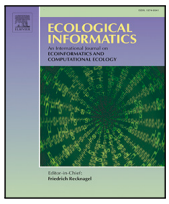
Ecological Informatics 97 (2026) 103874

Contents lists available at ScienceDirect

Ecological Informatics

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



Rapid fish sound detection using human-in-the-loop deep learning

Valentin Bordoux [a,*], Clea Parcerisas [b], Martin Jälmby [c], Elisabeth Debusschere [b], Albertinka J. Murk [a], Rosa M. van der Ven [a]

[a] Marine Animal Ecology, Wageningen University, Droevendaalsesteeg 1, Wageningen, 6708 PB, The Netherlands
[b] Flanders Marine Institute (VLIZ), Jacobsenstraat 1, Ostend, 8400, Belgium
[c] University of Hamburg, Mittelweg 177, Hamburg, 20148, Germany

ARTICLE INFO

Keywords:
Passive Acoustic Monitoring
Fish sounds
Deep learning
Active learning
Underwater acoustics
North Sea

ABSTRACT

Passive acoustic monitoring (PAM) can be used to detect and classify marine fauna sounds, providing a powerful, non-invasive tool for studying animal presence and behaviour across various temporal and spatial scales, offering unique insights into marine ecosystems. The use of PAM specifically for fish is currently limited by the intense manual effort to annotate the audio data. While deep learning has enabled automated PAM data processing in bird and marine mammal studies, these approaches depend on large annotated training datasets. Such datasets are not available for most fish species, especially in temperate waters where sounds are less frequent and harder to detect and identify. Therefore, this study evaluates an Agile Modelling workflow that incorporates a human-in-the-loop approach to efficiently train sound detectors with minimal effort. Previously applied to birds and reef sounds, we assess its applicability in two temperate marine environments, markedly different from prior test cases. The workflow allows for models to be trained in under two hours with no other initial training data than one example of the target sound. The detectors trained were evaluated against manually annotated datasets. Results show that the Agile Modelling workflow can effectively train models for detecting rare and putative fish sounds, significantly reducing annotation time. Different strategies were compared to offer practical guidelines and highlight method limitations. This approach enables quicker model development, promotes the sharing of annotated datasets, and could accelerate the broader adoption of automated fish PAM. Ultimately, such tools support improved monitoring, understanding and conservation of marine ecosystems.

1. Introduction

Passive Acoustic Monitoring (PAM) has been a cornerstone of ecological research in both terrestrial and marine environments for decades. In the marine realm, PAM has a 30-year history of efficacy, particularly for monitoring cetaceans, while in terrestrial ecosystems it has enabled the long-term study of birds, amphibians, and insects, especially in remote regions (Sugai et al., 2019). These successes have established PAM as a non-invasive, scalable method for biodiversity assessment across diverse biomes. However, while its use is well-established for marine mammals and various terrestrial taxa, the application of PAM for monitoring fish remains comparatively limited (McGeady et al., 2023).

Recent research suggests that many more fish species are vocal than previously assumed. Phylogenetic studies and recent field observations have shown that a wide variety of fish are capable of actively producing sounds, including species formerly assumed to be silent, such as, for example, multiple Batomorphi species (Rice et al., 2022; Fetterplace et al., 2022). These vocalisations are often tied to ecologically significant behaviours such as mating, spawning, territorial defence, and social interactions. The ability to detect such sounds offers a promising, non-invasive method for monitoring fish biodiversity, behaviour, and ecosystem health, as a complementary monitoring method. Passive Acoustic Monitoring has been successfully applied to monitor fish for multiple applications, including species presence assessment, invasive species detection, improving knowledge of red-listed species distribution, relative fish abundance estimation, and the detection and characterisation of fish spawning sites (Lindseth and Lobel, 2018; Bolgan et al., 2023; Amorim et al., 2023; Souza et al., 2023; Chérubin et al., 2020; Wilson et al., 2019).

PAM has potential for enabling non-invasive, scalable sensor networks across extended temporal and spatial scales, even in environments and conditions (e.g. at night) where conventional monitoring methods cannot be applied (Darras et al., 2025; Mooney et al., 2020).

* Corresponding author.
E-mail address: valentin.bordoux@gmail.com (V. Bordoux).

https://doi.org/10.1016/j.ecoinf.2026.103874
Received 20 August 2025; Received in revised form 22 May 2026; Accepted 6 June 2026
Available online 15 June 2026

However, applying PAM to study marine biodiversity remains challenging. Marine environments are inherently difficult to access and observe, especially in dynamic and turbid coastal regions. Ground-truthing acoustic recordings, linking detected sounds to specific species or behaviours, remains difficult due to limited visibility, logistical constraints, and the need for expensive and labour-intensive validation methods such as visual surveys or net sampling. These issues make it particularly hard to build reliable reference libraries of biological sounds and hamper the interpretation of acoustic datasets.

In order to successfully identify fish species based on sound, high-quality reference sounds are needed, which are only available for a fraction of fish species. Collecting new reference fish sounds is an active research topic, and efforts to fill reference databases are growing but remain scarce to date (Looby et al., 2024). Even without reference sounds, studying the abundance and diversity of putative – i.e., assumed to be – fish sounds has been shown to correlate with fish diversity or reef health in some environments (Bolgan et al., 2025; Di Iorio et al., 2021; Jarriel et al., 2024a). Gathering, storing, and sharing putative fish sounds can also help identify species by cross-referencing between locations (Vieira et al., 2024) or by comparing them with newly collected reference sounds (Parsons et al., 2024).

Research involving PAM on fish has mainly concentrated on coral reef systems in tropical and subtropical regions, which are often characterised by high species diversity, frequent vocal activity, and relative accessibility. In contrast, the application of PAM to temperate environments, particularly systematic soundscape research in the North Sea, is understudied (Basan et al., 2024). However, recent studies have begun addressing this knowledge gap through targeted regional investigations. Detailed soundscape analyses have been conducted in the Belgian North Sea (Calonge et al., 2024; Parcerisas et al., 2023b,a). In the Wadden Sea, comparative acoustic monitoring between natural and artificial reefs revealed significantly higher biotic acoustic diversity at natural and artificial reef sites compared to the surrounding sandflats (Watson et al., 2025). Other hard structures, wind-turbine foundations and oil-and-gas platforms, might also act as de facto artificial reefs. Still, initial acoustic surveys show that such sites are characterised by low call density and diversity (Bolgan et al., 2025). These observations suggest that while fish vocalisations are present, they may be infrequent, context-dependent, and hard to detect.

The southern North Sea is a highly industrialised and ecologically complex region, shaped by strong tidal currents, high turbidity, and intense human activity such as commercial fishing, dense shipping traffic, offshore wind farms, dredging, and oil and gas operations. This dynamic and heavily impacted environment poses considerable challenges for ecological observation. Although temperate marine ecosystems such as the North Sea can be biologically rich, centuries of overfishing and anthropogenic disturbance have substantially degraded their fish populations and habitats. Consequently, biological acoustic activity in this region is relatively sparse and remains poorly understood. While terabytes of PAM data have already been collected, primarily to monitor anthropogenic noise and marine mammals, realising their full potential for studying fish and other taxa is still constrained by the labour-intensive nature of data processing.

Advances in machine learning, particularly deep learning, have begun to address the challenge of labour-intensive data processing by automating bioacoustics data analysis. Deep neural networks have outperformed traditional methods for the classification of sounds from birds, marine mammals, and other taxa (Stowell, 2022). However, these models depend on large, annotated species- and location-specific datasets for training, which are lacking in understudied environments such as the southern North Sea. Extensive datasets exist for birds (e.g., Xeno-Canto (2026)) and marine mammals (e.g., the Watkins archive (Sayigh et al., 2017), Australian Antarctic Program Blue and Fin whale library (Miller et al., 2020)), but comparable resources are not yet developed for fish (Jarriel et al., 2024b). Furthermore, there are no general-purpose detection models for fish, highlighting a critical methodological gap. The high diversity of soniferous fish species and the complexity of marine soundscapes make it challenging and labour-intensive to document the full range of acoustic activity. Consequently, current automated fish sound detection models are narrowly tailored to specific species or habitats (e.g. Mouy et al. (2024), Ibrahim et al. (2024) and Laplante et al. (2022)). A recent attempt to develop a more broadly applicable model used a newly released dataset (ReefSet, which includes fish sounds) and additional datasets to classify reef-associated sounds (Williams et al., 2025), but its applicability beyond tropical ecosystems remains untested.

Transfer learning, repurposing pre-trained models for new, data-limited tasks, offers a potential solution to reduce annotation effort and improve model performance compared to training from scratch (Zhuang et al., 2021). The disparity between a model's original training task and its new application is called the *domain gap* or *domain shift*, and a model's ability to bridge this gap is referred to as *generalisation* (Liang et al., 2024; Farahani et al., 2021). In PAM, transfer learning has been increasingly applied to compensate for limited annotated data and to improve cross-species and cross-habitat detection, a strategy which has also been used for fish sounds (Ibrahim et al., 2020; Munger et al., 2022). For example, models trained on bird sounds have demonstrated better generalisation to other bioacoustics data from other taxa than models trained on non-biological data, Ghani et al. (2023), though this result can vary depending on fine-tuning strategies (Schwinger et al., 2025).

In addition to reusing pre-trained models, the annotation strategy can be adapted to reduce effort for training models. Active learning is a training paradigm in which models are iteratively refined by incorporating newly annotated data obtained through targeted human feedback, with the aim to reduce overall annotation effort (Settles, 2012). Despite growing evidence of its effectiveness in bioacoustics (Stowell, 2022), its implementation in the field is still limited, in part because evaluating methods that require human–model interaction remains a methodological challenge (Kath et al., 2024a). Active learning can be combined with transfer learning, whereby a pre-trained model is used for initialisation, and active learning is applied for fine-tuning. This combination has proven effective in reducing manual annotation effort across multiple taxa (Kath et al., 2024b). The recently proposed *Agile Modelling* workflow employs this joint strategy to detect rare bird calls and sounds in coral reefs using models pre-trained on both avian and reef audio (Dumoulin et al., 2025). The workflow operates on embeddings from a pre-trained model, computed only once, on top of which a lightweight classification layer is trained. This design substantially accelerates training and enhances scalability to large datasets because embeddings do not require repeated computation during iterative training. Although simulations indicate that the method is robust to substantial domain gaps (e.g., applying a bird-trained model to sounds of Anura), its performance in temperate marine environments has yet to be evaluated.

In environments like the southern North Sea, where biological sounds are rare, extracting biotic signals from long-term (i.e. several months) recordings is particularly labour-intensive and expensive. In a soundscape where calls are both infrequent and diverse, compiling enough examples may require hundreds of hours of manual annotation and screening several months of data before sufficient instances of a particular sound can be located. If applicable, the Agile Modelling workflow can overcome this limitation by training automatic sound detectors from a single example, facilitating the processing of large datasets and the collection of manually annotated sounds.

This study aims to help substantially reduce manual annotation efforts by validating the applicability of Agile Modelling for detecting biological sounds, particularly putative fish sounds, in temperate marine environments. Important to note that this study focuses on the detection of individual calls rather than fish choruses, which typically requires different data analysis methods due to the fundamental difference between these two types of biological signals (Kim et al., 2023).

