PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Check for updates

Cite this article: Barroso VR, Lessa AA, Ferreira CEL, Xavier FC. 2025 Optimal feature selection and model explanation for reef fish sound classification. *Phil. Trans. R. Soc. B* **380**: 20240055. https://doi.org/10.1098/rstb.2024.0055

Received: 10 June 2024
Accepted: 7 January 2025

One contribution of 11 to a theme issue 'Acoustic monitoring for tropical ecology and conservation'.

Subject Areas:
ecology, environmental science, ecosystem

Keywords:
bioacoustics, acoustic ecology, supervised learning, SHAP, XAI

Author for correspondence:
Viviane R. Barroso
e-mail: vivianerbarroso@gmail.com

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.7837885.

THE ROYAL SOCIETY PUBLISHING

Optimal feature selection and model explanation for reef fish sound classification

Viviane R. Barroso[1], Alexia A. Lessa[2], Carlos E. L. Ferreira[3] and Fabio C. Xavier[1]

[1]Marine Biotechnology Program, Instituto de Estudos do Mar Almirante Paulo Moreira, Arraial do Cabo, Rio de Janeiro 28930-000, Brazil
[2]Programa de Pós-Graduação em Ecologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ 21941-902, Brazil
[3]Departamento de Biologia Marinha, Universidade Federal Fluminense, Niterói, Rio de Janeiro 24210-201, Brazil

VRB, 0000-0003-0534-2432

Fish produce a wide variety of sounds that contribute to the soundscapes of aquatic environments. In reef systems, these sounds are important acoustic cues for various ecological processes. Artificial intelligence methods to detect, classify and identify fish sounds have become increasingly common. This study proposes the classification of unknown fish sounds recorded in a subtropical rocky reef using different feature sets, data augmentation and explainable artificial intelligence tools. We used different supervised algorithms (naive Bayes, random forest, decision trees and multilayer perceptron) to perform a multiclass classification of four classes of fish pulsed sounds. The proposed models showed excellent performances, achieving 98.1% of correct classification with multilayer perceptron using data augmentation. Explainable artificial intelligence allowed us to identify which features contributed to predict each sound class. Recognizing and characterizing these sounds is key to better understanding diel behaviours and functional roles associated with critical reef ecological processes.

This article is part of the theme issue 'Acoustic monitoring for tropical ecology and conservation'.

1. Introduction

Fish produce a wide variety of sounds, contributing to the soundscapes of aquatic environments. Communication sounds are used in different behavioural contexts, such as reproduction (courtship and mating), aggregation, feeding and agonistic interactions [1,2]. Sounds produced by fish have distinct spectral and temporal features, with many species producing pulsed and repetitive sounds [3,4]. These features can enable species identification, behavioural patterns and reproductive and spawning conditions [5]. In reef systems, sounds are important acoustic cues for different ecological processes, including larvae orientation and settlement and habitat use [6].

A common approach to study fish sounds is Passive Acoustic Monitoring (PAM), which allows continuous and long-term assessment of soundscapes [7,8]. However, these long-term recordings are producing an increasing amount of data, which makes manual analysis unfeasible and requires robust computational methods [9]. Therefore, the use of artificial intelligence (AI) to detect, classify and identify fish sounds is becoming increasingly common [10]. Although machine learning and deep learning have gained attention in this field in recent years, the interpretability of the models remains a gap.

One of the most challenging tasks concerning the classification problem using learning-based approaches is to explain the predictions generated by the models. The algorithms often operate like 'black boxes', and it is difficult to understand how the models achieve such performance, and which parameters were responsible for predicting each class [11]. Explainable Artificial Intelligence (XAI) has received increasing attention in various fields such as medicine [12], remote sensing [13], soundscape ecology [14] and many others. This is a method inspired in game theory that provides explanations for machine learning models by computing the contribution of each feature to the prediction [11]. This technique allows us to understand which features are important and how they specifically impact individual predictions of the model. Therefore, XAI provides the interpretability of the models aligned to the specificities of the studied phenomenon/problem.

Feature selection is a fundamental step in sound classification using machine learning [15] as it helps to reduce dimensionality and influence the accuracy of the model [16,17], and it must reflect the specificities of the problem domain. Thus, understanding which features are more effective to represent fish sounds can significantly improve the detection, classification and identification tasks. However, the study of feature selection is still in its early stages in the field of fish sounds, and it is possible that currently underrated features could be good descriptors of these sounds. Timbral texture features (TTFs), including the Mel Frequency Cepstral Coefficients (MFCCs), are widely used in music and human speech recognition. MFCCs provide a robust representation of sound structures [18] and have also been used in recent studies to detect and classify fish sounds [19–21]. However, few studies have demonstrated the individual contribution of each coefficient in predicting different sounds.

In this paper, we classified unknown fish sounds from subtropical rocky reefs in Arraial do Cabo (23°44′S–42°W), Brazil. Four supervised algorithms, naive Bayes (NB) [22], random forest (RF) [23], decision trees (DT) [24] and multilayer perceptron (MLP) [25] were used to classify four classes of fish pulsed sounds. Different feature sets were used, where the contribution of features for each class and the SHapley Additive exPlanations (SHAP) values were computed to understand the influence of the features on the classifiers. To the best of our knowledge, this is one of the first studies to use XAI to explain the classification model and assess the relationships between the features, allowing us to understand which features better describe each class of reef fish sounds.

2. Material and methods

(a) Data acquisition

We used acoustic data from the long-term monitoring of the BIOCOM project [26]. The study was conducted on the island of Cabo Frio, Arraial do Cabo (23°44′S–42°W), Rio de Janeiro, south-eastern Brazil. The region is characterized by extensive subtropical rocky shores influenced by seasonal upwelling events, which imparts unique characteristics of a clash between tropical and warm-temperate biodiversity [27]. Local upwelling brings intermediate deep waters with temperatures below 18° to shallow habitats, enriching trophic webs and promoting significant local hydrobiological variability [28,29]. The combination of the upwelling regime and the diversity of reef habitats favours a rich fish community, with tropical and subtropical components [30,31].

Acoustic data were collected using a recording system consisting of a stainless-steel pyramid structure with a 4-channel hydrophone (Hyd TP-1, Marsensing Ltda) deployed at a depth of 7.55 m and approximately 5 m from the local rocky shore. The recorder was configured with a sampling frequency of 52 734 Hz, 24-bit resolution and a duty cycle of 20% (i.e. a 1 min recording every 5 min). The sensitivity of the hydrophone was −174.9 dB re 1 μPa, with a flat response between 0.1 and 26 367 Hz. (For a brief description of the local soundscape, see electronic supplementary material, ‘Methods’.) [26,32].

(b) Exploratory data analysis and preprocessing

To classify fish sounds, we selected sound files during different seasons of the year from the night, dusk and dawn periods, as these are known for fish acoustic activity [33–35]. We grouped the sounds into four classes (figure 1) that were most common in the dataset. The classes are low-frequency sounds consisting of sequences of pulses: Class 1 is a pulse train with 3 to 9 pulses, call duration between 177 and 1387 ms, irregular inter-pulse intervals (IPIs) and a frequency range of 100−500 Hz; Class 2 is a pulse train with 4 pulses, call duration between 157 and 347 ms, almost regular IPIs and a frequency range of 100−600 Hz; Class 3 is a pulse train with 4 to 13 pulses, call duration between 195 and 910 ms, irregular IPIs and a frequency range of 100−400 Hz; and Class 4 is a pulse train with 3 to 7 pulses, with call duration between 216 and 939 ms, irregular IPIs and a frequency range of 100−700 Hz. (For more details regarding the sound classes see electronic supplementary material, ‘Methods’) [36–42].

