Review

A Methodological Literature Review of Acoustic Wildlife Monitoring Using Artificial Intelligence Tools and Techniques

Sandhya Sharma [1,*]**, Kazuhiko Sato** [1] **and Bishnu Prasad Gautam** [2]

1   The Department of Engineering, Muroran Institute of Technology, Muroran 050-0071, Japan
2   The Department of Economic Informatics, Kanazawa Gakuin University, Kanazawa 920-1392, Japan
*   Correspondence: 22096503@mmm.muroran-it.ac.jp

Abstract: Artificial intelligence (AI) has become a significantly growing field in the environmental sector due to its ability to solve problems, make decisions, and recognize patterns. The significance of AI in wildlife acoustic monitoring is particularly important because of the vast amounts of data that are available in this field, which can be leveraged for computer vision and interpretation. Despite the increasing use of AI in wildlife ecology, its future in acoustic wildlife monitoring remains uncertain. To assess its potential and identify future needs, a scientific literature review was conducted on 54 works published between 2015 and March 2022. The results of the review showed a significant rise in the utilization of AI techniques in wildlife acoustic monitoring over this period, with birds (N = 26) gaining the most popularity, followed by mammals (N = 12). The most commonly used AI algorithm in this field was Convolutional Neural Network, which was found to be more accurate and beneficial than previous categorization methods in acoustic wildlife monitoring. This highlights the potential for AI to play a crucial role in advancing our understanding of wildlife populations and ecosystems. However, the results also show that there are still gaps in our understanding of the use of AI in wildlife acoustic monitoring. Further examination of previously used AI algorithms in bioacoustics research can help researchers better understand patterns and identify areas for improvement in autonomous wildlife monitoring. In conclusion, the use of AI in wildlife acoustic monitoring is a rapidly growing field with a lot of potential. While significant progress has been made in recent years, there is still much to be done to fully realize the potential of AI in this field. Further research is needed to better understand the limitations and opportunities of AI in wildlife acoustic monitoring, and to develop new algorithms that can improve the accuracy and usefulness of this technology.

Keywords: artificial intelligence; bioacoustics; monitoring; review; wildlife

1. Introduction

Acoustic monitoring of wildlife is the process of capturing audio recordings of wildlife vocalizations and analyzing them to gain valuable information about the presence, distribution, and behavior of species [1]. This method is a powerful tool for wildlife conservation, as it provides a non-invasive way of monitoring wildlife populations, and can be used to gather information about species that are difficult to observe directly, such as those that are nocturnal, elusive [2–4] or live in remote areas [5]. In recent years, the progress of technology has caused the integration of artificial intelligence (AI) techniques in acoustic monitoring. AI algorithms are used to automate the detection and classification of wildlife sounds, reducing the time and effort required to manually process large amounts of audio data [2]. Because acoustic wildlife recordings are of low cost, processing the collected recordings is time-consuming while requiring extensive training to preprocess such recordings into meaningful data [1,6]. This has the potential to greatly improve the efficiency and accuracy of the process, allowing for the monitoring of multi-species and a greater number of sounds.

Since 1990, AI in bioacoustics monitoring has been conducted in marine habitats [7]; however, research on other species has appeared more recently [5]. Several different types of AI algorithms are used in acoustic monitoring including machine learning algorithms, such as Support Vector Machines (SVM), Random Forest (RF) classifiers, and deep learning algorithms such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) [6,8,9]. These algorithms can be trained on large datasets of audio recordings to detect and classify different types of wildlife sounds, such as calls, songs, and vocalizations [1,10]. Autonomous techniques work efficiently and successfully when processing huge vocalization datasets [11]. Despite the potential benefits of using AI in acoustic monitoring, there are several challenges and gaps in the research that need to be addressed. One major challenge is the need for larger and more diverse datasets to train the AI algorithms [12–14]. In order to ensure the accuracy and dependability of the output, it is important to have a large and diverse set of audio recordings that represent a wide range of species and vocalizations.

Another challenge is the need for standardization in the field. Given the diversity of species and habitats being monitored, as well as different types of recording equipment and AI algorithms being used, it is important to establish standard protocols for acoustic monitoring to ensure the accuracy and comparability of results. Finally, there is a need for further exploration of the potential of AI for monitoring a wider range of species and vocalizations. Currently, most of the research has focused on monitoring a limited number of species and types of vocalizations, and there is a need for further research to explore the potential of AI for the monitoring of other types of wildlife sounds and species.

In light of these challenges and gaps in the research, there is a strong need for a comprehensive examination of the current state of the art in acoustic monitoring of wildlife using AI. This review paper can provide a comprehensive overview of the various methods used for acoustic monitoring, the different types of AI algorithms used for detecting and identifying wildlife sounds, and the challenges and limitations associated with the methods. By conducting this review, the challenges and gaps in the research can be addressed, and the field can be advanced.

To enhance the use and development of AI-assisted bioacoustics monitoring in ecological studies, it is crucial to comprehend the prevailing trends, strengths, weakness, and research opportunities in this area. The purpose of this study is to offer a complete examination of AI technology employed in bioacoustics monitoring and delve into its utilization. This review aims to reflect the current development, patterns, trends, gaps, and innovations in the field of acoustic monitoring of wildlife as well as highlight any missing research areas and propose future research directions. By doing so, it promotes a more informed use of these innovative technologies and allows researchers to focus on the most recent achievements, providing an updated overview of the current state of acoustic wildlife monitoring. The review process encompasses a comprehensive examination of several species of AI-utilized acoustic surveillance work in marine and forested environments. Firstly, the review begins with a design that includes a review layout, research questions, search procedure, and selection benchmark. This is followed by the selection criteria benchmarking and evaluation to ensure the proper collection of studies. Next, the review provides an overview of recent studies in the field of AI-based acoustic monitoring in marine and forested ecosystems. Additionally, the review delves into the strengths and weaknesses of various bioacoustics features and AI classification methods used in wildlife monitoring. The review further compares different classification methods and examines simulation tools used in these studies. The review also touches upon the applications of AI-based acoustic monitoring in wildlife ecology and identifies the challenges and future needs in this field.

2. Materials and Methods

2.1. Review Procedure

This literature review utilized a scientific framework that was modified and improved based on the work of Bao and Xie [15] and Sharma et al. [16]. The revision included aspects such as review layout, gathering of appropriate review papers, conclusion, and discussion, resulting in a thorough and comparative outlook. The systematic approach employed in the review allowed for the identification of the most efficient and widely used AI classification algorithms (Figure 1).
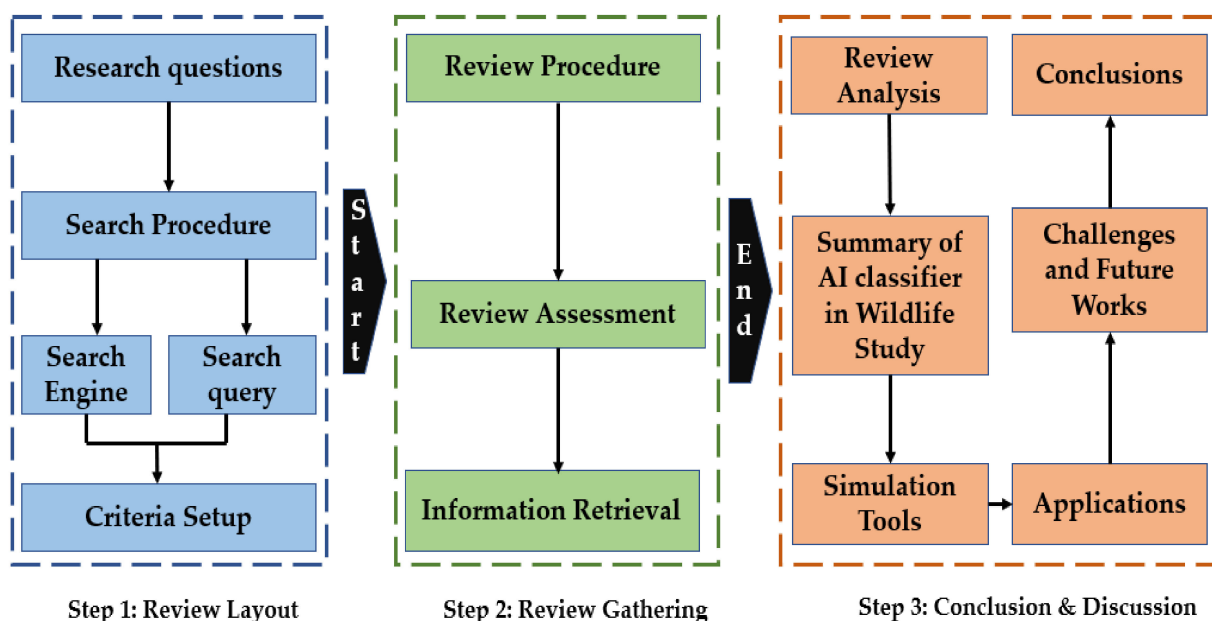
Figure 1. Workflow of the systematic literature review.

