Individual recognition and number prediction for Manx shearwater (*Puffinus puffinus*) by passive acoustic monitoring

Zhanyi Lin[1*], Yinka Adeniji[1*], William D Pearse[1], Stephanie Harris[2], Sue Anne Zollinger[3], Julia Schroeder[1]

1. Department of Life Sciences, Imperial College London, Silwood Park Campus, Berkshire, SL5 7PY. y.adeniji25@imperial.ac.uk https://orcid.org/0000-0001-5809-4888

2. University of Liverpool, 765 Brownlow Hl, Liverpool L69 7ZX

3. Department of Natural Sciences, Manchester Metropolitan University, School of Biological and Chemical Sciences, Manchester, M1 5GD. s.zollinger@mmu.ac.uk https://orcid.org/0000-0001-8819-2606

* Shared first authorship

Abstract

Seabirds can be notoriously difficult to monitor using conventional methodology, as many breed on remote islands, are active nocturnally and many breed in underground burrows. Current census methods such as counting burrows and using playbacks are time-consuming and labour-intensive. One novel solution is to use passive acoustic monitoring (PAM) followed by individual recognition and number estimation. Previous studies have successfully achieved individual bird recognition with high accuracy; however, precisely estimating population abundance via their acoustics remains an open problem. To bridge this gap, we tested whether individual Manx shearwaters (*Puffinus puffinus*) can be identified from acoustic recordings, and whether the numbers of individuals present can be estimated. We collected recordings of single-individual and multiple-individual calls from three colonies in the UK. We developed a pipeline to combine BirdNET, a one-dimensional convolutional neural network, and agglomerative clustering for individual identification and open-set individual counting. Closed-set recognition achieved high accuracy (89.3%) for Manx shearwater vocalisations at the individual level, and recognition accuracy showed no significant difference between sexes. Open-set individual number estimation achieved a mean relative error of 39.1%, and model performance was robust to noise and

recording conditions. These findings indicate that vocalisations of both sexes contain individual signatures and demonstrate the feasibility of counting unseen individuals from passive acoustic recordings. Importantly, this study supports population census using PAM in real field scenarios, where unseen individuals are abundant. We provide an easy, non-invasive approach for seabird population monitoring, which is transferable for monitoring other species, and has a high potential for wildlife conservation and ecological research.

Introduction

Seabirds are a diverse group of birds from multiple evolutionary lineages that have evolved adaptations to the marine environment. While these adaptations improve their survival in the marine niche, they may impair their ability to respond to climate and environmental change (Avaria-Llautureo et al., 2026; Haney et al., 2014; Meissner, 2025). As a result, 31% of seabird species are globally threatened, and nearly half of them show trends of decline (Dias et al., 2019). Therefore, monitoring seabird populations and their trends has become increasingly important for understanding the impacts of climate change and environmental pollution. To estimate the population size of seabirds, commonly used census techniques include directly counting nests or occupied sites for diurnal seabirds, counting occupied burrows for burrow-nesting seabirds, and capture–mark–recapture methods for nocturnal seabirds (Sanz-Aguilar et al., 2010; Walsh et al., 1995). Most of the approaches above are labour-intensive and time-consuming (Walsh et al., 1995), because many seabirds breed in dense colonies on cliffs, which is uneasy for humans to access. Furthermore, hidden burrows can easily be overlooked, and burrow occupancy needs to be checked manually, posing extra difficulty for accurate estimations. Techniques that involve capturing birds (e.g., ringing) can directly injure the bird (Griesser et al., 2012); other approaches can also cause disturbance and affect hatching success (Blackmer et al., 2004; Walsh et al., 1995), which is invasive for the population.

One potential solution that has received increasing attention in recent years is using non-invasive passive acoustic monitoring (PAM) to estimate density, population size, and other parameters (Bailey et al., 2021; Hensel et al., 2022). Nest density and local population size were found to correlate with acoustic activity in shearwaters (Oppel et al., 2014). However, Arneill et al. (2020) found that acoustic activity has no predictive effect on nest density, highlighting uncertainty in the relationship between calling activity and population size. Bird vocal activities also vary with weather and environmental conditions (Lengagne et al., 1999; Lengagne and Slater, 2002), limiting the reliability of population estimates based solely on acoustic indices. Therefore, new informative alternatives, such as individual number estimation by individual recognition, are highly needed. This is feasible because many nocturnal seabird species rely on vocalisations to

66 locate nests and their mates in darkness (Brooke, 1990; Cure et al., 2009), suggesting underlying
67 individual signals encoded in their calls.

68 Using machine learning to identify bird individuals in a closed set (i.e., all test set classes are
69 presented in the training set) has been successful in bioacoustics. Convolutional neural network
70 (CNN) is one of the most widely used and accurate architectures for individual recognition.
71 Closed-set individual identification accuracy has reached 78.6% for Chiffchaff (*Phylloscopus*
72 *collybita*) (Ptacek et al., 2016) and >90% for 5 bird species in different taxonomic groups (Huang
73 et al., 2025). However, accuracy drops when changing the background in the soundscape or using
74 another corpus for testing, suggesting background features may confound individual recognition
75 (Stowell et al., 2019; Xie et al., 2023). Open-set recognition (i.e., identifying novel targets not in
76 the training set) has reached an accuracy higher than 80% in different bird species, showing good
77 discriminative ability between seen and unseen individuals (Huang et al., 2025; Ntalampiras and
78 Potamitis, 2021). However, the study of open-set classification (i.e., identifying and classifying
79 new targets) remains scarce and has not been extended to bioacoustics. Bedoya and Molles (2021)
80 introduced a pipeline using CNN and fuzzy clustering to perform open-set classification and
81 achieved an accuracy of 88%. However, their examination was based on a small dataset with only
82 three novel classes. In the field of object detection, Zheng et al. (2022) introduced an open-set
83 object classification pipeline, using a similar CNN + clustering architecture. The study produces an
84 underestimation error of 20- 40% for open-set object counting. Similar algorithms are also
85 designed in facial recognition (Liu et al., 2018). These pipeline designs are important references
86 for this study, as all of them aim to separate and label unseen instances.

87 Despite recent research advances, no existing pipelines in bioacoustics can estimate the number of
88 calling individuals in real field scenarios, where unseen individuals are abundant. To bridge the
89 current gap, we used the Manx shearwater (*Puffinus puffinus*) as a model species to investigate
90 whether machine learning can be used to estimate the number of individuals present in audio
91 recordings. Manx shearwater is a nocturnal, burrow-nesting seabird whose breeding populations
92 are particularly difficult to estimate accurately using conventional census methods (Lee et al.,
93 2023). Most Manx shearwaters nest in grass slopes and cliff edges, which increases the difficulty

when accessing breeding sites. During breeding, a pair of breeding individuals share a burrow, so local population size can be estimated by counting the number of actively occupied burrows, which is checked by playback. However, response rates are typically below 80% and may vary by sex (Perkins et al., 2017). Furthermore, foraging adults are absent during the day, and they generally do not approach the colony until nightfall (Brooke, 1990). Consequently, the number of birds at an occupied burrow varies from time to time, making accurate assessments of burrow occupancy and breeding population size challenging. Manx shearwater is a good model for machine learning in bioacoustics because individual signatures have been identified in their calls, although principal components (PC), which interpret major directions of variation, show levels of overlap (Sun et al., 2024). This study has two objectives: (1) build a pipeline that can identify individual Manx shearwaters through their vocalisations, and (2) estimate individual numbers in Manx shearwater acoustic recordings, including those with unknown IDs.

