University of South-Eastern Norway
Faculty of Technology, Natural Sciences and Maritime Sciences
–

Master´s Thesis
Study programme:
Spring 2024

Emilio Antonio Joachín Godínez

A comparative analysis of Merlin and BirdNet applications for accurate bird species identification through passive acoustic monitoring

University of South-Eastern Norway
Faculty of Technology, Natural Sciences and Maritime Sciences …
Institute of Natural Sciences and Environmental Health.
PO Box 235
NO-3603 Kongsberg, Norway

http://www.usn.no

Abstract

In recent years, passive acoustic monitoring (PAM) has emerged as a powerful tool for monitoring avian biodiversity. However, a major challenge has been to develop algorithms that can process large amounts of data, and at the same time, correctly identify bird sounds to species. There are several algorithms available for the identification of bird sounds but, the most prominent ones are Merlin Audio ID and BirdNet. In this study, a comparative analysis of the Merlin and BirdNet applications was carried out to evaluate their accuracy and efficiency in the identification of bird species from PAM recordings. A human-trained ear was used as a baseline to evaluate the accuracy of these algorithms. The number of false negatives i.e. bird species detected by the human ear but not by the app and false positives i.e. bird species detected by the apps but not by the human ear were estimated. Merlin correctly identified 39.8% of the bird species identified by the human-trained ear. 44.2% were false negatives and 16% were false positives. BirdNet on the other hand, correctly identified 24.6% of the bird species identified by the trained- human ear. 70.6% were false negatives and 3% were false positives. A significative difference ($P < 0.05$) in the number of birds detected between Merlin, BirdNet, and the human-trained ear. Finally, inconsistency in bird detection of the Merlin application, when analyzing the same recordings for a second time, was found ($P = 0.006$). Although both applications show promising results, they still need many improvements for optimal performance.

Preface

In the realm of ecological research and conservation efforts, advancements in technology have revolutionized the methods available for monitoring and studying wildlife populations. One such advancement is the utilization of passive acoustic monitoring (PAM) techniques for the identification of and study of avian species. This master's thesis explores and compares two prominent applications, Merlin and BirdNet, for their efficacy in accurately identifying bird species through PAM.

The inspiration behind this study stems from the pressing need to enhance our understanding of bird populations and their dynamics, particularly in the face of environmental challenges such as habitat loss and climate change. Accurate bird species identification is crucial for ecological research, conservation planning, and policy-making. Traditional methods of bird identification, like counting points and line transects, are often labor-intensive, time-consuming, and limited in their scope. In contrast, PAM offers a non-invasive, cost-effective, and scalable alternative for monitoring avian biodiversity.

Merlin, a citizen science identification app, and BirdNet, a deep learning-based app were developed by the Cornell Lab of Ornithology. These represent two cutting-edge approaches to automated bird species identification using sound recordings. Both applications have gained considerable attention and adoption within the scientific community and citizen science initiatives. However, a comprehensive analysis of their performance, strengths, and limitations in real-world scenarios are currently lacking.

This thesis aims to address this gap by systematically evaluating Merlin and BirdNet. The findings of this study are expected to provide valuable insights into the capabilities and potential applications of these technologies in avian research and conservation.

The research in this thesis would not have been possible without the support and help of my supervisor Ph.D. Øyvind Steifetten, my family Patricia Joachín, Otoniel Joachín, and Fidelina Godínez, my colleague's Ph.D. José Soto, M.Sc. Claire Dallies Nusli and MPD Estuardo Girón, and my friends M.Sc. Irene García Cuesta, M.Sc. Malena Díaz, M.Sc. Eduard Codó, M.Sc. Ana Barrios, M.Sc. Camilo Bocanegra, M.Sc. Rachel Carboni, M.Sc. Adriana Sória Peris, M.Sc. Zuzana Maciejewska and B.S. Andrid Ramírez. I'm grateful for their encouragement, expertise, and unwavering assistance through this journey. Ultimately, I hope that this thesis serves as a resource for researchers, conservation practitioners, and policymakers seeking to leverage technology for the betterment of avian biodiversity monitoring and conservation efforts.

Bø

Emilio A. Joachín G.

Introduction

In recent years, the decline in bird populations worldwide has raised significant concerns among environmental scientists, conservationists, and policymakers, and it is estimated that 49% of the world's bird species are in decline or in serious danger of extinction (Less *et al.* 2022). For example, in the United States, 29% of the bird population has decreased since 1970, with grassland birds being the most affected (Li *et al*. 2020), and in Europe, 74% of ground-nesting bird species are in serious decline (MacMahon *et al*. 2020).

The main cause of these declines is linked to the main drivers of biodiversity loss, such as climate change, pollution, habitat loss and destruction, invasive species, and overexploitation (Less *et al*. 2022; Gregory *et al*. 2022). As we witness a rapid decline in bird populations, the need for effective monitoring methods to estimate population size and distribution becomes increasingly critical in order to develop effective conservation strategies (Leclère *et al*. 2020; Mathez *et al.* 2020).

There are currently several methods for estimating the richness and abundance of birds in various ecosystems, but the most widely used methods are point counts and line transects. These involve on-site visual and/or auditory observations, and this allow researchers to directly assess bird presence and abundance (Gutiérrez *et al*. 2020) As such, they are considered benchmarks in avian ecological studies (Darras *et al.* 2019). Although these methods provide satisfactory results, they are highly cost-inefficient, time-consuming, and local in scale. This makes them ineffective if we consider the accelerated rate of bird diversity loss (Bowler *et al*. 2019). With an even-accelerating problem, there is a need for methods that are inexpensive, that can collect data most of the year, and that can be analyzed quickly (Priyadarshani *et al*. 2018; Xu *et al*. 2018; Caro *et al*. 2022).

During the last 25 years, monitoring with acoustic loggers has been promoted. Bioacoustics has evolved through interdisciplinary collaboration between biologists, physicists, engineers, and computer scientists, advancing our understanding of acoustic phenomena in natural and anthropogenic environments (Green, 1995; Bowler *et al.* 2019; González *et al.* 2019).

Passive acoustic monitoring (PAM) has emerged as a promising technique for studying bird populations, offering a non-intrusive and efficient approach to assessing avian diversity (Green, 1995; Bowler *et al*. 2019; Xie *et al*. 2022). PAM relies on the recording and analysis of sounds emitted by birds, allowing for continuous monitoring in various habitats. PAM can be used for different purposes like studying species and populations, behaviour, community structure, and biodiversity (Pijanowski *et al*. 2011; Gutierrez *et al.* 2020). This method has proven to be particularly valuable in areas where visual observation may be challenging, such as dense forests and remote regions. Also, the implementation of bioacoustics in the study of birds is feasible because vocalizations are specific and unique to each species. Thus, it is possible to identify birds by sound alone (Lapp *et al*. 2023; Blake, 2021).