2.2. Research Questions

This literature review aimed to discover, quantify, and analyze published studies that utilized AI in wildlife acoustic monitoring. Four specific research questions were addressed in the review: (1) What recent advancements have been made in the use of AI for acoustic monitoring of wildlife in marine and forested ecosystems? (2) What are the strengths and weaknesses of using AI-based features and classification algorithms in wildlife acoustic monitoring? (3) What challenges exist and what future research is needed in AI-assisted wildlife acoustic monitoring? (4) What are the AI-powered simulation tools used in bioacoustics monitoring?

2.3. Selection Criteria Benchmarking

The study was to analyze the existing research on bioacoustics monitoring amidst AI. To achieve this goal, a thorough search was carried out using various search engines including Google Scholar and Sci-Hub. The keywords used during the search process included terms such as "acoustic monitoring", "wildlife", "marine", "aquatic", "terrestrial", "aerial", "bats", "whale", "birds", "elephant", "mammals", "invertebrates", "insects", "AI", "machine learning", "deep learning", and "big data". In selecting the relevant articles, we considered two important factors, including the focus of the article on the application of AI in acoustic monitoring of wildlife and its publication in recognized and credible sources such as peer-reviewed journals or related to conference papers. From a pool of over 222 articles, we finally reviewed only 73 articles (Table S1).

To ensure the relevance of the studies in the review, Kitchenham's [17] criteria were used to filter out applied research. This set of criteria consists of five important perspectives: a well-planned study with required data, relevant research methodology, clear research

objectives, accurate and critical analysis, and substantial technological advancement. As a result, 54 articles were selected for the final review based on their quality ratings. Until 2021, the most peer-reviewed studies on acoustic monitoring of wildlife using AI technology were related to avian species (Figure 2).
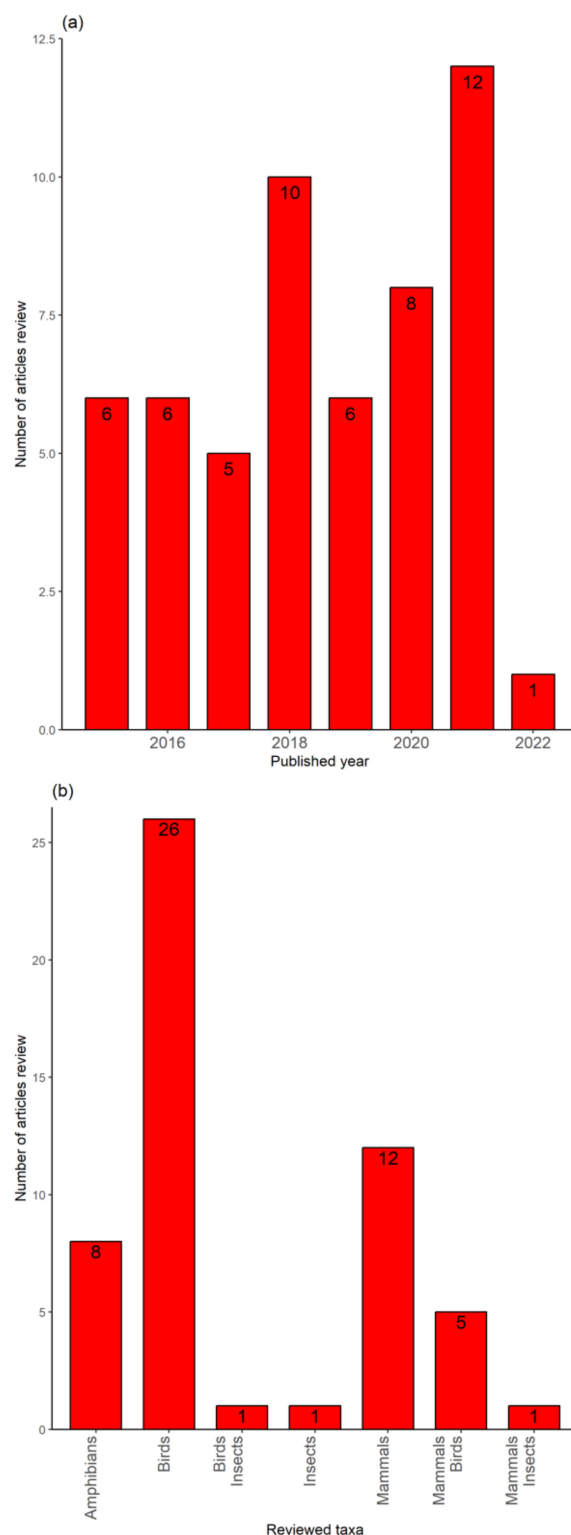

Figure 2. The number of articles reviewed: (**a**) 2015 through March 2022, (**b**) Taxa2.4. Review gathering.

2.4. Data Gathering Insights

In our systematic examination of the use of AI in bioacoustics monitoring, we analyzed relevant literature published between 2015 and March 2022. This allowed us to focus on recent advancements in the field and provide an updated overview. We collected information from each cited publication such as title, year of publication, journal name, species of animal studied, and AI classification features and methods used, as well as their performance. Our literature review analyzed the strengths, weaknesses, and research gaps surrounding the use of five acoustic wildlife monitoring features along with an AI classifier. The selection of five acoustic wildlife monitoring features was based on their prevalent usage in the peer-reviewed literature (most commonly utilized: 3, less frequently utilized: 2; however, with high performance). Only articles that utilized hybrid calcification techniques (combining multiple AI Algorithms) were considered in the comparison of models. The statistical graphs were created using R [18], and totals were calculated using the countif formula in an Excel spreadsheet.

3. Results

3.1. Summary of the Study Field with In-Depth Findings

Deep learning models, particularly deep learning neural networks are increasingly popular in the field of acoustic monitoring of wildlife in marine and forested ecosystems [19–26]. These models are capable of processing and analyzing large amounts of animal vocalizations amidst background noise, allowing them to detect and classify different species of animals [20,25]. However, one of the main challenges of using these models is the collection of high-quality datasets. To overcome this, researchers use a combination of automated and manual techniques to collect and annotate data, ensuring their models are trained on accurate and representative datasets [19,22,25,26]. In marine ecosystems, CNNs are used to monitor the vocalization of killer and humpback whales [19,20], while in forest ecosystems, CNN architecture has been used to analyze the call patterns of frogs [22,23]. Other AI methods used in the forest ecosystem include SVM classifiers [24,25] and multi-label learning approaches to identify rival frog species, and GMM for detecting the activity of frogs through their calls [26].

3.1.1. Summary of Current Work on AI-Assisted Wildlife Acoustic Monitoring in Marine Ecosystems

Acoustic monitoring is essential in the marine ecosystem as it helps researchers collect crucial information about species by identifying their unique vocalizations and evaluating their status and potential threats [19]. The data collected from bioacoustics detectors, mainly hydrophones, can span from hundreds to millions of acoustic monitoring datasets; however, only a small portion of the animal vocalizations can be used as they can be distinguished from the vast amount of background noise present. This makes it challenging to gather a sufficient number of vocalizations for in-depth analysis. To overcome this challenge, AI algorithms for marine sound recognition and classification have been developed. In recent studies, researchers have utilized CNNs to monitor the vocalizations of killer and humpback whales [19,20]. This approach was chosen due to the robust, automatic nature of CNNs, which enable the effective extraction of the concerned marine animal calls from large bioacoustics datasets in the presence of environmental noise. The use of AI algorithms in marine bioacoustics has greatly improved the accuracy and efficiency of vocalization analysis as AI models can process large amounts of acoustic data quickly and accurately and can be trained to identify specific vocalizations of interest, such as marine endangered or rare species. This has enabled researchers to study the behavior and distribution of marine mammals more effectively and has contributed to the development of conservation strategies to protect these species.

Acoustic wildlife monitoring in the marine environment is a challenging task due to the complex and dynamic nature of the ecosystem. AI provides significant advantages in this field, as it can process large amounts of acoustic data quickly over large areas, allowing for

the more accurate and reliable detection of wildlife. AI methods can be trained to recognize specific patterns in acoustic data, which is particularly useful in marine ecosystems where the environmental conditions can be challenging, such as studying the vocalization behavior of humpback whales [19]. However, there are several challenges associated with using AI methods in marine environments, such as the difficulty in collecting high-quality animal sounds that are often mixed with high levels of background noise, as seen in studying killer and humpback whales [19]. This can pose significant challenges as AI methods rely on large amounts of high-quality data to be effective. Additionally, the complexity of AI models can make them difficult to interpret, and false positives can occur, potentially leading to unnecessary interventions or disruptions to marine wildlife [19,20].