Material and methods

Study sites

We selected Lundy (51°10′N, 4°40′W), St. Agnes (49°53′N, 6°20′W), and Copeland (54°40′N, 5°32′W) as study sites, as they hold considerable breeding pairs of Manx shearwater (12,500, 4,900, and 225 pairs, respectively) and are geographically isolated, increasing data abundance and heterogeneity. In Lundy, 70 burrows in 5 colonies (Old Light, Virgin's Spring, Castle Hill & Benjamin's Chair, Logan Stone, and Dead Cow Point) were selected for data collection. In St. Agnes, 28 burrows in 2 colonies (Southwest of Gugh and North of Gugh) were selected for data collection. Each colony was preliminarily investigated to ensure sufficient burrows were present. Data collection in Copeland was conducted in 2025 by other members of the group, and per-burrow recordings are ready for use.

Data collection and pre-processing

When visiting the study site, we classified examined burrows as actively occupied if faeces, feathers, or individuals were present inside the burrow or at the burrow opening. These burrows

were selected for study. We deployed audio recording devices (Song Meter® Micro 2 and AudioMoth®) 10-30 cm from the burrow opening, with microphones facing inwards, and left them in situ for at least 48 hours before retrieval. To ensure monitoring overlaps with the most active calling period, devices were configured to work from 10 P.M. to 5 A.M. (British summer time), with 20 minutes on duty and 5 minutes off in each cycle. Gains were set as "medium" for audiomoths and "6-12 dB" for song meters. Each cycle produced a 16-bit depth monophonic waveform audio file (WAV) with a 22.05 kHz sampling rate, which covers the frequency range of Manx shearwater calls.

After data retrieval, audio files were pre-processed to generate the training set and the test set. We reviewed each audio file manually to select clear Manx shearwater call segments (> 5s each). Distinctively louder calls were from either one or two (a male and a female) individuals associated with the burrow, and the corresponding individual IDs were assigned. We used spectrogram patterns to judge the identity and assessed sex based on Lee et al. (2023). All calls with IDs were then imported into the BirdNet V2.4 acoustic model (Kahl et al., 2021) in Python (v3.9.23) for target detection using the default TensorFlow backend. Based on detection, Manx shearwater calls in auto-selected clips with confidence > 0.60 were saved as training samples. This threshold reduces bad samples (background calls and incomplete calls) while maximising detections. To generate the test set, we randomly chose 56 recording files each has a length of 20 minutes and contains multiple calling individuals from the active outside-burrow calling period (2-4 A.M.) from the same per-burrow recordings. For each file, the individual number and male/female count were estimated manually. Call structures and motif patterns (e.g., length, pitch range, etc.) in the spectrogram were used for identifying and distinguishing individuals. Faint calls that showed unclear call structures and motif patterns were removed from estimation. Background noise level was calculated using sound levels exceeding 90% of the measured time (L90) in Python. Then, we imported files into BirdNet (V2.4 acoustic model) to generate short clips containing target calls. Mean peak amplitude was calculated to characterise the calling loudness. Overlapping call proportion was estimated by sampling one-fifth of the clips and was calculated as the number of clips containing overlapping calls divided by the total number of clips. We saved the above indices as a comma-separated values (CSV) file for later analysis of model performance and robustness.

Pipeline design

Block design. To estimate individual numbers of Manx shearwaters in audio recordings, the pipeline was divided into three main blocks (Fig. 1). In the *feature extraction block*, BirdNet was used to detect target call clips and then generate a 1024-dimensional feature vector from each clip. Feature vectors were then imported into the *core model block* (a CNN classifier), which further learns the differences between individuals and creates new 256-d embeddings. The 256-d embeddings were then passed to the *cluster-and-count block* for dimensionality reduction and distance-based clustering, followed by final predictions of individual counts. The pipeline and the model were built in Python (v3.9.23) environment.

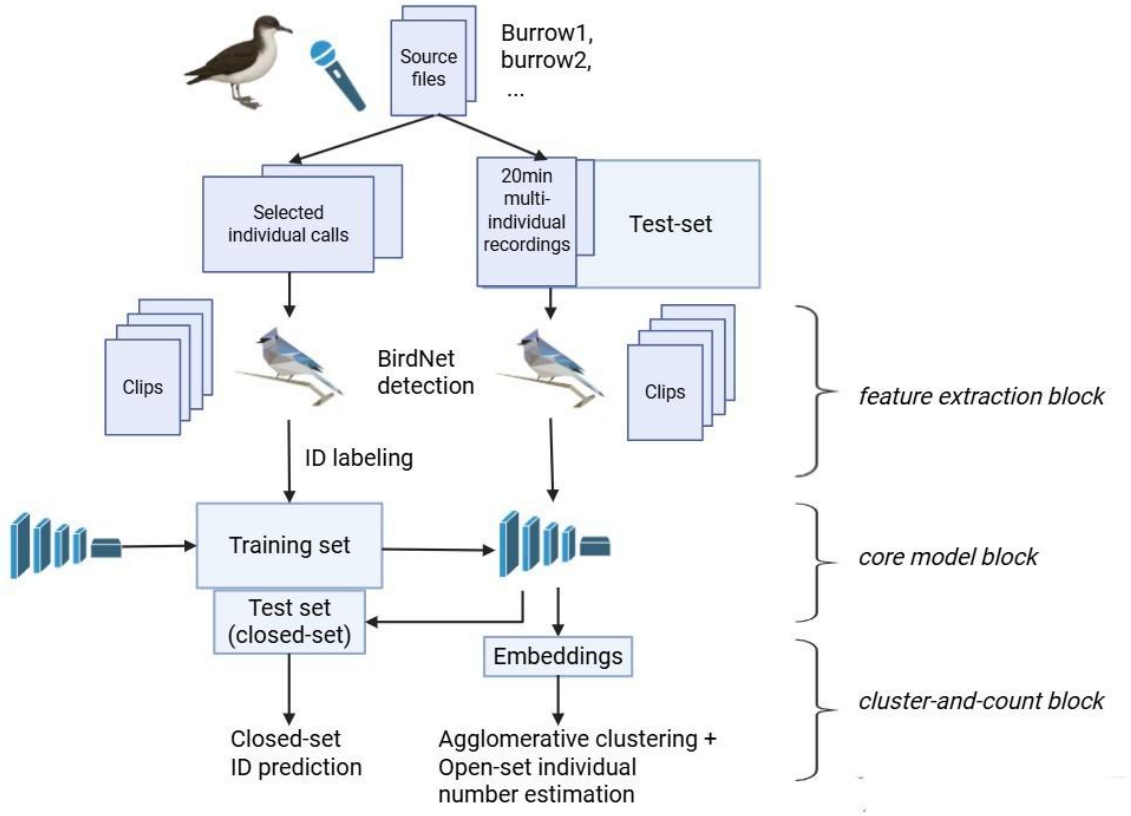

Figure 1. The pipeline structure. The pipeline is divided into three blocks. Source files from PAM were separated into calls with individual labels for later training and closed-set testing. 20-minute-long recordings were selected from the same files (but at different times) as the test set for open-set predictions. BirdNet was used for feature extraction, and the output embeddings were processed by the core convolutional neural network (CNN) model for either model training, closed-set testing, or 256-dimensional embedding generation. The 256-d embeddings were then used for agglomerative clustering and

open-set individual number estimation.

