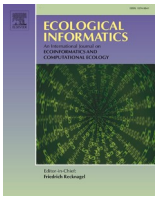
Ecological Informatics 92 (2025) 103474

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



Automated classification of albatross acoustic behaviour at sea: A free and open-source classifier for seabird sounds

Aline da Silva Cerqueira [a,b,*], Robin Freeman [b], Richard A. Phillips [c], Terence P. Dawson [a]

[a] Department of Geography, King's College London, London WC2R 2LS, UK
[b] Institute of Zoology, Zoological Society of London, London NW1 4RY, UK
[c] British Antarctic Survey, Natural Environment Research Council, High Cross, Madingley Road, Cambridge CB3 0ET, UK

ARTICLE INFO

Keywords:
Seabirds
Bioacoustics
Acoustic monitoring
Machine learning
Convolutional neural network
Conservation science
Google Colab

ABSTRACT

Advancements in acoustic data collection technologies have greatly increased their use in wildlife monitoring, but produce large volumes of data that are challenging to analyse manually. Recent developments in machine learning, particularly convolutional neural networks (CNNs), have transformed audio data analysis, enabling efficient and accurate sound classification. This study aimed to develop a method for automatic classification of behaviour (on-water activity, flight, vocalisation and preening) from sounds recorded by free-ranging albatrosses of two species equipped with audio recorders during foraging trips at sea. Using a manually labelled seabird audio dataset, a general-purpose CNN model was created and trained in Google Colab. The model development followed a structured workflow, including audio data preparation, pre-processing, model architecture and training, and performance evaluation. The model achieved a global accuracy and precision of 95 % during testing. Despite high overall accuracy, performance varied across sound categories due to the inherent complexity of distinguishing behaviours, leading to differences in prediction errors. This study primarily focused on developing and validating an accessible, high-performance workflow for automated acoustic classification, with the goal of enabling future ecological and conservation applications. It demonstrated that a generic web-based CNN model can effectively classify seabird sounds into different behaviours with high accuracy. The approach provides a foundation for future ecological and conservation applications, enabling detailed exploration of activities, interactions and environmental context of seabird behaviour using acoustic data. By leveraging open-source platforms and accessible tools, this work provides a foundation for future advancements in automated acoustic monitoring, making it accessible to a diverse range of researchers.

1. Introduction

Sounds emitted by animals reflect their behaviour, physiology and activities, offering first-hand insights into different aspects of their lives and environmental contexts (Bradbury and Vehrencamp, 2011; Snaddon et al., 2013; Tosa et al., 2021). The advent of autonomous recording devices has facilitated the sampling of animal soundscapes across diverse environments and time scales, allowing for the collection and storage of high-quality recordings in a non-intrusive manner (Duarte et al., 2021; Laiolo, 2010; Towsey et al., 2014). However, the analysis and classification of large volumes of acoustic data present challenges due to their inherent complexity and scale (Stowell et al., 2019). The importance of following best practices for acoustic data collection and analysis, including rigorous documentation, standardised annotation protocols, and transparent metadata handling, has been highlighted as critical for ensuring reproducibility and data utility (Oswald et al., 2022). Conventional methods involving auditory and visual inspection, as well as manual labelling of audio files, can be labour-intensive and error-prone, making the analysis process time-consuming and potentially inaccurate (Digby et al., 2013; Stowell et al., 2019; Swiston and Mennill, 2009).

Advancements in machine learning, particularly convolutional neural networks (CNNs), have revolutionised the way that audio recordings can be analysed and classified (Hershey et al., 2017; Purwins et al., 2019; Xie et al., 2018). CNNs are deep learning models designed for image processing that can be adapted to inspect spectrograms of audio recordings, enabling the automated identification and classification of sounds with high accuracy (Fairbrass et al., 2019; Salamon and Bello,

* Corresponding author at: Department of Geography, King's College London, London WC2R 2LS, UK.
E-mail address: aline.cerqueira@kcl.ac.uk (A. da Silva Cerqueira).

https://doi.org/10.1016/j.ecoinf.2025.103474
Received 2 December 2024; Received in revised form 3 August 2025; Accepted 10 October 2025
Available online 18 October 2025

2017). CNNs automatically extract features from audio files that describe their spectral and temporal characteristics such as audio time signal, peak frequency and frequency range (Browning et al., 2017; Dai et al., 2017). These features are then used by the neural network to recognise patterns in the audio data and to sort them into categories based on their characteristics. However, the training and optimisation of CNNs require large amounts of labelled data and computing resources, which is computationally intensive (Aodha et al., 2014; Goëau et al., 2016; Xie et al., 2018).

In recent years, several web-based interactive computing environments have been developed that offer an accessible and collaborative platform for researching and developing deep learning models. One such platform is Google's Colaboratory (hereafter "Colab"), a free, open-source, web-based interactive computing environment (Jupyter notebook) that allows computer code written in the Python language to be run on Google's cloud infrastructure, with support for both GPUs (Graphical Processing Units) and TPUs (Tensor Processing Units) (Dwivedi, 2025). This feature is particularly useful for training large-scale machine learning models when processing power on local machines is limited. Colab also comes with pre-installed machine learning libraries, including TensorFlow and PyTorch, used in deep learning frameworks for automated audio classification (Colaboratory, 2025; Dwivedi, 2025; Yalçın, 2020). Additionally, Colab's notebooks are stored in Google Drive, making it easy to share, edit, and collaborate on the framework from any location with internet access.

Leveraging Colab's accessibility and computational power offers a promising solution for overcoming the current challenges in classifying animal audio data. Animal-borne acoustic recorders provide a novel method for remote studies of ecology and behaviour, delivering valuable insights into activities and the environmental conditions in which sounds are produced (e.g., Clayton et al., 2023; Stowell et al., 2017; Thiebault et al., 2019; Wijers et al., 2018). Streamlining and accelerating the classification of animal audio data is essential for fully realising the potential of acoustic monitoring in ecological, behavioural, and conservation research. Developing open-source, cost-effective classification methods would greatly advance animal acoustic monitoring, making it more accessible and user-friendly for a broad range of researchers, from experts to non-experts.

Albatrosses are among the most threatened of all bird families, at particular risk from incidental mortality (bycatch) in fisheries, climate change and invasive species (Dias et al., 2019; Phillips et al., 2016). Analyses of acoustic data recorded by albatrosses at sea may therefore provide information on threats (e.g. proximity to fishing vessels, oil platforms or wind turbines), as well as on activity patterns or social interactions (Monier, 2024). With these applications in mind, we developed a general-purpose CNN model using Colab to automatically classify sounds recorded by free-ranging albatrosses of two species during their foraging trips at sea. The model was trained on a manually labelled subset of the audio dataset, which served as a benchmark for performance evaluation. We then assessed the accuracy and precision of the model in classifying seabird sounds. Results are discussed in terms of the broader application of this automated approach. As far as we are aware, the only previous studies that have analysed acoustic data from foraging seabirds were on penguins and cape gannets *Morus capensis* (McInnes et al., 2020; Thiebault et al., 2019b; Thiebault et al., 2021). Our study primarily focuses on developing and validating an accessible, high-performance workflow for automated acoustic classification, with the goal of enabling future ecological and conservation applications. It highlights the potential of integrating open-source deep learning platforms like Colab with animal-borne audio recorders to enhance our understanding of seabird behaviour and ecology. By streamlining the classification process, we aimed to make acoustic data analysis more accessible and efficient, thus paving the way for broader adoption of these methods in wildlife research and conservation.

