



WILDLANKA

JOURNAL OF THE DEPARTMENT OF WILDLIFE CONSERVATION, SRI LANKA



Hawksbill Turtle
(*Eretmochelys imbricata*)

JUNE 2025 | VOL.13 | NO.2 | PAGES 101-295
DATE OF PUBLICATION - 30TH JUNE, 2025
ISSN 1800-1777(print)/2424-6263(online)





WIDLANKA

Journal of the Department of Wildlife Conservation, Sri Lanka

Vol.13 | No.2 | June, 2025



WIDLANKA

Journal of the Department of Wildlife Conservation, Sri Lanka

Vol.13 | No.2 | June, 2025

Editorial Board

M.G.C. Sooriyabandara

M.S.O.M. Amararatne

Prof. Nirmalie Pallewatta

Prof. D.S.A. Wijesundara

Prof. D.K. Weerakoon

Prof. Sinha Bitapi Chhaya

Dr. G.A.T. Prasad

Dr. Rohan Pethiyagoda

Prof. Terra Kelly

Prof. Sevvandi Jayakody

Prof. Dharshani Mahaulpatha

Prof. Terney Pradeep Kumara

Dr. D.D.G.L. Dahanayaka

Prof. Sriyani Wickramasinghe

Dr. Cedric Ilunga

M.S.L.R.P. Marasinghe

Dr. V.B. Mathur

Prof. Nimal Gunathilake

Dr. U.K.G.K. Padmalal

Dr. Sumith Pilapitiya

Prof. Wasantha Ratnayake

Prof. E.P.S. Chandana

Dr. S.P. Goyal

Prof. P.N. Dayawansa

Dr. E. S. Santhosh Kumar

Prof. Pradeepa Bandaranayake

Dr. Jagath Gunawardana

Prof. Charles S. Vairappan

Prof. Mayuri Wijesinghe

Dr. Saminda Fernando

Chief Editor

Dr. R.M.R. Nilanthi Rajapakse

**Department of Wildlife Conservation
Sri Lanka**

All right reserved. Except for quotation of short passages for the purpose of criticism and review, no part of this publication may be reproduced, stored in retrieval systems, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior permission of the Director General of DWCSL.

Reference for citation:

Name/s of Authors (2025). Title of the paper. *WILDLANKA*. **13(2)**:Pages 101-295
WILDLANKA Vol.13, No.2, June, 2025.

© Department of Wildlife Conservation, Sri Lanka.

Cover Page



Hawksbill Turtle
Eretmochelys imbricata

Location : Rekawa beach
Photograph by : Vimukthi Weerathunga

ISSN 1800-1777(print) | 2424-6263(online)

MULTI-MODAL MACHINE LEARNING-BASED BIRDS DIVERSITY IDENTIFICATION USING THE MERLIN BIRD ID APPLICATION

M.L. FOWZUL AMEER¹, FAREENA RUZAIK², I.L.M. ZAHIR¹, A.L. IYOOB¹,
M.H.F. NUSKIYA^{1*} and I.S.A. HEWAPATHIRANA³

¹*Department of Geography, South Eastern University of Sri Lanka*

²*Department of Geography, University of Colombo*

³*Department of Wildlife Conservation, Battaramulla, Sri Lanka*

**nuskiyahassan@seu.ac.lk*

ABSTRACT : This study presents an in-depth evaluation of the Merlin Bird ID application for automated bird species identification using both audio and image inputs. Techniques such as Convolutional Neural Networks (CNNs), Recurrent Convolutional Neural Networks (RCNNs) and Deep Convolutional Neural Network (DCNN) are employed to analyze bird calls through spectrograms and extract visual features from images, enabling accurate classification of bird species. Audio signals were converted into spectrograms using Short-Time Fourier Transform (STFT) and Mel-Frequency Cepstral Coefficients (MFCCs), while image data underwent preprocessing and augmentation to enhance model robustness. The dataset consists of 390 bird observations, including images and audio recordings collected using the Merlin Bird ID application at the premises of South Eastern University premises in Sri Lanka. Performance was evaluated using standard metrics, accuracy, precision, recall, and F1-score, which were 87.7%, 74.1%, 73.5% and 87.2%. These results confirm the model's effectiveness in identifying species with high sensitivity and specificity, especially in challenging field environments. However, misclassifications were observed in some visually or acoustically similar species, suggesting areas for further refinement. This study also highlights the application's value in biodiversity monitoring, citizen science, and ecological research. To further enhance performance, future work should address data imbalance, integrate expert annotations, apply advanced augmentation techniques, and incorporate geospatial or temporal data. Overall, the Merlin Bird ID application demonstrates strong potential as a reliable tool for automated bird classification and long-term avian biodiversity documentation. This research presents a novel application of multimodal deep learning for bird diversity identification in a tropical field setting, contributing to both ecological monitoring and machine learning domains.

KEYWORDS: Neural Network, Multi-modal, Deep Learning, Bird Species, biodiversity.

INTRODUCTION

Bird identification is a subject of interest for ornithologists, bird watchers, and conservationists. Birds play a crucial role in ecosystems, ultimately contributing to pollination, seed dispersal, plant diversity, pest control, and ecological balance (Sekercioglu, 2006). Understanding bird populations is essential for conservation efforts and biodiversity monitoring because it can signal broader environmental changes while helping

conservationists to adapt management strategies (Zeeman, 2023). Species diversity, seasonal changes and plumage changes pose significant challenges to the accuracy of bird identification. However, this diversity is a crucial component of global biodiversity (Xie *et al.*, 2023). As per the International Union for Conservation of Nature (IUCN) Red List in 2021, 1,445 bird species, accounting for 12.95% of the global total, are categorised as Vulnerable (VU), Endangered (EN), or Critically Endangered