PAM has had an accelerated growth and development; it involves deploying audio recording devices in the field, and capturing the ambient soundscape over extended periods (Darras *et al.* 2017; González *et al*. 2020; Pérez *et al*. 2020). The collected data is then processed using advanced algorithms to identify and analyze the acoustic signatures of different bird species (Hill *et al*. 2018; Van *et al*. 2023). There are several applications for identifying bird sounds. But the most popular used are Merlin

Audio ID tool and BirdNet (Priyadarshini *et al*. 2018; Kahl *et al*. 2020; Brooker *et al*. 2020; Ware *et al*. 2023). PAM offers several advantages, including the ability to monitor birds in challenging terrains, conduct continuous observations, and cover larger geographical areas compared to traditional methods (Hutto & Stutzman, 2009; Hill *et al*. 2019; Smith *et al*. 2021).

However, like any scientific approach, PAM comes with its set of disadvantages that require careful consideration (Xu *et al*. 2018; Priyadarshani *et al*. 2018, Caro *et al.* 2022). The identification of bird species solely based on acoustic signals can be intricate due to variations in vocalizations, overlapping sounds, and the influence of environmental factors (Hill *et al*. 2019; Van *et al.* 2023). Additionally, distinguishing between similar-sounding species poses a considerable challenge. To address these limitations, it is crucial to explore and compare alternative methods, such as traditional bird counting points and line transects (Sueur *et al*. 2008; González *et al*. 2020), which could provide a valuable context for evaluating the accuracy and reliability of PAM techniques (Thompson *et al.* 2019).

The present study undertakes a comprehensive analysis of two prominent applications, namely Merlin Audio ID tool and BirdNet, with a specific focus on their efficacy and accuracy through the utilization of PAM. These applications were chosen because they are both at the forefront of automated bird recognition and because they are the two most popular among researchers and non-researchers (Andrejeff & Lehikoinen, 2023). In addition, both applications are designed for different purposes. BirdNet is designed for research purposes and can process large amounts of data. While, the Merlin Audio ID tool is designed with a citizen science approach, but cannot process large quantitatives of data (Denton *et al*. 2022; Andrejeff & Lehikoinen, 2023).

The overarching goal of this study is to contribute to the refinement of monitoring strategies that can effectively address the challenges posed by bird population decline.

Material and methods

Study area

This study was conducted in Midt-Telemark municipality, Norway, during the period 2023 25th of May to the 11th of June, 2023. The Telemark municipality is situated in the southern part of Norway. This area has varied nature with mountains, large lakes, rivers, deciduous and coniferous forests, agricultural landscapes, urban settlements, and others. Which results in a highly diverse avian fauna.

To obtain data for the comparison of the Merlin and BirdNet applications, 10 acoustic loggers were deployed in different habitats: three loggers were placed in a pine forest, one in a spruce forest patch, one in a deciduous forest, four in open agricultural landscape and one in a garden landscape. These different habitats were chosen for their climatic and physical properties to see if they could affect the performance of the loggers.

Field Procedures

AudioMoth version 1.2.0 recorders encased in a waterproof housing (IPX7) for the collection of bird songs were used. These devices are low-cost, broad-spectrum loggers that can be used as programmed recorders, i.e. the recording schedule and frequency of interest can be configured. They are optimized to record sounds within 200 to more than 800 meters away depending on the sound source (Piña-Covarrubias *et al*. 2019).

They use unidirectional microphones. In addition, the battery level and available storage space can be checked (Hill *et al*. 2019). Also, these recorders produce files in WAV format, which is an uncompressed file that allows all audio layers to be analyzed. Another feature is that if the device shows faults, anyone with knowledge of electronics could be able to repair it, which makes them versatile (Blake *et al*. 2021).

The factory settings of the recorders were used for this study (e.g. sample rate frequencies, gain level, date, and time), and only the recording schedule was modified. Once the loggers were configured, they were inserted into housings, which included a velcro strap for easier and safer installation. These were placed at a height of approximately 200 meters to ensure that the loggers captured as much audio as possible. They were set to record 24 hours for 2 weeks. Sampling was conducted 24 hours a day to detect nocturnal species. The time designated for recording was from the 25th to the 31st of May to the 5th to the 11th of June.

At the end of the first week of sampling, all recorders were checked to verify the condition of the recorders, the amount of battery remaining and to change the SD memory cards. All audio files were transferred from the SD memory cards to a bank.

Acoustic analysis

The collected audios were used to identify bird species with the Merlin and BirdNet applications. The original intention was to use BirdNet to analyze all the data collected, but preliminary analysis showed a very low detection rate for this application. It was decided to use the audio identification tool of Merlin developed by the Cornell Lab of Ornithology instead.

To identify bird songs to species, a Samsung A52s cell phone with updated versions of the Merlin application was used. The cell phone microphone was positioned next to the computer speaker for optimal recording. From a total of 42 days (1008 hours) of recording during the 14-day period, 6 days (144 hours or 8,640 1-minute files) were randomly selected for analysis.

Comparison of Merlin and Birdnet

Out of the 144 hours analyzed by Merlin, two hours (120 1-minute files) were randomly selected for further comparisons between the two applications. The two hours were analyzed by both Merlin and BirdNet to see which of the two applications would detect the most species. In addition, these results were compared with the number of species identified by an expert in bird song.

The objective was to verify the false negatives (bird species detected by human hearing that the applications failed to detect) and false positives (bird species detected by the applications, but not detected by the human ear) of both applications. The number of identified birds by the expert was designated as the baseline when comparing the results from the applications.

Consistency of Merlin in bird detection and identification

Out of the 144 hours analyzed by Merlin, another 2 hours (120 1-minute audio files) were randomly selected in order to demonstrate discrepancies in identification by Merlin application. It means, comparing a second round of identification with the original round of identification to see if the results match.

The audios corresponding to the original listening were ordered chronologically. That is, in July and August the audios corresponding to logger 8 were analyzed. In September and October, the audios from logger 9 were analyzed and finally, in November and December, the audios corresponding to logger 10 were analyzed. On the other hand, the detections corresponding to the replicate were analyzed in December.

The objective was to determine if Merlin could identify the same species in each round of identification. This information is important to verify the accuracy of the Merlin application for bird song detection and identification.

PAM vs field observations

To compare the effectiveness of passive acoustic monitoring and simulated counting points, the number of bird species detected by the Audiomoths during 24 hours and at the 5- and 10-minute intervals at the simulated count points was compared. These times were chosen for the intervals because they are the standard times used in the field for point counts (Roberts *et al*. 2018; Johnson *et al*. 2021).