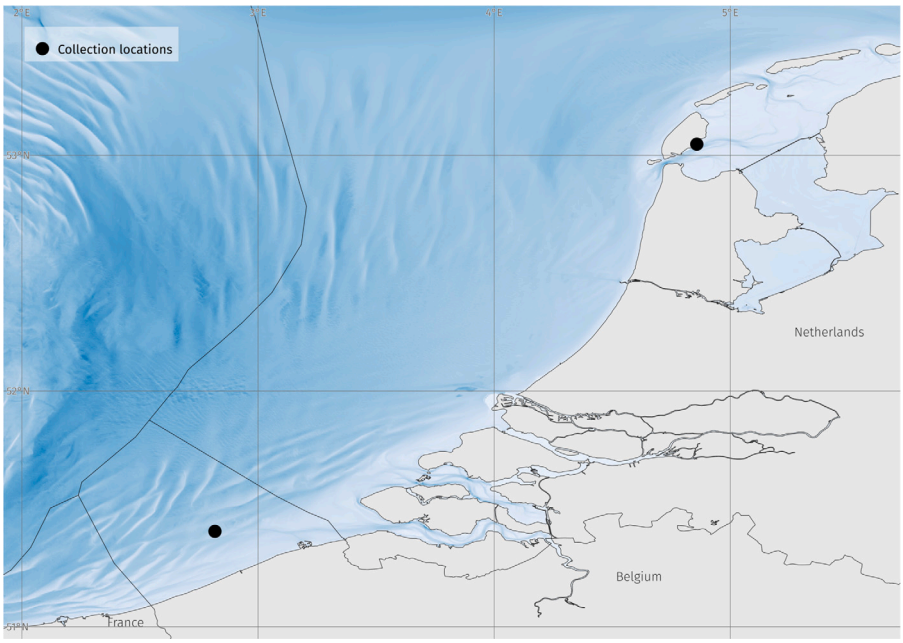


Fig. 1. Data collection locations at the coast of Belgium and the Netherlands. Black lines indicate the maritime borders. Map made by maps@vliz.be.

Specifically, we assess whether Agile Modelling can effectively be used to train automated sound detectors, starting with only one sound, and streamline the collection of sound instances in these ecosystems. We evaluate the method on datasets recorded in two geographically distinct southern North Sea locations, using manually annotated evaluation sets withheld from training. We compare key workflow strategies, including pre-trained models, sampling approaches, and training-set balances, to optimise detector performance. Additionally, we benchmark detectors trained with Agile Modelling against a conventional deep learning model trained from scratch on identical data, and test the temporal generalisability of the detectors by applying them to data from a different year and season at the same location. Based on these experiments, we provide practical guidelines and open-source code to enable researchers to apply Agile Modelling to their underwater acoustic datasets, while explicitly addressing limitations and constraints.

2. Methods

2.1. Data collection and processing

Data were collected from two separate fixed locations in the southern North Sea, one in the Texel-Oudeschild harbour on Texel in the Netherlands (NL) and the other midshore in the Belgian part of the North Sea (BE). The locations of data collection are displayed in Fig. 1. NL data were collected using a SoundTrap ST600 (Ocean Instruments, New Zealand, sensitivity −177 dB/V re 1 μPa) deployed from 3 September to 9 September 2024 with continuous recording at a sampling rate of 96 kHz. The recorder was mounted on a floating pontoon in the harbour, 1 m below the water surface. For this study, we will refer to this dataset as NL24. The data in BE were collected as part of the LifeWatch Broadband Acoustic Network (Parcerisas et al., 2021), using RESEA 320 recorders (RTSys, France) together with Colmar GP1190M-LP hydrophones (Colmar, Italy, sensitivity: −180 dB/V re 1 μPa, frequency range −3 dB: 10 Hz to 170 kHz), attached to steel mooring frames at 1 m above the sea bottom, with no moving parts. The deployment covered the period from 28th October to 7th November 2022, with a duty cycle of one day on, one day off, sampled at 48 kHz. For this study, we will refer to this dataset as BE22.

For each dataset, a 10-hour test set was created by randomly selecting sixty 10-minute files from across the entire deployment. These files were manually annotated via visual and aural inspection using Raven Pro 1.6 (The Cornell Lab of Ornithology, Ithaca, NY, USA) and were strictly excluded from the training data. The data from the test set were initially strongly annotated by drawing bounding boxes for a single class using Raven. To minimise false positives in the ground truth, only signals that were distinguishable from background noise via both visual inspection of the spectrogram and aural confirmation were labelled. These annotations were then used to generate weakly labelled annotations using a non-overlapping sliding window approach of 3 or 5 s duration, depending on the model evaluated. The windows were considered positive if 50% or more of at least one annotation is contained in a window, or if more than 20% of a window was covered by one or multiple annotations; the script of the method is available on GitHub (Bordoux, 2026). All the data, excluding the test set, were considered the training corpus.

Furthermore, a reference sound for each dataset and sound type was selected from the training corpus. A high-quality example, characterised by a complete call and a high signal-to-noise ratio relative to the rest of the dataset, was selected via rapid visual screening for each sound type. Reference sounds are assumed to be fish sounds based on aural and visual characteristics and temporal patterns. For the NL24 data, the selected sound was labelled 'Downsweep' (see Fig. 2a), and is characterised by a continuous sound with harmonics spanning from 100 to 700 Hz for a duration of 0.1 to 0.25 s, with a peak frequency around 250 Hz, slightly decreasing during the call at a rate varying per occurrence. For the BE22 data, the reference sound is described as a pulsed sound with a high repetition rate and a frequency spanning from around 100 to 4000 Hz. The number of pulses is not constant at each detected event, ranging from 2 to 70 pulses, but with a majority of them around 10–15 pulses. The repetition rate is around 14 pulses/second. The sound was previously described in Parcerisas et al. (2024) as 'Jackhammer', see Fig. 2b. A sound clip of each sound type can be found in the Supplementary Material.

For the experiment on the effect of domain shift (the generalisation capability of a model, see Section 2.3.5), performed on recordings from the Belgium location only, recordings from a different deployment were analysed. These recordings were collected using the same device and deployment setting, at the same location. This dataset, referred to as BE21, was collected from 10 March to 19 May 2021, and no previous knowledge was available on whether the Jackhammer sound was present or not.

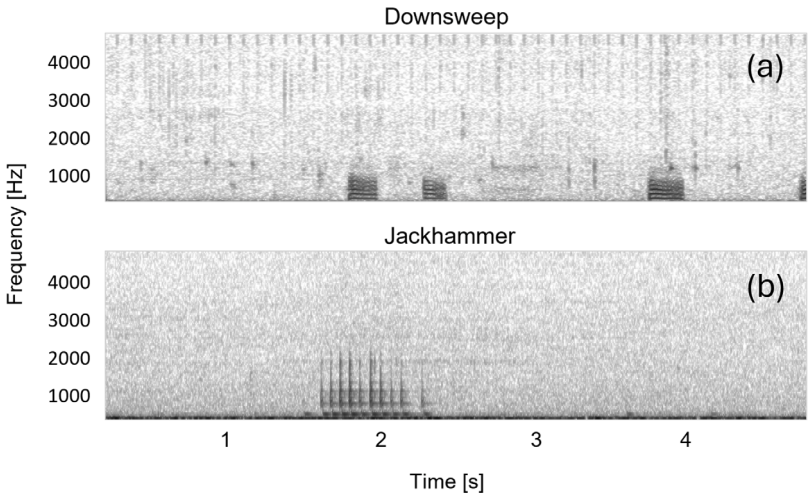


Fig. 2. Spectrograms of examples of the targeted sounds, the downsweep and the jackhammer, assumed to be produced by fish, for both (a) NL and (b) BE.

2.2. Agile modelling workflow

The workflow used in this study was adapted from Dumoulin et al. (2025). The code associated with this work is available on the project's GitHub repository (Bordoux, 2026), and the Agile Modelling workflow is illustrated in Fig. 3. Our objective was to validate the applicability of the Agile Modelling workflow to untested environments and provide guidelines on its use. Once this baseline is established, the workflow could be further optimised by incorporating automated selection strategies to reduce user effort and improve the reproducibility of sample selection. Agile Modelling is an iterative training approach, guided by user feedback, in which a one-layer linear classifier is trained on top of frozen, pre-trained models. Although the workflow used in this study shares the principle of iterative learning with active learning, it differs in how training samples are selected. The user guides sample selection at each iteration by choosing a logit score near which to retrieve samples (see details below). In contrast, active learning relies on an automated algorithm to select samples according to a predefined strategy, most commonly based on uncertainty or diversity criteria (Settles, 2012). Notably, the human-in-the-loop workflow used here can be converted into an active learning framework by replacing manual sample selection with an automated strategy (see step 6 in Fig. 3). However, the design and evaluation of active learning strategies is an active area of research and was beyond the scope of this study.

In this study, each training process was initialised using a single annotated sound, hereafter referred to as the reference sound. We focused on training binary classifiers, referred to as detectors from this point on, designed to target a single sound type. Although Agile Modelling can also support multi-class classification, provided additional annotation effort and at least one reference sound per class, this capability was not exploited here. While PAM datasets are generally multi-label and contain a wide variety of signals of interest, temperate marine environments such as the one used here can exhibit distinct characteristics: sounds of interest are comparatively rare, biotic sound diversity is low, and many sound types remain poorly characterised. In such contexts, the primary analytical bottleneck lies in detecting occurrences of target sounds within extensive background noise, rather than in discriminating among multiple sound classes, and valuable ecological data can be obtained through simple detection. For this reason, although a single model performing multi-species detection and classification would be preferable, the current state of knowledge and available data are insufficient to support reliable performance (Stowell, 2022). Even though the used framework supports multi-class classification, we instead propose to train a binary classifier per sound type that functions primarily as a call detector, making the process as efficient as possible so that training a model for each sound type of interest at a given location becomes feasible.

The workflow can be divided into 6 steps, visible in Fig. 3:

1. Audio recordings were provided as input to the workflow in WAV format with their native sampling rate. The audio data were resampled and segmented into windows of the duration required by the pre-trained model used (see Section 2.3.2 for model specifications). The pre-trained model generated a high-dimensional representation for each window, the embeddings, from the output of the last layer before the classification head. A simplified 2D representation of the embeddings is shown in step 3 of Fig. 3 for visualisation purposes only.
2. The user provided one example of the target sound, which was also converted into an embedding. This sound was then located in the multi-dimensional space with the embeddings of the rest of the dataset. In step 3 of Fig. 3, this is represented by a red cross.
3. A similarity search (using minimal Euclidean distance) was used to select samples resembling the target sound in the embedding space (red area in step 3 of Fig. 3, see also Allen-Ankins et al. (2025) for a similar approach).
4. Selected samples were suggested to the user for annotating as either "target sound" or "Unknown", or were left unannotated if the user was unsure.
5. Data annotated during the iteration were separated into a training set and a validation set with a proportion of 80% and 20%. In this study, five detectors were trained and evaluated using different splits of the training and validation data using a 5-fold cross-validation protocol. The performance of the five detectors was averaged to obtain a more reliable estimation of the quality of the detector.
6. Logit values (prediction scores) were computed for all the samples using the last detector trained in the previous step. The user selected a specific number of additional samples to annotate and the logit value around which the samples will be selected (in Fig. 3, represented as a red dotted line in step 6). Steps 4 to 6 were iterated until reaching the targeted annotation effort, which depended on the specific experiment (see Section 2.3).

