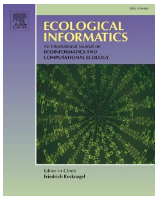
Contents lists available at ScienceDirect

Ecological Informatics

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



Automatic detection and taxonomic identification of dolphin vocalisations using convolutional neural networks for passive acoustic monitoring

Guilherme Frainer[a,b,*], Emmanuel Dufourq[c,d,e], Jack Fearey[a,b], Sasha Dines[b,f], Rachel Probert[b,g], Simon Elwen[b,f], Tess Gridley[b,f]

[a] *Centre for Statistics in Ecology, Environment and Conservation, University of Cape Town, South Africa*
[b] *Sea Search Research and Conservation, South Africa*
[c] *African Institute for Mathematical Sciences, South Africa*
[d] *Department of Mathematical Sciences, Stellenbosch University, South Africa*
[e] *National Institute for Theoretical and Computational Sciences, South Africa*
[f] *Department of Botany and Zoology, Stellenbosch University, South Africa*
[g] *School of Life Sciences, University of KwaZulu-Natal, South Africa*

ARTICLE INFO

Keywords:
Convolutional neural networks
Indian Ocean humpback dolphin
Machine learning
Passive acoustic monitoring
Sound detection
Species identification

ABSTRACT

A novel framework for acoustic detection and species identification is proposed to aid passive acoustic monitoring studies on the endangered Indian Ocean humpback dolphin (*Sousa plumbea*) in South African waters. Convolutional Neural Networks (CNNs) were used for both detection and identification of dolphin vocalisations tasks, and performance was evaluated using custom and pre-trained architectures (transfer learning). In total, 723 min of acoustic data were annotated for the presence of whistles, burst pulses and echolocation clicks produced by *Delphinus delphis* (~45.6%), *Tursiops aduncus* (~39%), *Sousa plumbea* (~14.4%), *Orcinus orca* (~1%). The best performing models for detecting dolphin presence and species identification used segments (spectral windows) of two second lengths and were trained using images with 70 and 90 dpi, respectively. The best detection model was built using a customised architecture and achieved an accuracy of 84.4% for all dolphin vocalisations on the test set, and 89.5% for vocalisations with a high signal to noise ratio. The best identification model was also built using the customised architecture and correctly identified *S. plumbea* (96.9%), *T. aduncus* (100%), and *D. delphis* (78%) encounters in the testing dataset. The developed framework was designed based on the knowledge of complex dolphin sounds and it may assists in finding suitable CNN hyper-parameters for other species or populations. Our study contributes towards the development of an open-source tool to assist long-term studies of endangered species, living in highly diverse habitats, using passive acoustic monitoring.

1. Introduction

Accurate remote sensing tools used to investigate wildlife populations are critical for long-term monitoring and effective conservation actions, especially for endangered species. Passive acoustic monitoring (PAM) has been extensively used to investigate endangered dolphin populations (Dong et al., 2017; Jaramillo-Legorreta et al., 2017; Munger et al., 2016), and machine learning techniques have been employed to improve the accuracy and speed of acoustic detection (Bergler et al., 2022; Caruso et al., 2020; White et al., 2022; Ziegenhorn et al., 2022). Despite the widespread use of PAM, relatively few tools are available that detect and identify these sounds in archived recordings (Bergler et al., 2022; Gillespie et al., 2009; Sugai et al., 2019). Additionally, the lack of annotated sounds in openly available datasets precludes further development of complex machine learning models for dolphin sounds detection and species identification as a large amount of data are needed (Jordan and Mitchell, 2015). Conservation actions and population monitoring using PAM are thereby limited for some species, particularly those living in noisy habitats where sympatric species emit similar acoustic signals. Effective classifiers are required to identify species of interest in highly complex ecosystems (Ziegenhorn et al., 2022).

The development of effective tools using PAM techniques, designed for the monitoring and conservation of the endangered Indian Ocean humpback dolphin (*Sousa plumbea*) in South Africa, was the catalyst for

[*] Corresponding author at: Centre for Statistics in Ecology, Environment and Conservation, University of Cape Town, South Africa.
E-mail address: gui.frainer@gmail.com (G. Frainer).

https://doi.org/10.1016/j.ecoinf.2023.102291
Received 24 June 2023; Received in revised form 25 August 2023; Accepted 5 September 2023
Available online 12 September 2023

this study. Toothed whales rely on acoustic communication for biological success, using a variety of functionally specific signals, such as tonal whistles or broadband pulse bursts in social interactions, as well as echolocation clicks for navigation and feeding. Humpback dolphins in South Africa inhabit shallow rocky and sandy-bottom shore zones of a very heterogeneous habitat along the southwesterly portion of this species' distribution (Best and Folkens, 2007), which could negatively affect the detectability of specific sounds due to the noisy environment (Shabangu et al., 2022). Additionally, the use of coastal habitats increases their interaction with human activities (Plön et al., 2015), such as boat traffic (Karczmarski et al., 1998), which not only contributes to the soundscape as noise (Schoeman et al., 2022), but can also mask the sounds produced by dolphins and interfere on both natural communication (Fouda et al., 2018; Jensen et al., 2009) and monitoring of wild populations. Despite the potential significance of passive acoustics in monitoring humpback dolphins (*Sousa* spp.) (Bopardikar et al., 2018; Dong et al., 2017; Yang et al., 2020), its application in long-term recordings is still constrained in South African waters, as there are no available automated classifiers to differentiate their sounds from other dolphin species that are present in the area. Humpback dolphins from the southern Indian Ocean have an overlapping distribution with at least three other whistling dolphin species, most commonly, the Indo-Pacific bottlenose dolphin (*Tursiops aduncus*), the common dolphin (*Delphinus delphis*), and the killer whale (*Orcinus orca*) (Peddemors, 1999).

The highly diverse vocal repertoire of delphinids (Odontoceti: Delphinidae) reflects the complexity of their cognitive abilities due to a strong social component (Fox et al., 2017). Despite this, their vocal production structures share similar morphological adaptations (Mead, 1975). However, slight variations in size (Jensen et al., 2018) and head shape of some species (e.g., *S. plumbea*) (Frainer et al., 2021; Song et al., 2022) may result in convergence on similar sound production capabilities with other species (e.g., *T. aduncus* and *D. delphis*) and potentially affect the accuracy of identification tasks (Yang et al., 2020). Humpback dolphins exhibit adaptations on the left side of their epicranial complex that may allow them to produce more directional and higher frequency communication sounds compared to bottlenose dolphins (*Tursiops* spp.) (Frainer et al., 2019). Such sounds, for example whistles, overlap in spectral frequency with those produced by *T. aduncus* and *D. delphis* (Erbs et al., 2017; Fearey et al., 2019; Gridley et al., 2014). Although most of the studies on this topic have investigated the differences across species using specific calls such as whistles (Erbs et al., 2017; Oswald et al., 2008; Oswald et al., 2021) or clicks (Buchanan et al., 2021; de Freitas et al., 2015; Luo et al., 2019; Temple et al., 2016; Yang et al., 2020), few studies have integrated multiple sound types as input for species classification tasks (Rankin et al., 2017).

In this study, we assessed the applicability of Convolutional Neural Networks (CNNs) for dolphin monitoring in long-term recordings using their complete vocal repertoire along with a model prediction and post-processing approach for automated taxonomic identification. Although prior studies have shown the effectiveness of CNNs in detecting and identifying whale (Allen et al., 2021) and dolphin sounds (Buchanan et al., 2021; Duan et al., 2022; Erbs et al., 2023; Luo et al., 2019; Nur Korkmaz et al., 2023), a multi-class classifier that encompasses all the dolphin species occurring in South African waters has yet to be developed. The proposed framework designed here combining biological knowledge on sound production in dolphins, and innovative machine learning tools, may enhance the use of PAM for target species in highly diverse areas. Improving remote sensing techniques to monitor the population dynamics of the endangered humpback dolphin in South Africa via their vocalisations (Longden et al., 2020; Wang et al., 2020) would be a critical stride towards the development of an automated and long-term monitoring system and effective conservation management strategies. This represents the first tool of this nature for South Africa and will be available for ecologists, management teams, and researchers.