These intervals correspond to a defined schedule from 4 am to 10 am. For the comparison, two random 5-minute intervals were chosen for each logger. From these selected intervals, an additional 5 minutes were used to obtain the 10-minute intervals. For the 24 hours, the number of species detected on the day sampled was considered.

Statistical analysis

R and Excel software were used for statistical analysis, data storage, and plotting (R Coreteam, 2024; Microsoft Corporation, 2024). In R, the packages “dunn.test" and “ggplot2” were used (Dinno, 2017; Kassambara 2020). To compare the BirdNet and Merlin applications against the human ear, a Kruskal-Wallis rank sum test was performed. Then, a pairwise comparison using the Wilcoxon sum rank test was performed. The nonparametric test was chosen because the data did not follow a normal distribution. In addition, to analyze the false negatives and false positives, a bar graph was made to show the percentages descriptively.

To compare the consistency of Merlin to detect birds by vocalization, a Welch two-sample T-test was performed. In the case of the comparison of the counting points with the 24 and 48 hours of recording, a Wilcoxon signed-rank test was performed. The nonparametric test was chosen because the data did not follow a normal distribution.

Results

Unfortunately, only three loggers of 10 were able to produce audible bird recordings. For the other seven loggers, six of them produced very faint sounds but with a massive static background noise, and one didn´t record any sounds at all. Of the three functional loggers, one was situated in deciduous forest (logger 8), One in garden landscape (logger 9), and one in open agricultural landscape (logger 10). In total for all three loggers, 59 bird species were detected. With 39 species detected by logger 8, 41 species by logger 9, and 36 species by logger 10. For a complete list of the recorded species see Table 1.

Table 1. Bird species detected by Merlin in logger 8 (deciduous forest), logger 9 (open agricultural landscape), and logger 10 (garden landscape). Likely refers to common bird species and unlikely to uncommon species in the habitat.

Name	**Logger 8**	**Logger 9**	**Logger 10**	**Name**	**Logger 8**	**Logger 9**	**Logger 10**
Likely (1), Unlikely (2)				Willow Tit (*Poecile montanus*) 1	x	x	x
Common Gull (*Larus canus*) 1		x		Eurasian Blue Tit (Cyanistes caeruleus) 1	x	x	x
Eurasian Oystercatcher (*Haematopus ostralegus*) 2		x		Great Tit (*Parus major*) 1	x	x	x
Eurasian Woodcock (*Scolopax rusticola*) 1			x	Eurasian Nuthatch (Sitta europaea) 1	x	x	x
Common Sandpiper (*Actitis hypoleucos*) 2			x	Eurasian Wren (*Troglodytes troglodytes*) 1	x	x	x
Common Greenshank (*Tringa nebularia*) 2			x	Icterine Warbler (*Hippolais icterina*) 1	x		
Common Wood-Pigeon (*Columba palumbus*) 1	x	x		Sedge Warbler (*Acrocephalus schoenobaenus*) 2			x
Osprey (*Pandion haliaetus*) 2		x		Common Reed Warbler (*Acrocephalus scirpaceus*) 2			x
Western Marsh Harrier (*Circus aeruginosus*) 2		x	x	Eurasian Blackcap (*Sylvia atricapilla*) 1	x	x	x
Eurasian Sparrowhawk (*Accipiter nisus*) 1	x			Garden Warbler (*Sylvia borin*) 1	x	x	x
Tawny Owl (*Strix aluco*) 1	x	x		Lesser Whitethroat (*Curruca curruca*) 1	x		
Common Swift (*Apus apus*) 1	x	x		Greater Whitethroat (*Curruca comunis*) 1	x		
Barn Swallow (*Hirundo rustica*) 1	x	x		Willow Warbler (*Phylloscopus trochilus*) 1	x	x	x
Common Cuckoo (*Cuculus canorus*) 1	x			Common Chiffchaff (*Phylloscopus collybita*) 1	x	x	x
Common Kingfisher (*Alcedo atthis*) 2		x	x	Goldcrest (*Regulus regulus*) 1	x	x	x
Eurasian Wryneck (*Jynx torquilla*) 1	x		x	Gray Wagtail (*Motacilla cinerea*) 1	x		
Great Spotted Woodpecker (Dendrocopos major) 1		x	x	White Wagtail (*Motacilla alba*) 1		x	x
Eurasian Jay (*Garrulus glandarius*) 1	x			Tree Pipit (*Anthus trivialis*) 1		x	
Eurasian Magpie (*Pica pica*) 1	x			Eurasian Skylark (*Alauda arvensis*) 1	x		x
Eurasian Nutcracker (*Nucifraga caryocatactes*) 1		x		Yellowhammer (Emberiza citrinella) 1	x	x	x
Hooded Crow (*Corvus cornix*) 1		x		Reed Bunting (*Emberiza schoeniclus*) 1			x
Spotted Flycatcher (*Muscicapa striata*) 1	x	x		House Sparrow (*Passer domesticus*) 1	x	x	
European Robin (*Erithacus rubecula*) 1	x	x	x	Eurasian Tree Sparrow (*Passer montanus*) 1		x	
Common Redstart (Phoenicurus phoenicurus) 2	x	x	x	Common Chaffinch (Fringilla coelebs) 1	x	x	x
Song Thrush (*Turdus philomelos*) 1	x	x	x	Hawfinch (*Coccothraustes coccothraustes) 1*		x	x
Redwing (*Turdus iliacus*) 1	x	x	x	Eurasian Bullfinch (*Pyrrhula pyrrhula*) 1	x	x	x
Eurasian Blackbird (*Turdus merula*) 1	x	x	x	European Greenfinch (*Chloris chloris*) 1	x	x	x
Fieldfare (*Turdus pilaris*) 1	x	x	x	European Goldfinch (*Carduelis carduelis*) 1	x	x	x
Dunnock (*Prunella modularis*) 1		x	x	Eurasian Siskin (*Spinus spinus*) 1	x	x	x
Marsh Tit (*Poecile palustris*) 1	x	x	x	European Starling (*Sturnus vulgaris*) 1			x

Comparison of Merlin and BirdNet

There was a significant difference in the number of bird species detected between BirdNet, Merlin, and human-trained ear (Kruskal-Wallis, n=753, $p < 0.05$) also applying for every comparative combination (Figure 1).