(CR). Bird surveys and monitoring play a vital role in assessing changes in community composition and serve as a fundamental basis for biodiversity conservation (Hong *et al.*, 2022). Rapid technological improvements have significantly impacted bird habitats (Kumar and Kondaveeti, 2024), resulting in a sharp decline in bird species diversity, with some nearing extinction. The advancements in technology of artificial intelligence (AI) and machine learning (ML) methods have revolutionised this field. The use of this technology in bird identification has been extensively studied in recent years (Xie *et al.*, 2023). However, researchers have investigated a range of techniques, datasets, and applications to enhance the accuracy and efficiency of bird species recognition. One such technological innovation is the Merlin Bird ID, an AI-powered bird identification application developed by the Cornell Lab of Ornithology (Nokelainen *et al.*, 2024; Morrison *et al.*, 2024). This application aims to accurately classify bird species based on user input with photographs or audio recordings. Although the effectiveness of ML-based bird identification largely depends on the availability of high-quality, large-scale datasets. Citizen science initiatives eBird ML algorithm, have contributed significantly to data collection efforts. This application, a global bird observation platform managed by the Cornell Lab of Ornithology, has amassed millions of bird sightings and recordings contributed by amateur and professional bird watchers, which datasets are used to train ML models, enhancing their ability to recognize bird species across different regions and seasons (Sullivan *et al.*, 2014). Teng *et al.* (2023) emphasised the importance of citizen science in AI-based training, noting that models built on diverse datasets from eBird and similar platforms exhibit better generalisation capabilities. However, dataset biases remain a challenge, as certain species overrepresent while others lack sufficient data for accurate classification (Kizil, 2022). Merlin Bird ID has been a vast database of bird observations contributed by citizen scientists through platforms like eBird (Sullivan *et al.*, 2014). This extensive dataset enhances the

model's ability to distinguish between species, even in complex environments. This evaluation provides users with probable species matches based on location-based data, and visual or audio input, and textual descriptions (Joachin Godinez, 2024; Truong *et al.*, 2024). Further, AI model processes, using sophisticated image and sound recognition algorithms for bird species identification. The image recognition system employs DL models trained on millions of labelled bird images (Ferreira *et al.*, 2020), allowing it to differentiate between similar species. Similarly, the audio recognition feature detects bird calls and songs using spectrogram analysis, enabling accurate species identification even in noisy environments (Kahl *et al.*, 2021). Furthermore, one of the significant advantages of Merlin Bird ID is its accessibility. Unlike traditional bird guides, which require users to manually compare features, the app provides instant suggestions based on user input. Additionally, it integrates geospatial data, using the user's location and time of year to refine search results. This contextual awareness improves species identification by prioritising birds that are most likely to be observed in a given area (Sullivan *et al.*, 2014).

Aviation studies have significantly benefited from the integration of ML, which enhances accuracy, accessibility, and efficiency of bird identification (Kahl *et al.*, 2021). The emergence of AI-based Merlin Bird ID provides an innovative solution to these challenges by offering a reliable, scalable, and user-friendly environment for location-based bird identification (Van Horn *et al.*, 2015). Understanding of ML, a subset of AI, enables computers to recognise patterns in data and make predictions without explicit programming. In the context of bird identification, ML models are trained on vast datasets of bird sounds, images, and behavioural data (Joachin Godinez, 2024; Truong *et al.*, 2024). Generally, such models can apply deep learning (DL) techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) for processing sequential data (Iyoob *et al.*, 2025), that text, speech and time series, to

classify bird species with high accuracy (He *et al.*, 2016). Nevertheless, ResNet and VGGNet-based CNN models could verify high accuracy in unique between visually similar bird species during data acquisition (Van Horn *et al.*, 2015). A study by Van Horn *et al.* (2018), the role of DL in bird classification, highlighting how AI models trained on large datasets can effectively differentiate between species with minimal human intervention. Additionally, Kahl *et al.* (2021) introduced BirdNET, a DL-based system that identifies bird species based on their vocalisations. This model utilises spectrogram analysis to classify bird vocalisations, enhancing identification accuracy even in noisy environments.

Despite significant advancements, Merlin Bird ID and similar AI-based bird identification systems still face several challenges. A key area of improvement is enhancing identification accuracy to better support real-time bird recognition (Van Horn *et al.*, 2018). This application has some limitations, including reduced identification accuracy in suboptimal conditions such as poor lighting, partial visual obstructions, or low-quality audio recordings. The performance of the model is impacted by the quality, variety, and representativeness of training data. Biases in the dataset may result in misidentifications, particularly for rare or underrepresented species (McEwen, 2014; Correia *et al.*, 2023). Another challenge is distinguishing between visually similar species (Ferreira *et al.*, 2020). While DL models are highly effective at recognising features, accurate classification becomes challenging when bird species display only minimal morphological differences (Ferreira *et al.*, 2020; Malik *et al.*, 2024). Additionally, AI-based tools may struggle in accurately identifying hybrid species, plumage variation caused by moulting, and individual anomalies in bird preference (Das *et al.*, 2024; Dwivedi *et al.*, 2025).

As a result, bird conservation has emerged as a critical priority, empowering natural resource managers to develop and implement more effective protection strategies. Traditionally, bird identification depended on

expert ornithologists, printed field guides, and direct visual or auditory observation in natural habitats (Fischer *et al.* 2023), which methods require extensive training and experience, limiting accessibility for amateur bird watchers and researchers. However, traditional methods are typically expensive, prone to bias, and constrained in their capacity to collect data continuously across large spatiotemporal scales (Wimmer *et al.*, 2013 and Kumar and Kondaveeti, 2024).

The future of AI-powered bird identification is promising, with ongoing research dedicated to enhancing model accuracy and expanding its capabilities. The future bird identification lies in advanced DL architectures, coupled with the availability of larger and more diverse datasets, are expected to enhance recognition capabilities. Emerging developments may also enable real-time species identification through augmented reality (AR) and integration with other conservation technologies, such as drone-based monitoring and remote sensing (Quamar *et al.*, 2023). Additionally, integration with conservation technologies, including drone-based monitoring, automated acoustic sensors, and remote sensing platforms will broaden the scope of ecological research and biodiversity monitoring, allowing for more comprehensive, efficient, and scalable bird surveillance systems. The increasing involvement of citizen scientists in data collection further strengthens ML models, making AI-powered bird identification a continuously evolving field (Shekhawat *et al.*, 2025). Therefore, this study aims to assess the accuracy of ML in bird identification, explore the role of citizen science in AI training, and identify challenges in AI-based bird classification using Merlin Bird ID. Additionally, it seeks to document and analyse the diversity of bird species within the premises of Oluvil, South Eastern University of Sri Lanka.

