Heliyon 9 (2023) e20275

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Review article

Systematic review of machine learning methods applied to ecoacoustics and soundscape monitoring

D.A. Nieto-Mora [a,*], Susana Rodríguez-Buritica [b], Paula Rodríguez-Marín [a], J.D. Martínez-Vargaz [c], Claudia Isaza-Narváez [d]

[a] MIRP-Instituto Tecnológico Metropolitano ITM, Cl. 54a N∘30-01, Medellín, Colombia
[b] Instituto Alexander Von Humboldt, Calle 28A N∘15-09, Bogotá, Colombia
[c] Universidad EAFIT, Cra 49, Cl. 7 Sur N∘50, Medellín, Colombia
[d] SISTEMIC-Universidad de Antioquia UdeA, Cl. 67 N∘53-108, Medellín, Colombia

ARTICLE INFO

Keywords:
Deep learning
Ecoacoustics
Ecological monitoring
Machine learning
Soundscape ecology

ABSTRACT

Soundscape ecology is a promising area that studies landscape patterns based on their acoustic composition. It focuses on the distribution of biotic and abiotic sounds at different frequencies of the landscape acoustic attribute and the relationship of said sounds with ecosystem health metrics and indicators (e.g., species richness, acoustic biodiversity, vectors of structural change, gradients of vegetation cover, landscape connectivity, and temporal and spatial characteristics). To conduct such studies, researchers analyze recordings from Acoustic Recording Units (ARUs). The increasing use of ARUs and their capacity to record hours of audio for months at a time have created a need for automatic processing methods to reduce time consumption, correlate variables implicit in the recordings, extract features, and characterize sound patterns related to landscape attributes. Consequently, traditional machine learning methods have been commonly used to process data on different characteristics of soundscapes, mainly the presence–absence of species. In addition, it has been employed for call segmentation, species identification, and sound source clustering. However, some authors highlight the importance of the new approaches that use unsupervised deep learning methods to improve the results and diversify the assessed attributes. In this paper, we present a systematic review of machine learning methods used in the field of ecoacoustics for data processing. It includes recent trends, such as semi-supervised and unsupervised deep learning methods. Moreover, it maintains the format found in the reviewed papers. First, we describe the ARUs employed in the papers analyzed, their configuration, and the study sites where the datasets were collected. Then, we provide an ecological justification that relates acoustic monitoring to landscape features. Subsequently, we explain the machine learning methods followed to assess various landscape attributes. The results show a trend towards label-free methods that can process the large volumes of data gathered in recent years. Finally, we discuss the need to adopt methods with a machine learning approach in other biological dimensions of landscapes.

* Corresponding author.
E-mail addresses: danielnieto@itm.edu.co (D.A. Nieto-Mora), drodriguez@humboldt.org.co (S. Rodríguez-Buritica), paularodriguez@itm.edu.co (P. Rodríguez-Marín), jdmartinev@eafit.edu.co (J.D. Martínez-Vargaz), victoria.isaza@udea.edu.co (C. Isaza-Narváez).

https://doi.org/10.1016/j.heliyon.2023.e20275
Received 25 February 2023; Received in revised form 12 September 2023; Accepted 18 September 2023

Available online 22 September 2023

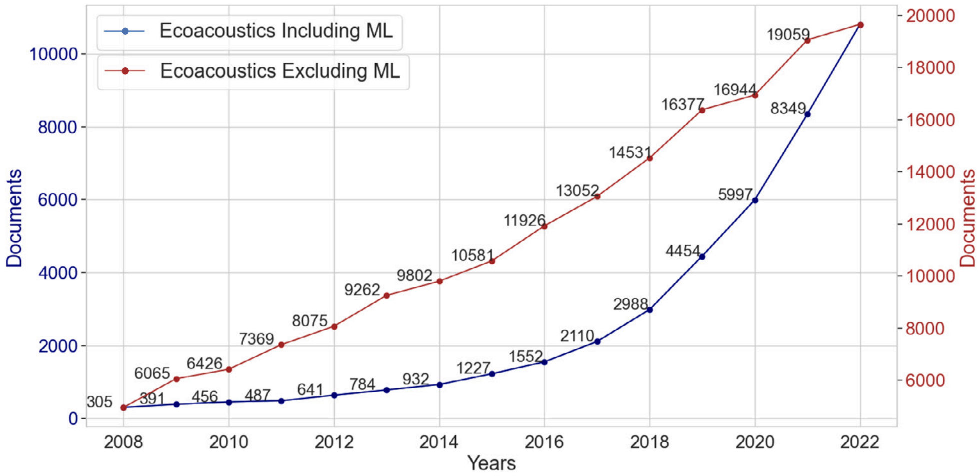

Fig. 1. Comparison of publication trends for ecoacoustic data analysis: Machine Learning (blue line) vs Non-Machine Learning Approaches (red line). Machine learning studies have exponential growth, while traditional methods have linear growth. The search equation was created using the interception of the words 'soundscape', 'ecoacoustics', and 'bioacoustics' with the sentence 'machine learning'.

1. Introduction

Habitat and biodiversity loss is a global concern that represents a threat to life on Earth. Different organizations such as the United Nations Environment Programme (UNEP), the International Union for Conservation of Nature (IUCN), and the World Wide Fund for Nature (WWF) report that human actions have severely damaged nature, causing a 68% decline in the populations of mammals, birds, fish, and reptiles [1]; disturbances in 75% of the landscapes and 66% of the seascapes [2]; and loss of 85% of the wetlands and 32% of the forests [3]. Consequently, many countries and organizations have signed international agreements to combat climate change [4–6], prioritizing conservation and restoration plans to mitigate the loss of nature.

One of the best methods to contribute to biodiversity conservation is monitoring [7]. It has been an important tool to study landscape elements such as heterogeneity [8], species richness [9,10], space occupancy [11], and structural composition [12]. Traditional methods have typically consisted of field campaigns conducted by experts who collect data on the behavior and natural patterns of the animal populations under study [13]. However, the development of sensing devices [14–16] and the implementation of automatic and remote techniques are enhancing the effectiveness of the studies.

Passive monitoring methods using sensing devices optimize time and efforts, reduce costs and operation risks, and increase data collection capacity, thus expanding available data for landscape ecology [17]. Most of these methods use satellites [18], Unmanned Aerial Vehicles (UAVs) [19], camera traps [20], and on-animal sensors to collect data at different distances; however, all of them have some limitations. Satellite imagery may be especially focused on the structural composition of a landscape, considering that it mainly captures information about cover heterogeneity, vegetation, water sources, and human settlements, while leaving aside the analysis of animal species. Moreover, satellite images have limited temporal continuity, which prevents a reliable analysis of the functional process of landscapes. Aerial monitoring is expensive compared to other methods and difficult to access [15,21]. Camera trapping lacks generality across geographies [17] and is particularly targeted at medium-sized and large species.

Recently, studies based on landscape-level acoustic recordings have proven helpful in identifying ecological patterns [10] in response to environmental changes, given that the acoustic behaviors are thought to reflect the state of the ecosystems. Passive Acoustic Monitoring (PAM) use acoustic recorder units (ARUs) deployed in the field for days or weeks [22] to capture landscape acoustic data at different spatial scale. PAM is non-invasive, cost-efficient compared to other methods, and flexible in terms of time and space [23]. According to this approach, acoustic activity is an attribute of the elements that make up an ecosystem and is present in the three dimensions of biodiversity: composition, structure, and function [24]. Thus, by means of ecoacoustic, bioacoustic, and soundscape analyses, researchers can detect and quantify the presence, reproductive activity, territorial behavior, and interactions of species, as well as compositional and structural characteristics of animal communities [25,26]. For this reason, new studies on the acoustic features of multiple landscapes are constantly being published, which favors biodiversity monitoring. However, this growing amount of information [27,28] poses the challenge of efficiently processing large volumes of data that favor ecosystem characterization and management [29,30].

Over the past 15 years, machine learning techniques have become an efficient tool to analyze big data, and their application in passive acoustic monitoring has demonstrated effectiveness in processing acoustic recordings for different ecological purposes. Fig. 1 shows the annual increase in ecoacoustic studies using machine learning since 2008 (blue line) compared to the annual increase of studies that do not use machine learning.

From the review of these studies, we identified three main methods for analyzing ecoacoustic, bioacoustic, and soundscape data: manual examination of acoustic events by listening to recordings or visually inspecting spectrograms [29], use of acoustic indices to summarize variations in acoustic energy [31,32], and automatic recognition of sonotypes using machine learning algorithms [33,34]. However, these methods have weaknesses that are worth mentioning. The manual processing of large acoustic datasets is impractical due to its time-consuming nature. To overcome this limitation, researchers often rely on acoustic indices as proxies for richness and diversity metrics [35]. However, the use of acoustic indices presents certain challenges. They can be sensitive to noise, require human expertise to select appropriate indices for a specific study, and there is a lack of consensus among researchers regarding their interpretation. Furthermore, the information captured and perceived by acoustic indices may vary depending on the specific ecosystem being studied. Finally, machine learning using supervised techniques achieves good results in ecoacoustic tasks such as species detection and noise reduction [36]; however, it requires data to be manually labeled, which is a time-consuming task that also hinders the comprehensive analysis of the recordings. As a result, various techniques have been proposed in the last three years to analyze acoustic recordings without supervision. These unsupervised methods are label-free and can process a broader range of biological information contained in the acoustic recordings, which makes them particularly useful for monitoring the soundscape [31].

Aware of the importance of acoustic monitoring for biodiversity conservation and restoration, and considering the current challenges in terms of data processing, noise reduction, characterization, information extraction, data interpretation, and linking information to biological questions, this review explores the implementation of machine learning techniques in soundscape analyses. In this sense, the objectives of the study are as follows:

1. Highlight the ecological components and issues commonly studied employing soundscape recordings to feed machine learning and statistical methods.
2. Identify the main machine learning algorithms and approaches used in soundscape ecology.
3. Find future directions for the use of machine learning techniques in the analysis of acoustic recordings and biodiversity conservation.

This paper provides a systematic review of the machine learning techniques used to analyze acoustic properties and patterns of landscapes. These analyses typically focus on landscape structure and composition, species identification, noise reduction, source separation, and sound-type segmentation techniques. In other words, they deal with a few components with particular time and frequency characteristics that make up an acoustic community, which include animal vocalizations and geophonic sources (e.g., rainfall, thunders, water, etc). This research paper aims to address the existing gap in the literature by providing a comprehensive review of machine learning methods applied to ecoacoustics and soundscape data analysis. Currently, there is a lack of a consolidated guide that comprehensively explores the state-of-the-art machine learning techniques utilized in this field. By conducting a systematic review, we aim to fill this gap and establish a reference point for researchers and practitioners. Furthermore, our review seeks to identify and describe the principal ecological attributes that have been assessed using these machine learning methods. The findings of this study will not only enhance our understanding of the application of machine learning in ecoacoustics but also guide future research endeavors in this rapidly evolving area.

2. Methods

For our research, we used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [37] to determine the search equation, eligibility criteria, and structure of this document so that the discussion and results would have a standard format. We systematically searched available databases accessible with institutional credentials such as EBSCO, IEEE Xplore, Taylor & Francis Online, and Scopus. We selected Science Direct to carry out the search thanks to its scope that supports the main topics of this research, which are related to engineering and environmental sciences and, specifically, to machine learning, ecoacoustics, and soundscape ecology.

We defined the search equation using an Excel tool designed by Elsevier to combine keywords according to exclusion criteria such as main words, secondary words, excluded words, word proximity, and word location in the document. To this end, we selected keywords by areas. For instance, relating to machine learning, we used *machine learning*, *supervised learning*, *unsupervised learning*, *reinforcement learning*, and *deep learning*. On the subject of soundscape, we used *soundscape, bioacoustics, ecoacoustics*, and *acoustic monitoring*. Thus, the general equation was: ((soundscape OR bioacoustics OR ecoacoustics OR acoustic monitoring) AND (machine learning OR supervised learning OR reinforcement learning OR unsupervised learning OR deep learning)). In addition, we implemented some word filters in the abstract and keywords. Specifically, we used *landscape passive monitoring, acoustic biodiversity*, *biophonies*, *geophonies*, and *anthropophonies*.

Having defined the equation, we conducted the search on ScienceDirect. The major search was done on December 28, 2021, obtaining 178 records. Initially, we checked the language of the publications, considering that the leading researchers in the two fields of study publish their papers in English. We included open-access and subscription publications accessible through the Instituto Tecnológico Metropolitano membership to academic databases. A total of 141 records were selected after the filtering criteria. We also updated the publications close to the submission of the manuscript maintaining the inclusion/exclusion criteria and filtering process. We included recent information in our discussion, figures, tables and results. The last search was done on November 1, 2022, where we included ten new records. Considering the last search, an overall of 194 papers were found.

A single reviewer performed the screening using ASReview, a tool designed for transparent systematic reviews [38]. Specifically, we screened 35 documents, of which we manually labeled 25 as relevant and 10 as irrelevant to our study. A machine learning algorithm then ranked the documents from most to least relevant in an Excel file. Once arranged, we checked each title and abstract for inclusion or exclusion. We found in the Excel file that, after a certain point, the papers were correctly marked as irrelevant, which helped us to accurately complete the screening process. Finally, we downloaded 111 full papers (update included).