Feature extraction block. To extract features of Manx shearwater vocalisations, we imported clips generated by BirdNet detections into BirdNet-Analyzer (v2.4.0) (Kahl et al., 2021), and extracted their feature vectors using the embedding extraction function. All settings were left at default (overlap = 0, audio speed modification = 0, batch size = 8, threads = 4). A CSV file containing information on clips and their corresponding embeddings was generated as output.

Input and output (core model). We implemented a one-dimensional convolutional neural network (1D CNN) in PyTorch (Paszke et al., 2019) as the core model. It takes the pre-extracted BirdNet acoustic feature arrays (comma-delimited string with 1024 float values) as input. When loading, they are imported as numerical arrays. Although BirdNet feature vectors are not time-series data, adjacent dimensions encode related acoustic information in the learned representations, making the 1D CNN available for feature extraction.

We developed three different modes in the model: TRAIN, TEST, and EMBED. In the TRAIN mode, the model was trained using input feature vectors as the training set. Individual labels are encoded in the embedding CSV file as a new column for validation. The trained model parameters were saved as a checkpoint file. The TEST mode predicts the class (i.e., individual ID) of new samples from closed sets by comparing the output 256-d embeddings of unlabelled recordings against stored per-individual prototype arrays. The EMBED mode exports the 256-d embeddings as an NPY file, which is suitable for downstream open-set classification.

Model scaffold and layers (core model). Four convolutional blocks, an average pooling layer, a dropout layer, a projection head, and an L2 normalisation layer were developed and stacked sequentially as the model's basic architecture. We adapted this architecture from SPEAKERNET, a 1D CNN model for speaker recognition similar to bird individual identification (Koluguri et al., 2020). To extract local patterns, the model transformed data from shape [32, 1, 1024] (batch size = 32, original embedding shape = [1024]) to [32, 256, 31] by four convolutional blocks. Each convolutional block above has three sequential layers: a 1-D convolution, a batch normalisation, and a Gaussian Error Linear Unit (GELU) activation function. To achieve higher performance in acoustic recognition, we used GELU instead of the rectified linear unit (RELU), although the

computational complexity was increased (Hendrycks and Gimpel, 2023).

To reduce dimensionality, adaptive average pooling averaged each feature map across its temporal dimension to collapse the data structure into [32, 256]. To reduce overfitting and prevent the model from over-relying on single features, we randomly dropped n% of activations in the dropout layer (n is tunable for parameter adjustment). The projection head comprised a fully connected layer to project data into a 256-d space, followed by a 1-d batch normalisation. Finally, the L2 normalisation layer unified all embeddings to ensure Euclidean distance between samples equals their cosine dissimilarity, which is used for label prediction.

Model training and validation (core model). Before each training epoch, we augmented data by adding Gaussian noise with standard deviation σ to reduce overfitting and enhance robustness. Embedding elements were dropped out to avoid over-relying on single features (σ and *dropout* are tuneable for parameter adjustment). Each batch has 32 samples, composed of 8 individuals with 4 samples per individual. We trained the model using supervised contrastive (SupCon) loss (Khosla et al., 2021). Compared to traditional loss functions (e.g., cross-entropy loss), using the SupCon loss learns and structures the geometry of high-dimensional space directly by enlarging the between-individual distance and squeezing the within-individual distance, which is more suitable for downstream clustering analysis and open-set predictions. The sampling of 4 samples per individual per batch meets the requirements of at least 2 samples per class per batch for supervised contrastive learning (Khosla et al., 2021). During training, weights were optimised using AdamW (Loshchilov and Hutter, 2019) with a learning rate = 0.001 and a weight decay = 1×10^{-4}. We set the SupCon temperature (τ) as a tuneable hyperparameter for model optimisation. We ran each trial using 200 epochs, which is sufficient for model convergence (Fig. 4). The final model was saved as a checkpoint .pt file, useful for later predictions.

Closed-set prediction (core model). Prototype representation vectors of individuals were saved in the final checkpoint file during training. For each sample, cosine similarity between its 256-d embedding and every prototype vector was computed. The prototype class with the highest cosine similarity was selected as the prediction.

Open-set individual number counting (cluster and count block). Individual number estimation

of an open set, which has untrained individuals in it (most are untrained individuals for this study), relies on the 256-d embeddings generated by the core CNN model as input. We loaded embedding vectors stored as an NPY file as a NumPy array. We conducted a principal component analysis (PCA) to reduce the dimensionality of embeddings to d components (d is tuneable for parameter adjustment). Then, we performed agglomerative clustering, beginning with a cluster number equal to the sample size ($k = n$), and iteratively merging the nearest two clusters until only two clusters exist. In each step, the silhouette score was computed to evaluate clustering quality. The silhouette score was then plotted as a function of cluster number. To reduce noise, the signal was filtered using a Savitzky-Golay filter with window length = $0.2n$ and polyorder = 3. After smoothing, the range of k corresponding to silhouette scores greater than or equal to 90% of the global maximum was raised as the range prediction, as shown as follows:

$$[\hat{k}_{\text{low}}, \hat{k}_{\text{high}}] = [\min\{k : S(k) \geq 0.90\, S_{\max}\},\ \max\{k : S(k) \geq 0.90\, S_{\max}\}]$$

where $S(x)$ = smoothed silhouette score; k = the cluster number; $S_{\max}$ = the global maximum.

The k corresponding to the first local max, which has a height ≥ 90% of the global max before the global max, if there are any, was raised as the point prediction. Otherwise, k to the global max was used as the point prediction. The formula is shown as follows:

$$\hat{k}_{\text{point}} = \begin{cases} \min\{k < k^* : S(k) \geq 0.90\, S_{\max},\ S(k) > S(k-1),\ S(k) > S(k+1)\} & \text{if non-empty} \\ k^* & \text{otherwise} \end{cases}$$

where $k^* = \arg\max_k S(k)$

Instead of using the cluster number corresponding to the global maximum as the point prediction, this formula has shown higher prediction accuracy when tested.

Parameter optimisation and model evaluation

Before evaluating model performance, we did parameter optimisation to increase prediction accuracy. Prediction accuracy was evaluated using a random 85:15 train-test split of the original training set. Performance of closed-set predictions was evaluated while changing each of the following parameters: SupCon temperature (τ), Gaussian noise standard deviation (σ), and dropout rate, and keeping other parameters at default values. We generated five replicates for each

evaluation using different seedlings. After each round of tuning, the best-performing value was retained while the next parameter was evaluated. We then tested open-set individual number prediction accuracy using the optimised model. Parameter optimisation was performed by changing each of the following parameters while keeping other parameters at default values: BirdNet recognition threshold for the test set (θ), sampling control parameters (deleting individuals with fewer than m sample numbers and removing excessive samples from individuals with more than n numbers), and reserved PCs (d). The final pipeline used the parameter combination with the highest open-set prediction accuracy. The default parameter values used before updating were: τ = 0.07, σ = 0.02, $dropout$ = 0.10, m = 12, n = 48, d = 2, θ = 0.60.