3.1.2. Summary on Bioacoustics Monitoring of Forest Environments Amidst AI Methods

The forest ecosystem is naturally filled with sounds from various sources, including wildlife. These sounds hold valuable information about the creatures' identity, status, size, and potential dangers. Moreover, calls can be heard from different directions, across barriers, and over long distances. As such, monitoring wildlife through sound is more effective than visual observation. In recent years, acoustic monitoring of wildlife in forests has gained considerable attention as it is a non-intrusive way to monitor a broad spectrum of wildlife species [21]. Auditory detectors have been employed in forest ecosystems to gather vast amounts of audio recordings over large geographic and temporal scales. However, most of the sounds captured are environmental noise, with only a few instances of wildlife calls included. To address this, AI algorithms have been created and applied to effectively identify wildlife calls amidst background noise in forested habitats. The study of Xie et al. [22,23] utilized CNN architecture to analyze the call patterns of frogs and found this approach was more accurate in recognizing the calls of the species in question compared to traditional methods (Table 1). Gan et al. [24] and Xie et al. [25] used an SVM classifier and multi-label learning approach, respectively, to identify the calls of two rival frog species (*Litoria olongburensis* and *Litoria fallax*) and their calling behavior. The accuracy of both classifiers was similar, ranging from 70 to 75%, as they both aimed to find a boundary or decision surface to distinguish the acoustic data into classes. Multi-label learning is a more comprehensive approach that can handle multiple classes, while SVM is a binary classifier that only separates two classes (Table 1). SVM is a popular supervised learning algorithm that uses both classification and regression analysis and can classify animal sounds based on various acoustic features such as frequency, duration, and aptitude [24].

Xie et al. [26] used GMM after extracting three bioacoustics indices to detect the activity of frogs through their calls (Table 1). GMM is widely used in a statistical model for unsupervised learning and clustering analysis. It assumes the acoustic data are generated from a mixture of Gaussian distributions and estimates the model parameters (mean, variance) of each Gaussian component to fit the data [26]. The CNN architecture is highly desirable and well-regarded among bioacoustics classification methods for frog detection and calling behavior as it achieves a high accuracy of over 95% (Table 1). CNNs are well-suited for identifying frog calls because they can handle large amounts of audio data even in the presence of background noise, quickly and accurately identify crucial features, and are relatively simple to use. The CRNN (Convolutional Neural Network + Recurrent Neural Network) architecture, which is a combination of CNN and RNN, was utilized to classify the activity patterns of Koalas through audio recordings. This method was found to be crucial in understanding Koalas' long-term behavior even in unknown settings, as it leverages the strength of both CNNs and RNNs to identify vocalizations and track behavior changes over time [22]. CRNNs combine convolutional and recurrent layers to extract local features from the acoustic spectrograms and capture the temporal dependencies between consecutive frames of the spectrograms. The convolutional layers extract spatial features from the spectrograms and the recurrent layers process the sequential information by using the outputs of previous frames to predict the features of the current frames [22]. Similarly,

various classifiers have been used for bird sound and calling behavior identification, including CNN architecture, the Gaussian mixture model, and Random Forest (RF) (Table 1). Out of these, CNN is particularly effective in classifying bird sounds and calling behavior due to its ability to extract significant acoustic features, handle background noise, differentiate between different bird calls, and attain high accuracy.

The CNN architecture has become increasingly popular in acoustic wildlife monitoring, delivering high accuracy rates of over 90% on unseen datasets (Table 1). Compared to traditional classifiers such as Simple Minded Audio Classifier in Python (SMACPY) and SVM, CNN is more robust and advanced, as it can extract relevant features automatically, handle complex patterns, and effectively learn from large amounts of vocalization data. Semi-automatic classification methods, such as Cubic SVM and RF, also have high accuracy rates of about 90% in identifying bird calls by combining human expertise with machine learning algorithms to handle complex patterns in acoustic data [27]. The deep Multiple Instance Learning (MIL) frameworks, however, have a lower accuracy of 0.77 (F1-score) compared to GMM and syllable-based models [28]. The hidden Markov model/Gaussian mixture model (HMM/GMM) classifier has the highest performance metrics for correct accuracy (VL02) due to its ability to handle complex vocalization patterns better [29] (Table 1). In acoustic wildlife monitoring, the HMM/GMM is a statistical model used for the sequence analysis and recognition of animal vocalizations. The model assumes that the acoustic data are generated from a sequence of hidden states modeled by Gaussian distributions and estimates their probability distribution using a hidden Markov model with a Gaussian mixture model for each state [29].

Acoustic monitoring of forest ecosystems involves capturing and analyzing audio recordings from the forest environment to study the biodiversity and behavior of wildlife species living in the forest. AI algorithms play a crucial role in this process by helping to accurately interpret the wildlife sounds and calls from the background noise. The use of AI algorithms in acoustic monitoring has become increasingly popular as they are capable of handling large amounts of audio data and can quickly and accurately identify important features in the recordings, even in the presence of background disturbances.

Table 1. Summary for the bioacoustics monitoring using AI methods.

Citations	Data	Taxa	AI Features	AI Classifiers
Himawan et al. [30]	Koala calls extracted	Mammals	Constant q-spectrogram	CNN + RNN or CRNN
Xie et al. [22]	24 h calls recordings	Amphibians	Multi-view spectrograms	CNN
Xie et al. [23]	1 h recordings	Amphibians	Mel-spectrogram	CNN-LSTM
Xie et al. [26]	24 h frog calls recordings	Amphibians	Shannon entropy, spectral peak track, harmonic index, and oscillation	GMM
Bergler et al. [19]	9000 h of vocalization recordings	Mammals	Mel-spectrogram	DNN
Adavanne et al. [31]	500 frames per clip	Bird	Dominant frequency, log Mel-band energy	CNN + RNN
Allen et al. [20]	187,000 h of vocalizations data	Mammals	Mel-spectrogram	CNN
Ruff et al. [11]	owl audio collections	Birds	Mel-spectrogram	CNN
Znidersic et al. [5]	30 days of audio recordings	Bird	LDFC spectrogram	RF Regression model
Prince et al. [32]	Bats: 128,000 samples of audio Cicada: 8192 vocalization data for training Gun-shoot: 32,768 vocalization data for training	Mammals, insects	Mel-spectrogram, MSFB, MFCC	CNN, HMM

Table 1. *Cont.*

Citations	Data	Taxa	AI Features	AI Classifiers
Zhang et al. [33]	audio at 44.1 kHz	Birds	Mel- spectrogram, STFT, MFCT, CT	SFIM based on deep CNN
Madhavi and Pamnani [34]	30 bird species vocalization recordings	Birds	MFCC	DRNN
Stowell et al. [35]	Data from Chernobyl Exclusion Zone (CEZ) repository	Birds	Mel-spectrograms, MFCCs	GMM, RF based on decision tree
Yang et al. [27]	Bird species 1200 syllabi	Birds	Wavelet packet decomposition	Cubic SVM, RF, RT
Castor et al. [28]	Bioacoustics with 48 KHz	Birds	MFCCs	Deep multi-instance learning architecture
Zhong et al. [36]	48 KHz audio recordings	Birds	Mel-spectrogram	Deep CNN
Zhao et al. [37]	Xeno-canto acoustic repository for vocalization	Birds	Mel-band-pass filter bank	GMM, event-based model
Ventura et al. [38]	Xeno-canto acoustic repository for vocalization	Birds	MFCCs	Audio parameterization, GMM, SBM, ROI model
de Oliveira et al. [29]	48 KHz acoustic recordings	Birds	Mel-spectrogram	HMM/GMM, GMM, SBM
Gan et al. [24]	24 h audio	Amphibians	LDFC spectrogram	SVM
Xie et al. [25]	512 audio samples	Amphibians	Linear Predictive Coefficient, MFCC, Wavelet-based features	Multi-label learning algorithm
Ruff et al. [39]	12 s calls recordings	Birds and mammals	Spectrogram	CNN
Ramli and Jaafar [40]	675 calls samples	Amphibians	Short-time energy (STE) and short-time average zero (STAZCR)	SBM
Cramer et al. [41]	vocalization of 14 birds classes	Birds	Mel-spectrogram	TaxoNet: deep CNN
Lostanlen et al. [42]	10 h audio recordings	Birds	Mel-spectrogram	Context-adaptive Neural Network (CN-NN)
Nanni et al. [43]	2814 audio samples	Birds, mammals	Spectrogram, harmonic	CNN
Nanni et al. [44]	Xeno-canto archives, online	Birds and mammals	Spectrogram	CNN
Nanni et al. [45]	Xeno-canto archives, online data	Birds and mammals	Spectrogram	CNN
Pandeya et al. [46]	Online database	Mammals	Mel-spectrogram	CNN, Convolutional Deep Belief Network (CDBN)
González-Hernández et al. [47]	Online database	Mammals	Whole spectrogram, the spectrogram of thump signal, octave analysis coefficient	Artificial Neural Network (ANN)
Incze et al. [48]	Xeno-canto archives for calls	Birds	Spectrogram	CNN
Zhang et al. [49]	2762 call events	Birds	Spectrogram	GMM, SVM
Sprengel et al. [50]	Above 33,000 calls events	Birds	Spectrogram	CNN
Lasseck [51]	Xeno-canto archive for birds vocalization	Birds	Spectrogram	Deep CNN
Zhang and Li [52]	Freesound repository for calls recordings	Birds	Mel-scaled wavelet packet decomposition sub-band cepstral coefficients (MWSCC)	SVM
Stowell et al. [53]	12 species audios	Birds	Mel-spectrogram	Random Forest, HMM
Salamon et al. [54]	5428 calls recordings	Birds	MFCCs	Deep CNN