Prior to the feature extraction step, the signals were pre-processed. Since the target sounds were of low frequency, filters were applied. For feature set 1, the raw sound files were filtered using the Raven Pro v. 1.6 [43] bandpass filter tool with a lower limit = 100 Hz and an upper limit = 1000 Hz. For feature set 2, we developed a Python code to detect the target sounds according to their start and end times, previously extracted in feature set 1. In this step, the signals were downsampled (from 52 734 Hz to 5273 Hz) and filtered by frequency using an elliptic (Cauer) bandpass filter from the SciPy library. The filter was configured with a maximum ripple of 1 dB, a minimum attenuation of 20 dB and an optimal filter order, based on frequency limits for each signal.

royalsocietypublishing.org/journal/rstb *Phil. Trans. R. Soc. B* **380:** 20240055

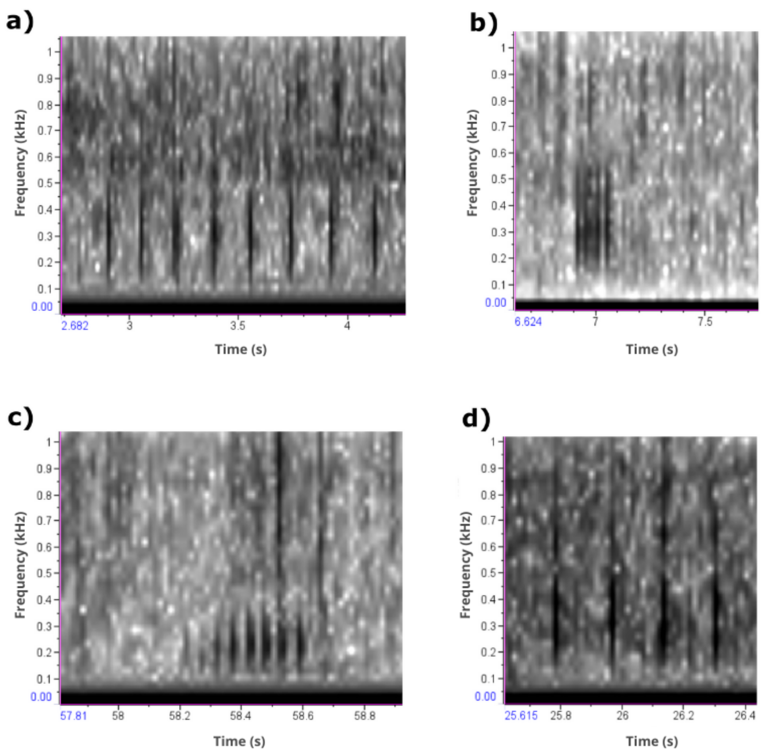

Figure 1. Spectrograms of the four classes of reef fish sounds. The four classes are low-frequency sounds, consisting of sequences of pulses. (a) Class 1, (b) Class 2, (c) Class 3, (d) Class 4.

(c) Feature extraction

With the aim of identifying an optimal feature set to classify the signals, we proposed three different feature sets, totalling 40 features from different domains (time, frequency, amplitude and cepstral domain). The first set has a total of nine features, consisting of spectral features (high frequency, low frequency, peak frequency and centre frequency), temporal features (call duration, pulse duration, interpulse interval and number of pulses) and energy. To build this set, we manually extracted features by inspecting both oscillograms and spectrograms using Raven Pro v. 1.6 (window size = 1768 fft, overlap = 50%, colour scale = 'Greyscale'). Energy and spectral parameters were extracted directly using Raven functions, while temporal parameters were extracted by using various selections in the signals and extracting 'delta time' [44].

To explore TTFs, we used the librosa Python package [45], which has specialized functions for this purpose. To do this, we developed a code to detect fish signals (previously manually selected in the spectrograms). Leveraging librosa functions, we also extracted tempo (beats) to obtain possible information about the underlying rhythm of these signals. For this set, we selected 31 features (spectral centroid, spectral rolloff, zero crossing rate, spectral flux, tempo and 26 MFCCs) [46–48]. The novelty of this approach is to evaluate the contribution of the MFCCs by interpreting the coefficients that were more important for each class. MFCCs provide information regarding energy distribution on the spectral envelope of signals [46]. Each coefficient corresponds to a specific frequency band, but in a perceptual scale (the Mel scale). Thus, MFCC1 can be interpreted as the total energy obtained, MFCC2 as the weighted ratio of the energy of a low to a high frequency band, MFCC3 as the ratio of mid to low + high frequencies, MFCC4 as the ratio of two specific bands to two higher frequency bands, and so on [49]. It is worth mentioning that the Mel scale provides higher resolution in the lower frequency bands, as it has an almost linear relationship with the Hertz scale below 1 KHz [50].

Feature set 3 is a combination of sets 1 and 2, with 40 features. The description of the features used in each set is shown in electronic supplementary material, table S1. In addition to the three main feature sets, we tested other combinations of features by domain (cepstral, spectral and temporal) to investigate their contributions. These sets were (i) MFCCS only (n = 26), (ii) spectral features only (n = 8), (iii) temporal features only (n = 5), (iv) TTFs only (n = 30) (electronic supplementary material, table S2).

(d) Data augmentation

Data augmentation is a technique commonly used to train machine learning algorithms on small and/or unbalanced datasets [51]. Considering this, we selected only a few samples of fish calls to test the contribution of data augmentation applied to tabular data to the performance of the classifiers. Therefore, a small dataset of fish sounds was used to assess the effectiveness of data augmentation on tabular data. This is important because in many situations larger datasets of fish sounds are not available. The original datasets contained 120 samples (30 samples for each class). We augmented each dataset by using the standard

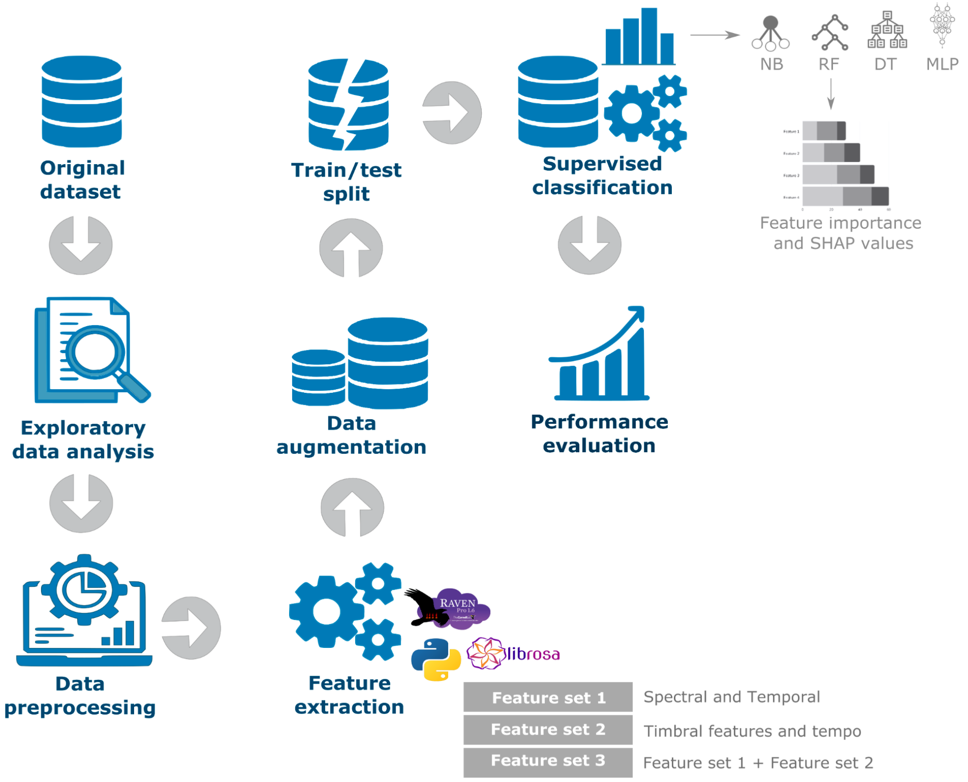

Figure 2. Workflow for the feature selection and classification processes for reef fish sounds.

deviations limits (summing and subtracting of each sample). After augmentation, the datasets had 360 samples (90 samples for each class). (For more details of the data augmentation process, see electronic supplementary material, 'Methods'.)