2. Material and methods

2.1. Data collection

To build the training library, boat-based focal follow recordings were used to record the vocalisations of four whistling coastal dolphin species that inhabit South African waters (Fig. 1). The recordings were made using a SoundTrap 300HF (flat frequency response of 20 Hz - 150 kHz ± 3 dB; Ocean Instruments Inc., New Zealand), or HTI-96-MIN hydrophones (flat frequency response of 2 Hz – 30 kHz ± 1 dB; High Tech Inc., U.S.) attached to a TASCAM DR 680 recorder (TASCAM, U.S.) (Supporting Information A) and were stored in .wav files. Hydrophones were set approximately four metres deep during dolphin encounters, with signals digitised at 96 kHz sample rate or higher in continuous recordings. Dedicated visual surveys were performed during all boat-based recordings to ensure that no other species were present in the area, i.e., data from mixed-species groups were not included in the analysis. Additionally, recordings made through moored instruments in Mossel Bay were obtained between March 19th and April 4th of 2021, using a SoundTrap 300HF sampling at 96 kHz at five meters depth (Fig. 1). The presence of *S. plumbea* and *T. aduncus* in the vicinity of the devices was confirmed by land-based observations from the harbour wall, located approximately 100 m away from the mooring. The close proximity of the dolphins to the recorder, combined with the simultaneous capture of strong signals by the devices during visual observations, confirmed the correlation between sound and species identification. Two confirmed *D. delphis* encounters between the 15th and 16th of May 2021, in False Bay, were recorded using a SoundTrap 300HF hydrophone attached four metres deep to a free-drifting buoy. Furthermore, to validate the single-species encounters, visual observations were conducted from a boat positioned roughly 400 m away from the drifting buoy. A moored SoundTrap 300HF sampling at 576 kHz at ~10 m depth was deployed between the 31st of January and the 2nd of February 2021, in Fish Hoek, Cape Town to record *O. orca* sounds during four days of a confirmed sighting in the area (i.e., reports from whale watching networks and personal observation) (Fig. 1). In this case, a male *O. orca* was sighted during consecutive days close to the moored hydrophone, through visual observations from a boat. The unique complex calls from *O. orca* (Miller and Bain, 2000) confirmed the species identification of the vocalisations. The moorings used in this study were attached to a rope that was suspended, along the water column, by a subsurface buoy. The moored hydrophones were then attached approximately two meters from the bottom, and all the moored and free-drifting recordings were made in continuous recordings (Supporting Information A).

2.2. Training dataset and testing dataset

Dolphin whistles, burst pulses, and echolocation click trains were inspected aurally and visually, using spectrograms (FFT length = 1300; hop size = 650; Hann window; with smoothing applied), and manually annotated using Raven Pro 1.6 (Cornell Lab of Ornithology, 2023). The labelled dolphin vocalisations varied from short whistles and burst pulses to long segments with more than one vocally active animal, including big groups (>100 animals, e.g., *D. delphis* and *T. aduncus*) (Fig. 2). Soundscapes, comprised of non-dolphin biological (e.g., fish, snapping shrimp, reef), geophonic (e.g., rough seas, rain), and anthropic sounds (e.g., chain noise, boats) were also manually annotated (Dufourq et al., 2021; Stowell et al., 2019) to represent the naturally occurring soundscape in the absence of dolphins. The start and end of each annotation were recorded, as well as the duration of each segment. For the testing dataset, vocalisations were categorised according to the amount of noise masking, interpolated from the signal-to-noise ratio graded from one (i.e., masked/weak signal) to three (i.e., strong and clear signal). The visually monitored recordings from moored hydrophones in Mossel Bay were used as ‘unseen data’ to test the

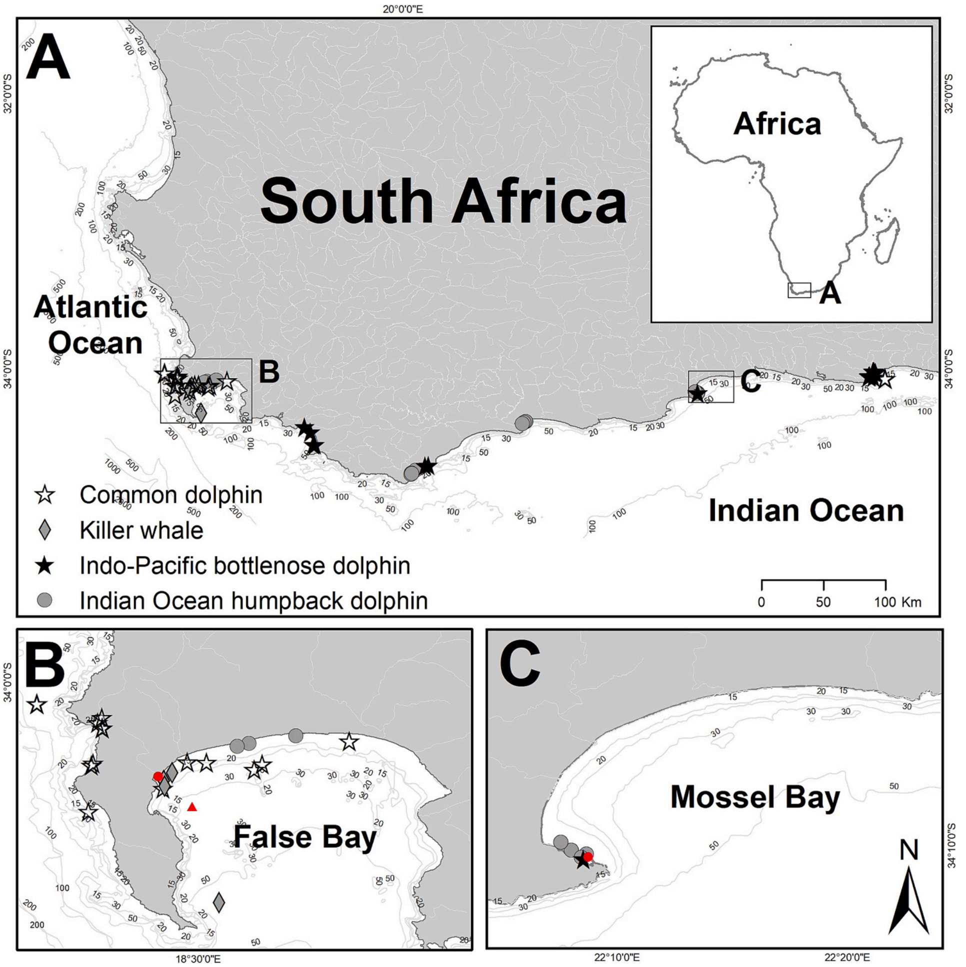


Fig. 1. Locations of the boat-based recordings of the common dolphin (*Delphinus delphis*), the killer whale (*Orcinus orca*), the Indo-Pacific bottlenose dolphin (*Tursiops aduncus*), and the Indian Ocean humpback dolphin (*Sousa plumbea*) used to build the training dataset. Recordings from moored (circles) and drifting buoy-attached (triangle) hydrophones used as the testing dataset are represented in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

generalisability of our tool. Similarly, *D. delphis* recordings from a free-drifting buoy, as well as *O. orca* sounds from the moored hydrophone, were only used to test the species identification model (Supporting Information A).

2.3. Pre-processing

To ensure consistent sampling rates, audio recordings with a sampling rate above 96 kHz were downsampled to 96 kHz. To create the training set, a sliding window approach was used to extract segments of sound with equal length (user defined hyper-parameter) from the annotated events (Dufourq et al., 2021), in which segments were sampled in series based on their start and end times. The segment start times were interspaced one second apart from each other to sample dolphin vocalisations in different contexts. We compared the accuracy of the models by varying the windows sizes (2, 3, 5, and 7 s) to determine the best parameter. These window sizes refer to the shortest segment possible (i.e., two seconds) and the longest segment that can cover at least the longest dolphin vocalization (e.g., *O. orca* complex calls). All segments were augmented by randomly mixing dolphin sounds with target soundscapes from where the classifier would be applied; in our case, Mossel Bay. The new segments contained a proportion of both dolphin (90%) and soundscape (10%) sounds; to elucidate a potential detection of species in the target area. The amount of augmentation for species was scaled up relative to the number of segments generated for the species with the largest amount of data, which was only duplicated due to the large number of clips generated (i.e., *D. delphis*, with 20,319 clips generated and 40,638 spectrograms created). We also balanced each species dataset per encounter to ensure equal distribution for the sounds produced in different contexts (see *Discussion* section). The class distribution was also balanced after the augmentation process, based on the class with the smallest dataset to ensure balanced datasets.

To test the efficacy of our models, we created several segments by using the same sliding window approach. Namely, we used the same window size that was used in training, and thus multiple segments were created across the entire testing file by moving the window by one second in the moored recording. We converted each of these testing segments into spectrograms (FFT length = 1024; hop size = 128; Hann window) which were used as input for subsequent model prediction. All generated spectrogram images were created as 5 × 5 in. but varied in their dpi configuration, ranging from 200 × 200 (40 dpi) to 500 × 500 (100 dpi) samples. The number of images used per class was constrained by our computational resources, and we used the maximum number of images possible in each case. We attempted a number of experiments and varied the number of classes. The largest dataset built comprised 80,000 images when combining three seconds window size and 40 dpi for the customised architecture (see *Convolutional neural networks* section), and the smallest one comprised 3900 images combining two seconds window size and 90 dpi for the transfer learning approach (Table 1).