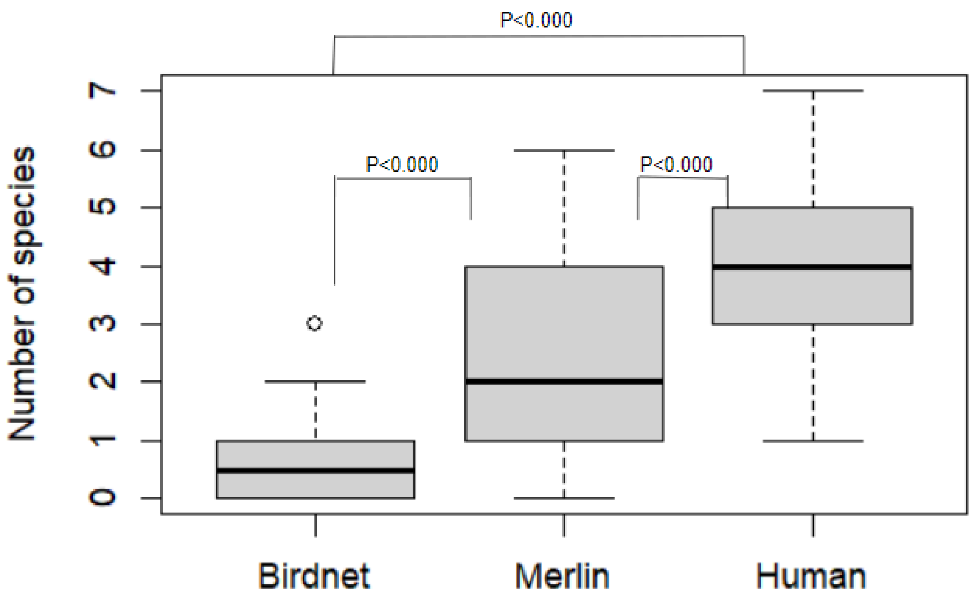

Figure 1. Box plot showing the number of bird species detected by BirdNet, Merlin, and human-trained ear.

Based on 753 observations Merlin and BirdNet correctly identified 39.8% (230) and 26.4% (151), respectively, of the bird species detected by an expert birder (Figure 2). However, compared to a human-trained ear, both applications showed a large number of false negatives. Merlin and BirdNet were unable to detect 44.2% (253) and 70.6% (405) of the baseline observations, respectively, indicating that both applications have a high potential for further improvement in their ability to detect bird species by vocalization (Figure 2). Regarding false positives, both applications did reasonably well, with Merlin overestimating the number of species by 16% (90), while BirdNet was least likely to record species that were not there (3% (17)) (Figure 2).

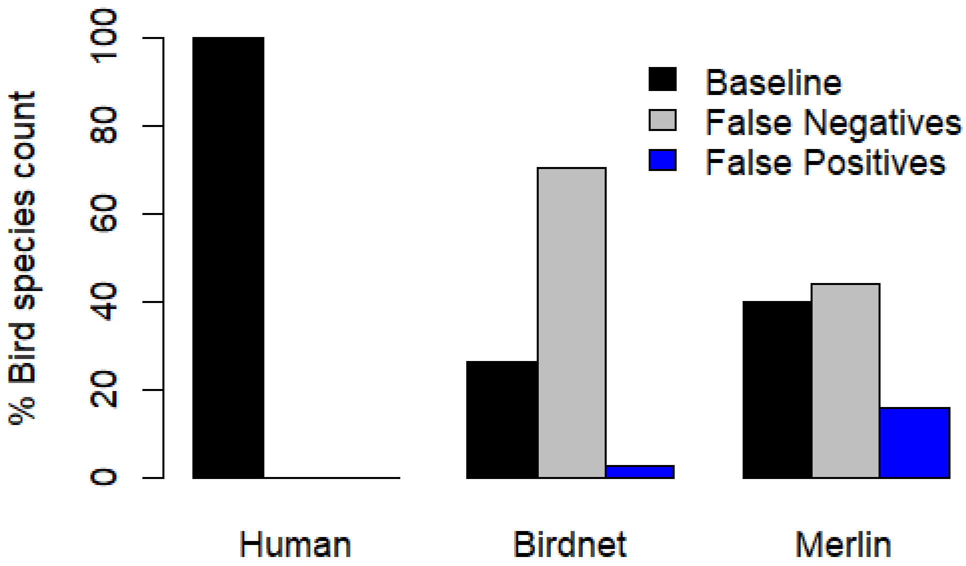

Figure 2. Bar plot showing the percentage of baseline observations, false negatives, and false positives by BirdNet and Merlin. The baseline was verified by human-trained ear.

Merlin Bird detection consistency

For the Merlin application a significant inconsistency in the number of bird species detected between the two observation trials was found (Welch 2 sample t-test, n= 24, p=0.006) (Figure 3). The first observation trial detected 22 bird species in 151 observations. More amount of birds than the second trial (16 bird species in 116 observations), indicating that Merlin is unable to detect and identify the same species on consecutive trials.

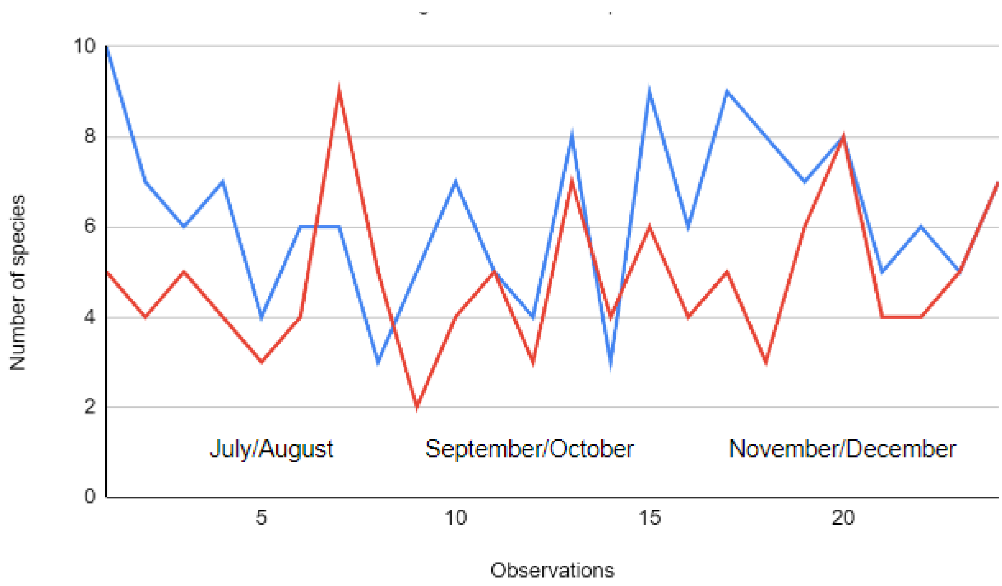

Figure 3. Graph showing the number of bird species detected by Merlin during two consecutive observation trials. Blue line: observation trial from July to November and Red line: observation trial from December.