Table 1. *Cont.*

Citations	Data	Taxa	AI Features	AI Classifiers
Noda et al. [55]	audio from 88 species of insects	Insects	MFCCs and linear frequency cepstral coefficients (LFCCs)	SVM
Ntalampiras [56]	Xeno-canto archives	Birds	Mel-scale filterbank, Short-time Fourier transform (STFT)	HMM, Randon Forest
Szenicer et al. [57]	8000 h of audio signal	Mammals	Spectrogram	CNN
Turesson et al. [12]	44.1 KHz frequency vocalization using microphone	Mammals	Time–frequency spectrogram	Optimum Path Forest, Multi-layer Artificial Neural Network, SVM, K-Nearest Neighbors, Logistic Regression, AdaBoost
Knight et al. [58]	Audio clip from 19 bird species	Birds	Spectrogram	CNN
Chalmers et al. [59]	Five distinct bird species audio clip 2104 individuals	Birds	Mel-frequency cepstrum (MFC)	Multi-layer Perceptrons
Zhang et al. [60]	1435 one-minute audio clips	Birds and Insects	Spectrogram	K-Nearest Neighbor, Decision Tree, Multi-layer Perceptrons
Jaafar and Ramli [10]	675 audios from 15 frog species	Amphibians	Multi-frequency cepstrum coefficient (MFCC)	K-Nearest Neighbor with Fuzzy Distance weighting
Bedoya et al. [13]	30 individuals with 54 recorders	Birds	Spectrogram	CNN
Brodie et al. [61]	512 audio samples per frames from 2584 total frames	Mammals	Spectrogram	SVM
Bermant et al. [9]	650 spectrogram images	Mammals	Spectrogram	CNN, LSTM-RNN
Zeppelzauer et al. [62]	335 min of audio recordings, annotated 635 rambles	Mammals	Spectrogram	SVM
Lopez-Tello et al. [6]	Audio recordings of birds and mammals	Birds and Mammals	MFCC	SVM
do Nascimento et al. [14]	24 h vocals recordings from six recorders	Mammals	Spectrogram with fast Fourier transformation	Linear Mixed Model
Yip et al. [63]	Vocalization during breeding season between 5:00–8:00 am	Birds	Spectrogram	K-Means Clustering
Dufourq et al. [64]	Eight song meter recorders from 1 March to 20 august 2016	Mammals	Spectrogram	CNN
Aodha et al. [65]	1024 audio samples between 5 KHz and 13 KHz frequency	Mammals	Spectrogram with fast Fourier transform	CNN

3.1.3. Performance Metrics: Different Performance Measures Are Frequently Used to Assess Acoustic Wildlife Monitoring Systems

The most popular performance statistic for a machine learning model used for acoustic wildlife monitoring is accuracy, which measures the ratio of the model's corrected predictions to all of its predictions. It is a good statistic for disseminating the findings of the animals' vocalization to a wider audience since it is an easy-to-understand, basic, and intuitive means of measuring performance [27,34,40,43,44,54–57,63,64]. However, accuracy can sometimes be a misleading metric, especially in imbalanced datasets where the number of instances of one class is significantly higher than the other. Therefore, it is

important to consider other metrics such as precision, recall, F1-score, and Area Under the Curve (AUC) and Receiver Operating Characteristic (AUC-ROC) to obtain a more complete picture of the model's performance. F1-score, however, evaluates the overall performance of the classification models through a balanced combination of precision and recall, which is another often-used performance statistic. This metric is a more thorough evaluation for identifying and categorizing animal vocalizations that consider both recall and accuracy. By considering both recall and accuracy, F1-score is a more robust metric for identifying and categorizing animal vocalizations in complex acoustic environments. Numerically, F1-score = 2 × (precision × recall)/(precision + recall) [22,23]. Xie et al. [22] conducted studies on amphibian species' call categorization, using a larger dataset with 24 h recordings that allowed for a more comprehensive representation of the acoustic features and variability in the target species, leading to a better-performing classification model, while the study conducted by Xie et al. [23] used only 1 hr recordings and achieved an F1-score of 83% indicating that the performance of classification model is reliant on the amount of data and the quality and diversity of the recordings, highlighting the need to consider these factors carefully in acoustic monitoring studies.

Both AUC and AUC-ROC assess the effectiveness of binary classifiers, which categorize a given species or group of species into one of two groups based on their acoustic signals: positive or negative. Between the two, AUC-ROC is the preferred statistic for acoustic wildlife monitoring because it provides a more extensive and in-depth evaluation of the model's performance [20,31]. A full view of the model performance is provided by precision when combined with recall (also called sensitivity) and F1-score. It is a crucial function used in acoustic wildlife monitoring used to prevent false positive predictions. It is especially helpful if the species is uncommon or on the edge of extinction [26]. Himawan et al. [30] obtained an AUC value of 87.46% in their study of the acoustic behavior of Koalas. Adavanne et al. [31] achieved an AUC value of 91.8% in their research on the categorization of bird calls. Meanwhile, Allen et al. [20] utilized AUC and AUC-ROC metrics to classify mammal sounds and obtained values of 97% and 99.2%, respectively. The study suggests that the classification models used in the studies of multi-species mammal sound outperformed those used in the research on bird calls. This difference could be attributed to the complexity of and variability in the acoustic signals produced by different species. Mammals are known to produce a wider variety of sounds with greater complexity, which may make them easier to distinguish using machine learning algorithms. Additionally, the habitats and behaviors of mammals may lend themselves better to acoustic monitoring than those of birds. However, further research is needed to fully understand the factors influencing the performance of classification models in acoustic monitoring. Recall (sensitivity) is used in conjunction with precision and F1-score to effectively identify all instances of positivity; but, unlike precision, which measures erroneous positive predictions, recall does not offer any guidance on how to do so [29]. Recall measures the proportion of actual positive instances that are correctly identified by the model, while precision measures the proportion of positive predictions made by the model that are actually true positive instances. F1-score is a harmonic mean of recall and precision, providing a balanced evaluation of the model's performance. Numerically, Recall = True Positive/(True Positive + False Negative); Precision = True Positive/(True Positive + False Positive); F1-score = 2 × (Precision × Recall)/(Precision + Recall). Here, true positives refer to the instances where the model correctly identifies positive cases such as detecting the presence of an animal in an audio recording, while false positives occur when the model incorrectly identifies a positive instance, such as identifying an animal in the recording where the species is not actually present. False negatives occur when the model incorrectly identifies a negative instance as a positive; for example, when the model fails to identify the presence of an animal in a recording where the animal is actually present [5,23,26].

3.2. Summary of Wildlife Acoustic Monitoring Features Amidst AI Classifiers

3.2.1. Summary of Strengths and Weakness for Bioacoustics Monitoring Amidst Bioacoustics' Features

A multi-view spectrum enhances overall identification performance as it suppresses the background noise to a greater extent and highlights specific features of the acoustic signals such as frequency, time, and amplitude, which are important for detection [22], whereas the mel-spectrogram in combination with the multi-view spectrum reduces background noise because it provides a comprehensive representation of the acoustic signals, which can differentiate between the target species vocalization (e.g., bird calls) and the background noises. This helps to improve the accuracy and reliability of acoustic wildlife monitoring, as it reduces false positive detections and enables detection (Table 2). The mel-spectrogram alone can make it harder to identify certain sounds made by wildlife such as killer whales. To achieve a better understanding of these vocalizations, it is important to use other techniques in addition to the mel-spectrogram. This will help improve the accuracy of identifying the target species' vocalization and overcome limitations [19]. The constant q-spectrogram method for analyzing wildlife vocalizations is more complicated and requires more computing power than other methods. It also has limitations in restoring original sounds and organizing the data. However, it can still aid in monitoring wildlife vocalizations by providing a closer look at the frequency of the target sounds. Researchers need to carefully consider the benefits and limitations of this method and may use additional techniques to gain a clear understanding of the target species' vocalizations. MFCC is a straightforward and versatile way to identify wildlife sounds. It assigns weights to specific data regions to reduce information loss while still capturing the important aspects of the acoustic in a simple way. Although MFCC may struggle with parallel processing and perform poorly in noisy environments, it can still be useful in monitoring wildlife vocalizations by using filters to enhance its performance [66]. In the extended readings, the LDFC spectrogram provides an in-depth acoustic structure. However, it is ineffective when the target species calls in the specific time period or when it appears with other species because the limited time window that is utilized to create a spectrogram may not capture species vocalization, resulting in data insufficiency. Similarly, the presence of other species leads to confusion and causes calls to overlap in the spectrogram simultaneously, making it difficult to distinguish the call of the target species [5] (Table 2).

Table 2. Summary of strengths and weaknesses with research gaps in the use of five features for wildlife acoustic monitoring.

Features	Strength	Weakness	Research Gap
Multi-view spectrogram	Improved accuracy, better sound differentiation, efficient detection and classification, multi-information capture, improved understanding of audio signals	reduces species recognition, not good quality for low-quality audio signals	recognize calling behavior, categorize species?
Mel-spectrogram	reduce background noises, higher signal-to-noise ratio, improved audio feature extraction	limited frequency range, computational complexity, lack of interpretability	applicability to the larger database?
Constant q-spectrogram	improved acoustic resolution reduces artifacts and distortions, better for complex sounds analysis	costly computation, lacks inverse transform, difficult data structure	replicability to other species

Table 2. *Cont.*

Features	Strength	Weakness	Research Gap
MFCC	simple, effective performance, adaptable, captured major features, weighted data value	less effective in noisy settings, filter influenced	applicable to larger databases?
LDFC spectrogram	auditory structure details, call recognition improved, species recognition boosted, improved signal details, robust noise reduction, better data understanding	limits detection accuracy, increased computation time, time-consuming analysis, limited frequency range, complex data interpretation	Multi-species supports?

