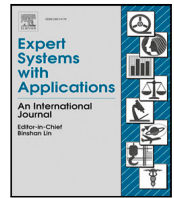




Contents lists available at ScienceDirect

# Expert Systems With Applications

journal homepage: [www.elsevier.com/locate/eswa](http://www.elsevier.com/locate/eswa)

## WILDetect: An intelligent platform to perform airborne wildlife census automatically in the marine ecosystem using an ensemble of learning techniques and computer vision

Kaya Kuru<sup>a,\*</sup>, Stuart Clough<sup>b</sup>, Darren Ansell<sup>a</sup>, John McCarthy<sup>c</sup>, Stephanie McGovern<sup>c</sup>

<sup>a</sup> School of Engineering, University of Central Lancashire, Fylde Rd, Preston, Lancashire, PR12HE, UK

<sup>b</sup> APEM Inc., 2603 NW 13th Street, 402, Gainesville, FL 32609-2835, USA<sup>1</sup>

<sup>c</sup> APEM Ltd., The Embankment Business Park, Stockport, SK4 3GN, UK<sup>2</sup>

### ARTICLE INFO

#### Keywords:

Marine ecosystems  
Machine learning  
Reinforcement learning  
Seabirds  
Gannets  
Object detection  
Image processing  
Aerial surveys  
Marine census  
Airborne bio census  
Wildlife census

### ABSTRACT

The habitats of marine life, characteristics of species, and the diverse mix of maritime industries around these habitats are of interest to many researchers, authorities, and policymakers whose aim is to conserve the earth's biological diversity in an ecologically sustainable manner while being in line with indispensable industrial developments. Automated detection, locating, and monitoring of marine life along with the industry around the habitats of this ecosystem may be helpful to (i) reveal current impacts, (ii) model future possible ecological trends, and (iii) determine required policies which would lead accordingly to a reduced ecological footprint and increased sustainability. New automatic techniques are required to observe this large environment efficiently. Within this context, this study aims to develop a novel platform to monitor marine ecosystems and perform bio census in an automated manner, particularly for birds in regional aerial surveys since birds are a good indicator of overall ecological health. In this manner, a new non-parametric approach, WILDetect, has been built using an ensemble of supervised Machine Learning (ML) and Reinforcement Learning (RL) techniques. It employs several hybrid techniques to segment, split and count maritime species – in particular, birds – in order to perform automated censuses in a highly dynamic marine ecosystem. The efficacy of the proposed approach is demonstrated by experiments performed on 26 surveys which include Northern gannets (*Morus bassanus*) by utilising retrospective data analysis techniques. With this platform, by combining multiple techniques, gannets can be detected and split automatically with very high sensitivity ( $Se$ ) ( $\approx 0.97$ ), specificity ( $Sp$ ) ( $\approx 0.99$ ), and accuracy ( $Acc$ ) ( $\approx 0.99$ ) — these values are validated by precision ( $Pr$ ) ( $\approx 0.98$ ). Moreover, the evaluation of the system by the APEM staff, which uses a completely new evaluation dataset gathered from recent surveys, shows the viability of the proposed techniques. The experimental results suggest that similar automated data processing techniques – tailored for specific species – can be helpful both in performing time-intensive marine wildlife censuses efficiently and in establishing ecological platforms/models to understand the underlying causes of trends in species populations along with the ecological change.

### 1. Introduction

The oceans cover two-thirds of the Earth's surface and the maritime economy has always been diverse and abundant. With the applications of emerging fields of science and technology in new and existing industries, prominent companies and research organisations have been recently developing and deploying evolving technologies supported by location-independent advanced maritime mechatronics

systems (AMMSs) (Kuru & Yetgin, 2019; Shi et al., 2017) to explore and exploit the resources in this tough landscape. This massively evolving industry, enabling enormous continuous human control in the maritime, has the potential to impact the marine ecosystem dramatically; in particular, the seabed, birds, turtles, and fish. Birds are an inseparable part of the maritime ecosystem. Seabird population changes are good indicators of long-term and large-scale change in marine ecosystems, and important because their populations are strongly influenced by

\* Corresponding author.

E-mail addresses: [kkuru@uclan.ac.uk](mailto:kkuru@uclan.ac.uk) (K. Kuru), [s.clough@apemltd.co.uk](mailto:s.clough@apemltd.co.uk) (S. Clough), [dansell@uclan.ac.uk](mailto:dansell@uclan.ac.uk) (D. Ansell), [J.McCarthy@apemltd.co.uk](mailto:J.McCarthy@apemltd.co.uk) (J. McCarthy), [S.McGovern@apemltd.co.uk](mailto:S.McGovern@apemltd.co.uk) (S. McGovern).

<sup>1</sup> <https://apem-inc.com>.

<sup>2</sup> <https://www.apemltd.co.uk>.

<https://doi.org/10.1016/j.eswa.2023.120574>

Received 12 February 2023; Received in revised form 10 May 2023; Accepted 27 May 2023

Available online 7 June 2023

0957-4174/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



Fig. 1. Example of a gannet in a high-resolution image.<sup>3</sup>

threats (e.g., entanglement in fishing gear, overfishing of food sources, climate change, pollution, disturbance, direct exploitation, development, energy production) to marine and coastal ecosystems (Paleczny et al., 2015). Considerable differences in population trajectories of offshore bird families have been documented, which suggests that overall offshore bird populations are decreasing (BOEM, 2022). The monitored portion of the global seabird population, representing approximately 19% of the global seabird population, has declined by nearly 70% between 1950 and 2010 (Paleczny et al., 2015), a net loss approaching 3 billion birds (u.e., %29) since 1970 (Rosenberg et al., 2019). This loss of bird abundance signals an urgent need to address threats to avert future avifaunal collapse and associated loss of ecosystem integrity, function, and services (Rosenberg et al., 2019).

One type of bird is the northern gannet (*Morus bassanus*), the largest seabird in the North Atlantic, having a wingspan of up to 180 cm and a length of up to 100 cm (RSBP, 2015). More specifically, gannets are large white birds with distinctive features including yellowish heads and black-tipped wings. They are distinctively shaped with a long neck and long pointed beak, long pointed tail, and long pointed wings (RSBP, 2015). An example is displayed in Fig. 1. The most important nesting ground for northern gannets is the UK with about half of the world's population (55.6%) (JNCC, 2015). APEM Ltd<sup>4</sup> has a wide range of gannet data with geographical positions obtained from all around the world and this species is the focus of this study which aims to test the developed approaches to help perform further autonomous bird censuses paving the way for automated classification of multispecies and counting them. The censuses of gannets have been undertaken since the 1980s (JNCC, 2015; Murray et al., 2015) and all Scottish colonies were surveyed in 2013 and 2014 via manual approaches (Murray, Harris et al., 2014; Murray et al., 2015; Murray, Smith et al., 2014). In a typical marine survey programme, there might be around half a million images taken over 12 months for a specific area and it is a labour-intensive task to separate this survey into positive images with targeted objects and negative images with no objects, and then count the objects in the images deemed positive. Many surveys acquired by APEM Ltd suggest that more than 95% of the images contain no targeted objects. The detection of small objects, particularly birds, in large-scale images with more than 50 million pixels is a non-trivial task when using manual approaches. Long-term data that utilises standardised and structured methodologies are ideal for quantifying change in species populations; Unfortunately, such data does not exist

for most biogeographic regions (Clements & Robinson, 2022) due to the difficulties and high cost of manual methods. Therefore, automation of this work using an automated intelligent computer system which would help the development of effective prospective environmental models with realistic inputs is highly beneficial.

Despite recent advances in computer vision and learning techniques as well as many attempts to monitor off-shore species in an automated manner, comprehensive large off-shore wildlife censuses are still conducted manually by experienced ecologists, ethologists, ornithologists (e.g., JNCC, 2022; Thompson, 2021) due to unmet expectations in accuracy rates for the counting and classification of species via automated methods as elaborated in Sections 2, 3 and 4.1. With this motivation in mind considering the challenges mentioned in Section 3, this study proposes a new supervised Machine Learning (ML) approach supported by Reinforcement Learning (RL) enabling user-model-data interaction that can detect, split and count birds, in particular, offshore gannets, in an automated decision-making way with high accuracy rates. To clarify the novelty of this paper, particular contributions are outlined as follows.

1. This is the first attempt that explicitly aims to implement maritime bio censuses in marine surveys automatically using an ensemble of supervised ML and RL techniques with a user-model-data interaction in finding the best analysis parameters for mitigating the highly dynamic characteristics of the maritime ecosystem.
2. The two phases of using ensemble techniques within the developed methodology can work successfully in performing the offshore bird censuses and most importantly, the methodology can be generalised to the automated classification and counting of broader maritime multispecies. The methodology can be expandable with more feature extraction techniques in addition to the employed three techniques to achieve higher accuracy rates.
3. The proposed approach shows a new direction for the detection of particular, small species with a diverse background and most importantly for the classification of multispecies even if there is a strong resemblance between them, as seen in bird species, where current techniques (i.e., off-the-shelf approaches (e.g., OBIA), Deep Neural Network (DNN) (e.g., CNN)) cannot converge to a desired solution with high accuracy rates based on the features of datasets.

The remainder of this paper is organised as follows. Section 2 surveys the related literature. Section 4 reveals how the methodology is built up. The implementation of the established methodology in splitting and counting the particular species in surveys is explained in Section 5. The results are presented in Section 6. Discussions are provided in Section 7. Finally, Section 8 draws conclusions and provides directions for potential future ideas.

## 2. Literature review

Wang et al. (2019) reviews studies regarding wild animal surveys based on multiple platforms, including satellites, manned aircraft, and unmanned aircraft systems (UASs), and focuses on the data used, animal detection methods, and their accuracies. The resolution of (sub-metre) satellite images is not sufficient to discern small (<0.6 m) animals at the species level; Manned aerial surveys have long been employed to capture the centimetre-scale images (with a spatial resolution of 2.5 cm Hollings et al., 2018) required for animal censuses over large areas whereas UASs can cover only small areas (Wang et al., 2019). Groom et al. (2013) analysed a very limited number of images (18 frames) within two offshore areas in the Irish Sea using an off-the-shelf object-based image analysis (OBIA) algorithm, aiming at combining manual and automated image analysis, to describe marine bird distributions and abundances. Similarly, Chabot et al. (2018) used OBIA to detect and count Lesser Snow Geese in large numbers of images

<sup>3</sup> Courtesy of the photographer and artist Rahul Alvares.

<sup>4</sup> APEM Ltd is a leading independent environmental consultancy specialising in freshwater and marine ecology. The company is the world's leading provider of digital aerial wildlife surveys for the offshore wind industry, having carried out over 2000 surveys in the North Sea, Irish Sea, Baltic Sea, Pacific, Atlantic, and Gulf of Mexico.

of breeding colonies across the Canadian Arctic, achieving better results compared to human counting. It is noteworthy to mention that the prevalent use of aerial thermal-infrared images for detecting large mammals is of limited applicability to seabirds because of the low pixel resolution of thermal cameras, the smaller size of birds (Chabot & Francis, 2016), and most importantly their low body temperature. Borowicz et al. (2019) established a semi-automated approach using deep learning networks for whale detection from satellite imagery with sub-metre resolution. Kellenberger et al. (2021) developed an approach to automatically detect and count seabirds in UAS imagery using deep convolutional neural networks (CNNs) resulting in low accuracy rates for some types of species regarding the insufficient number of training species for the CNN technique. Again, Dujon et al. (2021) developed a deep CNN using UAS imagery to detect three types of species, in particular, gannets with an overall precision of 0.74. Hong et al. (2019) employed several types of DNNs in non-marine bird detection, resulting in precision values ranging from 85.01% to 95.44%. Hayes et al. (2021) employed CNN in counting two types of birds on the shore in the sitting state using UAS at a close range, resulting in success rates of 97.66% for Black-browed Albatrosses, and 87.16% for Southern Rockhopper Penguins. Close-range use of UAS may disturb wildlife or disrupt their normal activities (Johnston, 2019), especially for flying birds. Akçay et al. (2020) conducted on-ground flying bird detection on bird population movement trends using several DNN techniques with precision values ranging from 0.86 and 0.94. Alqaysi et al. (2021) found the precision values ranging from 60% to 92% for bird detection around wind farms using DNN. There is no guarantee in achieving good accuracy rates using the most popular learning technique, the so-called DNNs. It can be concluded that these approaches require a huge amount of data samples to achieve a satisfactory training outcome (Delhez, 2022). The aforementioned techniques are discussed in Section 7 considering the proposed approach in this study. It is worth discussing the emerging promising approach, namely, Deep Reinforcement Learning (DRL) here as well. Recent revolutionary advances in artificial intelligence (AI) using the learning principles of biological brains and human cognition has fuelled the development and use of Deep Reinforcement Learning (DRL) in numerous fields such as Atari games (Mnih et al., 2015), poker (Moravčík et al., 2017), multiplayer games (Jaderberg et al., 2019), and board games (Silver et al., 2016; Silver, Hubert et al., 2017; Silver et al., 2018; Silver, Schrittwieser et al., 2017). DRL has surpassed human-level performance in many similar applications. It, with goal-directed behaviour and representation learning with the ability to learn different levels of abstraction from data, has emerged as a very effective approach by combining the strengths of two successful approaches – RL and DNN – to overcome the representation problem of RL as function approximators, which generalises knowledge to new unseen complex situations. More explicitly, DRL can be defined as a function approximation method in DNN to generalise past experiences to new situations in complex scenarios by mapping them to near-optimal decisions using scalable and generalisable optimal policies. DRL, in particular, with the most commonly used Deep Q-Networks (DQN), has been found successful in addressing high dimensional problems with less prior knowledge. However, to the best of our knowledge, DRL has been employed for generalising past experiences to a new situation to find the best optimal decision and has yet to be employed for a problem space similar to the one mentioned in this paper. Therefore, this method seems not applicable to our objectives considering the aforementioned problem space which is defined in Section 3.

### 3. Problem definition

Very large areas need to be surveyed in shorter time spans to understand the ecological footprint and to take necessary measures accordingly in a timely manner. Despite recent advances in computer vision and learning techniques as well as many attempts to monitor

off-shore species in an automated manner, comprehensive large off-shore wildlife censuses are still conducted manually by experienced ecologists, ethologists, ornithologists (e.g., JNCC, 2022; Thompson, 2021) due to unmet expectations in accuracy rates for the counting and classification of multispecies via automated methods. Manual approaches increase the cost of surveying large areas significantly and required regular surveys may not be conducted due to this high cost. New automated computer-based approaches are required to observe large areas efficiently and effectively to meet the desired objectives of the research community. We performed a literature survey analysis (Section 2) and conducted several preliminary experiments using the most commonly used techniques to develop the most appropriate approach that can meet the expectations of the research community. The outcomes of our preliminary tests are elaborated in Section 4.1. To summarise considering the survey analysis and preliminary tests specific to the airborne survey data, (i) template-matching approaches (e.g., SIFT) that requires no prior training are far from being able to realise any objectives desired by the research community due to the indistinct features of very small objects within very complex background, (ii) off-the-shelf computer vision techniques (e.g., OBIA) and off-the-shelf ML techniques that require prior training don't result in high accuracy rates due to the indistinct features of very small objects in very big images, and (iii) DNN (e.g., R-CNN), requiring prior training with a large number of data instances, do not converge to a desired solution due to the limited number of instances with the indistinct features of very small objects within a diverse background; Besides, the misclassification of multispecies is high with DNN where data instances in different groups resemble each other too closely as seen in bird species.