For the experiments, each detector was evaluated on the independent test set, after training (step 4). Training was done using the Adam optimiser (Kingma and Ba, 2017) with a learning rate of 0.001 and Binary Cross-Entropy loss provided by TensorFlow (v2.19.0), a batch size of 12, and 128 epochs. The selected test sets consisted of subsets of the same deployment, and therefore, the obtained results cannot

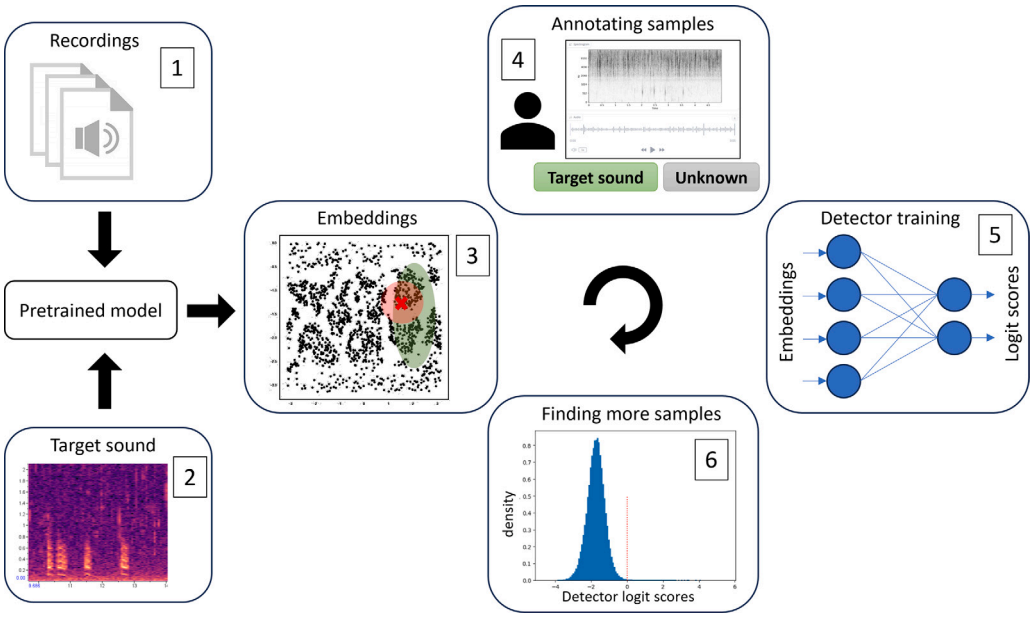


Fig. 3. Agile Modelling workflow. 1. Recordings were cut into short windows and projected in a high-dimensional embedding space by a pre-trained model (Williams et al., 2025). 2. The target sound was projected in the embedding space (red cross). 3. The closest samples to the target sound in the embedding space (red circle area) were shown to the user for annotation. 4. After annotating some samples, a detector constituted of a single-layer neural network was trained (the dimension of the input layer is reduced to four nodes for visualisation). 5. The detector computed a logit score for all points of the embedding space. 6. The user requested more samples to be annotated, around the logit score of their choice. The process was iterated to increase the size of the training set, therefore enhancing the detector until reaching the targeted annotation effort. The goal of workflow was to train a detector able to separate the target sound from the unknown sound samples in the embedding space (a simplified view is represented as the green area in 3).

be assumed to achieve equal performance in other deployments. The proposed evaluation on the test set aimed to evaluate the use of the model within that deployment, and to compare this performance to the one obtained on the validation set. To assess the model's generalisation capabilities to other deployments, one specific experiment was designed (see Section 2.3.5).

2.3. Experiments

We aimed to check if the Agile Modelling workflow was applicable for the detection of putative fish sounds in two temperate marine environments and evaluate the impact of different variations of the workflow. Agile Modelling contains many parameters and elements subject to modification, but we aimed to select the ones thought to have the largest impact. Three variations were identified:

- The pre-trained model used for creating the embeddings
- The strategy to select a logit value to pick new samples
- The ratio of negative and positive samples provided at each iteration

The training of detectors using Agile Modelling was performed on the training corpus of each location for the detection of the Jackhammer and the Downsweep sounds. Performances were evaluated in both the validation set, part of the samples annotated during the iterative training, and the test set (the part of the dataset excluded from the training corpus and manually annotated). The metrics used for evaluation were Precision, Recall, the area under the receiver operating characteristic curve (AUC-ROC), and the area under the precision–recall curve (AUC-PR). At each iteration, the score of each metric for the five detectors of the 5-fold cross-validation was estimated, and the average value was computed. The False Positive Rate was also computed for the baseline experiment, described thereafter, as it is not reflected in other metrics, but can help a user determine the relevance of a model for downstream analysis. We started by running a baseline experiment, then we experimented with variations of the workflow with different scenarios by changing only the selected criteria from the baseline.

2.3.1. Baseline experiment

The baseline experiment evaluated whether Agile Modelling is suitable for detecting fish sounds in temperate marine environments and whether validation set performance serves as a reliable indicator of detector quality. This baseline utilised the *SurfPerch* model (Williams et al., 2025) and a *Tail-Onset* sampling strategy. The user subjectively selected a logit value near the detector boundary—the region characterised by roughly equal proportions of positive and negative samples. In this study, this boundary was located at the right edge of the main density peak, marking the onset of the low-density positive-score tail (Step 6, Fig. 3). Training utilised a balanced 1:1 ratio of positive and negative samples. Following an initial annotation of 5 samples per class, we added 10 samples per iteration until reaching 50 samples per class; thereafter, 100 samples (50 per class) were added per iteration until reaching a total annotation effort of 300 samples. Performance was evaluated on both the validation and test sets. Because Agile Modeling aims to create effective models for specific conditions (e.g., a single deployment) without requiring initial manually annotated data, it can be deployed without a formal test set. In such cases, the workflow provides a performance estimate using a validation set separated from the training data annotated during the workflow. The validation score can be used as a stopping criterion of the iterative process. However, we hypothesised that this indicator is an unreliable estimate of model quality. We tested this by comparing validation performance against our manually annotated test sets.

For the subsequent experiments that compared different parameters of the workflow, the annotation effort was set to 150 annotated samples per run.

2.3.2. Experiment 1: embedding models

The Agile Modelling workflow is based on a pre-trained model. Models initially trained on bird sounds have demonstrated strong generalisation to out-of-domain taxa such as bats, frogs, and marine mammals (Ghani et al., 2023). Among these, BirdNET, Perch, and SurfPerch

have outperformed other bioacoustic embedding models when used for Agile Modelling (Dumoulin et al., 2025). In experiment 1 (model), we compared three pre-trained models used as feature extractors (i.e., with their final classification layer removed): SurfPerch, Perch, and BirdNET v2.4. Perch version 8 relies on an EfficientNet-B1 architecture and was trained on bird recordings from the Xeno-Canto database (Ghani et al., 2023). SurfPerch adopts the same architecture but was trained on a broader dataset encompassing bird sounds, diverse terrestrial sounds, and underwater recordings (ReefSet) (Williams et al., 2025). Both models operate on 5-second audio windows sampled at 32 kHz. Model weights were obtained from Kaggle (Google, 2026a,b). BirdNET v2.4 is based on an EfficientNet-B0 backbone with a custom front-end and classification head. It processes 3-second audio windows sampled at 48 kHz and is available via Zenodo (Kahl et al., 2025). The original birdNET model was trained on data from Xeno-Canto, the Macaulay library, and AudioSet (Kahl et al., 2021); however, the specific datasets used for BirdNET v2.4 are not publicly documented.

Visualisation of the embedding space for qualitative analysis. Another common way to qualitatively evaluate the capacity of a pre-trained model to perform on a different task is to use a visualisation of the embedding space. Agile Modelling heavily depends on the quality of the pre-trained model used to generate embeddings. This quality can be understood as the model's ability to distinguish between different types of sounds within its embedding space, therefore, enabling a simple linear model to effectively separate the target classes. To qualitatively assess the capacity of the pre-trained model, the projection of the test sets in the embedding space was visualised using Uniform Manifold Approximation and Projection (UMAP) on two principal components using the Euclidean metric, n_neighbors=15, and min_dist=0.1, and the code from the package umap-learn (v0.5.7) (McInnes et al., 2020).

2.3.3. Experiment 2: sampling strategies

This experiment compared two intuitive strategies that a user may employ to define the logit score-region from which to retrieve new samples for annotation at each iteration (step 6 in Fig. 3). A straightforward approach is to select examples at the extremities of the logit-score distribution, where the confidence of the classifier is highest; this mirrors classical certainty sampling strategies when done automatically, in active learning (Settles, 2012). However, previous work on active learning, including in bird acoustics, has shown that annotating high-uncertainty samples, those close to the decision boundary of the classifier, can accelerate model refinement, particularly in low-data or imbalanced regimes (Nguyen et al., 2022; McEwen et al., 2024). This motivates the use of an alternative approach in which the user targets the boundary region of the logit curve; margin sampling (Balcan et al., 2007) was selected as a high-uncertainty sampling approach based on previous work (Dumoulin et al., 2025). Accordingly, we evaluated two strategies: *Extremities Picking* and *Tail-Onset*. Extremities Picking selected samples from both ends of the logit distribution, where predictions were most confident, and examples were typically clear representatives of either the targeted sound or background noise. The Tail-Onset strategy targeted the region near the classifier's decision boundary; in our highly imbalanced datasets, this corresponded to the point where the dominant peak of the logit-density curve transitioned into a long, sparse right-hand tail (red dotted line in step 6 of Fig. 3). Within an automated active-learning workflow, these two approaches would best compare to certainty sampling and margin-based uncertainty sampling, respectively.

2.3.4. Experiment 3: training set balance

Balancing training sets is known to improve classifier performance (Ghosh et al., 2024); however, real data from temperate water regions, such as in this study, typically contain very few positive instances and a high proportion of background noise. To better learn to recognise the diversity of noise existing in the dataset, we assumed that increasing the amount of negative samples could enhance performance, as it reproduced a distribution closer to that of the real data (Schall et al., 2024). In experiment 3, the effect of different ratios of positive and negative samples was compared, specifically using 1:1, 1:2, and 1:5 positive:negative ratios. The same total amount of annotated samples was maintained to compare performance with similar annotation efforts. The number of samples chosen per iteration for the different ratios is shown in Table 1. To reach the desired class balance during iterative training, samples were annotated until the desired quota for each class was reached. When the quota for one class was met, additional samples of that class were temporarily ignored during querying until the other class reached its quota. This approach avoided discarding annotated samples while ensuring the correct balance in the training set for each iteration.

Table 1
Number of samples selected per class at each training iteration for different class balance ratios, along with the corresponding iteration numbers and ratios.

Ratio (pos:neg)	Label	Iteration						
		1	2	3	4	5	6	7
1:1	Positive	5	10	15	20	25	50	75
	Negative	5	10	15	20	25	50	75
	Total	10	20	30	40	50	100	150
1:2	Positive	5	10	15	20	25	50	
	Negative	10	20	30	40	50	100	
	Total	15	30	45	60	75	150	
1:5	Positive	5	10	15	20	25		
	Negative	25	50	75	100	125		
	Total	30	60	90	120	150		

2.3.5. Agile modelling for sourcing training data

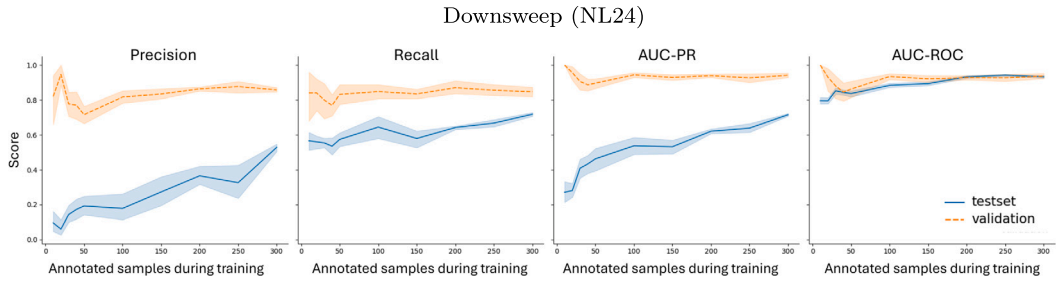
Transfer learning is known to improve performance in data-scarce conditions; however, some studies show that sometimes specialised classifiers can outperform more general application classifiers (Pérez-Granados, 2023). This experiment tests whether training a specialised convolutional neural network (CNN) from scratch with samples collected by using Agile Modelling could be an application, and whether the performance of such a model would outperform the obtained agile model. We used the package OpenSoundscape (v0.13.0) (Lapp et al., 2023) to train binary detectors based on the EfficientNet-B0 architecture (Tan and Le, 2019) using the samples found when training our baseline model. For a fair comparison, the training was performed using similar parameters: learning rate = 0.001, batch size = 12, 128 epochs, with no data augmentations or hyperparameter tuning. The 5-second audio clips (sampling rate: 32 kHz) were transformed into feature representations using a Mel-scale filterbank followed by Per-Channel Energy Normalisation (PCEN) (Wang et al., 2017). The Short-Time Fourier Transform (STFT) was computed using an FFT window size of 1024 samples ($n_{\text{stft}} = 513$), a window length of 640 samples, and a hop length of 320 samples. The linear frequency spectrum was mapped to 128 Mel bands ($n_{\text{mels}} = 128$) across a frequency range of 60 Hz to 16,000 Hz. PCEN was subsequently applied to the magnitude mel-spectrogram using the following hyperparameters: gain (α) = 0.8, bias (δ) = 10, power (r) = 0.25, time constant = 0.06, and eps $\epsilon = 10^{-6}$.