(e) Classification model

After the feature extraction process, fish sounds were classified using three supervised algorithms (NB, RF and DT) and a supervised neural network (MLP). With the aim of removing less important features (dimensionality reduction), we generated a random variable with a uniform distribution and added it to the feature vector to train an RF estimator to assess feature importance prior to the classification process. However, no feature was ranked below the random variable, and we assumed that all features were important. Tabular data of each feature set were used as input for these four algorithms. The classifiers were implemented using the Scikit-Learn Python package [52]. To train the supervised algorithms, the dataset was randomly split into 70% for training and 30% for testing. The supervised algorithms were run under Scikit-learn default parameters for all classifiers. The number of estimators for RF was set to 100. The MLP architecture developed for this work consisted of four hidden layers. The selected activation function was the Rectified Linear Unit (ReLU). The selected solver for weight optimization during the training step was the Adaptive Moment Estimation (*Adam*) algorithm. The regularization parameter (*alpha*), which helps to prevent overfitting, was set to 0.1. Training was performed over 200 epochs with a batch size of 8. For MLP, the dataset was also divided into 70% for training and 30% for testing.

Four metrics were used to evaluate the performance of the classification algorithms: accuracy, precision, recall and f1-score [20,53]. Accuracy is the ratio of the number of correct predictions to the total number of predictions, and it measures the overall correctness of a classification model; precision is calculated as the ratio of true positives to the total number of positive predictions made by the classifier, and it measures the correctness of positive predictions; recall is the ratio of the number of true positive predictions to the number of positive observations in the data; f1-score is the weighted average of precision and recall [54,55]. We also used a k-fold cross-validation method ($k = 5$) to evaluate the robustness of the classification model and the data augmentation effects. These results are presented in electronic supplementary material, table S3.

(f) Model explanation

To provide a comprehensive explanation of the proposed classifiers, we used SHAP. We used the shapExplainer algorithm from the SHAP Python package [56] in our RF classifier for the augmented datasets to compute the feature importance for each class and the impact on the model outputs. Figure 2 represents the workflow for the whole sound classification process, including the preprocessing, feature extraction and data augmentation steps [11,56,57] (For more details of the data augmentation process see electronic supplementary material, 'Methods'.)

3. Results

We have evaluated the performance of the models to determine their effectiveness in handling different feature representations, as well as the effects of data augmentation. Four metrics were used to evaluate the classification models (accuracy, precision, recall and f1-score), as shown in table 1. We observed that all classifiers had a satisfactory performance for all feature sets. In general, we observed the best results for RF and MLP, followed by NB for all feature sets. DT showed a less satisfactory performance; however, it performanced well for feature set 1, with an accuracy of 91.7%and precision, recall and f1-scores of 92.0%.

The data augmentation had a significant positive impact on the classifiers, as it increased the performance metrics. For instance, while the accuracy of the original feature set 3 was 94.4%, 98.1% of the calls were correctly classified by MLP after augmentation. The NB, MLP and RF classifiers were the most sensitive to data augmentation, while DT had less influence on its effects. Classification accuracy increased from 80.0% to 93.3% and from 88.9% to 93.5% for NB and RF from feature set 1, respectively. The cross-validation used to test our classifiers shows that the proposed architectures are satisfactory, with, for example, 95% ± 0.6% of correct classification by MLP (electronic supplementary material, table S3).

Regarding the feature sets, the third one was the most influential for all classifiers, reaching 98.1%accuracy for MLP, after augmentation. This set was the largest, with 40 features, and was the optimal feature set. Feature set 1 was the second most influential, with 95.4% of accuracy for the same classifier, after augmentation. The lowest performance for all classifiers was for feature set 2 compared with the other sets, but the correct classifications were superior to 75% for RF and MLP. It is worth mentioning that this set had a significant increase after augmentation, reaching 85.2%, 88.9%, 78.7% and 90.7% accuracy for NB, RF, DT and MLP, respectively. For the other features tested (electronic supplementary material, table S2), we observed that the TTFs set showed the best performance. The feature set with the second best performance was the MFCCs, followed by the spectral set. Temporal features presented the lowest performance metrics for the classifiers, but they increased significantly after augmentation.

We computed the SHAP values and the influence of each feature in predicting each class using the RF classifier for each set. The performance metrics (accuracy, precision, recall and f1-score) were also computed for each class for each feature set (electronic supplementary material, table S4). Accuracy was above 92% for all classes in sets 1 and 3, and above 86% for set 2, with Class 1 showing the highest accuracy in each set. For feature set 3—the optimal set—the most important features were interpulse interval (IPI), high-frequency, MFCC5, call duration and pulse duration (figure 3). In this feature set, we can see that high frequency and MFCC5 were the most important features for Class 3. IPI and call duration had a large contribution for Class 2. IPI and high frequency also contributed in predicting Classes 1 and 4.

Figure 4 shows the polar plots for the SHAP values for each class for feature set 3, highlighting the features with the most significant contributions to the classification. For Class 1, IPI, MFCC7 and MFCC10 emerged as the key drivers, indicating the strong influence of these features in predicting this class. For Class 2, IPI, call duration and pulse duration were the most important features in distinguishing this class. High frequency, MFCC5 and MFCC8 were most influential in predicting Class 3, whereas IPI, high frequency, MFCC7 and pulse duration were most influential in predicting Class 4. These results highlight the different roles that specific features play in the predictions made by the model. Electronic supplementary material, figure S1 provides a different type of visualization for the summary of SHAP values for feature set 3, showing how the low or high values of the features impacted the outputs of the model. SHAP summary plots for each class for features set 1 and 2 are in the electronic supplementary material, figures S2 and S3, respectively.

For TTFs, we used a dataset with MFCCs, spectral centroid, spectral rolloff, zero-crossing rate (ZCR) and spectral flux. The accuracy for this set was 85.2%, 88.9%, 77.8% and 91.7% after augmentation for NB, RF, DT and MLP, respectively (electronic supplementary material, table S2). This feature set showed the best performance among the other features tested. The main contributors to this performance were the MFCCs 8, 7, 5, as well as spectral centroid, ZCR and spectral rolloff. Spectral centroid influenced the prediction of Class 2, and ZCR and spectral rolloff contributed to the prediction of Class 1 (electronic supplementary material, figure S4). Spectral flux was not relevant for this set. However, for the set combining all of the spectral features, spectral flux was more important than centre and peak frequency (electronic supplementary material, figure S5).

4. Discussion and conclusions

The results presented in this work show the effectiveness of the model in classifying unknown reef fish sounds across the feature spaces considered. The proposed classifiers were able to discriminate classes with similar temporal patterns (actually the same call type: pulse trains). Previous work has trained supervised algorithms to detect, identify or classify fish sounds using different algorithms (such as k-nearest neighbours, random forest and support vector machine), architectures and feature sets [19,20,58,59]. Their best performances were, respectively, 96.0%, 95.6%, 96.9% and 82.7%. Compared with the performance reported in these works, our results are promising, as the proposed models achieved similar performance. It is worth noting that we proposed a simpler approach with very small datasets and achieved excellent results, especially after augmentation.