The literature, to the best of our knowledge, has a gap that can be filled with the research of computer-automated study analyses of species datasets acquired from the photogrammetry settings which use small aeroplanes to survey very large areas in shorter time spans when compared with other approaches that use static locations, ships or UAS. Due to low accuracy rates in detecting small animals in the marine ecosystem using several off-the-shelf computer vision techniques, off-the-shelf ML techniques, template-matching approaches, and DNN, which is elaborated in Section 4.1 regarding the preliminary experiments with our findings (e.g., the changing and complicated background of the sea, number of data samples in the training set, low-quality images of small species that lack clear features due to them being captured by small aeroplanes with remotely-sensed aerial monitoring photogrammetry settings), we developed a novel approach using an ensemble of ML and RL with a motivation to increase the detection accuracy to reach our target (>0.95) and classify multispecies for the further improvement of the application with multispecies training.

## 4. Methodology

### 4.1. Technical background

Repetitive surveying of very large areas for the purpose of observing trends and population fluctuations, which also use human-dependent approaches, may result in huge financial and time costs. Therefore, sampling is commonly employed to census species within representative sample areas using varying sampling strategies and a way of statistical prediction or projection to a whole figure to avoid high costs where the larger the sample of sites, the better the approximation. However, there can be many sampling biases in such datasets like spatial, taxonomic, or temporal leading to inaccurate inferences: Spatial bias refers to uneven sampling efforts across a region; Taxonomic bias can include over- or under-representation of certain species in the dataset; Temporal bias occurs when records are collected in one season only, or more often at certain times of the year (Jayadevan et al., 2022). Sampling may not be extrapolated to a reliable figure, in particular, for rare species, considering the high percentage of negative images

in whole surveys (> %95) and uneven density and variance in counts of species from one habitat to another, mostly, related to the habitat associations (e.g., food, breeding, sheltering) leading to poor sampling (i.e., oversampling, undersampling), which may produce misleading inferences. Several studies developed particular approaches to mitigate the effect of biases in surveys. For instance, Smyser et al. (2016) utilised a double-observer survey configuration to quantify and correct the bias caused by the failure of observers in aerial surveys. Monitoring all regions of interest and counting all species of interest is crucial to reach highly reliable outcomes and proper decisions with appropriate interpretations. Aerial surveys are an efficient survey platform, capable of collecting wildlife data rapidly across large spatial extents in short time frames; however, these surveys can yield unreliable data if not carefully executed (Davis et al., 2022). To this end, numerous approaches such as entropy-based information screening method (Li et al., 2021) and normalised double entropy (NDE) (Li et al., 2023) were developed to distinguish bad and redundant image data to increase the quality of sampling.

As an active research direction for decades, object recognition and detection have had increased importance within many fields such as nature, biometrics, medicine, and robotics. Current clustering algorithms, in which no prior training is performed, on visual datasets, are not successful in grouping similar objects with high rates of accuracy, particularly, for objects with very complex backgrounds (Kuru & Khan, 2018). One of the oldest methods of object recognition is the template-matching approach. It consists of sliding a particular template over the search area (usually an image in which we are trying to locate) and at each position, calculating a distortion or correlation measure that estimates the degree of dissimilarity or similarity between the template and the candidate (Reyes, 2014). Then, the minimum distortion or maximum correlation position (depending on the implementation) is taken to represent the instance of the template into the image under examination. There are various ways of calculating the degree of dissimilarity or similarity, such as the Sum of Absolute Differences (SAD) and the Sum of Squared Differences (SSD). The Normalised Cross-Correlation (NCC) is by far one of the most widely used correlation measures (Stefano et al., 2003; Yang, 2010). Recently, several well-advanced template-matching techniques have been developed to detect objects automatically. These off-the-shelf template-matching techniques are scale-invariant feature transform (SIFT), speeded-up robust features (SURF), features from accelerated segment test (FAST), binary robust independent elementary features (BRIEF), oriented FAST and rotated BRIEF (ORB), maximally stable extremal regions (MSER) and binary robust invariant scalable key points (BRISK). In these techniques, a similarity value regarding the specified number of most important key points is utilised to determine if there is a similarity between the reference object and the objects in images, videos, or real-time scenes given a threshold value. No pre-processing and training is required. We tested these approaches on our sample datasets and the preliminary results indicated that none of these approaches is successful enough to detect and split very small birds with many different postures in large-scale images against the changing and complicated background of the sea (Ex: Figs. 6, 16). It is noteworthy to mention that variations in sea-state, marine environments, atmospheric conditions, and solar illumination angles combine to produce a wide range of sea surface image patterns that form the background to the targets of a bird mapping operation (Groom et al., 2013).

The other approach is the supervised ML approach, which requires prior datasets to both determine the common features and train the system for further similar detections based on these features. Accuracy rates of detection are mainly dependent on the quality of datasets used in training in terms of representing the real environment by avoiding overfitting. In the training process, general features are acquired and these features are then compared to the features of objects in test datasets to observe how well the features are detected and to determine if these features are suitable to be employed in real life. Trained

models (i.e., detectors) are used for the detection of similar objects after the evaluation is conducted successfully by using an evaluation dataset. Our preliminary tests on the sample datasets using the supervised ML approaches showed promising results, which is elaborated in Section 4.2. The frequent low numbers of marine birds in any given area adds to the complexity of developing methods for large-scale operational surveys (Groom et al., 2013). Most of the time, there might be a single gannet in a large-scale image (Ex: Fig. 16) within our surveys. This makes detecting them highly difficult with regards to splitting the images with gannets from those without gannets, for aerial surveys with more than half a million images, into the positive folder. In other words, it would be easier to detect at least one gannet among several gannets in a large-scale image rather than detecting a single gannet in the image.

To summarise, as explained above, our preliminary test results showed that employing a template matching approach did not work for detecting and splitting birds in large-scale aerial images, because, despite their distinctive features (Ex: Fig. 1) the birds are not very clear in very complex and changing sea textures despite the high quality of the images with a very high camera resolution (i.e., > 50 Megapixels). Moreover, DNN techniques do not result in satisfactory outcomes where the number of instances in domain sets is not many as in our case in this study even though they are recently popular and successfully employed in many different types of application fields and these techniques have far exceeded the accuracy rates of current ML methods. More importantly, our preliminary test using DNN showed that the misclassification of multispecies is high if data instances in different groups resemble each other too closely as seen in bird species. Therefore, we have employed an ensemble of ML and RL techniques for automated recognition, splitting, and counting of birds in aerial surveys to both reach our goals in accuracy rates and classify multispecies in the further development of the proposed application and a user-friendly application was developed using Matlab Simulink MatWorks R2020,<sup>5</sup> as displayed in Fig. 2. The algorithms were developed to work on any size of bird objects using interpolation and extrapolation techniques, providing there is a training data set available. In particular, the methods of the sliding window (Forsyth & Ponce, 2012) and Gaussian pyramid (Witkin, 1984) are applied to detect any object that can appear in different regions of the image and in different scales. A detection window in the sliding window method slides over the image to extract the regions. The Gaussian pyramid (Witkin, 1984) method is primarily applied to the image during the detection stage of the sliding window to operate a scale search.

Three feature extraction techniques are employed in our methodology, namely Haar Cascades, Local Binary Patterns (LBP), and Histogram of Oriented Gradients (HOG). Each of these techniques acquires different features of objects using different mathematical modelling. We applied these techniques to establish the detectors in our implementation using Matlab ready-to-use commands along with the Viola-Jones matching technique.<sup>6</sup> (i) Haar cascade technique resembling Haar wavelets was first introduced by Papageorgiou et al. (1998) and Viola and Jones (2001). First, the pixel values inside the black area are added together; then the values in the white area are added together. Following that, the total value of the white area is subtracted from the total value of the black area. This result is used to categorise image sub-regions (Cruz et al., 2015), which requires a fair amount of time to train a classifier and generate the Haar training set. The calculation method of Haar-like features is faster by introducing an integral image or summed-area table (Viola & Jones, 2001), which makes the computing of Haar-cascade classifiers more efficient. (ii)

<sup>5</sup> MATLAB and Simulink are registered trademarks of the MathWorks, Inc. See [www.mathworks.com/trademarks](http://www.mathworks.com/trademarks).

<sup>6</sup> <https://uk.mathworks.com/help/vision/ug/train-a-cascade-object-detector.html>.

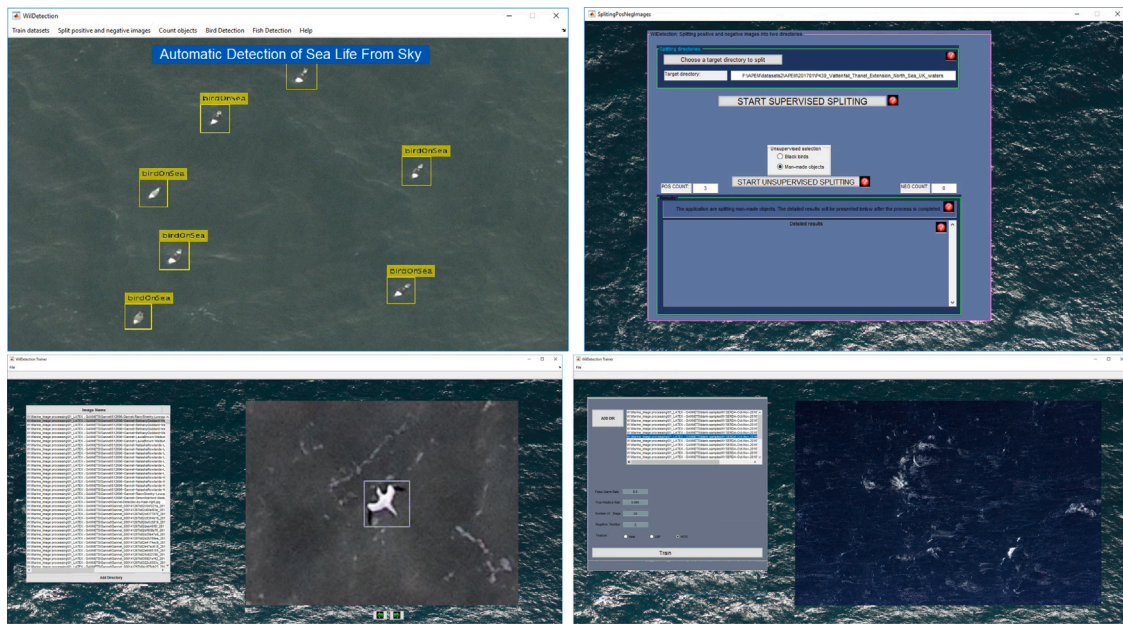


Fig. 2. Interfaces of the application from top to bottom: (i) the main, (ii) training for ROI selection, (iii) training for blank set and parameter selection and (iv) recognition/splitting.

LBP was first introduced by Wang and He (1990) and analysed in detail by Ojala et al. (1994). It has been improved by several other studies regarding object identification and recognition (Ojala et al., 2002; Trefný & Matas, 2010; Zhang et al., 2007). In the LBP technique, the texture is defined as a function of spatial variations in the pixel intensity of an image with a low computational cost by focusing on a small set of critical features, discarding most of the non-critical ones to increase the speed of the feature extraction and classification significantly without affecting accuracy; common features, such as edges, lines, points, flat areas, and corners can be represented by a value in a particular numerical scale (Cruz et al., 2015). Therefore, it is possible to recognise objects in an image using a set of values extracted a priori and several weak classifiers turn into a strong classifier regarding recognition (Cruz et al., 2015). (iii) HOG which explores gradient information and local shape information was first explored by McConnell (1986) and improved by Dalal and Triggs (2005). The technique counts occurrences of gradient orientation in localised portions of an image, which is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalisation by the distribution of intensity gradients or edge directions. Due to the strong texture and shape description ability, HOG can be used in the detection of many different types of objects. It is highly sensitive to object orientation. It responds rapidly to changing parameters of FAR and TPR based on its feature extraction method which uses histograms. (iv) The Viola–Jones technique that is included in Matlab Computer Vision System Toolbox (i.e., vision.CascadeObjectDetector) is used to match acquired features in detectors to those of the objects in images for comparison and detection. This technique along with feature extraction techniques is highly sensitive to different orientations of objects in images/videos. The main reasons for choosing Viola–Jones are its fast detection speed and its high accuracy detection rate regarding the large-scale aerial images on which we are working. How these techniques are employed in a novel approach in our methodology is explored in the following sections, particularly, Sections 4.2 and 5.

The main components of the platform, WILDetect, built in this study are depicted in Fig. 3. The phases are (i) data preparation/pre-processing (A.1), (ii) feature extraction/training (A.2), (iii) viability testing of the detectors and specifying the best detectors (A.3), (iv) implementation of the model in splitting and counting in surveys (A.4) (i.e., determining the best detectors in splitting and counting using the recursive RL approach (A.4.1), recognition and splitting (A.4.2.Phase1),

recognition and counting (A.4.2.Phase2)), and (v) database operations (A.5) that are explained in the following sections respectively.

#### 4.2. Establishment of the methodology

The defined problem space (Section 3), considering the literature analysis (Section 2) and the obtained results from the preliminary tests (Section 4.1) using off-the-shelf approaches, necessitates the development of a new approach to achieve the objectives of the research community while performing airborne wildlife census automatically in the marine ecosystem. With this in mind, the approach built here is explained step by step in the following subsections (Sections 4.2.1, 4.2.2 and 4.2.3) and the results of the implementation using large surveys are provided in Section 5).

##### 4.2.1. Data sets, data preprocessing/preparation (A.1)

The main subcomponents of this phase along with their interaction are illustrated in the dedicated section of Fig. 3 titled “A.1”. A dataset consisting of images with the object of interest and a dataset consisting of blank/background images that represent anything except the object of interest are needed to establish a supervised ML approach for training, testing, evaluation, and validation. Data preparation and data management in those steps are demonstrated in Fig. 4. The negative set typically contains more images than the positive set in order to complete the training phase where every positive image needs more background images that represent the real-world environment. APEM has many surveys in its repository in which almost 95% of the images are blank background images with no targeted object types. APEM conducts offshore digital wildlife surveys for the offshore renewables sector, reliably capturing imagery all year round in all lighting conditions and sea states up to four. The data is captured on a variety of sensor formats including both 35 mm and medium format from various manufacturers, in both single camera and multiple camera configurations, depending on the project requirements. The images are collected by these advanced cameras mounted in a small twin-engine aeroplane (Ex: Fig. 5) within a route in which all regions of interest are surveyed.