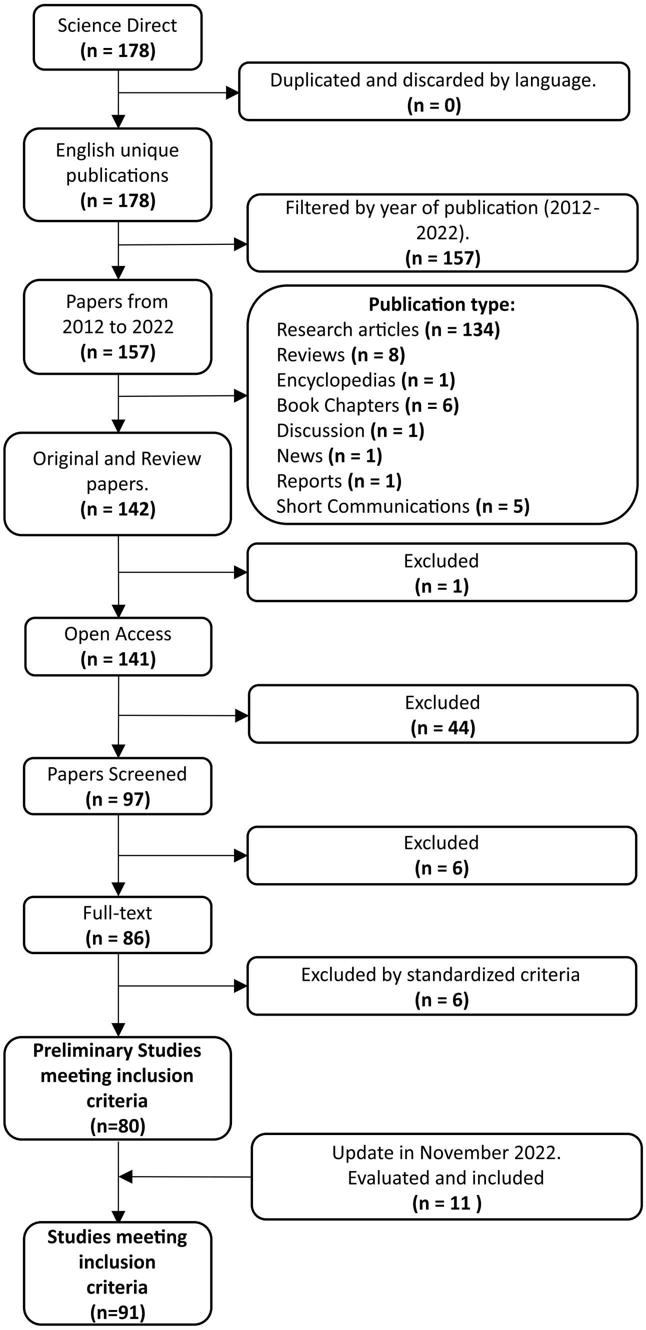

Fig. 2. Search and selection process of the systematic review following the PRISMA statement.

Before thoroughly reading the screened documents, we defined the eligibility criteria. Thus, we established the following contents as mandatory for an article to be included in the study: i) information about the assessed biocomplexity dimension (heterogeneity, connectivity, metapopulations, structural components); ii) recording information (sampling rates, recorders, recording periods, study sites, datasets); iii) audio processing or transformation; iv) methods section describing the used machine learning algorithms or statistical analysis performed. Consequently, we decided not to include papers with purely machine learning or purely biological approaches. Fig. 2 shows the selection process from the database search to the eligibility criteria. Finally, we chose 90 papers for the study. The reasons for exclusion were:

- **Purely machine learning approach or inadequate biological description (n = 5).** The usual misclassified topics were neuroscience and vocal learning.

- **Machine learning applied to irrelevant biological studies (n = 4).** The studies found were related to human perception of sounds.
- **Review with a slightly related approach (n = 4).** Reviews where ecoacoustics approach is mentioned but it is not the principal topic, or it is only used to compare with other methodologies.
- **Purely biological approach (n = 3).** The studies were mainly focused on urban planning.

3. Results

We present the findings of the study based on the analysis of the 90 papers selected to respond to the three main points declared in the introduction:

- Most addressed ecological components.
- Most used machine learning algorithms and approaches.
- Most commented future research topics.

Moreover, at the end of the section, we include some additional aspects that we found relevant to accomplish soundscape analysis using machine learning.

3.1. Ecological components

This section embraces the most common topics studied when analyzing soundscapes using machine learning approaches and some tools, such as acoustic indices, that have been widely used to relate landscape features from soundscapes.

3.1.1. Ecological and biological patterns studied using machine learning methods

We followed the classification proposed by [39] to connect ecological concepts and perspectives of automatic soundscape analysis. This approach describes three dimensions of ecosystem complexity: heterogeneity, connectivity, and historical contingency. Therefore, based on the abstract, keywords, and results, we classified the papers into these three biocomplexity axes. We found that most of them were related to heterogeneity, only two were slightly related to contingency or metapopulations dynamics [11], and one was related to connectivity [40]. The lack of research on these axes might be because analyzing connectivity and contingency requires a detailed analysis of time and space at a scale yet to be available through the soundscape, as they are associated with animal communities. In addition, they demand less dependence on species, bypassing the need for species identification to favor rapid, replicable, and scalable assessments of biodiversity change [11]. These findings indicate that soundscape studies are starting to explore the declared biocomplexity dimensions (heterogeneity); however, there is still a need to move towards more complex ecological attributes from the soundscape.

Regarding the first and most studied biocomplexity axis, heterogeneity is defined as the number and proportion of land covers and the complexity in their composition, which is related to the richness and diversity of both animal and plant species [41,42]. Consequently, heterogeneity is relevant to monitoring because it helps understand biodiversity dynamics. Mainly, the PAM uses the concept of acoustic heterogeneity (a simile of landscape heterogeneity) to refer to the quantity, diversity, and structure of sound types and acoustic sources [43].

Some authors state that the soundscape comprises sounds from three sources: biophonies, geophonies, and anthropophonies. Biophonies are sounds emitted or produced by biotic organisms [44], such as animal calls or whistles. Geophonies are sounds from physical phenomena and noises generated by natural processes [32], such as rivers, thunders, rainfall, and wind. Finally, anthropophonies are sounds that come from human activities [25]; for example, human voice, machinery, airplanes, boats, sirens, traffic, ring, and bells. As expected, biophonies were extensively studied, followed by anthropophonies and geophonies.

Biophonies were the main focus of analysis because they can potentially improve remote biodiversity monitoring [32]. Consequently, we collected information on species and structural patterns. According to [45], soundscape ecologists are currently adopting two different approaches. The first approach consists in implementing species presence/absence surveys, while the second approach involves the application of broad-spectrum metrics to recordings.

We identified two perspectives to classify our datasets: species-specific and multi-species studies. The first perspective deals with the study of species-specific sounds. It includes studies on specific animal taxa and investigations where no distinction is made between various types of sound (identification of animal vocalizations versus other sounds, segmentation of species-specific calls, and detection of activity at determined frequencies or time segments, among others). In the second perspective, we find multi-species studies that aim to identify, classify, or cluster multiple types of sound.

In both species-specific and multi-species studies, we identified a trend toward studying the calls of birds and anurans belonging to different genera. One of the reasons to choose these species is that not all animals emit sounds, which is a problem when studying the soundscape. However, birds and anurans, among other groups of animals, can be heard frequently and are indicators of ecosystem health [46], helping identify early changes in landscapes (e.g., slight changes in abundance, species presence, etc.) [8]. In [47], authors explain that birds tend to integrate a broad spectrum of ecological factors and are the best-known indicator group for conserving and managing natural resources in tropical ecosystems. Also, [48] states that birds live in most environments and occupy almost every niche within them. Similarly, anurans are excellent early indicators of ecological stress [49], although their populations steadily decrease worldwide [34]. Furthermore, bats can be associated with connectivity conditions [50]. Even

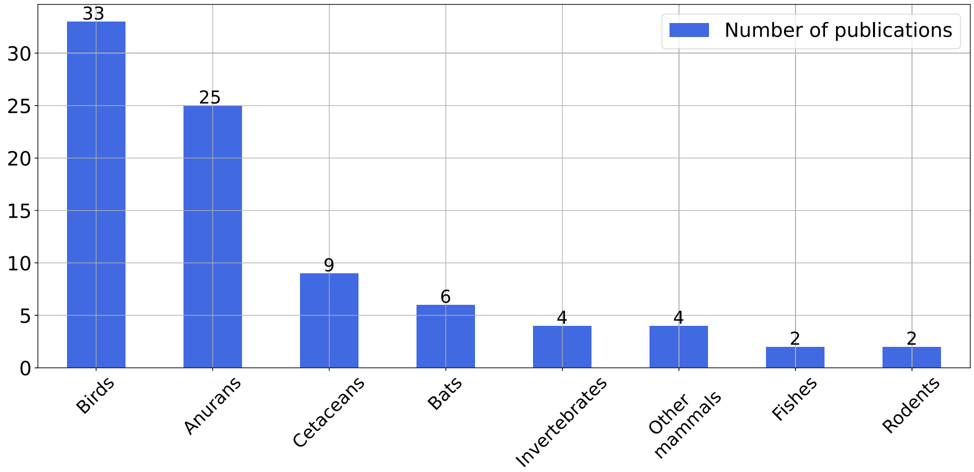

Fig. 3. Number of publications by species. We classified the species into eight general categories.

though birds and anurans are well-studied species and have intense acoustic activity, investigations that explore surrogate indicators and landscape ecology from ecoacoustics have obtained mixed results. Additionally, some authors highlight the need to explore understudied species such as fishes, insects and marine mammals [51,44]. Fig. 3 shows the variety of species found in our study.

While most papers focus on the sounds of animal species (see Table 3), some studies explore acoustic changes and similarities between landscape structures using plant species and compositional patterns [26]. For example, [12] evaluated restoration treatments using *Ochroma pyramidale* planting regimes; [52] studied oil palms plantations and the surrounding soundscapes; [42] compared intact and logged forests; [40] were interested in different forest types; [10] analyzed landscape gradients; [53] explored rocky savannas and high forests; and, lastly, [8] examined disturbed habitats. Moreover, marine structures such as coral reefs have recently been reported [54–56], which are of primary interest because they are indicators of ecosystem health and function [54]. In summary, 63 publications focused on terrestrial habitats, 12 on marine ecosystems (seascapes), and seven on farmland or urban areas.

3.1.2. Study sites

Ecological and landscape assessments using Passive Acoustic Monitoring (PAM) have been carried out in places with high biodiversity where ecological patterns need to be understood to contribute to their preservation and prevent biodiversity loss [57,56,26,25].

We found studies that use datasets from 24 countries. The country with the most publications referring to datasets collected in its territory is Australia with 16 studies [58,59,31,26,60–64,50,65,66,34,43,67,9]; followed by the United States with 14 publications [68,48,69–71,10,72,35,73,74,55,75,76,23]; Brazil with 13 publications [40,27,60,11,77,28,78–80,57,46,47,81]; Costa Rica [12,40,25,45], Malaysia [42,82–84] and the United Kingdom [22,85,32,86] with four publications each; Colombia [87,8,88] and France [85,89,90] with three publications each; Spain [85,91], Japan [54,92], Puerto Rico [93,94], and South Africa [95,96] with two publications; and, finally, Belize [97], Gabon [98], Portugal [44], South Korea [99], Sri Lanka [100], Madagascar [101], Canada [102], Ecuador [32], New Zealand [56], Argentina [33], Ivory Coast [103], Hungary [85], Thailand [82], Taiwan [104], Nepal [105], and Indonesia [52] with one publication each. Making a percentage analysis by continents, we found that 45.45% of the studies used data are from America; 20.45% from Oceania; 15.90% from Asia; 12.50% from Europe; and 5.70% from Africa.

The information above identifies the countries with the most monitoring programs which use machine learning to analyze acoustic data generated by PAM programs. Their purpose is to gather data to develop soundscape-based alternatives for monitoring and incorporate automatic and statistical analyses into their experiments. Fig. 4 shows the heat map of publications worldwide. This map includes publications that met inclusion criteria (90 articles).

Some of the most biodiverse countries, such as Brazil (the world's most diverse) and Australia, have been trying to find automatic methods to assess landscapes using acoustic recordings and data processing. However, there is a gap in the implementation of monitoring plans in other megadiverse countries such as Colombia (the second most biodiverse), Ecuador, Peru, and Venezuela, where there are no studies in progress in this field and whose ecosystems are being altered. Similarly, countries such as China, the Philippines, and Congo, which are also megadiverse, were not found in the datasets. Therefore, there is a need to bring the study of soundscapes to places with an enormous richness of species that are threatened.

3.1.3. Acoustic indices

In soundscape and ecoacoustics studies, acoustic indices play a crucial role in assessing and quantifying various acoustic properties within natural environments. While acoustic indices themselves are not classified as machine learning methods, they serve as essential tools utilized in conjunction with machine learning techniques to perform experiments and analyses. Acoustic indices provide valuable insights into the acoustic characteristics of ecosystems, aiding researchers in understanding environmental changes, biodiversity patterns, and species behavior.