2. Method

2.1. Audio data acquisition, classification, and dataset preparation

The acoustic datasets used in this study were collected from five black-browed albatrosses (*Thalassarche melanophris*) and five wandering albatrosses (*Diomedea exulans*) during the brood-guard period in austral summer 2014/15 at Bird Island (54°00′S, 38°03′W), South Georgia. Before departing their nests for foraging trips at sea, birds were equipped with Edic-mini Tiny Solar-300 h digital audio recorders (TS-Market Ltd., Moscow, Russia), IGotU GT-120 GPS loggers (Mobile Action Technology Inc., Taiwan), and Intigeo C250 combined light-level geolocator and immersion sensors (Migrate Technology Lt, Cambridge, United Kingdom). The audio recorders were set to record continuously at a of 22 kHz sampling rate in mono, using ADPCM compression internally to optimise storage. All devices were recovered from the birds upon their return to the colony, and the data were downloaded. Audio files were exported as standard WAV files.

A total of 436 h of seabird audio recordings made at sea were analysed manually and classified using WavePad Masters Sound Editor version 8.04 (NCH Software, Canberra, Australia, 2020). As a first step, each file was played in full to verify the presence of audible seabird sounds, and the at-sea segments were identified using GPS and immersion data. Sounds were identified through auditory examination and visual inspection of spectrograms and waveforms. Acoustic cues were used to assign labels as follows: flight was recognised by flapping sounds, rhythmic hops, or wind rushing over wings during gliding; vocalisations were calls, sometimes modulated or repeated, produced by the tagged bird or conspecifics; preening was characterised by repetitive tapping, scratching, or rubbing noises; and on-water activity included splashing, paddling, or water displacement. As studies of albatross acoustic behaviour at sea are extremely scarce – with most studies limited to colony-based observations (Makris et al., 2002; Pickering and Berrow, 2001) - this study represents a pioneering effort in characterising at-sea acoustic behaviour in these species.

Classifications were manually annotated in an Excel spreadsheet, with time-stamped descriptions of the identified sounds (e.g., "splash sound during landing" or "repeated tapping consistent with preening") to facilitate cross-referencing with GPS and immersion data. To support reproducibility and future reuse, metadata were systematically organised and documented retrospectively alongside the acoustic data, in alignment with best-practice recommendations for bioacoustic data collection (Oswald et al., 2022). Metadata included device settings (sampling frequency, compression format), recording context (at-sea versus colony-based segments identified through GPS and immersion data), time-stamped annotations of sound events, and behavioural classifications. All manual annotations were documented in structured spreadsheets linked to each audio file, providing a transparent framework for validation and further analysis.

File segments recorded before the first and after the last GPS fix were excluded, as they corresponded to time spent at the colony and the study focused on capturing seabird sounds at sea. A randomised quality control process was conducted independently by two researchers, who classified 10 % of the dataset without access to the original labels. Discrepancies were discussed and resolved by consensus to minimise subjectivity. For model training, only audio segments that clearly represented one of the four sound categories – flight, preening, vocalisation, and on-water activity - were selected from the main dataset (Table 1; Fig. 1).

The final set of manually classified audio segments was used to create distinct folders corresponding to the four target sound categories, which served as the training dataset for the automated seabird sound classifier. This approach ensured that the training data did not include overlapping seabird sounds, which was identified as a potential issue during the manual analysis phase. Although audio segments containing only one type of seabird sound were scarce, it was possible to create a subset

Table 1
Acoustic categories and descriptions of seabird sounds identified from bird-borne audio recordings of wandering and black-browed albatrosses at sea.

Sound category	Acoustic description
Flight	Flapping sounds during wing beats, hopping noises, and wind rushing over wing surfaces during gliding flight.
Preening	Repetitive tapping, scratching, or rubbing sounds associated with feather maintenance and grooming behaviours.
Vocalisation	Calls emitted by the tagged bird or conspecifics; may include individual or group calling events.
On-water activity	Splashing, paddling, water displacement, and intermittent submersion sounds, often linked to bathing, diving, or surface feeding attempts.

containing 8.2 h of 'pure' sounds to train and test the model, constituting 2 % of the entire dataset. This curated subset was considered suitable for supervised training, as it reduced within-class variability and eliminated background interference. To determine the ideal segment length for model training, clips of 1, 5, and 10 s were tested within the framework. An empirical comparison showed that 1-s segments yielded the highest classification accuracy and lowest validation loss, likely due to reduced variability within individual clips, offering the best balance between temporal resolution and model performance. The segmentation process produced a total of 29,584 audio clips, which were organised into four labelled category folders within Google Drive.

The original Colab audio classifier framework was designed to handle datasets with 10 distinct sound categories and approximately 8000 audio samples (Herring, 2018). For the present study, the framework was adapted to accommodate a smaller number of biologically meaningful classes, each supported by a larger and well defined training set. This modification aimed to build a focused seabird sound library that represented the principal behavioural sound categories identified during manual classification, enabling the development of a robust and generalisable model for classifying bird-borne audio. By reducing the number of classes, we intended to improve classification performance and facilitate the future application of the model in ecological and behavioural studies.

2.2. Model development

The development of the automated seabird sound classifier using a CNN architecture involved a multi-step process comprising audio data preparation, pre-processing, model training, and performance evaluation. The CNN model was built in Colab, leveraging components from Colab's *Audio Classifier Tutorial* (Herring, 2018) and *TensorFlow's Simple audio recognition: Recognizing keywords* (TensorFlow, 2020), with

