acid (PLA) microplastics in marine zooplankton. *Toxics* **2022**, *10*, 479. [CrossRef] [PubMed]
13. Templeman, M.A.; McKenzie, M.R.; Kingsford, M.J. The utility of jellyfish as marine biomonitors. *Mar. Pollut. Bull.* **2021**, *173*, 113056. [CrossRef] [PubMed]
14. Baiandina, I.; Grekov, A.; Vyshkvarkova, E. Black Sea Planktonic Organisms as Bioindicators for Biological Early Warning Systems: A Systematic Review. *Water* **2026**, *18*, 899. [CrossRef]
15. Templeman, M.A.; Kingsford, M.J. Trace element accumulation in Cassiopea sp. (Scyphozoa) from urban marine environments in Australia. *Mar. Environ. Res.* **2010**, *69*, 63–72. [CrossRef] [PubMed]
16. Epstein, H.E.; Templeman, M.A.; Kingsford, M.J. Fine-scale detection of pollutants by a benthic marine jellyfish. *Mar. Pollut. Bull.* **2016**, *107*, 340–346. [CrossRef] [PubMed]
17. Macali, A.; Bergami, E. Jellyfish as innovative bioindicator for plastic pollution. *Ecol. Indic.* **2020**, *115*, 106375. [CrossRef]
18. Rapp, J.; Herrera, A.; Bondyale-Juez, D.R.; González-Pleiter, M.; Reinold, S.; Asensio, M.; Martínez, I.; Gómez, M. Microplastic ingestion in jellyfish Pelagia noctiluca (Forsskal, 1775) in the North Atlantic Ocean. *Mar. Pollut. Bull.* **2021**, *166*, 112266. [CrossRef] [PubMed]
19. Praved, P.H.; Neethu, K.V.; Nandan, S.B.; Sankar, N.D.; Aravind, E.H.; Sebastian, S.; Marigoudar, S.R.; Sharma, K.V. Evaluation of microplastic pollution and risk assessment in a tropical monsoonal estuary, with special emphasis on contamination in jellyfish. *Environ. Pollut.* **2024**, *343*, 123158. [CrossRef] [PubMed]
20. Templeman, M.A. Investigating the Potential of Jellyfishes as Marine Biomonitors and Bioindicators of Metal Pollution. Ph.D. Thesis, James Cook University, Townsville, Australia, 2012. [CrossRef]
21. Yang, H.J.; Seo, H.J.; Kim, Y.H.; Yun, G.; Lee, M.J.; Du Yoo, Y.; Jang, S.H. Effects of harmful microalgae on the behavior and morphology of ephyrae of the moon jellyfish *Aurelia aurita*. *Mar. Pollut. Bull.* **2024**, *205*, 116640. [CrossRef] [PubMed]
22. Mercado, B.; Vila, B.; Roca-Pérez, L.; Duran-Giner, N.; Boluda-Hernández, R.; Andreu-Sánchez, O. *Aurelia aurita* as a Model for Ecotoxicologically Assessing Food Additives: 2-Methyl-1-phenylpropan-2-ol and 1-Phenylethan-1-ol. *Toxics* **2025**, *13*, 572. [CrossRef] [PubMed]
23. Echols, B.S.; Smith, A.J.; Gardinali, P.R.; Rand, G.M. The use of ephyrae of a scyphozoan jellyfish, *Aurelia aurita*, in the aquatic toxicological assessment of Macondo oils from the Deepwater Horizon incident. *Chemosphere* **2016**, *144*, 1893–1900. [CrossRef] [PubMed]
24. Olguín-Jacobson, C.; Pitt, K.A.; Carroll, A.R.; Melvin, S.D. Polyps of the Jellyfish *Aurelia aurita* Are Unaffected by Chronic Exposure to a Combination of Pesticides. *Environ. Toxicol. Chem.* **2020**, *39*, 1685–1692. [CrossRef] [PubMed]
25. Wu, X.; Zhang, X.; Liao, H.; Guo, J.; Ma, Z.; Fu, Z. Microplastics and tetracycline affecting apoptosis, enzyme activities and metabolism processes in the *Aurelia aurita* polyps: Insights into combined pollutant effects. *Front. Mar. Sci.* **2025**, *12*, 1545131. [CrossRef]
26. McHenry, M.J.; Jed, J. The ontogenetic scaling of hydrodynamics and swimming performance in jellyfish (*Aurelia aurita*). *J. Exp. Biol.* **2003**, *206*, 4125–4137. [CrossRef] [PubMed]
