Revista Brasileira de Ciências Ambientais
Brazilian Journal of Environmental Sciences

Advances in avian acoustic recognition through artificial intelligence: a systematic review of techniques and environmental applications

Avanços no reconhecimento acústico de aves por meio da inteligência artificial: uma revisão sistemática de técnicas e aplicações ambientais

André Maciel Pelanda[1], *Alysson Nunes Diógenes*[1]

ABSTRACT

The accelerating loss of biodiversity has driven the adoption of automated technologies for environmental monitoring, including the acoustic recognition of birds using artificial intelligence. This study aimed to systematically review the primary methods employed in automatic recognition of bird vocalizations, with an emphasis on the evolution of techniques and their environmental applications. The integrative literature review covered publications from 2013 to 2025, with searches conducted on the databases Scopus, Web of Science, IEEE Xplore, and Google Scholar, using terms related to machine learning and bioacoustics. After screening 2,435 publications, 25 studies were selected for in-depth analysis. The findings indicate a methodological shift from Mel-Frequency Cepstral Coefficients to Convolutional Neural Networks, highlighting improvements in classification accuracy and noise robustness. It is concluded that neural networks are becoming increasingly effective tools for biodiversity conservation, although challenges remain regarding model generalization and computational cost.

Keywords: convolutional neural networks; birdsong; machine learning.

RESUMO

A crescente perda de biodiversidade tem impulsionado o uso de tecnologias automatizadas para o monitoramento ambiental, entre elas o reconhecimento acústico de aves por meio de inteligência artificial. Esta pesquisa teve como objetivo revisar sistematicamente os principais métodos empregados no reconhecimento automático do canto de aves, com ênfase na evolução das técnicas e suas aplicações ambientais. A revisão integrativa da literatura abrangeu publicações de 2013 a 2025, com buscas realizadas nas bases Scopus, Web of Science, IEEE Xplore e Google Scholar, utilizando termos relacionados a aprendizado de máquina e bioacústica. Após a triagem de 2.435 publicações, 25 estudos compuseram o *corpus* da análise. Os resultados indicam uma transição dos Coeficientes Cepstrais de Frequência Mel para Redes Neurais Convolucionais, com destaque para ganhos em acurácia e robustez ao ruído. Conclui-se que as redes neurais vêm se consolidando como ferramentas promissoras para a conservação da biodiversidade, embora persistam desafios quanto à generalização dos modelos e ao custo computacional.

Palavras-chave: redes neurais convolucionais; canto de aves; aprendizado de máquina.

[1]Universidade Positivo – Curitiba (PR), Brazil.

Corresponding author: André Maciel Pelanda – Universidade Positivo – Programa de Pós-Graduação em Gestão Ambiental – Rua Professor Pedro Viriato Parigot de Souza, 5300 – CEP 81280-330 – Cidade Industrial de Curitiba – Curitiba (PR), Brazil. E-mail: andre.pelanda@yahoo.com.br

Conflicts of interest: the authors declare no conflicts of interest.

Funding: none.

Received on: 03/24/2025. Accepted on: 09/17/2025.

https://doi.org/10.5327/Z2176-94782514

Introduction

Human activities exert multiple environmental pressures, such as habitat modification, pollution, and climate change, which have caused unprecedented impacts on global biodiversity (Keck et al., 2025). Birds serve as effective bioindicators of environmental quality due to their sensitivity to ecological changes and their usefulness in monitoring both aquatic and terrestrial ecosystems (Maznikova et al., 2024). According to García et al. (2024), birds play a crucial role in maintaining ecological balance, contributing significantly to the control of pest insect populations as well as to the seed dispersal of numerous plant species. Blake and Loiselle (2024) report that Amazonian bird populations have undergone significant declines even in areas without direct human disturbance, indicating that more subtle environmental and ecological factors may also affect their population dynamics. Cooke et al. (2023) argue that many bird species' extinctions are driven by human actions, occurring either directly through activities such as hunting or indirectly through human-related impacts, including land-use change, wildfi\res, and the introduction of invasive species. The relative ease of detecting birds visually and acoustically, combined with the extensive knowledge available on their taxonomy, ecology, and conservation, makes them an ideal model group for research investigating species–environment interactions (Lees et al., 2022).

Stastny et al. (2018) reveal that birds use each vocal expression to convey various types of information, such as alerts about approaching dangers, territorial marking, and the attraction of females during the breeding season. Sick (1984) notes that many bird species exhibit morphological similarities and are distinguished by their vocalizations, making this trait a crucial element for species classification. Certain acoustic characteristics of bird vocalizations, including frequency parameters, are strongly linked to the evolutionary history and genetic basis of species, which makes them valuable for identification and classification analyses (Rivera et al., 2023).

The use of automatic recording devices for bird songs in field research offers several advantages over traditional methods, which require the continuous physical presence of the observer throughout the study period. Computational methods for representing and searching bioacoustic events in large audio datasets enable more efficient analyses that are less costly in terms of time and resources (Dong et al., 2013). Traditionally, point counts have been widely applied for bird monitoring. However, passive recording methods using autonomous units have become more economical and effective alternatives for monitoring bird communities in the field (Schuster et al., 2024).

Recent studies have shown that the use of autonomous recording units enables the large-scale, continuous collection of acoustic data, allowing both the identification of species occupancy patterns and the detection of impacts caused by human activities, such as intensive agricultural management (Molina-Mora et al., 2024). This approach is particularly effective given that birds communicate through vocalizations produced by the syrinx, a specialized vocal organ, and that in many passerine species, vocalizations occur around 4,000 Hz—a frequency close to the highest note of a piano (Sick, 1984).

The recording of bird sounds in natural environments can vary greatly depending on the distance between the sound source and the microphone and the type of landscape, both of which directly affect signal quality and the accuracy of automatic species identification (Somervuo et al., 2023). Xie et al. (2023) reveal that there is a growing interest in the automatic recognition of bird vocalizations, mainly due to the use of autonomous recording devices, which can be installed at strategic locations within study areas, enabling continuous sound capture 24 hours a day, 7 days a week. Hill et al. (2013) observed that significant differences may exist in vocalizations between bird subspecies inhabiting distinct regions, as in the case of the mainland New Zealand tui (*Prosthemadera novaeseelandiae novaeseelandiae*) and the Chatham Island tui (*Prosthemadera novaeseelandiae chathamensis*).

Modern computer science techniques, such as machine learning, have become increasingly popular in research involving wildlife, biodiversity, and ecosystems (Das et al., 2020). Acoustic monitoring has proven to be an effective, low-cost, and non-invasive tool for identifying priority areas for conservation, enabling the detection of endemic bird species and the analysis of environmental factors that influence their spatial distribution (Inoue et al., 2025).