MLP was the best classifier for these sounds, reaching an accuracy of 98.1% after augmentation for the optimal feature set. This classifier also showed high precision, recall and f1-score, which means a low number of false positives and false negatives [60]. This is a robust algorithm, capable of estimating nonlinear functions and with the ability to learn hierarchical representations of data [61,62]. RF was also a robust classifier, showing high accuracy for feature sets 1 and 3, especially after augmentation. This is an ensemble classifier with high accuracy and statistically robust results and it is currently one of the most powerful supervised algorithms [63,64]. MLP and RF classifiers showed similar performance for the original datasets. The

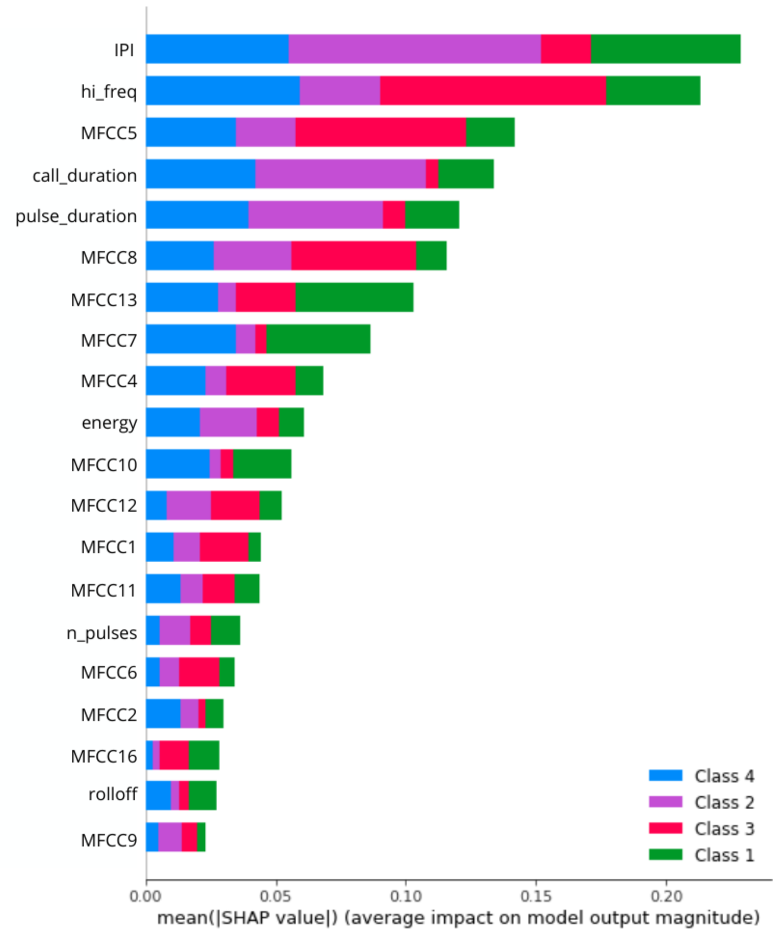

Figure 3. Summary of the contributions of features for each sound class for feature set 3.

Table 1. Performance of supervised classification for the main original and augmented feature sets. The performance metrics are Ac (% accuracy), Pc (% precision), Rc (% recall) and f1 (% f1-score). Values in bold indicate the best performances.

classifier	feature set	original				augmented			
		Ac	Pc	Rc	f1	Ac	Pc	Rc	f1
NB	**1**	80.6	89.1	82.3	82.2	**93.5**	**93.5**	93.4	93.2
	2	80.6	77.8	77.4	76.7	**85.2**	85.1	84.4	84.5
	3	91.7	91.7	92.0	90.5	**95.4**	95.1	95.0	95.1
RF	**1**	88.9	90.1	89.9	90.0	**93.5**	93.1	93.0	93.0
	2	77.8	73.9	74.6	74.0	**88.9**	88.8	88.3	88.3
	3	83.3	86.4	86.6	84.1	**92.6**	92.4	91.9	92.1
DT	**1**	91.7	92.0	92.0	92.0	**92.6**	92.0	**92.6**	92.2
	2	63.3	63.5	63.0	62.2	**78.7**	77.7	77.7	77.6
	3	72.2	70.6	73.7	70.6	**88.9**	88.7	88.2	88.3
MLP	**1**	88.9	91.5	88.0	89.0	95.4	95.4	**95.6**	95.5
	2	75.0	78.6	75.2	74.3	90.7	**90.8**	90.5	90.4
	3	94.4	94.2	93.3	93.6	98.1	98.1	**98.2**	98.1

high performance of RF can be explained by the ability of tree-based classifiers to handle tabular data. In fact, these models are known to outperform deep learning for this type of data [65]. However, after augmentation, MLP outperformed RF, as this technique optimizes the performance and robustness of deep learning models [66]. On the other hand, DT was not a good classifier for this type of fish sound, as it had a lower performance for the optimal feature set and was the least sensitive to data augmentation. This classifier has some limitations such as instability, where changes in any variable or in the size of the training data can affect the classification rules [67,68]. Although NB showed higher performance, this algorithm may not be appropriate for this problem given the complexity of the fish sounds. Although it learns fast, it is a simpler algorithm, based

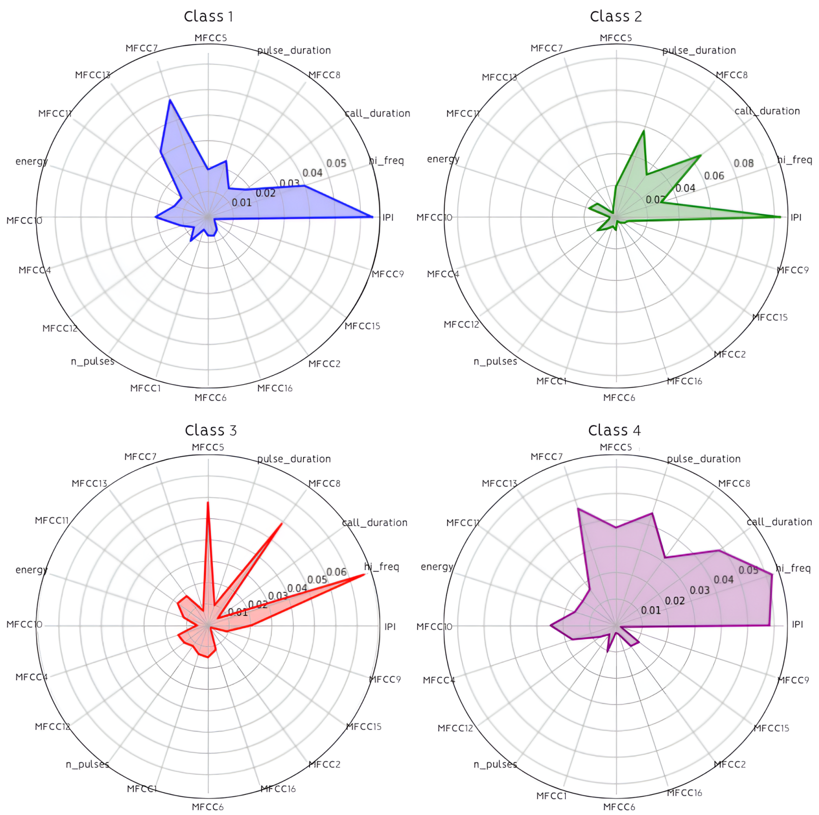

Figure 4. Polar plot for the summary of the SHAP values for each class for feature set 3. The amplitude of the polygons indicates the importance of features.

on probability, which considers all variables as independent [69]. Thus, the results show the importance of testing different classifiers, according to the specificities of the datasets.