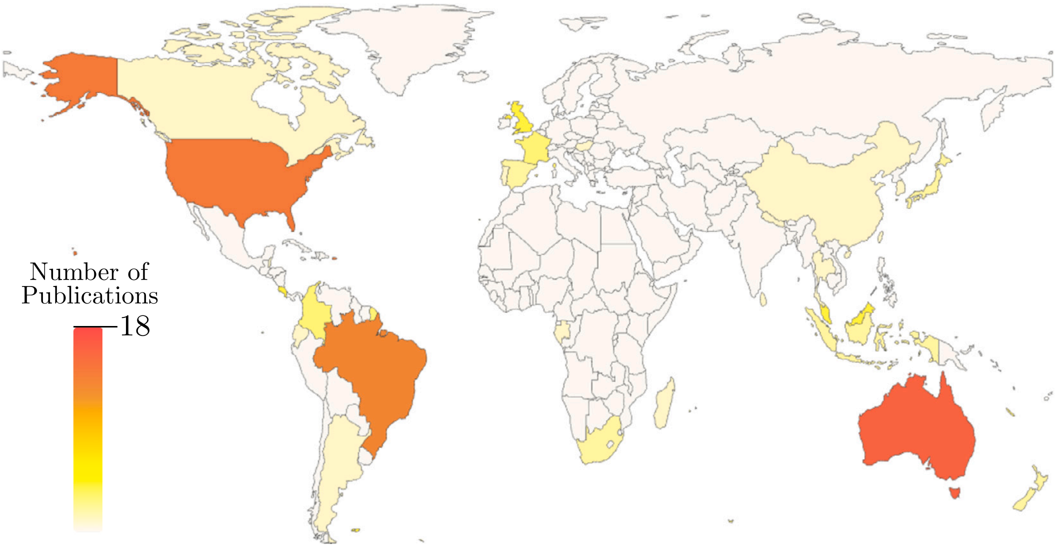

Fig. 4. Heat map of countries with soundscape datasets reported by authors.

Acoustic indices are a mathematical synthesis of the ecological and biological patterns contained in soundscape records and emerged as a practical tool for describing landscape attributes. These indices quantify amplitude and frequency differences in the recordings as a whole rather than a specific species [50]. A variety of acoustic indices has been developed by collapsing the soundscape into measures of energy distribution [11]; therefore, they can be categorized into two groups: alpha indices (within-group) and beta indices (between-group). Alpha indices, on the one hand, estimate the amplitude (intensity), evenness (relative abundance), richness (number of entities), and heterogeneity of the acoustic community [32]. On the other hand, beta indices are related to acoustic dissimilarity and express the differences between paired locations [52], providing a measure of community diversity [50]. Acoustic indices can also be classified as temporal or spectral. Temporal indices are derived from waveform envelopes, while spectral indices are based on the recordings' frequency [43].

In the publications reviewed, we identified 68 acoustic indices; however, only 27 studies (33.75%) employed these mathematical methods in their experimentation. Specifically, 21 investigations used the Acoustic Complexity Index (ACI), which compares sound intensity at subsequent time steps by calculating and summing their differences [56,26,101,10,23] to capture transient biophonic sounds [32]. Fifteen studies used the acoustic entropy (H) index (also known as ENT), related with habitat richness as it is based on the shannon information theory [59]. Nine studies employed the Normalized Difference Soundscape Index (NDSI), which quantifies disturbance based on the ratio of biophonies to anthropophonies in recordings [86]. In addition, the Acoustic Diversity Index (ADI), Acoustic Evenness Index (AEI), Bioacoustic Index (BI), Spectral Entropy (Hf) index, and Temporal Entropy (Ht) index [8] were found in more than five studies. These indices are commonly employed to find patterns in the sounds of biotic organisms. Similarly, the Background Noise (BGN), Signal-to-Noise Ratio (SNR), and amplitude (M) indices were related to anthropophonic or geophonic sounds. Other indices such as Acoustic Richness (AR), Entropy of the Spectral Peaks (EPS), Entropy of the Average Spectrum (EAS), and Horizontal Ridge Count (RHZ), among others, were reported in less than three studies. These findings indicate that, although a wide variety of indices exists, only a few are typically used. Moreover, there needs to be a consensus on how many indices and combinations achieve the best results.

Although acoustic indices are still widely used, some authors argue that they may present problems and challenges. For example, the ACI (the most frequently employed index) does not distinguish between changes in sound abundance and call diversity in some applications [56]. The H, ACI, AR, and AEI indices are broadband sensitive. Technophony (geophonies produced by human machinery) can artificially increase the ACI. Additionally, indices such as NDSI assume that species sounds are above the technophony threshold [50]. [101] found that various indices, such as the ADI, AEI, and H, are correlated, thus providing redundant information. Furthermore, results are inconsistent and vary depending on the habitat assessed [32].

3.2. Machine learning algorithms and approaches

Recently, machine learning has been employed as a paradigm for soundscape studies as it allows processing faster landscape and seascape audios to obtain high-precision results [28]. The implementation of machine learning in this field is partly due to its success in other related tasks such as audio tagging and musicology studies [45]. We found that machine-learning approaches applied to soundscape ecology involve audio preprocessing, feature extraction, classification/clustering, and source separation. These steps allow for analyzing high volumes of data gathered over weeks or months and, in some cases, collected by groups of ARUs in different landscapes. These conditions make the understanding of ecological patterns and the early detection of changes more complex.

Table 1
ARUs used to collect soundscape data and perform ecoacoustics studies.

Recorders	Count	Reference
ARBIMON	1	[11]
Audiomoth	7	[108,96,98,97,93,94,10]
Aural M2	1	[95]
AUSOMS-Mini	2	[99,54]
D-240X	1	[82]
DTAG3	1	[71]
High Tech 94 SSQ hydrophone	1	[44]
M500-384	1	[82]
Marantz PMD 620 MKII	1	[100]
Marantz PMD 660	1	[90]
Oceanpod	2	[77,80]
Olympus DM-420	1	[9]
Olympus LS-11/LS-12	1	[42]
Other devices	3	[108,91,94]
Pettersson D1000 bat detector	1	[28]
Rugged Swift Audio Recorder	2	[96,12]
SM1	1	[74]
SM2	15	[87,47,86,81,22,45,69,53,8,64,109,75,32,42,26]
SM3	9	[96,87,59,105,61,64,32,23,27]
SM4	4	[96,104,73,110]
SMBat	2	[86,28]
Total SM	31	N/A
Sony Stereo IC Recorder ICD-AX412F	1	[84]
SoundTrap ST300	4	[111,55,56,44]
Tascam DR-22WL	1	[52]
Tascam-DR05	1	[78]
Zoom H2n Handy Recorder	1	[35]
Not informed	33	N/A

3.2.1. Data acquisition and preprocessing

Nowadays, some of the main advantages of using ARUs for studying landscape patterns and changes are that their prices have decreased through the years [93] and new developments are frequently launched. In addition, the design of these devices is suitable for landscape monitoring, considering their resistance to environmental conditions, battery autonomy, memory capacity, size, and weight [27,106]. These facts are crucial because, in some investigations, ecologists and biologists have to locate many recorders in places with limited or difficult access; therefore, they require easy transport devices. Moreover, the battery autonomy reduces the frequency of equipment inspections, resulting in better soundscape analysis with flexible spatial and temporal resolution. We identified numerous recording devices, 25 documented; however, 33 publications did not specify the recorders employed. Table 1 lists all the reported devices and their references.

As shown in Table 1, SM hardware was the most frequently reported, with a total of 31 publications. Five versions of SM hardware (SM1, SM2, SM3, SM4, and SMBAT) have been employed for landscape data recording. Versions 1, 2, 3, and 4 are used for audible range applications, as are most other devices. For example, [81], [22], and [53] used SM2 recorders to detect bird activity, propose a bioacoustic occupancy model, and estimate animal acoustic diversity in tropical environments, respectively. [61] employed SM3 recorders to detect frog chorusing with acoustic indices. [55] used SoundTrap ST300 to identify spatio-temporal patterns of boat noise. [42] assessed tropical biodiversity using Olympus LS-11/LS-12. [78] analyzed vocal dialects for bird reintroductions using Tascam DR-05. Similarly, [35] used a Zoom H2n Handy Recorder for long-term acoustic monitoring, among other similar tasks. The frequency range varies between 16 kHz and 48 kHz mainly. Other studies report sampling rates of 2 kHz [102,107] and 8 kHz [57]. Some studies that subsample audios below 10 kHz suggest that low frequencies are related to marine ecosystems (fish and whales) [44,80] and anurans [58].

We also identified some ultrasonic applications. Recorders such as SMBAT are employed explicitly for high frequencies. For example, [86] used SM2 and SM2BAT to investigate urban acoustic diversity and disturbance by comparing low frequencies (SM2) and high frequencies (SM2BAT, set to record at 96 kHz). [28] employed SM2BAT, set to record at 384 kHz, for acoustic bat surveys. [104] used three different frequencies (50 kHz, 64 kHz, and 96 kHz) to detect dolphin whistles. Lastly, [56] described the sounds of a coral reef soundscape at a sampling rate of 96 kHz.

SM devices are widely used in these studies because they were specially developed for ecoacoustics applications. At the same time, recorders such as those manufactured by Sony and Tascam were designed for other tasks. In addition, SM1 was launched in 2007, compared to current devices such as Swift [96] and AudioMoth [108,96,98,97] released in 2017.

Once data has been gathered using ARUs, these audios need to be preprocessed and characterized to be used as inputs to further classification/clustering or source separation algorithms. As ecoacoustics recordings are collected in noisy environments, preprocessing is critical to remove artifacts or discard data that cannot be processed due to being corrupted or containing excessive noise. Rain, for example, is a geophony commonly reported as noise because it can mask vocalizations [66] and interfere with species detection, source separation, classification, and other tasks. To remove noisy recordings, we found that authors use automatic classification methods to identify saturation in the Power Spectrum Density or changes in acoustic indices of audios, such as in [88], where a threshold-based approach using Multi-Layer Perceptron with Acoustic Indices and MFCCs for rain detection is presented.

Preprocessing is also utilized to resample and filter audio recordings. In [46], authors proposed subsampling of soundscape data to reduce the computational and memory demands in the following processing steps. Data is passed to a band-pass filter, reducing the influence of wind noise and other low-frequency interferences. On the other hand, in [25], data is resampled to 8 kHz, and the DC component, which represents the constant or average level of the signal, was removed from the audios. Finally, z-Score normalization was conducted as a statistical measure to quantify the distance between each data point and the mean of the dataset in terms of standard deviation.

3.2.2. Characterization

Feature engineering, which relates to characterizing raw data, is an essential machine learning task. The success of the classification or clustering stage depends mainly on how good the features are at differentiating between classes or grouping similar samples.

In the case of time series such as audios, some mathematical representations could help researchers understand various ecological dynamics [12] and indicators that are temporally and/or spatially accentuated [42]. Therefore, an adequate representation is crucial for successful information extraction.

Spectral representations of signals are widely used in ecoacoustics analyses aided by machine learning techniques. Most reviewed publications used frequency and time-frequency-based features extracted from several implementations of the Fourier Transform as the Discrete Fourier Transform (DFT) [54], the Fast Fourier Transform (FFT) [44,35,23], and the Short-Time Fourier Transform (STFT) [70,106,34]. The parameterization of the Fourier transform was described in almost all publications where it was used. The primary reported parameters were window type, window length, number of frequency bins, and overlap. For overlap, there is no consensus among studies: some of the most common values are 25% [48,55,109,34]; 50% [85,65]; and 75% [40,112]. Regarding the windows, three types were reported: Hamming [25,86,46]; Hanning [40,64], also called Hann [93,94,75]; and Gaussian (reported in only one study using Fourier transform) [112].

The STFT is commonly used to calculate spectrograms and visually interpret the data. In this case, the image resolution is determined by the number of frames in time and frequency bins. In the studies analyzed, the number of frequency bins was 128 [31,75]; 256 [105,61,9]; and 2048 [35,85]. Likewise, the window lengths were 160, 256, 360, 400, 480, 512, 720, 1024, and 2000. With these parameters, the number and length of frames are returned. The most common frame lengths were 20 ms [103,93,65] and 10 ms [71,60,86], followed by 30 ms [113] and 5 ms [81]. However, the number of points depends on the duration of the audio, although one-minute or shorter audios can be used as a baseline.

Other types of time-frequency-based features used are Mel scales [103,35,93]; Mel Frequency Cepstral Coefficients (MFCCs) [51,65,84]; and Linear Frequency Cepstral Coefficients (LFCCs) [65,47]. Mel scales are preferred over linear spectrograms when implementing deep learning methods to avoid the spatially invariant nature of Convolutional Neural Networks (CNNs) [71]. MFCCs are classical features used for vocalizations in numerous speech recognition applications. They provide a better representation of vocalizations because the audio spectrum is more similar to the auditory systems [51], considering that human beings perceive frequencies on a logarithmic scale. LFCCs have been used together with MFCCs [65,33], acoustic indices [109], and temporal features [47]. Besides the traditional Fourier transform methods, [64] employs wavelet transforms.

Despite some articles suggesting that Mel-spectrograms are preferred for use with deep learning architectures in ecoacoustic studies, there is limited information available regarding the differences in classification and feature selection when using different representations and parameterizations. Therefore, it is crucial to explore and investigate the selection of data domains and parameters in these studies. By conducting in-depth analyses, researchers can gain a better understanding of how different representations, such as frequency-domain, time-frequency spectrograms, or other specialized feature representations, impact the performance and interpretability of machine learning models.