We tested the performance of the final model. Closed-set prediction accuracy was evaluated using a random 80:20 train-test split of the original training set and the updated model. The mean accuracy of each epoch after model convergence was computed to represent the prediction accuracy of each run. We selected the model with the highest prediction accuracy in five runs as the final model. Except for overall prediction accuracy, accuracy changes under the impact of sex and sample size were reported. We reported the overall open-set number estimation accuracy using models with updated parameters trained on the full training set. The model with the lowest sum of squared residuals (SSR) in five runs was selected as the final model. Model robustness was tested by fitting multiple linear models using environmental and acoustic parameters as independent variables.

Statistical analysis

To compare closed-set prediction accuracy differences, we performed a Kruskal–Wallis test to test differences across the groups. If a significant difference was detected, we then did a post hoc Dunn test for pairwise comparisons. To investigate whether sex and sample size affect prediction correctness, a generalised linear model (GLM) was built, with a binomial error distribution and a logit link function. The response variable was specified as counts of correct and incorrect predictions for each individual. To test how environmental and acoustic factors impact open-set individual counting accuracy, multiple linear regression models were fitted. The significance level was set as α = 0.05. All statistical analyses were performed in R (version 4.4.1) and RStudio

(version 2024.4.2.764).

Results and Discussion

Dataset summary

We collected Manx shearwater audio recordings using Automated Recording Units in Lundy, St. Agnes, and Copeland. After manual selection, the dataset comprised 166 individuals, including 103 males and 63 females. BirdNet detections generated a total of 3,939 clips, with 23.7 clips per individual on average. Clip number varied from 1 to 66 per individual. Different sampling controls generated multiple training sets that are either more balanced in sample size or with more samples per individual (Table 1). The test set selected from peak calling periods included 56 samples, with estimated real individual numbers ranging from 2 to 24, and male/female ratios from 0.13 to 5.0. Background noise, average call intensity, and overlapping call proportions across samples were also highly variable, providing heterogeneity for testing model accuracy under different acoustic conditions.

Sample number per individual (min-max)	Clip number	Individual number (male/female)
12-15	1844	125 (90/35)
12-20	2314	125 (90/35)
12-36	3082	125 (90/35)
12-48	3244	125 (90/35)
12-66	3287	125 (90/35)
20-24	2036	87 (71/16)
20-48	2670	87 (71/16)
20-66	2713	87 (71/16)

Table 1. Summary of training sets under different sampling controls. Clips less than 2.6s were removed before generating all sets. 66 was the maximum clip number per individual by the initial selection.

Parameter optimizations

We performed parameter optimisations to maximise closed-set prediction accuracy by changing τ, σ, and *dropout*, while keeping other parameters as default. All models with varied τ converged

after 170 epochs (Fig. 4), and model performance was evaluated using the mean prediction accuracy between epochs 170 and 200. Models with a τ from 0.04 to 0.15 showed no significant differences in prediction accuracy, but all performed better than using $\tau = 0.01$ (Fig. 2A). Model performance showed no significant differences among σ from 0 to 0.10 (Fig. 2B). Models using 10% to 30% dropouts during training led to a significantly higher performance than non-dropouts (Fig. 2C). The above tests show that models using the default model parameter settings have equivalent or higher close-set prediction accuracy to other tested parameter combinations.

We conducted open-set point prediction by using the best closed-set prediction model to generate embeddings, which were then processed by PCA and clustering. SSR between predicted values and estimated real values was used as the evaluation metric. Test set pre-processing using different BirdNet recognition thresholds ($\theta = 0.10$, 0.20, 0.40, and 0.60), produced no significant differences in model SSR (Fig. 3A). Different sampling controls of the training set, including limiting maximum sample numbers and removing individuals with insufficient samples, did not cause significant model SSR differences (Fig. 3B). When changing PC, models of 2 PC had a significantly lower SSR than models with 4 and 6 PC (Fig. 3C). These results show that the default parameter settings for the open-set prediction retain equal or better model performances than using other parameter combinations. Also, modifications of the training set and test set exhibit limited impact on open-set prediction accuracies.

In parameter testing, artificially added Gaussian noise (σ from 0.001 to 0.1) did not change the prediction accuracy, indicating the model is robust in dealing with signal noise. While so, embedding element dropout during training does improve closed-set recognition accuracy. Surprisingly, sampling controls do not significantly impact model SSR in open-set counting. It is possible that reserving 12-20 samples per individual is already sufficient in training the model for open-set predictions. When adjusting PCs, $d = 2$ works best compared with reserving more components for clustering, which suggests that the clustering algorithm works better for low-dimensional structures. Retaining more PCs may introduce more noise, thereby reducing the effectiveness of the clustering.

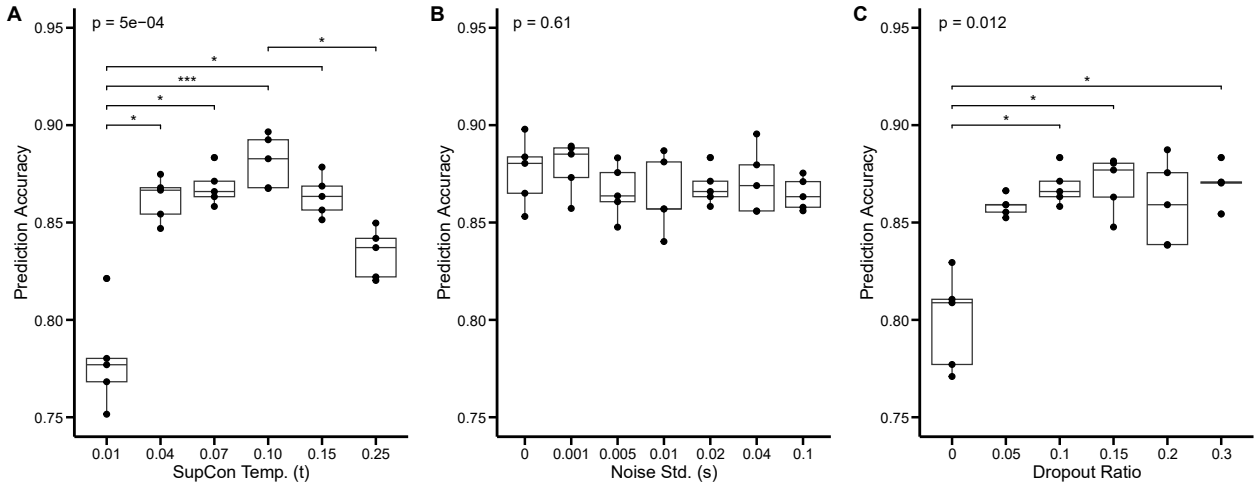

Figure 2. Close-set prediction accuracy varies with different model parameters. Each replicate represents the mean prediction accuracy between epochs 170 and 200 for independent runs, using a random sample of 85% of the original training set. Except for the independent variable, other parameters are set to default. Comparison between pairs was conducted using a Kruskal–Wallis test followed by post hoc Dunn tests. Only significant outcomes are shown. (A) Prediction accuracy by models using different SupCon temperatures. (B) Prediction accuracy by models adding Gaussian noise with different standard deviations. (C) Prediction accuracy by models randomly dropping embedding elements in different proportions.