2.4. Convolutional neural networks

Two CNN models were implemented to detect and identify dolphin sounds (Fig. 3). The first model (CNN1) was a binary classifier that was trained to detect the presence or absence of dolphin sounds. The second model (CNN2) was a multi-class classifier that was trained to differentiate between different species of dolphins. Two architectures were

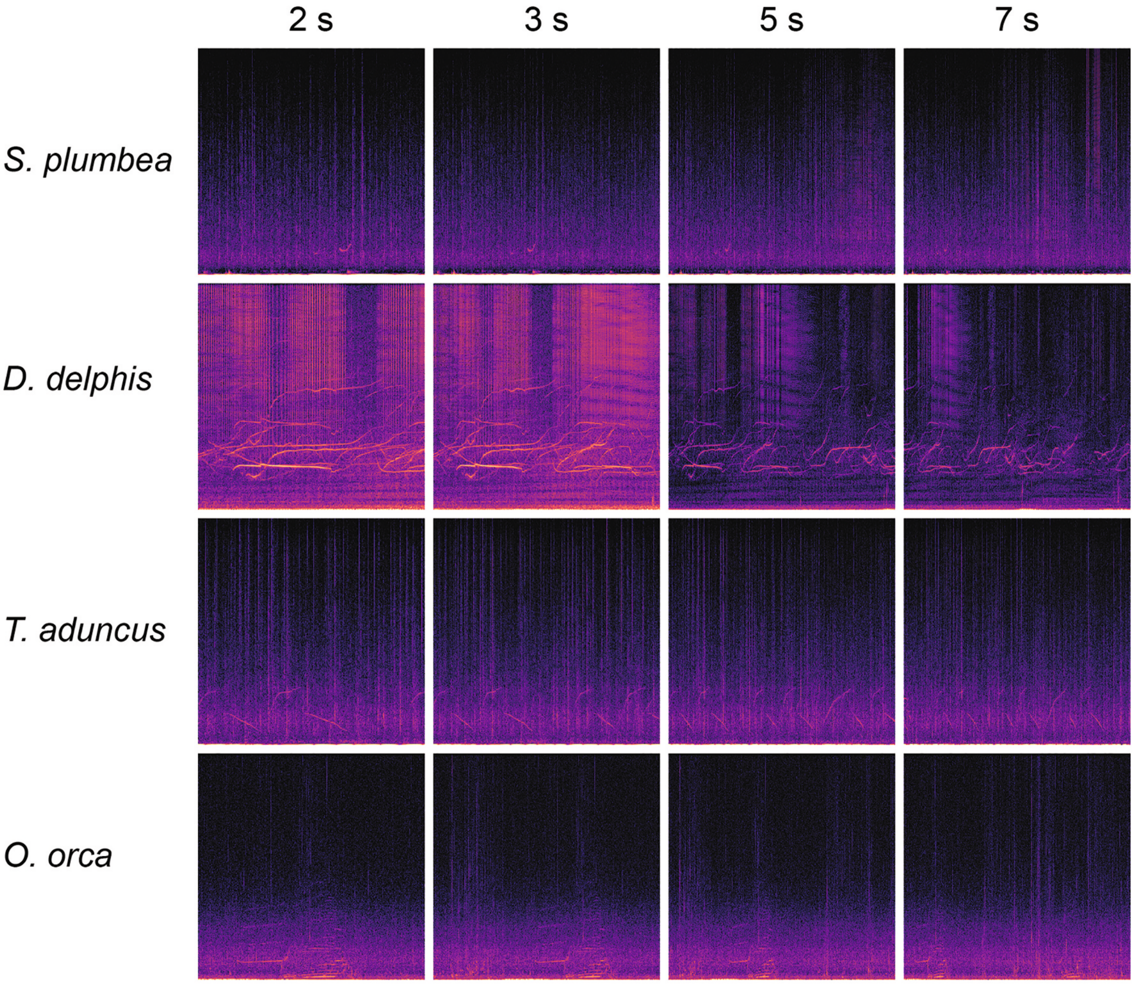


Fig. 2. Examples of spectrograms showing calls of all four species studied here built with distinct window sizes (two-, three-, five- and seven-seconds length). Sample rate 96 kHz (Nyquist frequency 48 kHz), Hann window size of 1024 samples, and a hop size of 128 samples (75% overlap).

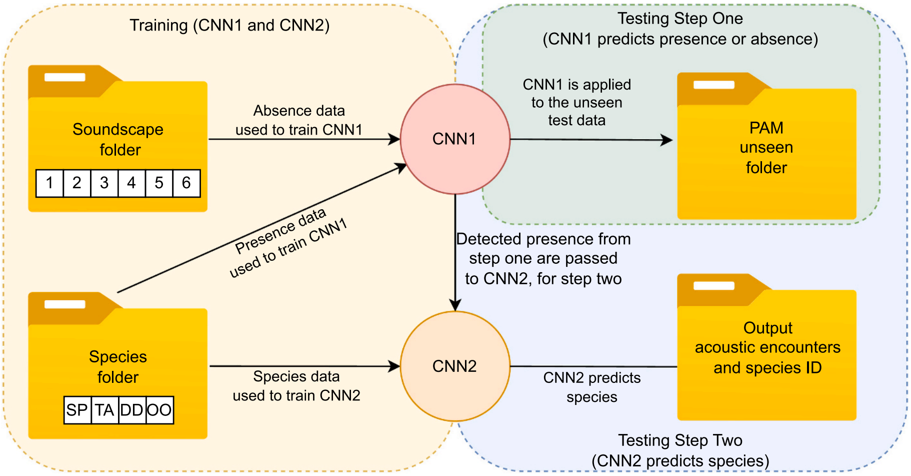

Training (CNN1 and CNN2)
Soundscape folder
1 2 3 4 5 6
Absence data used to train CNN1
Presence data used to train CNN1
Species folder
SP TA DD OO
Species data used to train CNN2
CNN1
CNN2
Testing Step One (CNN1 predicts presence or absence)
CNN1 is applied to the unseen test data
PAM unseen folder
Detected presence from step one are passed to CNN2, for step two
CNN2 predicts species
Output acoustic encounters and species ID
Testing Step Two (CNN2 predicts species)

Fig. 3. The general pipeline of the algorithm used to build (Training) and test (Testing) the models.

compared, namely, a customised CNN (based on preliminary hyperparameter tuning experiments), and a pre-trained ResNet152V2 architecture (He et al., 2016) that demonstrated good performance in animal sound classification tasks (Dufourq et al., 2022). The customised models were composed of three convolutional layers (32 filters, kernel size of 4 × 4, ReLU activation). Each convolutional layer was followed by dropout (rate of 0.4) and a max pooling (kernel size of 4 × 4) layer. This was followed by a fully connected layer with 64 ReLU units, dropout (rate of 0.4), and a softmax function (two units in the case of CNN1, and three or four units in the case of CNN2 depending on the number of species). The models were trained for 50 epochs using the Adam optimizer (Kingma and Ba, 2014), with a learning rate of 0.001 and a batch

Table 1
Dolphin detection models performance based on the comparison of the time assigned for the acoustic encounters and the ground-truth ($n = 18$ encounters, see Fig. 4 and Supporting Information B / "Post_processing_Human_detector.ipynb" file). Each row represents a combination of model architecture and the configurations used to build the image dataset for the training step such as window size and dpi. Spec, specificity; Sens, sensitivity; Prec, precision; Accu, accuracy. We also provide the total number of trainable network parameters. *Time to predict 10 min recording.

	Window size (s)	dpi	Library size	SNR ≥1					SNR ≥2					Model Param.	Prediction time* (min: sec)
				Spec (%)	Sens (%)	Prec (%)	Accu (%)	F1 score (%)	Spec (%)	Sens (%)	Prec (%)	Accu (%)	F1 score (%)		
Customised architecture	2	70	24 k	**96.4**	**56.7**	**87.6**	**84.4**	**68.9**	**92.3**	**76.3**	**67.9**	**89.5**	**71.9**	66,338	02:26.11
		80	24 k	99.3	40.7	96.2	81.5	57.3	94.7	48.6	66.1	86.6	56.0	84,900	02:29.12
		90	12 k	99.7	27.1	98.2	77.6	42.5	98.0	38.6	80.7	87.6	52.3	107,298	02:37.09
		100	20 k	95.2	44.5	80.4	79.8	57.3	90.7	52.6	54.8	84.0	53.7	107,493	02:34.64
	3	40	80 k	94.9	39.7	77.4	78.1	52.5	92.2	52.5	59.0	85.2	55.6	41,762	03:12.48
	7	40	80 k	96.3	14.1	62.6	71.3	23.0	96.2	21.4	54.7	83.1	30.8	41,762	06:49.82
Transfer Learning	2	90	3.9 k	94.1	24.9	65.1	73.1	36.0	93.8	37.2	56.1	83.9	44.8	921,602	06:14.78
	3	40	26 k	90.8	35.4	62.9	74.0	45.3	86.7	50.4	44.7	80.3	47.4	200,706	05:06.04

size of 32. The most suitable architecture was chosen based on the best validation accuracy (proportion of all correct predictions) and precision (number of true positives divided by true positives and false positives) obtained during training. The model training and prediction procedures were executed on Microsoft Azure using instance *NV12s v3* with 12 vCPUs and 112 GB RAM. The CNNs were implemented using TensorFlow (Abadi et al., 2016) and Python 3. The Ubuntu 20.04 operating system was used and obtained via the Ubuntu 20.04 Data Science Virtual Machine on Microsoft Azure. The algorithm scripts are available in Supporting Information B.