A snag library that consists of around 1 million snags (i.e., cropped images with objects of interest; ex: Fig. 6) has been established by APEM. We aimed to incorporate all possible targeted positive images

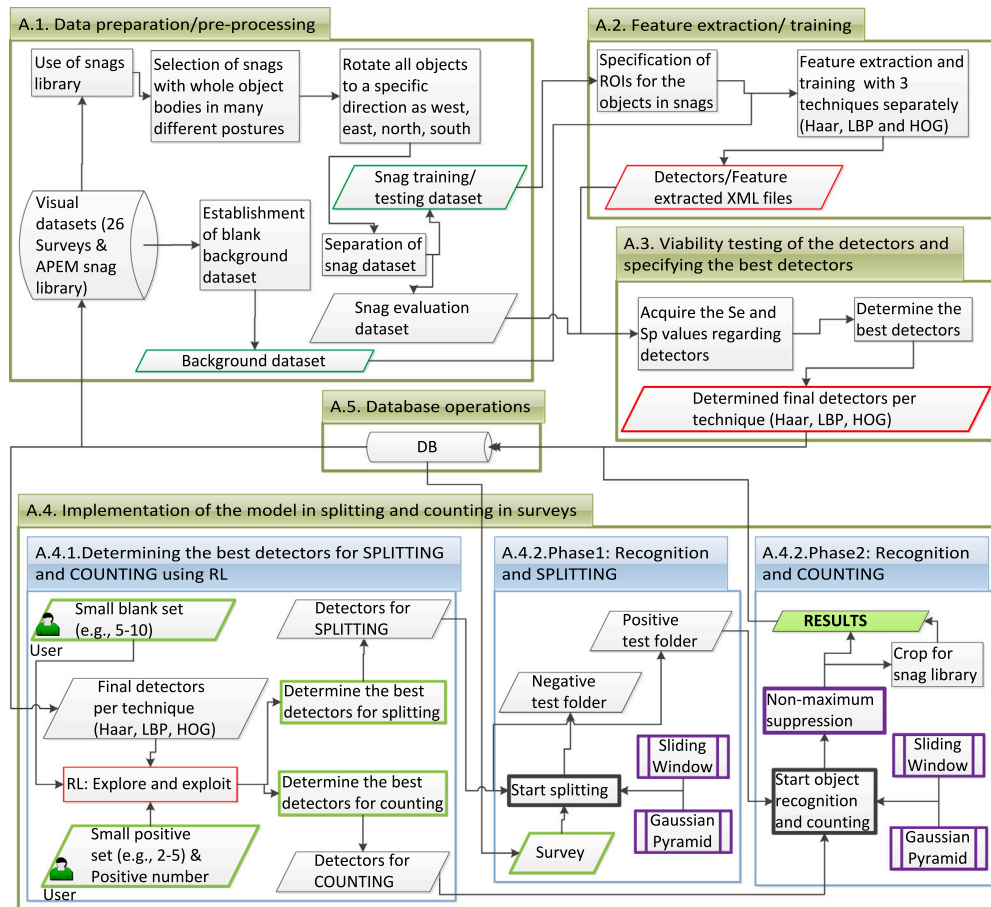


Fig. 3. Main components of the methodology.

into the methodology, either for training/testing or evaluation and validation to create a positive dataset that can represent the real-world object types by avoiding overfitting during the decision-making phase of the implementation in real-field tests. We pre-processed the gannets in this library by selecting the convenient gannet samples. Our preliminary tests showed that flying gannets with their partial body parts can be detected using whole body sets, but, a whole gannet body cannot be detected by a trained set that consists of various partial parts of gannets (e.g., only one wing). Furthermore, partial body parts can increase the false-positive (FP) rate. Therefore, in this phase, we aim to select as many gannets as possible that have whole bodies (i.e., two wings, head, and tail), but in all possible postures. With this in mind, we prepared two sets of gannets (50%/50%), one of which is for training/testing with 1073 snags (Fig. 4I) and the other one is for evaluation with again 1073 snags in many different postures (Fig. 4II). Our preliminary test results suggest that the detectors built using the three feature extraction techniques (i.e., Haar, LBP, HOG) based on the specific orientations (i.e., north, east, south, west) improve the accuracy rate significantly where these techniques are highly sensitive to different orientations of objects in images as explained in Section 4.1. Therefore, all the gannet objects in these sets are rotated into 4 directions automatically using the codes produced in this study for the data preprocessing phase, namely, north, east, south, and west, by which 4 sets of gannet objects totalling  $1073 \times 4 = 4292$  were generated for training/testing and evaluation, rather than separating them into these directions into 4 groups, which would reduce the number of objects substantially. In this way, 4 types of detectors are needed with the orientations north, south, east, and west, as well as a large number of negative images. The greater the variety of these snags/images representing the real environment, the better the detectors avoiding overfitting and consequently the higher

the accuracy of detecting targeted objects in images in real field tests. A sub-sample of the dataset in which all gannets are almost rotated to the north is presented in Fig. 6. More snag examples can be found in our technical report — MarineObjects\_Gannet\_Supplement\_2.pdf in the supplementary materials. Moreover, the gannet objects in large-scale images (Ex: Fig. 16) are presented in our technical report — MarineObjects\_Gannet\_Supplement\_3.pdf in the supplementary materials with many different postures and background textures.

In addition to the positive dataset, a blank/background/negative dataset was established using 26 surveys collected by APEM between 2014 and 2017. These surveys were acquired from different parts of the world in different seasons and time zones using different settings and types of image-capturing technologies. The texture of the negative images in these surveys differs from each other as displayed in Fig. 7, which makes the implementation more challenging. More examples specific to the surveys can be found in our technical report — MarineObjects\_Gannet\_Supplement\_1.pdf in the supplementary materials. We were given around 1 million images that are the subsets of these surveys. We used this large number of surveys, a volume of around 10 TB, to find out the general characteristics of aerial surveys. The diverse features revealed from these large surveys help make our approach strong and promising for further use of the application in any circumstances while separating targeted objects from their background. This large dataset was stored in high-powered servers and processed using these servers (A storage unit (12 TB), 2 Novatech servers and 5 HP servers connected to each other via the network. The storage unit is used for placing the big size of the datasets and applications on servers are run using the datasets placed in the storage unit for development, evaluation and validation. The specifications of the Novatech servers: Intel (R) Xeon (R) CPU E5 26300 2.30 GHz 2.30 GHz (2

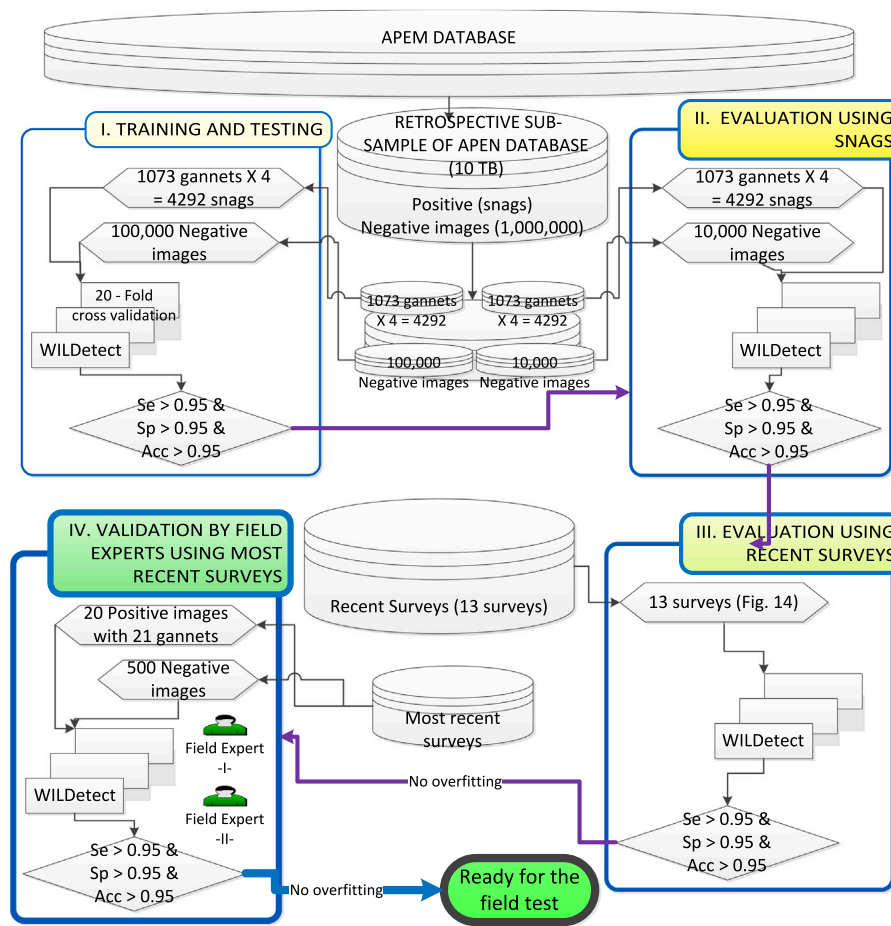


Fig. 4. Use of datasets during training, testing, evaluation, and validation of WILDetect.



Fig. 5. APEM aeroplane during a remotely-sensed aerial survey using advanced aerial high resolution photogrammetry.

processors), 64 bit, 64 GB RAM, GPU (NVIDIA GeForce GTX 680). The specifications of the HP servers: Intel (R) Xeon (R) CPU 5160 3.00 GHz, 64 bit, 8 GB RAM. We established a sub-sample set from the diverse surveys that consisted of 100,000 images (Fig. 4I) to use in the training process, with the aim of incorporating all the characteristics of the current and future surveys into implementation. It is worth emphasising that an equal number of negative images from all sub-surveys (107 sub-surveys), within the above-mentioned 26 surveys, were included considering the seasons and time zones to create a negative dataset that can represent the real-world circumstances. Rather than using 1 million images, this sub-sampled set would reduce the processing time of training significantly, in particular, while singling out the consecutive new sets for each following training iteration, which is elaborated in Section 4.2.2. Readers are referred to Fig. 4 in the related

sections below in which the evaluation and validation are explained after revealing the establishment of the methodology in the following sections.

#### 4.2.2. Feature extraction and training (A.2)

The main subcomponents of this phase along with their interaction are illustrated in the dedicated section of Fig. 3 titled “A.2”. Automatic detection systems usually require large and representative training datasets to achieve good detection rates with fewer FP rates (Vállez et al., 2015). The training phase is very important for the successful recognition of objects in the further use of the application. One badly trained file/classifier can cause the splitting process (A.4.2. Phase1 in Fig. 3) to function poorly and many positive images may be placed in the negative folder and vice versa, which we aim to avoid. The user interface developed for the training phase is displayed in Fig. 2ii and iii. With this interface, the detectors can be generated using several parameters such as true positive rate (TPR), false alarm rate (FAR), number of training stages, number of background images, and negative sample factor (NSF), with respect to the number of positive images in each training stage and the feature extraction techniques, i.e., Haar, LBP, and HOG. A mathematical model of the objects is extracted using these techniques as explained in Section 4.1. These techniques were selected, because, in addition to providing detectors with encouraging accuracy, they produce detectors that can function efficiently. For instance, objects can be detected in a few seconds in an image with 50 million pixels. The training interface lets the user feed the system with positive images for ROI selection and negative images for background analysis, as well as specify the parameter values. ROIs are specified in positive images by the user (at least one ROI in each image), and the feature descriptors are extracted based on ROIs

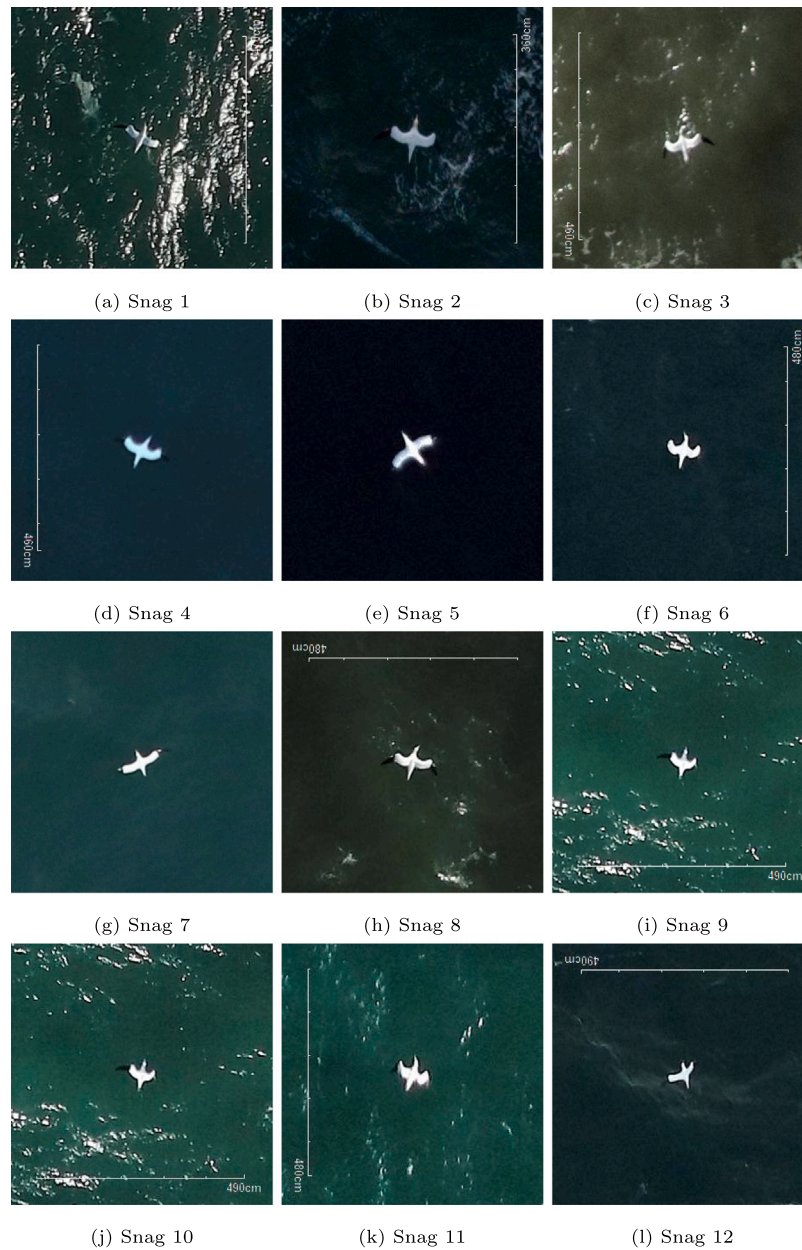


Fig. 6. Examples of the gannet snags for the north direction in several postures and background textures.

using the aforementioned techniques in the training process. Several training sets were acquired using different FAR and TPR parameters for each feature extraction technique. In each training, the number of training stages was 20 (i.e., 20-fold cross-validation) along with the number of the negative samples 3, which means that the number of the different negative images to be used in each training stage of the 20 iterations would be as many as 3 times the number of positive images. Our preliminary tests show that (1) decreasing the number of iterations (e.g., 10-fold) increases the training time significantly, (2) the recognition accuracy rate is almost the same with negative sample factors of 3 and 10; however, the processing time increases significantly with the value of 10. Therefore, the training parameters 20 for iterations rather than most commonly used 10-fold and 3 for negative sample factor were selected to decrease the training time. In each iteration, the techniques choose a set of different negative images in the negative dataset whose texture features are supposed to be different from the previously selected sets. The system stops if not sufficient negative images with different features are provided.

