Unlocking the soundscape of coral reefs with artificial intelligence

Ben Williams[ab], Santiago M. Belvanera[a], Sarab S. Sethi[ac], Timothy A.C. Lamont[d], Jamaluddin Jompa[e], Mochyudho Prasetya[f], Laura Richardson[g], Lucille Chapuis[h], Emma Weschke[h], Andrew Hoey[i], Ricardo Beldade[j], Suzanne C. Mills[kl], Anne Haguenauer[k], Frederic Zuberer[k], Stephen D. Simpson[h], David Curnick[b], Kate E. Jones[a]

[a] *Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, London, United Kingdom.* [b] *Zoological Society of London, Regents Park, London, UK.* [c] *Department of Plant Sciences, University of Cambridge, Cambridge, UK.* [d] *Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK.* [e] *Graduate School, Universitas Hasanuddin, Makassar, Indonesia.* [f] *MARS Sustainable Solutions, Makassar, Indonesia.* [g] *School of Ocean Sciences, Bangor University, Askew Street, Menai Bridge, Anglesey, LL59 5LR, UK.* [h] *School of Biological Sciences, University of Bristol, Bristol BS8 1TQ, UK.* [i] *Australian Research Council Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Queensland 4811, Australia.* [j] *Estación Costera de Investigaciones Marinas, Millennium Nucleus for Ecology and Conservation of Temperate Mesophotic Reef Ecosystems, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile.* [k] *PSL Research University: EPHE-UPVD-CNRS, UAR 3278 CRIOBE BP 1013, 98729 Papetoai, Moorea, French Polynesia.* [l] *Laboratoire d'Excellence "CORAIL", France.*

Abstract

Passive acoustic monitoring can offer insights into the state of coral reef ecosystems at low-costs and over extended temporal periods. However, lack of effective automated analysis has impeded progress in this field. Here, we show machine learning (ML) can be used to unlock greater insights from reef soundscapes. We showcase this on a diverse set of tasks using three biogeographically independent datasets, each containing fish community, coral cover or depth zone classes. Supervised and unsupervised ML models were effective at identifying relevant classes and individual sites within these. We also compare three different approaches for extracting feature embeddings from soundscape recordings: acoustic indices designed by ecologists, a pretrained convolutional neural network (P-CNN) and a CNN trained on each dataset (T-CNN). We report P-CNNs present a powerful tool for soundscape ecologists due to their strong performance and low computational cost. Our findings have implications for soundscape ecology on any habitat.

Keywords: Artificial intelligence, machine learning, coral reef, fish diversity, coral cover, mesophotic reef, passive acoustic monitoring, soundscape.

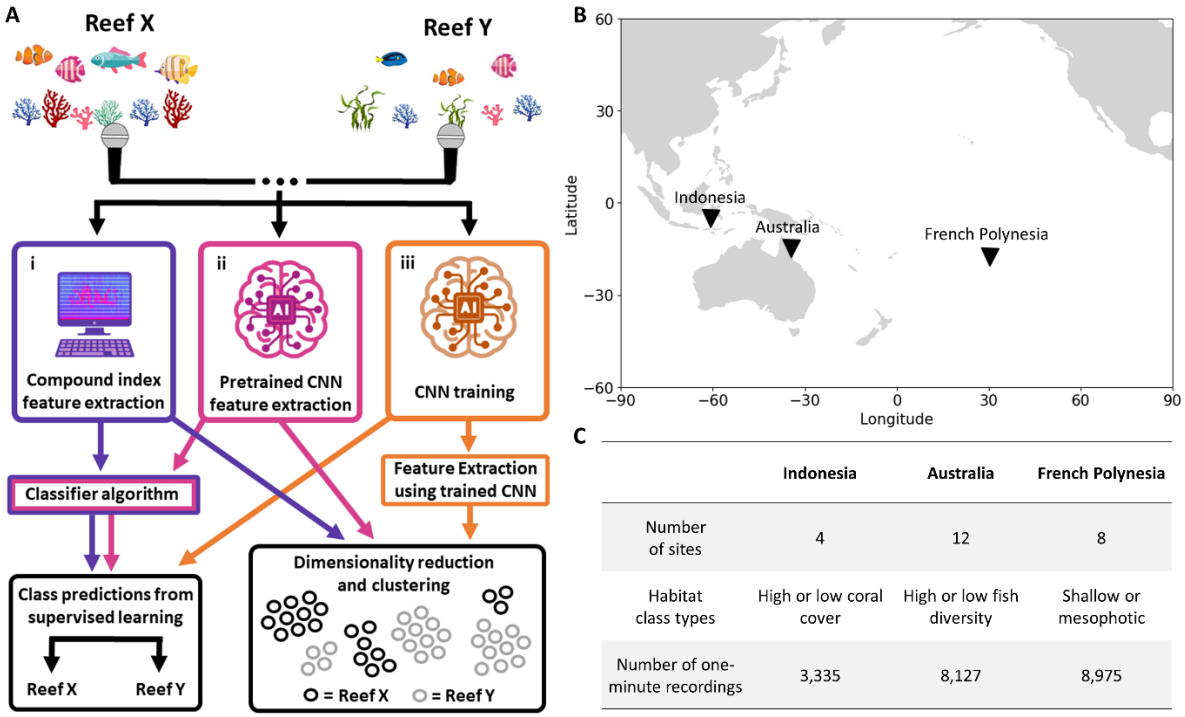
Introduction

Effective monitoring of coral reefs is essential to support their conservation and restoration[1]. Monitoring data is typically collected using diver-led underwater visual census surveys (UVC). However, UVCs incur high expertise, logistical and financial costs[1,2]. Passive acoustic monitoring (PAM) presents an alternative which can be collected with greater ease and over extended periods[3,4]. However, in contrast to ecosystems where PAM is well-established, the species identity for any given sound in reef PAM recordings is typically unknown[5,6]. Instead, coral reef PAM typically attempts to find relationships between the ecological community and its 'soundscape'; the full extent of environmental sounds present[7]. Information held within reef soundscapes often correlates with ecological community metrics (e.g., fish diversity, coral cover)[8–10] and also better captures temporal trends or the presence of cryptic organisms compared to traditional methods[11–14]. However, manual annotation of these data is highly labour-intensive, undermining labour costs saved in the field[6,15]. Previous attempts at automated analysis of soundscape data have primarily used acoustic indices; formulas designed to quantify properties of a spectrogram relevant to natural soundscapes (e.g., total amplitude, entropy across time). However, comparisons of individual handcrafted features, such as acoustic indices, provide a limited resolution[16,17]. This is further compounded by a lack of standardisation for the many parameters required when calculating indices[18,19]. Machine learning (ML) presents a powerful alternative to automate the analysis of bioacoustic data across multiple ecosystems and taxa, so far primarily focused on generating species or vocalisation detectors[15,20,21]. However, applications of ML for soundscape ecology are much less developed for most ecosystems and almost entirely untested on coral reefs[22]. The optimum techniques and insights these can generate are therefore unknown.