2.5. Inference and post-processing

CNN1 was applied to the unseen data to obtain softmax values indicating the likelihood of dolphin vocalisations within each testing segment. A post-processing technique was devised to group segments that were predicted as present and occurred within a 900 s timeframe of each other, and for which the model displayed a high degree of confidence (> 70%). The outcome of CNN1 determined the start and end times for each acoustic encounter (AE), which entails isolated calls occurring within at least 15 min of each other. The time between AE was determined based on ad hoc experimentation and can be easily adjusted during the inference step. Each AE was then assessed using CNN2 to assign a single species identification for all detected segments containing dolphin vocalisations. The taxonomic identification for an AE was determined by first using CNN2 to determine the species indications on each detected segment within the AE, and then the majority of taxonomic identification was assigned to the entire AE. The number of detections and the proportion of detections per species, as well as the start and end times (based on the files' name), and duration of the AE, are given in the output (see output example in Supporting Information B).

2.6. Model evaluation

The testing dataset was analysed by one experienced observer (GF) whereby dolphin echolocation clicks, burst pulses, and whistles were also manually annotated using Raven Pro 1.6 (Cornell Lab of Ornithology, 2023). A confusion matrix was then generated to compare the detected AEs by the CNN1 models against the ground-truth data, based on the time of correct/incorrect assignment (see Fig. 4). In this way, each second of the 24 h testing dataset was categorised as True Negative (TN), True Positive (TP), False Negative (FN), or False Positive (FP). The evaluation was performed for all dolphin sounds and, secondly, for all dolphins sounds with SNR higher than 1, which are considered useful for ecological studies (Gridley et al., 2015; Palmer et al., 2019). The models were assessed based on the accuracy, precision, sensitivity (recall), specificity, and F1 score:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN)$$

$$Precision = TP/(TP + FP)$$

$$Sensitivity = TP/(FN + TP)$$

$$Specificity = TN/(TN + FP)$$

$$F1\ score = 2*(Precision*Sensitivity)/(Precision + Sensitivity)$$

The performance of the species identification model (CNN2) was tested using moored or drifting recordings (see *Data collection* section) with verified species identification, and the accuracy for each species was reported.

To create the dataset, audio segments were extracted and augmented using the Microsoft Azure instance *E96ias v4* with 96 vCPUs and 672 GB RAM. We chose a high performance machine aiming to execute the algorithm with as much data as possible instead of sub-sampling the dataset. The software was implemented using various Python 3

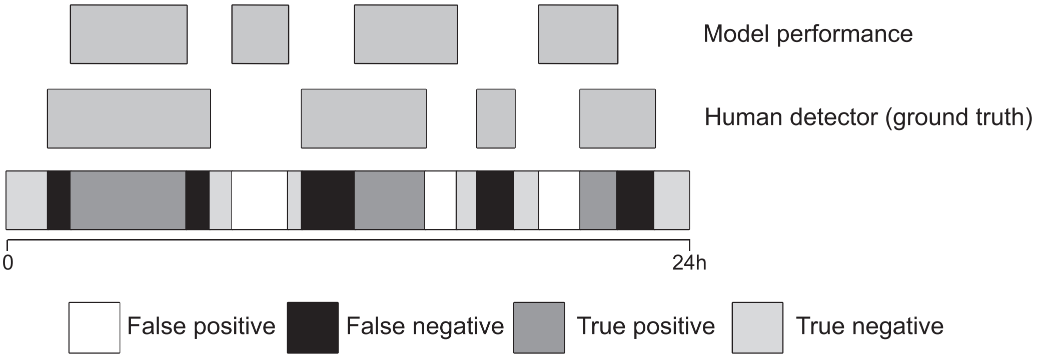


Fig. 4. Detection model evaluation based on acoustic encounters (AEs). The confusion matrix was built based on resultant AEs assigned by the model compared to manually annotated data (human detector/ground truth).

packages, including Librosa version 0.8.1 (McFee et al., 2015) and SciPy (Virtanen et al., 2020).

3. Results

The training dataset was based on 43 boat-based encounters (*D. delphis, n* = 8 encounters; *O. orca, n* = 4 encounters; *S. plumbea*, *n* = 19 encounters; *T. aduncus*, *n* = 12 encounters) and soundscape recordings from moored hydrophones (Fig. 1). Annotated sounds used to create the training dataset totalled 723 min of audio data for which the distribution was *D. delphis* (45.6%), *O. orca* (0.96%), *S. plumbea* (14.38%), *T. aduncus* (39%), as well as 772 min of the soundscape. The training library size varied based on the computing limitations (Table 1). The testing dataset for the detection model comprised 24 h of a day and contained 18 AE varying from less than one second to ~59 min. The testing dataset for the species identification model was based on 10 to 30 min of unseen data for each of the species studied here (Supplementary Information I). The varying length of the testing dataset was due to the number of vocalisations detected in the unseen data by the CNN1, which is potentially affected by the setup of the hydrophone deployment (moored or drifting buoy) and the behavioural biology of each species (see *Discussion* section). Except for *D. delphis*, in which we have mostly used 10 min of testing data due to the higher number of detections in those recordings, all other testing files listed in Supplementary Information A per species were used to evaluate the identification models. The best model weights for CNN1 (detection) and CNN2 (species identification) were obtained using two-second segments (windows) with images generated at 70 and 90 dpi, respectively (Fig. 5).

The customised CNN architecture achieved the highest accuracy for both models, outperforming the pre-trained ResNet152V2 model with faster predictions. The best CNN1 model exhibited an 84.4% accuracy (Precision = 87.6%, Sensitivity = 56.7%, Specificity =96.4%) in defining AEs based on all dolphin sounds in the test set and 89.5% accuracy (Precision = 67.9%, Sensitivity = 76.3%, Specificity = 92.3%) for sounds with an SNR higher than 1. On the other hand, the best ResNet152V2 model (using 90 dpi and two seconds window) achieved 83.9% accuracy (Precision = 56.1%, Sensitivity = 37.2%, Specificity = 93.8%) in a similar condition (i.e., detecting sounds with SNR > 1). Increasing the dpi in the training images improved the model's precision, but decreased its sensitivity, resulting in lower accuracy (Table 1). The best CNN1 model showed lower precision than the one built using 90 dpi but higher sensitivity (or recall), thus reflecting higher F1 score (Table 1). Notably, exploratory ad hoc tests investigating the duration of the segments (i.e., window size) used to build the dataset and the resolution of the training images were crucial in determining the best detection model.

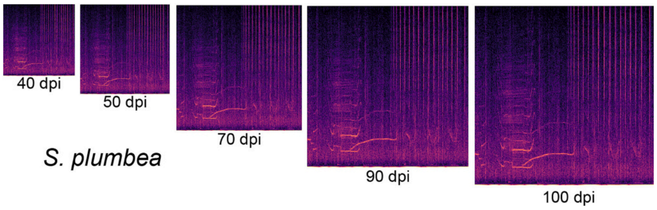


Fig. 5. Indian Ocean humpback dolphin (*Sousa plumbea*) vocalisations captured in a single two second window length segment and converted to a linear spectrogram in images with distinct dots per inch (dpi). Sample rate 96 kHz (Nyquist frequency 48 kHz), Hann window size of 1024 samples, and a hop size of 128 samples (75% overlap).

The species identification model (CNN2) only showed high accuracy when excluding one class (i.e., *O. orca*). The two best-performing models were obtained when using segments of two seconds and 90 dpi. Furthermore, these two models achieved the best testing results when trained on two (*S. plumbea* and *T. aduncus*) and three (*S. plumbea*, *T. aduncus* and *D. delphis*) classes (Fig. 6, Table 2). The only model showing >50% accuracy for *O. orca* sound classification was the one using the transfer learning approach, although it did not perform well when identifying *S. plumbea* sounds with only 9% accuracy. The highest accuracy for *S. plumbea* sound identification in PAM was achieved using a four-class model (including *O. orca*), but this model performed poorly in distinguishing *O. orca* sounds from other species (Fig. 6). The comparison of two two-class models (*S. plumbea* x *T. aduncus*) with distinct training library sizes (8 k and 12 k) demonstrated higher accuracy for the one built using a smaller training dataset. Inference using transfer learning was nearly twice as long as the custom CNN architectures (Table 1).