Among the most widely used methods for automatic bird song recognition are Mel-Frequency Cepstral Coefficients (MFCCs) and Convolutional Neural Networks (CNNs). Liu et al. (2022) demonstrated that a multi-scale CNN architecture significantly improves bird song classification performance compared to traditional methods. CNNs are networks capable of autonomously learning to identify patterns in sounds converted into images, such as spectrograms, as shown in the work by Giri et al. (2025), which achieved over 90.0% accuracy in species identification.

According to Mascorro and Torres (2013), the spectrogram is a graphical representation of the frequency spectrum of a sound signal, generated through the application of the Short-Time Fourier Transform. This technique makes it possible to visualize the distribution of sound energy over time, with the horizontal axis representing time and the vertical axis corresponding to frequency. The signal amplitude is expressed through color variations, allowing for a detailed analysis of the acoustic characteristics of bird vocalizations.

Considering the ecological importance of birds, the advances in sound recording technologies, and the potential of automatic recognition as support for environmental research, it is essential to expand studies that utilize vocalizations as a monitoring tool. In this context, the present study aimed to investigate the use of automatic bird vocalization recognition as an effective tool for environmental monitoring.

Methodology

This research was conducted through an integrative literature review, with searches carried out on the databases Scopus, Web of

Science, IEEE Xplore, and Google Scholar. The terms "birdsong recognition," "deep learning for bioacoustics," "CNN bird sound classification," and "automated species identification" were used, combined with Boolean operators. Articles published between 2013 and 2025 that addressed the use of machine learning in the automatic recognition of bird sounds were included. The initial search yielded 2,435 publications. After removing duplicates, studies lacking methodological descriptions, and documents outside the scope (e.g., theses, dissertations, and extended abstracts), 374 articles remained for analysis. At the end of the screening process, 25 studies composed the review body, with emphasis on the use of MFCCs, Mel spectrograms, and CNNs, revealing methodological patterns and research gaps in the field.

To ensure the quality and relevance of the review, exclusion criteria were defined regarding methodology, data presentation, and study timeframe. Articles that did not report quantitative performance metrics (such as accuracy, precision, or F1-score) were excluded, as were those that failed to provide methodological details, since the lack of such information compromises the replicability and reliability of the results. Studies that addressed exclusively manual methods of bird identification were also excluded, as the focus of this work is on automated approaches based on artificial intelligence.

Additionally, only studies published from 2013 onward were considered, as this period has been marked by significant advances in the field of automatic vocalization recognition. Articles published prior to this date were excluded, except when regarded as relevant theoretical milestones. Studies with small sample sizes for model training and testing were also disregarded, as they do not provide sufficient statistical robustness. As noted by Pereira and Centeno (2017), the larger the training sample size, the greater the potential accuracy of supervised classification.

Literature Review

Table 1 presents the distribution of scientific production by country, based on the institutional affiliation of the corresponding authors, as well as the degree of international collaboration. The analysis of publications indexed in the Scopus database—collected following the methodological criteria described earlier and processed using the Bibliometrix package in RStudio—reveals that China leads in scientific production on the topic with 27 articles (20.3%) but shows low international collaboration (multi-country publications [MCP] ratio=0.037). In contrast, Brazil and Malaysia, with four articles each, present a high proportion of multinational collaborations (MCP ratio=0.750). Germany and France display a balance between national and international production, while the United States, with 12 articles, stands out for its entirely autonomous production. These data highlight different strategies of scientific engagement among the countries analyzed.

The results not only reveal the countries with the highest scientific output and varying levels of international collaboration but also indicate methodological trends that have been guiding studies on the subject.

Table 1 – Distribution of scientific production by country (based on corresponding authors' affiliation) and degree of international collaboration.

Country	Articles	Frequency (%)	SCP (n)	MCP (n)	MCP (ratio)
China	27	20.30	26	1	0.037
India	13	9.77	10	3	0.231
United States	12	9.02	12	0	0.000
Germany	11	8.27	8	3	0.273
France	9	6.77	5	4	0.444
United Kingdom	8	6.02	6	2	0.250
Australia	6	4.51	2	4	0.667
Brazil	4	3.01	1	3	0.750
Malaysia	4	3.01	1	3	0.750
Netherlands	4	3.01	3	1	0.250

SCP: single-country publications; MCP: multi-country publications.

Within this context, the recurrent use of artificial intelligence-based techniques stands out, particularly for their ability to handle large volumes of data and extract complex patterns. Among these approaches, Artificial Neural Networks (ANNs) and their more advanced architectures have received special attention, becoming central tools in recent research.

According to Qamar and Zardari (2023), ANNs are highly parallel systems composed of a large number of interconnected basic processors. Braga et al. (2007) explain that artificial neurons receive input values, assign weights and other values (biases) to these inputs, and generate outputs. This process can be applied to a variety of tasks, such as classification, regression, prediction, or data clustering.

The recognition of visual features is a task naturally performed by the human brain, engaging a substantial proportion of the cerebral cortex in the analysis of visual information (Eickhoff et al., 2008; Himmelberg et al., 2022). CNNs stand out as effective architectures for image classification and object recognition, achieving performances comparable to those of humans in various applications (Celeghin et al., 2023).

Raschka and Mirjalili (2017) point out that the primary objective of the convolution operation in a CNN is to extract distinctive features from the input image. This operation preserves the spatial relationship among pixels, enabling the neural network to learn image features from different regions. As a result of the convolution, a feature map is generated that highlights the most relevant areas of the image for classification or recognition tasks.

Xie et al. (2023) observe that the structure of a call is often relatively simple, typically monosyllabic or disyllabic, although patterns vary across taxa, with bird vocalizations being simpler but more diverse than human speech. Somervuo et al. (2006) state that bird phonetics can be categorized into three main levels. The note is the smallest sound element in birdsong, characterized by the absence of intervals. A syllable consists of one or more grouped notes, separated by short intervals, usually maintaining a repetitive pattern. A phrase is composed

of one or more syllables, and when there is a change in the pattern of these syllables or a significant interval between them, it is understood that one phrase has ended and another has begun.

For the automatic recognition of birdsong, it is essential to consider the acoustic nonlinearity of vocalizations and to apply appropriate parameterization methods that allow the extraction of key features, even in environments with external noise (Stastny et al., 2018). In this process, the preprocessing of recordings is crucial for removing unwanted sounds and isolating the vocalizations of scientific interest. Xie et al. (2023) observe that this stage generally involves three main steps: pre-emphasis, noise reduction, and segmentation.