Our results also demonstrated that data augmentation had an expressive influence on the classifiers, improving their performance. Data augmentation has been widely used in AI, especially in deep learning, which requires large datasets for training. This method is more commonly applied directly on images or raw audio files to significantly increase the size of datasets [51,70]. Ibrahim *et al*. [71] have used data augmentation directly on images of spectrograms to increase their dataset of grouper species. Waddell *et al*. [72] also augmented their dataset of images of six different fish calls to optimize their classifier. Due to the complex structure of tabular data, data augmentation for this type of dataset is not well established and must be domain-specific [66,73]. However, our approach using a comparatively small increase in the number of samples of tabular data (33%) could significantly improve the model performance. It is worth noting that in this work we used a balanced dataset (30 samples per class) to ensure that the classifiers were not biased towards a particular class. In addition, we performed cross-validation on both the original and augmented datasets, so the models could handle unseen data. The increase in performance demonstrated that this type of augmentation can be used in situations where large datasets of images or sound files are not available.

Feature set 3, with 40 features, was the optimal feature set with the best results (94.4% accuracy for the original and 98.1% for the augmented dataset with the MLP classifier). This is consistent with the results of [20], where the best results (96.9% and 96.5% for the random forest and support vector machine classifiers, respectively) were for their set with the highest number of features (84 features). In terms of the effect of data augmentation, feature set 2 was the most sensitive as it had the lowest values for the performance metrics, but after augmentation, the performance of the models for this set increased significantly. This can be explained by the small size of this set to train comparatively more complex features, thus requiring more data samples to train the classifiers. Regarding the other feature sets tested (electronic supplementary material, table S2), the temporal feature set showed the lowest performance. Despite being the same call type (pulse trains), it is worth mentioning that classes 1, 3 and 4 varied in the number of pulses within the classes, and not only between classes, which increased the complexity of the classification. However, performance increased after augmentation, as the models were able to learn with more examples.

The results of the feature contributions for each class for feature set 3 using the SHAP Explainer showed that Class 3 was most influenced by high frequency, as it had the lowest frequency of the classes tested (figure 3). For this class, high frequency acted as a 'threshold' as it had the lowest frequency of all classes. As high frequency and MFCC5 influenced the prediction for Class 3, we generated a dependence plot for this class (electronic supplementary material, figure S6) to investigate possible interaction effects of these two variables. The plot shows the influence of high frequency on the vertical distribution of SHAP values for MFCC5. Higher values of high frequency have a small relationship with MFCC5, indicating that this class, characterized by a low frequency band, is related to this coefficient. In general, MFCCs were important in predicting Class 3, which may be related to the distinct characteristics of the spectral envelope in the frequency band of this class. IPI and call duration influenced the predictions of Class 2. In fact, this class was the only one without within-class variation, with a very clear temporal pattern (always four pulses), making IPI and call duration good descriptors for this class. The dependence plot for this class for IPI and call duration (electronic supplementary material, figure S7) shows that higher values of call duration are associated with higher values of IPI.

For feature set 1, high frequency, IPI and call duration were the most important features. High frequency had more influence in predicting Classes 3 and 4, while IPI and call duration had more influence in predicting Class 2 (electronic supplementary material, figure S8). In feature set 2, we can observe that the most important features were the MFCCs 7, 5 and 8. Although the temporal feature set had a lower performance, IPI, pulse duration and call duration showed significant contribution in predicting Class 1. For feature set 3, the summary plot for Class 1 (electronic supplementary material, figure S1a) shows that higher values of these features were one of the most important features in predicting this class. In general, the MFCCs had more influence in predicting Class 3 (electronic supplementary material, figure S1c), followed by Classes 4 and 1, and less influence on the prediction of Class 2 (electronic supplementary material, figure S9). Regarding the spectral feature set, it is possible to observe that high frequency and TTFs (spectral rolloff, spectral centroid and zero crossing rate) were the most important in predicting classes (electronic supplementary material, figure S5). The most important MFCCs by class were MFCC5, MFCC8 and MFCC4 for Class 3 (100–400 Hz); MFCC8 for Class 2 (100–600 Hz); MFCC13 for Class 1 (100–500 Hz); and MFCC7 for Classes 1 and 4 (100–500 and 100–700 Hz, respectively). Higher-order MFCCs (i.e. >14) were not significant in discriminating between classes.

Owing to the effectiveness of MFCCs in lower frequencies, these features are suitable for most fish sounds. This explains the recent popularity of MFCCs in the field of fish sounds. Previous studies have included MFCCs in their feature sets to detect, classify and identify fish sounds [19–21,58,59,74–76], achieving satisfactory results for their models. However, they have not provided a detailed discussion on the contribution of each coefficient for specific sound classes. The individual analysis of each coefficient may offer a more refined spectral representation of the signals. TTFs have also been explored in previous work for fish sound detection and classification [21,74,76]. For example, Ruiz-Blais *et al.* [76] used statistics of different timbral features to detect sciaenid species, achieving 89% correctness for their set with all features tested, after post-processing. In their work, they found that cepstral coefficients, spectral centroid and zero-crossing rate were the most relevant. Our results are consistent with theirs because the same features were most relevant, but we achieved 91.7% of accuracy for the augmented set for the MLP classifier. These results demonstrated that TTFs were also good descriptors for these sound types (see electronic supplementary material, table S2 and figure S4).

Fish are known to use sound for many purposes, including mating, territorial defence, feeding, schooling behaviour and many others [2,39]. However, a current limitation of fish sound classification is the lack of data to compare and match unknown sounds to specific species. Although over 1000 species from 175 families (including Haemulidae, Holocentridae, Batrachoididae, Serranidae, Sciaenidae, Pomacanthidae, Pomacentridae, among many others) are known to produce sounds [6,36,77], the repertoire of most soniferous species has not been totally characterized [78]. This gap in the acoustic ecology of fish is still a challenge for an ecological application of PAM in terms of identification and characterization of fish biodiversity patterns and to understand aspects of species behaviour. Although it was not possible to identify the species producing the sounds used in this paper, our results demonstrate the effectiveness of artificial intelligence in classifying pulsed fish sounds. Classes 2 and 3 show similarities to sounds described for the Pomacentridae family, especially those produced by the genus *Stegastes* [79,80]. In summary, for those classes of sounds possibly produced by *Stegastes*, the best features that described these sounds were IPI, call duration and pulse duration (Class 2) and high frequency, MFCC5 and MFCC8 (Class 3). Given that some species in this genus are known for their territorial behaviour and the crucial role that sound plays in agonistic interactions related to territory and nest defence in coral reefs [80], advances in the detection of these sounds are crucial. Improved sound classification can provide valuable insights into the dynamics of this keystone group and enhance our understanding of their ecological roles and interactions within their habitats.

Explainable AI allowed us to identify which features contributed to each sound class. Introducing the interpretability of the models in the complex problem of fish sound classification is an important step, as it provides insights into which features shape a particular sound type. In addition, it is important to understand the mechanisms that lead the model to achieve certain performance metrics in order to assess its reliability [81]. XAI also enabled us to evaluate the impact of each MFCC in predicting each class. MFCCs are widely used in speech recognition to study emotions—and even diseases (such as depression)—with distinct coefficients used as indicators of specific disorders [49,82]. For fish, these may be important descriptors for recognizing species-specific sounds or indicators of behaviours. Recognizing and characterizing these sounds are key to better understanding diel behaviours and functional roles linked to critical reef ecological processes [83]. This is a significant advance because most of the studies using AI to detect and classify fish sounds have focused only on target species and the majority of these sounds remain unidentified and uncharacterized in current research [78,84]. There is therefore a growing need to develop reliable detectors and classifiers that are capable of interpreting unknown fish sounds. By focusing on the development of such advanced detectors and classifiers, we can improve our ability to categorize and understand a wider range of fish acoustic signals. The ultimate knowledge of the spatial and temporal patterns of biophony and associated underwater soundscape are the next generation of data for the optimal management and conservation of marine ecosystems and associated biodiversity [85].