27. Bajcar, T.; Malačič, V.; Malej, A.; Širok, B. Kinematic properties of the jellyfish *Aurelia* sp. *Hydrobiologia* **2009**, *616*, 279–289. [CrossRef]
28. Boyd, W.; Ohdera, A.; Kaufman, L.; Herman, D. Chemoreception in *Aurelia aurita* studied by AI-enhanced image analysis. *J. Emerg. Investig.* **2025**, *8*, 1–7. [CrossRef] [PubMed]
29. Adriaens, I.; Friggens, N.C.; Ouweltjes, W.; Scott, H.; Aernouts, B.; Statham, J. Productive life span and resilience rank can be predicted from on-farm first-parity sensor time series but not using a common equation across farms. *J. Dairy Sci.* **2020**, *103*, 7155–7171. [CrossRef] [PubMed]
30. Martin-Abadal, M.; Ruiz-Frau, A.; Hinz, H.; Gonzalez-Cid, Y. Jellytoring: Real-Time Jellyfish Monitoring Based on Deep Learning Object Detection. *Sensors* **2020**, *20*, 1708. [CrossRef] [PubMed]

31. Zhang, W.; Rui, F.; Xiao, C.; Li, H.; Li, Y. JF-YOLO: The jellyfish bloom detector based on deep learning. *Multimed. Tools Appl.* **2024**, *83*, 7097–7117. [CrossRef]
32. Gao, M.; Li, S.; Wang, K.; Bai, Y.; Ding, Y.; Zhang, B.; Guan, N.; Wang, P. Real-Time Jellyfish Classification and Detection Algorithm Based on Improved YOLOv4-tiny and Improved Underwater Image Enhancement Algorithm. *Sci. Rep.* **2023**, *13*, 12989. [CrossRef] [PubMed]
33. Gao, M.; Bai, Y.; Li, Z.; Li, S.; Zhang, B.; Chang, Q. Real-Time Jellyfish Classification and Detection Based on Improved YOLOv3 Algorithm. *Sensors* **2021**, *21*, 8160. [CrossRef] [PubMed]
34. Pham, T.N.; Nguyen, V.H.; Kwon, K.R.; Kim, J.H.; Huh, J.H. Improved YOLOv5 Based Deep Learning System for Jellyfish Detection. *IEEE Access* **2024**, *12*, 87838–87849. [CrossRef]
35. Nawarathne, U.M.M.P.K.; Kumari, H.M.L.S.; Kumari, H.M.N.S. Comparative analysis of jellyfish classification: A study using YOLOv8 and pre-trained models. In Proceedings of the 2024 International Research Conference on Smart Computing and Systems Engineering (SCSE), Colombo, Sri Lanka, 4 April 2024; Volume 7, pp. 1–6. [CrossRef]
36. Pan, H.; Lan, J.; Wang, H.; Li, Y.; Zhang, M.; Ma, M.; Zhang, D.; Zhao, X. UWV-Yolox: A Deep Learning Model for Underwater Video Object Detection. *Sensors* **2023**, *23*, 4859. [CrossRef] [PubMed]
37. Philip, F.M.; Edappazham, G.; Jims, A.; Prabhullachandran, L.D. Deep Learning-based marine species detection and classification framework for biomonitoring in the Arctic fjords, Svalbard. *Czech Polar Rep.* **2023**, *13*, 182–196. [CrossRef]
38. Lopez-Marcano, S.; Jinks, E.L.; Buelow, C.A.; Brown, C.J.; Wang, D.; Kusy, B.; Ditria, E.M.; Connolly, R.M. Automatic detection of fish and tracking of movement for ecology. *Ecol. Evol.* **2021**, *11*, 8254–8263. [CrossRef] [PubMed]
39. Honda, M.; Suzuki, N. Toxicities of Polycyclic Aromatic Hydrocarbons for Aquatic Animals. *Int. J. Environ. Res. Public Health* **2020**, *17*, 1363. [CrossRef] [PubMed]
40. Jeong, H.; Ali, W.; Zinck, P.; Souissi, S.; Lee, J.-S. Toxicity of methylmercury in aquatic organisms and interaction with environmental factors and coexisting pollutants: A review. *Sci. Total Environ.* **2024**, *943*, 173574. [CrossRef] [PubMed]
41. Untersteiner, H.; Gretschel, G.; Puchner, T.; Napetschnig, S.; Kaiser, H. Monitoring Behavioral Responses to the Heavy Metal Cadmium in the Marine Shrimp Hippolyte inermis Leach (Crustacea: Decapoda) with Video Imaging. *Zool. Stud.* **2005**, *44*, 71–80.