3.2.2. Classification Methods' Comparison and Resources Required

Classifier in Acoustic Monitoring through Acoustics

In bioacoustics monitoring, the performance of different machine learning models is crucial in detecting a species' call accurately. One of the models used in this field is the CNN model. This model has been found to outperform the HMM in terms of F1-score, computational capabilities, and in detecting species' calls. This is because CNNs are capable of handling large amounts of data and background noises effectively, while HMMs struggle in noisy environments. The use of a CNN allows for the effective extraction of features from acoustic signals, which is crucial in species call detection. It utilizes convolutional layers to extract local features from spectrograms or other acoustic representations of sound, which are then down-sampled using pooling layers to generate feature maps. These maps are fed into fully connected layers to classify species or other acoustic events [22]. CNNs have demonstrated superior performance in a range of acoustic wildlife monitoring tasks, making them a popular choice for cutting-edge applications in ecology and conservation [21–23].

The combination of CNNs and RNNs in the CRNN architecture provides even more accurate results in species call detection. This architecture combines the strength of both models and allows for the consideration of the temporal structure of the signal is crucial in differentiating between species' calls and background noise, and the CRNN architecture effectively handles this aspect. The CRNN architecture has been found to provide more accurate results compared to HMMs, in detecting species' calls in ecological studies due to its ability to capture both temporal and spectral features of the acoustic signals, which is crucial for distinguishing between species and their calls. Additionally, CRNN can learn and extract features automatically from the input data, whereas HMM requires manual feature engineering, making it more time-consuming and less flexible [29,31]. Despite this, a syllable-based method is also a commonly used approach in bioacoustics monitoring that involves analyzing vocalizations to identify individual species and establish population estimates. This approach offers benefits such as accuracy, cost-effectiveness, and non-invasiveness. However, it also has limitations, including the need for specialized expertise, potential errors and biases, and vulnerability to environmental conditions [40].

Another model used in species call detection is the Long Short-Term Memory (LSTM) model, which focuses on sequential data. In comparison to the syllable-based model, the LSTM model performs better in detecting background noise. The LSTM model's ability to handle sequential data makes it a suitable choice for species call detection, where the temporal structure of the signal is important. However, despite its strength in handling sequential data, the LSTM model may not be as effective in handling large amounts of data and background noises as the CRNN architecture. ANNs are frequently used as a classifier for bioacoustics monitoring to enhance the accuracy and efficiency of acoustic wildlife monitoring, but their implementation demands thoughtful deliberation regarding the limitations and challenges such as the need for high-quality data, the potential for overfitting, and expertise in design and deployment, as demonstrated in the research

on marine mammals by González-Hernández et al. [47]. Overall, the choice of machine learning model depends on the specific requirements of the species call detection task and the characteristics of the acoustic signals being analyzed (Tables 1 and 2).

3.2.3. Resources Required for AI Monitoring of Wildlife through Acoustics

The use of AI algorithms in acoustic wildlife monitoring requires a large amount of data to achieve accurate results. Wildlife populations are often composed of a large number of different species, each with its own unique calls and sounds. To accurately identify these calls and distinguish between species, a large dataset is required to capture the variations and patterns in their sounds. Another reason for the need for a large dataset is the environmental variables. Environmental conditions, such as weather, background noise, and vegetation can impact the acoustic signals and make it difficult to accurately identify species. A large dataset can help account for these variables and improve accuracy. Similarly, for long-term monitoring, large datasets are required. Wildlife populations can change over time, and monitoring these changes requires a long-term dataset to track trends. For the purpose of using machine learning methods in wildlife acoustic monitoring, computational power is a crucial resource. In order to train and evaluate models using massive amounts of data, machine learning methods require a substantial number of computational capabilities. The training and testing process may be boosted and massive amounts of data can be handled with the help of High-Performance Computing (HPC) technologies such as computers and GPUs [19]. Access to cloud computing resources, such as Google Cloud Platform (GCP), can offer scalable and affordable computational capabilities for executing machine learning algorithms for monitoring animals through acoustics.

3.2.4. Summary of Strengths and Weakness for Acoustic Wildlife Monitoring Using AI Classifier

CNN can effectively analyze both small and large datasets containing species vocalization recordings and background noise. However, more research is needed to accurately count the number of sounds produced by specific species [11]. While the CRNN architecture is capable of identifying long-term patterns in the spectrograms of species' calls in unknown environments, it also has a higher rate of false positive identifications of multi-species' calls (Table 3). The GMM-based classifier has fewer parameters, but it is more susceptible to noise. On the other hand, the CNN-LSTM model is robust in species audio classification in noisy settings, but it requires each individual call to be labeled, making it time-consuming. HMMs have lower complexity, allowing them to run in real time, but they also have a high rate of false positives in the presence of background noise (Table 4). According to the particular issue and the data being examined, different acoustic monitoring classifiers are chosen and applied. For example, CNN is frequently used for audio classification issues and has good accuracy but high computing power requirements [11]. In contrast, CNN-RNN is employed for sequential acoustic monitoring datasets [23], whereas GMM is straightforward to implement but struggles with complicated situations [38].

Table 3. Strengths and weaknesses when comparing several classifiers.

Classification Methods	Strengths/Weakness
CNN and HMM	- CNN outperforms HMM, improving accuracy and complexity - CNN for strong image classification and object detection, while HMM is strong in time-series and sequence modeling
CNN and CNN + RNN	- CNN produces false positives in call recognition amidst background noise. - CNN is heavily used for image classification, while CRNN is for sequence learning

Table 3. *Cont.*

Classification Methods	Strengths/Weakness
CNN-LSTM and SBM	- CNN-LSTM accuracy is higher than that of the syllable-based classifier. - Unsupervised syllable noise detection, supervised CNN-LSTM noise detection - Syllable noise exclusion in training, CNN-LSTM includes noise as class - CNN-LSTM outperforms SBM
GMM and HMM	- GMM faster than HMM - GMM is more efficient than HMM

Table 4. Summary of strengths and weakness with research gap for five classifiers in bioacoustics monitoring.

Classification Methods	Strength	Weakness	Research Gaps
CNN	improves noise detection, automated feature learning, open-source software, reduce human error, improved species detection	high computational cost, need large labeled data, lacks temporal modeling	reliable on call counting?
CNN + RNN or CRNN	Accurate temporal modeling, reduces false positives, effectively filters noise	computationally demanding, requires labeled data, and may miss some signals.	enhancement of evaluation metrics
CNN-LSTM	improved temporal modeling, reduced false positive rate, better background noise handling	time-consuming labeling, complex network structure, high computational cost	unbalanced dataset suitability?
HMM	real-time processing speed, low false positive rate, simple model complexity	low accuracy rate, vulnerable to noise, simple modeling	evaluate multi-species performance
GMM	precise feature modeling, real-time processing ability, and fewer parameters needed	complexity in modeling, false positive with noise	conduct performance investigation detailed

3.3. Acoustic Wildlife Monitoring Simulation Tools

MATLAB provides a range of tools and functions for audio processing, such as filtering, and signal analysis. This makes it an ideal tool for cleaning and preparing audio data for analysis. A researcher can use MATLAB to perform various operations on audio recordings, such as removing noise, denoising, and transforming audio signals into a format suitable for analysis [30]. Audacity provides a range of audio editing and recording capabilities. In the context of acoustic wildlife monitoring, it can be used to visualize audio recordings, identify specific sounds, and apply various filters to enhance the audio quality [22]. Audacity also allows users to make annotations and labels, which can be useful for annotating specific sounds or segments of audio recordings for later analysis. Python is a powerful programming language that has a wide range of libraries for data analysis and machine learning. In wildlife, acoustic monitoring Python is commonly used to build and train machine learning models for tasks such as species identification and sound event detection, providing a range of tools for building and training neural networks [3]. Keras, on the other hand, provides a high-level interface to TensorFlow, allowing researchers to build and train machine learning models with less complexity and fewer coding efforts.