2.3.6. Generalisation of a detector during a domain shift

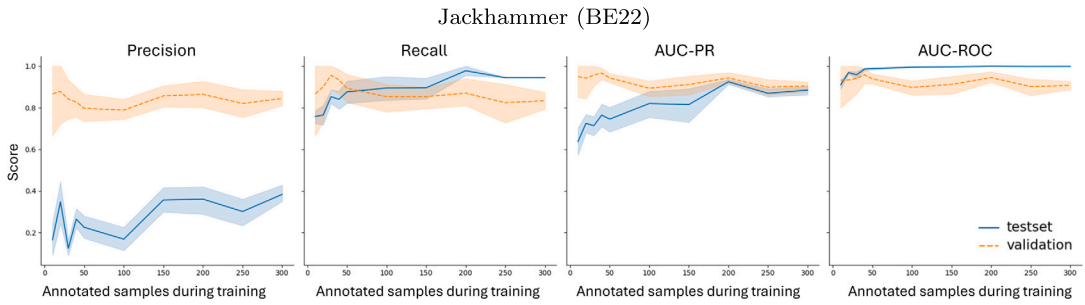
While Agile Modelling is designed to quickly train detectors on a given dataset, this process can still be time-consuming. It was therefore valuable to explore whether a detector trained for one deployment can be reused in a similar task that involved a domain shift. For example, a user might wonder whether a detector trained using data from a specific period of the year can be effectively applied to data from the same location recorded at a different period, where soundscapes may be similar due to geography but vary with seasonal or annual changes. In this experiment, we evaluated this possibility by applying

Table 2
Summary of the annotations of the three test sets.

Test set	Annotated events	Model	Total windows	Positive windows	% of positive window
BE22	44	(Surf)Perch	7,320	18	0.20
		BirdNET	12,217	23	0.18
BE21	102	SurfPerch	7,320	94	1.28
NL24	77	(Surf)Perch	7,198	47	0.65
		BirdNET	11,998	56	0.47



(a) Downsweep (NL24).



(b) Jackhammer (BE22).

Fig. 4. Results of the baseline experiment, comparing the performance of models on the validation set and test set during iterative training with Agile Modelling. The shaded areas and the lines represent, respectively, the 95% confidence intervals and the mean values obtained from the 5-fold cross-validation.

the detector trained during the baseline experiment for Jackhammer sound detection on the BE22 dataset to the BE21 dataset. Because no previous annotations were available for the Jackhammer sound in the BE21 deployment, the Agile Modelling approach was first used to identify one day with positive samples within the entire deployment. Exploration was done using two approaches: first, applying the Agile Modelling method using the reference sound from the BE22 dataset, and second, the baseline model trained on the BE22 dataset data was used to predict the presence of Jackhammer sounds. Once a positive day was identified, a randomly selected 10-hour test set was sampled from that day using the same strategy as for BE22 and NL24, and was then fully manually annotated.

3. Results

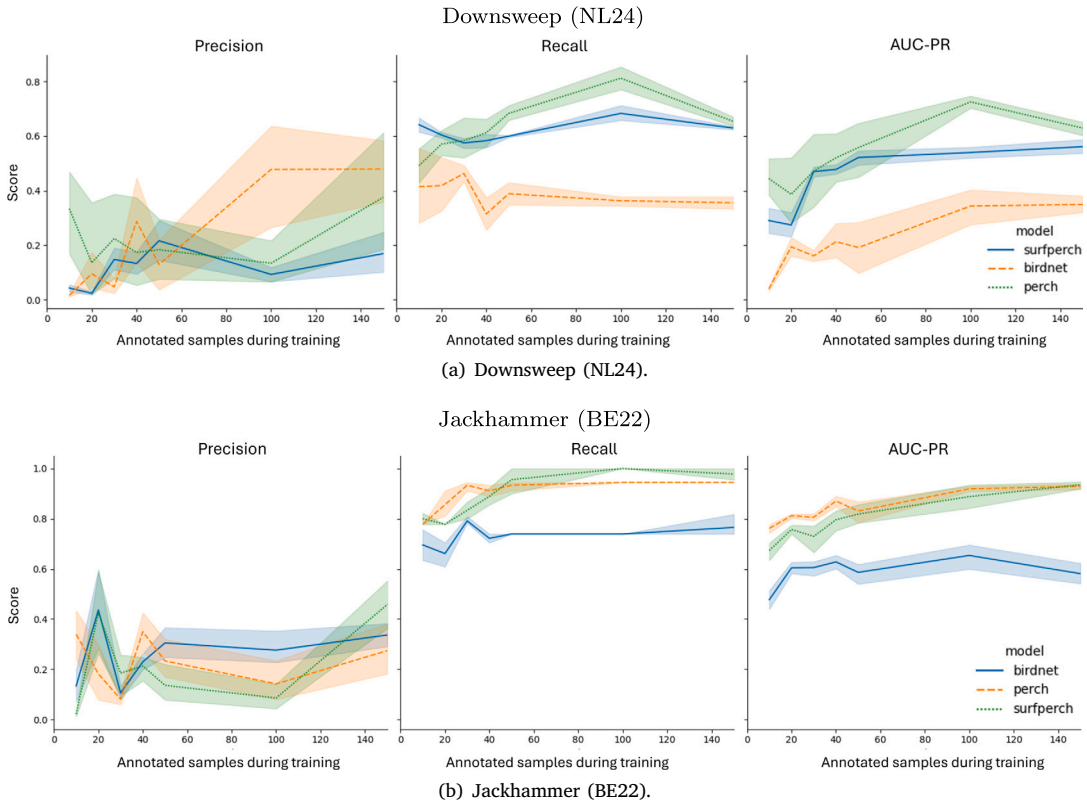
We studied the performance of Agile Modelling for detecting putative fish sounds in two temperate marine environments: Jackhammer sounds in Belgium, and Downsweep sounds in the Netherlands, and the impact of different variations of the workflow on performance. Manual annotation of the test sets revealed highly imbalanced class distributions, with positive samples comprising in all three datasets less than 1.5% positive samples (see Table 2 for further details).

3.1. Baseline experiment

For the baseline experiments, 300 samples were annotated in 10 iterations, with an equal proportion of positive and negative samples at each step, and detectors were trained using embeddings from SurfPerch. The results of the baseline experiment can be seen in Fig. 4.

The model's performance on the test set improved as more samples were added, as shown in Fig. 4 by increases in precision, recall, and AUC-PR. Although AUC-ROC is widely used for classifier evaluation in bioacoustics, it is known to be sensitive to class imbalance. In such cases, the metric can be dominated by the majority (negative) class, masking the true performance on the positive class and providing a misleading assessment for heavily imbalanced datasets. This effect was visible in the baseline experiments, when precision and recall were initially low, AUC-ROC remained high early in training. Subsequent improvements in other metrics (e.g., precision and recall) were not reflected in AUC-ROC, demonstrating its inadequacy for imbalanced scenarios. In contrast, AUC-PR offered a more informative measure by focusing exclusively on the positive class, making it better suited for the datasets of this study. The False Positive Rate for the final model was around 3% for the Downsweep and 1% for the Jackhammer sounds.

Across both tested datasets, BE22 and NL24, we observed that all metrics except AUC-ROC exhibit significant differences between the validation and test sets (Fig. 4). These results suggested that validation scores provided an over-optimistic assessment of model performance, probably due to the high imbalance in the data and the fact that the background noise is very variable. Furthermore, the AUC-ROC score on the test set did not reliably reflect the quality of the detectors, as demonstrated in the baseline experiment on the BE dataset, where the AUC-ROC score remained constant across iterations while other metric scores improved. For these reasons, we then focused solely on Precision, Recall, and AUC-PR scores from the test set in the following experiments.



(a) Downsweep (NL24).

(b) Jackhammer (BE22).

Fig. 5. Performance comparison of different embedding models on the BE and NL datasets as the number of annotated samples increases at each step of the sound detectors training. The shaded areas and the lines represent, respectively, the 95% confidence intervals and the mean values obtained from the 5-fold cross-validation.

3.2. Experiment 1: Embedding models

Comparing different embedding models, Perch and SurfPerch performed better than BirdNET on both the NL24 and BE22 datasets (Fig. 5). SurfPerch slightly outperformed Perch in Precision and Recall for Jackhammer detection (BE22), whereas Perch marginally surpassed SurfPerch in AUC-PR for Downsweep detection (NL24). Both models showed overall similar performance.

3.2.1. Visualisation of the embedding space for qualitative analysis

The target sound samples, Jackhammer or Downsweep, are not separated from negative samples in the UMAP projection of the embeddings from the different pre-trained models (see Fig. 6). Some local clustering is observed, particularly for the Jackhammer samples, but the overall distribution remains mixed. These patterns are consistent across the models evaluated.

3.3. Experiment 2: Sampling strategies

Fig. 7 shows that the sample picking strategy had an important effect on the performance of the detector. Detectors trained using the Extremities Picking strategy (selecting samples with high confidence) demonstrated no capability in detecting the target sounds, as shown by the precision score of nearly 0 across all iterations. This is most likely due to different data distribution between the test set and the training set, leading to overfitting of the model.

3.4. Experiment 3: Training set balance

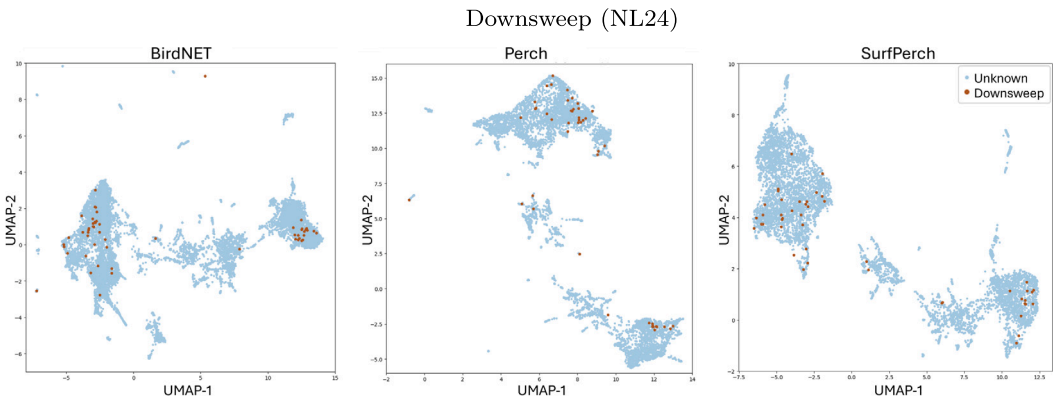
We observed large variations in the performances of Precision and Recall depending on the ratio of numbers of positive and negative samples used (see Fig. 8), which can be attributed to the detector's threshold, fixed at 0.50. This indicated a different compromise between precision and recall at different ratios. The AUC-PR score, independent from the detector's threshold, showed no difference in performance across the different balance ratios in the training set (See Fig. 8).