We identified three approaches broadly representative of the techniques that can be used to extract feature embeddings from soundscape data for input into ML algorithms[23]. The first approach used a 'compound index', generated by calculating multiple acoustic indices from recordings and combining these into a vector, which can be used to train computationally cheap ML algorithms. A limitation of compound indices is that they are hand crafted formulas that likely overlook other valuable but unidentified features. Our second approach used deep learning (DL) models custom trained on reef soundscape data. DL is a sub-field of ML which uses neural networks to learn the most relevant features from complex data types like audio[24]. A disadvantage of DL is that greater expertise and computational resources are required to train neural networks[25]. Our third approach attempted to retain the benefits of DL whilst overcoming its disadvantages by using a pretrained neural network. Pretrained networks are typically trained on large datasets to predict a broad selection of classes. If the final classification layers are removed, these networks can be used as computationally cheap embedding extractors which generalise well to new data, known as 'transfer learning'. Similar to a compound index, these embeddings can be used to train lower complexity ML algorithms, making this process significantly faster than training a custom neural network from scratch[26]. We used a convolutional neural network (CNN) for both the custom trained CNNs (T-CNN) and pretrained CNN (P-CNN) due to the success of CNNs in bioacoustic studies across a range of taxa[20,27,28].

We compared the performance of these three embedding extraction approaches at supervised and unsupervised ML tasks derived from coral reef soundscape recordings (Fig. 1a). Unsupervised ML was used to reduce the dimensions of embeddings and cluster these whilst blind to the true habitat or site identity class of recordings. The fidelity of clustered outputs were then compared to the known classes. For supervised learning, ML classifiers were trained on labelled recordings to directly predict the class of held-out recordings and the accuracy reported. We outline the ecological findings obtainable with these methods and discuss the benefits of each.

To test the performance of ML across a variety of datasets and tasks, we used recordings from three biogeographically distinct locations: Indonesia, Australia and French Polynesia (Fig. 1b)[29]. Each dataset was treated independently throughout to provide a suite of unique data and challenges. Each dataset was comprised of two unique habitat types: high or low coral cover for Indonesia, high or low fish diversity for Australia, and shallow or mesophotic depths for French Polynesia (Fig. 1c). Tasks were set to test the ability of models trained using the three embedding extraction approaches (compound index, P-CNN, T-CNN) to identify these habitat classes. We also tested each approaches ability to identify individual sites within each dataset, providing an additional means to compare performance. These tasks were explored using both supervised and unsupervised ML. We also compared our results to those generated using individual acoustic indices, representative of the most commonly used method for automated analysis of reef soundscape data[30,31].



	Indonesia	Australia	French Polynesia
Number of sites	4	12	8
Habitat class types	High or low coral cover	High or low fish diversity	Shallow or mesophotic
Number of one-minute recordings	3,335	8,127	8,975

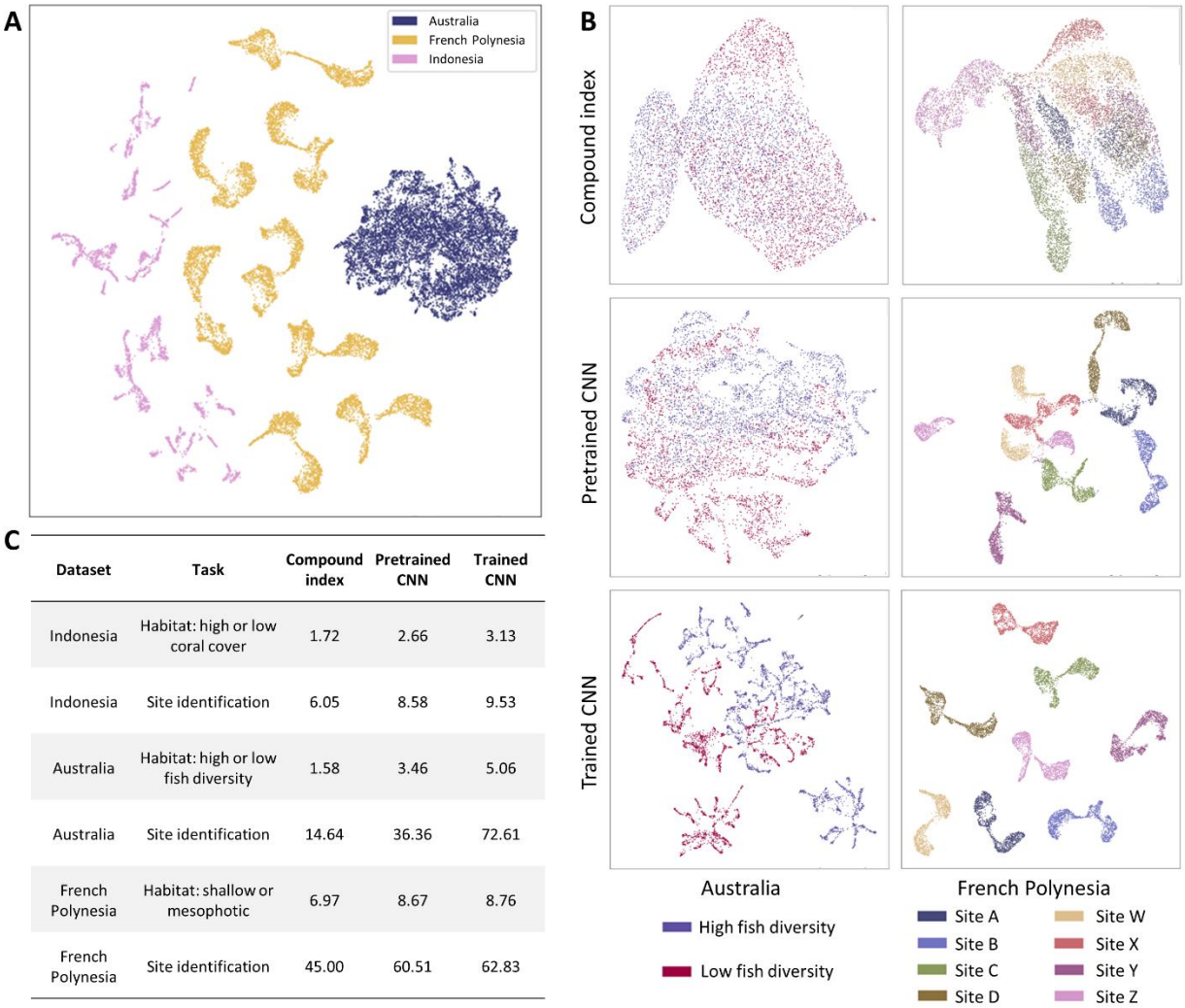
Fig. 1 (A) Analytical workflow undertaken for each coral reef soundscape dataset. Reef soundscape datasets are entered into three pathways. For both the (i) compound index and (ii) pretrained CNN pathways this first involves a feature extraction stage. For supervised learning tasks these extracted feature embeddings are then entered into a classifier algorithm (random forest) trained to predict the class of new recordings. For unsupervised learning, dimensionality reduction and visualisation are performed on extracted embeddings using uniform manifold approximation (UMAP) followed by a clustering algorithm (affinity propagation). The (iii) trained CNN pathway performs supervised classification tasks directly and, once trained, is adapted and used as a feature extractor on recordings. Dimensionality reduction and clustering is then performed on these learnt feature embeddings. **(B)** Geographic location of datasets used. **(C)** Table of the key information from each dataset. Indonesian sites were divided into high (91.2-93.1%) and low (2.1-17.6%) coral cover. Australian sites were divided into a class with high biomass (20.5-52.2kg) and species richness (45-53) scores per transect, and, low biomass (5.2-9.8kg) and species richness (30-35) scores per transect (fig. S6). French Polynesian sites were divided into shallow (10-15m) and mesophotic (55-65m) depth classes.