The presence of noise impairs the quality of bird vocalization signals, resulting in low performance in the automatic recognition of field recordings, as highlighted by Potamitis et al. (2014). Xie et al. (2020) proposed a model with an autoencoder, showing that noise and low signal-to-noise ratio negatively affect the accuracy of bird classification. In the frequency domain, a classical strategy for noise reduction is spectral subtraction, which consists of subtracting the noise spectrum from the noisy signal spectrum (Xie et al., 2015).

According to Priyadarshani et al. (2020), when a sound is correlated with the vocalizations of a specific species, it can be used as an important feature for the recognition of these vocalizations. However, in natural environments, as pointed out by Xie et al. (2023), sound tends to degrade due to the radiation effect during transmission. Specifically, high-frequency components attenuate more rapidly than low-frequency components. To compensate for the decay of high-frequency components, pre-emphasis is often applied as the first step in the preprocessing of recordings (Xie et al., 2018).

Priyadarshani et al. (2018) investigated how environmental factors and habitat characteristics influence the recording of bird vocalizations by autonomous recording units (ARUs) in the field. The results show that variables such as temperature, wind, cloud cover, and lunar phase directly affect recording quality and species detectability. Furthermore, environments with dense vegetation interfere with sound propagation, reducing the effectiveness of acoustic monitoring. These findings emphasize the importance of considering environmental and structural conditions when planning bioacoustic studies and monitoring programs using ARUs.

The BirdCLEF challenge was launched in 2014. Since then, many studies have been conducted to overcome the challenges associated with automatic birdsong recognition, given the large volume of data and the high diversity of species. As cited by Kahl et al. (2021), in the early editions of the event, the MFCC method was widely used by participating researchers. Koops et al. (2014) reported achieving birdsong recognition accuracy rates ranging between 10–20% in the test sets when using MFCC—results similar to those of other teams participating in the challenge.

Due to the large volume of data collected—over 14,000 recordings covering 501 species—many participants in BirdCLEF 2014 reported memory and processing constraints, which led them to reduce the dataset and select smaller samples during the feature extraction phase (Joly et al., 2014). These limitations highlight the need for developing more efficient methods and advanced classifiers capable of handling large volumes of data.

MFCCs are used to evaluate birdsong datasets separately, as mentioned by Stastny et al. (2018). This method not only enables the processing of bird vocalization data but can also be adapted from techniques originally developed for human speech processing.

In the 2016 BirdCLEF challenge, participants faced the task of analyzing sounds from approximately 1,000 bird species, with recordings in MP3 format sourced from the online library Xeno-Canto. Each file varied in duration and could contain vocalizations from multiple species. Piczak (2016) achieved a mean average precision of 41.2% in BirdCLEF 2016, while the best submission in the challenge reached approximately 70.0%. In the same year, Sprengel et al. (2016) employed a classifier based on a CNN for the automatic recognition of bird sounds. The authors trained spectrograms for this purpose, reserving 10.0% of the data for validation and using the remaining 90.0% for training the network. The results outperformed competing methodologies by more than 10.0%, highlighting the efficiency of CNNs in handling the diversity and complexity of bird sounds in challenging datasets such as BirdCLEF's.

Tóth and Czeba (2016), also participating in the 2016 BirdCLEF challenge, employed CNNs to analyze the spectrograms of bird sounds. During data preprocessing, irrelevant parts of the sounds were removed, and each spectrogram was divided into five-second segments, which were then used as input to the CNN. The results demonstrated that the deep learning-based approach is suitable for the task of automatic bird species recognition from sounds. However, the authors emphasized the need for fine-tuning to achieve higher accuracy rates, indicating that there is still room for methodological optimizations to address the diversity and complexity of the data.

Han and Peng (2023) demonstrated that the Error-Correcting Output Codes with Support Vector Machine (ECOC-SVM) model outperformed CNNs and other classifiers in bird vocalization classification, reaching 100% accuracy on a small dataset of 11 species. While this result highlights the method's efficiency, it also underscores limitations in generalization, as CNNs typically require larger training sets to achieve robust performance and avoid overfitting (Goodfellow et al., 2016; Salamon and Bello, 2017)—a condition in which the model becomes overly fitted to the training data and fails to generalize to new contexts.

In a study conducted by Lauha et al. (2022), 1,000 vocalization files from 101 species in the Macaulay Library were used to train a neural network. The data, containing both target species sounds and background noise, were processed with the Animal Sound Identifier software, which generated spectrograms and identified species-specific patterns. The automatic selections were manually refined, with colored boxes visually highlighting the species in the spectrograms.

Currently, there are mobile applications that use CNNs and spectrograms to classify more than 1,000 bird species, such as BirdNET. Kahl t al. (2021) state that BirdNET demonstrates high effectiveness in

identifying birdsong, enabling access to data that was previously difficult to obtain. The application is free and allows users to record bird sounds, select excerpts for identification, and store metadata anonymously on a server, facilitating studies on the vocalizations of recorded species.

McGinn et al. (2023) reveal that the BirdNET application can be considered a robust and applicable tool for the classification of acoustic events within a single species, as it enables the differentiation of acoustic events with important implications for ecology and conservation. An example is the differentiation of vocalizations between adults and juveniles of the great gray owl (*Strix nebulosa*): juveniles produce low-pitched food-begging calls, while adults emit low-frequency sounds used for long-range communication and territorial defense.

According to Kershenbaum et al. (2014), birds display a wide variability in their vocal repertoires. Moreover, data obtained through autonomous recordings may vary depending on the distance of the birds, resulting in changes in amplitude and frequency of the signals, which poses a challenge for the accurate identification of specific species. Lauha et al. (2022) indicate that training a neural network with global data can be effective, but fine-tuning with locally collected data can significantly improve recognition performance under specific environmental conditions. Such customized adjustments allow for more precise adaptation to the unique characteristics of a given environment, leading to more accurate and applicable results for targeted studies.

Table 2 shows the methodological trends used in research involving the automatic recognition of birdsong.

The analysis of the examined studies highlights a significant methodological evolution in the field of automatic recognition of bird vocalizations, driven by advances in artificial intelligence and acoustic signal processing techniques. This trajectory, which initially relied on MFCCs combined with traditional statistical methods, has progressively incorporated deep neural networks, resulting in substantial improvements in accuracy and scalability. However, despite these advances, challenges remain that must be addressed to ensure the applicability of these systems in real and ecologically diverse scenarios.

Table 2 – Studies and methods used in the recognition of bird vocalizations.