3.5. Agile modelling for sourcing training data

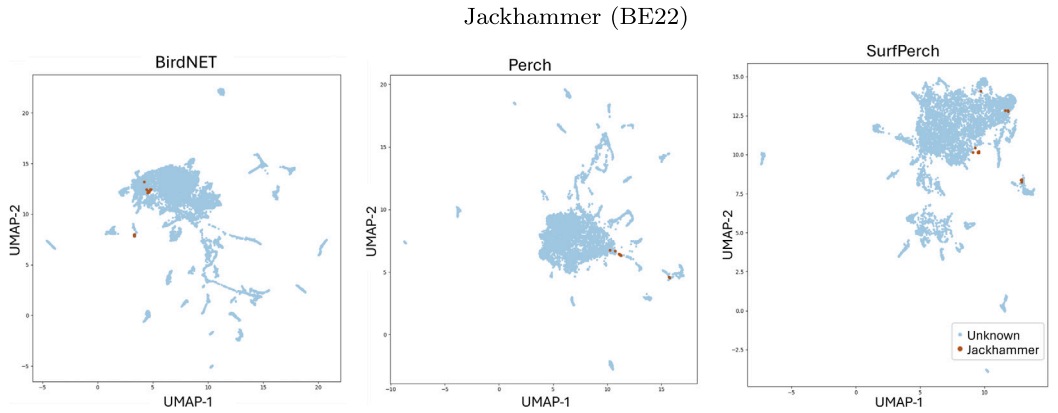
Training an EfficientNet-B0 model from scratch using agilely sourced data yielded lower F1 scores (Rijsbergen, 1979) compared to the Agile Modeling baseline across both target sounds (Table 3). This performance gap was primarily driven by lower Precision in the scratch-trained network. For example, on the Jackhammer dataset, the EfficientNet-B0 reached a Recall of 1.00 but a Precision of 0.20, compared to 0.38 Precision for the baseline. However, the AUC-PR remained comparable between the two approaches. For the Downsweep sound, AUC-PR was nearly identical (0.71 vs. 0.70), while the scratch-trained EfficientNet-B0 slightly outperformed the baseline on the Jackhammer sound (0.90 vs. 0.88). These contrasting metric trends indicate that while the EfficientNet-B0 successfully learned discriminative features, its default decision threshold was suboptimal for F1 score optimisation.

3.6. Generalisation of a detector during a domain shift

Before manually annotating an additional test set to evaluate the generalisation capacity of a detector trained with Agile Modelling, it was necessary to find periods with positive detections. Two methods were tested to detect Jackhammer sounds in the BE21 dataset. The first approach applied the Agile Modelling method using the Jackhammer reference sound from BE22 data, while the second utilised the model



(a) Downsweep (NL24).



(b) Jackhammer (BE22).

Fig. 6. Two-dimensional embedding space of the different pre-trained models applied to the test set of the NL24 and BE22 data, respectively, generated using the dimensionality reduction technique UMAP. The colour of the point indicates the class of the samples. Some negative outliers were removed to improve visualisation.

Table 3
Comparison of an Agile Modelling detector and an EfficientNet-B0 model's performance on the test set for the detection of the Downsweep and the Jackhammer sound, trained using the same samples, collected during the baseline experiment. Best results on each dataset are shown in bold. AM stands for Agile Modelling.

Sound	Approach	Precision	Recall	F1	AUC-PR
Downsweep (NL)	AM Baseline	**0.53**	0.72	**0.61**	**0.71**
Downsweep (NL)	EfficientNet-B0	0.36	**0.73**	0.47	0.70
Jackhammer (BE)	AM Baseline	**0.38**	**0.998**	**0.55**	0.88
Jackhammer (BE)	EfficientNet-B0	0.20	1.00	0.33	**0.90**

trained for the baseline experiment on BE22 data. Using the model from the baseline experiment was less time-consuming and provided better predictions, and it resulted in the detection of the 11th of March as a day with positive detections. Afterwards, 10 h were sampled from this day by randomly sampling sixty 10-minute files, and manually annotated using the same annotation protocol as for the test set of BE22 to constitute a new test set.

The model trained with 300 samples from the baseline was then evaluated on the obtained test set from the BE21 dataset. Results can be seen in Table 4. The obtained results in this deployment, which had not been seen by the model before, were lower than the ones obtained in the test set created from the same deployment. In more detail, only the recall performance decreased, meaning that the model missed more calls. The domain shift leading to the decreased performance could then be due to a decreased SNR in BE21, or a slight difference in the calls, which could be coming from another individual or population, or be attributed to seasonal differences.

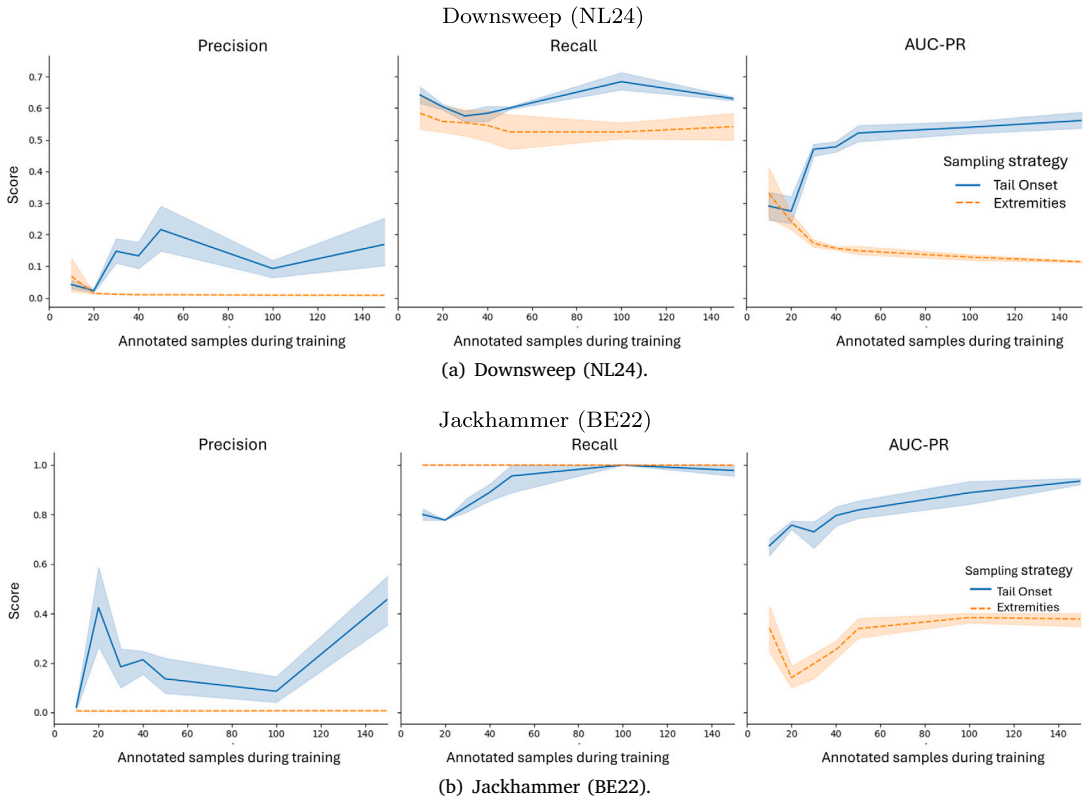
Table 4
Comparison of the results obtained in the test set from BE22 (same deployment as the data used for training) and the results obtained in the test set from BE21, which was never seen by the model.

	Precision	Recall	F1	AUC-PR
Test set BE22	0.38	0.998	0.55	0.88
Test set BE21	0.38	0.83	0.51	0.68

Due to the length of the datasets and the time-intensive nature of manual annotation, no complete manual analysis had been conducted initially. During the iterative training process, previously undetected non-target sound events were discovered, revealing sound types absent from the original annotations.

4. Discussion

The first aim of this study was to evaluate whether Agile Modelling, used in bird acoustics and in tropical marine soundscapes, can be applied to temperate marine environments to train automatic detectors of putative fish sounds. In addition, we aimed to provide guidelines on using the workflow in such environments using comparative experiments and potential applications. Specifically, we compared different pre-trained models, sampling strategies, training set balance, generalisation capacity and training data sourcing with a comparison between a model trained with the Agile Modelling workflow and trained from scratch.



(a) Downsweep (NL24).

(b) Jackhammer (BE22).

Fig. 7. Performance comparison of different sample-picking strategies on the BE and NL datasets as the number of annotated samples increases, at each step of the sound detectors training. The shaded areas and the lines represent, respectively, the 95% confidence intervals and the mean values obtained from the 5-fold cross-validation.

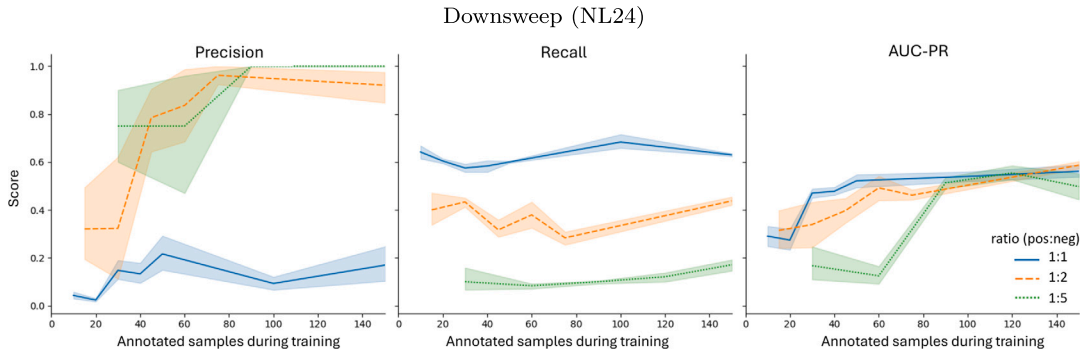
Results from the baseline experiment (Fig. 4) demonstrate that the Agile Modelling workflow effectively detects sparse putative fish sounds in temperate marine environments. Starting with only a single example and requiring less than two hours of training, including manual annotation, the binary detectors achieved F1 scores of 0.55 (BE22) and 0.61 (NL24). This is a meaningful result given the extremely low prevalence of target sounds in these environments (0.2% to 1.2% of test recordings), where traditional linear or random sampling methods are too labour-intensive to be usable.

While the F1 score is useful for model comparison, the True Positive Rate (TPR/Recall) and False Positive Rate (FPR) can be more meaningful for downstream applications. For complex bioacoustic tasks with severe class imbalance, a TPR superior to 0.7 and an FPR inferior to 0.01 are generally desired to consider a detector effective (Schall et al., 2024). Detectors trained in the baseline experiments achieved the target TPR (>0.7), with an FPR of 0.05 for Downsweep sounds and under 0.02 for Jackhammer sounds. While these baseline FPRs require refinement before full automation – potentially by raising the confidence threshold or introducing a brief manual validation step – the current models are highly effective as pre-screening tools. For instance, applying the NL24 detector as a pre-screening filter captures 72% of all true sound events. Although human validation is still required to separate the false positives, the workflow dramatically accelerates data curation: 53% of the pre-detected events are true positives, compared to a baseline prevalence of less than 2% in the raw data. Thus, the Agile Modelling approach can greatly accelerate the collection of training data, especially when the target sounds are rare.

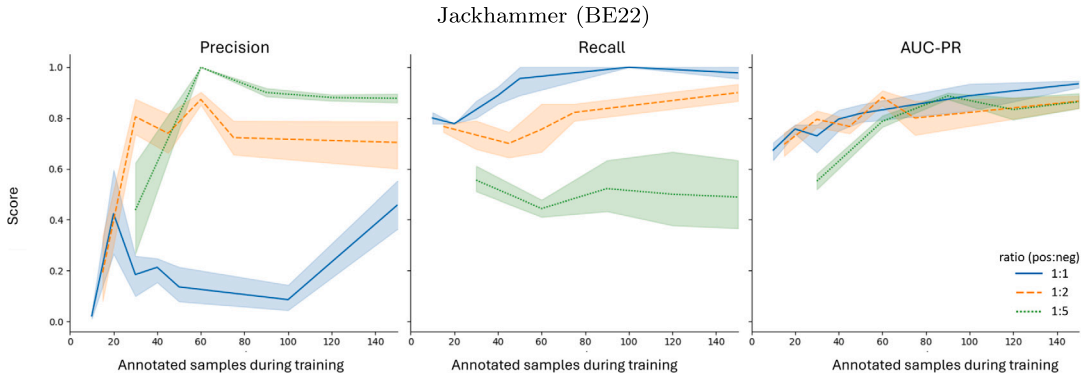
New methods have recently been developed to further accelerate annotation and data sourcing, particularly those combining active learning with label propagation in clusters generated by unsupervised learning (Parcerisas et al., 2024). These approaches allow for a rapid evaluation of different soundscapes and enable annotation across entire datasets, making them highly efficient. However, the performance of such unsupervised methods cannot be established or re-applied to analyse other data, and the clustering does not guarantee that the sound of interest will be considered a single class (Parcerisas et al., 2024).