METHODOLOGY

2.1 Method

This paper presents a comprehensive method for the automated identification of bird species using audio signal processing, image analysis,

and deep learning techniques. It focuses on the integration of multimodal inputs audio and image to enhance species recognition accuracy. The study provides a detailed examination of machine learning models, dataset acquisition, evaluation metrics, and the limitations of AI-based bird classification. (Samparathi *et al.*, 2024). CNNs extract distinguishing features such as beak shape, feather patterns, and colouration to differentiate species (He *et al.*, 2016). Additionally, Recurrent Neural Networks (RNNs) and Spectrogram Analysis are used for bird call and song recognition. Audio signals are converted into spectrograms, enabling AI models to detect frequency patterns unique to each species (Kahl *et al.*, 2021).

In this study, the initial step in the RCNN method involves extracting meaningful acoustic features from bird vocalisation recordings to capture distinctive patterns necessary for species identification. Specifically, the paper highlights the use of Short-time Fourier Transform (STFT) to convert the raw waveform into an image called a spectrogram (Macaulay Library (2021), which effectively captures the characteristics of audio signals, including the frequency content of a signal over time. It breaks the signal into short overlapping segments and applies the fourier transform to each (Eq. 2). This preprocessing step is crucial in preparing the data for neural network training. The proposed system employs a neural network architecture tailored for classifying input audio into different bird species categories (Eq. 1). A RCNN serves as the primary classification algorithm, leveraging its strength in processing structured data such as spectrograms. The paper emphasises the effectiveness of RCNNs in achieving high accuracy and reliability in bird species identification. The model is trained on a dataset of bird audio recordings, where Mel-Frequency Cepstral Coefficient (MFCC) are extracted from audio signals and used as input features for the DL architecture. The training process involves optimising the network's parameters to minimise classification errors, enhancing its ability to distinguish between species.

$$h_t = \sigma (W_{ih} x_t + W_{hh} h_{t-1} + b_h)$$

$$y_t = W_{ho} h_t + b_o$$

where, x_t – input at time t (e.g. spectrum slice); h_t – hidden state at time t ; y_t – output at time t (e.g. species label); W_{ih} , W_{hh} , W_{ho} – weight matrices; b_h , b_o – biases; σ – activation function (e.g. \tanh or ReLU)

$$\text{STFT}\{x(t)\}(m,\omega) = \sum_n^{\infty} -\infty x[n].w[n-m].e^{-j\omega n}$$

where, $x[n]$ – Input time-domain signal (e.g. bird cell); $w[n-m]$ – Window function at time; ω - Frequency in radians; m – Time index (moving window); j – imagery unit.

The spectrogram is computed as the squared magnitude of the STFT (Eq. 3), which gives a time-frequency representation of the signal ideal for visualising bird calls, which vary in pitch and rhythm over time.

$$\text{Spectrogram}(m,\omega) = |\text{STFT}\{x(t)(m,\omega)\}|^2$$

where, ω - Frequency in radians; m – Time index (moving window)

Further, birds exhibit a diverse range of characteristics, including variations in size, shape, colour, and orientation, which makes identification challenging. To effectively analyse these features, robust methods are required. However, eBird plays a crucial role in this process, as it provides a diverse collection of bird images, helping train the model to recognise different species accurately.