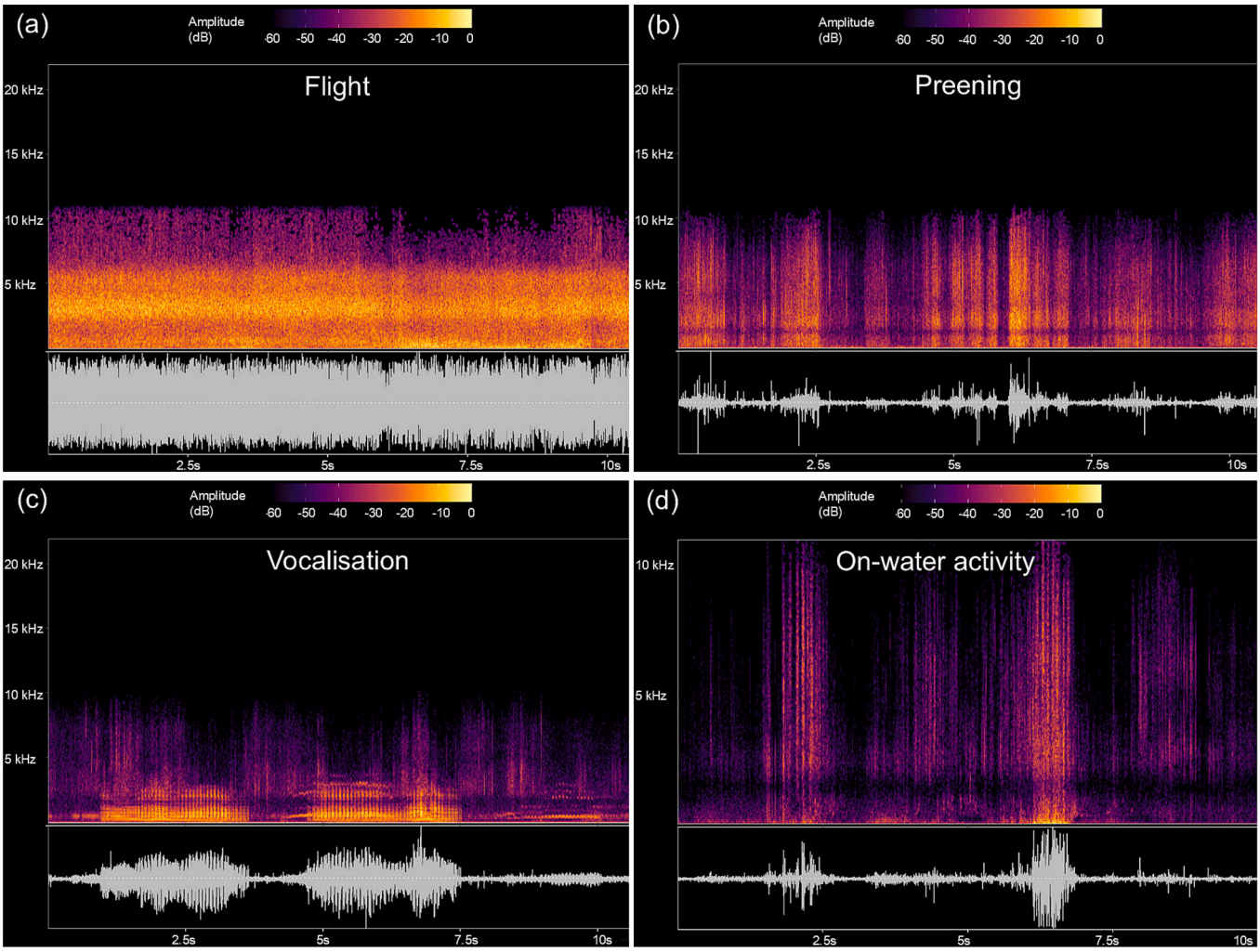


Fig. 1. Examples of the four behavioural sound categories used for model classification: (a) Flight, (b) Preening, (c) Vocalisation, and (d) On-water activity. Each panel shows a 10-s segment recorded from a foraging wandering albatross, with a spectrogram (top) and waveform (bottom). In the spectrograms, the x-axis represents time (seconds), the y-axis represents frequency (kHz), and colour indicates the amplitude (dB) of the sound signal, with brighter colours denoting higher intensity.

necessary modifications to suit the seabird sound dataset (see Supplementary Information).

2.2.1. Step 1 - audio data preparation

Labelled audio clips from the seabird dataset were imported into the Colab environment. The dataset was split into training, validation, and test sets in an 80:10:10 ratio, consistent with standard practices in deep learning for audio classification tasks (Gupta et al., 2021; Sun et al., 2022). This split resulted in 23,668 clips for training, and 2958 each for validation and testing. Given the limited number of sound classes and the need for sufficient samples per class to support robust training, this split provided a sufficient training size to ensure model convergence without overfitting. After splitting, the dataset was carefully curated to exclude overlapping sound categories.

2.2.2. Step 2 - audio data pre-processing

During model training, audio waveforms were transformed dynamically into spectrograms using a short-time Fourier transform (STFT) (Fairbrass et al., 2019). Audio files were resampled to 22,050 Hz to ensure compatibility with TensorFlow/Keras audio processing pipelines, which are optimised for this standard sample rate in deep learning workflows. This slight upsampling (~0.23 %) does not affect the effective spectral resolution (Nyquist ~11 kHz) but ensures consistent frame alignment during spectrogram computation, which is a common best practice in audio classification (Ibrahim, 2024; Velayudham, 2020). Audio files were then segmented into one-second clips, consistent with the optimised chunk length determined during pre-processing. Spectrograms, providing a 2D representation of the audio signal's frequency and amplitude over time, were computed using a 255-sample window, following TensorFlow's STFT convention of odd-length windows for symmetric centring, which also produced near-square spectrograms (171 × 129) optimised for CNN input (TensorFlow, 2020). The magnitude of the STFT was extracted without applying any filtering or log-scaling prior to model input. These parameters were selected to produce near-square image dimensions compatible with CNN input requirements and reflect common practice in audio classification pipelines (Hershey et al., 2017; Piczak, 2015). While an exhaustive parameter search was not conducted, preliminary evaluations confirmed that these settings supported strong model performance and convergence during training. This conversion enabled the model to learn spectral and temporal features essential for accurate sound classification.

2.2.3. Step 3 - model architecture and training

The CNN model was developed using the TensorFlow framework, consisting of 10 sequential layers designed for data pre-processing, feature extraction, and classification (Fig. 2). Full implementation details, including model architecture, data input format, and training pipeline, are openly available in the project's GitHub repository: https://github.com/SeabirdSoundscapes/Seabird-Audio-Classifier.

The initial Keras resizing and normalisation layers downsampled the input data and standardised pixel values, enhancing training speed and accuracy (Lamons et al., 2018; TensorFlow, 2020). The subsequent convolutional layers extracted spatial features from the spectrograms, while max pooling layers reduced data dimensionality, retaining essential information (Lecun et al., 2015). To mitigate overfitting, dropout layers were applied, and flatten layers transformed the multi-dimensional data into a one-dimensional feature vector for final classification (Hinton et al., 2012; Jeong, 2019). Fully-connected dense layers then interpreted these features to generate classification outputs (Rawat and Wang, 2017; Simonyan and Zisserman, 2015). The model was trained over 10 epochs using the Adam optimiser and a categorical cross-entropy loss function, as preliminary tests over 20 epochs showed no consistent improvements in validation accuracy and indicated early signs of overfitting. Backpropagation adjusted the model parameters by calculating gradients of the loss function, enhancing learning efficiency (Rawat and Wang, 2017). The Adam optimiser dynamically adapted the learning rates, which is particularly useful for handling large datasets (Dai et al., 2017). The categorical cross-entropy loss function quantified the difference between predicted and actual data distributions, further refining the model's accuracy (Purwins et al., 2019).