3.4. Uses of Acoustic Wildlife Monitoring Amidst AI Methods

AI technology has been applied in various aspects of acoustic wildlife monitoring including behavioral patterns [22,23,26], species recognition [24,25,29], species density/population estimation, species diversity [30], and illegal trade [32] (Table 1 and Figure 3). These applications have helped researchers understand the location and migration patterns of animals, as well as their social group affiliations. By monitoring animal behavior patterns, such as mating, diet, and predatory routines, AI systems can aid in wildlife conservation and protection against natural disasters or illegal activities. In addition, acoustic monitoring using AI can assist in the rescue and tracking of wildlife in wildlife sanctuaries and aid in the survival of endangered species. The approach also has benefits for surveillance, wildlife counts, and poaching control. AI-based acoustic monitoring offers valuable insights into the impacts of human activities on wildlife habitats, enabling real-time detection of changes in animal behavior and population dynamics that can inform conservation strategies. It also fosters collaboration among stakeholders, including researchers, wildlife managers, and conservation organizations, through data sharing. However, the approach faces challenges such as the need for accurate algorithms, ethical consideration, and potential biases in data collection and analysis, requiring responsible and transparent practices to ensure effective wildlife conservation.
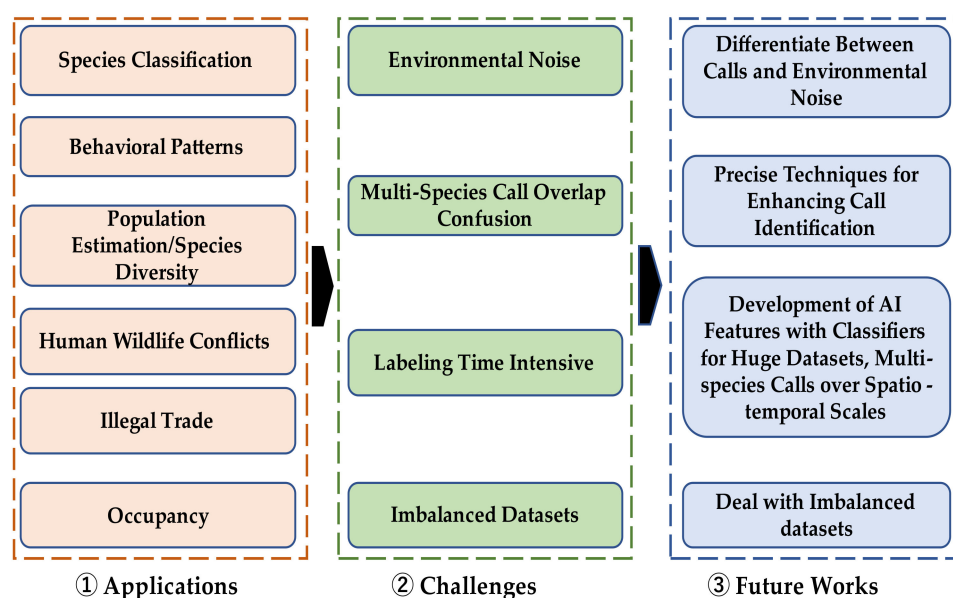

① Applications	② Challenges	③ Future Works
Species Classification	Environmental Noise	Differentiate Between Calls and Environmental Noise
Behavioral Patterns	Multi-Species Call Overlap Confusion	Precise Techniques for Enhancing Call Identification
Population Estimation/Species Diversity	Labeling Time Intensive	Development of AI Features with Classifiers for Huge Datasets, Multi-species Calls over Spatio-temporal Scales
Human Wildlife Conflicts		
Illegal Trade	Imbalanced Datasets	Deal with Imbalanced datasets
Occupancy		

Figure 3. Applications with challenges and future directions for bioacoustics monitoring amidst AI.

3.5. Active versus Passive Sensing Methods in Acoustic Monitoring in Wildlife

There are two main approaches to evaluating the population of animal species based on auditory monitoring: Passive Acoustic Monitoring (PAM) and Active Acoustic Monitoring (AAM). PAM involves capturing the sounds that animals emit naturally, whereas AAM utilizes a transmitted acoustic signal that bounces off the target animal and is detected upon return (Table 5). These techniques provide valuable insights into wildlife populations and are frequently used in bioacoustics sensing studies. The sensing methods presently use AI algorithms for wildlife automatic call recognition and the classification of marine and forested species, for instance SVM, RF, and CNNs [8,20,24]; however, we highly recommend the use of passive sensing methods for wildlife acoustic monitoring as it is cost-effective, relies on the detection of wildlife using their naturally produced sound, and does not require the use of transmitters or other equipment that alters the behavior of wildlife, which is crucial for the long-term monitoring purposes [67].

Table 5. The major differences between the two acoustic wildlife monitoring sensing methods.

Background	Active Acoustic Monitoring	Passive Acoustic Monitoring
Introduction	Detects the vocalization that wildlife emits naturally	Detects vocalization bounced off concerned species from the transmitted acoustic signal
Strengths	Non-invasive, less likely to disturb wildlife	Ability to target particular wildlife species
Weakness	Limited to vocalization produced naturally	Invasive, potential to affect wildlife species' behavior

3.6. Practitioner Consideration

Both in marine and forest habitats, collecting acoustic wildlife monitoring datasets is a challenging process. The researcher should employ data preparation techniques including filtering, spectrogram analysis, and machine learning approaches to extract useful information from the recordings as target species' vocalization records are influenced by environmental noise levels [52]. The geographical and temporal distribution of recordings, which is essential for comprehending the ecology and behavior of the concerned species, should be gathered by the researcher [22]. The researcher must have a thorough understanding of high-quality, appropriate acoustic monitoring equipment that can record the frequency range of the target species [4]. In our opinion, upholding moral standards and acquiring the required permits are crucial for ensuring the welfare of animals and avoiding legal repercussions. In order to guarantee the effectiveness and sustainability of the monitoring program, involvement with stakeholders such as local communities and conservation groups should also be carried out.

3.7. Challenges and Future Works

The challenges and future work for the acoustic monitoring of wildlife using AI involve addressing the current limitations and exploring new and innovative ways to improve the accuracy and efficiency of the process. One of the main challenges is the accuracy of AI algorithms. Currently, AI algorithms for acoustic monitoring are not always accurate in detecting and classifying wildlife sounds, particularly in noisy and complex environments. This can result in false positive and false negative detections, which can impact the accuracy of the results. To address these challenges, future work should focus on improving the accuracy of AI algorithms through the development of new and innovative techniques for feature extraction, data preprocessing, and model training. Many literature reviews [8,19,29,35] found background noise to be a sensitive factor that reduces classification accuracy and recommended robust feature extraction algorithms that perform well even for background noise. Due to the presence of more background noises than the bioacoustics training datasets, the detection and classification algorithms show poorer performances [25]. Background noise masks the target vocalizations, and generates false positives, although this is prior to filtering the background noise. The AI algorithms regularly fail to distinguish background noises from the target wildlife vocalizations, leading to low recall [68].

Another challenge is the scalability of the process. Acoustic monitoring generates large amounts of data, and processing these data can be time-consuming and computationally intensive. This can be a barrier to the widespread adoption of acoustic monitoring for wildlife conservation and management. To address this challenge, future work should focus on developing efficient and scalable AI algorithms for acoustic monitoring, as well as exploring new and innovative ways to process and store the data. A third challenge is the lack of standardization in the field. There is a need to establish standard protocols for acoustic monitoring, including the types of recording equipment used, the methods for collecting and processing audio data, and the AI algorithms used for detection and

classification. This will help to ensure the accuracy and comparability of results and will promote collaboration between researchers and wildlife organizations.

Next, the challenge is dealing with call overlap, where multiple species produce sounds at similar frequencies, making it difficult to distinguish between them. To effectively monitor multiple species, AI algorithms need to be able to differentiate and classify these overlapping calls accurately. Future work should focus on developing AI algorithms that can address these challenges and overcome the limitations of current methods. This could involve improving feature extraction techniques, training AI models on larger and more diverse datasets, and incorporating additional sources of information, such as spatial and metadata, to improve classification accuracy. Additionally, there is a need for interdisciplinary collaboration between computer scientists, wildlife biologists, and acoustic experts to advance the state of the art in AI-based acoustic wildlife monitoring.

Another challenge is labeling the image datasets to ensure that the data are accurate and comprehensive. This requires a large amount of manual effort to collect and label the data, as well as a thorough understanding of the species and their acoustic signals. Furthermore, there may be variations in the calls produced by different individuals of the same species, making it difficult to create a representative dataset. Future work in this field should focus on developing more efficient and automated methods for collecting and labeling image datasets. This could involve using existing data from wildlife monitoring programs, as well as developing new tools for data collection. Finally, it is important to consider ethical and privacy issues when collecting and using data from wildlife and to ensure the data are collected and used in a responsible manner.

4. Conclusions

This review encompasses the latest studies on the application of bioacoustics monitoring and AI learning in predicting several ecological outcomes. The results of all the research conducted in this area have been favorable. The AI algorithm assists in ecological studies by analyzing and interpreting a large amount of collected data, providing insights and conclusions based on the findings. The AI method focuses a larger emphasis on prediction than convolutional statistical methods and evaluates species identification and classification with better precision.

Species' vocalization predictions are challenging due to the limited presence of species' sounds and the high proportion of background noise in large-scale bioacoustics libraries, making manual retrieval of vocalizations for analysis difficult. Using AI algorithms can overcome these difficulties, resulting in more accurate identification and analysis of animal vocalizations. AI methods can also be used to analyze large and complex datasets, providing important insights into animal behavior and communication.