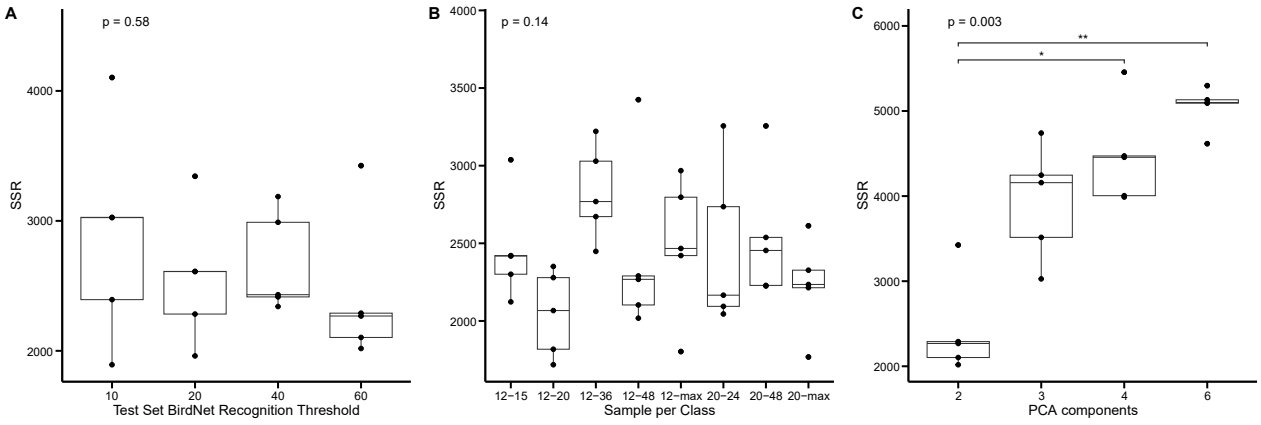

Figure 3. Comparison of open-set prediction sum of squared residuals (SSR) under different model parameters. Models were trained using default parameters and 200 epochs. Each replicate represents the model's SSR using random seedlings when training the same model. Comparison between pairs was conducted using a Kruskal–Wallis test followed by post hoc Dunn tests. Only significant outcomes are shown. (A) SSR by test set detection using different BirdNet recognition thresholds. (B) SSR by different

sampling controls for the training set. The independent variable range represents the minimum to maximum number of samples per class. (C) SSR by different principal components.

Closed-set predictions

Models with optimised parameters were trained in five replicates using a random 8:2 split of the training set as the new training and validation set. All five replicates achieved high prediction accuracy with low variation after convergence (mean = 88.6%, SD = 0.9%), with the top-1 reaching 89.3%, which was then selected as the final model. When examining sex-specific mismatches, only 4 of the 542 predictions (0.7%) had incorrect sex predictions (Fig. 5). A generalised linear model showed that sample size did not significantly affect prediction accuracy ($\beta = 0.03$, SE = 0.04, $z = 0.61$, $P = 0.54$). Sex also showed no significant effect ($\beta = -0.72$, SE = 0.49, $z = -1.48$, $P = 0.14$).

The high accuracy of closed-set predictions suggests that Manx shearwater vocalisations contain individual signatures, consistent with the findings by Sun et al. (2024). Furthermore, prediction accuracy is high and shows no difference between males and females, indicating both sexes contain sufficient individual signals for recognition. Earlier studies in a similar species, Yelkouan shearwaters (*Puffinus yelkouan*), report that both sexes are acoustically identifiable (Cure et al., 2009), and Sun et al. (2024) suggest that female Manx shearwater calls are identifiable by successfully discriminating 3 females, which had limited generalisability. We thus provide stronger evidence suggesting that both sexes of Manx shearwater contain individual signatures, by successfully discriminating 95 males and 35 females. The closed-set recognition accuracy was pushed to 89.3% but not higher, which may be due to limited training data for some individuals and incomplete call segmentation when generating clips. These limitations could be addressed in future studies by increasing sample size and developing new algorithms for improved call segmentation. Although not confirmed, the unitary non-linear calls present in the dataset may contain weaker individual signals and impact model performance, and future studies are needed to scrutinize call-type-specific individual signature differences.

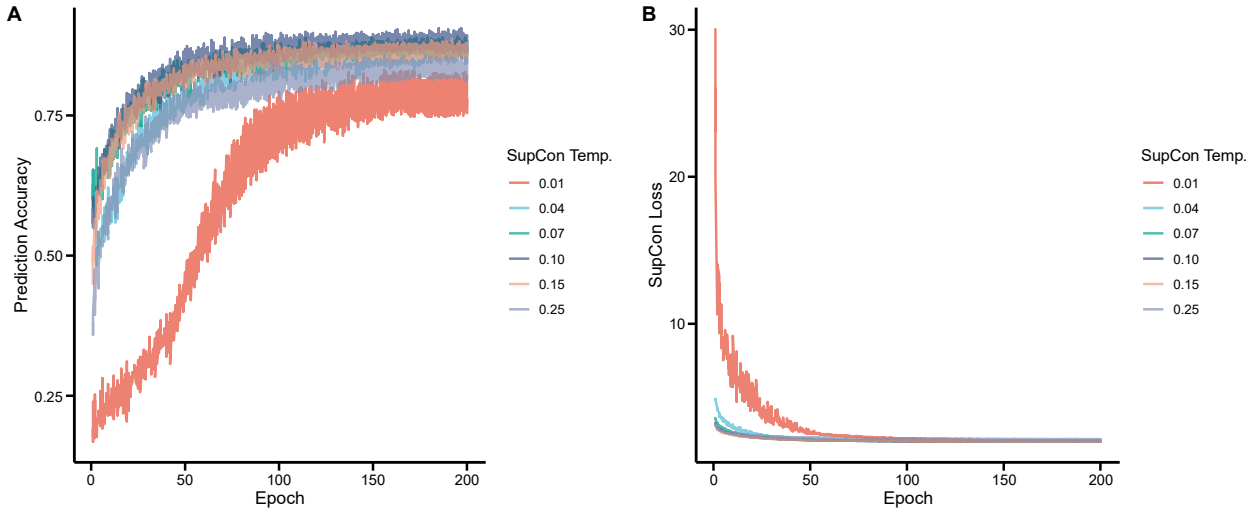

Figure 4. The model reaches convergence after 170 epochs. All trials converge in 170-200 epochs, as indicated by (A) prediction accuracy and (B) SupCon loss, regardless of SupCon temperatures. Each trial has 5 replicates. Different colours show different SupCon temperature values.