PAM vs the point count method

The number of birds detected during a 5 and 10-minute period, which is usually the time spent on a single counting point by a field observer, was for all three loggers substantially lower when comparing it to a full 24-hour recording period by PAM, and even more so for a 48-hour recording period (Figure 4 and 5). Although the 5 and 10-minute periods are not based on field observations *per se*, and might thus have underestimated the real number of species, the use of PAM shows great potential in assisting, or even replacing, manual bird counts. There was a significant difference between counting points with a 5-minute and 24-hour recording interval (Wilcoxon signed rank test, N=6, p= 0.03). Likewise, between counting points with 10 minutes and 24-hour recording intervals (p= 0.03). Also, the same was observed between 5 and 10-minute counting points and 48-hour recording interval.

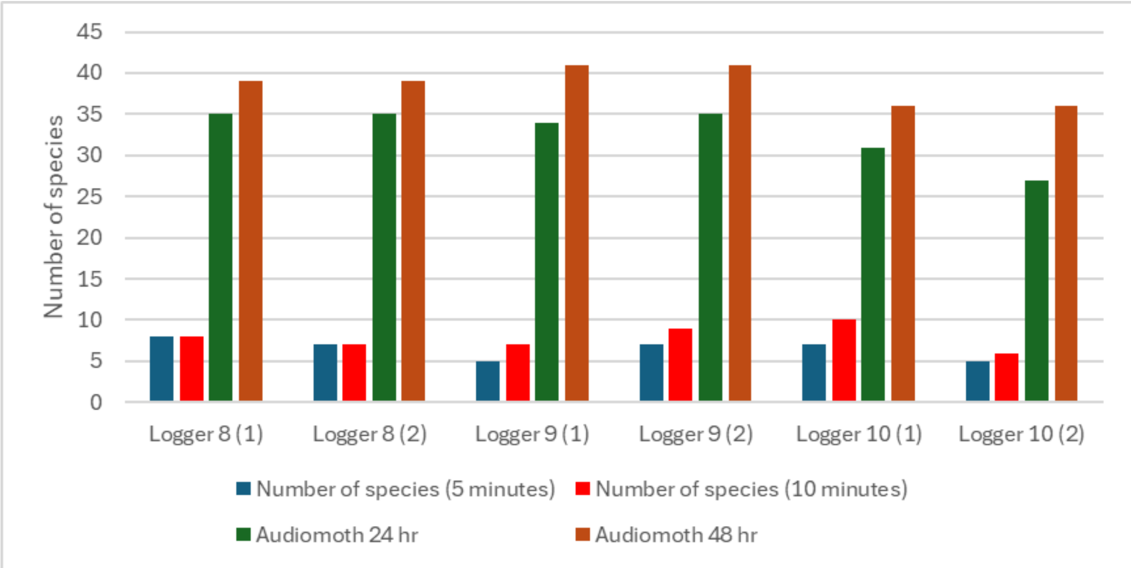

Figure 4. Bar plot showing the number of bird species detected within a 5 and 10-minute period compared to a full 24 and 48-hour period recorded by PAM.

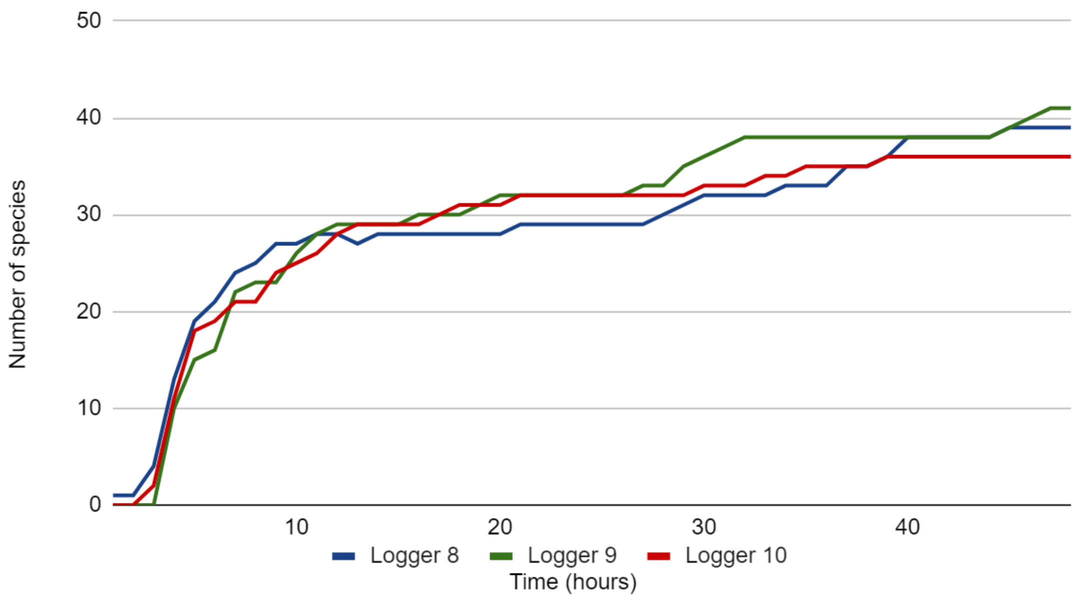

Figure 5. Accumulative detection curve for new bird species for each of the three loggers during a 48-hour period.

Discussion

The results indicate that the Merlin application demonstrates promising capabilities in identifying bird species from acoustic recordings. However, there were variations in its performance, with certain factors influencing their accuracy and effectiveness. Merlin exhibited higher accuracy in identifying common and well-documented species and multiple vocalizations in the landscape, likely due to its extensive database and machine learning algorithms trained on a wide range of bird vocalizations. In addition, it was able to detect a greater number of species compared to BirdNet. However, none of these applications were as effective compared to the baseline created. In addition, PAM is an effective monitoring method, allowing the detection of 59 bird species.

Merlin app shows better performance, with 39.8% closer to the baseline reference created. BirdNet, on the other hand, matched this same baseline reference by 26.4%. The remaining discrepancies are explained by false positives and false negatives. However, despite their capabilities, both applications displayed a significant number of false negatives when compared to human expertise. It was observed that Merlin missed detecting 44.2% (253) of the baseline observations, while BirdNet missed 70.6% (405), indicating substantial room for improvement in both applications' ability to accurately detect bird species solely based on vocalizations. This suggests that while these applications offer promising automated identification tools, they still lack the nuanced discernment of a trained human ear.

Furthermore, in terms of false positives, both applications performed reasonably well. Merlin tended to overestimate the number of species by 16% (90), while BirdNet was less prone to recording species that were not present, with only 3%

(17) false positives. These results imply that while false positives were relatively low, false negatives remained a significant challenge for both Merlin and BirdNet.

Overall, the findings underscore the potential of automated bird species identification applications like Merlin and BirdNet but also highlight the importance of ongoing refinement and improvement to enhance their accuracy and reliability, particularly in minimizing false negatives. Applications like Merlin and Birdnet are currently the most popular, so continued development and validation of these applications, in conjunction with human expertise, are crucial for their effective utilization in avian biodiversity monitoring and conservation efforts.