Study	Method	Period
Dufour et al. (2014)	MFCC	Predominant use of MFCC (2013–2015)
Koops et al. (2014)	MFCC	Predominant use of MFCC (2013–2015)
Verdin and Kumar (2015)	MFCC	Predominant use of MFCC (2013–2015)
Tóth and Czeba (2016)	CNNs	Transition to CNNs (2016–2020)
Sprengel et al. (2016)	CNNs	Transition to CNNs (2016–2020)
Ramirez et al. (2018)	MFCC/IMFCC	Transition to CNNs (2016–2020)
Incze et al. (2018)	CNNs	Transition to CNNs (2016–2020)
Xie et al. (2019)	CNNs	Transition to CNNs (2016–2020)
LeBien et al. (2020)	CNNs	Transition to CNNs (2016–2020)
Hidayat et al. (2021)	CNNs	Advancements in CNNs and new models (2021–2025)
Maegawa et al. (2021)	CNNs	Advancements in CNNs and new models (2021–2025)
Hong and Zabidi (2021)	CNNs	Advancements in CNNs and new models (2021–2025)
Permana et al. (2022)	CNNs	Advancements in CNNs and new models (2021–2025)
Lauha et al. (2022)	CNNs	Advancements in CNNs and new models (2021–2025)
Xie et al. (2022)	CNNs	Advancements in CNNs and new models (2021–2025)
Han and Peng (2023)	ECOC-SVM	Advancements in CNNs and new models (2021–2025)
Clark et al. (2023)	CNNs	Advancements in CNNs and new models (2021–2025)
García-Ordás et al. (2023)	FCN	Advancements in CNNs and new models (2021–2025)
Mohanty et al. (2023)	SNN	Advancements in CNNs and new models (2021–2025)
Uddin et al. (2024)	CNNs	Advancements in CNNs and new models (2021–2025)
Espejo et al. (2024)	ANNs applied to acoustic indices	Advancements in CNNs and new models (2021–2025)
Zhang et al. (2024)	CNNs	Advancements in CNNs and new models (2021–2025)
Krishna et al. (2024)	CNNs	Advancements in CNNs and new models (2021–2025)
He and Luo (2025)	CNNs (EfficientNet-B0 optimized with ECA and CBAM)	Advancements in CNNs and new models (2021–2025)
Márquez-Rodríguez et al. (2025)	CNNs + YOLOv8 + embeddings BirdNET	Advancements in CNNs and new models (2021–2025)

MFCCs: Mel-Frequency Cepstral Coefficients; IMFCC: Inverse MFCC; CNNs: Convolutional Neural Networks; ECOC-SVM: Error-Correcting Output Codes with Support Vector Machine; FCN: Fully Convolutional Networks; SNN: Spiking Neural Network; ANNs: Artificial Neural Networks applied to acoustic indices; ECA: Efficient Channel Attention; CBAM: Convolutional Block Attention Module.

Early studies, such as those by Dufour et al. (2014), Koops et al. (2014), and Verdin and Kumar (2015), employed MFCCs for acoustic feature extraction, exploring classifiers such as SVMs and k-Nearest Neighbor (k-NN). Although widely applied in audio signal analysis due to their ability to model human auditory perception, these methods exhibited limitations when confronted with high intraspecific variability and the growing volume of bioacoustic data.

Ramirez et al. (2018) compared MFCCs and Inverse Mel-Frequency Cepstral Coefficients (IMFCCs), showing that IMFCCs could outperform MFCCs in certain contexts. However, the reliance on conventional classifiers still limits the performance of these models when applied to more complex datasets.

Since 2016, studies have increasingly adopted CNNs as a more efficient alternative, replacing manual extraction of acoustic features with the automatic learning of patterns directly from spectrograms. This shift provided greater flexibility and improved performance in vocalization classification. Among the early advances in this direction, notable contributions include the studies of Tóth and Czeba (2016) and Sprengel et al. (2016), which applied CNNs in the BirdCLEF challenge using large datasets. The study by Sprengel et al. (2016) stood out by achieving a mean average precision of 0.686, reinforcing the potential of these networks for automatic species identification.

In the following years, the adoption of CNNs became consolidated as the standard in the field, as demonstrated by studies such as Incze et al. (2018), Xie et al. (2019), LeBien et al. (2020), and Maegawa et al. (2021). These works incorporated advanced preprocessing and data augmentation techniques, enhancing model robustness. A notable highlight was the study by Maegawa et al. (2021), which achieved 97,0% accuracy in classifying vocalizations of *Accipiter gentilis*, demonstrating the applicability of CNNs in the conservation of threatened species.

Recent studies have explored variations of CNNs and alternative architectures, aiming to improve both model accuracy and computational efficiency. Three main approaches stand out in this context. Xie and Zhu (2022) proposed a hybrid model that integrates 2D and 3D CNNs, enabling the processing of scale-frequency maps to more effectively represent vocalizations in both temporal and spectral domains. García-Ordás et al. (2023) employed fully convolutional networks (FCNs), which allowed for the classification of vocalizations without the need to define a fixed input size, thus providing greater flexibility in acoustic analysis.

Mohanty et al. (2023) introduced the use of Spiking Neural Networks (SNNs), inspired by the functioning of biological neurons, achieving 94.0% accuracy with lower computational cost. Espejo et al. (2024) applied ANNs to short-term acoustic indices for monitoring urban and natural environments, highlighting the applicability of artificial intelligence techniques in environmental bioacoustics. Márquez-Rodríguez et al. (2025) proposed a multi-stage semi-supervised pipeline to recognize bird vocalizations in complex acoustic environments, using a YOLOv8-based detector applied to spectrograms, followed by classifiers fine-tuned with BirdNET embeddings. This approach achieved a significant improvement in identification accuracy across 34 vocalization classes.

He and Luo (2025) proposed an optimized version of the EfficientNet-B0 network for bird song recognition, integrating the Efficient Channel Attention (ECA) and Convolutional Block Attention Module (CBAM) mechanisms and achieving 96.04% accuracy across ten species. Their model outperformed the original by 3.2% while reducing network complexity by 16.4%, demonstrating the effectiveness of alternative architectures to conventional CNNs.

Bioacoustics has increasingly been used beyond the simple identification of species, with applications in environmental monitoring and citizen science. Recent studies indicate that the acoustic structure of a community can reflect its ecological diversity, supporting the assessment of ecosystem health and guiding conservation actions (Chhaya et al., 2021). Permana et al. (2022) explored the detection of bird vocalizations as an early warning system for forest fires, demonstrating how variations in acoustic patterns can indicate environmental stress. Clark et al. (2023), in turn, employed CNNs to analyze the impact of anthropogenic noise on the classification of vocalizations in biodiversity monitoring projects.