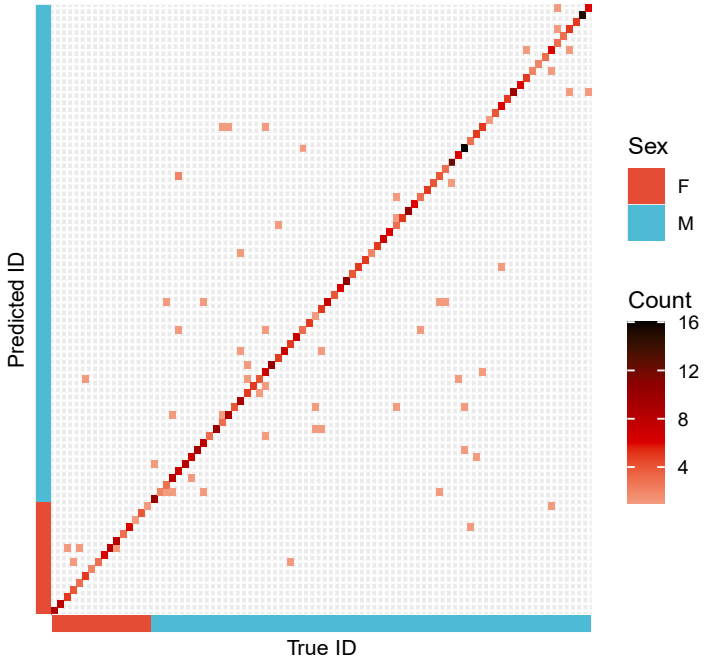

Figure 5. Confusion matrix showing the closed-set prediction quality of Manx shearwater individuals. Each coloured cell represents a prediction-real label pair, with darker colour for higher counts. Diagonal cells represent correct predictions. Coloured bars along the x- and y-axis show the sex of each individual.

Open-set predictions

Open-set individual number estimation was based on the model with the lowest SSR during the optimisation (SSR = 1718, $n = 56$) (Fig. 6). The mean absolute error was 4.46 ± 3.31 SD, and the mean relative error was 39.1% ± 30.4% SD, with seven samples greater than 60.0% for point predictions. True values fell into the predicted range for 48 out of 56 samples when using range predictions (as described in the methods). However, the mean prediction range was broad with low precision (mean width over the true value = 231.1%, SD = 130.1%). Multiple linear regression models were built to investigate the impact of recording conditions and environmental factors on prediction error (Fig. 7). The regression model showed that individual count had a negative relationship with model residual ($\beta = -0.81$, SE = 0.16, $t = -5.21$, $P < 0.001$), while having a positive correlation with prediction absolute error presented by the absolute value of residual ($\beta = 0.31$, SE = 0.11, $t = 2.97$, $P < 0.01$). Call quantity (clip number) had a positive relationship with residual ($\beta = 0.07$, SE = 0.03, t = 2.79, $P < 0.01$), and female/male ratio was negatively associated with absolute error ($\beta = -0.54$, SE = 0.27, $t = -2.02$, $P < 0.05$). This suggests that the model tends to overestimate the number of individuals in recordings containing fewer individuals and more calls. Furthermore, prediction accuracy increased when samples had fewer individuals and a greater female ratio. Other variables, including the overlapping call ratio, background noise, and average peak volume, did not show significant correlations with either residual or absolute error.

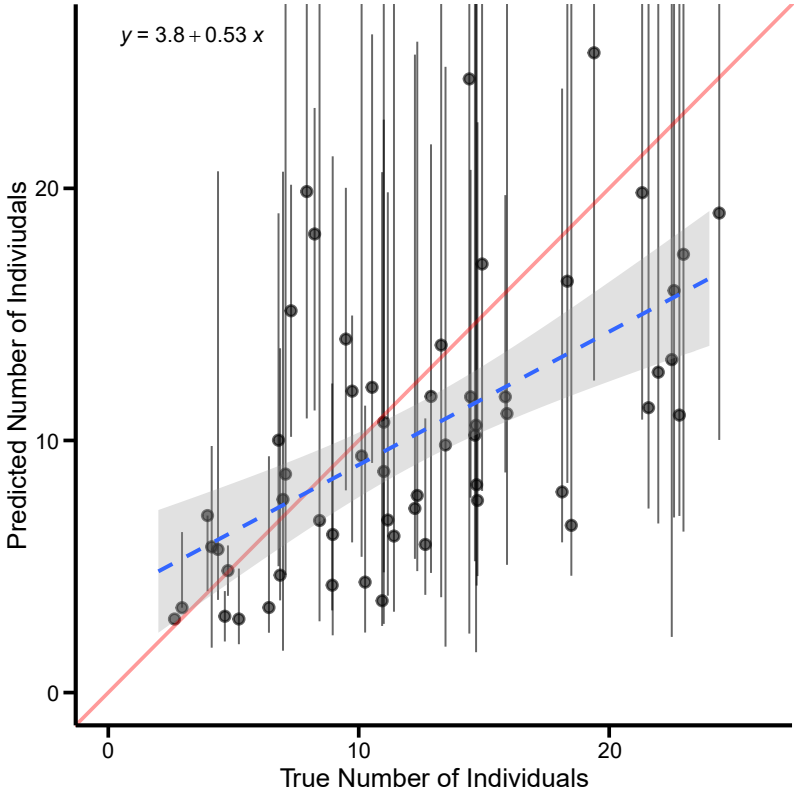

Figure 6. Predicted versus true numbers of individuals in 20-minute open-set Manx shearwater recordings. Each point represents the point prediction value, while each bar represents the predicted range (using a threshold of 90% of the silhouette score maximum). The red line shows perfect predictions ($y = x$). The blue dashed line and the grey area show the linear regression with 95% confidence band of point predictions. Coefficients are shown at the top. Points and bars are slightly jittered to avoid overlaps. Predicted ranges greater than 27 are not shown.

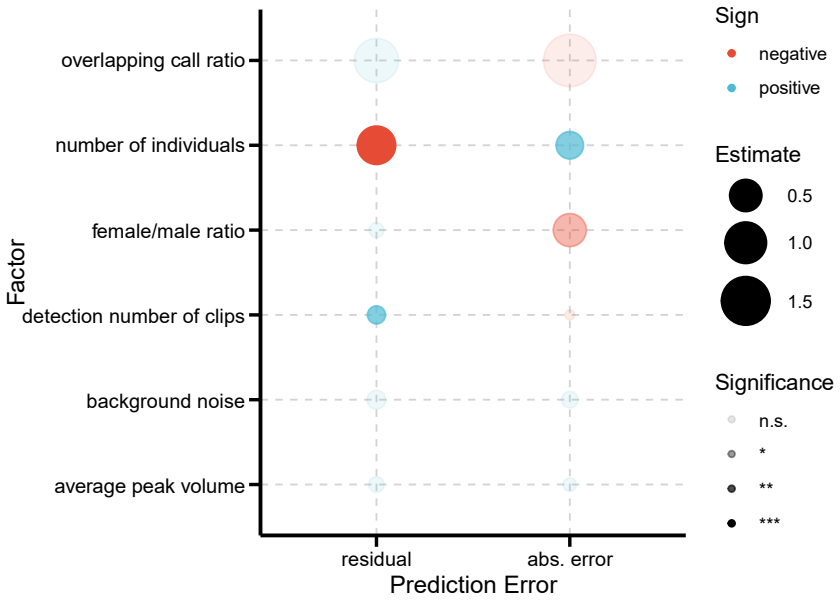

Figure 7. The correlation strength between factors and prediction errors. Two models were fitted using residues and the absolute value of residues (absolute error) as response variables. Y-axis terms are tested factors that may impact the prediction error. The linear regression estimate (β) is shown by the point size, and the sign of the slope is shown by colour. Stronger significance is indicated by stronger opacity.