2.2.4. Step 4 - model performance evaluation

Model performance was monitored throughout training, and after each epoch using several metrics for both the training and validation datasets: accuracy (the proportion of correctly classified audio clips), loss (the difference between predicted and true outputs, calculated using the cross-entropy algorithm), validation accuracy (accuracy measured on the validation set), and validation loss (loss computed on the validation set). After training, the model was run on the labelled test dataset to evaluate its performance. A confusion matrix table was produced to show how many audio clips were misclassified (Salamon and Bello, 2017). To assess the model's performance both globally and on each sound class, four metrics were calculated: (1) accuracy (proportion of the number of correct predictions over the number of audio clips analysed); (2) precision (proportion of correct predictions over the total

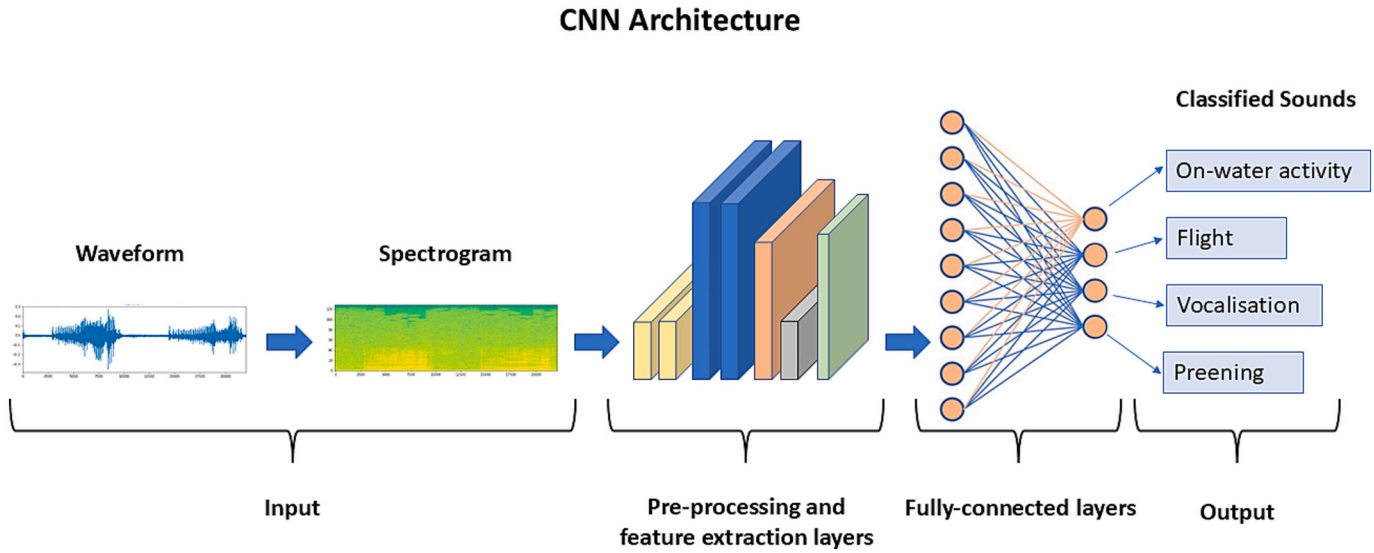


Fig. 2. Automated framework for the seabird sounds classifier.

number of positive predictions); (3) sensitivity or recall (proportion of correct predictions over the number of actual category occurrences); (4) specificity (proportion of actual negative predictions over the number of actual negative occurrences). These metrics were calculated using the following formulae: accuracy = (TP + TN)/(TP + TN + FP + FN), precision = TP/(TP + FP), sensitivity = TP/(TP + FN), specificity = TN / (TN + FP), where TP stands for true positives, TN for true negatives, FP for false positives and FN for false negatives. An additional metric, here defined as misclassification rate, was calculated to determine the proportion of the number of audio clips belonging to one category that were misclassified as another. Misclassification rate = xFN/ TP + xFN, where xFN stands for specific category false negative.

3. Results

The accuracy and loss function metrics were computed on both the training and validation datasets after each epoch. The model accuracy and model loss function were plotted as functions of the epoch number. The resulting training and validation accuracy curves (Fig. 3a) indicate that the model rapidly improved its accuracy during the first few epochs of training. By epoch 9, the training accuracy reached approximately 95 %, while the validation accuracy stabilised at around 93 %. The training loss function steadily decreased over the course of training, while the validation loss initially decreased but started to plateau after epoch 4 (Fig. 3b).

The model correctly classified the withheld test set with global accuracy and precision scores of 95.0 %, sensitivity of 94.6 %, and specificity of 98.2 %. The confusion matrix (Fig. 4) indicates that these metrics varied slightly across sound categories. Flight demonstrated the highest performance among the four categories, with accuracy of 99.7 %, precision of 100.0 %, sensitivity of 99.6 %, and specificity of 100.0 %. In contrast, on-water activity showed the lowest scores, except for sensitivity, with accuracy of 95.2 %, precision of 89.9 %, sensitivity of 96.5 %, and specificity of 94.6 % (Table 2). Preening sounds exhibited the highest misclassification rate, with 11.1 % of sounds misclassified as another category (8.5 % as on-water activity and 2.7 % as vocalisations). In contrast, on-water activity sounds had the lowest misclassification rates, with a total of 3.6 %, including 3.0 % misclassified as vocalisations and 0.5 % as preening (Table 3).

4. Discussion

This work demonstrates that it is possible to train a generic open-source web-based automated audio classification model to identify sounds collected using bird-borne audio recorders that are associated with different seabird behaviours, with high performance scores. The model performance, as shown by the accuracy and loss curves, indicates that the model steadily increased its accuracy scores over the first few epochs, reaching >90.0 % accuracy on both the training and the validation datasets after epoch 4. By epoch 8, training accuracy was 95.0 %, while validation accuracy dropped. This suggests that while the model generalised well to the data, the learning rate started to slow down after epoch 4 and further training would probably not lead to significant improvements in the accuracy metrics (Rawat and Wang, 2017).

Corroborating this conclusion, the training loss scores continued to decrease after epoch 4, whereas the validation loss scores started to plateau and then increased after epoch 8. This was probably due to overfitting after epoch 4, reducing the ability of the model to generalise. Model overfitting and reduced generalisation can be addressed by stopping the training process after a certain number of epochs or when the validation loss stops improving (Brownlee, 2018; Khan et al., 2018). However, regularisation techniques such as L1 (also known as Lasso regression), L2 (ridge regression), or dropout may also improve model performance during training without the need for early stopping (Hinton et al., 2012; Rawat and Wang, 2017; Srivastava et al., 2014). Adjusting the learning rate and fine-tuning of hyperparameters such as the number of layers, number of nodes in each layer, activation functions, optimiser, and loss function can also improve model training efficiency (Alto, 2019; Radhakrishnan, 2017; Simonyan and Zisserman, 2015).

Training convergence occurred rapidly, with validation accuracy stabilising by epoch 8. To evaluate the potential benefits of longer training, the model was also run for 20 epochs. Although training accuracy continued to increase, validation accuracy fluctuated and validation loss showed intermittent peaks after epoch 10, indicating unstable generalisation and early signs of overfitting. Based on these results, limiting training to 10 epochs provided the optimal balance between performance and computational efficiency. Each epoch required approximately 9–10 s on the GPU provided by the Colab environment (NVIDIA Tesla T4 GPU), and the full training process was completed in under two minutes. These results highlight the model's computational efficiency and accessibility, demonstrating that effective training can be achieved using freely available cloud-based resources. This low barrier to entry reinforces the model's value as a practical, open-source tool for researchers working in ecology and conservation.