Therefore, the images in the negative dataset must be different from each other with respect to their textures. A large number of images in the negative dataset increase the chance of finding a new set for each following training iteration. As explained earlier, 100,000 images selected for the negative datasets from different surveys provide enough distinctive iteration sets for our training iteration steps.

The training process is repeated to obtain several detectors using different parameters, in particular, reducing the values of TPR and FAR to flag fewer FPs. This is mainly beneficial to the analysis of different types of surveys with regard to their varying textures, as explained in the following sections. As soon as the detectors are generated, they are tested on the sample test dataset and the threshold parameters are reduced until almost all negative images are transmitted into the negative directory. This may cause several positive images to be missed with respect to each technique with reduced threshold parameters. However, these techniques use different features and if one detector with a technique misses one positive image, there is a high probability that one of the other two detectors using the other two techniques may



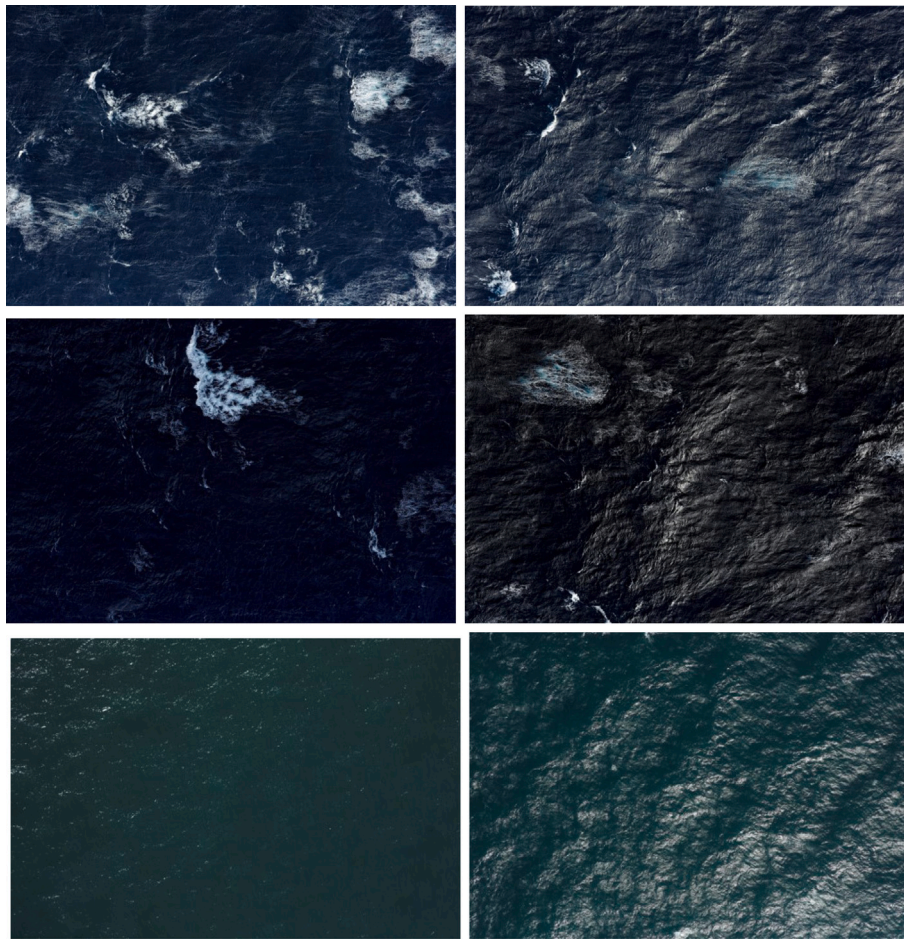


Fig. 7. Examples of blank images from 6 different surveys with various textures.

Table 1

Processing time of the detectors in days.

Techniques	0.30–0.985	0.35–0.985	0.40–0.985	0.40–0.995	0.45–0.995	0.50–0.995
Haar	51	45	38	31	28	23
LBP	31	26	22	18	16	13
HOG	43	38	33	27	25	21

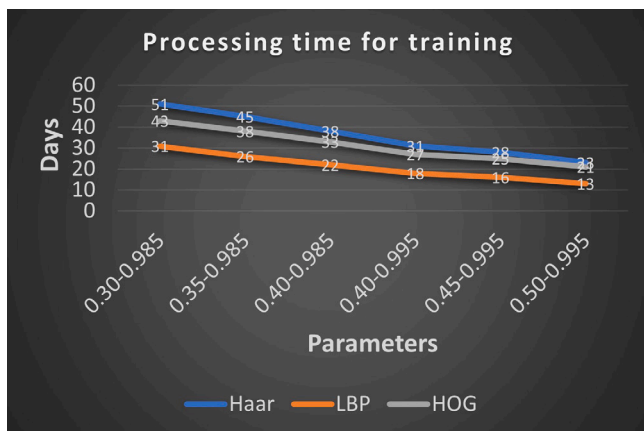


Fig. 8. Processing time of detectors in Table 1.

overcome the reduced sensitivity (Se) because of the FNs with respect to each technique in order not to miss any positive image, which is explained in Section 4.2.3 in detail with examples.

Detectors for the specific types of objects are created only once and can be used whenever needed to recognise, split and count specific objects in images for further analysis. Six trained sets — detectors consisting of 72 trained files (i.e., 6 threshold values × 3 techniques × 4 directions = 72) were created using 6 threshold values, as displayed in Table 1. In other words, 12 trained files were obtained for each trained set, 4 for each technique (i.e., Haar, LBP, HOG) and each of which represents the gannet sets in one of the four directions (i.e., north, east, south, west) (i.e., 12 trained files for each detector × 6 detectors = 72). The processing time of the training in terms of threshold values is shown in Table 1 and Fig. 8. The smaller the threshold values, the longer the training time.

4.2.3. Viability testing of the detectors and specifying min/max threshold parameters (A.3)

The acquired trained files were evaluated on the evaluation dataset (i.e., 1073 snags in four directions) spared for evaluation (Fig. 4II) as mentioned in Section 4.2.1. The evaluation results are presented in Table 2 and Fig. 9. As it is noticed in Fig. 9, the detection success of

specify this image as a positive image. Therefore, we are employing these three techniques at the same time for the splitting phase to

**Table 2**

Accuracy rates of the training phase with the snag dataset based on the detectors with 6 different parameters: all snags are recognised successfully by the parameters, FAR = 0.50 and TPR = 0.995 with the combination of 3 techniques.

Techniques	0.30–0.985	0.35–0.985	0.40–0.985	0.40–0.995	0.45–0.995	0.50–0.995
Haar	0.825	0.975	0.98	0.985	0.989	0.993
LBP	0.675	0.925	0.935	0.966	0.984	0.992
HOG	0.735	0.85	0.9	0.993	0.994	0.994
<b>Combined</b>	<b>0.84</b>	<b>0.99</b>	<b>0.992</b>	<b>0.995</b>	<b>0.996</b>	<b>1</b>

the feature extraction techniques varies depending upon the approaches followed in these techniques as elaborated in Section 4.1 as the parameters concerning the features of the datasets changed. For instance, the effect of the HOG technique is relatively poor when the parameters are small, and it increases rapidly after the values of parameters are increased. In this way, the drawbacks of one technique considering the features of data can be compensated by the other two techniques while the parameters need to be changed for achieving the desired goals, either for increasing  $Se$  or for increasing  $Sp$ . The trained files with the parameters FAR = 0.30 and TPR = 0.985 resulting in a  $Se$  value of 0.840 are excluded from the trained folder in order not to be used for further recognition and splitting process. Because the main objective of this research is to obtain a  $Se$  value greater than 0.95 which is one of the targeted success criteria, i.e., threshold level, as shown in Fig. 9 with the green line. In other words, we do not want to miss positive images at any cost even with small  $Sp$  values by achieving this success criterion. As explained in Sections 5.1 and 5.2, the system with established detectors was run on various evaluation and validation surveys (Fig. 4III and IV) with varying characteristics to find out the detectors' viability on further surveys based on the observed  $Se$  and  $Sp$  values, strictly speaking,  $Sp$  after achieving a satisfactory  $Se$  value with 5 threshold intervals, all of which are above the targeted sensitivity value, 0.95.

The use of three feature extraction techniques at a time is more important where the detectors with smaller threshold parameters are selected by the system with the RL approach as explained in the following Section 5. Some of the gannet objects detected by only one of the feature extraction techniques are presented in Fig. 10 where FAR = 0.35 and TPR = 0.85. These three gannet objects are detected by the three techniques at the same time with bigger threshold values where FAR = 0.50 and TPR = 0.95. We would like to note that these high threshold values may cause many FPs depending on the complexity of the background and it may not be a good option to use them for particular types of surveys, which is explained in the following sections in detail.

## 5. Implementation of the methodology in splitting and counting (A.4) using the recursive RL technique

Objects can appear in different regions of the image and in different scales. In order to solve this problem, the sliding window method (Fig. 3) is used (Forsyth & Ponce, 2012). It consists of a detection window that slides over an image extracting regions and classifying them. A Gaussian pyramid (Witkin, 1984) (Fig. 3) is also applied to the image to perform a scale search to detect similar objects in different sizes.

A multi-threaded approach was established to speed up the calculations and reduce the processing time. In this multi-threaded approach, jobs are distributed among the resources in the same network, particularly among the multi-core processors, with one job for each core. The user can choose one of the two processing options, either multi-threaded where powerful computing resources can be deployed to perform many tasks at once, or sequentially where operations are performed in order and results can be followed by the user per image. The multi-threaded option reduces the processing time significantly based on the power of the resources used. Some of the resources in use can be stopped to be used for other purposes, and vice versa, new

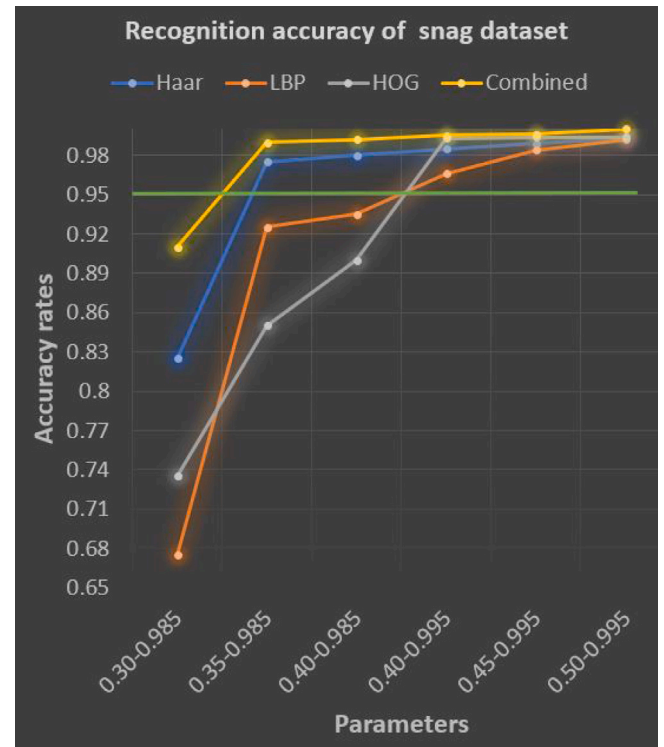


Fig. 9. Results for Table 2: The accuracy rate of recognition is increased by combining 3 techniques, which is depicted by the yellow line. Combination of 3 techniques is more important where the FAR and TPR parameters are smaller to acquire a satisfactory recognition rate. The horizontal green line drawn on 0.95  $Se$  is the objective threshold level; the  $Se$  values over this line are acceptable in terms of the yellow line.

resources can be incorporated into the system while the splitting or counting process is ongoing, using a novel flexible cloud computing approach built in this study.

It is worth noting that datasets are imbalanced -i.e., not uniform within surveys most of the time as mentioned earlier regarding the larger number of negative images (negative class) compared to a smaller number of positive images (positive class). This imbalance is mitigated using an ensemble of ML and RL techniques within the research in two phases of automated data analysis. The selection of the best detectors in the splitting phase is based on the features of the background to discard most of the negative images while aiming to place all the positive images in the positive folder whereas it is based on the features of the targeted objects in the counting phase to count all the objects in the images placed in the positive folder while aiming to discard all the remaining negative images placed in the positive folder during the splitting phase. Four values are measured to assess the obtained results, namely,

$$\begin{aligned}
 \text{Sensitivity} = Se &= \text{TruePositiveRate(TPR)} = TP / (TP + FN), \\
 \text{Specificity} = Sp &= \text{TrueNegativeRate(TNR)} = TN / (TN + FP), \\
 \text{Accuracy} = Acc &= (TP + TN) / (P + N) = (TP + TN) / (TP + TN + FP + FN), \\
 \text{Precision} = Pr &= TP / (TP + FP)
 \end{aligned} \tag{1}$$

The first three values —  $Se$ ,  $Sp$ , and  $Acc$  — are explained in Section 6 in detail based on the data analysis of the particular approaches.  $Pr$  is

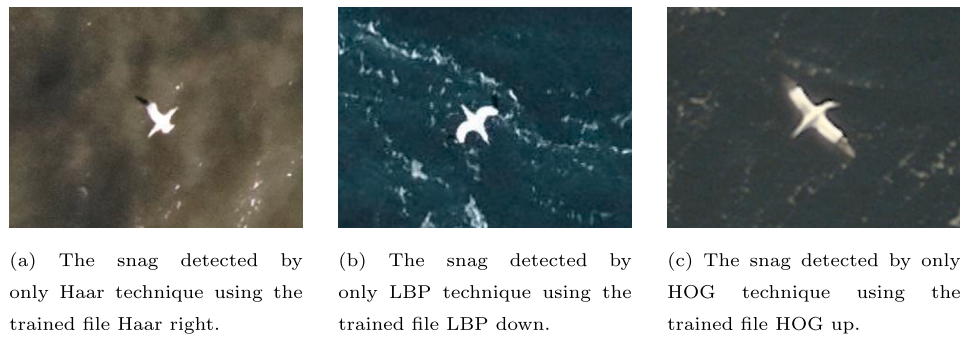


Fig. 10. Snag examples detected by only one of the techniques with the parameters of FAR = 0.35 and TPR = 0.85.

mainly employed to identify the class imbalance problem and assess how imbalanced data in favour of “negative images” that may lead to large FPs is influencing the results. More specifically, this assessment helps to understand (i) if the high values of  $Se$ ,  $Sp$  and  $Acc$  are biased and most importantly (ii) if the two phases of using an ensemble of learning techniques help alleviate the bias regarding the improvement in  $Pr$  through obtaining the final counting results. The low values of  $Se$ , e.g., < 0.80, require the implementation of cost-sensitive analysis (CSA), as we conducted in our previous research in Kuru et al. (2013) to get more reliable improved results. In CSA, classes have different costs associated with them using weights with respect to the number of instances; the classes with fewer instances, i.e., under-represented classes (positive cases in this research) are assigned higher costs (i.e., adding cost-sensitivity, e.g., P:N = 10:1) to reduce the number of false predictions, particularly in favour of the class with less number of instances, and consequently increase the reliability of the results related to that class by assigning different penalties to misclassification of samples (Kuru et al., 2013) in which there is a trade-off between  $Se$  and  $Pr$ .