A significant inconsistency in the number of bird species detected by the Merlin application was found. Such inconsistencies raise important considerations regarding the reliability and robustness of automated bird identification applications like Merlin. While the application may demonstrate promising capabilities in identifying bird species from acoustic recordings under certain conditions, the observed variability underscores the need for caution when interpreting results and relying solely on automated tools for avian biodiversity monitoring.

Possible factors contributing to this inconsistency could include variations in the algorithm itself, background noise levels, or the presence of competing sounds, which may impact the accuracy of Merlin's identification algorithms. Additionally, differences in recording quality or device positioning between trials could also influence the application's performance.

It should be mentioned that although the algorithms of both applications are practically the same, the applications are designed for different groups and purposes. Merlin was designed for citizen science and short observations. On the other hand, Birdnet was designed with a scientific focus and to analyze large amounts of data (Andrejeff & Lehikoinen, 2023). However, the way the two applications work is different. BirdNet uses a library of vocalizations of ~9,000 bird species, which are mainly distributed in North America and Europe. It works best with individual species recordings that do not have too much background noise pollution. If soundscape recordings are used, the detection accuracy decreases from 79% to 41% (Kahl *et al*. 2021).

Merlin, on the other hand, works with bird region species packages from around the world that include vocalizations, distribution, and migration information as mentioned above. This performance could explain the discrepancies in bird detection. It is possible that Merlin´s bird species packs may remove species depending on the time of year the app is used. Thus, birds common in summer, for migration reasons, would no longer be detected in winter. In addition, some bird species may be removed due to updates in the official taxonomic lists and updates in the application. These factors would generate discrepancies in the detection of birds. At the moment, the only current way to avoid this is to analyze the audios once they are compiled (Rahaman, 2023).

In terms of effectiveness, a total of 59 species were identified for the 3 loggers. 38 species were identified for logger 8, 41 for logger 9, and 36 for logger 10. This makes PAM an effective tool with potential for avifauna monitoring. It allows us to record more species in continuous monitoring. However, in the case of this study, the total number of species identified could be lower due to the difficulties and discrepancies in detections of Merlin.

Also, it should be noted that PAM is an efficient method that allows us to identify more species compared to methods such as point counts in which very few species are detected. It is an innovative methodology due to the cost-efficient approach, continuous monitoring, remote operation, non-intrusive, detection of rare/elusive spp., and long-term studies (Priyadarshani *et al.* 2018; Hoefer *et al*. 2023). In the case of birds, it should be remembered that they stop their activity when they encounter disturbances in their environment such as noise, movement, and odors (Stuchbury & Morton, 2022). Which can lead to a loss of data in invasive methods such as point counts (Hoefer *et al*. 2023). Combining PAM with point counts and/or line transects is currently the most common way to monitor birds (Darras *et al*. 2019; Blake, 2021).

While the shorter observation periods may not capture the full bird activity in a given area, they provide insights into the potential limitations of traditional manual bird counts. Despite the likelihood of underestimating the true number of species present, the results underscore the considerable potential of PAM in complementing and potentially even substituting manual bird counting methodologies. Statistical analyses revealed significant differences between the number of birds detected at counting points with 5-minute intervals and those observed during 24-hour recording sessions. Similarly, significant differences were noted between counting points with 10-minute intervals and 24-hour recording periods. These findings further highlight the efficacy of PAM in providing comprehensive avian biodiversity data over extended durations, surpassing the limitations inherent in traditional field observations with brief timeframes. The results support the notion that integrating PAM into avian monitoring protocols could enhance the efficiency and effectiveness of biodiversity assessments, contributing to more robust conservation and management strategies.

There were failures with the loggers and the audio files. That is to say, the recordings presented a lot of continuous static noise. Although sounds from the environment could be perceived, they were very distant, almost indistinguishable, and infrequent. This specific fault may be associated with memory, configuration, the position where the device was placed (direction of the microphone, or loggers failures (Rogers, 2024). Although tests were performed to determine the source of this fault, no particular cause was found. It is important to take these observations into account to implement improvements in future versions of audiomoths. This will allow us to have better recorders with the benefits already mentioned.

Another apparent cause is the battery level. It seems that when the battery level drops too low, the performance of the Audiomoth is greatly affected and the quality of the audio files is greatly diminished. It is therefore recommended to check the battery levels regularly to ensure optimal performance of the Audiomoth (Lapp *et al*. 2023). In addition, there may be failures related to memory cards. Please note that not all memory cards are compatible with these recorders. In addition, sudden movements or improper removal of the memory cards may cause the files to become corrupted, resulting in loss of data. To minimize these issues, it's advisable to use high-quality, reputable brands, regularly format the cards, and handle them carefully to prevent damage or corruption. Additionally, maintaining backup SD cards during field deployments can serve as a precautionary measure against any card-related complications (Lapp *et al*. 2023; Hill *et al*. 2018; Hill *et al*. 2019).

These applications offer cost-effective and accessible tools for monitoring avian biodiversity, especially in remote or inaccessible areas where counting points and transect surveys may be challenging to conduct. Additionally, their user-friendly interfaces make them suitable for contributing to large-scale data collection efforts. Furthermore, the ability of these applications to rapidly identify bird species can facilitate timely conservation interventions and habitat management strategies. This approach might result in significant economic and time savings for stakeholders.

Also, data analysis can provide better and more accurate information because it can better describe the dynamics of bird populations when taken over a long period of time and the method itself is standardized, which means that there is no need to rely on the expertise of the surveyors (Hill *et al*. 2018). It should be considered that these savings may vary depending on the equipment to be used, the complexity of the ecosystem studied, and the costs of the area where it is decided to implement the research (Hutto *et al*. 2009).

Also, the analysis of audio-related data is often complicated by two aspects, the difficulty of storing a large amount of data and the complexity of the files. In the case of the study of birds, another difficulty is added, which consists of identifying bird species by their vocalizations. These aspects together make it difficult to apply methods based only on passive acoustic monitoring (Khal *et al*. 2021). While there are applications that can identify bird vocalizations, they only work with one vocalization at a time. In addition, not all species are included (Hill *et al*. 2018).

In recent years, proposals have been developed that use artificial intelligence to create tools that can identify all the bird species present in an audio file. This would be a major innovation for biological monitoring based on remote sensors (Olmedo, 2022).

Regarding failures with Audiomoth recorders, It should be mentioned that these recorders are low-cost, so the quality of their materials and circuitry may fail. Also, other factors should be evaluated, such as the required features for batteries and SD cards, or else they could generate performance problems.