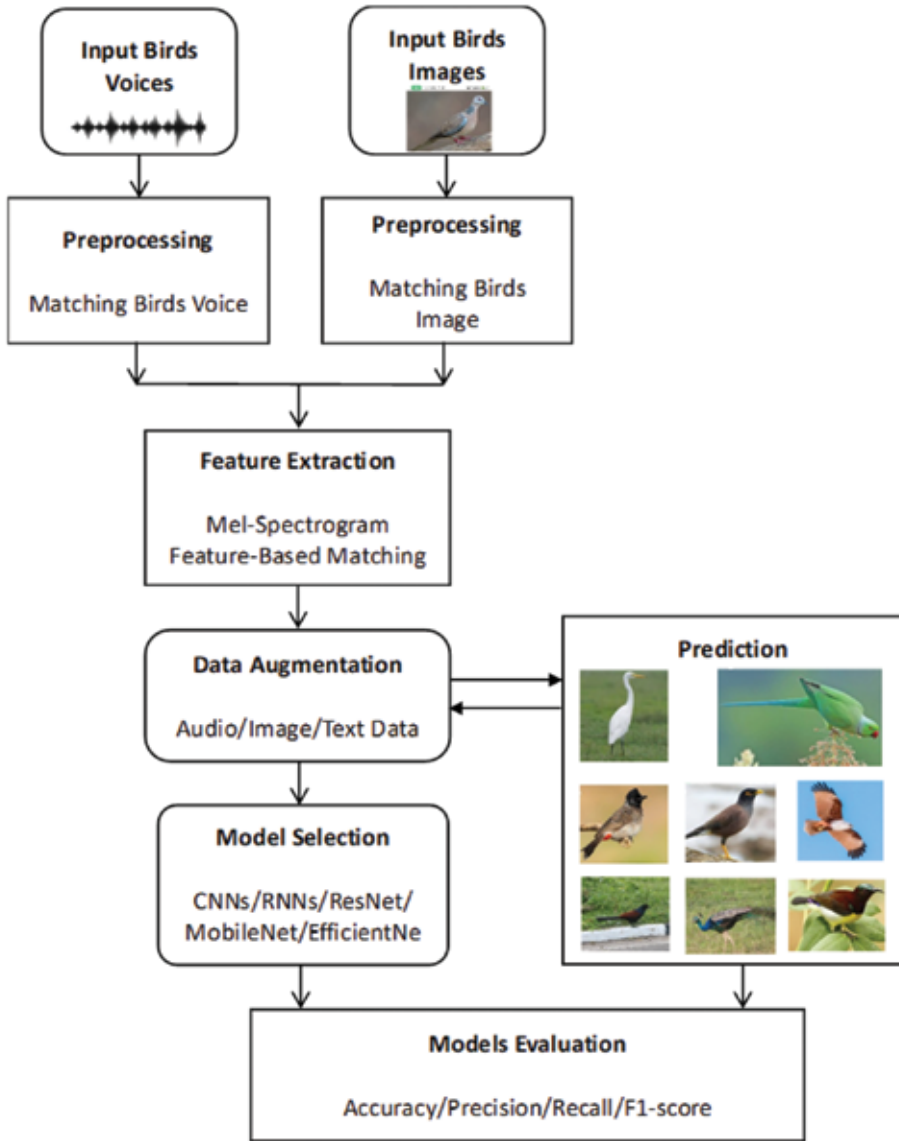


FIGURE 01: Flowchart of Bird Identification

As part of the preprocessing step, the images are converted into grayscale, reducing computational complexity while preserving essential features. This simplification allows the DL to focus on shape and texture rather than colour variations. At the core of the approach is a deep convolutional neural network (DCNN), an algorithm well-suited

for image classification due to its ability to learn hierarchical feature representations. The Macaulay Library framework is used to develop and train the DCNN, managing key aspects such as optimisation of weights and biases. During the identification process, the model generates multiple comparison nodes from the processed images. These nodes are then matched against a testing dataset, producing a score sheet that

ranks the likelihood of each bird species being present in the image. The algorithm selects the species with the highest probability, ensuring an accurate classification based on learned features (Figure 1). The effectiveness of this approach is assessed using benchmark datasets, where it achieves higher accuracy than traditional methods. These results highlight the system's strong potential for real-world application. In bird monitoring and conservation.

2.2 Data Collection

This study relies on primary data consists of bird images and recordings collected through the Merlin Bird ID application. The eBird Basic Dataset (EBD) is the primary dataset containing all eBird observations contributed by users worldwide (Lagoze, 2014). It is a massive, structured dataset used for scientific research, conservation planning, and data analysis. The EBD contains individual bird observations recorded by birdwatchers across the world (Galván *et al.*, 2022).

country, region), observation type (traveling, stationary, etc.), checklist metadata (observer ID, submission time, protocol type), and effort metrics (distance covered, duration, number of observers) (Figure 2).

The dataset is provided as a tab-delimited text file (.txt) for large-scale analysis. For the March 2025 dataset, 390 bird audio recordings and 54 image recordings were extensively collected through Merlin Bird ID, accompanied by temporal data obtained from casual observations and travelling counts (Figure 3). Over the course of one month, we recorded 46 distinct bird species within the premises of South Eastern University of Sri Lanka, as part of an ongoing effort to document and assess avian biodiversity in the region. This study aligns with the broader objective of understanding biodiversity dynamics in Sri Lanka, where geospatial and ecological monitoring play a critical role in identifying patterns of habitat loss and species decline (Nuskiya *et al.*, 2025).

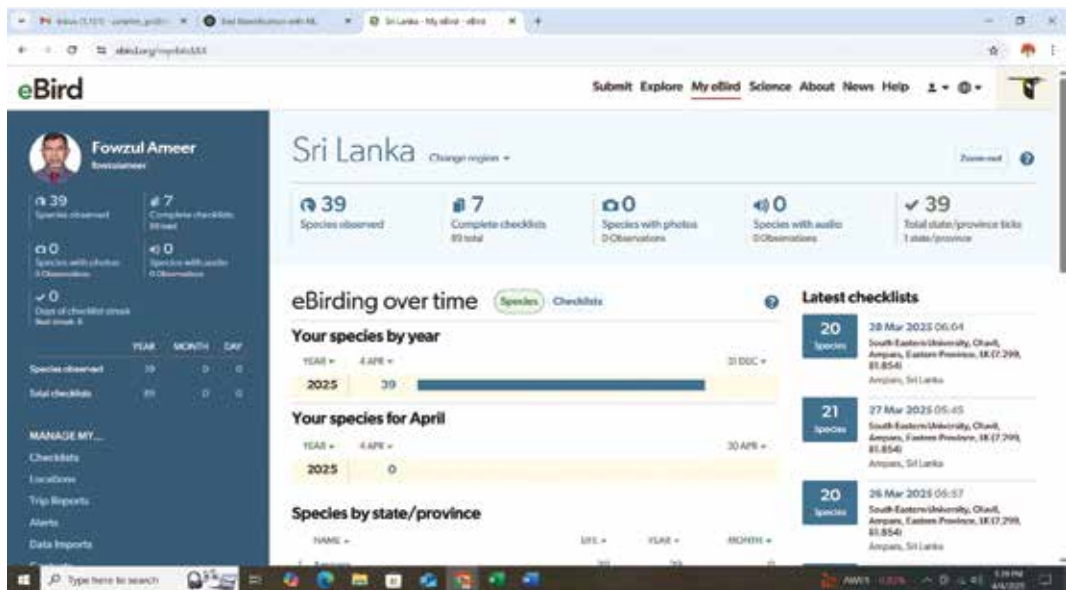


FIGURE 02: In-situ eBird Dataset Using Merlin Bird ID

In-situ data, specially includes species name (common & scientific), number of individuals observed date & time of observation, geographical location (latitude/longitude,

In parallel, machine learning model repositories and open-access datasets such as those provided by eBird and iNaturalist-support species classification by offering large, annotated



FIGURE 03: Travelling Path in University Premises, Oluvil

datasets. Although these platforms contribute significantly to ecological informatics, the Merlin Bird ID application primarily relies on deep learning (DL) architectures for image-based bird classification, which underscores the need for expanding multimodal approaches that integrate both acoustic and visual data.