42. Shuman-Goodier, M.E.; Propper, C.R. A meta-analysis synthesizing the effects of pesticides on swim speed and activity of aquatic vertebrates. *Sci. Total Environ.* **2016**, *565*, 758–766. [CrossRef] [PubMed]
43. Alimba, C.G.; Faggio, C. Microplastics in the marine environment: Current trends in environmental pollution and mechanisms of toxicological profile. *Environ. Toxicol. Pharmacol.* **2019**, *68*, 61–74. [CrossRef] [PubMed]
44. Ge, J.; Li, B.; Liao, M.; Zhang, Z.; Chen, S.; Xia, B.; Wang, Y. Ingestion, egestion and physiological effects of polystyrene microplastics on the marine jellyfish *Rhopilema esculentum*. *Mar. Pollut. Bull.* **2023**, *187*, 114609. [CrossRef] [PubMed]
45. Muharemi, F.; Logofătu, D.; Leon, F. Machine learning approaches for anomaly detection of water quality on a real-world data set. *J. Inf. Telecommun.* **2019**, *3*, 294–307. [CrossRef]
46. Dogo, E.M.; Nwulu, N.I.; Twala, B.; Aigbavboa, C. A survey of machine learning methods applied to anomaly detection on drinking-water quality data. *Urban Water J.* **2019**, *16*, 235–248. [CrossRef]
47. Zhu, M.; Wang, J.; Yang, X.; Zhang, Y.; Zhang, L.; Ren, H.; Wu, B.; Ye, L. A review of the application of machine learning in water quality evaluation. *Eco-Environ. Health* **2022**, *1*, 107–116. [CrossRef] [PubMed]
48. Grekov, A.N.; Kabanov, A.A.; Vyshkvarkova, E.V.; Trusevich, V.V. Anomaly Detection in Biological Early Warning Systems Using Unsupervised Machine Learning. *Sensors* **2023**, *23*, 2687. [CrossRef] [PubMed]
49. Leigh, C.; Alsibai, O.; Hyndman, R.J.; Kandanaarachchi, S.; King, O.C.; McGree, J.M.; Neelamraju, C.; Strauss, J.; Talagala, P.D.; Turner, R.D.R.; et al. A framework for automated anomaly detection in high frequency water-quality data from in situ sensors. *Sci. Total Environ.* **2019**, *664*, 885–898. [CrossRef] [PubMed]
50. Page, E.S. Continuous inspection schemes. *Biometrika* **1954**, *41*, 100–115. [CrossRef]
51. Dillon, T.M. Effects of acute changes in temperature and salinity on pulsation rates in ephyrae of the scyphozoan *A. aurita*. *Mar. Biol.* **1977**, *42*, 31–35. [CrossRef]
52. Banerjee, C.; Nguyen, K.; Fookes, C.; Karniadakis, G.E. Physics-Informed Computer Vision: A Review and Perspectives. *ACM Comput. Surv.* **2024**, *57*, 17. [CrossRef]
53. Chen, X.; Xin, Z.; Zhang, H.; Wu, Y.; Wei, C.; Postolache, O. Vision Transformer-Based Image Dehazing for Climate-Resilient Maritime Navigation. *IEEE Trans. Intell. Transp. Syst.* **2026**, 1–13. [CrossRef]