Also, When the recorders are deployed in the field, sudden movements caused during installation, by wind, rain, and/or animals can disable them. Other factors such as temperature and humidity can affect battery performance. Since these situations are uncontrollable for the researcher, it is suggested that recurrent inspections be made to the recorders to verify their correct operation (Hill *et al*. 2019). These recorders were chosen because of their good performance, low cost, and ease of use (Hill *et al*. 2018). Therefore, it is possible to purchase more than one device and monitor more locations simultaneously. To avoid these inconveniences, it is suggested that a trial round of monitoring be performed. However, there are a wide variety of recorders available and it will depend on the budget.

The main challenge in this study was the analysis of the audio files and the identification of species by vocalizations. While Merlin and BirdNet work well, they present some difficulties. First, Merlin outperformed BirdNet. However, one of its disadvantages is data storage. If you want to use Merlin, the identification process is manual. That is, identifying file by file, which makes it time-consuming. BirdNet, on the other hand, allows the analysis of several audios simultaneously. But the library must be trained first with the desired species. Second, the identification of bird vocalizations. Both applications had difficulties. In the case of Merlin, inconsistencies in identifications were obtained.

Future studies should focus on refining the algorithms and expanding the databases of Merlin and BirdNet to improve their accuracy and taxonomic coverage, especially for species-rich regions with diverse vocalizations. Additionally, exploring the integration of multiple acoustic monitoring technologies, such as bioacoustic sensors and machine learning algorithms, could enhance the robustness of bird species identification systems. Moreover, collaborative efforts between researchers, developers, and citizen scientists are crucial for the continued development and validation of these applications in diverse ecological contexts (Smith *et a*l. 2022).

Currently, the way to optimize bird monitoring is to combine PAM or bioacoustic with point counts or line transects. In this case, an exploratory analysis can be carried out to know the peaks of activity and general activity in birds in a specific place or habitat. This is to perform point counts or line transects at those times with greater activity to make visual confirmations of rare species or of which a presence is sought. PAM would help to have the most complete listing of a habitat by monitoring continuously). In developing countries, where there are not many resources available for environmental monitoring. PAM proves to be a very good tool. It allows us to monitor, at least, a part of the biodiversity.

Conclusions

Passive acoustic monitoring (PAM) is a potential and effective method for improving conservation strategies for bird populations. The reason is that PAM has the following advantages: cost-efficient, less time-consuming, continuous and long-term recordings, large geographical scale, non-invasive and standardized method. This study has provided a comparative analysis of the Merlin and BirdNet applications for accurate bird species identification using passive acoustic monitoring. By evaluating key metrics such as overall accuracy and precision, as well as considerations of processing time and ease of use, valuable insights have been gained into the performance of these applications in real-world scenarios. Also, the results of the analysis indicated that both Merlin and BirdNet have strengths and weaknesses in their ability to accurately identify bird species from acoustic recordings. Although both applications perform acceptably well, there are variations in performance across species and environmental conditions. Merlin, with its large database and species-specific models, excels at identifying common and well-documented species. In addition, it works well with recordings where several birds vocalise at the same time. On the other hand, BirdNet's machine-learning algorithms show promise in distinguishing between similar and/or rare species. In addition, they can handle acoustic environments where there is not too much noise pollution.

In addition, the effectiveness of each application is influenced by factors such as geographic location, recording quality, condition of the recorders and their components (SD memory and batteries), and the user's knowledge. Therefore, practitioners and researchers should take these factors into account when selecting the most appropriate tool for their specific monitoring objectives. Despite their differences, both Merlin and BirdNet contribute significantly to advancing the field of avian biodiversity monitoring and conservation. By leveraging the capabilities of these applications, citizen scientists, conservationists, and researchers can collect valuable data on bird populations, contributing to our understanding of ecological dynamics and informing conservation efforts.

Going forward, further research is warranted to address the limitations identified in this study and explore potential synergies between Merlin and BirdNet. Furthermore, current advances in artificial intelligence and machine learning hold promise for improving the accuracy and efficiency of bird species identification using PAM.

In conclusion, the comparative analysis presented in this thesis underscores the importance of leveraging technological advances in PAM for effective monitoring and conservation of avian biodiversity. By harnessing the capabilities of tools such as Merlin and the Bird Network, we can continue to move towards the conservation and protection of our natural world for future generations.

References

Andrejeff, S., & Lehikoinen, P. (2023). Bird Sounds Global: automated species recognition with machine learning. Neotropical Birding, 33, 53-58.

Blake, J. G. (2021). Acoustic monitors and direct observations provide similar but distinct perspectives on bird assemblages in a lowland forest of eastern Ecuador. PeerJ, 9, e10565.

Bowler, D. E., Heldbjerg, H., Fox, A. D., de Jong, M., & Böhning-Gaese, K. (2019). Long-term declines of European insectivorous bird populations and potential causes. Conservation Biology, 33(5), 1120-1130.

Caro, T., Rowe, Z., Berger, J., Wholey, P., & Dobson, A. (2022). An inconvenient misconception: Climate change is not the principal driver of biodiversity loss. Conservation Letters, 15(3), e12868.

Darras, K., Batáry, P., Furnas, B. J., Grass, I., Mulyani, Y. A., & Tscharntke, T. (2019). Autonomous sound recording outperforms human observation for sampling birds: a systematic map and user guide. Ecological Applications, 29(6), e01954.

Darras, K., Batáry, P., Furnas, B., Fitriawan, I., Mulyani, Y., & Tscharntke, T. (2017). Autonomous bird sound recording outperforms direct human observation: Synthesis and new evidence. BioRxiv, 117-119.

Denton, T., Wisdom, S., & Hershey, J. R. (2022, May). Improving bird classification with unsupervised sound separation. In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 636-640). IEEE.

Dinno, A., & Dinno, M. A. (2017). Package 'dunn. test'. CRAN Repos, 10, 1-7.

Ebird (2023). Database. Extracted from: https://ebird.org/explore

Gonzalez, C., Johnson, C. D., Thompson, W. L., & Roberts, S. M. (2020). Evaluating the effectiveness of passive acoustic monitoring in bird surveys: A case study in tropical rainforests. Biodiversity and Conservation, 29(9), 2727-2743.

Green, R. E. (1995). Diagnosing causes of bird population declines. Ibis, 137, S47-S55.

Gregory, R. D., Eaton, M. A., Burfield, I. J., Grice, P. V., Howard, C., Klvaňová, A., ... & Burns, F. (2023). Drivers of the changing abundance of European birds at two spatial scales. Philosophical Transactions of the Royal Society B, 378(1881), 20220198.

Hill, A. P., Prince, P., Piña Covarrubias, E., Doncaster, C. P., Snaddon, J. L., & Rogers, A. (2018). AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment. Methods in Ecology and Evolution, 9(5), 1199-1211.