Traditional statistical methods have limited precision when dealing with complex datasets. As the number of inputs increases, these methods become less accurate. To address these limitations, AI techniques are utilized to effectively analyze complex data and overcome the challenges posed by traditional statistical models. The application of AI in bioacoustics monitoring has mainly focused on avian and mammalian species, with the highest number of articles published in the year 2021 and the least in 2022. Among the AI learning methods, CNN showed high accuracy and was more frequently used than other methods. The highest F1 score of 99.60% was achieved by combining multi-view features with CNN. However, multi-view features have only been used once in bioacoustics monitoring and more research is recommended to determine their accuracy.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data associated with this study are included as Supplementary Material.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Campos, I.B.; Landers, T.J.; Lee, K.D.; Lee, W.G.; Friesen, M.R.; Gaskett, A.C.; Ranjard, L. Assemblage of Focal Species Recognizers—AFSR: A technique for decreasing false indications of presence from acoustic automatic identification in a multiple species context. *PLoS ONE* **2019**, *14*, e0212727. [CrossRef] [PubMed]
2. Digby, A.; Towsey, M.; Bell, B.D.; Teal, P.D. A practical comparison of manual and autonomous methods for acoustic monitoring. *Methods Ecol. Evol.* **2013**, *4*, 675–683. [CrossRef]
3. Knight, E.C.; Hannah, K.C.; Foley, G.J.; Scott, C.D.; Brigham, R.M.; Bayne, E. Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs. *Avian Conserv. Ecol.* **2017**, *12*, 14. [CrossRef]
4. Jahn, O.; Ganchev, T.; Marques, M.I.; Schuchmann, K.-L. Automated Sound Recognition Provides Insights into the Behavioral Ecology of a Tropical Bird. *PLoS ONE* **2017**, *12*, e0169041. [CrossRef]
5. Znidersic, E.; Towsey, M.; Roy, W.; Darling, S.E.; Truskinger, A.; Roe, P.; Watson, D.M. Using visualization and machine learning methods to monitor low detectability species—The least bittern as a case study. *Ecol. Inform.* **2019**, *55*, 101014. [CrossRef]
6. Lopez-Tello, C.; Muthukumar, V. Classifying Acoustic Signals for Wildlife Monitoring and Poacher Detection on UAVs. In Proceedings of the 21st Euromicro Conference on Digital System Design (DSD), Prague, Czech Republic, 29–31 August 2018; pp. 685–690. [CrossRef]
7. Steinberg, B.Z.; Beran, M.J.; Chin, S.H.; Howard, J.H., Jr. A neural network approach to source localization. *J. Acoust. Soc. Am.* **1991**, *90*, 2081–2090. [CrossRef]
8. Gibb, R.; Browning, E.; Glover-Kapfer, P.; Jones, K.E. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods Ecol. Evol.* **2018**, *10*, 169–185. [CrossRef]
9. Bermant, P.C.; Bronstein, M.M.; Wood, R.J.; Gero, S.; Gruber, D.F. Deep Machine Learning Techniques for the Detection and Classification of Sperm Whale Bioacoustics. *Sci. Rep.* **2019**, *9*, 12588. [CrossRef]
10. Jaafar, H.; Ramli, D.A. Effect of Natural Background Noise and Man-Made Noise on Automated Frog Calls Identification System. *J. Trop. Resour. Sustain. Sci. (JTRSS)* **2015**, *3*, 208–213. [CrossRef]
11. Ruff, Z.J.; Lesmeister, D.B.; Duchac, L.S.; Padmaraju, B.K.; Sullivan, C.M. Automated identification of avian vocalizations with deep convolutional neural networks. *Remote Sens. Ecol. Conserv.* **2019**, *6*, 79–92. [CrossRef]
12. Turesson, H.K.; Ribeiro, S.; Pereira, D.R.; Papa, J.P.; de Albuquerque, V.H.C. Machine Learning Algorithms for Automatic Classification of Marmoset Vocalizations. *PLoS ONE* **2016**, *11*, e0163041. [CrossRef]
13. Bedoya, C.L.; Molles, L.E. Acoustic Censusing and Individual Identification of Birds in the Wild. *bioRxiv* **2021**. [CrossRef]
14. Nascimento, L.A.D.; Pérez-Granados, C.; Beard, K.H. Passive Acoustic Monitoring and Automatic Detection of Diel Patterns and Acoustic Structure of Howler Monkey Roars. *Diversity* **2021**, *13*, 566. [CrossRef]
15. Bao, J.; Xie, Q. Artificial intelligence in animal farming: A systematic literature review. *J. Clean. Prod.* **2022**, *331*, 129956. [CrossRef]
16. Sharma, S.; Sato, K.; Gautam, B.P. Bioacoustics Monitoring of Wildlife using Artificial Intelligence: A Methodological Literature Review. In Proceedings of the 2022 International Conference on Networking and Network Applications (NaNA), Urumqi, China, 3–5 December 2022; pp. 1–9.
17. Kitchenham, B.A.; Charters, S.M. Guidelines for Performing Systematic Literature Review in Software Engineering. EBSE Technical Report, EBSE-2007-01. 2007. Available online: https://www.researchgate.net/publication/302924724_Guidelines_for_performing_Systematic_Literature_Reviews_in_Software_Engineering (accessed on 22 April 2023).
18. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021.
19. Bergler, C.; Schröter, H.; Cheng, R.X.; Barth, V.; Weber, M.; Nöth, E.; Hofer, H.; Maier, A. ORCA-SPOT: An Automatic Killer Whale Sound Detection Toolkit Using Deep Learning. *Sci. Rep.* **2019**, *9*, 10997. [CrossRef]
20. Allen, A.N.; Harvey, M.; Harrell, L.; Jansen, A.; Merkens, K.P.; Wall, C.C.; Cattiau, J.; Oleson, E.M. A Convolutional Neural Network for Automated Detection of Humpback Whale Song in a Diverse, Long-Term Passive Acoustic Dataset. *Front. Mar. Sci.* **2021**, *8*, 607321. [CrossRef]
21. Sugai, L.S.M.; Silva, T.S.F.; Ribeiro, J.W.; Llusia, D. Terrestrial Passive Acoustic Monitoring: Review and Perspectives. *Bioscience* **2018**, *69*, 15–25. [CrossRef]
22. Xie, J.; Zhu, M.; Hu, K.; Zhang, J.; Hines, H.; Guo, Y. Frog calling activity detection using lightweight CNN with multi-view spectrogram: A case study on Kroombit tinker frog. *Mach. Learn. Appl.* **2021**, *7*, 100202. [CrossRef]
23. Xie, J.; Hu, K.; Zhu, M.; Guo, Y. Bioacoustic signal classification in continuous recordings: Syllable-segmentation vs sliding-window. *Expert Syst. Appl.* **2020**, *152*, 113390. [CrossRef]

24. Gan, H.; Zhang, J.; Towsey, M.; Truskinger, A.; Stark, D.; van Rensburg, B.J.; Li, Y.; Roe, P. A novel frog chorusing recognition method with acoustic indices and machine learning. *Future Gener. Comput. Syst.* **2021**, *125*, 485–495. [CrossRef]

25. Xie, J.; Michael, T.; Zhang, J.; Roe, P. Detecting Frog Calling Activity Based on Acoustic Event Detection and Multi-label Learning. *Procedia Comput. Sci.* **2016**, *80*, 627–638. [CrossRef]

26. Xie, J.; Towsey, M.; Yasumiba, K.; Zhang, J.; Roe, P. Detection of anuran calling activity in long field recordings for bio-acoustic monitoring. In Proceedings of the 10th International Conference on Intelligence Sensors, Sensor Network and Information (ISSNIP), Singapore, 7–9 April 2015.

27. Yang, S.; Friew, R.; Shi, Q. Acoustic classification of bird species using wavelets and learning algorithm. In Proceedings of the 13th International Conference on Machine Learning and Computing (ICMLC), New York, NY, USA, 26 February–1 March 2021.

28. Castor, J.; Vargas-Masis, R.; Alfaro-Rojan, D.U. Understanding variable performance on deep MIL framework for the acoustic detection of Tropical birds. In Proceedings of the 6th Latin America High Performance Computing Conference (CARLA), Cuenca, Ecuador, 2–4 September 2020.

29. de Oliveira, A.G.; Ventura, T.M.; Ganchev, T.D.; de Figueiredo, J.M.; Jahn, O.; Marques, M.I.; Schuchmann, K.-L. Bird acoustic activity detection based on morphological filtering of the spectrogram. *Appl. Acoust.* **2015**, *98*, 34–42. [CrossRef]

30. Himawan, I.; Towsey, M.; Law, B.; Roe, P. Deep learning techniques for koala activity detection. In Proceedings of the 19th Annual Conference of the International Speech Communication Association (INTERSPEECH), Hyderabad, India, 2–6 September 2018.

31. Adavanne, S.; Drossos, K.; Cakir, E.; Virtanen, T. Stacked convolutional and recurrent neural networks for bird audio detection. In Proceedings of the 25th Europena Signal Processing Conference (EUSIPCO), Greek Island, Greece, 28 August–2 September 2017.

32. Prince, P.; Hill, A.; Piña Covarrubias, E.; Doncaster, P.; Snaddon, J.L.; Rogers, A. Deploying Acoustic Detection Algorithms on Low-Cost, Open-Source Acoustic Sensors for Environmental Monitoring. *Sensors* **2019**, *19*, 553. [CrossRef] [PubMed]

33. Zhang, F.; Zhang, L.; Chen, H.; Xie, J. Bird Species Identification Using Spectrogram Based on Multi-Channel Fusion of DCNNs. *Entropy* **2021**, *23*, 1507. [CrossRef]

34. Madhavi, A.; Pamnani, R. Deep learning based audio classifier for bird species. *Int. J. Sci. Res.* **2018**, *3*, 228–233.