As demonstrated for the detection of the Jackhammer sound in BE21 (see Section 2.3.5), Agile Modelling can also be particularly valuable in identifying ecologically relevant periods (e.g. seasonal spawning) where sounds may occur only during a specific week or month within a year. Identifying these periods can support important conservation decisions. Visualising long spectrograms can help detect transient events when they produce noticeable changes in sound pressure levels across some frequency bands, but this method does not indicate the type of sound detected (Ryan et al., 2021; Duane et al., 2024; Schoeman et al., 2022). The obtained detectors trained using Agile Modelling can then be used to identify periods when specific sound events occur. Even though further analysis is necessary, the generalisation experiment suggests that a pre-trained model can be applied to data at the same location from a different season or year (see Table 4). One main limitation of the approach remains that one would have to train one detector for each sound of interest, which can be time-consuming even given the efficiency of the workflow.

The identification of previously overlooked sound events highlights a secondary benefit of using a human-in-the-loop workflow, an advantage likely shared by other iterative active learning processes. Rather than following a standard linear analysis, this iterative annotation loop directs human verification toward the model's uncertainty zone, effectively filtering out both obvious background noise and clear target sounds. This targeted sampling makes the discovery of novel or rare acoustic classes highly probable during the training phase itself. This effect is particularly valuable in understudied environments, such as temperate marine habitats, where undocumented acoustic signals are



(a) Downsweep (NL24).



(b) Jackhammer (BE22).

Fig. 8. Performance comparison of different ratios of the number of positive and negative samples in the training set, on the BE and NL datasets, as the number of annotated samples increases at each step of the sound detectors training. The shaded areas and the lines represent, respectively, the 95% confidence intervals and the mean values obtained from the 5-fold cross-validation.

likely to emerge. Ultimately, while all deep learning outputs require human verification, integrating active learning into the training pipeline can accelerate the discovery of these hidden sound types. Further work is needed to quantify the extent of this effect compared to traditional sampling methods.

Our comparative analysis demonstrates that the models Perch and SurfPerch exhibit superior performance over BirdNET, when used as pre-trained models for Agile Modelling to detect putative fish sounds in temperate environments, suggesting better generalisation capabilities 5. This finding contrasts with recent work on cross-domain feature extractor performance using frog vocalisations (Kather et al., 2025), highlighting the domain-specific nature of bioacoustic generalisation. Several acoustic factors may represent challenges for the generalisation from bird to fish sounds, as temporal and spectral characteristics differ substantially. For example, most fish sounds occur below 4 kHz while bird-focused models analyse bandwidths extending to 16–24 kHz. Contrary to BirdNET, Perch and SurfPerch are not only trained on bird data but also on amphibians, insects, and mammals, which could result in a smaller domain gap and better generalisation. Nonetheless, despite being trained on underwater sounds, representing a smaller domain gap, SurfPerch did not outperform Perch. This suggests that SurfPerch's specialisation in tropical reef sound may not provide an advantage when generalising to other marine environments, compared to the better generalisability of Perch (Williams et al., 2025).

Displaying the embedding space (Fig. 6) shows that the feature extractors provide limited separability for these datasets, likely constraining detector performances. However, since UMAP is a non-linear projection technique that may not preserve global distances, this observation does not necessarily imply inseparability in the original feature space. Models pre-trained on general tasks could benefit from further training on fish sound data, which exhibit unique characteristics such as low-frequency and variable durations, including short pulses. However, annotated fish sound datasets, especially from temperate marine environments, are nearly non-existent to date. While reference sounds for some species exist, they are typically limited to a few examples per sound type, insufficient for training deep learning models. More generalist models developed from diverse bioacoustics downstream applications, such as AVES (Hagiwara, 2023), BioLingual (Robinson et al., 2023), or Nature-LM (Robinson et al., 2024), are promising candidates for improved performance and should be evaluated in future work.

The proposed method, as most of the single-call deep learning detectors for fish sounds (Barroso et al., 2023), performs detection at the window level of a duration pre-set by the feature extractor, in weak labels, i.e., where the sound is known to occur within a window but not precisely localised. Weak labelling has been identified as a limitation for training high-performance classification models in bird acoustics (Ghani et al., 2023, 2025). Other feature extractors could allow for shorter windows, better suited to fish sounds. We recommend using at least 1 to 2 s, as shorter durations can be hard for humans to interpret during training. Transformer-based models that support variable input lengths (Hagiwara, 2023; Robinson et al., 2023) may offer a ready-to-use alternative. Alternatively, post-processing steps can help isolate relevant sound segments or define bounding boxes more accurately. Techniques from computer vision, such as YOLO models, have already proven useful in underwater acoustics (Parcerisas et al., 2024; Best et al., 2025). Other strategies, based on the selection of the most relevant parts of recordings, such as max-pooling or energy-based selection of segments, have been used to improve the use of weak labels (Ghani et al., 2025) and could serve as effective preprocessing for training new models.

Training set balance did not affect overall AUC-PR but influenced precision and recall, indicating a trade-off that can be adjusted through

decision thresholds rather than class rebalancing, depending on the detector's purpose or study objectives. Because maintaining class balance is time-consuming, we recommend annotating all queried examples. However, limiting negative samples risks underrepresenting the diversity of background sounds (e.g., boat noise, currents, anthropogenic interference), which may reduce detection performance and generalisation. Confirming this hypothesis will require further research.

Sampling from the least-confidence region improved performance, whereas training on the most confident samples failed to produce a functional classifier. This outcome contrasts with the simulated experiment done in a previous study, in which high-confidence sampling accelerated learning in low-data regimes with rare call occurrences (Dumoulin et al., 2025); although our dataset also contains few positive instances, no quantification of class imbalance in the original study could be found to compare the data regimes. Selecting high-confidence samples likely yields low performance because all positive samples contain distinct background noise or specific patterns (e.g., echosounders in the NL data) alongside the target signal. Additionally, the severe dataset imbalance prevents the selected negative samples from fully representing the true diversity of the background noise

The successful strategy, sampling in the uncertainty region, depends on the user accurately identifying this region in the logit score distribution, which limits reproducibility. Automating this step through an active-learning sampling strategy could both remove user-defined thresholds and potentially improve performance, provided that the method ensures the acquisition of both positive and negative samples for training convergence. Combining quantile-based sampling with selecting the most informative samples yielded the best results across different data regimes in previous simulated experiments (Dumoulin et al., 2025). Including diversity-based sampling to explore the entire embedding space is also crucial, as leaving some positive samples undiscovered can bias the classifier. Future work should therefore evaluate hybrid uncertainty- and diversity-based sampling strategies, together with more advanced active-learning approaches for agile modelling workflows, following recent evaluation frameworks (Ren et al., 2021; Lüth et al., 2023).

One of the challenges in applying Agile Modelling lies in reliably assessing detector performance. While using dedicated test sets is effective, it undermines the method's primary advantage of minimising annotation efforts, particularly for highly imbalanced datasets, due to the necessity of having sufficient positive samples for reliable estimation of the performance. Nevertheless, monitoring model performance remains essential to identify when to stop the iterative training if performance stagnates, to optimise the time investment in model development, and for ecological applications. Our results demonstrate performance improvements with additional samples, showing a trend toward convergence between validation and test metrics (see Fig. 4), though confirming the trend and finding a recommended sample size will require more research and will likely be task-specific. Future work should also explore regularisation strategies, such as hyperparameter optimisation, data augmentation, dropout, and label smoothing, to reduce discrepancies between validation and test scores and potentially mitigate overfitting.

Current approaches to estimating performance face distinct limitations for precision versus recall. While precision can be reasonably estimated through manual verification of detected events, accurately measuring recall in imbalanced datasets proves far more challenging, as it requires examining impractically large volumes of data to confirm undetected target sounds. In the study presenting Agile Modelling (Dumoulin et al., 2025), using Bayesian reasoning and validation of samples with different levels of certainty, it is possible to obtain an estimate of bird sounds detector performance and a direct estimate of call density, even with imperfect detectors, suggesting similar potential applications for aquatic acoustic monitoring (Navine et al., 2024). However, this approach proves incompatible with our study context, where AUC-ROC is inefficient and where AUC-PR does not admit such a probabilistic interpretation.

In practical applications, we recommend a hybrid solution combining Agile Modelling with targeted validations. For long-term monitoring scenarios requiring reliable precision/recall metrics, researchers might first employ quickly trained agile models to identify candidate periods containing target sounds, then perform focused manual annotation on these subsets. This strategy was applied successfully to constitute the dataset BE21 at the BE location, where a comprehensive manual review would otherwise have been extremely time-consuming. However, it is important to note that any pre-detection step introduces selection bias that may influence performance estimation when applied to real data.

The results in Section 3.4 (see Table 3) show that the model trained from scratch achieves AUC-PR performance comparable to the detector trained using Agile Modelling when both use the same data. This similarity suggests that the performance limits observed in the Agile Modelling workflow are unlikely to stem solely from a lack of information in the embedding space. Instead, the performance differences highlight a distinct architectural trade-off. The Agile Modelling baseline benefits from a frozen SurfPerch feature extractor that leverages rich, pre-trained bioacoustic representations. Visual inspection of the embedding space also reveals only partial data separability, suggesting that this fixed embedding space contributes to limiting the performance of the agile modelling detectors. In contrast, training an EfficientNet-B0 updates the entire parameter space, offering greater structural flexibility. However, given the small number of training samples available from the agile modelling session, full-network optimisation from scratch introduces a severe risk of overfitting, which likely explains the combination of maximum Recall and low Precision observed for the Jackhammer target.

Ultimately, these experiments do not yet demonstrate a clear advantage in training a classifier directly on data sourced through Agile Modelling versus using the classifier obtained within the workflow. However, the model trained from scratch was evaluated without hyperparameter optimisation, data augmentation, or additional regularisation techniques; implementing these standard strategies presents a potential pathway to mitigate overfitting and improve precision. Looking ahead, a promising direction for future work is to fine-tune a pretrained model (without freezing the feature extractor weights) using annotations collected through Agile Modelling, as this approach has been shown to outperform training from scratch in data-scarce bioacoustic contexts (Ghani et al., 2023). If effective, this fine-tuning step could be incorporated as a final stage in the workflow when a specialised, high-performance classifier is the end goal.

Several constraints currently collectively still hinder the application of automated data analysis methods to PAM for fish: existing models are narrowly applicable to specific contexts, their development demands substantial manually annotated datasets, and their implementation requires specialised machine learning expertise (Barroso et al., 2023; Ibrahim et al., 2024; Laplante et al., 2022). When applying Agile Modelling to the detection of putative fish sounds, these limitations are addressed by providing an accessible entry point that eliminates the need for initial training data while leveraging advanced techniques (e.g., transfer learning, active learning) through a user-friendly script (Bordoux, 2026). By reducing both the annotation burden and technical threshold, this approach facilitates access to state-of-the-art detection methods, enabling researchers to efficiently develop context-specific detectors without prerequisite datasets or extensive machine learning training. While not eliminating the need for eventual validation, this paradigm shift lowers the barrier to implementing performant bioacoustic monitoring across diverse marine ecosystems.