In other studies, authors extracted features from the temporal domain of the signals. For example, [57] computed exponentially-weighted features based on energy and a zero-crossing rate to identify syllables. [62] also proposed a temporal method for syllable segmentation, but using 1D-CNN. [79] employed comprehensive temporal acoustic Low-Level Descriptors (LLDs) to segment anuran calls. [100] integrated bioacoustic signals, DNA barcoding, and niche modeling for anuran conservation. [69,45] introduced the use of musicology to contribute and explore new temporal (and even spectral) features by assessing soundscape data. Furthermore, [33,104,28,80] combined temporal and spectral features. These approaches compute spectral and temporal features separately and then concatenate them to be fed into a classifier, clustering algorithm, or statistical framework. Other tasks and machine learning methods are described in section 3.2. We show some common representations for audio data in Fig. 5.

Once the characterization was done, some authors used feature extraction/selection methods to improve the features used in soundscape ecology. For instance, traditional methods, such as PCA and Fisher discriminant ratio, were used to select appropriate features to identify anuran calling behavior within species [65]. [74] used dissimilarity measures between bags (regions of interest in spectrograms) to improve the classification of 10-second birdsongs. [46] employed morphological operators on spectrograms. [33] proposed LLDs based on spectral and temporal data as validated features on a dataset of 25 species of the Furnariidae birds family. Likewise, [80] highlighted the usefulness of the Power Spectral Density (PSD), cepstral analysis, spectral peaks, pulse repetitions, and other time–frequency descriptors.

We also identified recent and novel characterization techniques. For example, [70] employed directional embedding for bird vocalizations; [109] used a fusion of frequency and time domain features for classifying anurans; [27] extracted features implementing a gray level co-occurrence matrix; and [92] combined features with the Recurrence Quantification Analysis (RQA) for urban

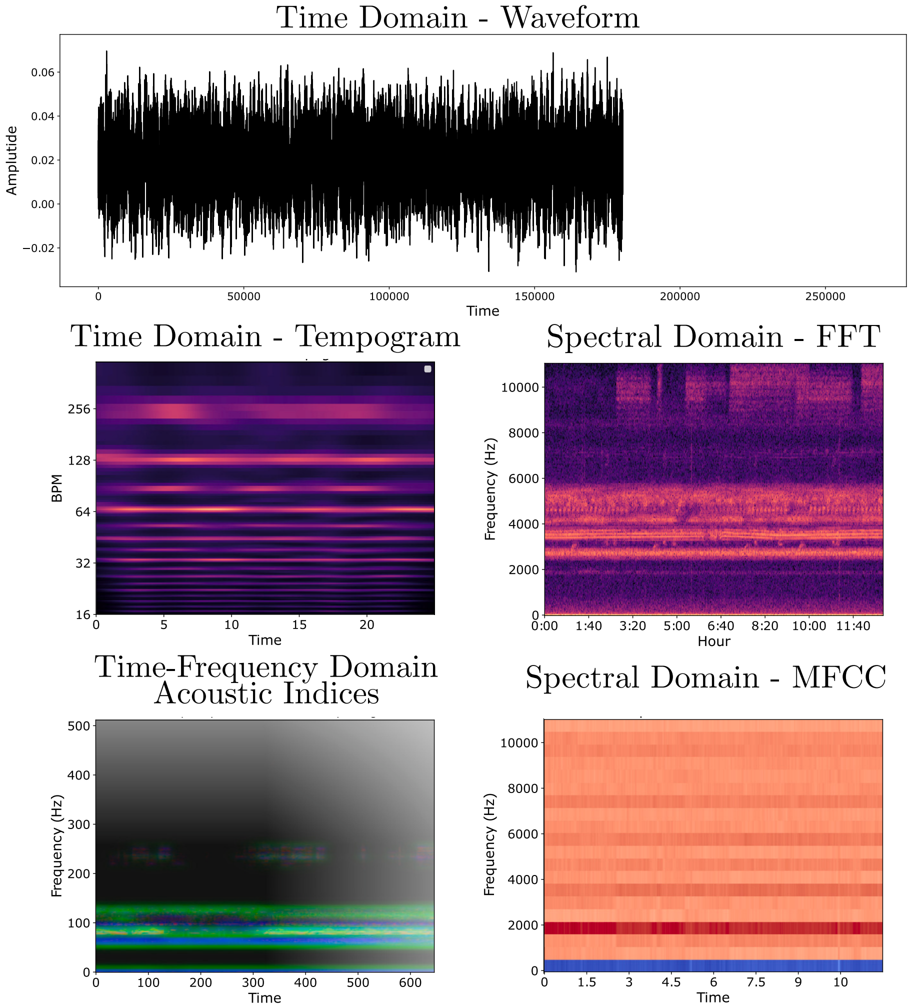

Fig. 5. Audio data representations using different domains.

sound identification. However, some authors have suggested that unsupervised learning frameworks are more suitable for soundscape analysis. In this sense, deep learning architectures are employed to extract features automatically [93,71]. These architectures can be trained for specific tasks, and the trained model can be then reused for related tasks thanks to transfer learning [48,113]. Nevertheless, deep learning represents a challenge due to the high volume of data needed to train models [93]. Moreover, supervised and semi-supervised models require labels for each sample or group of samples [70,89].

3.2.3. Soundtypes and sources identification

After data preprocessing and characterization, features are usually taken to the identification stage, where authors are mainly interested in the presence/absence of species, as mentioned in 3.1.1. Therefore, classification is commonly performed at the end of the machine learning approach, where algorithms such as Support Vector Machines (SVMs) [59,106,34,91]; K-Nearest Neighbors (KNNs) [69,66,65,34]; Random Forest [77,63,33]; Decision Trees [43,9]; and Gaussian Mixture Models (GMMs) [106,25,47,46] are used to identify species vocalizations. Authors typically compare the performance of multiple algorithms. For example, [85] compared KNNs and deep networks (VGG, ResNet, AlexNet, DenseNet) to identify European woodpeckers. Similarly, [94] compared VGG16 and ResNet50 to identify multiple species. Although classification and clustering are widely used in machine learning frameworks, audio segmentation is also studied, especially to replace manual labeling. In this regard, [67] annotated faunal acoustic events employing a "naïve" decision support tool; [81] detected bird acoustic activity using GMMs; and [104] based their investigation on dolphin whistles and clicks.

We present several machine learning methods discussed in our review, along with the corresponding datasets used in the experiments, and the reported metrics and scores, as summarized in Table 2. Methods associated with unsupervised or semi-supervised work are highlighted in bold. To ensure a focused analysis, we selected articles that specifically employed classification, clustering, and segmentation approaches for soundtypes identification. We excluded papers on denoising, source separation and focused on features. Papers addressing denoising, source separation, and data featuring were excluded. We also excluded papers that did not

Table 2
Review of the performance of machine learning methods on various reported datasets.

ML Methods	Dataset	Score Reported	Reference
1D CNN, 2D CNN, 1D-2D CNN	Two datasets of australian and brazilian frogs. 136171 total samples.	F1-score = 0.89 and 0.93	[60]
1D CNN-LSTM	Recordings including one frog species with the duration ranging from eight to fifty-five seconds.	F1-score = 0.88	[62]
Autoencoders, **tSNE**	320,000 1-s audio of 20 days continuous audio recording.	bcubed = 0.03 purity = 0.03	[31]
BatNET	817 audio-files from 678 individuals of 35 bat species.	acc = 0.91	[82]
CNN, VGGISH	Two 24-h frog recordings. 6 Sec. Segments.	AUC average = 0.98	[58]
CNNs	97900 1-min soundscape recordings of multiple species.	prec = 0.90 recall = 0.71	[93]
CNNs	53,292 unique clips of vocalizations from the 14 target species.	accuracies greater than 0.8 for several species	[72]
CNNs	Presence/absence of gunshots. More than 2 Million files, 749 of gunshots.	prec = 0.85 recall = 0.95	[97]
CNNs, Transfer learning	Eight different sites in mountains. Recordings of 5 min. Duration.	prec = 0.90	[105]
Continuous Region Analysis, Histogram of Orient Gradients	Two datasets of right whale up-calls. Kaggle dataset with more than 200000 samples.	AUC = 0.96	[107]
Dictionaries, **CNNs**	MLSP 2013 database. 645 field-recordings each containing vocalizations of 0–5 different species.	acc = 0.84	[76]
GLMMs, **tSNE**	300 recordings of 30 habitats. 40 sec. length of each audio.	p-value up to 2.95 E−41. Significative differences among analyzed sites	[12]
GMM	SM2 recorded in a frequency of 5-min every 10 min for 5 days and stopping 5 days.	F1-Score = 0.92	[87]
HMMS	521 audio recordings of species of whales.	acc from 0.82 (9 classes) to 1.00 (1–3 classes)	[51]
KNN, AlexNet, VGG, Inception ResNet, DenseNet	Xeno-Canto, Tierstimmen and a private collection. (Multi-Species).	acc 0.9	[85]
MLP	67,000 songs and calls from 7147 birds. Treehugger dataset.	acc = 0.81	[112]
MLP, KNN, DT	Samford Ecological Research Facility (SERF) dataset.	acc = 0.85	[43]
One-class SVM, NN	1 min every 20 mins and 1 min every 10 mins recordings between 2012 and 2016.	acc = 0.92	[8]
PCA, **RDA**	262 noise-clean recordings of seascapes.	91.7% (0.8, mean SE) success rate at correctly classifying individual recordings	[44]
Reservoir Network, Transfer learning, Decision Trees, HMM, RF	GTZAN Corpus, Xeno Canto.	acc = 0.925	[113]
RF	A total of 1,088,940 sound files. 650,000 bat passes identified.	acc = 0.85	[28]
RF, GLMs	Whale calls collected between July 2014 and January 2017.	about 80% of acoustic detectability	[95]
SVM, SMO	469 measurement points of urban soundscapes.	AUC = 0.913	[91]
Transfer learning, **CNNs**, **pseudo-labeling**	700 sampling sites from 2016 to 2019 1-minute recording every 10 min.	AUC = 0.99	[94]

report the used dataset, the employed metrics, or the achieved score. These works showed that deep learning techniques are a trend, especially using Convolutional Neural Networks (CNNs).

Moreover, another approach of deep learning as transfer learning facilitates the training process by re-utilizing existing neural networks [113,105,94], and more recently, the use of autoencoders has been shown to be useful in extracting relevant information from acoustic data. Deep learning scores are commonly over 0.90 [82,85], and up to 0.99 [58,94] in accuracy. On the other hand, traditional machine learning methods were applied in publications with major intuition about data characterization [91,8,28], giving more information about how data is distributed and relating features to ecological, biological, or acoustic patterns. In this way, deep learning generally overpasses traditional ML scores such as RF, SVMs, Decision Trees (DT), and HMMs that are widely explored for ecoacoustic data. However, these traditional methods are used in conjunction with prior information on landscape and specific features, such as acoustic indices, which could relate some acoustic patterns to environmental variables.

Also, we identify that unsupervised learning methods are commonly employed in ecoacoustic studies to analyze a wide range of soundtypes and capture interactions between different places or environments [31]. Unsupervised learning techniques, such as clustering and dimensionality reduction algorithms [12,31,44], allow researchers to explore and discover patterns within acoustic data without the need for labeled training examples. These methods are particularly useful for identifying unique acoustic signatures, understanding the acoustic characteristics of different habitats, and uncovering hidden relationships between sound sources and environmental variables.

On the other hand, supervised learning methods are often focused on species identification tasks in ecoacoustics. With labeled training data, classifiers and deep learning models can be trained to recognize specific species based on their acoustic features. This approach enables researchers to build models that can accurately classify and distinguish different species' vocalizations, contributing to biodiversity monitoring and conservation efforts.

Datasets in Table 2 show researchers commonly take their own data using different criteria as shown in section 3.2.1 from scratch in different regions. It is crucial to emphasize the importance of fostering collaboration networks and knowledge-sharing in this field. Currently, there appears to be limited cohesion among researchers and research groups from different countries, despite the availability of valuable data. To address this issue, it is imperative to promote the generation and expansion of cooperation networks, facilitating the sharing of data and knowledge within the research community. By collaborating and pooling resources, researchers can leverage the wealth of data generated from different regions, enabling comprehensive analyses and further advancements in the field of ecoacoustics.

In this regard, there are several popular datasets reported in the literature that have proven to be valuable resources for ecoacoustic studies. Some notable examples include the GTZAN Corpus [113], Xeno Canto [85,113], and SERF [107]. Additionally, data from worldwide challenges, such as the Challenge on Detection and Classification of Acoustic Scenes and Events (DCASE), have played a significant role in advancing acoustic research [103,70,48]. These challenges cover a wide range of topics, including acoustic classification, acoustic scene detection, audio captioning, sound synthesis, and bioacoustics. For instance, DCASE 2023 introduced the Few-shot Bioacoustic Event Detection task, further enriching the available datasets and fostering exploration in this area.

Although classification, clustering, and segmentation are widely studied and fundamental approaches, source separation is also important considering that soundscapes are taken in environments with an extreme variation of acoustic sources. However, we did not find many studies devoted to this matter. Geophonies, anthropophonies, and biophonies are primary sound sources boarded; nevertheless, sound types need to be further studied individually. Each sound type may represent an information source, but identifying multiple sources in thousands of audio files is an exhaustive task. In addition, recordings are usually made in noisy acoustic environments [83]. Consequently, machine learning techniques help optimize time consumption and accurate identification.