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. Codes and datasets available at: [86]

Supplementary material is available online [87].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. V.R.B.: conceptualization, data curation, formal analysis, investigation, methodology, project administration, visualization, writing—original draft; A.A.L.: investigation, visualization, writing—original draft, writing—review and editing; C.E.L.F.: conceptualization,

funding acquisition, investigation, resources, supervision, writing—review and editing; F.C.X.: conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. This study was funded in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001. C.E.L.F. is continously supported by grants from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

Acknowledgements. We thank the BIOCOM team, especially Dr. Sérgio Jesus (Universidade do Algarve) and Dr. Eduardo Fagundes Netto (Instituto de Estudos do Mar Almirante Paulo Moreira—IEAPM), for providing the acoustic dataset used in this paper. We also would like to thank the reviewers; we sincerely appreciate the constructive feedback, which contributed to improving the manuscript.

References

1. Kasumyan AO. 2009 Acoustic signaling in fish. *J. Ichthyol*. **49**, 963–1020. (doi:10.1134/s0032945209110010)
2. Ladich F. 2019 Ecology of sound communication in fishes. *Fish Fish.* **20**, 552–563. (doi:10.1111/faf.12368)
3. Wall C, Simard P, Lembke C, Mann D. 2013 Large-scale passive acoustic monitoring of fish sound production on the West Florida Shelf. *Mar. Ecol. Prog. Ser*. **484**, 173–188. (doi:10.3354/meps10268)
4. Putland R, Mackiewicz A, Mensinger A. 2018 Localizing individual soniferous fish using passive acoustic monitoring. *Ecol Inf*. **48**, 60–68. (doi:10.1016/j.ecoinf.2018.08.004)
5. Bolgan M, Crucianelli A, Mylonas CC, Henry S, Falguière JC, Parmentier E. 2020 Calling activity and calls' temporal features inform about fish reproductive condition and spawning in three cultured *Sciaenidae* species. *Aquaculture* **524**, 735243. (doi:10.1016/j.aquaculture.2020.735243)
6. Lobel PS, Kaatz IM, Rice AN. 2010 Acoustical behavior of coral reef fishes. In *Reproduction and sexuality in marine fishes: patterns and processes* (ed. K Cole), pp. 307–386. Oakland, CA: California Scholarship Online. (doi:10.1525/california/9780520264335.003.0010)
7. McWilliam JN, McCauley RD, Erbe C, Parsons MJG. 2017 Patterns of biophonic periodicity on coral reefs in the Great Barrier reef. *Sci. Rep.* **7**, 17459. (doi:10.1038/s41598-017-15838-z)
8. Lima JAM *et al*. 2024 The Brazilian Santos basin underwater soundscape monitoring project (PMPAS-BS). *Front. Mar. Sci*. **11**, 1416590. (doi:10.3389/fmars.2024.1416590)
9. Bianco MJ, Gerstoft P, Traer J, Ozanich E, Roch MA, Gannot S, Deledalle CA. 2019 Machine learning in acoustics: theory and applications. *J. Acoust. Soc. Am*. **146**, 3590–3628. (doi:10.1121/1.5133944)
10. Barroso VR, Xavier FC, Ferreira CEL. 2023 Applications of machine learning to identify and characterize the sounds produced by fish. *ICES J. Mar. Sci*. **80**, 1854–1867. (doi:10.1093/icesjms/fsad126)
11. Molnar C. 2019 *Interpretable machine learning: a guide for making black box models explainable*, 2nd edn. See https://christophm.github.io/interpretable-ml-book/.
12. Hulsen T. 2023 Explainable artificial intelligence (XAI): concepts and challenges in healthcare. *AI* **4**, 652–666. (doi:10.3390/ai4030034)
13. Ishikawa S-n, Todo M, Taki M, Uchiyama Y, Matsunaga K, Lin P, Ogihara T, Yasui M. 2023 Example-based explainable AI and its application for remote sensing image classification. *Int. J. Appl. Earth Obs. Geoinf.* **118**, 103215. (doi:10.1016/j.jag.2023.103215)
14. Parcerisas C, Roca IT, Botteldooren D, Devos P, Debusschere E. 2023 Categorizing shallow marine soundscapes using explained clusters. *J. Mar. Sci. Eng*. **11**, 550. (doi:10.3390/jmse11030550)
15. Sharma G, Umapathy K, Krishnan S. 2020 Trends in audio signal feature extraction methods. *Appl. Acoust*. **158**, 107020. (doi:10.1016/j.apacoust.2019.107020)
16. Abayomi-Alli OO, Damaševičius R, Qazi A, Adedoyin-Olowe M, Misra S. 2022 Data augmentation and deep learning methods in sound classification: a systematic review. *Electronics* **11**, 3795. (doi:10.3390/electronics11223795)
17. Jasim WN, Saddam SAW, Harfash EJ. 2022 Wind sounds classification using different audio feature extraction techniques. *Informatica* **45**, 57–65. (doi:10.31449/inf.v45i7.3739)
18. Chérubin LM, Dalgleish F, Ibrahim AK, Schärer-Umpierre M, Nemeth RS, Matthews A, Appeldoorn R. 2020 Fish spawning aggregations dynamics as inferred from a novel, persistent presence robotic approach. *Front. Mar. Sci* **6**, 779. (doi:10.3389/fmars.2019.00779)
19. Noda J, Travieso C, Sánchez-Rodríguez D. 2016 Automatic taxonomic classification of fish based on their acoustic signals. *Appl. Sci*. **6**, 443. (doi:10.3390/app6120443)
20. Malfante M, Mars JI, Dalla Mura M, Gervaise C. 2018 Automatic fish sounds classification. *J. Acoust. Soc. Am*. **143**, 2834–2846. (doi:10.1121/1.5036628)
21. Harakawa R, Ogawa T, Haseyama M, Akamatsu T. 2018 Automatic detection of fish sounds based on multi-stage classification including logistic regression via adaptive feature weighting. *J. Acoust. Soc. Am*. **144**, 2709–2718. (doi:10.1121/1.5067373)
22. Lewis DD. 1998 Naive (Bayes) at forty: the independence assumption in information retrieval (eds C Nédellec, C Rouveiro). In *Machine Learning: ECML-98. Lecture Notes in Computer Science, vol. 1398*, pp. 4–15. Berlin, Heidelberg, Germany: Springer. (doi:10.1007/BFb0026666)
23. Breiman L. 2001 Random forests. *Mach. Learn.* **45**, 5–32. (doi:10.1023/A:1010933404324)
24. Breiman L, Friedman JH, Olshen RA, Stone CJ. 2017 Regression trees. In *Classification and regression trees*, pp. 216–265, 1st edn. New York, NY: Chapman and Hall/CRC. (doi:10.1201/9781315139470-8)
25. Rumelhart DE, McClelland JL. 1987 A general framework for parallel distributed processing. In *Parallel distributed processing: explorations in the microstructure of cognition. Volume 1: Foundations* (eds DE Rumelhart, JL McClelland), pp. 45–76. Cambridge, MA: MIT Press.
26. Jesus SM, Xavier FC, Vio RP, Osowsky J, Simões MVS, Netto EBF. 2020 Particle motion measurements near a rocky shore off Cabo Frio Island. *J. Acoust. Soc. Am*. **147**, 4009–4019. (doi:10.1121/10.0001392)
27. Cordeiro CAMM, Harborne AR, Ferreira CEL. 2014 Patterns of distribution and composition of sea urchin assemblages on Brazilian subtropical rocky reefs. *Mar. Biol*. **161**, 2221–2232. (doi:10.1007/s00227-014-2500-0)
28. Carriere O, Hermand JP, Calado L, de Paula AC, da Silveira ICA. 2009 Feature-oriented acoustic tomography: Upwelling at Cabo Frio (Brazil). In *OCEANS 2009 MTS/IEEE Biloxi - Marine Technology for Our Future: Global and Local Challenges*, *Biloxi, MS*, pp. 1–8. Biloxi, MI: IEEE. (doi:10.23919/OCEANS.2009.5422337)
29. Valentin JL. 1984 Analyse des paramètres hydrobiologiques dans la remontée de Cabo Frio (Brésil). *Mar. Biol*. **82**, 259–276. (doi:10.1007/bf00392407)
30. Ferreira C, Gonçalves J, Coutinho R. 2001 Community structure of fishes and habitat complexity on a tropical rocky shore. *Environ. Biol. Fishes* **61**, 353–369. (doi:10.1023/A:1011609617330)