Exploring coral reef soundscapes with unsupervised learning

Unsupervised machine learning (ML) proved to be an effective tool for exploration of coral reef soundscapes. We show this technique to be valuable when ecological or predefined class information is unavailable, enabling the grouping of recordings based on their acoustic characteristics. Dimensionality reduction was performed on embeddings of one-minute recordings extracted by each method (compound index, P-CNN, T-CNN) using uniform manifold approximation (UMAP)[32]. For the T-CNN embeddings were obtained by training the network to identify sites, given site origin should always be available for a recording, before inferencing on the full dataset. A visualisation of the three full datasets (Indonesia, Australia, French Polynesia) using the P-CNN UMAP embeddings showed each formed distinct groups separate from the other datasets, indicating the properties of recordings from each location were unique from the other datasets. Moreover, further sub-groups within the Indonesian and French Polynesian datasets are visible (Fig. 2a).

Qualitative inspection of UMAP visualisations from each individual dataset revealed habitat and site classes were a key driver behind the clusters that were formed, indicating the influence of these on the soundscape (Fig. 2b; Fig. S1). Across all three datasets, recordings were most clearly separated into discrete clusters that conformed to habitat classes by the T-CNN, followed by the P-CNN (Fig. 2b; Fig. S1). For fair comparison, we applied uniform settings for local and global structure in all UMAP plots. Whilst in some cases the compound index conserves global similarity between sites, adjusting UMAP settings for the CNNs can help achieve this. In contrast, the compound index may reach its limits earlier in finer grained analysis (e.g adding more sites). The output also revealed a unique soundscape 'fingerprint' for many sites, with the French Polynesian dataset exhibiting the clearest discrete clusters for each site (Fig. 2b). This performance on the French Polynesian dataset is likely due to the datasets ecological attributes, broader spatial distribution (Fig. S5), or inability to rotate instruments in this dataset (Text S1) which highlights the importance of controlling for this bias. As well as identifying site and habitat classes, interactive UMAP plots also revealed temporal patterns within the data (Text S1; Supp. 2). The interactive plots showed that within each dataset recordings from similar temporal periods typically grouped closer together. For example, crepuscular periods link separate night and day clusters for the French Polynesian dataset, whilst the soundscape of two sites in Indonesia converged around the new moon (Text S1; Supp. 2).

A quantitative assessment of unsupervised performance was undertaken using the affinity propagation clustering algorithm on UMAP embeddings. Chi-squared tests were used to determine how well clusters split the data non-randomly into ground-truth habitat and site classes. Across all three datasets, the T-CNNs reported the highest fidelity to both the habitat and site classes. In every instance this was followed by the P-CNN and then compound index (Fig. 2c). However, Chi-squared scores for the T-CNNs were only marginally better than the P-CNN for the French Polynesian habitat and site identification tasks (Fig. 2c). The UMAP plots (Fig. 2b) reveal that highly distinct groups between classes in the French Polynesian dataset were already established using P-CNN embeddings, leaving minimal scope for the T-CNNs to improve despite having been explicitly trained to identify sites.



C

Dataset	Task	Compound index	Pretrained CNN	Trained CNN
Indonesia	Habitat: high or low coral cover	1.72	2.66	3.13
Indonesia	Site identification	6.05	8.58	9.53
Australia	Habitat: high or low fish diversity	1.58	3.46	5.06
Australia	Site identification	14.64	36.36	72.61
French Polynesia	Habitat: shallow or mesophotic	6.97	8.67	8.76
French Polynesia	Site identification	45.00	60.51	62.83

Fig. 2 (A) Uniform manifold approximation (UMAP) dimensionality reduction and plotting applied to P-CNN embeddings of all 20,437 one-minute recordings (3,335 from Indonesia; 8,127 from Australia; 8,925 from French Polynesia). Individual points represent a one-minute recording. **(B)** Selected plots from uniform manifold approximation (UMAP) dimensionality reduction performed on the compound index, pretrained CNN and trained CNN embeddings. Plots were produced from the Australian and French Polynesian datasets and are labelled with colours corresponding to habitat class (high or low fish diversity) and site, respectively. Remaining plots can be found in Supplementary 1. **(C)** Chi-squared scores (x 10^3) generated from contingency tables of each recordings' true class and the cluster that each recording was assigned to by affinity propagation clustering. Affinity propagation was performed on UMAP embeddings generated using each of the three methods (compound index, pretrained CNN and trained CNNs). Higher scores indicate clusters better represented true classes. All Chi-squared scores were highly significant ($p<0.001$).

Predicting habitat class and site identity with supervised learning

To predict the habitat or site identity classes for each dataset, random forest classifiers were trained on the compound index or P-CNN embeddings, whereas class predictions were output directly the T-CNNs. To separate training and evaluation data, entire sites were excluded from training for the Australian and French Polynesian datasets and entire recording blocks for the Indonesian set (Text S2). For site classification, full recording blocks from each site (e.g., a contiguous 24 hr period) were held out for evaluation to mitigate against temporal autocorrelation.

Classifiers performed well at predicting habitat and individual site classes of one-minute reef soundscape recordings. The six tasks exhibited a range of difficulty, with a mean accuracy of 0.56 (±0.05) to 1.00 (±0.0) and performance above random classification ranging from 0.23 (±0.11) to 0.87 (±0.0)(Fig. 3; Table S1). This highlights the potential of supervised learning to inference the habitat state of new sites from their soundscape. The strongest performance at site classification tasks was reported for the French Polynesian dataset and the weakest for the Australian. However, classifiers did not always perform consistently across repeats, for example the French Polynesian habitat task had a mean accuracy >0.99, but some outliers failed to discriminate classes. Differing performance across repeats was likely due to the combination of sites held out for testing (Text S2), highlighting the importance of gathering adequate training and test data across a representative selection of reefs in the given system. Between datasets, performance differences were likely driven by the difficulty of the task, sampling regime (Text S2) and spatial scale (Fig. S5).