4. Discussion

The algorithm developed in this study assisted in finding optimal parameters to construct a suitable training dataset to be used as input to CNNs for classification tasks on complex dolphin sounds. We found that using shorter window sizes generated more accurate models for both tasks (Tables 1 and 2). With a constant dpi, we investigated the impact of window size on the classification of dolphin calls to determine if it was necessary to encompass the longest annotated call (e.g., *O. orca* whistle) as proposed in previous studies (Dufourq et al., 2021). The comparison of two two-class (i.e., *S. plumbea* x *T. aduncus*) models both built using customised architecture and 40 dpi, but with different window sizes (3 s and 7 s, Tables 1 and 2), demonstrated better performance for models built using smaller window size, specifically two or three seconds in length (Table 1). The best model was built using a two-second window length. Smaller window size yields a more nuanced representation of dolphin sounds, allowing for the detection of rapid frequency modulation patterns that may not be discernible in longer windows (see Fig. 2). Additionally, we demonstrated that fine-tuning the dpi parameter had a significant impact on both models' accuracy as the optimal dpi differed

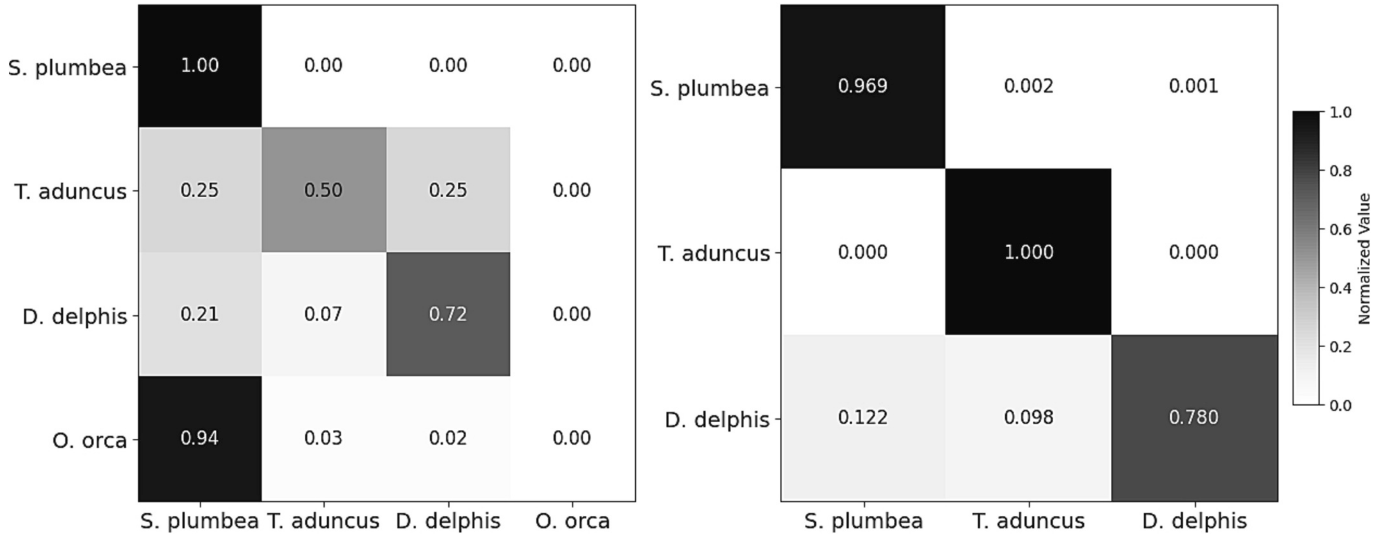


Fig. 6. Comparison of confusion matrices for a four-class (left) and a three-class (right) species identification model applied to the testing, unseen dataset (Supplementary Information A). Both models were trained using a customised architecture, two seconds window size to extract the annotated sounds from boat-based recordings, and the resulting spectrograms (images) used to train the models maintained a resolution of 90 dpi.

Table 2

Species identification models performance. n, the number of segments in the testing file detected by the CNN1 model (see *Material and Methods* section) that was used to assign species identification. Each row represents a combination of model architecture, the configurations used to build the image dataset for the training step such as window size and dpi, and the classes (i.e., species) used to build the model. Differences on accuracy related to library size was evaluated between two two-class models (*S. plumbea* x *T. aduncus*) using customised architecture, two seconds window and 70 dpi. *Four ten-minutes files were used for this testing. **Total dataset size of 12 k images. ***Total dataset size of 8 k images.

Architecture	Window size (s)	dpi	Accuracy (%)			
			S. plumbea	*T. aduncus*	*D. delphis*	*O. orca*
Customised architecture	2	70	63.9 (n = 183)**	67.8 (n = 171)**	–	–
			43.3 (n = 127)***	97.0 (n = 171)***	–	–
			78.7 (n = 127)	97.0 (n = 171)	43.9 (n = 599)	–
			53.8 (n = 130)	95.8 (n = 192)	60.8 (n = 2396)*	0 (n = 897)
		80	72.9 (n = 48)	88.6 (n = 106)	75.7 (n = 598)	0.3 (n = 595)
		90	87.3 (n = 166)	80.0 (n = 10)	–	–
			96.9 (n = 166)	**100.0** (n = 10)	**77.9** (n = 599)	–
			100.0 (n = 33)	50.0 (n = 4)	71.7 (n = 598)	0.3 (n = 309)
		100	60.7 (n = 51)	94.0 (n = 101)	83.5 (n = 492)	4.5 (n = 22)
	3	40	50.4 (n = 121)	99.2 (n = 136)	–	–
			4.1 (n = 121)	2.2 (n = 136)	9.0 (n = 598)	–
			15.7 (n = 121)	2.2 (n = 136)	0.5 (n = 598)	1.2 (n = 894)
	5	50	0.0 (n = 96)	100.0 (n = 9)	–	–
			22.2 (n = 27)	100.0 (n = 9)	23.1 (n = 596)	0.0 (n = 592)
	7	40	0.0 (n = 44)	100.0 (n = 4)	76.4 (n = 594)	0.4 (n = 227)
			36.3 (n = 44)	75.0 (n = 4)	–	–
Transfer learning	2	90	1.2 (n = 166)	100.0 (n = 10)	–	–
			10.2 (n = 166)	80.0 (n = 10)	76.1 (n = 599)	–
			9.0 (n = 166)	100.0 (n = 10)	79.1 (n = 599)	51.1 (n = 432)
	3	40	0.0 (n = 284)	36.5 (n = 41)	9.5 (n = 598)	1.9 (n = 827)
	7	40	0.0 (n = 292)	96.5 (n = 29)	83.3 (n = 588)	4.6 (n = 434)

between the best CNN1 (dpi = 70) and CNN2 (dpi = 90) models, and higher or lower dpi settings were not effective for both tasks. Furthermore, our results in Table 2 reveal that the differences in model accuracy, due to window size and dpi, may have accounted for variations in the number of detections considered for species identification in the different CNN2 models. Although presenting higher precision compared to the best CNN1 model described before, the model built using customised architecture, two seconds window, and 90 dpi showed lower sensitivity, thus potentially depending on strong signals from dolphin vocalisations (SNR > 1) to be detected and then classified at the species level.

The best CNN2 model successfully identified *S. plumbea*, *T. aduncus* and *D. delphis* sounds in a three-class classification model in the unseen data (Table 2). However, it was unable to perform well when including *O. orca* that, interestingly, produces distinct echolocation click train patterns and complex calls including biphonic whistles with multiple harmonics (Miller and Bain, 2000), which are quite distinguishable from other species with mostly single contour whistle repertoires (Erbs et al., 2017). The inefficiency of the four-class CNN2 model can likely be attributed to the small sample size for *O. orca*, representing only ~0.9% of all annotated dolphin sounds which was potentially limited by a small diversity of calls and behavioural contexts (Oswald et al., 2008; Quick and Janik, 2008). Nevertheless, the three-class CNN2 model represents a significant advance in dolphin sound classification tasks for taxonomic identification, especially for *S. plumbea* monitoring in South African waters. It is worth stressing that *O. orca* is not as common as *T. aduncus* or *D. delphis* (Best et al., 2010; Melly et al., 2018). They also produce visually distinguishable sounds from the other dolphins investigated, allowing them to be manually checked in a post-hoc analysis of results. Future investigation may address transfer learning using the same optimal window size and dpi found for the best CNN2 model as this approach, in our case, performed better than any other model for *O. orca* sounds (Table 2).

The approach proposed in this study presents a promising framework for future assessments on dolphin detection and identification using PAM recordings as the algorithm was based on the biology of dolphin sounds. The nature of vocal production varies considerably among dolphins as some species are more actively vocal than others, potentially driven by group size dynamics (Oswald et al., 2008; Quick and Janik, 2008) (Fig. 2), resulting in a different number of sound detections extracted from the training dataset for each species, despite a similar number of boat-based encounters (see *Material and Methods* section, Supporting Information A). We balanced the dataset to account for the imbalance of the total number of detections per species, to match the largest dataset for a class (i.e., *D. delphis*). Also, dolphin vocal production is dependent on its behavioural context (Quick and Janik, 2008), and thus we also balanced each species dataset per encounter to ensure equal weights for the sounds produced in different contexts. Indian Ocean humpback dolphins, for example, presented long periods of echolocation click trains while on other occasions only a few whistles (personal observation on the training dataset). This approach ensured a better representation of whistles in the dataset for this species.