We found four proposals on source identification. [83] compared Fast Fixed-Point Independent Component Analysis (FastICA), Principal Component Analysis (PCA), and Non-Negative Matrix Factorization (NMF) as Blind Source Separation (BSS) methods. These methods were evaluated using Signal to Distortion Ratio (SDR), Signal to Interference Ratio (SIR), and Signal to Artifacts Ratio (SAR) applied to a dataset of anuran vocalizations. [89] also used NMF to estimate road traffic noise levels. Similarly, [2] investigated the best release areas for bird populations in Brazil using PCA analysis. Finally, [49] proposed the implementation of Singular Spectrum Analysis (SSA) to decompose acoustic signals into oscillatory components and identify anuran calls using amplitude, entropy, spectral entropy, and permutation entropy as quantifiers. These studies show an open challenge in the search for a methodology to separate bioacoustic sources correctly. Other studies have focused on separating vocalizations and other types of sound in seascapes. Particularly, [71] classified vocalizations of cetaceans and fish, natural sounds, and unidentified ocean noise recordings. In this case, authors implemented two classifiers based on deep learning architectures, CNNs, and Recurrent Neural Networks (RNNs), achieving 96.1% accuracy.

Fig. 6 shows a flow diagram from data acquisition to the identification of sound types. It shows some variations or methods typically implemented for soundscape analysis.

3.2.4. Software tools

The development of new methodologies and the increasing use of audio recordings to explore ecological patterns have led to a synergy between biologists and professionals from fields such as data science, statistics, and electronics. Consequently, more and more software and computational tools are becoming available for the analysis of soundscape data. However, our review found that more than 57% of the studies used three main programs. Python was the software of choice in 21.6% of the papers, being the preferred software tool in machine learning applications. Python modules commonly used were librosa for audio processing and PyTorch to develop deep learning architectures. R is the second most used software with 19.6%. Some modules specially used in soundscape ecology are Seewave, Soundecology, MuMIn, and MASS. MATLAB was employed in 18.6% of the studies, being the

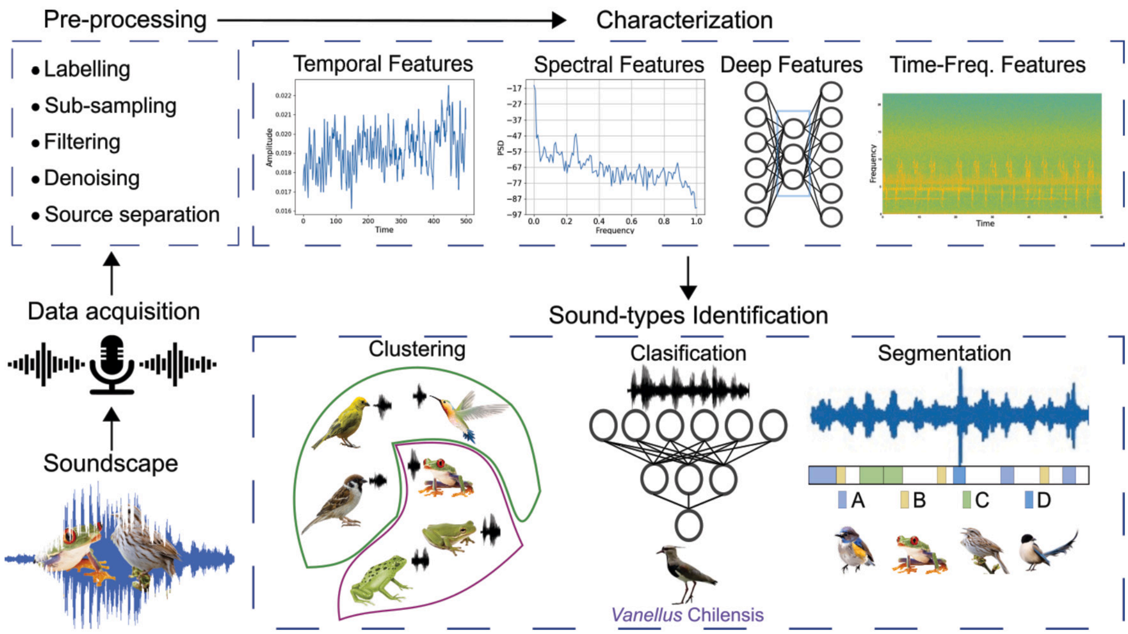

Fig. 6. Soundscape ecology data processing using machine learning frameworks.

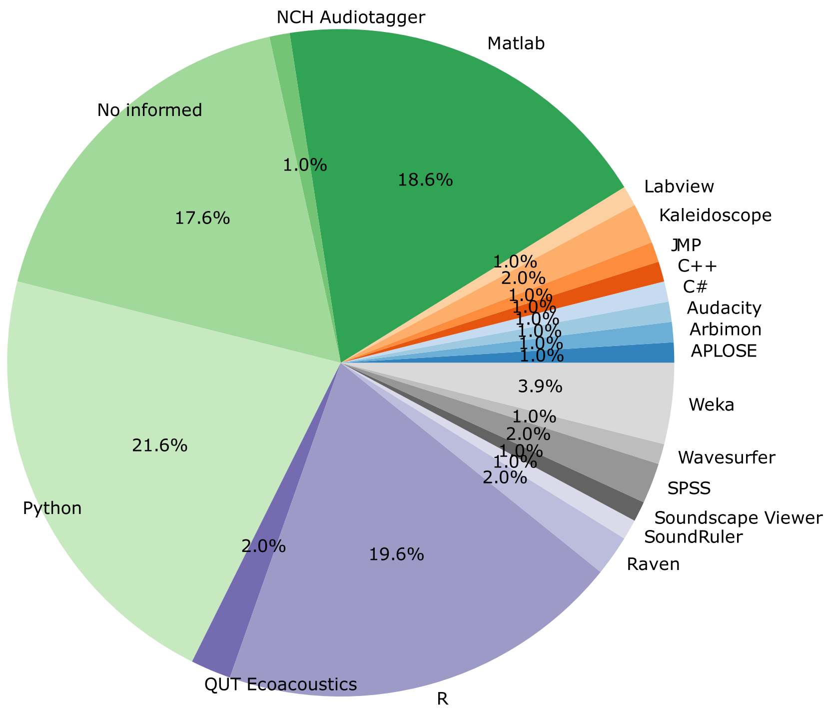

Fig. 7. Software tools used for ecoacoustics and bioacoustics data processing.

third most used software. Fig. 7 shows the software tools employed in the publications. We identified a total of 19 computational applications. In addition, we included an item corresponding to those papers that did not report any program or tool (17.6%).

3.3. Trends and perspectives of soundscape ecology using machine learning

Compiling information from the last year (2022), we identified that publications have diversified in terms of the sounds analyzed. Some studies focus on anthropogenic sources, others on biophonies, and some have expanded the scale to study communities. Even within biophony studies, there is an increase in the study of marine species such as dolphins and whales. We classified publications

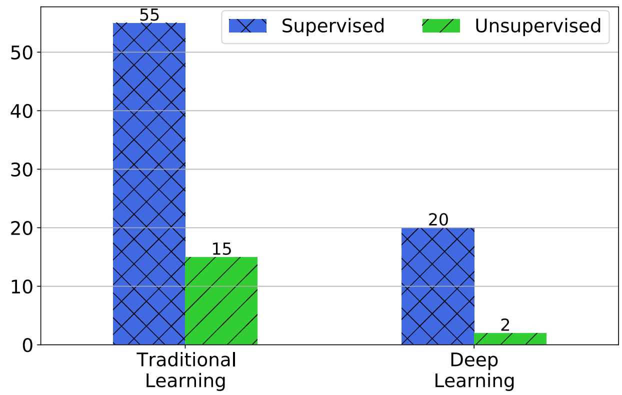

Fig. 8. Number of publications that employ traditional learning vs deep learning methods. Each approach was also divided into supervised (blue) and unsupervised (green) learning.

of the last year according to the acoustic source; we obtained that three works focused on anthropophonies, one terrestrial (9.09%) and two marine studies (18.18%), and eight on biophonies, five terrestrial (45.45%) and three marine studies (27.27%).

Authors have exposed the need to develop robust methodologies to assess high volume and noisy recordings due to the nature of the collected data [97]. Promissory studies such as in [98] establish opportunities using soundscape methodologies to monitor landscape-level changes in biodiversity and the complementarity of species coverage to other monitoring methods. This approach of integrating several kinds of sensors has been exposed as an essential way to complement the information of soundscape [98,111,22,11]. Nevertheless, the first step to integrating sensors is to promote the use of ARUs and expand the spatial range. In this way, there are initiatives to deploy numerous recorders and analyze spatial patterns, e.g., in [94], 700 sites of Puerto Rico were sampled, in [114] four datasets were referenced, of which three of them contain information collected with multiple ARUs, and [108] combines personal devices from citizen scientists and Audiomoths, showing potential cooperation by people feeding datasets using common devices such as a smartphone.

In Fig. 8, we show the number of publications that use traditional learning compared to deep learning. We also compare studies based on supervised and unsupervised methods. It is notorious that traditional and supervised learning are widely used for several tasks in ecoacoustics data processing and analysis. Even when we observe the publication date, it is notorious that these methods are current, and there are still improvements to be made to outperform species identification, vocalizations segmentation in noisy environments, and increase the accuracy for sound types source separation distinguishing between human, biotic and abiotic sources, among other tasks. Nevertheless, we have observed in works published in the last two years a preference for unsupervised methods. These methods have been used for characterization, identification, and segmentation tasks. However, deep learning architectures have also been recently presented as a new trend to explore ecoacoustic information [31,105,82,62], which indicates a need to process high volumes of data at all stages automatically; therefore, end-to-end models might be more appropriate.

On the other hand, recent advancements in machine learning methodologies for audio analysis and classification have been greatly influenced by transformative techniques developed by leading companies such as Meta, Google, and DeepMind. Of particular significance are transformer-based models, which have revolutionized various domains, including natural language processing. These models have been successfully adapted to audio tasks, allowing for the capture of long-range dependencies and complex relationships within audio data. Furthermore, encoder-decoder architectures and transfer learning have played a crucial role in leveraging pre-trained models on extensive audio datasets, leading to remarkable improvements in audio feature extraction and classification accuracy.

These transformative technologies are primarily based on cloud computing infrastructure. Leveraging the power of cloud computing, researchers and practitioners can efficiently explore and analyze ecoacoustic data. By utilizing cloud-based platforms, such as those provided by Meta, Google, and DeepMind, the scalability and computational resources required for processing large-scale ecoacoustic datasets are readily available. These technologies enable the extraction of valuable insights from ecoacoustic data, facilitating research and advancements in areas such as soundscape analysis, species identification, and environmental monitoring. The cloud-based nature of these technologies ensures accessibility, scalability, and the ability to collaborate on ecoacoustic research projects, fostering innovation and driving the understanding of acoustic ecosystems to new heights.

4. Discussion

This study systematically reviewed and summarized the methods and machine-learning-based approaches for ecoacoustics and soundscape monitoring. With this review, we aimed to: i) find the ecological components and issues from soundscapes that have been most studied using machine learning, ii) identify the most used machine learning algorithms used for such tasks, and iii) identify new possible applications of machine learning that allows responding biological questions from soundscapes.

Regarding the first question, current soundscape ecology schemes using machine learning usually focus on species detection, which facilitates the study of species richness. Moreover, some studies approach the analysis of sound types and the link between the soundscape and the landscape heterogeneity. These studies are essential because, including a proper spatio-temporal scale, they will allow scientists to use PAM to track changes in ecosystems at risk of disappearance and recommend to the competent authorities conservation or restoration plans. Moreover, it would be interesting to include in such studies an analysis of the entire spectrum of sounds, including biophonies, geophonies, and anthropophonies, as they are complementary to understanding the soundscape.

In response to the second question, we found that researchers have proposed several methodologies using machine learning algorithms to analyze soundscape and ecoacoustic data. Several complete pipelines of supervised learning comprised of audio pre-processing, characterization, and classification or source separation have been proposed for soundscape studies. In the first stage, audio pre-processing aims to discard recordings that will damage the training process of the machine learning models. This stage is commonly carried out based on energy-threshold methods. However, some authors have also trained models to classify proper and noisy recordings. For the characterization stage, the most used approach is to analyze spectral and temporal features, such as spectrograms and Mel coefficients. Once the audio data has been pre-processed and characterized, source separation techniques such as PCA and Matrix factorization are commonly employed to identify the sound types that compose the acoustic space, extract the information of interest, and discard noising or irrelevant components. Finally, classification allows for identifying several sound types. Some traditional machine learning methods reported to segment audio vocalizations, classify sound types, and detect species from the obtained features are Support vector machines, K-nearest Neighbor, Random Forest, Gaussian Mixture Models, and Hidden Markov Models. Moreover, recent reports have shown a trend to include deep learning models improving detection results and avoiding featuring extraction. Even new architectures such as Birdent have been explicitly developed for ecoacoustic data.

Regarding the kind of machine learning algorithms used, the literature highlights the need to use unsupervised methods to analyze the soundscape. Unsupervised methods would allow us to explore the different types of sound in a recording without requiring labeled recordings and reduce the time the experts will have to invest in providing biological insights from large amounts of data. Regarding the trend in using machine learning for analyzing soundscape recordings, most published papers focus on improving a specific pipeline stage. As a result, coming studies should focus on responding to new biological questions that link the information the machine learning algorithms provided with conservation strategies.