Hill, A. P., Prince, P., Snaddon, J. L., Doncaster, C. P., & Rogers, A. (2019). AudioMoth: A low-cost acoustic device for monitoring biodiversity and the environment. HardwareX, 6, e00073. https://doi.org/10.1016/j.ohx.2019.e00073

Hoefer, S., McKnight, D. T., Allen-Ankins, S., Nordberg, E. J., & Schwarzkopf, L. (2023). Passive acoustic monitoring in terrestrial vertebrates: a review. Bioacoustics, 32(5), 506-531.

Hutto, R. L., & Stutzman, R. J. (2009). Humans versus autonomous recording units: A comparison of point-count results. Journal of Field Ornithology, 80(4), 387-398.

Johnson, C. D., Smith, A. B., Thompson, W. L., & Roberts, S. M. (2021). Comparing point counts and passive acoustic monitoring for avian surveys in forested habitats. Ecological Applications, 31(3), e02257.

Kahl, S., Wood, C. M., Eibl, M., & Klinck, H. (2021). BirdNET: A deep learning solution for avian diversity monitoring. Ecological Informatics, 61, 101236.

Kassambara, A. (2020). ggpubr:'ggplot2'based publication-ready plots. R package version 0.4. 0.

Lapp, S., Stahlman, N., & Kitzes, J. (2023). A quantitative evaluation of the performance of the low-cost AudioMoth acoustic recording unit. Sensors, 23(11), 5254.

Lees, A. C., Haskell, L., Allinson, T., Bezeng, S. B., Burfield, I. J., Renjifo, L. M., ... & Butchart, S. H. (2022). State of the world's birds. Annual Review of Environment and Resources, 47, 231-260.

Leclère, D., Obersteiner, M., Barrett, M., Butchart, S. H., Chaudhary, A., De Palma, A., ... & Young, L. (2020). Bending the curve of terrestrial biodiversity needs an integrated strategy. Nature, 585(7826), 551-556.

Li, F., Xin, J., Chen, T., Xin, L., Wei, Z., Li, Y., ... & Liao, H. (2020). An automatic detection method of bird's nest on transmission line tower based on faster_RCNN. IEEE Access, 8, 164214-164221.

McMahon, B. J., Doyle, S., Gray, A., Kelly, S. B., & Redpath, S. M. (2020). European bird declines: Do we need to rethink approaches to the management of abundant generalist predators? Journal of Applied Ecology, 57(10), 1885-1890.

Mathez-Stiefel, S. L., Mulanovich, A. J., Jaquet, S., Bieri, S., Lojas, J., Breu, T., & Messerli, P. (2020). Establishing a science-management-society interface for biodiversity conservation and human well-being in the Amazon: the case of Madre de Dios, Peru. Ecosystems, 29(1), 1882-1882.

Microsoft Corporation (2024). Leading software technology company. Retrieved from https://www.microsoft.com/

Olmedo Becerra, G. E. (2022). Identification of Ecuadorian bird species based on the recognition of their vocalizations using neural networks for the promotion of bioacoustics as a method of bird monitoring (Master's thesis, Quito: Universidad Tecnològica Indoamèrica).

Pérez-Granados, C., & Traba, J. (2021). Estimating bird density using passive acoustic monitoring: a review of methods and suggestions for further research. Ibis, 163(3), 765-783.

Pijanowski, B. C., Farina, A., Gage, S. H., Dumyahn, S. L., & Krause, B. L. (2011). What is soundscape ecology? An introduction and overview of an emerging new science. Landscape ecology, 26(9), 1213-1232.

Piña-Covarrubias, E., Hill, A. P., Prince, P., Snaddon, J. L., Rogers, A., & Doncaster, C. P. (2019). Optimization of sensor deployment for acoustic detection and localization in terrestrial environments. Remote Sensing in Ecology and Conservation, 5(2), 180-192.

Priyadarshani, N., Marsland, S. & Castro, I. (2018). Automated birdsong recognition in complex acoustic environments: a review. Journal of Avian Biology 49, jav-01447.

Rahaman, A. (2023, November 21). Personal communication. BirdNet app questions. Retrieved from gmail.com

R Core Team (2024). R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Ruiz-Gutiérrez, V., Berlanga, H.A., Calderón-Parra R., Savarino-Drago, A., Aguilar-Gómez, M.A. and Rodríguez-Contreras, V. (2020). Illustrated Manual for Bird Monitoring. PROALAS: Latin American Program for Wild Birds. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad /Initiative for the Conservation of North American Birds, Mexico and Cornell Lab of Ornithology Mexico City and Ithaca N. Y. 104 pp.

Roberts, S. M., Johnson, C. D., Thompson, W. L., & Smith, A. B. (2018). A comparison of point counts and passive acoustic monitoring for avian species detection and richness in lowland forest of Papua New Guinea. The Condor, 120(2), 371.

Rogers A. (2024, April 2). Personal communication. Audiomoth audio files problems. Retrieved from gmail.com

Salas-Correa, Á. D., & Mancera-Rodríguez, N. J. (2020). Birds as ecological indicators of successional stages in a secondary forest, Antioquia, Colombia. Journal of Tropical Biology, 68(1), 23-39.

Smith, A. B., Johnson, C. D., Thompson, W. L., & Roberts, S. M. (2022). Advances in passive acoustic monitoring: Applications for bird surveys. Journal of Ornithology, 45(2), 120-135.

Stutchbury, B. J., & Morton, E. S. (2022). Behavioural ecology of tropical birds. Academic press.

Sueur J., Pavoine S., Hamerlynck O., Duvail S. (2008). «Rapid Acoustic Survey for Biodiversity Appraisal». PLoS ONE 3. 12 pp.

Thompson, W. L., Smith, A. B., Johnson, C. D., & Roberts, S. M. (2019). Bird surveys using point counts: A review of current scientific evidence. Journal of Avian Biology, 50(5), e02153.

Van Doren, B. M., Lostanlen, V., Cramer, A., Salamon, J., Dokter, A., Kelling, S., ... & Farnsworth, A. (2023). Automated acoustic monitoring captures timing and intensity of bird migration. Journal of Applied Ecology, 60(3), 433-444.

Ware, L., Mahon, C. L., McLeod, L., & Jetté, J. F. (2023). Artificial intelligence (BirdNET) supplements manual methods to maximize bird species richness from acoustic data sets generated from regional monitoring. Canadian Journal of Zoology, 101(12), 1031-1051.

Xie, J., Zhong, Y., Zhang, J., Liu, S., Ding, C., & Triantafyllopoulos, A. (2022). A review of automatic recognition technology for bird vocalizations in the deep learning era. Ecological Informatics, 101927.

Xu, X., Xie, Y., Qi, K., Luo, Z., & Wang, X. (2018). Detecting the response of bird communities and biodiversity to habitat loss and fragmentation due to urbanization. Science of the total environment, 624, 1561-1576.