Moreover, the integrated use of bioacoustic techniques with tools such as machine learning and environmental DNA analysis has significantly expanded monitoring potential. An example is the study by Müller et al. (2023), conducted in tropical forests of Ecuador, which demonstrated that acoustic indices and CNN-based models were effective in tracking the regeneration of degraded areas. The combination of sound data with genetic analyses validated the results and reinforced the usefulness of bioacoustics as a reliable, accessible, and promising method for assessing the success of ecological restoration strategies.

Figure 1 provides a comparison between MFCCs combined with traditional classifiers and CNNs, illustrating the evolution of average accuracy of these techniques over the years and highlighting the significant improvement of CNNs after 2016.

Table 3 presents a comparison between these methodologies, highlighting their advantages and limitations.

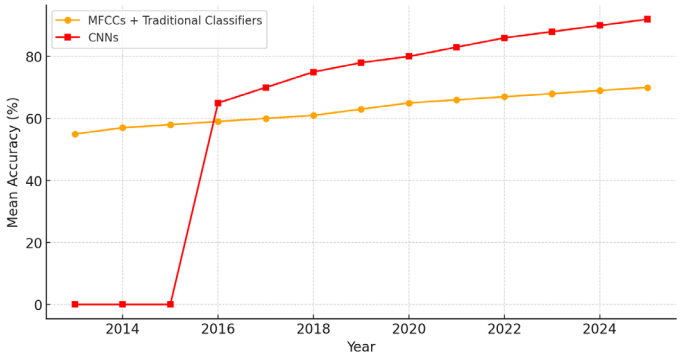

Figure 1 – Performance comparison between Mel-Frequency Cepstral Coefficients (MFCCs) and Convolutional Neural Networks (CNNs) (2013–2025).

Table 3 – Comparison between Mel-Frequency Cepstral Coefficients + Traditional Classifiers and Convolutional Neural Networks.

Characteristic	MFCCs + Traditional Classifiers	CNNs
Average accuracy	55–72% (variation between 2013–2025)	65–97% (from 2016 onward)
Feature extraction	Required prior to classification	Automatic learning from spectrograms
Preprocessing dependence	High (noise removal, manual segmentation)	Low (models can train on raw data)
Scalability	Limited to smaller datasets	High, suitable for large-scale datasets
Robustness to environmental noise	Low (sensitive to geophony and anthropophony)	High (able to learn patterns even under noisy conditions)
Computational cost	Low (can run on standard machines)	High (requires Graphics Processing Unit and advanced processing)
Real-time application	Limited	Possible, with optimizations

MFCCs: Mel-Frequency Cepstral Coefficients; CNNs: Convolutional Neural Networks.

Table 3 shows that, although methods based on MFCCs and traditional classifiers were widely used until 2015, the introduction of CNNs led to significant improvements in accuracy, scalability, and robustness to noise. However, deep learning-based models require greater computational power, which remains a limiting factor in applications that demand energy efficiency or deployment on embedded devices.

Despite these advances, several challenges still need to be addressed for automatic vocalization recognition systems to become more robust and widely applicable. One of the main challenges is model generalization, as most studies rely on controlled recordings that do not fully reflect real field conditions. Models need to be more adaptable to different habitats and acoustic variations in order to reduce errors caused by environmental noise. Another critical issue is the requirement for large annotated datasets, since neural network training depends on manually labeled data, making the process time-consuming and costly. Strategies such as semi-supervised learning and transfer learning may offer viable alternatives to reduce this dependency.

In addition, there are limitations related to computational efficiency and real-time applications, as deep neural networks require high processing power, which hinders their use on low-power devices such as field sensors. However, lighter and more energy-efficient models, such as SNNs and FCNs, have emerged as promising solutions. Finally, it is worth highlighting the importance of integration with other technologies, such as climate and air quality sensors, enabling more comprehensive analyses of environmental impacts on bird populations.

Conclusion

Technological advances in the field of automatic bird vocalization recognition have consolidated bioacoustics as an essential tool for biodiversity conservation and environmental monitoring. The transition from traditional methods, based on MFCCs and statistical classifiers, to modern approaches using CNNs has led to substantial improvements in accuracy, scalability, and robustness to environmental noise. However, challenges such as the need for large, labeled datasets and the high computational cost still represent barriers to the widespread application of these models.

The recent evolution of the field points to new solutions, such as self-supervised learning, which reduces dependence on manually annotated data, and hybrid models that combine CNNs and Recurrent Neural Networks (RNNs), enhancing the systems' generalization capacity. In addition, the development of more efficient models for mobile devices and real-time applications creates new opportunities for continuous and accessible wildlife monitoring.

In this context, it is evident that automatic vocalization recognition will continue to evolve, driven by advances in artificial intelligence and the growing need for effective conservation strategies. Future research should focus on the integration of more efficient techniques, the improvement of rare species identification, and the strengthening of citizen science, ensuring that these technologies are widely applicable and accessible across different environmental contexts.

Authors' Contributions

Pelanda, A.M.: conceptualization; data curation; formal analysis; investigation; methodology; software; visualization; writing – original draft. **Diógenes**, A.N.: supervision; validation; project administration; writing – review & editing.

References

Blake, J.G.; Loiselle, B.A., 2024. Sharp declines in observation and capture rates of Amazon birds in absence of human disturbance. Global Ecology and Conservation, v. 51, e02902. https://doi.org/10.1016/j.gecco.2024.e02902.

Braga, A.P.; Ferreira, A.C.P.L.; Ludermir, T.B., 2007. Redes neurais artificiais: teorias e aplicações. LTC, Rio de Janeiro, 226 p.

Celeghin, A.; Borriero, A.; Orsenigo, D.; Diano, M.; Méndez Guerrero, C.A.; Perotti, A.; Petri, G.; Tamietto, M., 2023. Convolutional neural networks for vision neuroscience: significance, developments, and outstanding issues. Frontiers in Computational Neuroscience, v. 17, 1153572. https://doi.org/10.3389/fncom.2023.1153572.

Chhaya, V.; Lahiri, S.; Jagan, M.A.; Krishnan, A., 2021. Community bioacoustics: studying acoustic community structure for ecological and conservation insights. Frontiers in Ecology and Evolution, v. 9, 706445. https://doi.org/10.3389/fevo.2021.706445.

Clark, M.L.; Salas, L.; Baligar, S.; Quinn, C.; Snyder, R.L.; Leland, D.; Schakwitz, W.; Goetz, S.J.; Newsam, S., 2023. The effect of soundscape composition on bird vocalization classification in a citizen science biodiversity monitoring project. Ecological Informatics, v. 75, 102065. https://doi.org/10.1016/j.ecoinf.2023.102065.

Cooke, R.; Sayol, F.; Andermann, T.; Blackburn, T.M.; Steinbauer, M.J.; Antonelli, A.; Faurby, S., 2023. Undiscovered bird extinctions obscure the true magnitude of human-driven extinction waves. Nature Communications, v. 14, 8116. https://doi.org/10.1038/s41467-023-43445-2.