35. Stowell, D.; Wood, M.D.; Pamuła, H.; Stylianou, Y.; Glotin, H. Automatic acoustic detection of birds through deep learning: The first Bird Audio Detection challenge. *Methods Ecol. Evol.* **2018**, *10*, 368–380. [CrossRef]

36. Zhong, M.; Taylor, R.; Bates, N.; Christey, D.; Basnet, H.; Flippin, J.; Palkovitz, S.; Dodhia, R.; Ferres, J.L. Acoustic detection of regionally rare bird species through deep convolutional neural networks. *Ecol. Inform.* **2021**, *64*, 101333. [CrossRef]

37. Zhao, Z.; Zhang, S.-H.; Xu, Z.-Y.; Bellisario, K.; Dai, N.-H.; Omrani, H.; Pijanowski, B.C. Automated bird acoustic event detection and robust species classification. *Ecol. Inform.* **2017**, *39*, 99–108. [CrossRef]

38. Ventura, T.M.; de Oliveira, A.G.; Ganchev, T.D.; de Figueiredo, J.M.; Jahn, O.; Marques, M.I.; Schuchmann, K.-L. Audio parameterization with robust frame selection for improved bird identification. *Expert Syst. Appl.* **2015**, *42*, 8463–8471. [CrossRef]

39. Ruff, Z.J.; Lesmeister, D.B.; Appel, C.L.; Sullivan, C.M. Workflow and convolutional neural network for automated identification of animal sounds. *Ecol. Indic.* **2021**, *124*, 107419. [CrossRef]

40. Ramli, D.A.; Jaafar, H. Peak finding algorithm to improve syllable segmentation for noisy bioacoustics sound signal. *Procedia Comput. Sci.* **2016**, *96*, 100–109. [CrossRef]

41. Cramer, A.L.; Lostanlen, V.; Farnsworth, A.; Salamon, J.; Bello, J.P. Chirping up the Right Tree: Incorporating Biological Taxonomies into Deep Bioacoustic Classifiers. In Proceedings of the 45th International Conference on Acoustic, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020.

42. Lostanlen, V.; Salamon, J.; Farnsworth, A.; Kelling, S.; Bello, J.P. Robust sound event detection in bioacoustic sensor networks. *PLoS ONE* **2019**, *14*, e0214168. [CrossRef]

43. Nanni, L.; Costa, Y.M.G.; Aguiar, R.L.; Mangolin, R.B.; Brahnam, S.; Silla, C.N., Jr. Ensemble of convolutional neural networks to improve animal audio classification. *EURASIP J. Audio Speech Music Process.* **2020**, *2020*, 8. [CrossRef]

44. Nanni, L.; Maguolo, G.; Paci, M. Data augmentation approaches for improving animal audio classification. *Ecol. Inform.* **2020**, *57*, 101084. [CrossRef]

45. Nanni, L.; Maguolo, G.; Brahnam, S.; Paci, M. An Ensemble of Convolutional Neural Networks for Audio Classification. *Appl. Sci.* **2021**, *11*, 5796. [CrossRef]

46. Pandeya, Y.R.; Kim, D.; Lee, J. Domestic Cat Sound Classification Using Learned Features from Deep Neural Nets. *Appl. Sci.* **2018**, *8*, 1949. [CrossRef]

47. González-Hernández, F.R.; Sánchez-Fernández, L.P.; Suárez-Guerra, S.; Sánchez-Pérez, L.A. Marine mammal sound classification based on a parallel recognition model and octave analysis. *Appl. Acoust.* **2017**, *119*, 17–28. [CrossRef]

48. Incze, A.; Jancsa, H.; Szilagyi, Z.; Farkas, A.; Sulyok, C. Bird sound recognition using a convolutional neural network. In Proceedings of the 16th International Symposium on Intelligent Systems and Informatics (SISI), Subotica, Serbia, 13–19 September 2018; pp. 000295–000300.

49. Zhang, S.-H.; Zhao, Z.; Xu, Z.-Y.; Bellisario, K.; Pijanowski, B.C. Automatic Bird Vocalization Identification Based on Fusion of Spectral Pattern and Texture Features. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 271–275.

50. Sprengel, E.; Jaggi, M.; Kilcher, Y.; Hofmann, T. *Audio Based Species Identification Using Deep Learning Techniques*; CEUR: Berlin, Germany, 2016.

51. Lasseck, M. *Audio Based Bird Species Identification Using Deep Convolutional Neural Networks*; CEUR: Berlin, Germany, 2018.

52. Zhang, X.; Li, Y. Adaptive energy detection for bird sound detection in complex environments. *Neurocomputing* **2015**, *155*, 108–116. [CrossRef]

53. Stowell, D.; Benetos, E.; Gill, L.F. On-Bird Sound Recordings: Automatic Acoustic Recognition of Activities and Contexts. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2017**, *25*, 1193–1206. [CrossRef]

54. Salamon, J.; Bello, J.P.; Farnsworth, A.; Kelling, S. Fusing shallow and deep learning for bioacoustic bird species classification. In Proceedings of the 2017 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP), New Oreans, LA, USA, 5–9 March 2017; pp. 141–145.

55. Noda, J.J.; Travieso, C.M.; Sanchez-Rodriguez, D.; Dutta, M.K.; Singh, A. Using bioacoustic signals and Support Vector Machine for automatic classification of insects. In Proceedings of the IEEE International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 11–12 February 2016; pp. 656–659. [CrossRef]

56. Ntalampiras, S. Bird species identification via transfer learning from music genres. *Ecol. Inform.* **2018**, *44*, 76–81. [CrossRef]

57. Szenicer, A.; Reinwald, M.; Moseley, B.; Nissen-Meyer, T.; Muteti, Z.M.; Oduor, S.; McDermott-Roberts, A.; Baydin, A.G.; Mortimer, B. Seismic savanna: Machine learning for classifying wildlife and behaviours using ground-based vibration field recordings. *Remote Sens. Ecol. Conserv.* **2021**, *8*, 236–250. [CrossRef]

58. Knight, E.C.; Hernandez, S.P.; Bayne, E.M.; Bulitko, V.; Tucker, B. Pre-processing spectrogram parameters improve the accuracy of bioacoustic classification using convolutional neural networks. *Bioacoustics* **2019**, *29*, 337–355. [CrossRef]

59. Chalmers, C.; Fergus, P.; Wich, S.; Longmore, S.N. Modelling Animal Biodiversity Using Acoustic Monitoring and Deep Learning. In Proceedings of the 2021 International Joint Conference on Neural Network (IJCNN), Shenzhen, China, 18–22 July 2021; pp. 1–7. [CrossRef]

60. Zhang, L.; Towsey, M.; Xie, J.; Zhang, J.; Roe, P. Using multi-label classification for acoustic pattern detection and assisting bird species surveys. *Appl. Acoust.* **2016**, *110*, 91–98. [CrossRef]

61. Brodie, S.; Allen-Ankins, S.; Towsey, M.; Roe, P.; Schwarzkopf, L. Automated species identification of frog choruses in environmental recordings using acoustic indices. *Ecol. Indic.* **2020**, *119*, 106852. [CrossRef]

62. Zeppelzauer, M.; Hensman, S.; Stoeger, A.S. Towards an automated acoustic detection system for free-ranging elephants. *Bioacoustics* **2014**, *24*, 13–29. [CrossRef]

63. Yip, D.A.; Mahon, C.L.; MacPhail, A.G.; Bayne, E.M. Automated classification of avian vocal activity using acoustic indices in regional and heterogeneous datasets. *Methods Ecol. Evol.* **2021**, *12*, 707–719. [CrossRef]

64. Dufourq, E.; Durbach, I.; Hansford, J.P.; Hoepfner, A.; Ma, H.; Bryant, J.V.; Stender, C.S.; Li, W.; Liu, Z.; Chen, Q.; et al. Automated detection of Hainan gibbon calls for passive acoustic monitoring. *Remote Sens. Ecol. Conserv.* **2021**, *7*, 475–487. [CrossRef]

65. Mac Aodha, O.; Gibb, R.; Barlow, K.E.; Browning, E.; Firman, M.; Freeman, R.; Harder, B.; Kinsey, L.; Mead, G.R.; Newson, S.E.; et al. Bat detective—Deep learning tools for bat acoustic signal detection. *PLoS Comput. Biol.* **2018**, *14*, e1005995. [CrossRef]

66. Mohammed, R.A.; Ali, A.E.; Hassan, N.F. Advantages and disadvantages of automatic speaker recognition systems. *J. Al-Qadisiyah Comput. Sci. Math.* **2019**, *11*, 21–30.

67. Melo, I.; Llusia, D.; Bastos, R.P.; Signorelli, L. Active or passive acoustic monitoring? Assessing methods to track anuran communities in tropical savanna wetlands. *Ecol. Indic.* **2021**, *132*, 108305. [CrossRef]

68. Goeau, H.; Glotin, H.; Vellinga, W.P.; Planque, R.; Joly, A. LifeCLFF bird identification task 2016: The arrival of deep learning. In *CLEF: Conference and Labs of the Evaluation Forum, Évora, Portugal, 5–8 September 2016*; CEUR: Berlin, Germany, 2016; pp. 440–449.