A mean relative error of 39.1% shows the pipeline is promising, although it is not sufficiently accurate for reliable abundance estimation. Compared to the open-set object counting by Zheng et al. (2022), which achieved a lower prediction error (mean relative error = 21%), estimating population abundance using acoustics has greater difficulty because individual vocalisations are highly similar, frequently with overlaps and environmental noise. Although the same individual tends to be clustered, boundaries between them are often vague, given that within-individual calls can be heterogeneous across short time (Singer et al., 2026). The presence of non-typical calls and overlaps is hypothesised to generate random noise in the embedding space, which would lead to failures in producing clustered local structures and the resulting biased estimates. Nevertheless, background noise, call volume, and call overlaps have limited effects on model performance, demonstrating model robustness in varied recording conditions. Model performance is impacted by the number of individuals, with reduced accuracy and a trend of underestimation when more calling individuals are presented. One possible explanation is that more individuals generate more overlapping calls and greater difficulty for clustering in a limited high-dimensional space. Pre-

processing by BirdNet is another source of prediction bias. Low prediction confidence for non-typical female calls can lead to reduced female clip numbers used for training and prediction, leading to underrepresentation of females and underestimation of caller numbers. Interestingly, an increased female sample ratio did not significantly increase residual but did correlate with reduced absolute error in this study, showing improved accuracy. It is possible that female calls contain more repetitive patterns and are less variable than males, leading to greater open-set prediction certainty.

From a conservation perspective, this study helps estimate local seabird population size in a non-invasive, time-efficient, and easy way. With further optimisation, the pipeline has the potential to achieve higher accuracy in estimating population size in comparison with the traditional approach of counting occupied burrows using playback. When applying this pipeline to real scenarios, data from different monitoring devices can be combined for estimating per-colony and per-island populations, which is an advance from the current methodologies. However, limitations of this approach exist, as flight calls are released predominantly by females (Brooke, 1978; James, 1985), so males in flight will be neglected, leading to bias. Future research directions could try fitting regressions of predicted values versus real population numbers to calibrate the prediction.

In Conclusion, in this study we demonstrate that machine learning approaches can identify Manx shearwater from their vocalisations with high accuracy and show potential for estimating the minimum number of calling individuals, representing a promising advance in bioacoustics for individual recognition. The approach is robust against noise and external recording conditions. The findings highlight its potential to support Manx shearwater monitoring and, with further validation, can be adapted to the study of other seabird species and potentially other vocal taxa. We also provide an approach to detect individual signatures in animals' vocalisations, and present a new method for understanding their population dynamics through PAM.

Acknowledgements

We thank Simon Parker, Warden of Lundy Island, and Hester Odgers, Warden in the Isles of Scilly,

for their support and assistance during fieldwork. We also thank the staff and wardens at Copeland Island for their valuable logistical and field support. We gratefully acknowledge the Royal Society of Biology for providing travel grant support.

Funding Statement

This work was supported by a President's PhD Scholarship from Imperial College London awarded to YJA, a master's student research project budget awarded to ZL, and a Lundy Field Grant awarded to ZL.

Author Contributions

ZL: Conceptualisation; Investigation; Methodology; Visualisation; Writing – original draft; Writing – review and editing. YJA: Conceptualisation; Investigation; Methodology; Visualisation; Writing – review and editing. WDP: Supervision; Writing – review and editing. SAZ: Conceptualisation; Supervision; Writing – review and editing. JS: Conceptualisation; Supervision; Writing – review and editing. All authors reviewed and approved the manuscript.

Conflict of Interest Statement

We declare no conflict of interest.

Data and Code Availability

Data used in developing and testing the model, including model checkpoint files and training set and test set embedding inputs, can be found in the following archive: "Dataset for Manx shearwater (Puffins puffins) individual acoustic identification". Zenodo. Available at: https://doi.org/10.5281/zenodo.21297859.

Python and R code used for building the model and statistical analysis can be found in the

following GitHub repository: https://github.com/Antongsky/Shearwater_individual_recognition.

References

Arneill, G.E., Critchley, E.J., Wischnewski, S., Jessopp, M.J., Quinn, J.L., 2020. Acoustic activity across a seabird colony reflects patterns of within-colony flight rather than nest density. *Ibis 162*, 416–428. https://doi.org/10.1111/ibi.12740

Avaria-Llautureo, J., Rivadeneira, M.M., Venditti, C., Luna-Jorquera, G., 2026. Seabird range contraction and dispersal under climate change. *Nat. Clim. Change* 1–7. https://doi.org/10.1038/s41558-026-02655-4

Bailey, H., Fandel, A.D., Silva, K., Gryzb, E., McDonald, E., Hoover, A.L., Ogburn, M.B., Rice, A.N., 2021. Identifying and predicting occurrence and abundance of a vocal animal species based on individually specific calls. *Ecosphere* 12, e03685. https://doi.org/10.1002/ecs2.3685

Bedoya, C.L., Molles, L.E., 2021. Acoustic censusing and individual identification of birds in the wild. https://doi.org/10.1101/2021.10.29.466450

Blackmer, A.L., Ackerman, J.T., Nevitt, G.A., 2004. Effects of investigator disturbance on hatching success and nest-site fidelity in a long-lived seabird, Leach's storm-petrel. *Biol. Conserv.* 116, 141–148. https://doi.org/10.1016/S0006-3207(03)00185-X

Brooke, M., 1990. The Manx shearwater. London : Poyser ; San Diego, CA : Academic Press.