Das, N.; Mondal, A.; Chaki, J.; Padhy, N.; Dey, N., 2020. Machine learning models for bird species recognition based on vocalization: a succinct review. Information Technology and Intelligent Transportation Systems. IOS Press, Amsterdam, p. 117-124. https://doi.org/10.3233/FAIA200052.

Dong, X.; Towsey, M.; Zhang, J.; Banks, J.; Roe, P., 2013. A novel representation of bioacoustic events for content-based search in field audio data. In: International Conference on Digital Image Computing: Techniques and Applications (DICTA), Hobart, pp. 1-6. https://doi.org/10.1109/DICTA.2013.6691473.

Dufour, O.; Artières, T.; Glotin, H.; Giraudet, P., 2014. Clusterized mel filter cepstral coefficients and support vector machines for bird song identification. In: Glotin, H. (Ed.), Proceedings of the First International Workshop on Machine Learning for Bioacoustics (ICML4B), joint to ICML 2013, Atlanta, GA, USA, p. 89-93. InTech. https://doi.org/10.5772/56872.

Eickhoff, S.; Rottschy, C.; Kujovic, M.; Palomero-Gallagher, N.; Zilles, K., 2008. Organizational principles of human visual cortex revealed by receptor mapping. Cerebral Cortex, v. 18 (11), 2637-2645. https://doi.org/10.1093/cercor/bhn024.

Espejo, D.; Vargas, V.; Viveros-Muñoz, R.; Labra, F.A.; Huijse, P.; Poblete, V., 2024. Short-time acoustic indices for monitoring urban-natural environments using artificial neural networks. Ecological Indicators, v. 160, 111775. https://doi.org/10.1016/j.ecolind.2024.111775.

García, D.; Rumeu, B.; Illera, J.C.; Miñarro, M.; Palomar, G.; González-Varo, J.P., 2024. Common birds combine pest control and seed dispersal in apple orchards through a hybrid interaction network. Agriculture, Ecosystems & Environment, v. 365, 108927. https://doi.org/10.1016/j.agee.2024.108927.

García-Ordás, M.T.; Rubio-Martín, S.; Benítez-Andrades, J.A.; Alaiz-Moretón, H.; García-Rodríguez, I., 2023. Multispecies bird sound recognition using a fully convolutional neural network. Applied Intelligence, v. 53 (20), 23287-23300. https://doi.org/10.1007/s10489-023-04704-3.

Giri, G.; Kc, I.; Khatiwada, P.; Adhikari, S.K.; Shakya, S., 2025. CNN-based bird sound detection: a comparative performance study. International Journal on Engineering Technology, v. 2 (2), 176-187. https://doi.org/10.3126/injet.v2i2.78615.

Goodfellow, I.; Bengio, Y.; Courville, A., 2016. Deep learning. MIT Press, Cambridge, MA (Accessed March 09, 2025) at:. https://www.deeplearningbook.org.

Han, X.; Peng, J., 2023. Bird sound classification based on ECOC-SVM. Applied Acoustics, v. 204, 109245. https://doi.org/10.1016/j.apacoust.2023.109245.

He, H.; Luo, H., 2025. An improved lightweight method based on EfficientNet for birdsong recognition. Scientific Reports, v. 15, 23727. https://doi.org/10.1038/s41598-025-07875-w.

Hidayat, A.A.; Cenggoro, T.W.; Pardamean, B., 2021. Convolutional neural networks for scops owl sound classification. Procedia Computer Science, v. 179, 81-87. https://doi.org/10.1016/j.procs.2020.12.010.

Hill, S.D.; Ji, W.; Parker, K.A.; Amiot, C.; Wells, S.J., 2013. A comparison of vocalisations between mainland tui (*Prosthemadera novaeseelandiae novaeseelandiae*) and Chatham Island tui (*P. n. chathamensis*). New Zealand Journal of Ecology, v. 37 (2), 214-223 (Accessed July 18, 2025) at:. https://newzealandecology.org/nzje/3085_/pdf.

Himmelberg, M.M.; Winawer, J.; Carrasco, M., 2022. Linking individual differences in human primary visual cortex to contrast sensitivity around the visual field. Nature Communications, v. 13 (1), 3309. https://doi.org/10.1038/s41467-022-31041-9.

Hong, T.Y.; Zabidi, M., 2021. Bird sound detection with convolutional neural networks using raw waveforms and spectrograms. International Symposium on Applied Science and Engineering, Erzurum, Turkey, 7-9 (Accessed March 07, 2025) at:. https://www.researchgate.net/publication/350725575_Bird_Sound_Detection_with_Convolutional_Neural_Networks_using_Raw_Waveforms_and_Spectrograms.

Incze, A.; Janczó, H.B.; Szilágyi, Z.A.; Farkas, A.; Sulyok, C., 2018. Bird sound recognition using a convolutional neural network. Proceedings of IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY), 295-300. https://doi.org/10.1109/SISY.2018.8524677.

Inoue, T.; Okura, Y.; Yoshida, T.; Washitani, I., 2025. Passive acoustic monitoring for assessing forest bird distribution and identifying conservationally important areas in a subtropical forest landscape. Ecological Research, v. 40 (4). https://doi.org/10.1111/1440-1703.12543.

Joly, A.; Champ, J.; Buisson, O., 2014. Instance-based bird species identification with undiscriminant features pruning – LifeCLEF 2014. CLEF Working Notes, LifeCLEF 2014 (Accessed July 07, 2025) at:. https://ceur-ws.org/Vol-1180/CLEF2014wn-Life-JolyEt2014b.pdf.

Kahl, S.; Wood, C.M.; Eibl, M.; Klinck, H., 2021. BirdNET: A deep learning solution for avian diversity monitoring. Ecological Informatics, v. 61, 101236. https://doi.org/10.1016/j.ecoinf.2021.101236.

Keck, F.; Peller, T.; Alther, R.; Barouillet, C.; Blackman, R.; Capo, E.; Chonova, T.; Couton, M.; Fehlinger, L.; Kirschner, D.; Knüsel, M.; Muneret, L.; Oester, R.; Tapolczai, K.; Zhang, H.; Altermatt, F., 2025. The global human impact on biodiversity. Nature, v. 641, 395-400. https://doi.org/10.1038/s41586-025-08752-2.