The use of AEs to define a time period of dolphin detections not only assisted in species identification by handling potential false positives but also defined periods of dolphin activity near moored hydrophones that may be useful for future ecological studies. Here, we built a framework to test the efficiency of the detection model (i.e., CNN1) based on AEs (Fig. 4) as the identification model (i.e., CNN2) was dependent on the sounds captured within each AE. In other words, we assessed the taxonomic identification of dolphin sounds based on the proportion of classified segments for each species in a certain time period (i.e., AE). We used this approach as, for certain species, the classification tasks based on one call may not be recommended (Rankin et al., 2017) due to the time-frequency characteristics of vocalisations overlapping with other species in the area, thus contributing to decreased accuracy in classification models (Yang et al., 2020). Killer whales are known to be able to mimic other dolphin species (Musser et al., 2014) and other marine mammal species (Foote et al., 2006). As such, it is necessary to consider the context in which those sounds were produced, instead of identifying single clicks, burst pulses, or whistles. Although our algorithm does not identify mixed species groups, it might assist future dedicated research on this complex task. One can still verify the proportion of classified detections for each AE that is given in the output, and even experiment with more conservative times between AE thus assigning species identification based on more individualised groups of vocalisations. In this context, it is important to emphasize that a drawback of employing CNNs, in the manner our algorithm was designed is a limitation of identifying only one species per second. Consequently, the model is unable to distinguish between detections where two species are vocalizing simultaneously. However, this topic needs to be further investigated in detail.

Our study showcases the exceptional performance of CNNs in accurately classifying complex biological patterns such as click trains across species. Specifically, the testing data for *S. plumbea* was composed of a few whistles and a long series of click trains, for which the model correctly assigned 96.9% of the detected dolphin sounds ($n = 166$, Table 2). In this way, most of the click detections were correctly assigned at the species level. It is worth noting that the sample rate used here (i.e., 96 kHz) did not capture all of the dolphin click energy that can reach up to 150 kHz (Au, 2000). However, the decision to use a sample rate of 96 kHz was made as this is a widely used sampling frequency that captures the entire frequency range of most dolphin whistles (Au, 2000) while maximizing the deployment time for moored hydrophones, compared to full bandwidth recordings.

5. Conclusion

This study aimed to develop a sound classifier to acoustically monitor the critically endangered humpback dolphin in South African waters. As this species coexists with three other whistling dolphin species in the study region (Findlay et al., 1992), a species identification model was deemed essential. Our findings are encouraging and can greatly assist conservation efforts by providing a tool for ecologists and researchers. The algorithm holds significant promise as a tool to be further developed for the monitoring and research on Indian Ocean humpback dolphin acoustics in long-term recordings. The spatiotemporal definition of AEs to investigate Indian Ocean humpback dolphins' activity may assist studies on habitat use (Caruso et al., 2020) and those using individually distinctive signature whistles (Deecke and Janik, 2006; Janik et al., 2013) as input to mark-recapture approaches for population dynamics studies (Longden et al., 2020). The proposed framework can be adapted to other similar tasks involving PAM and species identification tasks, especially on cetaceans. The automated adjustment of main parameters such as sample rate, dpi, and window size enhances the adaptability of the application. The output of the application may define the time of dolphin activity near a moored hydrophone, with a customisable time period between AEs that can be tailored to other locations and studies. Dolphins mostly live in a fission-fusion society, so the AE definition (see *Material and Methods* section) can be adapted for other species to assist with social-network studies based on group composition within a time frame (Whitehead, 2008).

We demonstrated the power of CNNs on the taxonomic identification of dolphin sounds. The open-source application presented here advances the research in improving the detection and identification of dolphin vocalisations in audio recordings and will be valuable for monitoring the endangered Indian Ocean humpback dolphin in South African waters. The effective performance of the algorithm provided here encourages future research on using customisable CNNs and algorithms for the identification of complex signals. The proposed framework was designed to easily fine-tune classification tasks of biological sounds and may increase the use of CNNs through a near-friendly, Linux operating system interface. Future research may address further improvement on the detectability of dolphin vocalisations, enhancing identification accuracy, and categorising these sounds to potentially assign specific behavioural activity for each AE. Moreover, further research should be conducted to reduce processing time and facilitate real-time monitoring, thereby expanding the potential applications of this algorithm. The utilization of a high-performing application for dolphin identification in low-cost devices with "low" sample rates (i.e., 96 kHz) could prove invaluable for PAM in low-income countries (Lamont et al., 2022), particularly for optimising battery and deployment time. The development of effective remote sensing tools to monitor endangered dolphin species with optimised sampling rates may help expand hydrophone networks and cover larger areas in longer periods. There are fewer than 500 humpback dolphins remaining in South African waters (Vermeulen et al., 2018) and the population is under severe threat from anthropogenic (Plön et al., 2015) and natural impacts (Frainer et al., 2022). The proposed framework could be further refined by incorporating a new class into CNN2 to identify potential threats to the Indian Ocean humpback dolphin such as boat traffic, while assisting population dynamics and habitat use studies on this endangered species. The long-term monitoring of this species using acoustics may ensure a replicable way to evaluate changes in population dynamics in historic sites of occurrence.

Declaration of Competing Interest

None.

Data availability

All code for training and testing the neural networks is available at https://github.com/Gui-Frainer/CetusID. A subset of the acoustic recordings used for the demonstration notebook, including labels for the acoustic data used for training, as well as the testing dataset with confirmed species identification has been stored in Zenodo and can be accessed at DOI: https://doi.org/10.5281/zenodo.8074949.

Acknowledgments

This study was supported by a National Research Foundation grant awarded to T.G. (Grant number: 116047) and including G.F.'s postdoctoral fellowship and S.D.'s Ph.D. scholarship. Authors would like to thank Oceans Research Institute, the University of the Witwatersrand, and the Mammal Research Institute Whale Unit (University of Pretoria) for their logistical support during data collection. The research was conducted under research permits issued by the Department of Environment Forestry and Fisheries (and its predecessors) to S.E. (RES 2015/94 to 2021/26). We are grateful to Microsoft Azure for the sponsorship in computing time for the use of virtual machines through the program 'AI for Earth' research grants, which was fundamental to running all the analyses performed here. E.D. is supported by a research chairship from the African Institute for Mathematical Sciences South Africa. E.D.'s work was carried out with the aid of a grant from the International Development Research Centre, Ottawa, Canada, www.idrc.ca, and with financial support from the Government of Canada, provided through Global Affairs Canada (GAC), www.international.gc.ca.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2023.102291.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., 2016. Tensorflow: large-scale machine learning on heterogeneous distributed systems arXiv preprint arXiv:160304467.

Allen, A.N., Harvey, M., Harrell, L., Jansen, A., Merkens, K.P., Wall, C.C., Cattiau, J., Oleson, E.M., 2021. A convolutional neural network for automated detection of humpback whale Song in a diverse, long-term passive acoustic dataset. Front. Mar. Sci. 8 https://doi.org/10.3389/fmars.2021.607321.

Au, W.L., 2000. Hearing in whales and dolphins: An overview. In: Au, W.L., Richard, R.F. (Eds.), Hearing by Whales and Dolphins. Springer, New York, pp. 1–42. https://doi.org/10.1007/978-1-4612-1150-1_1.

Bergler, C., Smeele, S.Q., Tyndel, S.A., Barnhill, A., Ortiz, S.T., Kalan, A.K., Cheng, R.X., Brinkløv, S., Osiecka, A.N., Tougaard, J., Jakobsen, F., Wahlberg, M., Nöth, E., Maier, A., Klump, B.C., 2022. ANIMAL-SPOT enables animal-independent signal detection and classification using deep learning. Sci. Rep. 12 (1), 21966. https://doi.org/10.1038/s41598-022-26429-y.

Best, P.B., Folkens, P.A., 2007. Whales and Dolphins of the Southern African Subregion. Cambridge University Press, Cape Town, South Africa, p. 338.

Best, P.B., Meÿer, M.A., Lockyer, C., 2010. Killer whales in south African waters—a review of their biology. Afr. J. Mar. Sci. 32 (2), 171–186. https://doi.org/10.2989/1814232X.2010.501544.

Bopardikar, I., Sutaria, D., Sule, M., Jog, K., Patankar, V., Klinck, H., 2018. Description and classification of Indian Ocean humpback dolphin (*Sousa plumbea*) whistles recorded off the Sindhudurg coast of Maharashtra, India. Marine Mamm. Sci. 34 (3), 755–776. https://doi.org/10.1111/mms.12479.