5. Future research

Identifying fish species by sound is fundamentally limited by the scarcity of reference recordings in accessible databases, which remains a major challenge for applying PAM to fish. Recent methodological advances employ innovative techniques such as audio-video arrays and cross-referencing sounds from different locations to identify novel species-specific sounds (Mouy et al., 2023; Vieira et al., 2024; Dantzker et al., 2025). However, these approaches are still constrained by their dependence on manual sound annotation, which currently limits their scalability, and rely on enough visibility (which is a rarity in the southern part of the North Sea) to be able to visually assign a ground-truth. Improved analytical methods like Agile Modelling, shown here to be effective across diverse sound types and environments, can significantly accelerate the annotation of recordings. The collection of reference sounds should be a collaborative effort, making one or multiple annotated examples per labelled sound type accessible, e.g. FishSounds (Looby et al., 2023), MarineSoundLib (Flanders Marine Institute (VLIZ), 2026), which can in turn be used as input for detection modelling.

Agile Modelling strongly relies on the performance of the feature extractor used for training high-performing detectors. As demonstrated previously (lack of separability in UMAP embeddings), current feature extractors may be limited for putative fish sounds in temperate marine environments. Enhancing feature extractors is key and can be done in several ways. Models like SurfPerch trained on a combination of sounds from birds, tropical reef sounds, and other sources show improved in-domain classification, suggesting that a fish-specific extractor would better support cross-environment and cross-species performance. Building general-purpose fish sound detectors will require broad collaboration to assemble representative datasets across diverse soundscapes. Infrastructure for large-scale data sharing, such as that proposed by Parsons et al. (2024) or Darras et al. (2025), could facilitate this effort. Due to the limited availability of annotated datasets for fish sounds, self-supervised learning presents a particularly promising approach, enabling models to learn from large volumes of unannotated audio data. While several general-purpose models have incorporated data from a range of taxa—including those trained using self-supervised paradigms (Hagiwara, 2023; Robinson et al., 2023, 2024), fish sounds have not been included in either training or evaluation. This omission reflects both the recent state of research in this field and the current lack of publicly available fish sound datasets.

Finally, given that the method has proven effective in two ecologically and acoustically distinct marine environments, tropical coral reefs (Dumoulin et al., 2025), which are rich and diverse in biophony, and temperate coastal marine environments, which are comparatively sparse and less variable, it is reasonable to infer that the method is broadly applicable across a spectrum of marine soundscapes. Preliminary studies on the detection of putative fish sounds in the Adriatic Sea suggest the applicability of Agile Modelling to other, unrelated marine environments (Bordoux et al. in preparation).

6. Conclusion

This study demonstrates, for the first time, that Agile Modelling is a highly effective approach for detecting putative fish sounds in temperate marine environments. With only a single annotated example, models can be developed in under two hours and then applied to weeks or months of acoustic data, underscoring the potential of Agile Modelling to accelerate fish sound research and monitoring. For contexts similar to the datasets presented here, we recommend using Perch or Surfperch models, sampling from boundary regions during iterations, and annotating both classes without strict balance requirements. When detector performance evaluation is required, a subset of the deployment can serve as a test set, for which we provide scripts for generation and evaluation. In cases of severe class imbalance, AUC-PR offers a robust performance metric.

Because state-of-the-art general models cannot currently achieve high-performance detection across most fish sounds, it remains highly relevant to develop distinct, specialised detectors for individual call types, which is traditionally time-consuming. This study validates that Agile Modelling can be an efficient method for mitigating this bottleneck. Nonetheless, further improvements to this approach are needed to overcome the low precision scores and high false-positive rates inherent to detecting rare signals in severely imbalanced datasets, which will ultimately help to further reduce the manual labour required for detection validation.

Beyond enabling efficient model training, this workflow can facilitate rapid sourcing of training data, and detectors may be reused for similar acoustic conditions, though further research is needed to quantify these benefits. A major bottleneck for advancing the field remains the lack of available annotated datasets, which makes sharing and open repositories especially important. By reducing the time and cost of creating annotated datasets, Agile Modelling makes dataset production more feasible and scalable, encouraging greater data sharing and collaboration. In lowering barriers to annotation and detector development, this work helps overcome a critical limitation in passive acoustic monitoring of fish, particularly in temperate marine systems, thereby enabling broader spatial and temporal studies.

CRediT authorship contribution statement

Valentin Bordoux: Writing – review & editing, Writing – original draft, Visualization, Software, Investigation, Data curation, Conceptualization. **Clea Parcerisas:** Writing – review & editing, Writing – original draft, Software, Investigation, Data curation. **Martin Jälmby:** Writing – review & editing, Software, Investigation. **Elisabeth Debusschere:** Writing – review & editing, Resources, Funding acquisition. **Albertinka J. Murk:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Rosa M. van der Ven:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Next Level Animal Science innovation program of Wageningen University and Research. We would like to thank Lea Kornau and Lodewijk van Walraven for the data collection at the NL location. The authors thank the Simon Stevin RV crew for their help with the deployment and retrieval of the acoustic instruments to collect the data in the BE location. The work of CP was funded by FWO, Belgium, grant number 1281026N.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ecoinf.2026.103874.

Data availability

The code used for the experiments and for reuse, as well as the full results of the iterative training experiments, are available on the GitHub repository Bordoux (2026). Recordings, metadata, and annotations for the NL location are available Bordoux and van der Ven (2025). The BE recordings and metadata are available at Parcerisas et al. (2021).

References

Allen-Ankins, S., et al., 2025. The use of BirdNET embeddings as a fast solution to find novel sound classes in audio recordings. Front. Ecol. Evol. 12, http://dx.doi.org/10.3389/fevo.2024.1409407, Publisher: Frontiers.

Amorim, M.C.P., et al., 2023. Detection of invasive fish species with passive acoustics: Discriminating between native and non-indigenous sciaenids. Mar. Environ. Res. 188, 106017. http://dx.doi.org/10.1016/j.marenvres.2023.106017.

Balcan, M.-F., et al., 2007. Margin based active learning. In: Bshouty, N.H., Gentile, C. (Eds.), Learning Theory. Springer, Berlin, Heidelberg, pp. 35–50. http://dx.doi.org/10.1007/978-3-540-72927-3_5.

Barroso, V.R., et al., 2023. Applications of machine learning to identify and characterize the sounds produced by fish. ICES J. Mar. Sci. 80 (7), 1854–1867. http://dx.doi.org/10.1093/icesjms/fsad126.

Basan, F., et al., 2024. The underwater soundscape of the North Sea. Marine Poll. Bull. 198, 115891. http://dx.doi.org/10.1016/j.marpolbul.2023.115891.

Best, P., et al., 2025. Analysing vocal complexity in relation to sociality in orcas of British Columbia: An application of long-term computational passive acoustics. Ecol. Inform. 90, 103211. http://dx.doi.org/10.1016/j.ecoinf.2025.103211.

Bolgan, M., et al., 2023. Use of passive acoustic monitoring to fill knowledge gaps of fish global conservation status. Aquat. Conserv.: Mar. Freshw. Ecosyst. 33 (12), 1580–1589. http://dx.doi.org/10.1002/aqc.4020.

Bolgan, M., et al., 2025. Soundscape and fish passive acoustic monitoring around a North Sea gas-production platform in the Dogger Bank. PLoS One 20 (4), e0319536. http://dx.doi.org/10.1371/journal.pone.0319536, Publisher: Public Library of Science.

Bordoux, V., 2026. vbordoux/SuPer_sound_detection_AL: Training classifier on top of Surfperch as feature extractor using an active learning approach. URL: https://github.com/vbordoux/SuPer_sound_detection_AL.

Bordoux, V., van der Ven, R.M., 2025. Passive acoustic recordings from Texel-Oudeschild harbour during september 2024. http://dx.doi.org/10.5281/zenodo.16900684, URL: https://zenodo.org/records/16900684.

Calonge, A., et al., 2024. Revised clusters of annotated unknown sounds in the Belgian part of the North sea. Front. Remote. Sens. 5, http://dx.doi.org/10.3389/frsen.2024.1384562, Publisher: Frontiers.

Chérubin, L.M., et al., 2020. Fish spawning aggregations dynamics as inferred from a novel, persistent presence robotic approach. Front. Mar. Sci. 6.

Dantzker, M.S., et al., 2025. Deciphering complex coral reef soundscapes with spatial audio and 360° video. Methods Ecol. Evol. 16 (11), 2622–2637. http://dx.doi.org/10.1111/2041-210X.70149, _eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.70149.

Darras, K.F.A., et al., 2025. Worldwide soundscapes: A synthesis of passive acoustic monitoring across realms. Glob. Ecol. Biogeogr. 34 (5), e70021. http://dx.doi.org/10.1111/geb.70021.

Di Iorio, L., et al., 2021. Biogeography of acoustic biodiversity of NW Mediterranean coralligenous reefs. Sci. Rep. 11 (1), 16991. http://dx.doi.org/10.1038/s41598-021-96378-5, Publisher: Nature Publishing Group.

Duane, D., et al., 2024. Moonlight-driven biological choruses in Hawaiian coral reefs. PLoS One 19 (3), e0299916. http://dx.doi.org/10.1371/journal.pone.0299916, Publisher: Public Library of Science.

Dumoulin, V., et al., 2025. The search for squawk: Agile modeling in bioacoustics. http://dx.doi.org/10.48550/arXiv.2505.03071, URL: http://arxiv.org/abs/2505.03071, arXiv:2505.03071 [eess].

Farahani, A., et al., 2021. A brief review of domain adaptation. In: Stahlbock, R., et al. (Eds.), Advances in Data Science and Information Engineering. Springer International Publishing, Cham, pp. 877–894. http://dx.doi.org/10.1007/978-3-030-71704-9_65.

Fetterplace, L.C., et al., 2022. Evidence of sound production in wild stingrays. Ecology 103 (11), e3812. http://dx.doi.org/10.1002/ecy.3812.

Flanders Marine Institute (VLIZ), 2026. Marine SoundLib. URL: https://www.marinesoundlib.org/.

Ghani, B., et al., 2023. Global birdsong embeddings enable superior transfer learning for bioacoustic classification. Sci. Rep. 13 (1), 1–14. http://dx.doi.org/10.1038/s41598-023-49989-z, Publisher: Nature Publishing Group.

Ghani, B., et al., 2025. Impact of transfer learning methods and dataset characteristics on generalization in birdsong classification. Sci. Rep. 15 (1), 16273. http://dx.doi.org/10.1038/s41598-025-00996-2, Publisher: Nature Publishing Group.

Ghosh, K., et al., 2024. The class imbalance problem in deep learning. Mach. Learn. 113 (7), 4845–4901. http://dx.doi.org/10.1007/s10994-022-06268-8.

Google, 2026a. Google | Perch | Kaggle. URL: https://www.kaggle.com/models/google/bird-vocalization-classifier.

Google, 2026b. Google | SurfPerch | Kaggle. URL: https://www.kaggle.com/models/google/surfperch.

Hagiwara, M., 2023. AVES: Animal vocalization encoder based on self-supervision. In: ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, pp. 1–5. http://dx.doi.org/10.1109/ICASSP49357.2023.10095642.

Ibrahim, A.K., et al., 2020. Transfer learning for efficient classification of grouper sound. J. Acoust. Soc. Am. 148 (3), EL260–EL266. http://dx.doi.org/10.1121/10.0001943.

Ibrahim, A.K., et al., 2024. Fish acoustic detection algorithm research: a deep learning app for Caribbean grouper calls detection and call types classification. Front. Mar. Sci. 11, http://dx.doi.org/10.3389/fmars.2024.1378159, Publisher: Frontiers.

Jarriel, S., et al., 2024a. Unidentified fish sounds as indicators of coral reef health and comparison to other acoustic methods. Front. Remote. Sens. 5, http://dx.doi.org/10.3389/frsen.2024.1338586.

Jarriel, S., et al., 2024b. Compilation of existing underwater PAM repositories, libraries, and applications for sound processing. http://dx.doi.org/10.5281/zenodo.12096525, URL: https://zenodo.org/records/12096525.

Kahl, S., et al., 2021. BirdNET: A deep learning solution for avian diversity monitoring. Ecol. Inform. 61, 101236. http://dx.doi.org/10.1016/j.ecoinf.2021.101236.