Future work could investigate how alternative training/validation/test splits, such as a 50:50 configuration, influence model convergence

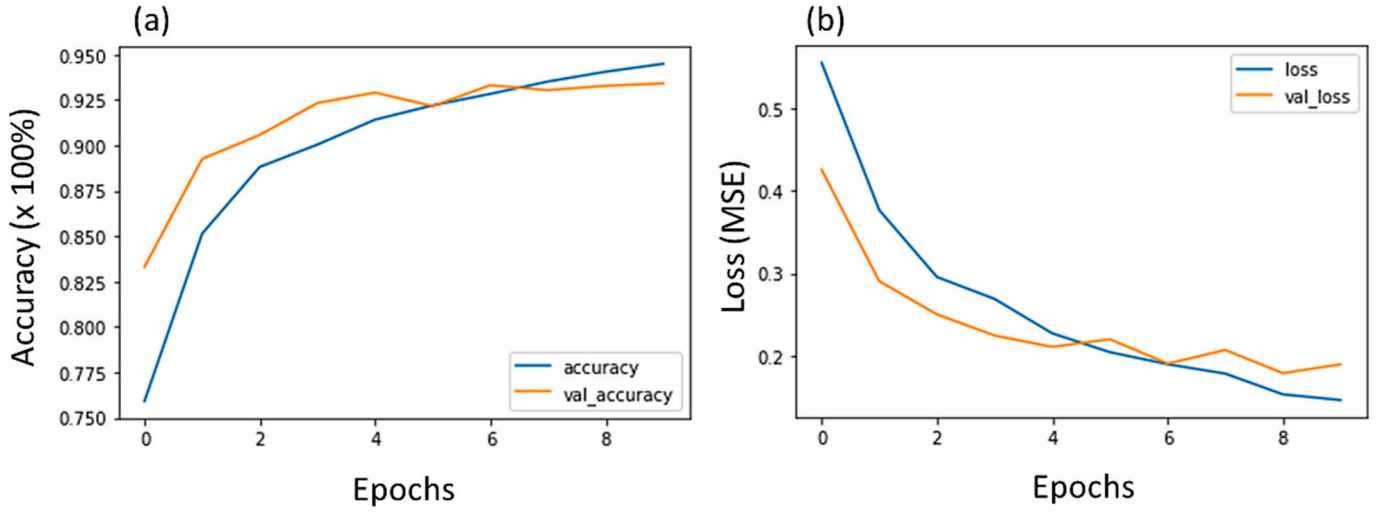


Fig. 3. (a) training and validation accuracy curves and (b) training and validation loss curves as functions of the number of epochs during training of a model to classify behaviour based on audio data from foraging albatrosses. The horizontal axis represents the epochs, indicating the number of complete passes through the training dataset, while the vertical axis shows the accuracy values (in percentage) for the accuracy curves and the loss values (mean squared error) for the loss curves.

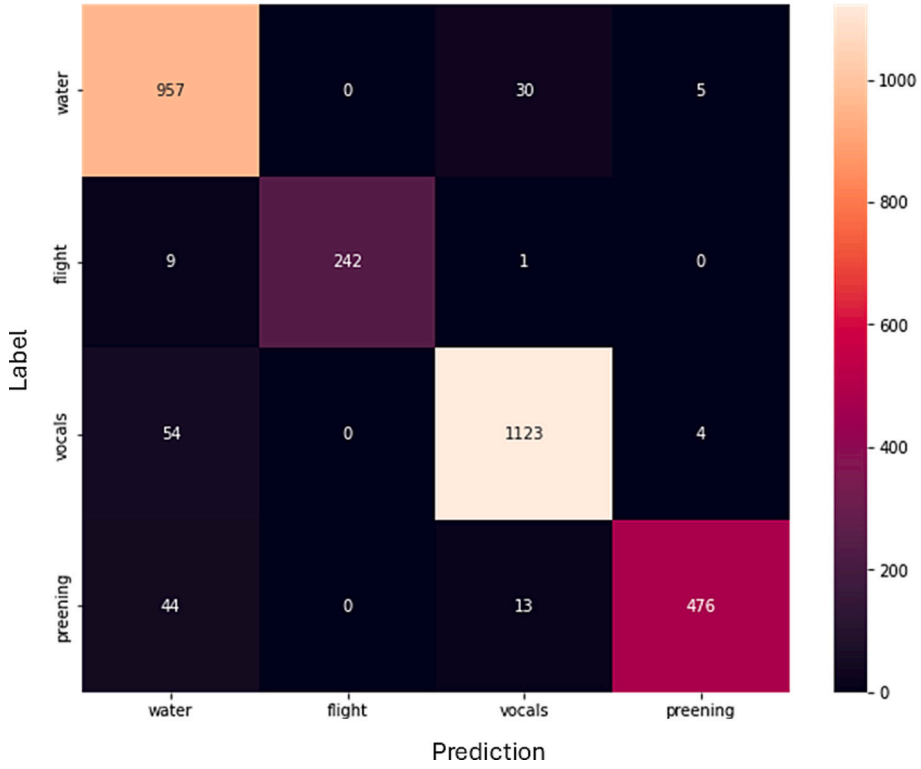


Fig. 4. Performance of an automated seabird sound classifier applied to test data to identify behaviour of foraging albatrosses. The confusion matrix shows the number of correctly classified audio clips (true positives - TP) for each sound category on diagonal, the number of audio clips incorrectly classified as true positives (false positives - FP) per column for each category (excluding the value on diagonal), and the number of audio clips incorrectly classified as true negatives (false negatives - FN) per row for each category (excluding the value on diagonal).

Table 2
Automated seabird sound classifier model performance metrics calculated on the test dataset for each seabird sound category: Accuracy = (TP + TN)/ (TP + TN + FP + FN), Precision = (TP/ (TP + FP)), Sensitivity or Recall = (TP/ (TP + FN)) and Specificity = (TN / (TN + FP)).

Sound category	Accuracy (%)	Precision (%)	Sensitivity/ Recall (%)	Specificity (%)
Flight	99.7	100.0	99.6	100.0
Vocalisation	96.6	96.2	95.1	97.6
Preening	97.8	98.1	89.3	99.6
On-water	95.2	89.9	96.5	94.6

and generalisation. Such comparisons may provide insights into the minimum training data requirements and offer further guidance for researchers applying similar classification models under data-limited conditions.

The model achieved high overall performance during testing, with accuracy and precision scores of 95.0 %, sensitivity of 94.6 %, and specificity of 98.2 %. These metrics indicate that the model effectively captured key features of the audio dataset. However, performance varied across different seabird sound categories (behaviours). Notably, the on-water activity category had the lowest precision (89.9 %) despite an overall high accuracy of 95 %. This suggests that while the model could generally classify on-water activity sounds correctly, it occasionally struggled with precise identification, though misclassifications were relatively infrequent. Misclassification rates varied among classes; for example, on-water activity was occasionally mistaken for vocalisation (3.0 %) and preening (0.5 %). Similarly, flight sounds were sometimes misclassified as on-water activity (3.6 %) or vocalisation (0.4 %), and vocalisation was occasionally confused with on-water activity (4.6 %) and preening (0.4 %). The preening class, in particular, showed the highest misclassification rate, being confused with on-water activity (8.5 %) and vocalisation (2.7 %). These variations highlight the challenges in distinguishing overlapping behaviours and acoustic signals in recordings of seabirds.