### 5.1. Implementation of the platform in splitting

Most of the time, more than 95% of images in a survey contain no targeted objects, and therefore this phase of the implementation aims to separate out the images with no targeted objects in a reduced overall processing time. Strictly speaking, the main objective of this phase is to perform the best splitting between negative and positive images based on the parameters specified in Section 4.2.3. The negative images are placed in the negative folder and the positive images are placed in the positive folder. Then, the images in the positive directory are analysed in detail to locate all targeted objects, which is explained in Section 5.2. The methodology selects a set of detectors for each feature extraction technique to deploy during the splitting process based on the particular characteristics and specific patterns of the images in surveys. This step is explained in Section 5.1.1. Then, how the splitting is performed is explored using these selected detectors in Section 5.1.2.

#### 5.1.1. Pattern recognition and specification of the best feature extraction detectors for splitting using RL (A.4.1)

The methodology chooses the best detectors regarding separating negative images from positive images successfully based on the texture patterns and characteristics of the images in the surveys using the user-model-data interaction as illustrated in Fig. 3, A.4.1. The components of the recursive RL algorithm employed in this phase are demonstrated in a broader perspective in Fig. 11 and the main steps are explained as follows.

First, a very small subset of the negative images (i.e., 5–10) representing the whole of the negative images (i.e. background) in the survey is selected by the user. The characteristics of this very small set play an important role in determining the best convenient detectors. Therefore, the user is expected to choose blank images that have

diverse background textures in the survey. For instance, at least a blank image taken from each camera mounted on the aeroplane and blank images taken from different time intervals should be placed in this set in order to represent the background characteristics of the whole survey. Alternatively, processing of the images from different cameras or in different time intervals – subsets of surveys – can be conducted separately, which can increase the efficacy of the platform further.

Second, the blank images selected by the user are processed by the approach to determine the best detectors for each technique based on the observed  $Sp$  (i.e.,  $TN/(TN+FP)$ ) values. In this process, a screening test is performed with preferably higher  $Sp$  values to increase the chance of placing images with no targeted object in the negative folder. In other words, the FP cases are reduced to a minimum resulting in very high  $Sp$  values with an ability to correctly place the negative images in the negative folder and this means that if an image is tagged as negative, it is a high probability that there is no object in that image. The RL algorithm makes the detectors run on the sample negative images fed by the user using the Viola–Jones matching technique to single out the successful detectors for splitting based on the characteristics of the background texture. This process starts from the detectors with the highest threshold values (i.e., 050–0995 in Table 2) that may result in many FPs reducing  $Sp$  whereas the background has a complicated texture. However, no FP may be obtained if the background has a clear texture. This iterative process using predetermined nominated detectors (Fig. 9) proceeds (Fig. 11) until no FP is obtained per detector where  $Sp$  is 1. In other words, the process stops per detector where  $Sp$  is 1 and the detector is selected at this stage in which a satisfactory pattern is observed and learned by the system. Otherwise, the last detectors with the smallest threshold values are processed where the  $Sp$  may be slightly smaller than 1 and they are selected for splitting.

Finally, the methodology determines the most suitable detectors for each technique (i.e., Haar, LBP, HOG) through the detector sets trained previously as depicted in Table 2 that are above the green line in Fig. 9. The results of the RL process for the 13 surveys regarding the selection of the detectors for splitting are explained in Section 6.1.

#### 5.1.2. Object recognition and splitting (A.4.2.Phase1)

The detectors determined by the RL approach at the start of the splitting process as explained in Section 5.1.1 are utilised in this phase. The methodology makes these detectors run on all the images using the Viola–Jones matching technique and the images are placed in the negative directory if they are specified as negative; in other words, these are the images in which no object is detected by any of these detectors. The images are readily placed into the positive directory when an object is detected by any detector without screening the image for other objects using the remaining detectors. The main aim is to increase  $Sp$  by reducing FPs with respect to each technique, but to increase  $Se$  using 3 techniques at the same time by reducing FNs regarding the number of positive images (see Fig. 10). The higher the number of objects in an image, the more likely that the image will be put into the positive directory. The splitting phase was evaluated on several surveys (Fig. 4III) and the results (Table 3) are explained in Section 6.1.

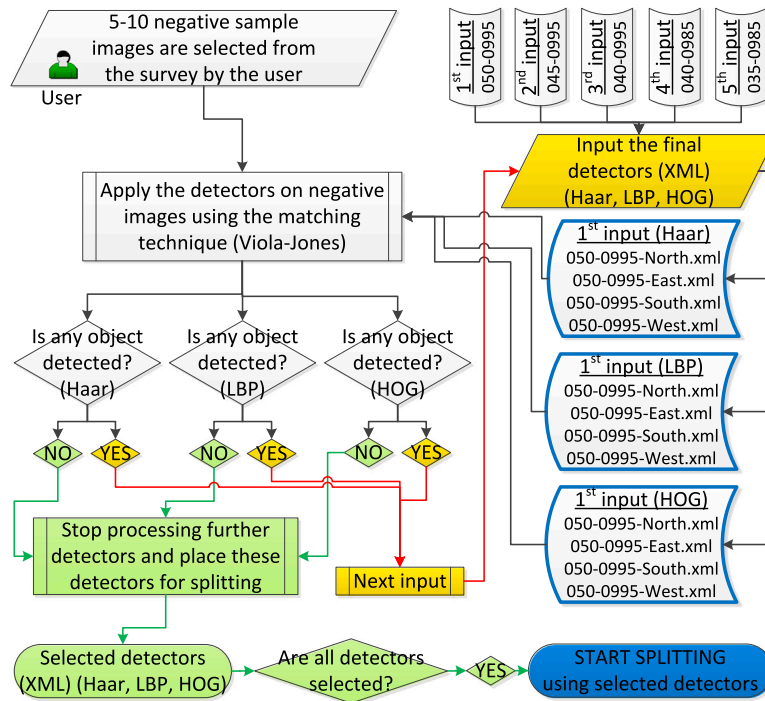


Fig. 11. Use of Reinforcement Learning (RL) for selecting the best detectors for splitting.

### 5.2. Implementation of the platform in counting objects

In the splitting phase, the application places any image into the positive directory when an object is detected without screening the image for other objects using the remaining detectors. In this way, the processing time of the splitting is reduced significantly. On the other hand, the aim of the counting phase is to detect every targeted object in images placed in the positive directory. New detectors are selected to complete this task using a similar recursive RL approach explored above, but differently as explained in Section 5.2.1 in order not to miss any targeted object in the positive images.

#### 5.2.1. Pattern recognition and specification of the best feature extraction detectors for counting using RL (A.4.1)

The methodology chooses the best detectors regarding counting objects in the images placed in the positive folder based on the particular patterns and characteristics of the objects in those images as illustrated in Fig. 3, A.4.1. The main objective of this phase is to detect and count all the targeted objects successfully. The components of the recursive RL algorithm employed in this phase are demonstrated in Fig. 12 and the main steps are explained as follows.

First, a very small subset of the positive images (i.e., 2–5) in the survey are selected and all the objects in these images are outlined with a bounding box along with the targeted object counts per image by the user upon the interface provided in the application during the selection of a subset of the negative samples at the start of the survey analysis as mentioned in Section 5.1. In other words, the detectors for both splitting and counting are designated by the recursive RL approach before the survey analysis starts. In this way, the methodology carries out the counting process automatically after the splitting phase is completed.

Second, these selected positive images are processed with respect to the user-specified objects by the RL approach to determine the best detectors for each technique based on the observed  $Se$  (i.e.,  $TP/(TP+FN)$ ) values. In this step, an object recognition test is performed with preferably higher  $Se$  values to increase the chance of detecting a targeted object in the positive folder. In other words, the FN cases are reduced substantially with respect to the targeted objects, preferably to zero,

resulting in very high  $Se$  values with an ability to correctly detect the objects in the images. The RL algorithm makes the detectors run on the sample positive images using the Viola–Jones matching technique to single out the successful detectors for counting by referencing the users’ object inputs from the selected positive images. This time, different from the splitting phase, the process starts from the detectors with the lowest threshold values (i.e., 035–0985 in Fig. 9) that may result in many FNs which may reduce  $Se$ . This iterative process proceeds until no FN is obtained per detector where  $Se$  is 1. In other words, the process stops per detector where  $Se$  is 1 and the detector is selected at this stage in which a satisfactory pattern is observed and learned by the system. Otherwise, the last detectors with the highest threshold values are processed where the  $Se$  may be slightly smaller than 1 and they are selected for counting. Additionally, the last detector with the highest threshold values may result in several FP where the images have complex backgrounds, which may reduce  $Sp$  of the system at this stage. But, the objective is to detect all targeted objects successfully with a high  $Se$ , preferably 1, as specified earlier even though compromising  $Sp$  slightly. The use of multiple designated detectors at a time in a collective way aims to ensure a high  $Se$  — one of the other two detectors can detect an object if it is missed by a detector.

Finally, the methodology places the most convenient detectors for each technique (i.e., Haar, LBP, HOG) through the detector sets trained previously as depicted in Table 2 that are above the green line in Fig. 9. The results of the RL process for the last survey (Table 3) regarding the selection of the detectors for counting are explained in Section 6.2 (Fig. 4III).

#### 5.2.2. Object recognition and counting of objects in surveys (A.4.2.Phase2)

In this phase, the aim is to detect all targeted objects in the positive images with an increased  $Se$  by giving several FPs if necessary, in order not to miss any targeted objects. Every image in the positive directory is processed by the Viola–Jones technique using each designated detector and objects are tagged wherever they are detected and coordinates for one or more rectangular ROI (coloured bounding box around the recognised object (e.g., Fig. 13)) are returned. These coordinates are mainly utilised for both counting each detected object once using the non-maximum suppression technique (Fig. 3), as explained in the

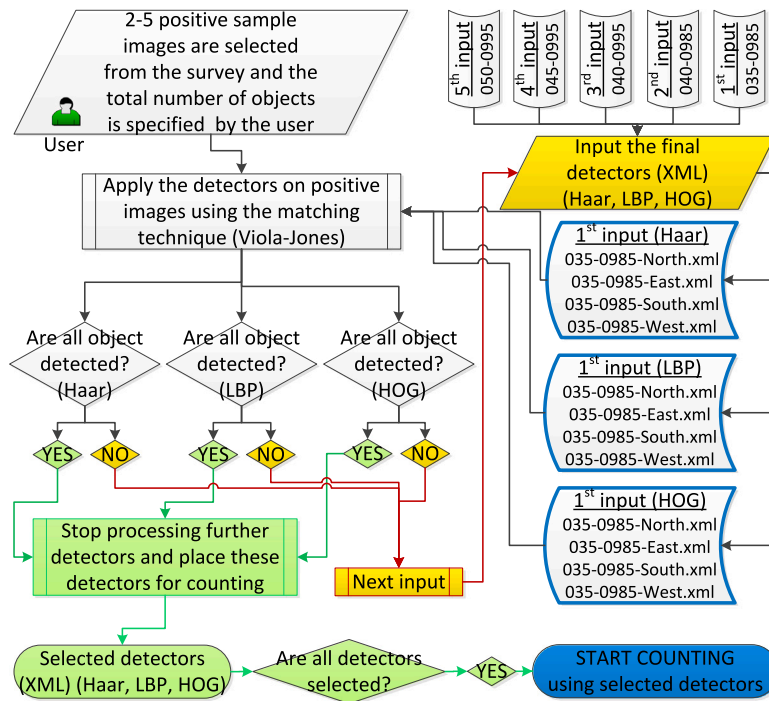


Fig. 12. Use of Reinforcement Learning (RL) for selecting best detectors for counting.



Fig. 13. Two magnified gannet objects detected by several detectors (left) and counted only once (right).

following paragraph, and cropping the tagged objects automatically for further analysis.

Due to the fact that detection windows overlap each other, the same object can be counted more than once. The main reason for this is that 12 detectors are applied for detecting objects in any direction, which may detect and specify an object several times. For instance, a gannet object is detected by 3 detectors and consequently counted 3 times and likewise, another gannet object is detected by 5 detectors and counted 5 times in Fig. 13(left). The non-maximum suppression technique, in which windows with a local maximum classifier response suppress nearby windows (Forsyth & Ponce, 2012), is employed to

count the same object only once as shown in Fig. 13(right). Two gannets are located by the detectors several times and they are counted as 2 objects in a whole image in Fig. 14 using the non-maximum suppression technique.

## 6. Results for splitting and counting

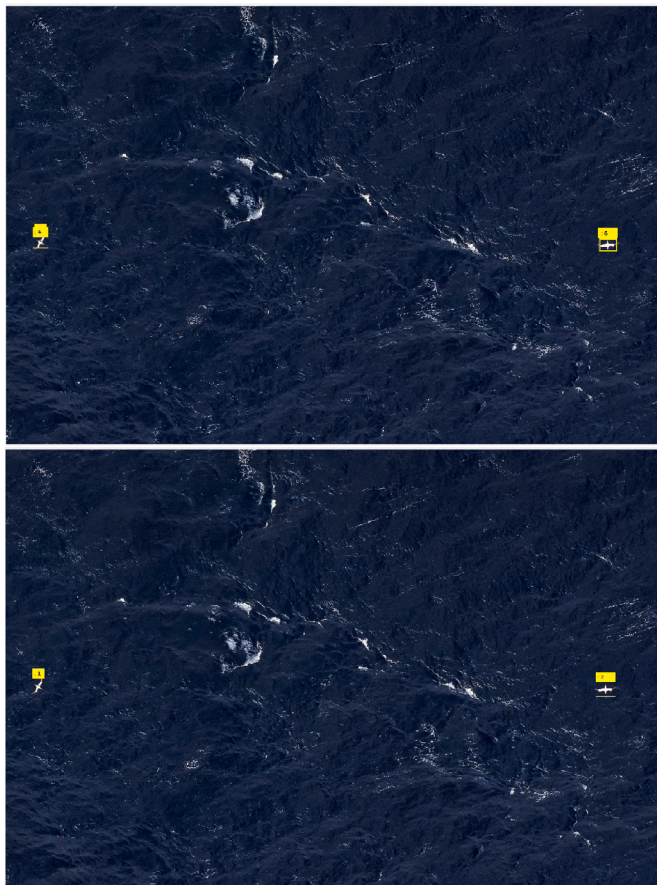
### 6.1. Results for phase 1: splitting

The methodology was evaluated on each of the 13 surveys (Fig. 4III) in which gannet objects exist to observe the success rates of splitting.

**Table 3**

Accuracy rates of the snag dataset based on the trained files of 4 different parameters for splitting images into positive and negative categories: all snags are recognised successfully by the training parameters, FAR = 0.50 and TPR = 0.995 with the combination of 3 techniques.