Kahl, S., et al., 2025. BirdNET Model V2.4. Publisher: Zenodo, http://dx.doi.org/10.5281/zenodo.15050749.

Kath, H., et al., 2024a. Active learning in multi-label classification of bioacoustic data. In: Hotho, A., Rudolph, S. (Eds.), KI 2024: Advances in Artificial Intelligence. Springer Nature Switzerland, Cham, pp. 114–127. http://dx.doi.org/10.1007/978-3-031-70893-0_9.

Kath, H., et al., 2024b. Leveraging transfer learning and active learning for data annotation in passive acoustic monitoring of wildlife. Ecol. Inform. 82, 102710. http://dx.doi.org/10.1016/j.ecoinf.2024.102710.

Kather, V.S., et al., 2025. Clustering and novel class recognition: evaluating bioacoustic deep learning feature extractors. http://dx.doi.org/10.48550/arXiv.2504.06710, URL: http://arxiv.org/abs/2504.06710, arXiv:2504.06710 [cs].

Kim, E.B., et al., 2023. SoundScape learning: An automatic method for separating fish chorus in marine soundscapesa). J. Acoust. Soc. Am. 153 (3), 1710–1722. http://dx.doi.org/10.1121/10.0017432.

Kingma, D.P., Ba, J., 2017. Adam: A method for stochastic optimization. http://dx.doi.org/10.48550/arXiv.1412.6980, URL: http://arxiv.org/abs/1412.6980, arXiv:1412.6980 [cs].

Laplante, J.-F., et al., 2022. Deep learning for marine bioacoustics and fish classification using underwater sounds. In: 2022 IEEE Canadian Conference on Electrical and Computer Engineering. CCECE, (ISSN: 2576-7046) pp. 288–293. http://dx.doi.org/10.1109/CCECE49351.2022.9918242.

Lapp, S., et al., 2023. OpenSoundscape: An open-source bioacoustics analysis package for Python. Methods Ecol. Evol. 14 (9), 2321–2328. http://dx.doi.org/10.1111/2041-210X.14196.

Liang, J., et al., 2024. Mind the domain gap: A systematic analysis on bioacoustic sound event detection. In: 2024 32nd European Signal Processing Conference. EUSIPCO, (ISSN: 2076-1465) pp. 1257–1261. http://dx.doi.org/10.23919/EUSIPCO63174.2024.10714948.

Lindseth, A.V., Lobel, P.S., 2018. Underwater soundscape monitoring and fish bioacoustics: A review. Fishes 3 (3), 36. http://dx.doi.org/10.3390/fishes3030036, Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

Looby, A., et al., 2023. FishSounds Version 1.0: A website for the compilation of fish sound production information and recordings. Ecol. Inform. 74, 101953. http://dx.doi.org/10.1016/j.ecoinf.2022.101953.

Looby, A., et al., 2024. Fish sound production research: Historical practices and ongoing challenges. In: The Effects of Noise on Aquatic Life. Springer, Cham, pp. 109–128. http://dx.doi.org/10.1007/978-3-031-50256-9_92.

Lüth, C., et al., 2023. Navigating the pitfalls of active learning evaluation: A systematic framework for meaningful performance assessment. Adv. Neural Inf. Process. Syst. 36, 9789–9836.

McEwen, B., et al., 2024. Active few-shot learning for rare bioacoustic feature annotation. Ecol. Inform. 82, 102734. http://dx.doi.org/10.1016/j.ecoinf.2024.102734.

McGeady, R., et al., 2023. A review of new and existing non-extractive techniques for monitoring marine protected areas. Front. Mar. Sci. 10, http://dx.doi.org/10.3389/fmars.2023.1126301, Publisher: Frontiers.

McInnes, L., et al., 2020. UMAP: Uniform manifold approximation and projection for dimension reduction. http://dx.doi.org/10.48550/arXiv.1802.03426, URL: https://arxiv.org/abs/1802.03426, arXiv:1802.03426.

Miller, B.S., et al., 2020. An annotated library of underwater acoustic recordings for testing and training automated algorithms for detecting Antarctic blue and fin whale sounds. http://dx.doi.org/10.26179/5e6056035c01b.

Mooney, T.A., et al., 2020. Listening forward: approaching marine biodiversity assessments using acoustic methods. R. Soc. Open Sci. 7 (8), 201287. http://dx.doi.org/10.1098/rsos.201287.

Mouy, X., et al., 2023. Identification of fish sounds in the wild using a set of portable audio-video arrays. Methods Ecol. Evol. 14 (8), 2165–2186. http://dx.doi.org/10.1111/2041-210X.14095.

Mouy, X., et al., 2024. Automatic detection of unidentified fish sounds: a comparison of traditional machine learning with deep learning. Front. Remote. Sens. 5, http://dx.doi.org/10.3389/frsen.2024.1439995.

Munger, J.E., et al., 2022. Machine learning analysis reveals relationship between pomacentrid calls and environmental cues. Mar. Ecol. Prog. Ser. 681, 197–210. http://dx.doi.org/10.3354/meps13912.

Navine, A.K., et al., 2024. All thresholds barred: direct estimation of call density in bioacoustic data. Front. Bird Sci. 3, http://dx.doi.org/10.3389/fbirs.2024.1380636, URL: https://www.frontiersin.orghttps://www.frontiersin.org/journals/bird-science/articles/10.3389/fbirs.2024.1380636/full.

Nguyen, V.-L., et al., 2022. How to measure uncertainty in uncertainty sampling for active learning. Mach. Learn. 111 (1), 89–122. http://dx.doi.org/10.1007/s10994-021-06003-9.

Parcerisas, C., et al., 2021. Broadband acoustic network dataset. URL: https://www.vliz.be/en/imis?module=dataset&dasid=7879, https://www.lifewatch.be/en/broadband-acoustic-network.

Parcerisas, C., et al., 2023a. Studying the soundscape of shallow and heavy used marine areas: Belgian part of the North Sea. In: Popper, A.N., et al. (Eds.), The Effects of Noise on Aquatic Life. Springer International Publishing, Cham, pp. 1–27. http://dx.doi.org/10.1007/978-3-031-10417-6_122-1.

Parcerisas, C., et al., 2023b. Categorizing shallow marine soundscapes using explained clusters. J. Mar. Sci. Eng. 11 (3), 550. http://dx.doi.org/10.3390/jmse11030550, Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

Parcerisas, C., et al., 2024. Machine learning for efficient segregation and labeling of potential biological sounds in long-term underwater recordings. Front. Remote. Sens. 5, http://dx.doi.org/10.3389/frsen.2024.1390687, Publisher: Frontiers.

Parsons, M.J.G., et al., 2024. A global library of underwater biological sounds (GLUBS): An online platform with multiple passive acoustic monitoring applications. In: Popper, A.N., et al. (Eds.), The Effects of Noise on Aquatic Life: Principles and Practical Considerations. Springer International Publishing, Cham, pp. 2149–2173. http://dx.doi.org/10.1007/978-3-031-50256-9_123.

Pérez-Granados, C., 2023. BirdNET: applications, performance, pitfalls and future opportunities. Ibis 165 (3), 1068–1075. http://dx.doi.org/10.1111/ibi.13193.

Ren, P., et al., 2021. A survey of deep active learning. ACM Comput. Surv. 54 (9), 180:1–180:40. http://dx.doi.org/10.1145/3472291.

Rice, A.N., et al., 2022. Evolutionary patterns in sound production across fishes. Ichthyol. Herpetol. 110 (1), 1–12. http://dx.doi.org/10.1643/i2020172, Publisher: The American Society of Ichthyologists and Herpetologists.

Rijsbergen, C.J.V., 1979. Information retrieval, second ed. Butterworths, London, Boston, Open Library ID: OL4739711M.

Robinson, D., et al., 2023. Transferable models for bioacoustics with human language supervision. URL: http://arxiv.org/abs/2308.04978, arXiv:2308.04978 [cs, eess, q-bio].

Robinson, D., et al., 2024. NatureLM-audio: an audio-language foundation model for bioacoustics. http://dx.doi.org/10.48550/arXiv.2411.07186, URL: http://arxiv.org/abs/2411.07186, arXiv:2411.07186 [cs].

Ryan, J.P., et al., 2021. Reduction of low-frequency vessel noise in monterey bay national marine sanctuary during the COVID-19 pandemic. Front. Mar. Sci. 8.

Sayigh, L., et al., 2017. The watkins marine mammal sound database: An online, freely accessible resource. Proc. Meet. Acoust. 27 (1), 040013. http://dx.doi.org/10.1121/2.0000358.

Schall, E., et al., 2024. Deep learning in marine bioacoustics: a benchmark for baleen whale detection. Remote. Sens. Ecol. Conserv. http://dx.doi.org/10.1002/rse2.392.

Schoeman, R.P., et al., 2022. Analysis of soundscapes as an ecological tool. In: Erbe, C., Thomas, J.A. (Eds.), Exploring Animal Behavior Through Sound: Volume 1: Methods. Springer International Publishing, Cham, pp. 217–267. http://dx.doi.org/10.1007/978-3-030-97540-1_7.

Schwinger, R., et al., 2025. Foundation models for bioacoustics – a comparative review. http://dx.doi.org/10.48550/arXiv.2508.01277, URL: http://arxiv.org/abs/2508.01277, arXiv:2508.01277 [cs].

Settles, B., 2012. Active learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, Springer International Publishing, Cham, http://dx.doi.org/10.1007/978-3-031-01560-1.

Souza, Jr., P.M., et al., 2023. Paired passive acoustic and gillnet sampling reveal the utility of bioacoustics for monitoring fish populations in a turbid estuary. ICES J. Mar. Sci. 80 (5), 1240–1255. http://dx.doi.org/10.1093/icesjms/fsad085.

Stowell, D., 2022. Computational bioacoustics with deep learning: a review and roadmap. PeerJ 10, e13152. http://dx.doi.org/10.7717/peerj.13152, Publisher: PeerJ Inc..

Sugai, L.S.M., et al., 2019. Terrestrial passive acoustic monitoring: Review and perspectives. BioScience 69 (1), 15–25. http://dx.doi.org/10.1093/biosci/biy147.

Tan, M., Le, Q., 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. In: Proceedings of the 36th International Conference on Machine Learning. PMLR, (ISSN: 2640-3498) pp. 6105–6114.

Vieira, M., et al., 2024. Cross-referencing unidentified fish sound data sets to unravel sound sources: a case study from the Temperate Northern Atlantic. Front. Remote. Sens. 5, http://dx.doi.org/10.3389/frsen.2024.1377206, Publisher: Frontiers.

Wang, Y., et al., 2017. Trainable frontend for robust and far-field keyword spotting. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, (ISSN: 2379-190X) pp. 5670–5674. http://dx.doi.org/10.1109/ICASSP.2017.7953242.

Watson, M.S., et al., 2025. The biological soundscape of temperate reefs in the Wadden sea. Sci. Rep. 15 (1), 9216. http://dx.doi.org/10.1038/s41598-025-92955-0, Publisher: Nature Publishing Group.

Williams, B., et al., 2025. Using tropical reef, bird and unrelated sounds for superior transfer learning in marine bioacoustics. Phil. Trans. R. Soc. B 380 (1928), 20240280. http://dx.doi.org/10.1098/rstb.2024.0280, Publisher: Royal Society.

Wilson, K.C., et al., 2019. Development and evaluation of a passive acoustic localization method to monitor fish spawning aggregations and measure source levels. In: OCEANS 2019 MTS/IEEE SEATTLE. (ISSN: 0197-7385) pp. 1–7. http://dx.doi.org/10.23919/OCEANS40490.2019.8962663.

Xeno-Canto, T., xeno-canto :: Sharing wildlife sounds from around the world. URL: https://xeno-canto.org/.

Zhuang, F., et al., 2021. A comprehensive survey on transfer learning. Proc. IEEE 109 (1), 43–76. http://dx.doi.org/10.1109/JPROC.2020.3004555.