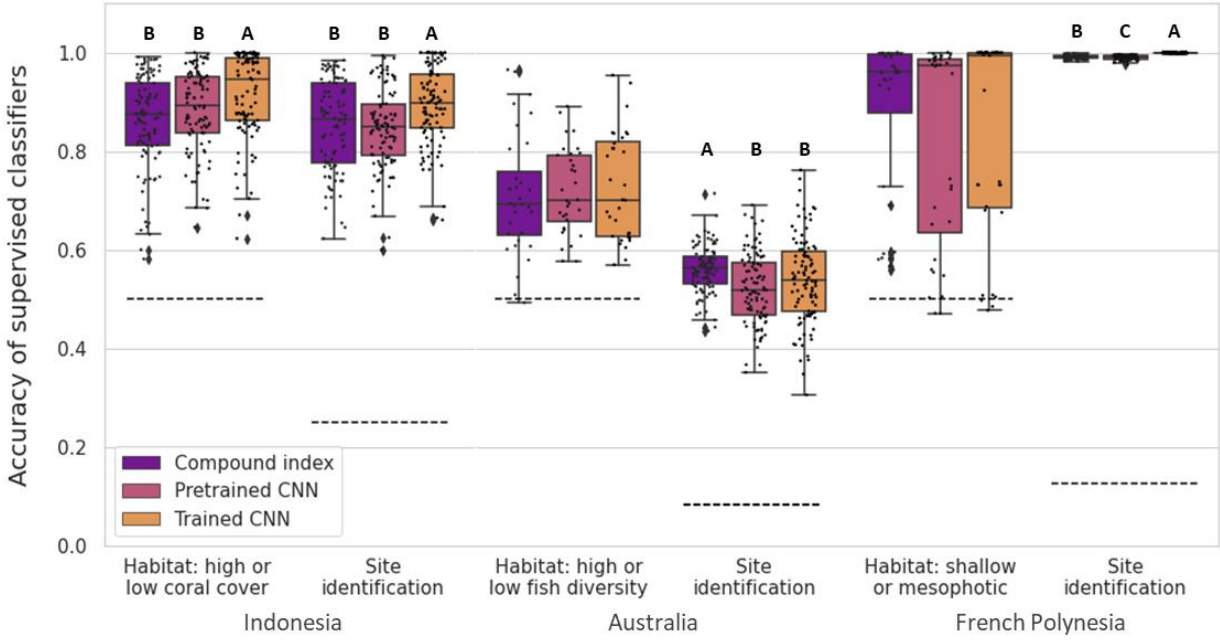


Fig. 3. Boxplots of supervised classifier accuracies for the six different tasks across repeated training instances using each of the three embedding extraction methods (compound index, pretrained CNN and trained CNN). Boxes and their bars represent the 25th, 50th and 75th quartiles. Black dotted lines represent expected accuracy using random classification (1 / number of classes). For each task, letters indicates a significant difference between these groupings by the ANOVA test. The direction of significant differences reported by the Tukey HSD are indicated, with 'A' for the group with the highest accuracy, 'B' for the second highest accuracy, and where present, 'C' for a third group with the lowest accuracy. N = 100 for the number of repeats performed for all tasks, except the habitat 'high or low fish diversity and 'shallow or mesophotic' tasks, where N = 32 (Text S2).

Each of the three feature embedding and ML methods reported similar accuracies when compared to one another at any given task, with ANOVA tests reporting no significant difference between any of the three methods for the habitat identification tasks set from the Australian and French Polynesian datasets (Fig. 3; Table S2). However, differences in accuracy were reported between each method for the Indonesian habitat classifier and for classifiers trained to identify sites for each dataset. For three of the four tasks where a difference was reported, the T-CNNs outperformed the other two methods: the habitat and site classifiers for the Indonesian dataset, and the site identity classifier for the French Polynesian dataset. Whereas, the compound index had a significantly higher accuracy than the P-CNN and T-CNN for the Australian site classifier.

Inspection of confusion matrices used to interpret classifier performance revealed the three embedding extraction methods reported similar patterns of misclassification for each task, where samples from any given class were assigned to the same incorrect classes across methods (Fig. S2). The only exception was the French Polynesian habitat classification task, with 84.1 % and 67.6 % of the mesophotic recordings misclassified as shallow by the compound index and P-CNN respectively, while this was only 42.3 % with the T-CNN. The only instance where a class was misclassified the majority of the time across all repeats occurred for site A in the Australian dataset where samples were more likely to be assigned to class G (Fig. S2). This misclassification was likely due to their close proximity (Fig. S5) and habitat attributes, both being a 'low' fish diversity site.

Benchmarking against current automation: acoustic indices

Statistical comparisons of individual acoustic indices represent the predominant method for automated analysis of reef soundscape data[30,31]. To assess their performance on our data, we deliberately selected the individual acoustic indices that reported the highest significant differences between each habitat class. Despite this biased selection, our analysis showed these indices were still unable to perform classification of individual recordings to the accuracy obtainable with ML (Fig 4). The proportion of recordings which could be unambiguously classified as the correct habitat type were only 5.2%, 13.0%, and 12.2% for each dataset respectively. Even the lowest-performing ML classifiers across all combinations of tasks and repeats achieved accuracies of 58.1%, 57.7% and 47.1% for each embedding type (Fig. 3). Furthermore, individual acoustic indices were unable to classify the site of origin for recordings (Fig. S3).

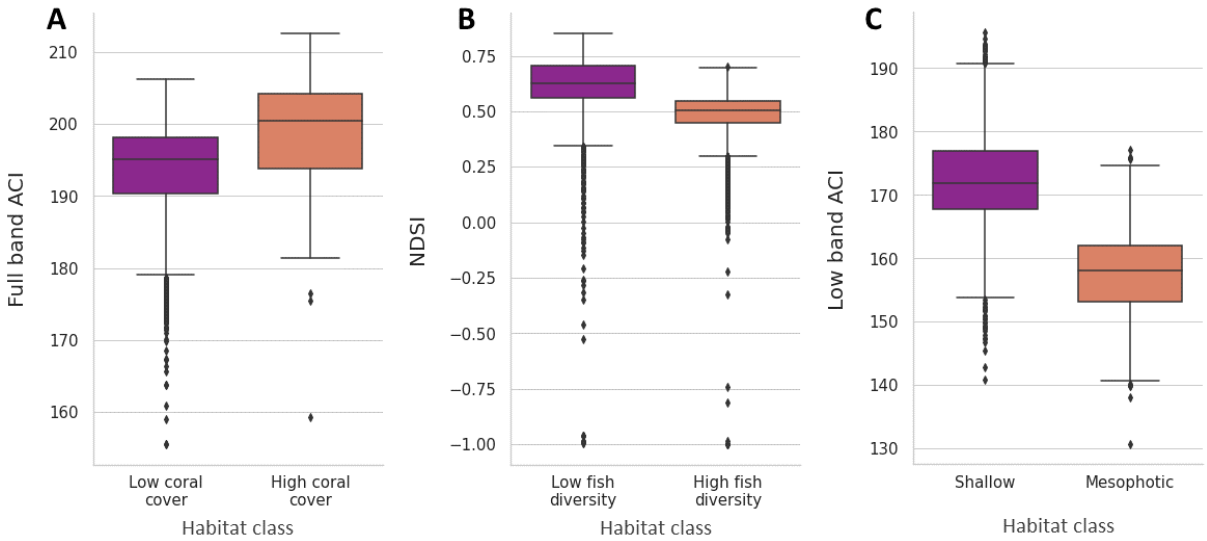


Fig. 4. Boxplots of the acoustic index for which the highest significant difference between habitat classes was reported for the **(A)** Indonesian, **(B)** Australian and **(C)** French Polynesian datasets. These were the full band acoustic complexity Index (ACI), normalised difference soundscape index (NDSI), and low band acoustic complexity index (ACI) respectively, which reported Mann-Whitney U test results of: $u = 2.00 \times 10^6$, $p<0.0001$, $u = 7.71 \times 10^6$, $p<0.0001$, and, $u = 3.64 \times 10^6$, $p<0.0001$. Boxes and their bars represent the 25th, 50th and 75th quartiles.

Discussion

Machine learning applications in coral reef soundscapes

Using three biogeographically independent datasets we demonstrate the potential of machine learning (ML) to automate the analysis of coral reef soundscape data at scale (20,437 one-minute recordings in the present study). Previous work using automated techniques has predominantly used individual acoustic indices which have revealed relationships between the soundscape and key ecological attributes on reefs[9,10,33–36]. However, comparing individual indices only provides a weak signal and often fails in new contexts[2,23,33,37,38]; supported by our findings (Fig. 4, Fig S4). Instead, we show that combining features into embeddings and inputting these into ML algorithms provides a powerful alternative, especially when using embeddings output by deep-learning models. But, care is needed when designing the sampling regimes as well as dividing training and evaluation data.

Our results provide further support for the hypothesis that the soundscape shares a relationship with ecological properties of reefs such as the fish community and coral cover. We also provide further evidence that shallow and mesophotic reefs often, but not always, exhibit distinct soundscapes[39–41] (Fig. 3). Furthermore, ML was often able to differentiate individual sites based on their acoustic 'fingerprint', an insight which could not have been elucidated using acoustic indices (Fig. S3). Our findings using unsupervised learning provide further evidence that temporal patterns influence reef soundscapes[11,42,43], with clusters typically grouping recordings from similar time periods (Text S2, Supp. 2). When considering the performance across datasets, the unique biogeography and sampling regime (Text S2) of each likely interact, making it difficult to disentangle these effects. Unsupervised learning further confirms each dataset exhibits unique properties (Fig. 2a), highlighting the need for caution if attempting to inference ecological variables using models or findings from another dataset[44].