31. Cordeiro CAMM, Mendes TC, Harborne AR, Ferreira CEL. 2016 Spatial distribution of nominally herbivorous fishes across environmental gradients on Brazilian rocky reefs. *J. Fish Biol*. **89**, 939–958. (doi:10.1111/jfb.12849)
32. Campbell D, Xavier FC, Melo Junior UG, Silveira NG, Versiani LL, Netto EBF. 2019 Underwater soundscape pattern during high season of nautical tourism in Cabo Frio Island, Brazil. In *5th International Conference on the Effects of Noise on Aquatic Life*, Den Haag, The Netherlands, 7–12 July 2019. *POMA*. vol. **37, 070003**. AIP Publishing. (doi:10.1121/2.0001100)
33. Mann D, Grothues T. 2009 Short-term upwelling events modulate fish sound production at a mid-Atlantic Ocean observatory. *Mar. Ecol. Prog. Ser*. **375**, 65–71. (doi:10.3354/meps07720)
34. Rice A, Soldevilla M, Quinlan J. 2017 Nocturnal patterns in fish chorusing off the coasts of Georgia and eastern Florida. *Bull. Mar. Sci*. **93**, 455–474. (doi:10.5343/bms.2016.1043)
35. Siddagangaiah S, Chen CF, Hu WC, Danovaro R, Pieretti N. 2021 Silent winters and rock-and-roll summers: the long-term effects of changing oceans on marine fish vocalization. *Ecol. Indic*. **125**, 107456. (doi:10.1016/j.ecolind.2021.107456)
36. Looby A. 2023 FishSounds version 1.0: a website for the compilation of fish sound production information and recordings. *Ecol. Informatics* **74**, 101953. (doi:10.1016/j.ecoinf.2022.101953)
37. Froese R, Pauly D. 2021 *FishBase*. See https://www.fishbase.org.version.
38. Fish M, Mowbray W. 1970 *Sounds of western North Atlantic fishes, a reference file of biological underwater sounds*. Baltimore, MD: Johns Hopkins Press.
39. Amorim M. 2006 Diversity of sound production in fish. In *Communication in fishes* (eds F Ladich, SP Collin, P Moller, BG Kapoor), pp. 71–104. Enfield, NH: Science Publishers.
40. Ladich F. 2014 Fish bioacoustics. *Curr. Opin. Neurobiol.* **1**, 121–127. (doi:10.1016/j.conb.2014.06.013)
41. Carriço R, Silva MA, Meneses GM, Fonseca PJ, Amorim MCP. 2019 Characterization of the acoustic community of vocal fishes in the Azores. *PeerJ* **7**, e7772. (doi:10.7717/peerj.7772)
42. Desiderà E, Guidetti P, Panzalis P, Navone A, Valentini-Poirrier C, Boissery P, Gervaise C, Di Iorio L. 2019 Acoustic fish communities: sound diversity of rocky habitats reflects fish species diversity. *Mar. Ecol. Prog. Ser*. **608**, 183–197. (doi:10.3354/meps12812)
43. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology2019 *Raven Pro: interactive sound analysis software*. Ithaca, NY: The Cornell Lab of Ornithology. See https://www.ravensoundsoftware.com/software/raven-pro/.
44. Charif RA, Strickman LM, Waack AM. 2010 *Raven Pro 1.4 User's Manual*. Ithaca, NY: The Cornell Lab of Ornithology. See https://www.ravensoundsoftware.com/knowledge-base/.
45. McFee B, ColinR, DawenL, Ellis DP, McVicar M, Battenberg E, Nieto E. 2015 librosa: Audio and Music Signal Analysis in Python. In *Proceedings of the 14th Python in Science Conference (SciPy 2015)* (eds K Huff, J Bergstra), pp. 18–24. Austin, TX: SciPy.
46. Zelkowitz M. 2010 *Advances in computers: improving the web*, vol. **78**, 1st edn. Oxford, UK: Academic Press.
47. Giannakopoulos T, Pikrakis A. 2014 *Introduction to Audio Analysis: a MATLAB® approach*. Oxford, UK: Academic Press. (doi:10.1016/B978-0-08-099388-1.00001-7)
48. Torres-García A, Garcia C, Villasenor-Pineda L, Mendoza-Montoya O (eds). 2021 *Biosignal processing and classification using computational learning and intelligence: principles, algorithms, and applications*. London, UK: Academic Press.
49. Tracey B *et al*. 2023 Towards interpretable speech biomarkers: exploring MFCCs. *Sci. Rep*. **13**, 22787. (doi:10.1038/s41598-023-49352-2)
50. Rao K, Manjunath K. 2017 *Speech recognition using articulatory and excitation source features*. Fort Lee, NJ: Springer.
51. Stowell D. 2022 Computational bioacoustics with deep learning: a review and roadmap. *PeerJ* **10**, e13152. (doi:10.7717/peerj.13152)
52. Pedregosa F. 2011 Scikit-learn:Machine Learning in Python. *J. Mach. Learn. Res* **12**, 2825–2830. (doi:10.48550/arXiv.1201.0490)
53. Laplante JF, Akhloufi MA, Gervaise C. 2022 Deep learning for marine bioacoustics and fish classification using underwater sounds. In *2022 IEEE Canadian Conf. on Electrical and Computer Engineering (CCECE)*, Halifax, Canada, 18–20 September 2022, pp. 288–293. Halifax, Canada: IEEE. (doi:10.1109/CCECE49351.2022.9918242)
54. Sokolova M, Lapalme G. 2009 A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag*. **45**, 427–437. (doi:10.1016/j.ipm.2009.03.002)
55. Cai W, Zhu J, Zhang M, Yang Y. 2022 A parallel classification model for marine mammal sounds based on multi-dimensional feature extraction and data augmentation. *Sensors* **22**, 7443. (doi:10.3390/s22197443)
56. Lundberg SM, Lee SI. 2017 A unified approach to interpreting model predictions. *Adv. Neur. Inf. Proc. Sys* **30**, 4768–4777. (doi:10.48550/arXiv.1705.07874)
57. Sundararajan M, Najmi A. 2020 *The many Shapley values for model explanation. Proc. 37th Int. Conf. on Machine Learning (PMLR), Online* **119**, 9269–9278. https://proceedings.mlr.press/v119/sundararajan20b.html
58. Sattar F, Cullis-Suzuki S, Jin F. 2016 Identification of fish vocalizations from ocean acoustic data. *Appl. Acoust*. **110**, 248–255. (doi:10.1016/j.apacoust.2016.03.025)
59. Ibrahim AK, Chérubin LM, Zhuang H, Schärer Umpierre MT, Dalgleish F, Erdol N, Ouyang B, Dalgleish A. 2018 An approach for automatic classification of grouper vocalizations with passive acoustic monitoring. *J. Acoust. Soc. Am*. **143**, 666–676. (doi:10.1121/1.5022281)
60. Salman A, Maqbool S, Khan A, Jalal A, Shafait F. 2019 Real-time fish detection in complex backgrounds using probabilistic background modelling. *Ecol. Informatics* **51**, 44–51. (doi:10.1016/j.ecoinf.2019.02.011)
61. Reifman J, Feldman E. 2002 Multilayer perceptron for nonlinear programming. *Comput. Oper. Res*. **29**, 1237–1250. (doi:10.1016/S0305-0548(01)00027-2)
62. Bengio Y. 2011 Deep learners benefit more from out-of-distribution examples. In *Proc. Fourteenth Int. Conf. Artificial Intelligence and Statistics*, pp. 164–172. Fort Lauderdale, FL: JMLR Workshop and Conference Proceedings.
63. Parmar A, Katariya R, Patel V. 2019 A review on random forest: an ensemble classifier. In *International conference on intelligent data communication technologies and internet of things (icici) 2018. icici 2018. lecture notes on data engineering and communications technologies, vol. 26* (eds J Hemanth, X Fernando, P Lafata, Z Baig), pp. 758–763. Cham, Switzerland: Springer International Publishing. (doi:10.1007/978-3-030-03146-6_86)
64. Géron A. 2022 *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 3rd edn. Sebastopol, CA: O'Reilly Media, Inc.
65. Grinsztajn L, Oyallon E, Varoquaux G. 2022 Why do tree-based models still outperform deep learning on typical tabular data? *Adv. Neural Inf. Process. Syst* **35**, 507–520. (doi:10.48550/arXiv.2207.08815)
66. Onishi S, Meguro S. 2023 Rethinking data augmentation for tabular data in deep learning. *arXiv* 2305.10308. (doi:10.48550/arXiv.2305.10308)
67. Li RH, Belford GG. 2002 Instability of decision tree classification algorithms. In *Proc. of the Eighth ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining*, Edmonton, Alberta, Canada, 23–26 July 2002, pp. 570–575. New York, NY: Association for Computing Machinery. (doi:10.1145/775047.775131)
68. Kumar P, Kumar D. 2020 Decision tree classifier: a detailed survey. *Int. J. Inf. Decis. Sci* **12**, 246–269. (doi:10.1504/IJIDS.2020.108141)
69. Dimitoglou G, Adams JA, Jim CM. 2012 Comparison of the C4.5 and a naïve Bayes classifier for the prediction of lung cancer survivability. *J. Comp* **4**, 1–9. (doi:10.48550/arXiv.1206.1121)
70. Stowell D, Petrusková T, Šálek M, Linhart P. 2019 Automatic acoustic identification of individuals in multiple species: improving identification across recording conditions. *J. R. Soc. Interface* **16**, 20180940. (doi:10.1098/rsif.2018.0940)
71. Ibrahim AK, Zhuang H, Chérubin LM, Schärer-Umpierre MT, Nemeth RS, Erdol N, Ali AM. 2020 Transfer learning for efficient classification of grouper sound. *J. Acoust. Soc. Am*. **148**, EL260–EL266. (doi:10.1121/10.0001943)