Kershenbaum, A.; Blumstein, D. T.; Roch, M. A.; Akçay, Ç.; Backus, G.; Bee, M. A.; Bohn, K.; Cao, Y.; Carter, G.; Cäsar, C.; Coen, M.; DeRuiter, S. L.; Doyle, L.; Edelman, S.; Ferrer-i-Cancho, R.; Freeberg, T. M.; Garland, E. C.; Gustison, M.; Harley, H. E.; Huetz, C.; Hughes, M.; Hyland Bruno, J.; Ilany, A.; Jin, D. Z.; Johnson, M.; Ju, C.; Karnowski, J.; Lohr, B.; Manser, M. B.; McCowan, B.; Mercado III, E.; Narins, P. M.; Piel, A.; Rice, M.; Salmi, R.; Sasahara, K.; Sayigh, L.; Shiu, Y.; Taylor, C.; Vallejo, E. E.; Waller, S.; Zamora-Gutierrez, V., 2014. Acoustic sequences in nonhuman animals: a tutorial review and prospectus. Biological Reviews, v. 91 (1), 13-52. https://doi.org/10.1111/brv.12160.

Koops, H.V.; Van Balen, J.; Wiering, F., 2014. A deep neural network approach to the LifeCLEF 2014 bird task. CLEF2014 Working Notes, v. 1180, 634-642 (Accessed July 18, 2025) at:. https://ceur-ws.org/Vol-1180/CLEF2014wn-Life-KoopsEt2014.pdf.

Krishna, B.; Kondle, P.; Vankdothu, R., 2024. Automated system for identifying bird species. African Journal of Biological Sciences, v. 6 (Si2), 367-385 (Accessed March 07, 2025) at:. https://www.afjbs.com/uploads/paper/3887c3d484b83ff1a68b53586f2fd925.pdf.

Lauha, P.; Somervuo, P.; Lehikoinen, P.; Geres, L.; Richter, T.; Seibold, S.; Ovaskainen, O., 2022. Domain-specific neural networks improve automated bird sound recognition already with small amount of local data. Methods in Ecology and Evolution, v. 13 (12), 2799-2810. https://doi.org/10.1111/2041-210X.14003.

LeBien, J.; Zhong, M.; Campos-Cerqueira, M.; Velev, J.P.; Dodhia, R.; Lavista Ferres, J.; Aide, T.M., 2020. A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network. Ecological Informatics, v. 59, 101113. https://doi.org/10.1016/j.ecoinf.2020.101113.

Lees, A.C.; Haskell, L.; Allinson, T.; Bezeng, S.B.; Burfield, I.J.; Renjifo, L.M.; Rosenberg, K.V.; Viswanathan, A.; Butchart, S.H.M., 2022. State of the World's Birds. Annual Review of Environment and Resources, v. 47, 231-260. https://doi.org/10.1146/annurev-environ-112420-014642.

Liu, J.; Zhang, Y.; Lv, D.; Lu, J.; Xie, S.; Zi, J.; Yin, Y.; Xu, H., 2022. Birdsong classification based on ensemble multi-scale convolutional neural network. Scientific Reports, v. 12, 8636. https://doi.org/10.1038/s41598-022-12121-8.

Maegawa, Y.; Ushigome, Y.; Suzuki, M.; Taguchi, K.; Kobayashi, K.; Haga, C.; Matsui, T., 2021. A new survey method using convolutional neural networks for automatic classification of bird calls. Ecological Informatics, v. 61, 101164. https://doi.org/10.1016/j.ecoinf.2020.101164.

Márquez-Rodríguez, A.; Rodríguez-Gómez, C.; León-Ortega, M.; Guzmán, J.; Baños-Guerrero, C., 2025. A bird song detector for improving bird identification through deep learning: a case study from Doñana. Ecological Informatics, v. 75, 103254. https://doi.org/10.1016/j.ecoinf.2025.103254.

Mascorro, G.A.M.; Torres, G.A., 2013. Reconocimiento de voz basado en MFCC, SBC y espectrogramas. Ingenius, (10), 12-20. ISSN: 1390-650X.

Maznikova, V.N.; Ormerod, S.J.; Gómez Serrano, M.Á., 2024. Birds as bioindicators of river pollution and beyond: specific and general lessons from an apex predator. Ecological Indicators, v. 158, 111366. https://doi.org/10.1016/j.ecolind.2023.111366.

McGinn, K.; Kahl, S.; Peery, M.Z.; Klinck, H.; Wood, C.M., 2023. Feature embeddings from the BirdNET algorithm provide insights into avian ecology. Ecological Informatics, v. 74, 101995. https://doi.org/10.1016/j.ecoinf.2023.101995.

Mohanty, R.; Bhuyan, H.K.; Pani, S.K.; Ravi, V.; Krichen, M., 2023. Bird species recognition using spiking neural network along with distance based fuzzy co-clustering. International Journal of Speech Technology, v. 26 (3), 681-694. https://doi.org/10.1007/s10772-023-10040-1.

Molina-Mora, I.; Ruiz-Gutiérrez, V.; Vega-Hidalgo, Á.; Sandoval, L., 2024. The utility of passive acoustic monitoring for using birds as indicators of sustainable agricultural management practices. Frontiers in Bird Science, v. 3. https://doi.org/10.3389/fbirs.2024.1386759.

Müller, J.; Mitesser, O.; Schaefer, H.M.; Seibold, S.; Busse, A.; Kriegel, P.; Rabl, D.; Gelis, R.; Arteaga, A.; Freile, J.; Leite, G.A.; De Melo, T.N.; LeBien, J.; Campos Cerqueira, M.; Blüthgen, N.; Tremlett, C.J.; Böttger, D.; Feldhaar, H.; Grella, N.; Falconí López, A.; Donoso, D.A.; Morinière, J.; Buřivalová, Z., 2023. Soundscapes and deep learning enable tracking biodiversity recovery in tropical forests. Nature Communications, v. 14, 6191. https://doi.org/10.1038/s41467-023-41693-w.

Pereira, G.H.A.; Centeno, J.A.S., 2017. Avaliação do tamanho de amostras de treinamento para redes neurais artificiais na classificação supervisionada de imagens utilizando dados espectrais e laser scanner. Boletim de Ciências Geodésicas, v. 23 (2), 268-283. https://doi.org/10.1590/S1982-21702017000200017.

Permana, S.D.H.; Saputra, G.; Arifitama, B.; Yaddarabullah; Caeserenda, W.; Rahim, R., 2022. Classification of bird sounds as an early warning method of forest fires using convolutional neural network (CNN) algorithm. Journal of King Saud University – Computer and Information Sciences, v. 34 (11), 4345-4357. https://doi.org/10.1016/j.jksuci.2021.04.013.

Piczak, K. J., 2016. Recognizing bird species in audio recordings using deep convolutional neural networks. In: Working Notes of CLEF 2016 – Conference and Labs of the Evaluation Forum, 534-543. CEUR-WS. (CEUR Workshop Proceedings, v. 1609), Aachen (Accessed July 18, 2025) at:. https://ceur-ws.org/Vol-1609/16090534.pdf.