Performance and comparison of machine learning approaches

In regards to selecting the optimum feature extraction approach for ML, a clear performance difference between the methods was revealed by the unsupervised learning tasks. Here, the T-CNNs produced clusters with a notably greater fidelity to true classes for every task compared to the compound index and in most tasks when compared to the P-CNN (Fig. 2C, Fig. S1), which in turn outperformed the compound index. These results reveal that feature embeddings obtained from deep learning yield improved outputs compared to hand crafted compound index embeddings. Notably, this represents the first demonstration that transfer learning using a network pretrained on data from an entirely different audio domain (5.2m hrs of YouTube audio) can be successfully applied to coral reef soundscapes.

For supervised classification tasks, our findings show that the difference in performance between the three embedding extraction methods on reef soundscape data was minimal. Multiple algorithms and their respective hyperparameters are available. However, it is well documented that as dataset size grows the choice of algorithm matters less[45]. Our findings support this in the context of soundscapes given the broadly similar performance exhibited at the scale of data used here. Furthermore, datasets of this scale are easily achieved using raw soundscape data[23]. Alternatively, future work could identify the minimum sampling regime necessary, enabling efficient handling of equipment resources and field time by soundscape ecologists. Further accuracy improvements to supervised classifiers are more likely obtainable by considering recording periods longer than the one-minute samples used here. Whilst short samples (e.g., one-minute) are common practice in the

literature[17,46,47], a preliminary comparison using the most common classification pooled over 24 hrs in our data highlights the potential of considering longer periods (Fig. S4).

Moving forward, for a guaranteed optimal performance we recommend users to develop T-CNNs specific to their dataset. However, the T-CNN approach requires significantly more expertise and computational overhead (Text S4). These overheads are greatly reduced for the P-CNN at the price of only a small performance dip in the tasks presented here. The P-CNN also presents a more standardised tool compared to a compound index which requires curating from an extensive set of indices and parameters for each index that currently vary from study to study[31,38]. The standardised features of a P-CNN also enable comparisons of new datasets more readily, whereas a T-CNN would require retraining. We therefore recommend the P-CNN for most use cases.

When selecting an unsupervised or supervised approach, the former has the advantage that the natural overlap which can occur between soundscapes across time can be handled, rather than being reported as a classification error. Additionally, supervised learning is limited by the sensitivity of classification accuracy to the training and evaluation split of the data, requiring careful curation of this alongside extensive repeats to ensure confidence in performance (Fig. 3). When using unsupervised approaches, additional measures such as centroid distance in the embedding space, percentage overlap between cluster allocations of samples or connectivity maps using UMAP could be used to compare and contrast treatments (e.g., level of protection or construction damage). Furthermore, an existing unsupervised model able to separate treatments can have new data from additional sites passed through and cluster allocation used as a form of classification.

Future directions

We present only a limited set of problem tasks, further experimentation with ML will help disentangle the relationship between soundscapes and the attributes of reefs which drive this. These experiments could be performed across ecological gradients, used to track temporal cycles, reveal patterns in functions reliant on soundscapes[48–50], or to explore novel applications. To train models that can identify relationships with raw biodiversity measures, ground-truth ecological survey from a representative selection of sites will be essential. However, the impact of different treatments (e.g., restoration or destructive fishing) across sites can often be compared using soundscape data alone[22,46]. Prospective users should take care to follow best practices when assembling datasets by controlling for confounding variables that can impact machine learning algorithms including instrument bias, temporal autocorrelation, careful curation of training and test sets which ideally reserve entirely separate sites for evaluation, alongside other considerations[51–53]. Future work could also aim to reap the benefits of a P-CNN whilst bridging the performance gap to the T-CNN by developing a generalisable bioacoustics or reef specific P-CNN[54].

The availability of low-cost recording technology[6,55,56] and rapid insightful analysis now possible using ML has the potential to greatly increase the scale of ecological assessment on coral reefs. As long as careful sampling design and evaluation protocols are in place, artificial intelligence represents a powerful tool to process this data and unlock new insights. The potential demonstrated here also has implications for other marine and terrestrial habitats where these techniques could be applied. While there remains much to discover, soundscape ecology has a lot to offer in terms of understanding, protecting and restoring coral reefs and other habitats around the world.

Methods

Datasets

The datasets used in this study were collected from three different coral reef biogeographic realms: (1) South Sulawesi, Indonesia, in the Tropical Indo-Pacific; (2) Lizard Island, Australia, in the Coral Sea; and (3) French Polynesia, in the Mid-South Tropical Pacific (Fig. 1B)[29]. Each dataset was collected using SoundTrap hydrophone recorders (SoundTrap 300ST, Ocean Instruments, Auckland, NZ) which were calibrated by the manufacturer with a flat frequency response. These were suspended 0.5 m above the seabed using a sub-surface buoy. For each dataset, the individual recorders were frequently rotated between sites to mitigate against recorder bias. To minimise the introduction of geophonic noise, all recordings were taken during seastates between 0-2 on the Beaufort scale and never during periods forecasted to rain. All recordings were taken in remote areas away from frequently trafficked boating channels and mooring points to minimize the presence of motorboat noise, ensuring models had to focus on features in the biophony (Text S3). Full recording schedules and instrument rotations are available in Text S2.

The Indonesian dataset was collected as part of the monitoring programme of the Mars Coral Reef Restoration Project (www.buildingcoral.com). This monitoring programme included recordings collected from four study sites across two reefs (Fig. S5b); two high coral cover reefs (Fig. S5c-d) and two low coral cover reefs (Fig. S5c). All sites were at a depth between 2–3.3m at low tide. One-hour blocks of soundscape recordings were taken at each site for five days either side of the full moon (26 August 2018) and three days either side of the following new moon (10 September 2018) during daylight (09:00–15:00), twilight (05:30–06:30 and 17:30–18:30) and night-time (23:30–00:30) periods. These were made in a counterbalanced blocking design, such that there was a similar number of recordings taken from each site, comprising an approximately even spread of time points and lunar phases (see Lamont et al., (2021) for more detail). Recordings were split into one-minute segments, providing a total of 3,335 one-minute recordings.

The Australian dataset consisted of recordings from 12 randomly selected sites (A–L) located around Lizard Island on the Great Barrier Reef; sites were separated by at least 500 m and constituted >250 x 5 m of contiguous reef (Fig. S5e). All sites were at a depth between 2–3m at low tide. Six recording blocks of 19–24 hr duration were taken at each site between 23 October to 14 December 2018; blocks from each site were spaced 9 days (SE: ±3.49) apart on average (Text S2). Recordings were not made during the Nov 11-20th period due to elevated sea states which could introduce geophonic noise. Recordings were subsampled to take one-minute clips separated by nine minutes of unused recording, providing 8,127 one-minute recordings in total. Fish community biomass and species richness data were collected along three transects at each site during this period following methods from Richardson et al., (2018). Of the 12 sites, the four with the highest and four with the lowest species richness values were identified and split into two classes, totalling 6289 one-minute recordings after excluding the other four sites (Fig. S6),. The sites from these two classes were also found to have non-overlapping fish biomass values and were therefore split into high and low fish diversity classes.

The third dataset was collected in February and March 2021 at four sites in French Polynesia, consisting of reefs around the islands of Mo'orea, Tikehau and Tahiti (Fig. S5f-h). At each site, soundscape recordings were taken at a shallow (10–15 m depth) and mesophotic (55–65 m depth) site, totalling four shallow sites (A–D) and four mesophotic sites (W–Z). Recordings were taken

simultaneously for each shallow–mesophotic pair for a continuous period of 89–97 hrs (Text S2). Recordings were subsampled to take one-minute clips separated by four minutes of unused recording, providing 8,975 one-minute recordings in total.