These results likely reflect the natural overlap of different behaviours performed by the albatrosses at sea. Seabirds often vocalise while feeding, interacting with conspecifics on the water, or during brief flights and landings (Monier, 2024; Thiebault et al., 2019). Similarly, preening typically occurs at the sea surface, where vocalisations and splashing sounds are also present. Despite efforts to select only “pure” sound segments, some audio labelled as a single category may have contained multiple sound types, contributing to misclassifications.

Table 3
Behaviour misclassification rates by seabird sound category indicating the proportion of audio clips belonging to one category that were misclassified as another category (xFN/ TP + xFN, where xFN).

Sound category	Correct predictions	Misclassification rate (%)	True category misclassified as (%):			
			On-water	Flight	Vocalisation	Preening
On-water	957	3.6	–	0	3.0	0.5
Flight	242	4.0	3.6	–	0.4	0
Vocalisation	1123	4.9	4.6	0	–	0.4
Preening	476	11.1	8.5	0	2.7	–

Further refinement of the dataset or the application of more sophisticated model architectures may be required to improve classification accuracy. Although this study addressed a four-class classification task, the combination of natural background noise, overlapping behaviours, and environmental variability introduced meaningful complexity. Unlike controlled laboratory datasets, these bird-borne field recordings capture authentic ecological conditions, where vocalisations, water interactions, and ambient sounds frequently co-occur. This ecological realism strengthens the study's relevance for applied conservation but also presents inherent challenges for automated classification.

4.1. Automated seabird audio classifier: limitations and opportunities

The dataset used in this study was limited in terms of complexity, but its information resolution could be improved by refining the labels of the samples. The sound category 'flight' encompasses multiple modes of movement, including take-off, landing, gliding and wing beats during sustained flight, and dynamic soaring. Additionally, complementary sensors such as accelerometers and magnetometers can provide valuable context for distinguishing flight modes. For instance, accelerometers are effective at identifying flapping and soaring through kinematic signals, while magnetometers can reveal subtle variations in dynamic soaring patterns, such as heading adjustments and angular velocity around the yaw axis (Conners et al., 2021). Integrating audio data with these sensors could enhance classification accuracy and provide a more comprehensive understanding of albatross flight dynamics.

Vocalisation includes calls emitted by the tagged bird, multiple conspecifics or possibly other albatross species. On-water activity includes sounds caused by diving, splashing, bathing and potential feeding. Preening sounds vary from gentle feather grooming to loud tapping and scratching sounds. Besides the bird-specific sound categories targeted in this study, other aspects of seabird soundscapes hold ecological and behavioural importance which could be explored in future investigations. Environmental (e.g., wind, rain, storm), anthropogenic (e.g., boat engine, off-shore wind farm, human voice) and sounds made by other animals (biophony), provide context to important aspects of seabird life, helping to identify environmental interactions and potential threats (Darby et al., 2024; Dias et al., 2019; Phillips et al., 2016). However, although refining audio data resolution has the potential to increase the range of information that can be extracted from seabird soundscapes, this requires considerable effort to annotate and label the audio files. This drawback can be addressed by combining the use of supervised and unsupervised deep learning models to identify similar features and patterns in unlabelled audio datasets and to group audio samples together in clusters based on specific parameters (Purwins et al., 2019; Sethi et al., 2020).

The seabird audio classifications produced in this study have a temporal resolution of 1 s, making them a valuable dataset for future investigations. While activity budgets can already be determined using immersion data (flights and landings) or accelerometer data (Conners et al., 2021), the addition of acoustic data provides complementary information. This approach not only allows for the calculation of time-activity budgets but also offers insights into the behavioural context of sound production, such as vocalisations or water interactions, which are not captured by immersion or accelerometer data alone. Such detailed understanding is an important step toward exploring the energetics and communication strategies of free-ranging individuals (Thiebault et al., 2021). By matching the audio data classifications with other data streams such as GPS and saltwater immersion it is possible to examine the distribution and rate of occurrence of seabird sounds across different scales of time and space, providing the opportunity to pinpoint hotspots of specific acoustic behaviours. It is also possible to calculate the duration of activity bouts and determine the potential drivers. By examining the relationship between the audio data classifications and environmental data layers, it is possible to predict the effects of environmental factors on seabird sounds distribution and how the investigated relationships may change over time and under different scenarios (Frankish et al., 2020).

The results of this study demonstrate that it is possible to effectively resolve bottlenecks in animal acoustic data classification through the utilisation of advanced open-source machine learning technologies. Such an approach not only enhances the understanding of animal sounds but also extends their applications to the classification of entire soundscapes, thereby providing insights into the surrounding environment. As demonstrated in Sethi et al. (2020), general-purpose audio classifier Convolutional Neural Network (CNN) models can be tailored to identify anomalous events within large datasets over extended periods in an unsupervised manner. This approach allows for the extraction of detailed information about the natural environment based on its soundscape. These advanced models are capable of processing larger and more diverse acoustic datasets, thereby enhancing the potential of automated classifier applications for biodiversity and ecological monitoring across various scales. However, it is worth noting that the audio classifier developed by Sethi et al. was based on the VGG architecture and trained using the TensorFlow framework, an open-source machine learning framework developed by the Google Brain team (Simonyan and Zisserman, 2015). Such powerful open-source deep learning resources hold promise as fundamental technologies for extensive monitoring efforts.

Finally, this automated seabird sound classifier was developed with the benefit of access to a free open-source web-based machine learning research environment, without the need for using a local machine with high computational power or prior specialised technical knowledge on CNNs by the user. Readily accessible tutorials on how to develop CNN models for audio classification in Colab were used extensively during the development of this study, highlighting the great importance that open source and open access information resources have in advancing scientific research.

5. Conclusion

Acoustic monitoring of wildlife using animal-borne instrumentation is a relatively new and as yet under-exploited tool with potentially wide applications in ecology, animal behaviour and conservation research, facilitating access to a wealth of information on biodiversity and the surrounding environment. After data collection, the ability to analyse and classify audio data correctly and efficiently is arguably the most crucial step in the process. Automated audio classification systems such as deep learning and machine learning models have the capacity to process large volumes of audio data quickly and more accurately compared to traditional manual audio analysis and classification methods. This study demonstrated that a generic web-based deep learning CNN model designed to classify audio data can be trained to classify seabird sounds accurately. It also highlighted the importance of access to free open-source information resources. The automatically classified seabird sounds produced in this work can form the basis for a series of subsequent investigations using acoustic-based calculation of time-activity budgets, and the spatial distribution and environment drivers of sounds produced at sea. Furthermore, machine learning and deep learning-based automated audio classifiers serve as tools to explore broader soundscapes across various scales, offering opportunities to uncover diverse aspects of animal lives, including behaviour, interactions with the environment, and exposure to specific threats.

Future work could extend this workflow by integrating automated signal detection to identify and extract candidate audio segments containing seabird sounds. Such an approach would enable full automation of the pipeline, from data curation to classification, using the same open-source machine learning framework described in this study.

Funding sources

This work was supported by the Natural Environment Research

Council (Grant No. NE/L002485/1) to Aline da Silva Cerqueira, as part of the London NERC Doctoral Training Partnership.