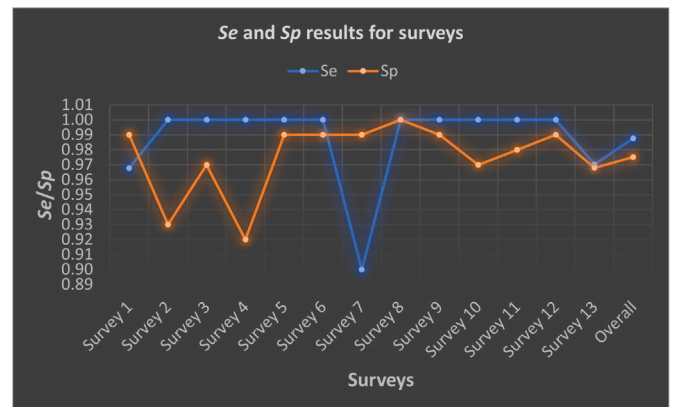
Surveys	NegativeImages	PositiveImages	GannetsTot #	TP	TN	Se	Sp	Selected parameters for three techniques					
								Haar		LBP		HOG	
								FAR	TPR	FAR	TPR	FAR	TPR
Survey 1	100	31	31	30	99	0.968	0.990	0.350	0.985	0.400	0.995	0.350	0.985
Survey 2	100	7	7	7	93	1.000	0.930	0.350	0.985	0.350	0.985	0.350	0.985
Survey 3	100	3	3	3	97	1.000	0.970	0.350	0.985	0.400	0.995	0.400	0.985
Survey 4	100	2	2	2	92	1.000	0.920	0.350	0.985	0.350	0.985	0.350	0.985
Survey 5	100	1	1	1	99	1.000	0.990	0.500	0.995	0.500	0.995	0.500	0.995
Survey 6	100	2	2	2	99	1.000	0.990	0.350	0.985	0.400	0.995	0.400	0.985
Survey 7	100	10	10	9	99	0.900	0.990	0.350	0.985	0.400	0.995	0.400	0.985
Survey 8	100	3	3	3	100	1.000	1.000	0.400	0.995	0.400	0.995	0.500	0.985
Survey 9	100	1	1	1	99	1.000	0.990	0.500	0.995	0.500	0.995	0.500	0.995
Survey 10	100	1	1	1	97	1.000	0.970	0.400	0.995	0.400	0.995	0.350	0.985
Survey 11	100	10	10	10	98	1.000	0.980	0.350	0.985	0.400	0.995	0.450	0.995
Survey 12	100	3	3	3	99	1.000	0.990	0.400	0.995	0.500	0.995	0.400	0.995
Survey 13	500	202	256	196	484	0.970	0.968	0.350	0.985	0.350	0.985	0.400	0.995
<b>Split Avg</b>						<b>0.988</b>	<b>0.975</b>						



**Fig. 14.** Gannets detected by the convenient detectors: Multiple gannet objects detected by several detectors (top) and counted only once (bottom).

The number of gannets and the negative images along with the success rates are presented in Table 3 and Fig. 15 with respect to the surveys. The detectors selected by the system for each feature extraction technique are shown in the column titled “selected parameters for three techniques” of Table 3 regarding each survey. For instance, the values FAR = 0.350 and TPR = 0.985 for survey 1 correspond to the four detectors, namely, 0.35–0.985-North.xml, 0.35–0.985-East.xml, 0.35–0.985-South.xml, and 0.35–0.985-West.xml determined for the Haar technique.

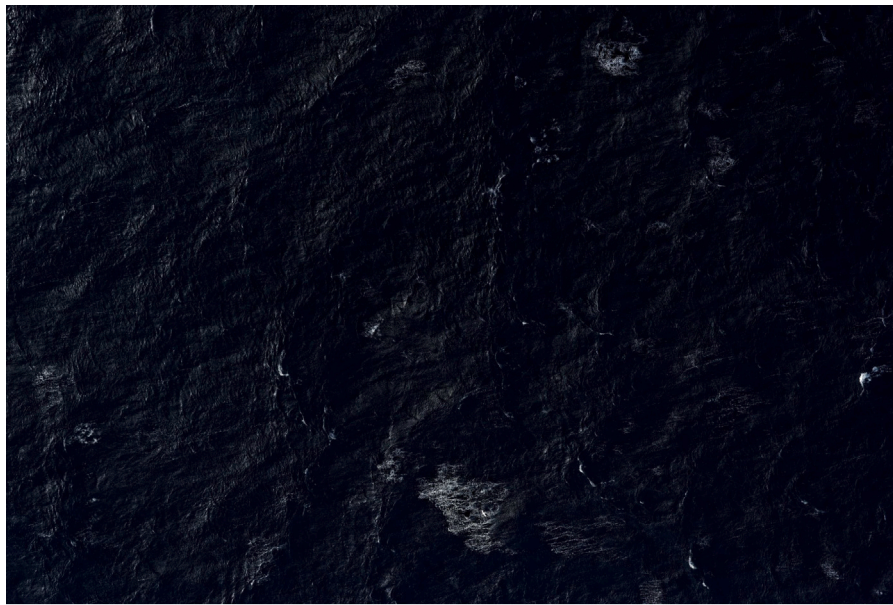
The large images with gannets that were not detected as positive are presented in our technical report — MarineObjects\_Gannet\_



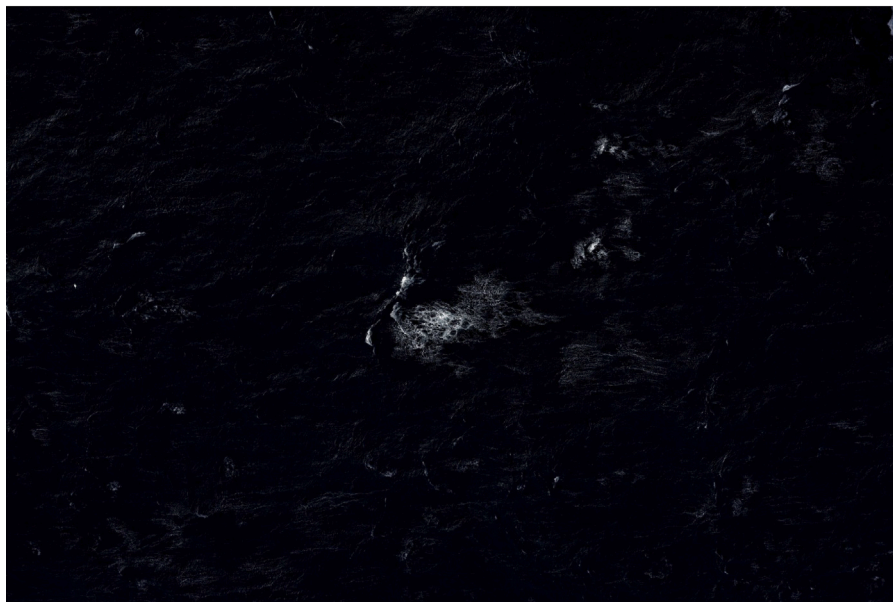
**Fig. 15.** Visualisation of Se and Sp results of the surveys in Table 3 for splitting images into positive and negative categories.

Supplement\_4.pdf in the supplementary materials. Additionally, the blank images with no gannets that were detected as positive are presented in our technical report — MarineObjects\_Gannet\_Supplement\_5.pdf as well. The average Se of the system concerning the Se results of 13 surveys based on the number of images (i.e., the column titled Se in Table 3) is 0.988. The average Sp of the system concerning the Sp results of 13 surveys based on the number of images (i.e., the column titled Sp in Table 3) is 0.975.

Se – correctly-detected-positive-images/all-positive-images – shows the power of the techniques used in the paper in giving assurance that if an image is tagged as a positive image, with at least one bird, that image most probably comprises at least one bird with a belief, an average confidence level of 0.988. In other words, we can conclude that there is a chance that this image does not comprise a bird with an average confidence level of 0.012, which is significantly low in a sense of showing high confidence when a decision is given about an image that is determined as “positive”. How the splitting process is implemented successfully can be noticed in Table 3 in the column “TP” compared to the column “Positive Images”. Almost all positive images with birds are placed in the positive folder for further processing (e.g., counting). This success is clearer in Survey 13 with many negative images and positive images with multiple targeted objects. On the other hand, Sp – correctly-detected-negative-images/all-negative-images – shows the power of the techniques in giving assurance that if an image is tagged as a negative image, that image most probably comprises no bird with a belief, an average confidence level of 0.975. In other words, we can conclude that there is a chance that this image is not a bird-free



(a) Example 1: the object with very difficult background texture; Bottom centre.



(b) Example 2: the object during diving in exactly different shape; at the middle left.

Fig. 16. Examples of gannet objects in whole images not detected by the trained classifiers.

image with an average confidence level of 0.025, which is significantly low in a sense of showing high confidence when a decision is given about an image that is determined as “negative”.

## 6.2. Results for phase 2: counting

The last survey – Survey 13 – in Table 3 was used to evaluate the viability of the object recognition and counting phase (Fig. 4III). The reason for selecting this survey is that it is the largest survey and has multiple gannets in some of the images, which can help quantify the obtained results more realistically with less bias. The detectors with the parameters of 0.40–0.995, 0.45–0.995, and 0.40–0.985 were selected respectively for Haar, LBP, and HOG techniques by the recursive RL technique. These parameters are bigger than the parameters selected by the RL algorithm in phase 1 (i.e., splitting with 0.35–0.985, 0.35–0.985, and 0.40–0.995) as explained in Section 5.2.1. This shows that

different detectors may be chosen for different purposes (i.e., splitting and counting) by the same recursive RL technique using two different approaches to realise the two different objectives, higher  $Sp$  with a high level of splitting and higher  $Se$  with a high level of object detection respectively. 248 objects out of 256 objects in 202 images were tagged as positive successfully resulting in a  $Se$  value of 0.968 which is 0.976 during the splitting phase regarding the number of objects. 6 objects are missed during the splitting phase within 6 different positive images (Table 3) and 2 objects within 2 different images are missed here during the counting phase. The two objects not detected by the application are shown in Fig. 16. The difference in  $Se$ , i.e., 0.08 (0.976–0.968), is not found to be significant ( $p > 0.01$  using the statistical paired t-test).  $Se$  – correctly-detected-positive-objects/all-positive-objects – shows the power of the techniques used in the paper in giving assurance that if an object is tagged as a positive gannet, that object most probably is a gannet with a belief, a confidence level of 0.968. In other words, we

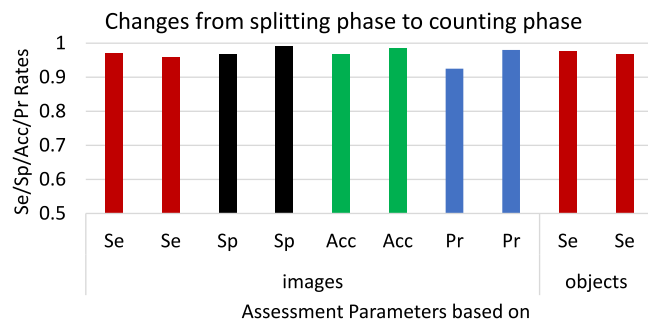


Fig. 17. Performance of the counting phase with respect to splitting regarding Survey 13 presented in Table 3.

can conclude that there is a chance that this object is not a gannet with a confidence level of 0.032, which is significantly low in a sense of showing high confidence when a decision is given about an object that is determined as “positive”.

On the other hand, it could be highly informative to compare the results between the splitting and counting phases based on the number of images rather than the number of objects for assessing how the counting phase is performing in further splitting, particularly, in handling the imbalanced data. 194 images out of 202 images were tagged as positive successfully resulting in a  $Se$  value of 0.960 which is 0.970 during the splitting phase. 6 positive images are missed during the splitting phase (Table 3) and 2 positive images are missed here during the counting phase. The difference in  $Se$ , is not found to be significant ( $p > 0.01$ , i.e., 0.1578). There were 4 FPs where waves were shaped similarly to the shape of gannets in the snags included in the training process. 496 out of 500 negative images are detected correctly as TN after the counting process whereas it is 484 during the splitting phase for survey 13 (Table 3). This results in a  $Sp$  value of 0.992 whereas it is 0.968 during the splitting phase based on the number of images. The difference, 0.024 (0.992–0.968), was found to be statistically significant ( $p < 0.01$ , i.e., 0.0005015) considering the number of negative images, i.e., 500, using the statistical paired t-test. The reduction of FP regarding the increased  $Sp$  is highly important, particularly for the surveys that are comprised of a great majority of bird-free negative images (e.g., >95%) leading to imbalanced data distribution and bias on the obtained results as elaborated above in Section 5. Moreover,  $Pr$  is increased slightly from the splitting  $Pr$ , 0.925 ( $TP / (TP + FP) = 196 / (196 + 16) = 0.925$ ), to the counting  $Pr$ , 0.980 ( $196 / (196 + 4)$ ) based on the number of images, which is statistically significant ( $p < 0.01$ ). Finally, overall  $Acc$  rises from 0.969 ( $(196 + 484) / (202 + 500)$ ) to 0.985 ( $(196 + 496) / (202 + 500)$ ) based on the number of images, which is statistically significant ( $p < 0.01$ ) as well. The results are presented in Fig. 17 for better visualisation. To summarise, the techniques used during the counting phase provide (i) a successful way of object detection leading to counting objects correctly, and (ii) further successful splitting leading to discarding the FP images substantially as well. The high value of  $Pr$  indicates that there is still a large room to perform CSA by which  $Se$  can be increased while compromising  $Pr$  slightly if  $Se$ , resulting from the minority positive class, is not deemed as satisfactory (<95% for our research) due to the imbalanced data class distribution that may cause unreliable results. These outcomes demonstrate that the two phases of using ensemble techniques proposed in this study can work successfully in performing the offshore bird censuses even without needing to perform CSA (Section 5) and most importantly, the proposed approaches can be generalised to the automated counting of broader species.

A comprehensive field test with a completely new survey has not yet been completed. The system was validated by two field experts from APEM using a completely new evaluation dataset with a decent number of example species (i.e., 20 positive images with 21 gannets

and 500 negative images (Fig. 4IV) taken from other recent surveys at the end of the project at the UCLan Intelligent Systems Laboratory before a comprehensive field survey is conducted using the established system in this study. There was a single juvenile gannet in this dataset and this was not detected as positive where all other gannets (i.e., 21) were detected correctly without missing a single one and without producing any FP by excluding other types of flying birds such as terns (i.e., 2 terns) and shearwaters (i.e., 11 shearwaters) as TN. The reason for not detecting this juvenile gannet is that the features of the juvenile gannets seem significantly different from their mature ones, e.g., first-year juvenile gannets are almost black, and subsequent sub-adult plumages show increasing amounts of white (SeabirdCentre, 2017). It is noteworthy to mention that there were no juvenile gannets either in our training nugget dataset or in our surveys. We suggest the construction of new classifiers specific to juvenile gannets to increase the chance of their detection. The correct labelling of the other images with other types of species (e.g., terns and shearwaters) as TN indicates that the classifiers established for gannets perform perfectly for detecting gannets as anticipated, and particular classifiers need to be established for the other species as mentioned in Sections 4.2.1–4.2.3 to identify them. This outcome confirms that the designed techniques in this research enable the automated classification of multispecies and counting them since every targeted species has its particular classifiers.