Extracting feature embeddings

Guidance from previous literature was used to assemble a compound index. The aim was to include a broad selection of features, whilst preventing this from including unnecessary noise which can reduce the performance of machine learning algorithms. This begun with the identification of suitable acoustic indices using the following criteria: (1) previous use in the literature on marine soundscape recordings, (2) do not require recorder calibration (precluding the use of more widely accessible recorders[56]); (3) could be computed using existing toolkits in one programming language (e.g., Python). Of the identified indices, the acoustic diversity index (ADI) and acoustic evenness (AEI) have previously been reported to strongly covary[22]. Therefore, to reduce the introduction of noise only AEI was used[57] (Table S3). Seven of the eight indices were then calculated across three different frequency bands: a low-frequency (0.05–2 kHz) where fish sounds dominate[5], a medium-frequency band (2–8 kHz) where snapping shrimp sound dominates[10], and a full band (spanning 0.05–8 kHz). We excluded frequencies below 0.05 kHz from the low- and broad-frequency band recordings to remove geophonic noise and self-noise from the recording system[58]. The exception was the normalised difference soundscape index (NDSI), which requires the input of two bands, for which 0.05–1 kHz and 2–5 kHz bands were used as in Williams et al., (2022). Recordings were split into 0.96 second segments (to match CNNs), totalling 62 per minute, and indices were calculated for each segment. The mean and standard deviation of these were then taken for each minute[17], providing a 44 dimension embedding which was used as the compound index to represent each one-minute recording. All processing was performed in Python (v3.7) using the scikit-maad package (v1.3)[59].

For the pre-trained convolutional neural network (P-CNN), we selected VGGish due to its successful application in previous work on terrestrial soundscape data[47,60]. VGGish was trained using the YouTube-100M dataset, a diverse set of YouTube audio clips totalling 5.2m hrs in length, producing a highly generalisable audio and embedding extractor[61]. It uses a version of the Visual Geometry Group object recognition architecture that was adapted for audio input[62]. The P-CNN pre-processing first downsamples recordings to 16 kHz, and divides them into non-overlapping 960 ms audio frames, processed through a Short-time Fourier Transform with 25 ms windows every 10 ms. These are integrated into 64 mel-spaced frequency bins which are log-transformed to produce a 96×64 bin log-mel spectrogram which can be passed to the network. Further details are available in Hershey et al., (2017). This P-CNN can be set to output the 128-dimension feature set from the penultimate layer in place of the classification head. We averaged these feature values from each one-minute recording to produce a single embedding. All processing was performed in Python (v3.7).

To ensure comparability with the P-CNN, we trained classifiers (T-CNN) for each respective reef soundscape dataset and task using the same architecture. The pre-processing protocol used for the P-CNN was replicated to produce 960 ms log-mel spectrogram samples. The order of these samples was shuffled and minibatches of 32 samples used for network for training. Processing was performed using Tensorflow (v1.15) and scikit-learn (v0.22) in Python (v3.7).

Unsupervised clustering

Unsupervised clustering was used to reveal inherent structures and patterns in the data without relying on pre-labeled classes. Embeddings from the compound index and P-CNN were extracted from recordings. To generate embeddings using the T-CNNs, the VGGish CNN was trained for 50

epochs on each dataset to predict which site recordings originated from within the respective dataset (see supervised classifiers for more detail). The final two layers of the trained models were then removed to produce three networks comparable to the P-CNN, except these were now trained on reef soundscape recordings from their respective datasets. These T-CNNs were then used as embedding extractors on all recordings from their respective datasets in the same manner as the P-CNN. T-CNNs trained to classify from which site recordings originated were used for both the site and habitat classification tasks, given that the site of origin, but not habitat type, should be known for any similar dataset.

To improve clustering, dimensionality reduction was performed on each embedding using uniform manifold approximation (UMAP) via its associated Python package (v0.5.3)[32]. To determine which of the three embedding extraction methods was most proficient at producing UMAP embeddings that represented known properties of the data, a qualitative assessment was first performed using UMAP visualisation in two dimensions. For all six tasks performed by each method, a plot was produced where points were labelled with their known class and the fidelity of groupings to their true class compared between each method.

A quantitative assessment was also undertaken using affinity propagation clustering[63] on each UMAP embedding. This algorithm was selected as there is no requirement to predetermine the number of classes, instead the affinity propagation algorithm identifies this, simulating analysis with an entirely unlabelled dataset where number of classes is unknown. For a given method and task, the cluster to which each recording was assigned was entered into a contingency table against its known class. Models most proficient at clustering recordings from the same class, while excluding recordings from other classes, generated more discrete clusters. Those less proficient assigned recordings more randomly across clusters. A chi-squared test was performed on contingency tables to assess this, with a higher score indicating models with a higher fidelity to true classes. Processing was performed using scikit-learn (v0.22) in Python (v3.7).

Supervised classifiers

All three ML methods were set two classification tasks (habitat and individual site identity) from each of the three soundscape datasets. The habitat classification tasks were as follows: high or low coral cover sites for the Indonesian dataset, high or low fish diversity sites for the Australian dataset, and shallow or mesophotic sites for the French Polynesian dataset. For each task, recordings were split into training, validation and test sets, where 66-75% of the data was used for training and the remaining data was evenly split for validation and testing of each task (see Text S2 for specific details for each dataset and task). Validation and test accuracies were carefully monitored to ensure clear over and underfitting did not occur.

For the compound index and P-CNN embeddings, random forest ML classifiers were trained for each classification task. Random forests were selected due to their previous successful implementation in the soundscape literature[17,47], robustness to overfitting, and, low computational costs which enable calculation on personal computing devices; a likely consideration for future users (Text S4). For each task, fifty random forest classifiers were trained and then used for inference on the validation data. The instance which reported the highest validation accuracy was used for inference on the test data, and the accuracy of this reported. As classes were well balanced within each task, raw accuracy (the proportion of one-minute recordings correctly classified) was used as the performance metric. For each task, this process was repeated multiple times using different splits of the data into training, validation and test sets as initial trials showed the split impacted performance of the three embedding approaches differently. One-hundred of these repeats were performed for each task,

except for the Australian and French Polynesian habitat classification tasks, where only 32 were used as this was the maximum number of possible combinations which enabled the exclusion of entire sites from the training data (Text S2). Analysis was performed using scikit-learn (v0.22) in Python (v3.7).

For the T-CNNs, these were trained on the reef soundscape dataset for each individual task. The number of output nodes was adjusted to match the number of target classes in each task. The T-CNN was trained for 50 epochs, with inference on the validation data every epoch. The epoch which reported the highest validation accuracy was then used to inference on the test data. The most common class prediction across all 960 ms segments in a one-minute recording was used as the class prediction for each minute in the test data. This process was repeated 100 times using the same training, validation and test combinations as the compound index and pretrained T-CNN. This workflow was performed using Tensorflow (v1.15) in Python (v3.7).

Analysis of variance (ANOVA) tests were used to determine whether any significant difference between the accuracy of each embedding approach was present across all repeats performed for each task. Where significant differences were detected ($p<0.05$), post hoc Tukey tests were used to determine which methods differed; performed using the scipy library (v1.73) in Python (v3.7).

Supplementary information

Supplementary 1 contains additional details on results and methods.

Supplementary 2 contains interactive UMAP plots for each dataset. These come in '.html' format and can be opened in a web browser. Further details are specified in Supplementary 1 Text S1. Supplementary 2 can be downloaded as a zip file from: https://doi.org/10.5281/zenodo.10532242.