Cure, C., Aubin, T., Mathevon, N., 2009. Acoustic convergence and divergence in two sympatric burrowing nocturnal seabirds. *Biol. J. Linn. Soc.* 96, 115–134. https://doi.org/10.1111/j.1095-8312.2008.01104.x

de L. Brooke, M., 1978. Sexual differences in the voice and individual vocal recognition in the Manx shearwater (*Puffinus puffinus*). *Anim. Behav.* 26, 622–629. https://doi.org/10.1016/0003-3472(78)90074-X

Dias, M.P., Martin, R., Pearmain, E.J., Burfield, I.J., Small, C., Phillips, R.A., Yates, O., Lascelles, B., Borboroglu, P.G., Croxall, J.P., 2019. Threats to seabirds: A global assessment. *Biol. Conserv.* 237, 525–537. https://doi.org/10.1016/j.biocon.2019.06.033

Griesser, M., Schneider, N.A., Collis, M.-A., Overs, A., Guppy, M., Guppy, S., Takeuchi, N., Collins, P., Peters, A., Hall, M.L., 2012. Causes of Ring-Related Leg Injuries in Birds – Evidence and Recommendations from Four Field Studies. *PLOS ONE* 7, e51891. https://doi.org/10.1371/journal.pone.0051891

Haney, J.C., Geiger, H.J., Short, J.W., 2014. Bird mortality from the Deepwater Horizon oil spill. I. Exposure probability in the offshore Gulf of Mexico. *Mar. Ecol. Prog. Ser.* 513, 225–237. https://doi.org/10.3354/meps10991

Hendrycks, D., Gimpel, K., 2023. Gaussian Error Linear Units (GELUs). *arXiv.org.* https://doi.org/10.48550/arXiv.1606.08415

Hensel, A.L.J., Dobney, S.L., Moran, I.G., Thomas, I.P., Burant, J.B., Woodworth, B.K., Doucet, S.M., Newman, A.E.M., Norris, D.R., Williams, H., Mennill, D.J., 2022. Passive acoustic monitoring provides predictable and reliable underestimates of population size and longevity in wild Savannah Sparrows. *Ornithol. Appl.* 124, duac018. https://doi.org/10.1093/ornithapp/duac018

Huang, L., Clarke, R.H., Teixeira, D., Chiaradia, A., Meyer, B., 2025. Acoustic recognition of individuals in

closed and open bird populations. *Ecol. Inform.* 90, 103330. https://doi.org/10.1016/j.ecoinf.2025.103330

James, P.C., 1985. The Vocal Behaviour of the Manx Shearwater *Puffinus puffinus*. *Z. Für Tierpsychol.* 67, 269–283. https://doi.org/10.1111/j.1439-0310.1985.tb01394.x

Kahl, S., Wood, C.M., Eibl, M., Klinck, H., 2021. BirdNET: A deep learning solution for avian diversity monitoring. *Ecol. Inform.* 61, 101236. https://doi.org/10.1016/j.ecoinf.2021.101236

Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D., 2021. Supervised Contrastive Learning. *arXiv.org.* https://doi.org/10.48550/arXiv.2004.11362

Koluguri, N.R., Li, J., Lavrukhin, V., Ginsburg, B., 2020. SpeakerNet: 1D Depth-wise Separable Convolutional Network for Text-Independent Speaker Recognition and Verification [WWW Document]. *arXiv.org.* https://arxiv.org/abs/2010.12653v1 (accessed 5.21.26).

Lee, D.S., Haney, J.C., Carboneras, C., Jutglar, F., Kirwan, G.M., 2023. Manx Shearwater (Puffinus puffinus), version 1.1. *Birds World.* https://doi.org/10.2173/bow.manshe.01.1

Lengagne, T., Aubin, T., Lauga, J., Jouventin, P., 1999. How do king penguins (Aptenodytes patagonicus apply the mathematical theory of information to communicate in windy conditions? *Proc. R. Soc. B Biol. Sci.* 266, 1623–1628. https://doi.org/10.1098/rspb.1999.0824

Lengagne, T., Slater, P.J.B., 2002. The effects of rain on acoustic communication: tawny owls have good reason for calling less in wet weather. *Proc. R. Soc. B Biol. Sci.* 269, 2121–2125. https://doi.org/10.1098/rspb.2002.2115

Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., Song, L., 2018. SphereFace: Deep Hypersphere Embedding for Face Recognition. *arXiv.org.* https://doi.org/10.48550/arXiv.1704.08063

Loshchilov, I., Hutter, F., 2019. Decoupled Weight Decay Regularization. *arXiv.org.* https://doi.org/10.48550/arXiv.1711.05101

Meissner, W., 2025. Significant Reduction in the Impact of Oil Spills and Chronic Oil Pollution on Seabirds: A Long-Term Case Study from the Gulf of Gdańsk, Southern Baltic Sea. *Sustainability* 17, 8037. https://doi.org/10.3390/su17178037

Ntalampiras, S., Potamitis, I., 2021. Acoustic detection of unknown bird species and individuals. *CAAI Trans. Intell. Technol.* 6, 291–300. https://doi.org/10.1049/cit2.12007

Oppel, S., Hervias, S., Oliveira, N., Pipa, T., Silva, C., Geraldes, P., Goh, M., Immler, E., McKown, M., 2014. Estimating population size of a nocturnal burrow-nesting seabird using acoustic monitoring and habitat mapping. *Nat. Conserv.* 7, 1–13. https://doi.org/10.3897/natureconservation.7.6890

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv.org.* https://doi.org/10.48550/arXiv.1912.01703

Perkins, A.J., Douse, A., Morgan, G., Cooper, A., Bolton, M., 2017. Using dual-sex calls improves the playback census method for a nocturnal burrow-nesting seabird, the Manx Shearwater *Puffinus puffinus*. *Bird Study* 64, 146–158. https://doi.org/10.1080/00063657.2017.1307940

Ptacek, L., Machlica, L., Linhart, P., Jaska, P., Muller, L., 2016. Automatic recognition of bird individuals on an open set using as-is recordings. *Bioacoustics* 25, 55–73. https://doi.org/10.1080/09524622.2015.1089524

Sanz-Aguilar, A., Tavecchia, G., Minguez, E., Massa, B., Lo Valvo, F., Ballesteros, G.A., Barberá, G.G., Amengual, J.F., Rodriguez, A., McMinn, M., Oro, D., 2010. Recapture processes and biological inference in monitoring burrow-nesting seabirds. *J. Ornithol.* 151, 133–146.

https://doi.org/10.1007/s10336-009-0435-x
Singer, D., Kamp, J., Hondong, H., Schuldt, A., Hagge, J., 2026. Diel and seasonal vocal activity patterns revealed by passive acoustic monitoring suggest expert recommendations for breeding bird surveys need adjustment. *J. Ornithol*. 167, 165–180. https://doi.org/10.1007/s10336-025-02307-y
Stowell, D., Petrusková, T., Šálek, M., Linhart, P., 2019. Automatic acoustic identification of individuals in multiple species: improving identification across recording conditions. *J. R. Soc. Interface* 16, 20180940. https://doi.org/10.1098/rsif.2018.0940
Sun, Y., Dunning, J., Taylor, T., Schroeder, J., Anne Zollinger, S., 2024. Calls of Manx shearwater Puffinus puffinus contain individual signatures. *J. Avian Biol.* 2024, e03170. https://doi.org/10.1111/jav.03170
Walsh, P., Nevo, A. de, Halley, D.J., Sim, I.W.M., Harris, M.P., 1995. Seabird monitoring handbook for Britain. Joint Nature Conservation Committee, Peterborough.
Xie, J., Zhang, L., Zhang, J., Zhang, Y., Schuller, B.W., 2023. Cross-corpus open set bird species recognition by vocalization. *Ecol. Indic.* 154, 110826. https://doi.org/10.1016/j.ecolind.2023.110826
Zheng, J., Li, W., Hong, J., Petersson, L., Barnes, N., 2022. Towards Open-Set Object Detection and Discovery. *arXiv.org.* https://doi.org/10.48550/arXiv.2204.05604