72. Waddell EE, Rasmussen JH, Širović A. 2021 Applying artificial intelligence methods to detect and classify fish calls from the Northern Gulf of Mexico. *J. Mar. Sci. Eng*. **9**, 1128. (doi:10.3390/jmse9101128)
73. Machado P, Fernandes B, Novais P. 2022 Benchmarking data augmentation techniques for tabular data (eds H Yin, D Camacho, P Tino). In *Int. Conf. on Intelligent Data Engineering and Automated Learning, Manchester, UK, 24–26 November 2022*, pp. 104–112. Cham, Switzerland: Springer International Publishing. (doi:10.1007/978-3-031-21753-1_1)
74. Vieira M, Fonseca PJ, Amorim MCP, Teixeira CJC. 2015 Call recognition and individual identification of fish vocalizations based on automatic speech recognition: an example with the Lusitanian toadfish. *J. Acoust. Soc. Am*. **138**, 3941–3950. (doi:10.1121/1.4936858)
75. Vieira M, Pereira BP, Pousão-Ferreira P, Fonseca PJ, Amorim MCP. 2019 Seasonal variation of captive meagre acoustic signalling: a manual and automatic recognition approach. *Fishes* **4**, 28. (doi:10.3390/fishes4020028)
76. Ruiz-Blais S, Rivera-Chavarria MR, Camacho A. 2012 Autonomous detection of neotropical sciaenid fishes. *Proc. Mtngs Acoust* **18**, 010001. (doi:10.1121/1.4792734)
77. Rice A. 2022 Evolutionary patterns in sound production across fishes. *Ichthyol. Herpetol*. **110**, 1–12. (doi:10.1643/i2020172)
78. Parsons MJG *et al*. 2022 Sounding the call for a global library of underwater biological sounds. *Front. Ecol. Evol.* **10**, 39. (doi:10.3389/fevo.2022.810156)
79. Spanier E. 1979 Aspects of species recognition by sound in four species of damselfishes, genus *Eupomacentrus* (Pisces: Pomacentridae). *Z. Für Tierpsychol*. **51**, 301–316. (doi:10.1111/j.1439-0310.1979.tb00691.x)
80. Frédérich B, Parmentier E. 2016 *Biology of damselfishes*, p. 322, 1st edn. Boca Raton, FL: CRC Press.
81. Ribeiro MT, Singh S, Guestrin C. 2016 'Why should I trust you?' Explaining the predictions of any classifier. In *Proc. of the 22nd ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining*, San Francisco, CA, 13–17 August 2016, pp. 1135–1144. New York, NY: Association for Computing Machinery. (doi:10.1145/2939672.2939778)
82. Wang J, Zhang L, Liu T, Pan W, Hu B, Zhu T. 2019 Acoustic differences between healthy and depressed people: a cross-situation study. *BMC Psychiatry* **19**, 1–12. (doi:10.1186/s12888-019-2300-7)
83. Bellwood DR, Streit RP, Brandl SJ, Tebbett SB. 2019 The meaning of the term 'function' in ecology: a coral reef perspective. *Funct. Ecol*. **33**, 948–961. (doi:10.1111/1365-2435.13265)
84. Mouy X *et al*. 2024 Automatic detection of unidentified fish sounds: a comparison of traditional machine learning with deep learning. *Front. Remote Sens.* **22**, 1439995. (doi:10.3389/frsen.2024.1439995)
85. Duarte CM *et al*. 2021 The soundscape of the Anthropocene ocean. *Science* **371**, eaba4658. (doi:10.1126/science.aba4658)
86. vivianeocn/. 2024 *fish_sounds_shap*. See https://github.com/vivianeocn/fish_sounds_shap.
87. Barroso VR, Lessa A, Ferreira CEL, Xavier F. 2025 Supplementary material from: Optimal feature selection and model explanation for reef fish sound classification. Figshare. (doi:10.6084/m9.figshare.c.7837885)