Code availability

All the code required to run this analysis is available at: https://doi.org/10.5281/zenodo.10532242. This code is available as Google Collaboratory notebooks which can be run from a web browser on sample data through the following link: https://github.com/BenUCL/Reef-acoustics-and-AI.

Data availability

Feature embeddings extracted from the full datasets, T-CNN predictions and a link to sample audio are available from: https://doi.org/10.5281/zenodo.10532242.

We share the full datasets of raw recordings used in this study, totalling >340 hrs. Due to their size, the data is split across three Zenodo repositories. For specific details beyond those shared in this manuscript, please contact the corresponding author (BW). The Indonesian dataset can be found in the Zenodo repository cited above. The remaining data can be found in these repositories:

Australia: https://doi.org/10.5281/zenodo.10533066

French Polynesia: https://doi.org/10.5281/zenodo.10539938

Acknowledgements

For providing the Indonesian dataset the authors would like to thank Mars Global; Hasanuddin University; the Department of Marine Affairs and Fisheries of the Province of South Sulawesi; the Government Offices of the Kabupaten of Pangkep, Pulau Bontosua and Pulau Badi; and the communities of Pulau Bontosua and Pulau Badi for their support. For the Australian dataset the authors would like to acknowledge the Dingaal Aboriginal people (original land owners of Lizard Island) and thank staff from Australian Museum Lizard Island Research Station. For supporting the collection of the French Polynesian dataset the authors would like to thank staff from the Centre de Recherches Insulaires et Observatoire de l'Environnement (CRIOBE). This work was funded by the following grants: Fisheries Society of the British Isles PhD studentship award to B.Ws; Natural Environment Research Council Research Grant to S.D.S (NE/ P001572/1); Natural Environment Research Council–Australian Institute of Marine Science CASE GW4+ Studentship (NE/L002434/1) and a Royal Commission for the Exhibition of 1851 Research Fellowship to T.A.C.L; Swiss National Science Foundation Early Postdoc Mobility fellowship to L.C (P2SKP3-181384); Agence National de la Recherche to Glenn Almany, S.C.M and R.B (ANR-14-CE02-0005-01/Stay or Go); Haut-Commissariat de la République en Polynésie Française to S.C.M (HC/3041/DIE/BPT/); Pacific Funds (BLEACH & ALAN) to S.C.M; Millenium Nucleus for the Ecology and Conservation of Temperate Mesophotic Reef Ecosystem (NUTME), Chile to R.B.

Author Information

Contributions

B.W and K.E.J conceptualised the study. T.A.C.L, J.J, M.P, L.R, L.C, E.W, A.H, R.B, S.C.M, A.H, F.Z and B.W were responsible for the design and delivery of data collection. B.W performed the analysis of the data with support from S.M.B, S.S.S, S.D.S, D.C and K.E.J. The manuscript was written by B.W with comments from T.A.C.L, L.R, L.C, A.H, R.B, S.C.M, S.S.S, S.D.S, D.C and K.E.J.

References

1. Obura, D. O. *et al.* Coral reef monitoring, reef assessment technologies, and ecosystem-based management. *Front. Mar. Sci.* **6**, 580 (2019).

2. Mooney, T. A. *et al.* Listening forward: approaching marine biodiversity assessments using acoustic methods. *R. Soc. Open. Sci.* **7**, 201287 (2020)

3. Van Hoeck, R. V. *et al.* Passive acoustic monitoring complements traditional methods for assessing marine habitat enhancement outcomes. *Ecosphere* **12**, e03840 (2021).

4. Elise, S. *et al.* Combining Passive Acoustics and Environmental Data for Scaling Up Ecosystem Monitoring: A Test on Coral Reef Fishes. *Rem. Sens.* **14**, 2394 (2022).

5. Lamont, T. A. C. *et al.* The sound of recovery: Coral reef restoration success is detectable in the soundscape. *J. Appl. Ecol.* **59**, 742-756 (2021).

6. Parsons, M. J. G. *et al. The Effects of Noise on Aquatic Life Ch. 1: A Global Library of Underwater Biological Sounds (GLUBS): An Online Platform with Multiple Passive Acoustic Monitoring Applications* (Springer, Switzerland, 2024)

7. Pijanowski, B. C., Farina, A., Gage, S. H., Dumyahn, S. L. & Krause, B. L. What is soundscape ecology? An introduction and overview of an emerging new science. *Landsc. Ecol.* **26**, 1213–1232 (2011).

8. Kennedy, E. V., Holderied, M. W., Mair, J. M., Guzman, H. M. & Simpson, S. D. Spatial patterns in reef-generated noise relate to habitats and communities: Evidence from a Panamanian case study. *J. Exp. Mar. Biol. Ecol.* **395**, 85–92 (2010).

9. Nedelec, S. L. *et al.* Soundscapes and living communities in coral reefs: Temporal and spatial variation. *Mar. Ecol. Prog. Ser.* **524**, 125–135 (2015).

10. Elise, Simon. *et al.* Assessing key ecosystem functions through soundscapes: A new perspective from coral reefs. *Ecol. Indic.* **107**, 105623 (2019).

11. Staaterman, E. *et al.* Celestial patterns in marine soundscapes. *Mar. Ecol. Prog. Ser.* **508**, 17–32 (2014).

12. Staaterman, E. *et al.* Bioacoustic measurements complement visual biodiversity surveys: Preliminary evidence from four shallow marine habitats. *Mar. Ecol. Prog. Ser.* **575**, 207–215 (2017).

13. Brandl, S. J., Goatley, C. H. R., Bellwood, D. R. & Tornabene, L. The hidden half: ecology and evolution of cryptobenthic fishes on coral reefs. *Biol. Rev.* **93**, 1846–1873 (2018).

14. Parsons, M. *et al.* Long-term monitoring of soundscapes and deciphering a usable index: Examples of fish choruses from Australia. *Proc. Mtgs. Acoust.* **27**, 010023 (2016).

15. Kim, E. B. *et al.* SoundScape learning: An automatic method for separating fish chorus in marine soundscapes. *J. Acoust. Soc. Am.* **153**, 1710–1722 (2023).

16. Lecun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

17. Bradfer-Lawrence, T. *et al.* Guidelines for the use of acoustic indices in environmental research. *Methods. Ecol. Evol.* **10**, 1796–1807 (2019).

18. Raick, X., Di Iorio, L., Lecchini, D., Bolgan, M. & Parmentier, E. “To Be, or Not to Be”: Critical Assessment of the Use of α-Acoustic Diversity Indices to Evaluate the Richness and Abundance of Coastal Marine Fish Sounds. *J. Ecoacoust.* **7**, 1 (2023).

19. Bradfer-Lawrence, T., Desjonqueres, C., Eldridge, A., Johnston, A. & Metcalf, O. Using acoustic indices in ecology: Guidance on study design, analyses and interpretation. *Methods. Ecol. Evol.* **14,** 2192-2204 (2023).

20. Stowell, D. Computational bioacoustics with deep learning: a review and roadmap. *PeerJ* **10**, e13152 (2022).

21. Parcerisas, C., Roca, I. T., Botteldooren, D., Devos, P. & Debusschere, E. Categorizing Shallow Marine Soundscapes Using Explained Clusters. *J. Mar. Sci. Eng.* **11**, 550 (2023).

22. Williams, B. *et al.* Enhancing automated analysis of marine soundscapes using ecoacoustic indices and machine learning. *Ecol. Indic.* **140**, 108986 (2022).