## 7. Discussion

Prevention of regional and global extinction of species during industrial developments and environmental changes (e.g., climate change, habitat loss with rapid urbanisation and coastal disturbance, toxic pesticide use) is a social responsibility from a conservationist point of view. In this sense, a species whose population is in decline needs to be identified urgently and should be protected with higher priority before it is too late. Data science is considered by Gibert et al. (2018) as the multidisciplinary field that combines data analysis with data processing methods and domain expertise, transforming data into understandable and actionable knowledge relevant to informed decision-making. Interdisciplinary efforts will help precipitate the shift towards increased use of computer-automated aerial photographic species census techniques (Chabot & Francis, 2016). Within this context, this study by bringing domain expertise and data scientists together in a fruitful collaborative team aims to develop a novel environmental platform for monitoring the marine ecosystem and performing bio censuses in an automated manner at regular intervals to track changes in a particular species population. Birds are sensitive indicators of biological richness, environmental health, ecosystem integrity, and environmental trends and fulfil many key ecological functions; they contribute to our understanding of natural processes (Bibby et al., 1998; Burger & Gochfeld, 2004; Morelli, 2015). Extinction of the passenger pigeon (*Ectopistes migratorius*), once likely the most numerous bird on the planet, provides a poignant reminder that even abundant species can go extinct rapidly (Rosenberg et al., 2019). Continuously, automated monitoring of species is of paramount importance which requires the use of advanced tools equipped with effective intelligent surveillance techniques. In this sense, a new non-parametric platform composed of an ensemble of supervised ML and RL techniques, WILDetect, is built to segment, split and count maritime species, in particular, birds for performing automated censuses in a highly dynamic maritime environment. Typically, parameter selection to mitigate the variations in datasets and obtain the best possible outcome in an intelligent autonomous system are carried out by users based on several predictions and trials and the success rates of the systems are highly associated with the wisdom of this assumption and implementation of trials correctly, which is a non-trivial task, specifically for ordinary users. Furthermore, there is no single best approach that suits every type of problem space based on the changing characteristics of datasets (e.g., quantity,



quality, attributes) and many other environmental dynamics (e.g., different seasons and time zones, different weather conditions, different settings and types of image-capturing technologies). It can be concluded based on the preliminary tests, as elaborated in Section 4.1, and the current research attempts in the literature to count species and classify multispecies, as elaborated in Sections Section 2, that (i) there is no computer-automated study that analyses datasets of small species acquired from the photogrammetry settings using small aeroplanes to survey very large areas in shorter times compared to the other on-ground, ships or UAS platforms, (ii) The most popular learning technique, the so-called DNN, yield the precision values ranging from 60% to 97.66% for bird detection using the aforementioned platforms, (iii) Large data samples with distinctive features (e.g., species that contrast distinctively from image backgrounds) may result in high accuracy rates in using DNN, (iv) The inner states of the DNN approaches are accepted as black boxes by the research community and these approaches do not let the researchers intervene in their inner states which may help increase their efficacy if they do not produce desired outcomes, (v) The misclassification of multispecies is high using DNN and clustering techniques if data instances in different groups resemble each other too closely as seen in bird species. In the proposed intelligent platform – within a dynamic approach that adjust its parameters according to the features of targeted objects, their background and the targeted accuracy rate – the best possible parameters, resulting in the best outcome, are chosen by the platform itself through the automated selection of pre-trained models, in which the parameters are instilled, using the user-model-data interaction solution that is implemented within a new recursive RL technique for mitigating the highly dynamic characteristics of the maritime ecosystem as well as the concerns mentioned with the aforementioned approaches. Additionally, the use of multiple trained models at a time, focusing on different features, ensures a high accuracy rate where one of the other two detectors/models can detect an object if it is missed by the other detector/model in use as elaborated in Section 4.2.2.

The validation of the platform, as summarised in Fig. 4, has been performed on several aerial maritime domains resulting in successful empirical evidence for the viability of the model. During the splitting phase, a positive image is most likely to be placed in a positive folder if there are several targeted objects in that image. Strictly speaking, there is a very high probability that one of the objects in an image will be detected by at least one of the three techniques using 12 detectors regarding the orientation of the objects during the splitting phase. Therefore, the more targeted objects in images, the higher the success rate of splitting. We would like to emphasise that the success rates are very high even though there is mostly only one gannet object in images in the surveys in this study (Table 3 and Fig. 15). The main reason for not detecting 2 of the gannet objects depicted in Fig. 16 in the second phase (i.e., recognition and counting) is that one of them does not look like the shape of a gannet in the training set, because, it is in the diving position, while the other one was not detected because of the very complex background texture behind the gannet. The training snag set should have more similar object types to be able to represent the real-world better and in this way, these types of objects are not missed by the trained detectors.

The trained files established for the gannets do not detect other types of birds as TP, such as common gulls, shearwaters, or terns. Therefore, if the objective is to count other types of birds as well, all bird types should be trained independently as explained in Section 4.2.2 to increase the accuracy of the system. In this way, the classification of other bird species becomes possible using the specific classifiers trained for these types of species. The methodology developed for the detection, splitting, and counting of birds, particularly gannets, in large-scale aerial images may be used for the UK marine gannet census since the most important nesting ground for northern gannets is in the UK with about half of the world's population (55.6%) (JNCC, 2015). Furthermore, multiple types of species of interest can be classified and counted

at once using the methodology (as concluded in Section 6.2) with the multiple classifiers that can be obtained as explained in Section 4. It is worth mentioning that the methodology can be expandable with more feature extraction techniques in addition to the feature extraction techniques (i.e., Haar, LBP, and HOG) that we employ in this study.

Given the current pace of global environmental change, quantifying change in species abundances is essential to assess ecosystem impacts. Evaluating the magnitude of declines requires effective long-term monitoring of population sizes and trends, data that are rarely available for most species (Rosenberg et al., 2019). Models perform better as they are attributed to the results of more realistic/recent-data analysis on particular domains. With the proposed platform, current labour-intensive and costly censuses of species conducted in longer time intervals can be replaced with cost-effective and more robustly automated computerised systems and can be repeated in an automated manner at regular intervals. Hence, cycles of the census can be conducted more frequently in shorter intervals over time, and incorporation of near-real-time results along with the prior results (e.g., population fluctuations) attributed to shorter intervals into these models paves the way for developing more effective ecological environmental models with realistic data trends and future projections. This, in turn, can boost the decision-making and prediction abilities of these data-driven simulation models, particularly, about the ecological footprint of human activities on the environment, specifically, on areas/offshores that are being turned into industrial zones, for both assessing the likely impact of the industrial developments on nature (e.g., habitat associations) and constraining/alleviating their potential damaging effects.

## 8. Conclusions and future work

Advanced tools, enabling effective monitoring of species, are needed to observe and predict the likely effects of environmental changes on species, mostly caused by indispensable industrial developments to take urgent proper actions, e.g., rebuilding natural habitats to maintain/increase species counts. Birds have been demonstrated to serve as good indicators of biodiversity and environmental change and as such can be used to make strategic conservation planning decisions for the wider environment (Bibby et al., 1998). Based on the literature reviewed in Chabot and Francis (2016), a major shift to computer-automated aerial photographic bird censusing is not yet underway and investigators are encouraged to study for potential approaches to automate animal detection and enumeration in aerial images. In this study, a novel supervised ML platform supported by a new recursive RL approach using several off-the-shelf feature extraction techniques and a matching algorithm were developed to conduct marine bird censuses in an automated manner. In the proposed approach, the uncertainties within a highly dynamic maritime environment and inconsistencies/variability in the characteristics of datasets attributed to the diverse sets of image-capturing technologies used in the maritime ecosystem have been mitigated using the recursive RL technique with the user-model-data interaction. In this technique, the most available parameters based on the characteristics of the dataset to be analysed are selected within the platform by the direction of the user at the start of the analysis to result in the best possible outcome. In this way, the developed approach adapts itself to the characteristics of the dataset concerning targeted objects and background and the environmental dynamics, which leads to resulting a desired solution to the current problem space in hand. The methodology has been evaluated and validated by field experts using several surveys and datasets that are independent of the dataset used in the training phase as outlined in Fig. 4. Experimental results on many aerial surveys demonstrate that the proposed methodology is effective and efficient in the detection and segmentation of targeted objects in the maritime ecosystem. The efficacy of the proposed approach can be increased as the techniques are trained with larger datasets for particular species.

The outcome of the study is expected to benefit the entire environmental modelling community. In particular, the proposed techniques can shed light on similar object detection implementations in finding the best possible parameters for analysis in an automated manner by employing the user-model-data interaction solution. Moreover, the platform can be employed to detect all types of birds after these species are pre-processed and trained, as mentioned in Sections 4.2.1–4.2.3. The outcomes elaborated in Section 6 demonstrate that the proposed approaches can be generalised to the automated counting of a broader number of species in a given area and these automated approaches can help track population changes of particular species at different specific locations on a regular basis with a true picture. Strictly speaking, it can be primarily deployed by environmentalists, researchers, authorities, and policymakers to monitor the marine ecosystem for fulfilling their goals effectively.

Within a holistic view, we aim to study other bird species and other marine species (e.g., turtles) as well as man-made maritime objects (Kuru et al., 2022) to be able to observe the bio marine ecosystem with the possible environmental footprint in the short, mid, and long-term. Moreover, the automatic classification of maritime ecosystems based on a variety of species will be in our future plans to support all types of environmental models with near-real-time information with multiple species.

## 9. Limitations of the study

The established environmental platform can work for other bird species, but using the specific detectors that can be trained for each species as explained in Sections 4.2.1 and 4.2.2. The higher the quality of the datasets representing the real environment, the higher the accuracy rates. We aim to share our results with other papers about our ongoing research on multispecies census of other species such as shearwaters, terns, gulls, scooters, fulmars.

## CRediT authorship contribution statement

**Kaya Kuru:** Conceptualization, Methodology, Software, Writing – original draft. **Stuart Clough:** Validation, Resources, Data curation. **Darren Ansell:** Investigation, Project administration. **John McCarthy:** Validation, Resources, Data curation. **Stephanie McGovern:** Validation, Resources, Data curation.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgements

We would like to thank APEM Ltd for providing the maritime aerial surveys used in this study. Anyone who wants to use any of the images either in this manuscript or in our technical reports in the supplementary materials should request permission from APEM Ltd, which retains the copyright and intellectual property contained within. The authors would like to thank the anonymous reviewers for their constructive comments which doubtlessly improved the quality of the paper significantly.

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.eswa.2023.120574>.

## References

- Akçay, H. G., Kabasakal, B., Aksu, D., Demir, N., Öz, M., & Erdoğan, A. (2020). Automated bird counting with deep learning for regional bird distribution mapping. *Animals*, 10(7), <http://dx.doi.org/10.3390/ani10071207>, URL: <https://www.mdpi.com/2076-2615/10/7/1207>.
- Alqaysi, H., Fedorov, I., Qureshi, F. Z., & O'Nils, M. (2021). A temporal boosted YOLO-based model for birds detection around wind farms. *Journal of Imaging*, 7(11), <http://dx.doi.org/10.3390/jimaging7110227>, URL: <https://www.mdpi.com/2313-433X/7/11/227>.
- Bibby, C., Jones, M., & Marsden, S. (1998). *Expedition field techniques: Bird surveys*. London: Royal Geographical Society.
- BOEM (2022). Ocean wind 1 offshore wind farm draft environmental impact statement. URL: <https://www.boem.gov/sites/default/files/documents/renewable-energy/state-activities/OceanWind1-DEIS.pdf>.
- Borowicz, A., Le, H., Humphries, G., Nehls, G., Höschle, C., Kosarev, V., & Lynch, H. J. (2019). Aerial-trained deep learning networks for surveying cetaceans from satellite imagery. *PLOS ONE*, 14(10), Article e0212532. <http://dx.doi.org/10.1371/journal.pone.0212532>.
- Burger, J., & Gochfeld, M. (2004). Marine birds as sentinels of environmental pollution. *EcoHealth*, 1(3), <http://dx.doi.org/10.1007/s10393-004-0096-4>.
- Chabot, D., Dillon, C., & Francis, C. M. (2018). An approach for using off-the-shelf object-based image analysis software to detect and count birds in large volumes of aerial imagery. *Avian Conservation and Ecology*, 13(1), <http://dx.doi.org/10.5751/ace-01205-130115>.
- Chabot, D., & Francis, C. M. (2016). Computer-automated bird detection and counts in high-resolution aerial images: a review. *Journal of Field Ornithology*, 87(4), 343–359. <http://dx.doi.org/10.1111/jof.12171>, URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jof.12171>.
- Clements, N., & Robinson, W. (2022). A re-survey of winter bird communities in the Oregon Coast Range, USA, initially surveyed in 1968–1970. *Biodiversity Data Journal*, <http://dx.doi.org/10.3897/arphapreprints.e91575>.
- Cruz, J. E. C., Shiguemori, E. H., & Guimaraes, L. N. F. (2015). A comparison of Haar-like, LBP and HOG approaches to concrete and asphalt runway detection in high resolution imagery. *The Journal of Computational Interdisciplinary Sciences*, 6(3), 121–136.
- Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Proceedings of the 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) - Volume 1 - Volume 01 CVPR '05*, (pp. 886–893). Washington, DC, USA: IEEE Computer Society, <http://dx.doi.org/10.1109/CVPR.2005.177>.
- Davis, K. L., Silverman, E. D., Sussman, A. L., Wilson, R. R., & Zipkin, E. F. (2022). Errors in aerial survey count data: Identifying pitfalls and solutions. *Ecology and Evolution*, 12(3), <http://dx.doi.org/10.1002/ece3.8733>.
- Delhez, A. (2022). *Multiple flying bird and bird keypoint detection toolbox for processing bird videos* (Master's thesis), Belgium: University of Louvain.
- Dujon, A. M., Ierodiaconou, D., Geeson, J. J., Arnould, J. P. Y., Allan, B. M., Katselidis, K. A., & Schofield, G. (2021). Machine learning to detect marine animals in UAV imagery: effect of morphology, spacing, behaviour and habitat. *Remote Sensing in Ecology and Conservation*, 7(3), 341–354. <http://dx.doi.org/10.1002/rse2.205>, URL: <https://zslpublications.onlinelibrary.wiley.com/doi/abs/10.1002/rse2.205>.
- Forsyth, D., & Ponce, J. (2012). *Computer vision* (2nd ed.). London: Pearson.
- Gibert, K., Horsburgh, J. S., Athanasiadis, I. N., & Holmes, G. (2018). Environmental data science. *Environmental Modelling & Software*, 106, 4–12. <http://dx.doi.org/10.1016/j.envsoft.2018.04.005>, URL: <http://www.sciencedirect.com/science/article/pii/S1364815218301269>. Special Issue on Environmental Data Science. Applications to Air quality and Water cycle.
- Groom, G., Stjernholm, M., Nielsen, R. D., Fleetwood, A., & Petersen, I. K. (2013). Remote sensing image data and automated analysis to describe marine bird distributions and abundances. *Ecological Informatics*, 14, 2–8. <http://dx.doi.org/10.1016/j.ecoinf.2012.12.001>, URL: <https://www.sciencedirect.com/science/article/pii/S1574954112001185>. The analysis and application of spatial ecological data to support the conservation of biodiversity.
- Hayes, M. C., Gray, P. C., Harris, G., Sedgwick, W. C., Crawford, V. D., Chazal, N., Crofts, S., & Johnston, D. W. (2021). Drones and deep learning produce accurate and efficient monitoring of large-scale seabird colonies. *Ornithological Applications*, 123(3), <http://dx.doi.org/10.1093/ornithapp/duab022>.
- Hollings, T., Burgman, M., van Andel, M., Gilbert, M., Robinson, T., & Robinson, A. (2018). How do you find the green sheep? A critical review of the use of remotely sensed imagery to detect and count animals. *Methods in Ecology and Evolution*, 9(4), 881–892. <http://dx.doi.org/10.1111/2041-210X.12973>, URL: <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12973>.
- Hong, S.-J., Han, Y., Kim, S.-Y., Lee, A.-Y., & Kim, G. (2019). Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery. *Sensors*, 19(7), <http://dx.doi.org/10.3390/s19071651>, URL: <https://www.mdpi.com/1424-8220/19/7/1651>.
- Jaderberg, M., Czarnecki, W. M., Dunning, I., Marris, L., Lever, G., Castañeda, A. G., Beattie, C., Rabinowitz, N. C., Morcos, A. S., Ruderman, A., Sonnerat, N., Green, T., Deason, L., Leibo, J. Z., Silver, D., Hassabis, D., Kavukcuoglu, K., & Graepel, T. (2019). Human-level performance in 3D multiplayer games with population-based