Finally, related to the third question, we found that some studies are willing to elucidate: spatial patterns of the soundscape by placing recording devices in several locations simultaneously, temporal patterns by placing recordings placed during extended periods (months, years), or spatio-temporal patterns mixing both recording strategies. These kinds of studies will allow, in the upcoming years, to analyze from ecoacoustics other dimensions of biological complexity such as population dynamics, acoustic dynamics related to landscape cover, and landscape connectivity.

5. Conclusions

This paper aims to present a systematic review of the principal methodologies and applications of machine learning to soundscape ecology. To this aim, we searched for the most related and cited papers in these thematics over the last fourteen years. From the review, we identified that the most frequently addressed tasks are: analyzing species' presence (absence), call segmentation, and grouping sound sources. Based on such patterns, which are mainly related to the acoustic composition of the ecosystem, researchers have calculated ecological indicators as measures of biodiversity. Nevertheless, such tasks are highly biased toward species-specific analysis and do not take advantage of the several sonotypes available on the recordings.

Most reviewed studies use supervised learning as a tool to answer the above-mentioned ecology-related questions. However, using such algorithms implies the availability of a considerable amount of labeled data, which demands time, resources, and staff training for using software to label data. As a result, in recent years, unsupervised learning, mainly based on neural networks, has emerged as an alternative to avoid labeling data. However, these efforts have been focused on answering the same addressed tasks extending some of them to several species (as there are no labels).

It is well known that beyond species or multispecies questions, other biological complexity dimensions have not been addressed from machine learning and ecoacoustics perspectives such as connectivity and historical contingency. For this reason, in future work, we plan to use unsupervised learning to provide insights that allow biologists to analyze such landscape patterns from audio recordings.

CRediT authorship contribution statement

D.A Nieto-Mora: Conceived and designed the experiments, performed the experiments, analyzed and interpreted the data, wrote the paper.

Susana Rodríguez-Buritica: Conceived and designed the experiments, analyzed and interpreted the data, wrote the paper.

Paula Rodríguez-Marín: Contributed reagents, materials, analysis tools or data, wrote the paper.

J.D Martínez-Vargaz: Conceived and designed the experiments, analyzed and interpreted the data, wrote the paper.

Claudia Isaza-Narváez: Conceived and designed the experiments, analyzed and interpreted the data, Contributed reagents, materials, analysis tools or data, wrote the paper.

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.

Data availability

The data associated with this study, comprising the reviewed articles and the analyzed dataset, is available upon request. We are dedicated to fostering transparency and facilitating access to our research findings. Interested parties are encouraged to contact us for access to the relevant data to further support the findings presented in this research.

Acknowledgements

This work was supported by Universidad de Antioquia, Instituto Tecnológico Metropolitano de Medellín, Alexander von Humboldt Institute for Research on Biological Resources and Colombian National Fund for Science, Technology and Innovation, Francisco Jose de Caldas - MINCIENCIAS (Colombia). [Grant No. 111585269779].

Appendix A

See Table 3.

Table 3
Table of reported species divided into Genus/Family categories.

Genus/Family	Count	Reference	Category
Coccyzus, Coereba, Loxigilla, Megascops, Melanerpes, Nesospingus, Setophaga, Spindalis, Todus	1	[93]	Birds
Margarops, Patagioenas	2	[93], [94]	Birds
Turdus	5	[93], [113], [112], [35], [94]	Birds
Vireo	2	[93], [70]	Birds
Burhinus, Chrysodeixis, Chenonetta, Coracina, Cracticus, Eopsaltria, Geopelia, Lichenostomus, Melithreptus, Myiagra, Pardalotus, Philemon, Rhipidura, Sericornis, Todiramphus, Trichoglossus, Zosterops	1	[31]	Birds
Corvus	3	[31], [67], [35]	Birds
Meliphaga, Myzomela, Oriolus	2	[31], [67]	Birds
Vanellus	3	[31], [47], [81]	Birds
Phylloscopus	2	[105], [112]	Birds
Spelaeornis	1	[105]	Birds
Lagopus	1	[90]	Birds
Furnariidae	1	[33]	Birds
Acrocephalus, Calidris, Carduelis, Emberiza, Falco, Lanius, Larus, Parus	1	[113]	Birds
Sylvia	2	[113], [22]	Birds
Caprimulgus, Lullula	1	[22]	Birds
Dendrocopos, Dryocopus, Picoides, Picus	1	[85]	Birds
Dryobates	2	[85], [35]	Birds
Colluricincla, Pheucticus, Piranga, Pyrrholaemus, Toxostoma	1	[70]	Birds
Catharus, Certhia, Contopus, Empidonax, Ixoreus, Junco	1	[74]	Birds
Poecile	2	[74], [35]	Birds
Sitta	1	[74], [112], [35]	Birds
Troglodytes	2	[74], [86]	Birds
Actitis, Hippolais, Locustella, Numenius, Upupa	1	[112]	Birds
Hippolais	1	[112]	Birds
Locustella	1	[112]	Birds
Numenius	1	[112]	Birds
Upupa	1	[112]	Birds

Table 3 (*continued*)

Genus/Family	Count	Reference	Category
Ixobrychus	1	[23]	Birds
Eupsittula	*1*	[78]	Birds
Accipiter, Alcedo, Amaurornis, Calyptorhynchus, Todiramphus, Pachycephala	1	[67]	Birds
Cardinalis, Cyanocitta, Melospiza, Poecile, Spinus	1	[35]	Birds
Litoria	7	[59], [34], [65], [63], [62], [61], [60]	Anurans
Assa	3	[34], [62], [60]	Anurans
Crinia	5	[34], [63], [62], [61], [60]	Anurans
Limnodynastes	4	[34], [63], [62], [60]	Anurans
Mixophyes, Neobatrachus	2	[34], [62]	Anurans
Philoria, Pseudophryne	1	[34]	Anurans
Uperoleia	4	[34], [65], [63], [62]	Anurans
Leptodactylus	3	[57], [88], [93]	Anurans
Ameerega, Dendropsophus, Osteocephalus	2	[57], [79]	Anurans
Adenomera	3	[57], [79], [49]	Anurans
Rhinella	4	[57], [63], [88], [79]	Anurans
Cyclorana, Platyplectrum	1	[63]	Anurans
Aromobatidae, Centrolenidae, Craugastoridae, Dendrobatidae, Eleutherodactylidae, Hemiphractidae, Hylidae, Leptodactylidae, Microhylidae	1	[25]	Anurans
Colostethus, Dendrobates, Diasporus, Engystomops, Leucostethus, Pristimantis	1	[88]	Anurans
Scinax	2	[88], [79]	Anurans
Hyla	3	[88], [79], [49]	Anurans
Eleutherodactylus	2	[88], [93]	Anurans
Pseudophryne, Rheobatrachus	1	[62]	Anurans
Brachycephalus	1	[79]	Anurans
Aplastodiscus	2	[79], [49]	Anurans
Taudactylus	1	[58]	Anurans
Atelopus, Uperodon	1	[100]	Anurans
Duttaphrynus, Fejervarya, Hylarana, Kaloula, Microhyla, Odorrana, Philautus, Phrynoidis, Polypedates	1	[84]	Anurans
Bufonidae, Dendrobatidae, Hemiphractidae, Hylidae, Hyperoliidae, Leptodactylidae, Mantellidae, Microhylidae, Myobatrachidae, Ranidae, Rhacophorus, Scaphiopodidae	1	[109]	Anurans
Sotalia guianensis	1	[77]	Dolphines
Sousa chinensis, *Stenella* longirostris	1	[104]	Dolphines
Balaenoptera	1	[68], [51], [71]	Whales
Eubalaena	1	[107], [51]	Whales
Argyrosomus regius	1	[44]	Fish
Halobatrachus didactylus	1	[44]	Fish
Aselliscus, Hipposideros, Hypsugo, Kerivoula, Miniopterus, Murina, Phoniscus, Rhinolophus, Scotomanes, Tylonycteris	1	[82]	Bats
Myotis	2	[82], [28]	Bats

(*continued on next page*)

Table 3 (*continued*)

Genus/Family	Count	Reference	Category
Centronycteris, Cormura, Emballonuridae, *Furipterus*, Molossidae, *Peropteryx, Promops, Pteronotus, Rhynchonycteris, Saccopteryx,* Vespertilionidae	1	[28]	Bats
Pan troglodytes	1	[103]	Other Mammals
Cicada	2	[66,9]	Invertebrates
Balaena	2	[71], [51]	Whales
Delphinapterus, Globicephala, Megaptera, Monodon, Orcinus, Pseudorca	1	[51]	Whales
Eschrichtius, Delphinus, Pseudorca	1	[71]	Whales
Erignathus, Histriophoca, Phoca, Phocoenoides	1	[71]	Other Mammals
Alpheidae, Amphiprioninae, *Evechinus* chloroticus, Palinuridae, *Pogonias*, *Salvelinus*, *Sciaenops*, *Stomatopoda*	1	[71]	Marinne invertebrates
Gryllus pennsylvanicus, *Neotibicen* canicularis	1	[35]	Invertebrates
Tamias striatus, *Sciurus* carolinensis	1	[35]	Rodents
Sciurus carolinensis	1	[86]	Rodents
Pipistrellus	1	[86]	Bats
Vulpes	1	[86]	Other Mammals

References

[1] WWF, Informe Planeta Vivo 2020: Revertir la curva de la pérdida de biodiversidad. Resumen, Unknown, 2020, p. 130.
[2] H.G. Martinez, Discusiones POST 2020, Unknown, 2020.
[3] El IPBES, Informe de la Evaluacion Mundial sobre la Diversidad Biologica y los Servicios de los Ecosistemas, Unknown, 2019.
[4] UNFCCC, COP26 the Glasgow climate pact, in: COP26 Explained, 2021, p. 28.
[5] UNFCCC, COP26 explained, in: COP26 Explained, 2021, p. 25. [Online]. Available: https://ukcop26.org/wp-content/uploads/2021/07/COP26-Explained.pdf, 2021.
[6] S. Maginnis, S. Sandeep, United Nations framework convention on climate change, Environ. Health Perspect. 108 (8) (2021).
[7] J.J. Lahoz-Monfort, M.J. Magrath, A comprehensive overview of technologies for species and habitat monitoring and conservation, Bioscience 71 (10) (2021) 1038–1062.
[8] W.E. Gómez, C.V. Isaza, J.M. Daza, Identifying disturbed habitats: a new method from acoustic indices, Ecol. Inform. 45 (2018) 16–25.
[9] M. Towsey, J. Wimmer, I. Williamson, P. Roe, The use of acoustic indices to determine avian species richness in audio-recordings of the environment, Ecol. Inform. 21 (100) (2014) 110–119.
[10] H. Shamon, Z. Paraskevopoulou, J. Kitzes, E. Card, J.L. Deichmann, A.J. Boyce, W.J. McShea, Using ecoacoustics metrices to track grassland bird richness across landscape gradients, Ecol. Indic. 120 (jan 2021).
[11] D.I. Rappaport, J.A. Royle, D.C. Morton, Acoustic space occupancy: combining ecoacoustics and lidar to model biodiversity variation and detection bias across heterogeneous landscapes, Ecol. Indic. 113 (February) (2020) 106172, https://doi.org/10.1016/j.ecolind.2020.106172. [Online].
[12] Á. Vega-Hidalgo, E. Flatt, A. Whitworth, L. Symes, Acoustic assessment of experimental reforestation in a Costa Rican rainforest, Ecol. Indic. 133 (December) (2021).
[13] B.D. Suárez-Tangil, A. Rodríguez, Integral assessment of active and passive survey methods for large-scale monitoring of mammal occurrence in Mediterranean landscapes, Ecol. Indic. 125 (2021).
[14] J. Som-Ard, C. Atzberger, E. Izquierdo-Verdiguier, F. Vuolo, M. Immitzer, Remote sensing applications in sugarcane cultivation: a review, Remote Sens. 13 (20) (2021) 1–46.
[15] L.G. Howell, J. Clulow, N.R. Jordan, C.T. Beranek, S.A. Ryan, A. Roff, R.R. Witt, Drone thermal imaging technology provides a cost-effective tool for landscape-scale monitoring of a cryptic forest-dwelling species across all population densities, Wildl. Res. 49 (1) (2022) 66–78.
[16] Y.P. Lin, H.J. Chu, Y.L. Huang, C.H. Tang, S. Rouhani, Monitoring and identification of spatiotemporal landscape changes in multiple remote sensing images by using a stratified conditional Latin hypercube sampling approach and geostatistical simulation, Environ. Monit. Assess. 177 (1–4) (2011) 353–373.
[17] D. Tuia, B. Kellenberger, S. Beery, B.R. Costelloe, S. Zuffi, B. Risse, A. Mathis, M.W. Mathis, F. van Langevelde, T. Burghardt, R. Kays, H. Klinck, M. Wikelski, I.D. Couzin, G. van Horn, M.C. Crofoot, C.V. Stewart, T. Berger-Wolf, Perspectives in machine learning for wildlife conservation, Nat. Commun. 13 (1) (2022) 1–15.
[18] N. Demirel, M.K. Emil, H.S. Duzgun, Surface coal mine area monitoring using multi-temporal high-resolution satellite imagery, Int. J. Coal Geol. 86 (1) (2011) 3–11, https://doi.org/10.1016/j.coal.2010.11.010. [Online].
[19] L. Demarchi, A. Kania, W. Ciezkowski, H. Piórkowski, Z. Oświecimska-Piasko, J. Chormański, Recursive feature elimination and random forest classification of natura 2000 grasslands in lowland river valleys of Poland based on airborne hyperspectral and LiDAR data fusion, Remote Sens. 12 (11) (2020).
[20] R.T. Buxton, P.E. Lendrum, K.R. Crooks, G. Wittemyer, Pairing camera traps and acoustic recorders to monitor the ecological impact of human disturbance, Glob. Ecol. Conserv. 16 (2018) e00493, https://doi.org/10.1016/j.gecco.2018.e00493. [Online].
[21] H. Yu, X. Liu, B. Kong, R. Li, G. Wang, Landscape ecology development supported by geospatial technologies: a review, Ecol. Inform. 51 (January) (2019) 185–192, https://doi.org/10.1016/j.ecoinf.2019.03.006. [Online].
[22] C. Abrahams, M. Geary, Combining bioacoustics and occupancy modelling for improved monitoring of rare breeding bird populations, Ecol. Indic. 112 (may 2020).
[23] E. Znidersic, M. Towsey, W.K. Roy, S.E. Darling, A. Truskinger, P. Roe, D.M. Watson, Using visualization and machine learning methods to monitor low detectability species—the least bittern as a case study, Ecol. Inform. 55 (jan 2020).
[24] R.F. NOSS, Indicators for monitoring biodiversity: a hierarchical approach, Conserv. Biol. 4 (4) (1990) 355–364.
[25] J.B. Alonso, J. Cabrera, R. Shyamnani, C.M. Travieso, F. Bolaños, A. García, A. Villegas, M. Wainwright, Automatic anuran identification using noise removal and audio activity detection, Expert Syst. Appl. 72 (2017) 83–92.