23. Nieto-Mora, D. A., Rodríguez-Buritica, S., Rodríguez-Marín, P., Martínez-Vargaz, J. D. & Isaza-Narváez, C. Systematic review of machine learning methods applied to ecoacoustics and soundscape monitoring. *Heliyon* **9**, e20275 (2023).

24. Lecun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

25. Wang, Y., Wei, G.-Y., Brooks, D. & Paulson, J. A. Benchmarking TPU, GPU, and CPU Platforms for Deep Learning. Preprint at: https://doi.org/10.48550/arXiv.1907.10701 (2019).

26. Kong, Q. *et al.* PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition. *IEEE/ACM Trans Audio Speech Lang Process* **28**, 2880–2894 (2020).

27. Frazao, F., Padovese, B. & Kirsebom, O. S. Workshop Report: Detection and Classification in Marine Bioacoustics with Deep Learning. Preprint at: https://doi.org/10.48550/arXiv.2002.08249 (2020).

28. Kahl, S., Wood, C. M., Eibl, M. & Klinck, H. BirdNET: A deep learning solution for avian diversity monitoring. *Ecol. Inform.* **61**, 101236 (2021).

29. Costello, M. J. *et al.* Marine biogeographic realms and species endemicity. *Nat. Commun.* **8**, 1–10 (2017).

30. Lindseth, A. V. & Lobel, P. S. Underwater soundscape monitoring and fish bioacoustics: A review. *Fishes* **3**, (2018).

31. Pieretti, N. & Danovaro, R. Acoustic indexes for marine biodiversity trends and ecosystem health. *Philos. Trans. R. Soc*. **375**, 20190447 (2020)

32. Mcinnes, L., Healy, J., Saul, N. & Großberger, L. Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint available at: https://doi.org/10.48550/arXiv.1802.03426 (2018).

33. Kaplan, M., Mooney, T., Partan, J. & Solow, A. Coral reef species assemblages are associated with ambient soundscapes. *Mar. Ecol. Prog. Ser.* **533**, 93–107 (2015).

34. Bertucci, F., Parmentier, E., Lecellier, G., Hawkins, A. D. & Lecchini, D. Acoustic indices provide information on the status of coral reefs: an example from Moorea Island in the South Pacific. *Sci. Rep.* **6**, 33326 (2016).

35. Freeman, L. A. & Freeman, S. E. S. Rapidly obtained ecosystem indicators from coral reef soundscapes. *Mar. Ecol. Prog. Ser.* **561**, 69–82 (2016).

36. Gordon, T. A. C. *et al.* Habitat degradation negatively affects auditory settlement behavior of coral reef fishes. *Proc. Natl. Acad. Sci.* **115**, 5193–5198 (2018).

37. Bohnenstiehl, D. R., Lyon, R. P., Caretti, O. N., Ricci, S. W. & Eggleston, D. B. Investigating the utility of ecoacoustic metrics in marine soundscapes. *J. Ecoacoust.* **2**, 1156 (2018).

38. Harris, S. A., Shears, N. T. & Radford, C. A. Ecoacoustic indices as proxies for biodiversity on temperate reefs. *Methods Ecol. Evol.* **7**, 713–724 (2016).

39. Bertucci, F. *et al.* Snapshot recordings provide a first description of the acoustic signatures of deeper habitats adjacent to coral reefs of Moorea. *PeerJ* **2017**, 1–16 (2017).

40. Lin, T. H. *et al.* Using Soundscapes to Assess Deep-Sea Benthic Ecosystems. *Trends Ecol. Evol.* **34**, 1066–1069 (2019).

41. Raick, X. *et al.* Fish sounds of photic and mesophotic coral reefs: variation with depth and type of island. *Coral Reefs* **42**, 285–297 (2023).

42. Kaplan, M. B., Lammers, M. O., Zang, E. & Mooney, T. A. Acoustic and biological trends on coral reefs off Maui, Hawaii. *Coral Reefs* **37**, 121–133 (2018).

43. Bertucci, F., Parmentier, E., Berten, L., Brooker, R. M. & Lecchini, D. Temporal and Spatial Comparisons of Underwater Sound Signatures of Different Reef Habitats in Moorea Island, French Polynesia. *PLoS One* **10**, e0135733 (2015).

44. Elise, Simon. *et al.* An optimised passive acoustic sampling scheme to discriminate among coral reefs' ecological states. *Ecol. Indic.* **107**, 105627 (2019).

45. Halevy, A., Norvig, P. & Pereira, F. The unreasonable effectiveness of data. *IEEE Intell. Syst.* **24**, 8–12 (2009).

46. Müller, J. *et al.* Soundscapes and deep learning enable tracking biodiversity recovery in tropical forests. *Nat. Commun.* **14**, (2023).

47. Sethi, S. S. *et al.* Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set. *Proc. Natl. Acad. Sci.* **117**, 17049–17055 (2020).

48. Gordon, T. A. C. *et al.* Acoustic enrichment can enhance fish community development on degraded coral reef habitat. *Nat Commun* **10**, 1–7 (2019).

49. Simpson, S. D., Meekan, M., Montgomery, J., McCauley, R. & Jeffs, A. Homeward Sound. *Science* **308**, 221 (2005).

50. Pysanczyn, J. W. *et al.* The role of acoustics within the sensory landscape of coral larval settlement. *Front. Mar. Sci.* **10,** 1111599 (2023).

51. Borowiec, M. L. *et al.* Deep learning as a tool for ecology and evolution. *Methods Ecol. Evol.* **13** 1640–1660 (2022).

52. Pichler, M. & Hartig, F. Machine Learning and Deep Learning-A review for Ecologists. *Methods Ecol. Evol.* **14**, 994-1016 (2023).

53. Christin, S., Hervet, É. & Lecomte, N. Applications for deep learning in ecology. *Methods Ecol. Evol.* **10**, 1632–1644 (2019).

54. Ghani, B., Denton, T., Kahl, S. & Klinck, H. Global birdsong embeddings enable superior transfer learning for bioacoustic classification. *Sci. Rep.* **13**, 22876 (2023).

55. Lamont, T. A. C. *et al.* HydroMoth: Testing a prototype low-cost acoustic recorder for aquatic environments. *Rem. Sens. Ecol. Conserv.* **8**, 362–378 (2022).

56. Chapuis, L., Williams, B., Gordon, T. A. C. & Simpson, S. D. Low-cost action cameras offer potential for widespread acoustic monitoring of marine ecosystems. *Ecol. Indic.* **129**, 107957 (2021).

57. Kohavi, R. & Sommerfield, D. Feature Subset Selection Using the Wrapper Method: Overfitting and Dynamic Search Space Topology. *KDD* **95**, 192–197 (1995).

58. Curtis, K. R., Howe, B. M. & Mercer, J. A. Low-frequency ambient sound in the North Pacific: Long time series observations. *J. Acoust. Soc. Am.* **106**, 3189–3200 (1999).

59. Ulloa, J. S., Haupert, S., Latorre, J. F., Aubin, T. & Sueur, J. scikit-maad: An open-source and modular toolbox for quantitative soundscape analysis in Python. *Methods Ecol. Evol.* **12,** 2334-2340 (2021).

60. Sethi, S. S. *et al.* Soundscapes predict species occurrence in tropical forests. *Oikos* **3**, e08525 (2021).

61. Hershey, S. *et al.* CNN architectures for large-scale audio classification. *ICASSP* 131–135 (2017).

62. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. Preprint at: https://doi.org/10.48550/arXiv.1409.1556 (2014).

63. Frey, B. J. & Dueck, D. Clustering by Passing Messages Between Data Points. *Science* **315**, 972-976 (2007).