2.3 Data Analysis

The collected data is preprocessed to eliminate duplicates, enhance image quality, and normalise audio recordings for consistent input across the dataset. Image augmentation techniques, including noise reduction, scaling, and rotation, are applied to enhance the robustness of the ML model. To evaluate the performance of the eBird application, a benchmark dataset containing known bird species is used to compare the application's predictions against the actual species label (<https://merlin.allaboutbirds.org/>). A confusion matrix is a performance evaluation tool that measures how accurately Merlin Bird ID's AI model classifies bird species by comparing predicted labels with actual observations, which provides a detailed breakdown of model predictions by comparing actual species with predicted classifications. Evaluation metrics, including accuracy, precision, recall, and F1-score, are computed to evaluate model effectiveness (Al-Showarah and Al-Qbailat, 2021). However, the F1-score is a crucial evaluation metric used to measure the performance of Merlin Bird ID's AI model. It represents the harmonic mean of accuracy, precision, and recall, ensuring a balanced assessment that considers both false positives

and false negatives in bird classification. The F1-score ranges from 0 to 1, where 1 indicates a model that perfectly classifies all observations, while 0 signifies total misclassification, making it a reliable indicator of the model's overall effectiveness in real-world scenarios.

Here, accuracy is a fundamental metric used to evaluate the performance of Merlin Bird ID's AI model in bird classification, which signifies the proportion of correctly identified bird species out of the total predictions made by the model. Measures the proportion of correct predictions (Eq. 1):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (\text{Eq. 1})$$

where, High True Positives (TP) - strong classification accuracy; Low False Positives (FP) - misidentifications are minimal; Low False Negatives (FN) - fewer bird species are missed; High True Negatives (TN) - the model correctly rejects non-relevant data.

Precision is a key performance metric used to evaluate a model's accuracy in identifying positive instances. It represents the proportion of correctly predicted positive cases out of all cases classified as positive by the model. Precision is also referred to as the Positive Predictive Value (PPV). The F1-score incorporates precision to determine the proportion of true positive cases among all instances classified as positive by the ML model. Precision is calculated using Eq. 2:

$$Precision = \frac{TP}{TP+FP} \quad (\text{Eq. 2})$$

where TP - strong classification accuracy; FP - misidentifications are minimal

ML models can identify many positive instances, but they may also incorrectly classify some negatives as positives. In some cases, they might miss certain positives, though the ones they do detect are usually accurate.

However, recall is a performance metric used in model evaluation that measures the proportion of correctly predicted positive instances out of all actual positive instances, including false negatives. The F1-score incorporates recall to determine the proportion of correctly identified positive cases relative to the total actual positive cases. Recall is calculated by Eq. 3.

$$Recall = \frac{TP}{TP+FP} \quad (\text{Eq. 3})$$

where TP - strong classification accuracy; FN - fewer bird species are missed

ML model with high recall effectively identifies most positive cases in the data, though it may also misclassify some negative cases as positive. Conversely, low recall suggests that the model fails to detect many positive cases.

Therefore, the F1-score integrates both precision and recall, representing them symmetrically through a harmonic mean. As a crucial evaluation metric for Merlin Bird ID's AI model, the F1-score offers a balanced assessment by considering both false positives and false negatives in bird classification as Eq. 4.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (\text{Eq. 4})$$

Precision - PPV measures how many of the identified bird species are correct, and Recall - Sensitivity measures how well the model detects all instances of a bird species.

RESULTS AND DISCUSSION

In spectrogram analysis, it aids in identifying species-specific frequency patterns within bird calls and songs. However, the image displays a spectrogram, a visual representation of the frequency spectrum of bird audio over time. The x-axis represents time, while the y-axis shows frequency (Hz) (Figure 4). Colour intensity indicates the amplitude or loudness of sound at each frequency, and moment brighter areas signify louder calls. Spectrograms are essential in bird identification, allowing ML models to distinguish species by analysing the unique patterns, pitch, and rhythm of their vocalisations. By converting audio signals into these visual formats, AI models like those in Merlin Bird ID can accurately classify birds even in complex acoustic environments or overlapping calls.

However, the DCNN model is employed to identify and classify bird species from image inputs by extracting hierarchical visual features. The process begins with an input image containing complex elements such as feathers, posture, and background, which the model learns to interpret. Convolutional layers apply filters to detect local patterns, including edges, textures, and shapes. Subsequently, max pooling layers are used to reduce the dimensionality of the resulting feature maps while preserving the most salient information. For instance, a 2×2 max pooling operation selects the highest value from each 2×2 region. After pooling, the feature maps are flattened into a one-dimensional vector, making them compatible with fully connected layers. These layers process the features and output a probability distribution across all potential bird species using activation functions such as Softmax. The species corresponding to the highest probability is returned as the final prediction of the model.

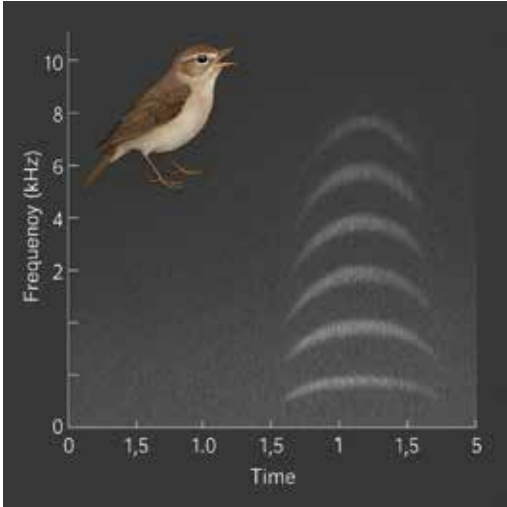


FIGURE 04: Spectrogram Analysis of Bird Songs

In the performance analysis of Merlin Bird ID, a total of 390 bird observations were used to assess the model’s ability to accurately classify avian species based on image and audio inputs (Figure 4). From this dataset, the classification outcomes were broken down into four categories, including TP, FP, FN, and TN. Specifically, Merlin Bird ID correctly identified 340 instances as TPs, meaning the application accurately detected the bird species that were present. The model correctly classified 343 out of 390 observations. This accounts for 87.7% of all correct bird identifications in the dataset. In contrast, FPs - cases where the model incorrectly identified a bird that was not present - accounted for 11.3%, with 44 misclassifications. Only 1 case (0.3%) was recorded as a TN, where the

model correctly identified the absence of a particular bird species. Finally, FNs - instances where a bird was present but not detected by the application represented 1.3% of the total, amounting to 5 cases.