Buchanan, C., Bi, Y., Xue, B., Vennell, R., Childerhouse, S., Pine, M.K., Briscoe, D., Zhang, M., 2021. Deep convolutional neural networks for detecting dolphin echolocation clicks. In: 2021 36th International Conference on Image and Vision Computing New Zealand (IVCNZ), 9–10 Dec. 2021, pp. 1–6. https://doi.org/10.1109/IVCNZ54163.2021.9653250.

Caruso, F., Dong, L., Lin, M., Liu, M., Gong, Z., Xu, W., Alonge, G., Li, S., 2020. Monitoring of a nearshore small dolphin species using passive acoustic platforms and supervised machine learning techniques. Front. Mar. Sci. 7 https://doi.org/10.3389/fmars.2020.00267.

Deecke, V.B., Janik, V.M., 2006. Automated categorization of bioacoustic signals: avoiding perceptual pitfalls. J. Acoust. Soc. Am. 119 (1), 645–653.

Dong, L., Liu, M., Dong, J., Li, S., 2017. Acoustic occurrence detection of a newly recorded indo-Pacific humpback dolphin population in waters southwest of Hainan Island, China. J. Acoust. Soc. Am. 142 (5), 3198–3204. https://doi.org/10.1121/1.5011170.

Duan, D., Lü, L.-g., Jiang, Y., Liu, Z., Yang, C., Guo, J., Wang, X., 2022. Real-time identification of marine mammal calls based on convolutional neural networks. Appl. Acoust. 192, 108755 https://doi.org/10.1016/j.apacoust.2022.108755.

Dufourq, E., Durbach, I., Hansford, J.P., Hoepfner, A., Ma, H., Bryant, J.V., Stender, C.S., Li, W., Liu, Z., Chen, Q., Zhou, Z., Turvey, S.T., 2021. Automated detection of Hainan gibbon calls for passive acoustic monitoring. Remote Sens. Ecol. Conserv. 7 (3), 475–487. https://doi.org/10.1002/rse2.201.

Dufourq, E., Batist, C., Foquet, R., Durbach, I., 2022. Passive acoustic monitoring of animal populations with transfer learning. Eco. Inform. 70, 101688 https://doi.org/10.1016/j.ecoinf.2022.101688.

Erbs, F., Elwen, S.H., Gridley, T., 2017. Automatic classification of whistles from coastal dolphins of the southern African subregion. J. Acoust. Soc. Am. 141 (4), 2489–2500. https://doi.org/10.1121/1.4978000.

Erbs, F., Gaona, M., van der Schaar, M., Zaugg, S., Ramalho, E., Houser, D., André, M., 2023. Towards automated long-term acoustic monitoring of endangered river dolphins: a case study in the Brazilian Amazon floodplains. Sci. Rep. 13 (1), 10801. https://doi.org/10.1038/s41598-023-36518-1.

Fearey, J., Elwen, S.H., James, B.S., Gridley, T., 2019. Identification of potential signature whistles from free-ranging common dolphins (*Delphinus delphis*) in South Africa. Anim. Cogn. 22 (5), 777–789. https://doi.org/10.1007/s10071-019-01274-1.

Findlay, K.P., Best, P.B., Ross, G.J.B., Cockcroft, V.G., 1992. The distribution of small odontocete cetaceans off the coasts of South Africa and Namibia. S. Afr. J. Mar. Sci. 12 (1), 237–270. https://doi.org/10.2989/02577619209504706.

Foote, A.D., Griffin, R.M., Howitt, D., Larsson, L., Miller, P.J.O., Rus Hoelzel, A., 2006. Killer whales are capable of vocal learning. Biol. Lett. 2 (4), 509–512. https://doi.org/10.1098/rsbl.2006.0525.

Fouda, L., Wingfield, J.E., Fandel, A.D., Garrod, A., Hodge, K.B., Rice, A.N., Bailey, H., 2018. Dolphins simplify their vocal calls in response to increased ambient noise. Biol. Lett. 14 (10), 20180484. https://doi.org/10.1098/rsbl.2018.0484.

Fox, K.C.R., Muthukrishna, M., Shultz, S., 2017. The social and cultural roots of whale and dolphin brains. Nat. Ecol. Evol. 1 (11), 1699–1705. https://doi.org/10.1038/s41559-017-0336-y.

Frainer, G., Plön, S., Serpa, N.B., Moreno, I.B., Huggenberger, S., 2019. Sound generating structures of the humpback dolphin *Sousa plumbea* (Cuvier, 1829) and the directionality in dolphin sounds. Anat. Rec. 302 (6), 849–860. https://doi.org/10.1002/ar.23981.

Frainer, G., Huggenberger, S., Moreno, I.B., Plön, S., Galatius, A., 2021. Head adaptation for sound production and feeding strategy in dolphins (Odontoceti: Delphinida). J. Anat. 238 (5), 1070–1081. https://doi.org/10.1111/joa.13364.

Frainer, G., Elwen, S., Dines, S., James, B., Vermeulen, E., Penry, G., Vargas-Fonseca, O. A., Atkins, S., Conry, D., Gridley, T., 2022. Rostrum abnormalities in the endangered Indian Ocean humpback dolphin (*Sousa plumbea*) in South Africa. Integr. Zool. https://doi.org/10.1111/1749-4877.12685 n/a (n/a).

de Freitas, M., Jensen, F.H., Tyne, J., Bejder, L., Madsen, P.T., 2015. Echolocation parameters of Australian humpback dolphins (*Sousa sahulensis*) and indo-Pacific bottlenose dolphins (*Tursiops aduncus*) in the wild. J. Acoust. Soc. Am. 137 (6), 3033–3041. https://doi.org/10.1121/1.4921277.

Gillespie, D., Mellinger, D.K., Gordon, J., McLaren, D., Redmond, P., McHugh, R., Trinder, P., Deng, X.Y., Thode, A., 2009. PAMGUARD: Semiautomated, open source software for real-time acoustic detection and localization of cetaceans. J. Acoust. Soc. Am. 125 (4), 2547. https://doi.org/10.1121/1.4808713.

Gridley, T., Cockcroft, V.G., Hawkins, E.R., Blewitt, M.L., Morisaka, T., Janik, V.M., 2014. Signature whistles in free-ranging populations of indo-Pacific bottlenose dolphins, *Tursiops aduncus*. Marine Mamm. Sci. 30 (2), 512–527. https://doi.org/10.1111/mms.12054.

Gridley, T., Nastasi, A., Kriesell, H.J., Elwen, S.H., 2015. The acoustic repertoire of wild common bottlenose dolphins (*Tursiops truncatus*) in Walvis Bay, Namibia. Bioacoustics 24 (2), 153–174. https://doi.org/10.1080/09524622.2015.1014851.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

Janik, V.M., King, S.L., Sayigh, L.S., Wells, R.S., 2013. Identifying signature whistles from recordings of groups of unrestrained bottlenose dolphins (*Tursiops truncatus*). Mar. Mamm. Sci. 29 (1), 109–122. https://doi.org/10.1111/j.1748-7692.2011.00549.x.

Jaramillo-Legorreta, A., Cardenas-Hinojosa, G., Nieto-Garcia, E., Rojas-Bracho, L., Ver Hoef, J., Moore, J., Tregenza, N., Barlow, J., Gerrodette, T., Thomas, L., Taylor, B., 2017. Passive acoustic monitoring of the decline of Mexico's critically endangered vaquita. Conserv. Biol. 31 (1), 183–191. https://doi.org/10.1111/cobi.12789.

Jensen, F.H., Bejder, L., Wahlberg, M., Aguilar Soto, N., Johnson, M., Madsen, P.T., 2009. Vessel noise effects on delphinid communication. Mar. Ecol. Prog. Ser. 395, 161–175.

Jensen, F.H., Johnson, M., Ladegaard, M., Wisniewska, D.M., Madsen, P.T., 2018. Narrow acoustic field of view drives frequency scaling in toothed whale biosonar. Curr. Biol. 28 (23), 3878–3885.e3873. https://doi.org/10.1016/j.cub.2018.10.037.

Jordan, M.I., Mitchell, T.M., 2015. Machine learning: trends, perspectives, and prospects. Science 349 (6245), 255–260. https://doi.org/10.1126/science.aaa8415.

Karczmarski, L., Cockcroft, V.G., McLachlan, A., Winter, P.E.D., 1998. Recommendations for the conservation and management of humpback dolphins *Sousa chinensis* in the Algoa Bay region, South Africa, 41(2), p. 9. https://doi.org/10.4102/koedoe.v41i2.257.

Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization arXiv preprint arXiv:14126980.

Lamont, T.A.C., Chapuis, L., Williams, B., Dines, S., Gridley, T., Frainer, G., Fearey, J., Maulana, P.B., Prasetya, M.E., Jompa, J., Smith, D.J., Simpson, S.D., 2022.

HydroMoth: testing a prototype low-cost acoustic recorder for aquatic environments. Remote Sens. Ecol. Conserv. 8 (3), 362–378. https://doi.org/10.1002/rse2.249.