[26] B. Holgate, R. Maggini, S. Fuller, Mapping ecoacoustic hot spots and moments of biodiversity to inform conservation and urban planning, Ecol. Indic. 126 (jul 2021).
[27] L.M. Huancapaza Hilasaca, L.P. Gaspar, M.C. Ribeiro, R. Minghim, Visualization and categorization of ecological acoustic events based on discriminant features, Ecol. Indic. 126 (jul 2021).
[28] A. López-Baucells, L. Torrent, R. Rocha, P.E.D. Bobrowiec, J.M. Palmeirim, C.F.J. Meyer, Stronger together: combining automated classifiers with manual post-validation optimizes the workload vs reliability trade-off of species identification in bat acoustic surveys, Ecol. Inform. 49 (August 2018) (2019) 45–53, https://doi.org/10.1016/j.ecoinf.2018.11.004. [Online].
[29] A. Digby, M. Towsey, B.D. Bell, P.D. Teal, A practical comparison of manual and autonomous methods for acoustic monitoring, Methods Ecol. Evol. 4 (7) (2013) 675–683.
[30] D. Stowell, J. Sueur, Ecoacoustics: acoustic sensing for biodiversity monitoring at scale, Remote Sens. Ecol. Conserv. 6 (3) (2020) 217–219.
[31] B. Rowe, P. Eichinski, J. Zhang, P. Roe, Acoustic auto-encoders for biodiversity assessment, Ecol. Inform. 62 (August 2020) (2021) 101237, https://doi.org/10.1016/j.ecoinf.2021.101237. [Online].
[32] A. Eldridge, P. Guyot, P. Moscoso, A. Johnston, Y. Eyre-Walker, M. Peck, Sounding out ecoacoustic metrics: avian species richness is predicted by acoustic indices in temperate but not tropical habitats, Ecol. Indic. 95 (dec 2018) 939–952.
[33] E.M. Albornoz, L.D. Vignolo, J.A. Sarquis, E. Leon, Automatic classification of Furnariidae species from the Paranaense Littoral region using speech-related features and machine learning, Ecol. Inform. 38 (mar 2017) 39–49.
[34] J. Xie, M. Towsey, J. Zhang, P. Roe, Acoustic classification of Australian frogs based on enhanced features and machine learning algorithms, Appl. Acoust. 113 (dec 2016) 193–201.
[35] M.M. Morgan, J. Braasch, Long-term deep learning-facilitated environmental acoustic monitoring in the Capital Region of New York State, Ecol. Inform. 61 (September 2020) (2021) 101242, https://doi.org/10.1016/j.ecoinf.2021.101242. [Online].
[36] S.A. Nitoslawski, K. Wong-Stevens, J.W. Steenberg, K. Witherspoon, L. Nesbitt, C.C. Konijnendijk van den Bosch, The digital forest: mapping a decade of knowledge on technological applications for forest ecosystems, Earth's Future 9 (8) (2021) 1–28.
[37] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, S.E. Brennan, R. Chou, J. Glanville, J.M. Grimshaw, A. Hróbjartsson, M.M. Lalu, T. Li, E.W. Loder, E. Mayo-Wilson, S. McDonald, L.A. McGuinness, L.A. Stewart, J. Thomas, A.C. Tricco, V.A. Welch, P. Whiting, D. Moher, The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, BMJ 372 (2021) 2020–2021.
[38] R. van de Schoot, J. de Bruin, R. Schram, P. Zahedi, J. de Boer, F. Weijdema, B. Kramer, M. Huijts, M. Hoogerwerf, G. Ferdinands, A. Harkema, J. Willemsen, Y. Ma, Q. Fang, S. Hindriks, L. Tummers, D.L. Oberski, An open source machine learning framework for efficient and transparent systematic reviews, Nat. Mach. Intell. 3 (2) (feb 2021) 125–133, https://doi.org/10.1038/s42256-020-00287-7. [Online].
[39] M.L. Cadenasso, S.T. Pickett, J.M. Grove, Dimensions of ecosystem complexity: heterogeneity, connectivity, and history, Ecol. Complex. 3 (1) (2006) 1–12.
[40] F.F. Dias, H. Pedrini, R. Minghim, Soundscape segregation based on visual analysis and discriminating features, Ecol. Inform. 61 (mar 2021).
[41] L.I. Ben, W. Zhang, Z. Wang, H. Xie, X. Yuan, E. Pei, T. Wang, Effects of landscape heterogeneity and breeding habitat diversity on rice frog abundance and body condition in agricultural landscapes of Yangtze River Delta, China, Curr. Zool. 66 (6) (2020) 615–623.
[42] S.L. Mitchell, J.E. Bicknell, D.P. Edwards, N.J. Deere, H. Bernard, Z.G. Davies, M.J. Struebig, Spatial replication and habitat context matters for assessments of tropical biodiversity using acoustic indices, Ecol. Indic. 119 (August) (2020) 106717, https://doi.org/10.1016/j.ecolind.2020.106717. [Online].
[43] L. Zhang, M. Towsey, J. Xie, J. Zhang, P. Roe, Using multi-label classification for acoustic pattern detection and assisting bird species surveys, Appl. Acoust. 110 (sep 2016) 91–98.
[44] M. Vieira, P.J. Fonseca, M.C.P. Amorim, Fish sounds and boat noise are prominent soundscape contributors in an urban European estuary, Mar. Pollut. Bull. 172 (nov 2021).
[45] K.M. Bellisario, J. VanSchaik, Z. Zhao, A. Gasc, H. Omrani, B.C. Pijanowski, Contributions of MIR to Soundscape Ecology. Part 2: spectral timbral analysis for discriminating soundscape components, Ecol. Inform. 51 (February) (2019) 1–14, https://doi.org/10.1016/j.ecoinf.2019.01.008. [Online].
[46] T.M. Ventura, A.G. De Oliveira, T.D. Ganchev, J.M. De Figueiredo, O. Jahn, M.I. Marques, K.L. Schuchmann, Audio parameterization with robust frame selection for improved bird identification, Expert Syst. Appl. 42 (22) (2015) 8463–8471, https://doi.org/10.1016/j.eswa.2015.07.002. [Online].
[47] T.D. Ganchev, O. Jahn, M.I. Marques, J.M. De Figueiredo, K.L. Schuchmann, Automated acoustic detection of Vanellus chilensis lampronotus, Expert Syst. Appl. 42 (15–16) (2015) 6098–6111, https://doi.org/10.1016/j.eswa.2015.03.036. [Online].
[48] S. Kahl, C.M. Wood, M. Eibl, H. Klinck, BirdNET: a deep learning solution for avian diversity monitoring, Ecol. Inform. 61 (mar 2021).
[49] J.G. Colonna, E.F. Nakamura, Unsupervised selection of the singular spectrum components based on information theory for bioacoustic signal filtering, Digit. Signal Process., Rev. J. 82 (nov 2018) 64–79.
[50] B. Doohan, S. Fuller, S. Parsons, E.E. Peterson, The sound of management: acoustic monitoring for agricultural industries, Ecol. Indic. 96 (July 2018) (2019) 739–746, https://doi.org/10.1016/j.ecolind.2018.09.029. [Online].
[51] M.B. Trawicki, Multispecies discrimination of whales (cetaceans) using Hidden Markov Models (HMMS), Ecol. Inform. 61 (January) (2021) 101223, https://doi.org/10.1016/j.ecoinf.2021.101223. [Online].
[52] K. Hayashi Erwinsyah, V.D. Lelyana, K. Yamamura, Acoustic dissimilarities between an oil palm plantation and surrounding forests: analysis of index time series for beta-diversity in South Sumatra, Indonesia, Ecol. Indic. 112 (February) (2020) 106086, https://doi.org/10.1016/j.ecolind.2020.106086. [Online].
[53] J.S. Ulloa, T. Aubin, D. Llusia, C. Bouveyron, J. Sueur, Estimating animal acoustic diversity in tropical environments using unsupervised multiresolution analysis, Ecol. Indic. 90 (jul 2018) 346–355.
[54] T.H. Lin, T. Akamatsu, F. Sinniger, S. Harii, Exploring coral reef biodiversity via underwater soundscapes, Biol. Conserv. 253 (jan 2021).
[55] J.P. Dinh, J.J. Suca, A. Lillis, A. Apprill, J.K. Llopiz, T.A. Mooney, Multiscale spatio-temporal patterns of boat noise on U.S. Virgin Island coral reefs, Mar. Pollut. Bull. 136 (nov 2018) 282–290.
[56] S.A. Dimoff, W.D. Halliday, M.K. Pine, K.L. Tietjen, F. Juanes, J.K. Baum, The utility of different acoustic indicators to describe biological sounds of a coral reef soundscape, Ecol. Indic. 124 (2021) 107435, https://doi.org/10.1016/j.ecolind.2021.107435. [Online].
[57] J.G. Colonna, M. Cristo, M. Salvatierra, E.F. Nakamura, An incremental technique for real-time bioacoustic signal segmentation, Expert Syst. Appl. 42 (21) (jun 2015) 7367–7374.
[58] J. Xie, M. Zhu, K. Hu, J. Zhang, H. Hines, Y. Guo, Frog calling activity detection using lightweight CNN with multi-view spectrogram: a case study on Kroombit tinker frog, Mach. Learn. Appl. 7 (October 2021) (2022) 100202, https://doi.org/10.1016/j.mlwa.2021.100202. [Online].
[59] H. Gan, J. Zhang, M. Towsey, A. Truskinger, D. Stark, B.J. van Rensburg, Y. Li, P. Roe, A novel frog chorusing recognition method with acoustic indices and machine learning, Future Gener. Comput. Syst. 125 (2021) 485–495, https://doi.org/10.1016/j.future.2021.06.019. [Online].
[60] J. Xie, K. Hu, Y. Guo, Q. Zhu, J. Yu, On loss functions and CNNs for improved bioacoustic signal classification, Ecol. Inform. 64 (July 2020) (2021) 101331, https://doi.org/10.1016/j.ecoinf.2021.101331. [Online].
[61] H. Gan, J. Zhang, M. Towsey, A. Truskinger, D. Stark, B.J. van Rensburg, Y. Li, P. Roe, Data selection in frog chorusing recognition with acoustic indices, Ecol. Inform. 60 (nov 2020).
[62] J. Xie, K. Hu, M. Zhu, Y. Guo, Bioacoustic signal classification in continuous recordings: syllable-segmentation vs sliding-window, Expert Syst. Appl. 152 (2020) 113390, https://doi.org/10.1016/j.eswa.2020.113390. [Online].
[63] S. Brodie, S. Allen-Ankins, M. Towsey, P. Roe, L. Schwarzkopf, Automated species identification of frog choruses in environmental recordings using acoustic indices, Ecol. Indic. 119 (August) (2020) 106852, https://doi.org/10.1016/j.ecolind.2020.106852. [Online].

[64] J.C. Bishop, G. Falzon, M. Trotter, P. Kwan, P.D. Meek, Livestock vocalisation classification in farm soundscapes, Comput. Electron. Agric. 162 (August 2018) (2019) 531–542, https://doi.org/10.1016/j.compag.2019.04.020. [Online].

[65] J. Xie, K. Indraswari, L. Schwarzkopf, M. Towsey, J. Zhang, P. Roe, Acoustic classification of frog within-species and species-specific calls, Appl. Acoust. 131 (feb 2018) 79–86.

[66] A. Brown, S. Garg, J. Montgomery, Automatic rain and cicada chorus filtering of bird acoustic data, Appl. Soft Comput. J. 81 (aug 2019).

[67] A. Truskinger, M. Towsey, P. Roe, Decision support for the efficient annotation of bioacoustic events, Ecol. Inform. 25 (jan 2015) 14–21.

[68] P. Nguyen Hong Duc, M. Torterotot, F. Samaran, P.R. White, O. Gérard, O. Adam, D. Cazau, Assessing inter-annotator agreement from collaborative annotation campaign in marine bioacoustics, Ecol. Inform. 61 (mar 2021).