- reinforcement learning. *Science*, 364(6443), 859–865. <http://dx.doi.org/10.1126/science.aau6249>.
- Jayadevan, P., Nameer, P. P. O., Jha, A., Aravind, A., Dilip, K. G., Karuthedathu, D., Tom, G., Mavelikara, H., Mannar, H., Palot, M., Johnson, J., Raveendran, J., Rodrigues, M., Mujeeb, M., Namassivayan, L., Payyeri, N., Nesrudheen, P., Narayanan, S. P., Prasanth, S., & Yathumon, M. A. (2022). Kerala Bird Atlas 2015–20: features, outcomes and implications of a citizen-science project. *Current Science*, 122, 298–309. <http://dx.doi.org/10.18520/cs/v122/i3/298-309>.
- JNCC (2015). Northern Gannet *Morus bassanus*. URL: <http://jncc.defra.gov.uk/page-2875>.
- JNCC (2022). Seabirds count – the fourth breeding seabird census. URL: <https://jncc.gov.uk/our-work/seabirds-count/>.
- Johnston, D. W. (2019). Unoccupied aircraft systems in marine science and conservation. *Annual Review of Marine Science*, 11(1), 439–463. <http://dx.doi.org/10.1146/annurev-marine-010318-095323>, PMID: 30020850.
- Kellenberger, B., Veen, T., Folmer, E., & Tuia, D. (2021). 21 000 Birds in 4.5 h: efficient large-scale seabird detection with machine learning. *Remote Sensing in Ecology and Conservation*, 7(3), 445–460. <http://dx.doi.org/10.1002/rse2.200>, URL: <https://zslpublications.onlinelibrary.wiley.com/doi/abs/10.1002/rse2.200>.
- Kuru, K., Clough, S., Ansell, D., McCarthy, J., & McGovern, S. (2022). Intelligent airborne monitoring of irregularly shaped man-made objects in the marine ecosystem using statistical machine learning techniques. *SSRN Electronic Journal*, <http://dx.doi.org/10.2139/ssrn.4283676>.
- Kuru, K., Girgin, S., Arda, K., & Bozlar, U. (2013). A novel report generation approach for medical applications: The SISDS methodology and its applications. *International Journal of Medical Informatics*, 82(5), 435–447. <http://dx.doi.org/10.1016/j.ijmedinf.2012.05.019>, URL: <https://www.sciencedirect.com/science/article/pii/S138650561200113X>.
- Kuru, K., & Khan, W. (2018). Novel hybrid object-based non-parametric clustering approach for grouping similar objects in specific visual domains. *Applied Soft Computing*, 62, 667–701. <http://dx.doi.org/10.1016/j.asoc.2017.11.007>.
- Kuru, K., & Yetgin, H. (2019). Transformation to advanced mechatronics systems within new industrial revolution: A novel framework in automation of everything (AoE). *IEEE Access*, 7, 41395–41415. <http://dx.doi.org/10.1109/ACCESS.2019.2907809>.
- Li, Y., Yang, J., & Wen, J. (2021). Entropy-based redundancy analysis and information screening. *Digital Communications and Networks*, <http://dx.doi.org/10.1016/j.dcan.2021.12.001>.
- Li, Y., Yang, J., Zhang, Z., Wen, J., & Kumar, P. (2023). Healthcare data quality assessment for cybersecurity intelligence. *IEEE Transactions on Industrial Informatics*, 19(1), 841–848. <http://dx.doi.org/10.1109/TII.2022.3190405>.
- McConnell, R. (1986). Method of and apparatus for pattern recognition. URL: <http://www.google.co.uk/patents/US4567610>. US Patent 4, 567, 610.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fiedelnd, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533. <http://dx.doi.org/10.1038/nature14236>.
- Moravčík, M., Schmid, M., Burch, N., Lisý, V., Morrill, D., Bard, N., Davis, T., Waugh, K., Johanson, M., & Bowling, M. (2017). DeepStack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*, 356(6337), 508–513. <http://dx.doi.org/10.1126/science.aam6960>.
- Morelli, F. (2015). Indicator species for avian biodiversity hotspots: Combination of specialists and generalists is necessary in less natural environments. *Journal for Nature Conservation*, 27, 54–62. <http://dx.doi.org/10.1016/j.jnc.2015.06.006>, URL: <https://www.sciencedirect.com/science/article/pii/S1617138115300078>.
- Murray, S., Harris, M. P., & Wanless, S. (2014). An aerial survey of Northern Gannets *Morus bassanus* on Scar Rocks, southwest Scotland, in 2014. *Seabird*, 27(1), 104–109.
- Murray, S., Harris, M. P., & Wanless, S. (2015). The status of the gannet in Scotland in 2013–14. *Scottish Birds*, 35(1), 3–18.
- Murray, S., Smith, I., & Smith, A. (2014). Gannet and Guillemot breeding on Rockall, North Atlantic. *Scottish Birds*, 34(1), 13–15.
- Ojala, T., Pietikainen, M., & Harwood, D. (1994). Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. In *Proceedings of 12th international conference on pattern recognition*, Vol. 1 (pp. 582–585). <http://dx.doi.org/10.1109/ICPR.1994.576366>.
- Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 971–987. <http://dx.doi.org/10.1109/TPAMI.2002.1017623>.
- Paleczny, M., Hammill, E., Karpouzi, V., & Pauly, D. (2015). Population trend of the world's monitored seabirds, 1950–2010. *PLOS ONE*, 10(6), Article e0129342. <http://dx.doi.org/10.1371/journal.pone.0129342>.
- Papageorgiou, C. P., Oren, M., & Poggio, T. (1998). A general framework for object detection. In *Sixth international conference on computer vision (IEEE Cat. No.98CH36271)* (pp. 555–562). <http://dx.doi.org/10.1109/ICCV.1998.710772>.
- Reyes, E. (2014). *A comparison of image processing techniques for bird detection* (Master's thesis), the Faculty of California Polytechnic State University.
- Rosenberg, K. V., Dokter, A. M., Blancher, P. J., Sauer, J. R., Smith, A. C., Smith, P. A., Stanton, J. C., Panjabi, A., Helft, L., Parr, M., & Marra, P. P. (2019). Decline of the North American avifauna. *Science*, 366(6461), 120–124. <http://dx.doi.org/10.1126/science.aaw1313>, URL: <https://www.science.org/doi/abs/10.1126/science.aaw1313>. arXiv:https://www.science.org/doi/pdf/10.1126/science.aaw1313.
- RSBP (2015). Gannet. URL: <https://www.rspb.org.uk/birds-and-wildlife/wildlife-guides/bird-a-z/gannet/>.
- SeabirdCentre (2017). Gannet. URL: <https://seabird.org/wildlife/seabirds/gannet/12/26/47>.
- Shi, Y., Shen, C., Fang, H., & Li, H. (2017). Advanced control in marine mechatronic systems: A survey. *IEEE/ASME Transactions on Mechatronics*, 22(3), 1121–1131. <http://dx.doi.org/10.1109/TMECH.2017.2660528>.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489. <http://dx.doi.org/10.1038/nature16961>.
- Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T. P., Simonyan, K., & Hassabis, D. (2017). Mastering chess and shogi by self-play with a general reinforcement learning algorithm. ArXiv, [abs/1712.01815](https://arxiv.org/abs/1712.01815).
- Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, 362(6419), 1140–1144. <http://dx.doi.org/10.1126/science.aar6404>, arXiv:https://www.science.org/doi/pdf/10.1126/science.aar6404.
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., & Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, 550(7676), 354–359. <http://dx.doi.org/10.1038/nature24270>.
- Smyser, T. J., Guenzel, R. J., Jacques, C. N., & Garton, E. O. (2016). Double-observer evaluation of pronghorn aerial line-transect surveys. *Wildlife Research*, 43(6), 474. <http://dx.doi.org/10.1071/wr16006>.
- Stefano, L. D., Maccoccia, S., & Mola, M. (2003). An efficient algorithm for exhaustive template matching based on normalized cross correlation. In *12th international conference on image analysis and processing, 2003. Proceedings* (pp. 322–327). Mantova, Italy: IEEE.
- Thompson, D. (2021). Seabird survey methods for offshore installations: Black-legged kittiwakes. URL: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/974338/Kittiwake\\_survey\\_advice\\_v2.1.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/974338/Kittiwake_survey_advice_v2.1.pdf).
- Trefný, J., & Matas, J. (2010). Extended set of local binary patterns for robust object detection. In *Computer vision winter workshop 2010. Czech Society for Cybernetics and Informatics*.
- Vállez, N., Deniz, O., & Bueno, G. (2015). Sample selection for training cascade detectors. *PLOS ONE*, 10(7), 1–16. <http://dx.doi.org/10.1371/journal.pone.0133059>.
- Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, Vol. 1* (pp. 1–511–518). <http://dx.doi.org/10.1109/CVPR.2001.990517>.
- Wang, L., & He, D.-C. (1990). Texture classification using texture spectrum. *Pattern Recognition*, 23(8), 905–910.
- Wang, D., Shao, Q., & Yue, H. (2019). Surveying wild animals from satellites, manned aircraft and unmanned aerial systems (UASs): A review. *Remote Sensing*, 11(11), <http://dx.doi.org/10.3390/rs11111308>, URL: <https://www.mdpi.com/2072-4292/11/11/1308>.
- Witkin, A. (1984). Scale-space filtering: A new approach to multi-scale description. In *ICASSP '84. IEEE international conference on acoustics, speech, and signal processing*, Vol. 9 (pp. 150–153). <http://dx.doi.org/10.1109/ICASSP.1984.1172729>.
- Yang, Z. (2010). Fast template matching based on normalized cross correlation with centroid bounding. In *2010 international conference on measuring technology and mechatronics automation*, Vol. 3 ICMTMA, (pp. 224–227). Changsha, China: IEEE.
- Zhang, L., Chu, R., Xiang, S., & Liao, S. (2007). *Lecture notes in computer science, Facts and fallacies: A book of definitive mistakes and misguided predictions in advances in biometrics*. Berlin Heidelberg: Springer.



**Kaya Kuru** received the B.Sc. degree from National Defense University, the major/ADP degree in computer engineering from Middle East Technical University (METU), the M.B.A. degree from Selcuk University, and the M.Sc. and Ph.D. degrees in computer science from METU. He completed his postdoctoral studies with the School of Electronics and Computer Science, University of Southampton, U.K. He worked as a DBA, SW Developer, SW Development Manager, and IT Manager with the IT Department in a university, for a period of 15 years. He is currently an Associate Professor in computer-information systems engineering. He has recently

engaged in the implementation of numerous AI-based real-world systems within various funded projects. His research interests include the development of autonomous intelligent systems, using FL, DL, and DRL with CPSS.



**Stuart Clough** has a B.Sc. in Environmental Biology from the University of Sunderland and a Ph.D. in Animal Behaviour from the University of St Andrews. (Thesis: migration and habitat use of dace in an English chalk stream). He is an aquatic ecologist with over 20 years of experience in the environmental consulting industry. Stuart joined APEM as a director in 2006 and took on responsibility for the company's remote sensing division in 2009. Under Stuart's leadership, the division has become a leading provider of ultra-high-resolution aerial surveys to the offshore wind industry globally, delivering over 2,000 aerial surveys for many of the world's leading wind farm developers. Stuart began his career in 1992 with the Institute of Freshwater Ecology in the UK, publishing a number of scientific papers and becoming a specialist in fish migration. In addition to his domestic remote sensing and ecology work, Stuart also has responsibility for APEM's overseas work in both Germany, Vietnam, Australia and the USA. Stuart is a Fellow of the Institute of Fisheries Management and a member of the Institute of Water.



**Daren Ansell** received the B.Sc. from the University of Manchester Institute of Science and Technology (UMIST) in electrical and electronic engineer and Ph.D. degree from Cranfield University in Antenna Optimisation Using Evolutionary Algorithms. He is the engineering lead for Space and Aerospace and professor in the School of Engineering, University of Central Lancashire. He specialises in applied autonomous and intelligent systems research. He previously worked in industry at BAE Systems in research management and research and development roles, specialising in Mission Systems and Autonomy. Darren is research active



within the area of digital engineering and is a member of the Applied Digital Signal and Image Processing Research Centre (ADSIP). He is leading collaborative research projects with industry partners, developing intelligent software for the aerospace, medical and nuclear sectors.

**John McCarthy** has a B.Sc. in Environmental Plant Biotechnology from University College Cork and an M.Sc. in Environmental Assessment and Management from the University of Salford. He is the head of APEM's data operations and works closely with the aviation and remote sensing teams, helping to plan aerial surveys and ensure that our cutting-edge hardware is working optimally and that the thousands of ultra-high-resolution images and other data gathered is downloaded, processed and stored correctly. As well as working on surveys all around the UK, John has been the technical lead on surveys in Ireland, New York, Hawaii, Texas, Florida, Vietnam and New England.



**Stephanie McGovern** Stephanie McGovern has a B.Sc. in Animal Behaviour from the University of Liverpool, and M.Sc. in Ecology, Evolution and Conservation from Imperial College London and a Ph.D. in Assessment of Environmental Change from Bangor University. She has 10 years of experience working on environmental projects in APEM's ornithology team. Stephanie has undertaken density surface modelling to predict bird distribution, data analysis work and collision risk assessments for a number of aerial survey and offshore wind farm EIA projects across the UK, Ireland, Germany and the US. Stephanie has gained experience on a number of offshore wind farm projects, has developed an in-depth knowledge of survey methodologies used for ornithology and marine mammal surveys and, in 2011, created APEM's bespoke bird migration model MIGROPATH, which has subsequently been used in multiple Round 3 EIAs.