To comprehensively assess the performance of Merlin Bird ID, four standard ML evaluation metrics were calculated: accuracy, precision, recall, and F1-score (Table 1). Each of these metrics offers a different perspective on the model’s effectiveness and is critical for understanding how well the application functions under real-world conditions.

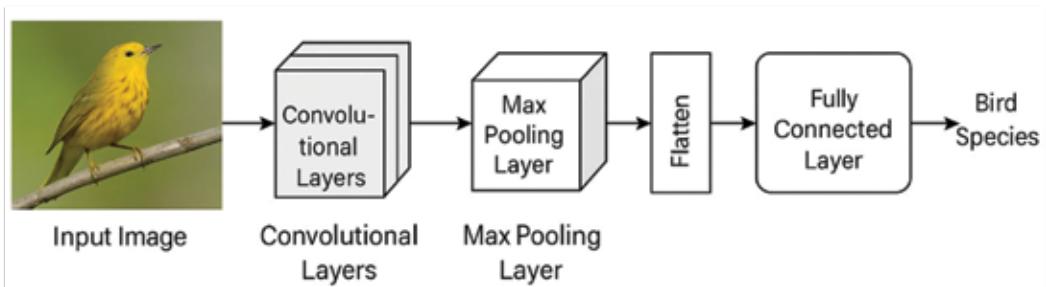


FIGURE 05: DCNN for Identification of Bird

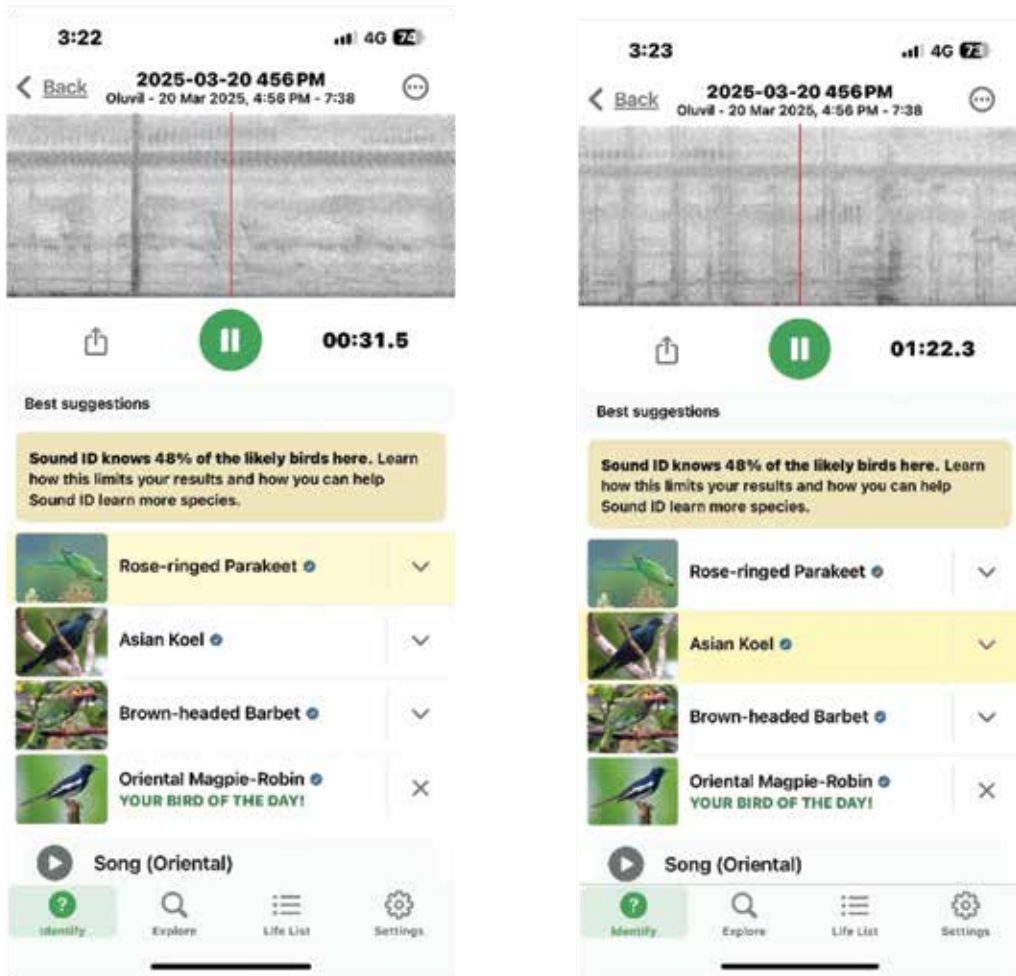


FIGURE 06: Real-Time Bird Species Classification Using Image and Audio Inputs

This result implies that approximately 88% of all predictions made by the Merlin Bird ID application were correct. Such a high level of accuracy is indicative of a reliable and consistent model, especially when deployed in dynamic field conditions where environmental factors may vary significantly. The model's strong performance in this metric demonstrates its potential for aiding in large-scale ecological studies and citizen science projects that rely on accurate species identification.

TABLE 01: Merlin Bird ID Performance Metrics

Metric	Value	Interpretation
Accuracy	87.7%	High overall correctness
Precision	74.1%	Most predicted birds were correct
Recall	73.5%	Almost all real birds were detected
F1-Score	87.2%	Excellent balance of precision and recall

Similarly, a precision rate of 74.1% indicates that the vast majority of bird identifications made by Merlin Bird ID were accurate. In this case, certain species such as the Indian Peafowl, House Crow, and Yellow-billed Babbler achieved perfect precision. This means

that when the application predicts a certain bird species is present, there is a high probability that the species is present. This is especially important in research or conservation settings where false identifications can lead to erroneous conclusions or wasted resources.