CRediT authorship contribution statement

Aline da Silva Cerqueira: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Robin Freeman:** Writing – review & editing, Supervision, Methodology, Formal analysis, Data curation. **Richard A. Phillips:** Writing – review & editing, Resources, Data curation. **Terence P. Dawson:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgements

This research formed part of the PhD thesis of Aline da Silva Cerqueira at the Department of Geography, King's College London, and the Institute of Zoology, Zoological Society of London. We are grateful to all fieldworkers who assisted with device deployment and retrieval at Bird Island. This study contributes to the Ecosystems component of the British Antarctic Survey's Polar Science for a Sustainable Planet Programme, funded by NERC. We thank the journal's Handling Editor and anonymous reviewers for their thoughtful comments and constructive suggestions, which greatly enhanced the clarity, rigour, and overall quality of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2025.103474.

Data availability

The datasets analysed in this study, along with the Colab notebook, are available in the Seabird-Audio-Classifier GitHub repository (https://github.com/SeabirdSoundscapes/Seabird-Audio-Classifier). Interested users can adapt the code and replace the data to use this method on their own seabird acoustic data.

References

Alto, V., 2019. Neural Networks: parameters, hyperparameters and optimization strategies [WWW Document]. Towar. Data Sci. URL https://towardsdatascience.com/neural-networks-parameters-hyperparameters-and-optimization-strategies-3f0842fac0a5 (accessed 1.13.23).

Aodha, O. Mac, Stathopoulos, V., Brostow, G.J., Terry, M., Girolami, M., Jones, K.E., 2014. Putting the scientist in the loop - accelerating scientific Progress with interactive machine learning. Proc. - Int. Conf. Pattern Recognit. 9–17.

Bradbury, J.W., Vehrencamp, S.L., 2011. Principles of Animal Communication, Second Ed. Animal communication, Sunderland, Massachusetts : Sinauer Associates.

Browning, E., Gibb, R., Glover-Kapfer, P., Jones, K.E., Bas, G., Billington, Z., Burivalova, D., Clink, J., De Ridder, J., Halls, T., Hastings, D., Jacoby, A., Kalan, A., Kershenbaum, S., Linke, S., Lucas, R., Machado, P., Owens, C., Sutter, P., Trethowan, R., Whytock, P., 2017. Passive acoustic monitoring in ecology and conservation. WWF Conserv. Technol. Ser. 1, 1–75.

Brownlee, J., 2018. Use Early Stopping to Halt the Training of Neural Networks at the Right Time [WWW Document]. Mach. Learn, Mastery. https://machinelearningmastery.com/how-to-stop-training-deep-neural-networks-at-the-right-time-using-early-stopping/ (accessed 1.23.23).

Clayton, H., Cade, D.E., Burnham, R., Calambokidis, J., Goldbogen, J., 2023. Acoustic behavior of gray whales tagged with biologging devices on foraging grounds. Front. Mar. Sci. 10, 1–9.

Colaboratory, 2025. Audio Classifier Tutorial [WWW Document]. Google. https://colab.research.google.com/github/pytorch/tutorials/blob/gh-pages/_downloads/audio_classifier_tutorial.ipynb#scrollTo=q_9yv3PFt1Mq.

Conners, M.G., Michelot, T., Heywood, E.I., Orben, R.A., Phillips, R.A., Vyssotski, A.L., Shaffer, S.A., Thorne, L.H., 2021. Hidden Markov models identify major movement modes in accelerometer and magnetometer data from four albatross species. Mov. Ecol. 9, 1–16.

Dai, W., Dai, C., Qu, S., Li, J., Das, S., 2017. Very deep convolutional neural networks for raw waveforms. In: In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 421–425.

Darby, J., Phillips, R.A., Weimerskirch, H., Xavier, C., Wakefield, E.D., Pereira, J.M., Patrick, S.C., Xavier, C., Pereira, J.M., 2024. Report strong winds reduce foraging success in albatrosses strong winds reduce foraging success in albatrosses. Curr. Biol. 34, 1–7.

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.

Digby, A., Towsey, M., Bell, B.D., Teal, P.D., 2013. A practical comparison of manual and autonomous methods for acoustic monitoring. Methods Ecol. Evol. 4, 675–683.

Duarte, C.M., Chapuis, L., Collin, S.P., Costa, D.P., Devassy, R.P., Eguiluz, V.M., Erbe, C., Gordon, T.A.C., Halpern, B.S., Harding, H.R., Havlik, M.N., Meekan, M., Merchant, N.D., Miksis-Olds, J.L., Parsons, M., Predragovic, M., Radford, A.N., Radford, C.A., Simpson, S.D., Slabbekoorn, H., Staaterman, E., Van Opzeeland, I.C., Winderen, J., Zhang, X., Juanes, F., 2021. The soundscape of the Anthropocene Ocean. Science 80-.). 371, eaba4658.

Dwivedi, H., 2025. ML model development: How to use Google Colab for deep learning – complete tutorial [WWW document]. Neptune.ai Blog, 29 April. https://neptune.ai/blog/how-to-use-google-colab-for-deep-learning-complete-tutorial (accessed 1.7.25).

Fairbrass, A.J., Firman, M., Williams, C., Brostow, G.J., Titheridge, H., Jones, K.E., 2019. CityNet—Deep learning tools for urban ecoacoustic assessment. Methods Ecol. Evol. 10, 186–197.

Frankish, C.K., Phillips, R.A., Clay, T.A., Somveille, M., Manica, A., 2020. Environmental drivers of movement in a threatened seabird: insights from a mechanistic model and implications for conservation. Divers. Distrib. 26, 1315–1329.

Goëau, H., Glotin, H., Vellinga, W.-P., Planqué, R., Joly, A., Lifeclef Bird, A.J., 2016. LifeCLEF Bird identification task 2016: the arrival of deep learning. CLEF Conf. Labs Eval. Forum 440–449.

Gupta, G., Kshirsagar, M., Zhong, M., Gholami, S., Ferres, J.L., 2021. Comparing recurrent convolutional neural networks for large scale bird species classification. Sci. Rep. 11, 1–12.

Hershey, S., Chaudhuri, S., Ellis, D.P.W., Gemmeke, J.F., Jansen, A., Moore, R.C., Plakal, M., Platt, D., Saurous, R.A., Seybold, B., Slaney, M., Weiss, R.J., Wilson, K., 2017. CNN architectures for large-scale audio classification. ICASSP. IEEE Int. Conf. Acoust. SpeechSignal Process. Proc. 131–135.

Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R., 2012. Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors 1–18.

Ibrahim, M., 2024. An Introduction to Audio Classification with Keras [WWW Document]. Weight, Biases. https://wandb.ai/mostafaibrahim17/ml-articles/reports/An-Introduction-to-Audio-Classification-with-Keras–Vmlldzo0MDQzNDUy.

Herring, W., 2018. Audio Classifier Tutorial [WWW Document]. GitHub: PyTorch Tutorials. https://colab.research.google.com/github/pytorch/tutorials/blob/gh-pages/_downloads/audio_classifier_tutorial.ipynb#scrollTo=GbvbodAFgr_E (accessed 8.22.22).