[69] K.M. Bellisario, T. Broadhead, D. Savage, Z. Zhao, H. Omrani, S. Zhang, J. Springer, B.C. Pijanowski, Contributions of MIR to soundscape ecology. Part 3: tagging and classifying audio features using a multi-labeling k-nearest neighbor approach, Ecol. Inform. 51 (February) (2019) 103–111, https://doi.org/10.1016/j.ecoinf.2019.02.010. [Online].

[70] A. Thakur, P. Rajan, Directional embedding based semi-supervised framework for bird vocalization segmentation, Appl. Acoust. 151 (2019) 73–86, https://doi.org/10.1016/j.apacoust.2019.02.023. [Online].

[71] B. Mishachandar, S. Vairamuthu, Diverse ocean noise classification using deep learning, Appl. Acoust. 181 (2021) 108141, https://doi.org/10.1016/j.apacoust.2021.108141. [Online].

[72] Z.J. Ruff, D.B. Lesmeister, C.L. Appel, C.M. Sullivan, Workflow and convolutional neural network for automated identification of animal sounds, Ecol. Indic. 124 (July 2020) (2021) 107419, https://doi.org/10.1016/j.ecolind.2021.107419. [Online].

[73] C. Flowers, F.M. Le Tourneau, N. Merchant, B. Heidorn, R. Ferriere, J. Harwood, Looking for the -scape in the sound: discriminating soundscapes categories in the Sonoran desert using indices and clustering, Ecol. Indic. 127 (2021).

[74] J.F. Ruiz-Muñoz, G. Castellanos-Dominguez, M. Orozco-Alzate, Enhancing the dissimilarity-based classification of birdsong recordings, Ecol. Inform. 33 (2016) 75–84, https://doi.org/10.1016/j.ecoinf.2016.04.001. [Online].

[75] J.C. Ross, P.E. Allen, Random Forest for improved analysis efficiency in passive acoustic monitoring, Ecol. Inform. 21 (2014) 34–39, https://doi.org/10.1016/j.ecoinf.2013.12.002. [Online].

[76] I. Potamitis, Unsupervised dictionary extraction of bird vocalisations and new tools on assessing and visualising bird activity, Ecol. Inform. 26 (P3) (2015) 6–17, https://doi.org/10.1016/j.ecoinf.2015.01.002. [Online].

[77] O.M. Serra, F.P. Martins, L.R. Padovese, Active contour-based detection of estuarine dolphin whistles in spectrogram images, Ecol. Inform. 55 (August 2019) (2020) 101036, https://doi.org/10.1016/j.ecoinf.2019.101036. [Online].

[78] B.A. Martins, G.S.R. Rodrigues, C.B. de Araújo, Vocal dialects and their implications for bird reintroductions, Perspect. Ecol. Conserv. 16 (2) (2018) 83–89, https://doi.org/10.1016/j.pecon.2018.03.005. [Online].

[79] J.G. Colonna, E.F. Nakamura, O.A. Rosso, Feature evaluation for unsupervised bioacoustic signal segmentation of anuran calls, Expert Syst. Appl. 106 (2018) 107–120, https://doi.org/10.1016/j.eswa.2018.03.062. [Online].

[80] I. Sánchez-Gendriz, L.R. Padovese, Temporal and spectral patterns of fish choruses in two protected areas in southern Atlantic, Ecol. Inform. 38 (2017) 31–38.

[81] A.G. De Oliveira, T.M. Ventura, T.D. Ganchev, J.M. De Figueiredo, O. Jahn, M.I. Marques, K.L. Schuchmann, Bird acoustic activity detection based on morphological filtering of the spectrogram, Appl. Acoust. 98 (2015) 34–42, https://doi.org/10.1016/j.apacoust.2015.04.014. [Online].

[82] X. Chen, J. Zhao, Y. hua Chen, W. Zhou, A.C. Hughes, Automatic standardized processing and identification of tropical bat calls using deep learning approaches, Biol. Conserv. 241 (jan 2020).

[83] N. Hassan, D.A. Ramli, A comparative study of blind source separation for bioacoustics sounds based on FastICA, PCA and NMF, Proc. Comput. Sci. 126 (2018) 363–372, https://doi.org/10.1016/j.procs.2018.07.270. [Online].

[84] D.A. Ramli, H. Jaafar, Peak finding algorithm to improve syllable segmentation for noisy bioacoustic sound signal, Proc. Comput. Sci. 96 (2016) 100–109, https://doi.org/10.1016/j.procs.2016.08.105. [Online].

[85] J. Florentin, T. Dutoit, O. Verlinden, Detection and identification of European woodpeckers with deep convolutional neural networks, Ecol. Inform. 55 (August 2019) (2020) 101023, https://doi.org/10.1016/j.ecoinf.2019.101023. [Online].

[86] A.J. Fairbrass, P. Rennett, C. Williams, H. Titheridge, K.E. Jones, Biases of acoustic indices measuring biodiversity in urban areas, Ecol. Indic. 83 (February) (2017) 169–177, https://doi.org/10.1016/j.ecolind.2017.07.064. [Online].

[87] N. Rendon, S. Rodríguez-Buritica, C. Sanchez-Giraldo, J.M. Daza, C. Isaza, Automatic acoustic heterogeneity identification in transformed landscapes from Colombian tropical dry forests, Ecol. Indic. 140 (May) (2022) 109017, https://doi.org/10.1016/j.ecolind.2022.109017. [Online].

[88] C. Bedoya, C. Isaza, J.M. Daza, J.D. López, Automatic recognition of anuran species based on syllable identification, Ecol. Inform. 24 (2014) 200–209, https://doi.org/10.1016/j.ecoinf.2014.08.009. [Online].

[89] J.R. Gloaguen, A. Can, M. Lagrange, J.F. Petiot, Road traffic sound level estimation from realistic urban sound mixtures by non-negative matrix factorization, Appl. Acoust. 143 (2019) 229–238, https://doi.org/10.1016/j.apacoust.2018.08.018. [Online].

[90] T. Marin-Cudraz, B. Muffat-Joly, C. Novoa, P. Aubry, J.F. Desmet, M. Mahamoud-Issa, F. Nicolè, M.H. Van Niekerk, N. Mathevon, F. Sèbe, Acoustic monitoring of rock ptarmigan: a multi-year comparison with point-count protocol, Ecol. Indic. 101 (October 2018) (2019) 710–719.

[91] A.J. Torija, D.P. Ruiz, Á.F. Ramos-Ridao, A tool for urban soundscape evaluation applying support vector machines for developing a soundscape classification model, Sci. Total Environ. 482–483 (1) (2014) 440–451, https://doi.org/10.1016/j.scitotenv.2013.07.108. [Online].

[92] J. Ye, T. Kobayashi, M. Murakawa, Urban sound event classification based on local and global features aggregation, Appl. Acoust. 117 (2017) 246–256, https://doi.org/10.1016/j.apacoust.2016.08.002. [Online].

[93] J. LeBien, M. Zhong, M. Campos-Cerqueira, J.P. Velev, R. Dodhia, J.L. Ferres, T.M. Aide, A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network, Ecol. Inform. 59 (April 2020) (2020) 101113, https://doi.org/10.1016/j.ecoinf.2020.101113. [Online].

[94] M. Zhong, J. LeBien, M. Campos-Cerqueira, R. Dodhia, J. Lavista Ferres, J.P. Velev, T.M. Aide, Multispecies bioacoustic classification using transfer learning of deep convolutional neural networks with pseudo-labeling, Appl. Acoust. 166 (2020) 107375, https://doi.org/10.1016/j.apacoust.2020.107375. [Online].

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

[96] B. Williams, T.A. Lamont, L. Chapuis, H.R. Harding, E.B. May, M.E. Prasetya, M.J. Seraphim, J. Jompa, D.J. Smith, N. Janetski, A.N. Radford, S.D. Simpson, Enhancing automated analysis of marine soundscapes using ecoacoustic indices and machine learning, Ecol. Indic. 140 (January) (2022) 108986, https://doi.org/10.1016/j.ecolind.2022.108986. [Online].

[97] L.K. Katsis, A.P. Hill, E. Piña-Covarrubias, P. Prince, A. Rogers, C. Patrick Doncaster, J.L. Snaddon, Automated detection of gunshots in tropical forests using convolutional neural networks, Ecol. Indic. 141 (July) (2022).

[98] J.A. Zwerts, J.N. (Yannick) Wiegers, E.H. Sterck, M. (Marijke) van Kuijk, Exploring spatio-temporal variation in soundscape saturation of an African tropical forest landscape, Ecol. Indic. 137 (August 2021) (2022) 108712, https://doi.org/10.1016/j.ecolind.2022.108712. [Online].

[99] C. Jin, M. Kim, S. Jang, D.G. Paeng, Semantic segmentation-based whistle extraction of Indo-Pacific Bottlenose Dolphin residing at the coast of Jeju island, Ecol. Indic. 137 (December 2021) (2022) 108792, https://doi.org/10.1016/j.ecolind.2022.108792. [Online].

[100] N. Wijayathilaka, G. Senevirathne, C. Bandara, S. Rajapakse, R. Pethiyagoda, M. Meegaskumbura, Integrating bioacoustics, DNA barcoding and niche modeling for frog conservation – the threatened balloon frogs of Sri Lanka, Glob. Ecol. Conserv. 16 (2018) e00496, https://doi.org/10.1016/j.gecco.2018.e00496. [Online].

[101] S. Dröge, D.A. Martin, R. Andriafanomezantsoa, Z. Burivalova, T.R. Fulgence, K. Osen, E. Rakotomalala, D. Schwab, A. Wurz, T. Richter, H. Kreft, Listening to a changing landscape: acoustic indices reflect bird species richness and plot-scale vegetation structure across different land-use types in north-eastern Madagascar, Ecol. Indic. 120 (2021).
[102] X.C. Hong, G.Y. Wang, J. Liu, L. Song, E.T. Wu, Modeling the impact of soundscape drivers on perceived birdsongs in urban forests, J. Clean. Prod. 292 (apr 2021).
[103] F. Anders, A.K. Kalan, H.S. Kühl, M. Fuchs, Compensating class imbalance for acoustic chimpanzee detection with convolutional recurrent neural networks, Ecol. Inform. 65 (nov 2021).
[104] S. Siddagangaiah, C.F. Chen, W.C. Hu, T. Akamatsu, M. McElligott, M.O. Lammers, N. Pieretti, Automatic detection of dolphin whistles and clicks based on entropy approach, Ecol. Indic. 117 (February) (2020) 106559, https://doi.org/10.1016/j.ecolind.2020.106559. [Online].
[105] M. Zhong, R. Taylor, N. Bates, D. Christey, H. Basnet, J. Flippin, S. Palkovitz, R. Dodhia, J. Lavista Ferres, Acoustic detection of regionally rare bird species through deep convolutional neural networks, Ecol. Inform. 64 (January) (2021) 101333, https://doi.org/10.1016/j.ecoinf.2021.101333. [Online].
[106] Z. Zhao, S. hua Zhang, Z. yong Xu, K. Bellisario, N. hua Dai, H. Omrani, B.C. Pijanowski, Automated bird acoustic event detection and robust species classification, Ecol. Inform. 39 (April) (2017) 99–108, https://doi.org/10.1016/j.ecoinf.2017.04.003. [Online].
[107] P. Dugan, M. Pourhomayoun, Y. Shiu, R. Paradis, A. Rice, C. Clark, Using high performance computing to explore large complex bioacoustic soundscapes: case study for right whale acoustics, Proc. Comput. Sci. 20 (2013) 156–162.
[108] C.A. Quinn, P. Burns, G. Gill, S. Baligar, R.L. Snyder, L. Salas, S.J. Goetz, M.L. Clark, Soundscape classification with convolutional neural networks reveals temporal and geographic patterns in ecoacoustic data, Ecol. Indic. 138 (April) (2022) 108831, https://doi.org/10.1016/j.ecolind.2022.108831. [Online].
[109] J.J. Noda, C.M. Travieso, D. Sánchez-Rodríguez, Methodology for automatic bioacoustic classification of anurans based on feature fusion, Expert Syst. Appl. 50 (2016) 100–106.
[110] Z. Hao, H. Zhan, C. Zhang, N. Pei, B. Sun, J. He, R. Wu, X. Xu, C. Wang, Assessing the effect of human activities on biophony in urban forests using an automated acoustic scene classification model, Ecol. Indic. 144 (July) (2022) 109437, https://doi.org/10.1016/j.ecolind.2022.109437. [Online].
[111] L. Wilson, R. Constantine, T. van der Boon, C.A. Radford, Using timelapse cameras and machine learning to enhance acoustic monitoring of small boat sound, Ecol. Indic. 142 (July) (2022) 109182, https://doi.org/10.1016/j.ecolind.2022.109182. [Online].
[112] R. Pahuja, A. Kumar, Sound-spectrogram based automatic bird species recognition using MLP classifier, Appl. Acoust. 180 (sep 2021).
[113] S. Ntalampiras, Bird species identification via transfer learning from music genres, Ecol. Inform. 44 (January) (2018) 76–81, https://doi.org/10.1016/j.ecoinf.2018.01.006. [Online].
[114] E. Dufourq, C. Batist, R. Foquet, I. Durbach, Passive acoustic monitoring of animal populations with transfer learning, Ecol. Inform. 70 (May) (2022) 101688, https://doi.org/10.1016/j.ecoinf.2022.101688. [Online].