Potamitis, I.; Ntalampiras, S.; Jahn, O.; Riede, K., 2014. Automatic bird sound detection in long real-field recordings: applications and tools. Applied Acoustics, v. 80, 1-9. https://doi.org/10.1016/j.apacoust.2014.01.001.

Priyadarshani, N.; Marsland, S.; Castro, I., 2018. The impact of environmental factors in birdsong acquisition using automated recorders. Ecology and Evolution, v. 8, 5016-5033. https://doi.org/10.1002/ece3.3889.

Priyadarshani, N.; Marsland, S.; Juodakis, J.; Castro, I.; Listanti, V., 2020. Wavelet filters for automated recognition of birdsong in long-time field recordings. Methods in Ecology and Evolution, v. 11 (3), 403-417. https://doi.org/10.1111/2041-210X.13357.

Qamar, R.; Zardari, B.A., 2023. Artificial neural networks: an overview. Mesopotamian Journal of Computer Science, v. 2023, 130-139. https://doi.org/10.58496/MJCSC/2023/015.

Ramirez, A.D.P.; De la Rosa Vargas, J.I.; Valdez, R.R.; Becerra, A., 2018. A comparative between Mel Frequency Cepstral Coefficients (MFCC) and Inverse Mel Frequency Cepstral Coefficients (IMFCC) features for an automatic bird species recognition system. IEEE Latin American Conference on Computational Intelligence (LA-CCI), Guadalajara, México, 1-4 (Accessed July 18, 2025) at:. https://ieeexplore.ieee.org/document/8625230.

Raschka, S.; Mirjalili, V., 2017. Python machine learning. Packt Publishing Ltd., Birmingham, 622 p.

Rivera, M.; Edwards, J.A.; Hauber, M.E.; Woolley, S.M.N., 2023. Machine learning and statistical classification of birdsong link vocal acoustic features with phylogeny. Scientific Reports, v. 13, 7076. https://doi.org/10.1038/s41598-023-33825-5.

Salamon, J.; Bello, J.P., 2017. Deep convolutional neural networks and data augmentation for environmental sound classification. IEEE Signal Processing Letters, v. 24 (3), 279-283. https://doi.org/10.1109/LSP.2017.2657381.

Schuster, G.E.; Walston, L.J.; Little, A.R., 2024. Evaluation of an autonomous acoustic surveying technique for grassland bird communities in Nebraska. PLOS ONE, v. 19 (7), e0306580. https://doi.org/10.1371/journal.pone.0306580.

Sick, H., 1984. Ornitologia Brasileira. Universidade de Brasília, Brasília, 827 p.

Somervuo, P.; Härmä, A.; Fagerlund, S., 2006. Parametric representations of bird sounds for automatic species recognition. IEEE Transactions on Audio, Speech and Language Processing, v. 14 (6), 2252-2263. https://doi.org/10.1109/TASL.2006.872624.

Somervuo, P.; Lauha, P.; Lokki, T., 2023. Effects of landscape and distance in automatic audio based bird species identification. Journal of the Acoustical Society of America, v. 154 (1), 245-254. https://doi.org/10.1121/10.0020153.

Sprengel, E.; Jaggi, M.; Kilcher, Y.; Hofmann, T., 2016. Audio based bird species identification using deep learning techniques. CLEF Working Notes, LifeCLEF 2016, v. 1609, 534-543 (Accessed July 18, 2025) at:. https://ceur-ws.org/Vol-1609/16090547.pdf.

Stastny, J.; Munk, M.; Juranek, L., 2018. Automatic bird species recognition based on birds vocalization. EURASIP Journal on Audio, Speech and Music Processing, v. 2018, art. 19, 1-19. https://doi.org/10.1186/s13636-018-0143-7.

Tóth, B.P.; Czeba, B., 2016. Convolutional neural networks for large-scale bird song classification in noisy environment. CLEF Working Notes, LifeCLEF 2016, v. 1609, 560-568 (Accessed July 18, 2025) at:. https://ceur-ws.org/Vol-1609/16090560.pdf.

Uddin, M.R.; Asaduzzaman, A.; Soza, R.; Minkler, C., 2024. Avian song identification using CNN. IEEE Green Technologies Conference (GreenTech), Springdale, AR, USA, 43-47. https://doi.org/10.1109/GreenTech58819.2024.10520499.

Verdin, R.; Kumar, A., 2015. Musical segmentation techniques for bird song classification. [S.l.] (Accessed March 09, 2025) at:. https://regisverdin.github.io.

Xie, J.; Hu, K.; Zhu, M.; Yu, J.; Zhu, Q., 2019. Investigation of different CNN-based models for improved bird sound classification. IEEE Access, v. 7, 175353-175361. https://doi.org/10.1109/ACCESS.2019.2957572.

Xie, J.; Li, W.; Zhang, J.; Ding, C., 2018. Bird species recognition method based on chirplet spectrogram feature and deep learning. Journal of Beijing Forestry University, v. 40 (3), 122-127. https://doi.org/10.13332/j.1000-1522.20180008.

Xie, J.; Towsey, M.; Eichinski, P.; Zhang, J.; Roe, P., 2015. Acoustic feature extraction using perceptual wavelet packet decomposition for frog call classification. IEEE 11th International Conference on e-Science, 237-242. https://doi.org/10.1109/eScience.2015.47.

Xie, J.; Yang, J.; Ding, C.; Li, W., 2020. High accuracy individual identification model of Crested Ibis (Nipponia nippon) based on autoencoder with self-attention. IEEE Access, v. 8, 41062-41070. https://doi.org/10.1109/ACCESS.2020.2973243.

Xie, J.; Zhong, Y.; Zhang, J.; Liu, S.; Ding, C.; Triantafyllopoulos, A., 2023. A review of automatic recognition technology for bird vocalizations in the deep learning era. Ecological Informatics, v. 73, 101927. https://doi.org/10.1016/j.ecoinf.2022.101927.

Xie, J.; Zhu, M., 2022. Sliding-window based scale-frequency map for bird sound classification using 2D- and 3D-CNN. Expert Systems with Applications, v. 207, 118054. https://doi.org/10.1016/j.eswa.2022.118054.

Zhang, Q.; Hu, S.; Tang, L.; Deng, R.; Yang, C.; Zhou, G.; Chen, A., 2024. SDFIE-NET – A self-learning dual-feature fusion information capture expression method for birdsong recognition. Applied Acoustics, v. 221, 110004. https://doi.org/10.1016/j.apacoust.2024.110004.