Jeong, J., 2019. The Most Intuitive and Easiest Guide for Convolutional Neural Network [WWW Document]. Towar. Data Sci. URL https://towardsdatascience.com/the-most-intuitive-and-easiest-guide-for-convolutional-neural-network-3607be47480 (accessed 2.22.23).

Khan, S., Rahmani, H., Shah, S.A.A., Bennamoun, M., 2018. A guide to convolutional neural networks for computer vision. Synth. Lect. Comput. Vis. 8, 1–207.

Laiolo, P., 2010. The emerging significance of bioacoustics in animal species conservation. Biol. Conserv. 143, 1635–1645.

Lamons, M., Kumar, R., Nagaraja, A., 2018. Python Deep Learning Projects: 9 Projects Demystifying Neural Network and Deep Learning Models for Building Intelligent Systems. Packt Publishing, Limited, Birmingham.

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Makris, M.G., Nikolopoulos, L.A.A., Lambropoulos, P., 2002. Courtship behaviour of the wandering albatross Diomedea exulans at Bird Island, South Georgia. Phys. Rev. A - At. Mol. Opt. Phys. 66, 7.

McInnes, A.M., Thiebault, A., Cloete, T., Pichegru, L., Aubin, T., McGeorge, C., Pistorius, P.A., 2020. Social context and prey composition are associated with calling behaviour in a diving seabird. Ibis (Lond. 1859) 162, 1047–1059.

Monier, S.A., 2024. Social interactions and information use by foraging seabirds. Biol. Rev. 99, 1717–1735.

Oswald, J.N., Van Cise, A.M., Dassow, A., Elliott, T., Johnson, M.T., Ravignani, A., Podos, J., 2022. A Collection of Best Practices for the Collection and Analysis of Bioacoustic Data. Appl, Sci, p. 12.

Phillips, R.A., Gales, R., Baker, G.B., Double, M.C., Favero, M., Quintana, F., Tasker, M.L., Weimerskirch, H., Uhart, M., Wolfaardt, A., 2016. The conservation status and priorities for albatrosses and large petrels. Biol. Conserv. 201, 169–183.

Pickering, S.P.C., Berrow, S.D., 2001. Courtship behaviour of the wandering albatross Diomedea exulans at Bird Island, South Georgia. Mar. Ornithol. 29, 29–37.

Piczak, K.J., 2015. Environmental sound classification with convolutional neural networks. In: 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE, pp. 1–6.

Purwins, H., Li, B., Virtanen, T., Schluter, J., Chang, S.-Y., Sainath, T., 2019. Deep learning for audio signal processing. IEEE J. Sel. Top. Signal Process. 13, 206–219.

Radhakrishnan, P., 2017. What Are Hyperparameters ? and how to Tune the Hyperparameters in a Deep Neural Network? [WWW Document]. Towar, Data Sci.

https://towardsdatascience.com/what-are-hyperparameters-and-how-to-tune-the-hyperparameters-in-a-deep-neural-network-d0604917584a.
Rawat, W., Wang, Z., 2017. Deep convolutional neural networks for image classification: a comprehensive review. Neural Comput. 29, 2352–2449.
Salamon, J., Bello, J.P., 2017. Deep convolutional neural networks and data augmentation for environmental sound classification. IEEE Signal Process. Lett. 24, 279–283.
Sethi, S.S., Jones, N.S., Fulcher, B.D., Picinali, L., Clink, D.J., Klinck, H., Orme, C.D.L., Wrege, P.H., Ewers, R.M., 2020. Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set. Proc. Natl. Acad. Sci. USA 117, 17049–17055.
Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. In: 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc, pp. 1–14.
Snaddon, J., Petrokofsky, G., Jepson, P., Willis, K.J., 2013. Biodiversity technologies: tools as change agents. Biol. Lett. 9, 3–5.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.
Stowell, D., Benetos, E., Gill, L.F., 2017. On-Bird sound recordings: automatic acoustic recognition of activities and contexts. IEEE/ACM Trans. Audio Speech Lang. Process. 25, 1193–1206.
Stowell, D., Wood, M.D., Pamuła, H., Stylianou, Y., Glotin, H., 2019. Automatic acoustic detection of birds through deep learning: the first Bird audio detection challenge. Methods Ecol. Evol. 10, 368–380.
Sun, Y., Midori Maeda, T., Solís-Lemus, C., Pimentel-Alarcón, D., Buřivalová, Z., 2022. Classification of animal sounds in a hyperdiverse rainforest using convolutional neural networks with data augmentation. Ecol. Indic. 145, 0–9.
Swiston, K.A., Mennill, D.J., 2009. Comparison of manual and automated methods for identifying target sounds in audio recordings of pileated, pale-billed, and putative ivory-billed woodpeckers. J. F. Ornithol. 80, 42–50.
TensorFlow, 2020. Simple audio recognition: Recognizing keywords. TensorFlow Core Tutorials, 14 October. Available at: https://www.tensorflow.org/tutorials/audio/simple_audio (accessed 22 August 2022).
Thiebault, A., Charrier, I., Aubin, T., Green, D.B., Pistorius, P.A., 2019a. First evidence of underwater vocalisations in hunting penguins. PeerJ 2019, 1–16.
Thiebault, A., Charrier, I., Pistorius, P., Aubin, T., 2019b. At sea vocal repertoire of a foraging seabird. J. Avian Biol. 50, 1–14.
Thiebault, A., Huetz, C., Pistorius, P., Aubin, T., Charrier, I., 2021. Animal-borne acoustic data alone can provide high accuracy classification of activity budgets. Anim. Biotelemetry 9, 28.
Tosa, M.I., Dziedzic, E.H., Appel, C.L., Urbina, J., Massey, A., Ruprecht, J., Eriksson, C.E., Dolliver, J.E., Lesmeister, D.B., Betts, M.G., Peres, C.A., Levi, T., 2021. The rapid rise of next-generation natural history. Front. Ecol. Evol. 9, 1–18.
Towsey, M., Wimmer, J., Williamson, I., Roe, P., 2014. The use of acoustic indices to determine avian species richness in audio-recordings of the environment. Ecol. Inform. 21, 110–119.
Velayudham, V., 2020. Audio Data Processing— Feature Extraction — Essential Science & Concepts behind them — Part I [WWW Document. Anal. Vidhya]. In: https://medium.com/analytics-vidhya/audio-data-processing-feature-extraction-science-concepts-behind-them-be97fbd587d8.
Wijers, M., Trethowan, P., Markham, A., du Preez, B., Chamaillé-Jammes, S., Loveridge, A., Macdonald, D., 2018. Listening to lions: animal-borne acoustic sensors improve bio-logger calibration and behaviour classification performance. Front. Ecol. Evol. 6, 1–8.
Xie, J., Ding, C., Li, W., Cai, C., 2018. Audio-Only Bird Species Automated Identification Method with Limited Training Data Based on Multi-Channel Deep Convolutional Neural Networks 1.
Yalçın, O.G., 2020. 4 Reasons why you Should Use Google Colab for your Next Project [WWW Document]. Towar. Data Sci. https://towardsdatascience.com/4-reasons-why-you-should-use-google-colab-for-your-next-project-b0c4aaad39ed.