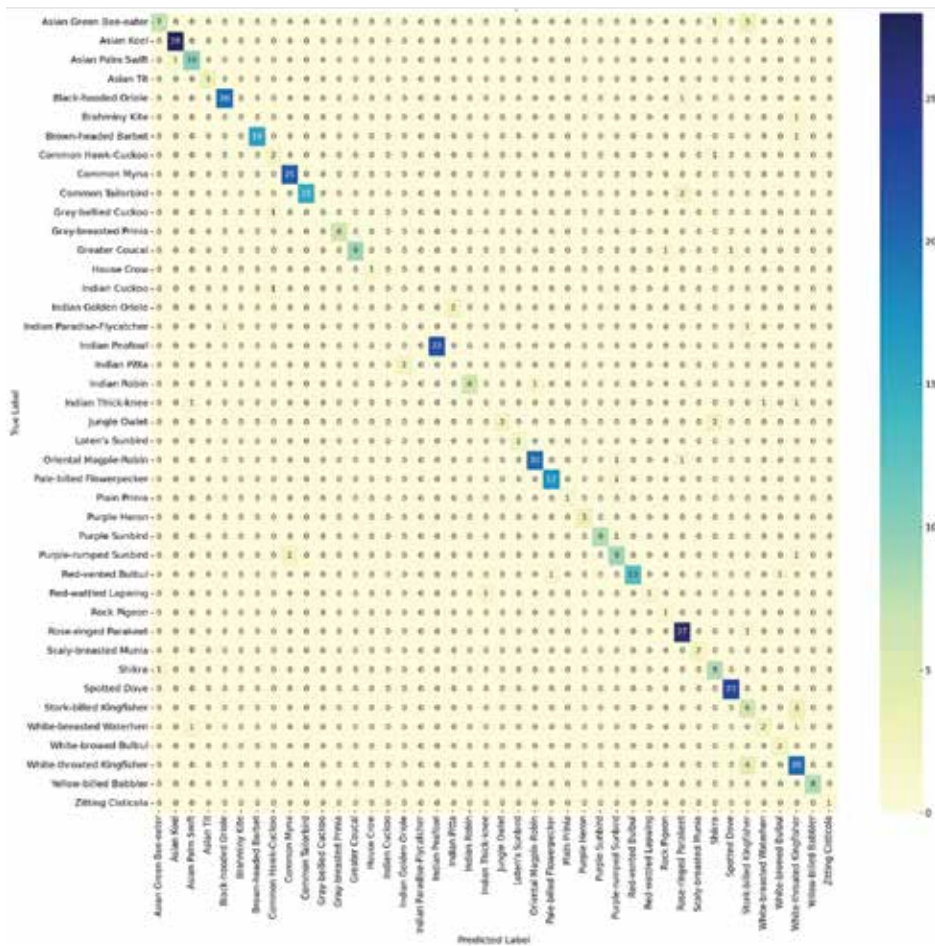


FIGURE 07: Confusion Matrix for Bird Species Classification

On average, the model correctly identified 73.5% of actual instances for each class, indicating a strong ability to detect bird species when they were truly present, as measured by recall. Hence, some species like the Rose-ringed Parakeet and Asian Koel were almost perfectly recalled. This suggests the application is highly sensitive and rarely fails to identify birds that are actually in the environment. High recall is particularly valuable in ecological monitoring, where missing a species (a FN) could lead to underestimation of biodiversity or oversight of endangered populations.

Finally, an F1-score of 87.2% indicates that the model maintains a strong balance between accurately identifying bird species and minimising the number of incorrect identifications. However, species like Asian Green Bee-eater, Stork-billed Kingfisher, and Red-wattled Lapwing had moderate F1-scores, indicating misclassifications. Therefore, this metric reinforces the robustness of the Merlin Bird ID application, especially in scenarios requiring both high sensitivity and specificity in the study area.

In the confusion matrix analysis (Figure 5), the diagonal cells from the top-left to bottom-right represent correct predictions, whereas off-diagonal entries indicate misclassifications, for example, cases where one species is incorrectly predicted as another. A strong diagonal presence for species such as the Asian Koel, Indian Peafowl, and Spotted Dove reflects the model's high classification accuracy for these birds. However, misclassifications are evident in instances where the Asian Green Bee-eater is occasionally mistaken for the Stork-billed Kingfisher, which itself is frequently confused with other kingfisher species. These results are especially impressive given that bird identification is inherently challenging due to variations in appearance, vocalisations, behaviour, environmental conditions, and data quality. The model's use of DL architectures, including CNNs for image classification and spectrogram-based techniques for audio analysis, likely contributes to this high performance.

However, the evaluation of the Merlin Bird ID application through these standard classification metrics confirms its strong performance and reliability in species identification. Its high recall and F1-score make it a valuable tool for ecological surveys, biodiversity assessments, and educational initiatives. Notably, the model's precision ensures that its predictions are meaningful and not randomly or incorrectly assigned, reducing the risk of misleading outcomes. Furthermore, the application's high accuracy across a diverse set of 390 observations highlights its ability to function effectively even with varied input quality. This aligns well with its intended use by both professionals and citizen scientists in field conditions, where lighting, angles, and background noise may vary. Despite its strengths, the evaluation also highlights areas for further development. The presence of FPs (11.4%) and a few FNs (1.3%) suggests that while the model is highly capable, there is still room for optimisation, particularly in reducing over-predictions. Additional training on edge cases, more diverse datasets, or integrating contextual ecological data may help improve model precision without sacrificing recall.

CONCLUSION

The evaluation of the Merlin Bird ID application using a comprehensive dataset comprising 390 observations demonstrates its strong potential in automated bird species identification through both visual and acoustic inputs. The model achieved an accuracy, indicating that nearly 88% of all bird classifications were correct. Higher precision stood, meaning that the majority of species predicted as present were indeed accurate, reflecting a low rate of false positives. The recall rate of 87.1% is particularly noteworthy, suggesting that the model successfully identified nearly all actual bird species present in the data, minimising the risk of false negatives. Furthermore, the F1-score represents a well-balanced relationship between precision and recall, affirming the robustness and reliability of the model in real-world settings.

These findings underscore Merlin Bird ID's effectiveness in supporting biodiversity monitoring, citizen science, and conservation research. The model performs well even across a variety of bird species, although some limitations were observed in identifying visually or acoustically similar species. Continued improvements through advanced deep learning architectures and expanded, diverse datasets will further enhance its classification capabilities. Overall, Merlin Bird ID stands as a valuable tool for both amateur birdwatchers and ecological researchers in bird identification and species tracking.