Longden, E.G., Elwen, S.H., McGovern, B., James, B.S., Embling, C.B., Gridley, T., 2020. Mark–recapture of individually distinctive calls—a case study with signature whistles of bottlenose dolphins (*Tursiops truncatus*). J. Mammal. 101 (5), 1289–1301.

Luo, W., Yang, W., Zhang, Y., 2019. Convolutional neural network for detecting odontocete echolocation clicks. J. Acoust. Soc. Am. 145 (1), EL7–EL12. https://doi.org/10.1121/1.5085647.

McFee, B., Raffel, C., Liang, D., Ellis, D.P., McVicar, M., Battenberg, E., Nieto, O., 2015. librosa: Audio and music signal analysis in python. In: Proceedings of the 14th Python in Science Conference, pp. 18–25.

Mead, J.G., 1975. Anatomy of the external nasal passages and facial complex in the Delphinidae (Mammalia: Cetacea). Smithsonian Contrib. Zool. 207, 1–35.

Melly, B.L., McGregor, G., Hofmeyr, G.J.G., Plön, S., 2018. Spatio-temporal distribution and habitat preferences of cetaceans in Algoa Bay, South Africa. J. Mar. Biol. Assoc. U. K. 98 (5), 1065–1079. https://doi.org/10.1017/S0025315417000340.

Miller, P.J.O., Bain, D.E., 2000. Within-pod variation in the sound production of a pod of killer whales, *Orcinus orca*. Anim. Behav. 60 (5), 617–628. https://doi.org/10.1006/anbe.2000.1503.

Munger, L., Lammers, M.O., Cifuentes, M., Würsig, B., Jefferson, T.A., Hung, S.K., 2016. Indo-Pacific humpback dolphin occurrence north of Lantau Island, Hong Kong, based on year-round passive acoustic monitoring. J. Acoust. Soc. Am. 140 (4), 2754–2765. https://doi.org/10.1121/1.4963874.

Musser, W.B., Bowles, A.E., Grebner, D.M., Crance, J.L., 2014. Differences in acoustic features of vocalizations produced by killer whales cross-socialized with bottlenose dolphins. J. Acoust. Soc. Am. 136 (4), 1990–2002. https://doi.org/10.1121/1.4893906.

Nur Korkmaz, B., Diamant, R., Danino, G., Testolin, A., 2023. Automated detection of dolphin whistles with convolutional networks and transfer learning. Front. Artif. Intell. 6 https://doi.org/10.3389/frai.2023.1099022.

K. Lisa Yang Center for Conservation Bioacoustics, 2023. Raven Pro: Interactive Sound Analysis Software (Version 1.6.5) [Computer software]. The Cornell Lab of Ornithology, Ithaca, NY. Available from. https://ravensoundsoftware.com/.

Oswald, J.N., Rankin, S., Barlow, J., 2008. To whistle or not to whistle? Geographic variation in the whistling behavior of small odontocetes. Aquat. Mamm. 34 (3), 288–302.

Oswald, J.N., Walmsley, S.F., Casey, C., Fregosi, S., Southall, B., Janik, V.M., 2021. Species information in whistle frequency modulation patterns of common dolphins. Philos. Trans. R. Soc. B 376 (1836), 20210046. https://doi.org/10.1098/rstb.2021.0046.

Palmer, K.J., Brookes, K.L., Davies, I.M., Edwards, E., Rendell, L., 2019. Habitat use of a coastal delphinid population investigated using passive acoustic monitoring. Aquat. Conserv. 29 (S1), 254–270. https://doi.org/10.1002/aqc.3166.

Peddemors, V.M., 1999. Delphinids of southern Africa: a review of their distribution, status and life history. J. Cetacean Res. Manag. 1 (2), 157–165. https://doi.org/10.47536/jcrm.v1i2.463.

Plön, S., Cockcroft, V.G., Froneman, W.P., 2015. The natural history and conservation of Indian Ocean humpback dolphins (*Sousa plumbea*) in south African waters. In: Jefferson, T.A., Curry, B.E. (Eds.), Advances in Marine Biology, vol. 72. Academic Press, Oxford, pp. 143–162.

Quick, N.J., Janik, V.M., 2008. Whistle rates of wild bottlenose dolphins (*Tursiops truncatus*): influences of group size and behavior. J. Comp. Psychol. 122 (3), 305.

Rankin, S., Archer, F., Keating, J.L., Oswald, J.N., Oswald, M., Curtis, A., Barlow, J., 2017. Acoustic classification of dolphins in the California current using whistles, echolocation clicks, and burst pulses. Mar. Mamm. Sci. 33 (2), 520–540. https://doi.org/10.1111/mms.12381.

Schoeman, R.P., Erbe, C., Plön, S., 2022. Underwater chatter for the win: a first assessment of underwater soundscapes in two bays along the eastern Cape Coast of South Africa. J. Marine Sci. Eng. 10 (6) https://doi.org/10.3390/jmse10060746.

Shabangu, F.W., Yemane, D., Best, G., Estabrook, B.J., 2022. Acoustic detectability of whales amidst underwater noise off the west coast of South Africa. Mar. Pollut. Bull. 184, 114122 https://doi.org/10.1016/j.marpolbul.2022.114122.

Song, Z., Zhang, C., Fu, W., Gao, Z., Ou, W., Zhang, J., Zhang, Y., 2022. Investigation on whistle directivity in the indo-Pacific humpback dolphin (*Sousa chinensis*) through numerical modeling. J. Acoust. Soc. Am. 151 (6), 3573–3579. https://doi.org/10.1121/10.0011513.

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 (3), 368–380. https://doi.org/10.1111/2041-210X.13103.

Sugai, L.S.M., Silva, T.S.F., Ribeiro Jr., J.W., Llusia, D., 2019. Terrestrial passive acoustic monitoring: review and perspectives. BioScience 69 (1), 15–25. https://doi.org/10.1093/biosci/biy147.

Temple, A.J., Tregenza, N., Amir, O.A., Jiddawi, N., Berggren, P., 2016. Spatial and temporal variations in the occurrence and foraging activity of coastal dolphins in Menai Bay, Zanzibar, Tanzania. PLoS One 11 (3), e0148995. https://doi.org/10.1371/journal.pone.0148995.

Vermeulen, E., Bouveroux, T., Plön, S., Atkins, S., Chivell, W., Cockcroft, V., Conry, D., Gennari, E., Hörbst, S., James, B.S., Kirkman, S., Penry, G., Pistorius, P., Thornton, M., Vargas-Fonseca, O.A., Elwen, S.H., 2018. Indian Ocean humpback dolphin (*Sousa plumbea*) movement patterns along the South African coast. Aquat. Conserv. Mar. Freshwat. Ecosyst. 28 (1), 231–240. https://doi.org/10.1002/aqc.2836.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., Vijaykumar, A., Bardelli, A.P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C.N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D.A., Hagen, D.R., Pasechnik, D.V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G.A., Ingold, G.-L., Allen, G.E., Lee, G.R., Audren, H., Probst, I., Dietrich, J.P., Silterra, J., Webber, J.T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J.L., de Miranda Cardoso, J.V., Reimer, J., Harrington, J., Rodríguez, J.L.C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N.J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P.A., Lee, P., McGibbon, R.T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T.J., Robitaille, T.P., Spura, T., Jones, T.R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y.O., Vázquez-Baeza, Y., SciPy C, 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17 (3), 261–272. https://doi.org/10.1038/s41592-019-0686-2.

Wang, W., Yin, Y., Xie, Q., Fan, S., Gui, D., Wang, D., 2020. Applying machine learning method to identify indo-pacific humpback dolphin click signals. In: 2020 IEEE/OES Autonomous Underwater Vehicles Symposium (AUV), 30 Sept.-2 Oct. 2020, pp. 1–6. https://doi.org/10.1109/AUV50043.2020.9267907.

White, E.L., White, P., Bull, J., Risch, D., Beck, S., Edwards, E., 2022. More than a whistle: automated detection of marine sound sources with a convolutional neural network. Front. Mar. Sci. 9.

Whitehead, H., 2008. Analyzing Animal Societies: Quantitative Methods for Vertebrate Social Analysis. University of Chicago Press.

Yang, L., Sharpe, M., Temple, A.J., Jiddawi, N., Xu, X., Berggren, P., 2020. Description and classification of echolocation clicks of Indian Ocean humpback (*Sousa plumbea*) and indo-Pacific bottlenose (*Tursiops aduncus*) dolphins from Menai Bay, Zanzibar, East Africa. PLoS One 15 (3), e0230319. https://doi.org/10.1371/journal.pone.0230319.

Ziegenhorn, M.A., Frasier, K.E., Hildebrand, J.A., Oleson, E.M., Baird, R.W., Wiggins, S. M., Baumann-Pickering, S., 2022. Discriminating and classifying odontocete echolocation clicks in the Hawaiian islands using machine learning methods. PLoS One 17 (4), e0266424. https://doi.org/10.1371/journal.pone.0266424.