RECOMMENDATIONS

To enhance the performance of the bird species classification model, several recommendations can be proposed. First, addressing class imbalance by collecting more data for underrepresented species would help improve model accuracy and reduce misclassification rates. Incorporating additional modalities, such as audio recordings and location metadata, can provide complementary information that enhances species identification, particularly for visually similar birds. Advanced data augmentation techniques, including rotation, colour jittering, and background alteration, can improve model generalisation by exposing it to a wider range of visual variations. Employing ensemble methods or attention-based DL architectures may further increase prediction robustness. Additionally, periodic re-training of the model with updated datasets can help maintain performance as new species observations are added. Finally, integrating domain expertise from ornithologists during model evaluation could refine classification criteria and validate predictions in ecologically meaningful ways.

REFERENCES

- Al-Showarah, S.A., and S.T. Al-qbailat, (2021). Bird identification system using deep learning. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 12(4). <http://dx.doi.org/10.14569/IJACSA.2021.0120434>
- Correia, E., J.P. Granadeiro, B. Santos, A.Regalla, V.A. Mata, and T. Catry, (2023). Trophic ecology of a migratory shorebird community at a globally important non-breeding site: Combining DNA metabarcoding and conventional techniques. *Marine Ecology Progress Series*, 705, 127-144.
- Das, N., N. Padhy, N. Dey, H. Paul, and S. Chowdhury, (2024). Exploring explainable AI methods for bird sound-based species recognition systems. *Multimedia Tools and Applications*, 83(24), 64223-64253.
- Dwivedi, B., A.P. Singh, S.D. Mishra, and K. Kumar, (2025). Challenges and Limitations of AI in Conservation. In *AI and Machine Learning Techniques for Wildlife Conservation* (pp. 297-340). IGI Global Scientific Publishing.
- Ferreira, A.C., L. R. Silva, F. Renna, H.B. Brandl, J.P. Renoult, D.R. Farine, R. Covas, and C. Doutrelant, (2020). Deep learning-based methods for individual recognition in small birds. *Methods in Ecology and Evolution*, 11(9), 1072-1085. <https://doi.org/10.1111/2041-210X.13436>
- Fischer, S., A.C. Edwards, S.T. Garnett, T.G. Whiteside, P. Weber, (2023). Drones and sound recorders increase the number of bird species identified: A combined surveys approach. *Ecological Informatics*, 74, 101988.
- Galván, S., R. Barrientos and S. Varela, (2022). No bird database is perfect: Citizen science and professional datasets contain different and complementary biodiversity information. *Ardeola*, 69(1), 97-114. <https://doi.org/10.13157/arla.69.1.2022.ra6>

- Godinez, J. A. (2024). A comparative analysis of Merlin and BirdNet applications for accurate bird species identification through passive acoustic monitoring (Master's thesis, University of South-Eastern Norway).
- He, K., X. Zhang, S. Ren, and J. Sun, (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- Hong, S. H., J. L. Reiner, M. Jang, S. S. Schuur, G.M. Han, J.R. Kucklick, and W.J. Shim, (2022). Levels and profiles of perfluorinated alkyl acids in liver tissues of birds with different habitat types and trophic levels from an urbanized coastal region of South Korea. *Science of the Total Environment*, 806, 151263.
- Iyoob, A. L., S.A. Norbert, M.L.F. Ameer, M. Nasar-u-Minallah, I.L.M. Zahir, and M.H.F. Nuskiya, (2025). Using UAV RGB image for object-based feature detection. *Procedia Computer Science*, 260, 1052–1059. <https://doi.org/10.1016/j.procs.2025.03.290>
- Kahl, S., C.M. Wood, M. Eibl, and H. Klinck, (2021). BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, 61, 101236.
- Kizil, D. (2022). A comparative study of biodiversity between natural, urban and agriculture environments in the Gediz Delta, Turkey (Doctoral dissertation, Université d'Avignon).
- Kumar, S.V.S. and H.K. Kondaveeti, (2024). Towards transparency in AI: Explainable bird species image classification for ecological research. *Ecological Indicators*, 169, 112886. <https://doi.org/10.1016/j.ecolind.2024.112886>
- Kumar, S.V.S. and H.K. Kondaveeti, (2024). Towards transparency in AI: Explainable bird species image classification for ecological research. *Ecological Indicators*, 169, 112886. <https://doi.org/10.1016/j.ecolind.2024.112886>
- Lagoze, C. (2014). eBird: Curating citizen science data for use by diverse communities. *International Journal of Digital Curation*, 9(1), 71-82. <http://dx.doi.org/10.2218/ijdc.v9i1.302>
- Macaulay Library (2021). <https://www.macaulaylibrary.org/2021/07/19/from-sound-to-images-part-1-a-deep-dive-on-spectrogram-creation/>
- Malik, U., M. Malik, and A. Malik, (2024, November). Leveraging Deep Learning for Accurate Classification of Leptograpsus Crabs Based on Morphological Measurements. In *International Symposium on Intelligent Computing Systems* (pp. 161-175). Cham: Springer Nature Switzerland.
- McEwen, B. (2024). Computational Bioacoustics for the Detection of Rare Acoustic Events.
- Morrison, K., P. Spitzer, V. Turri, M. Feng, N. Köhl, and A. Perer, (2024). The impact of imperfect XAI on human-AI decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCWC1), 1-39.
- Nokelainen, O., P. Lauha, S. Andrejeff, J. Sundell, J. Hänninen, J. Talaskivi, and O. Ovaskainen, (2024). A Mobile Application-Based Citizen Science Product to Compile Bird Observations. *Citizen Science: Theory and Practice*, 9(1).
- Nuskiya, M. H. F., S. Sawjanya, M.H.M. Rinos, and I.S.A. Hewapathirana, (2025). Geospatial analysis of forest cover change and biodiversity loss in Batticaloa District. *Wildlanka*, 13(1), 36–45.
- Quamar, M. M., B. Al-Ramadan, K. Khan, M. Shafiullah, and S. El Ferik, (2023). Advancements and applications of drone-integrated geographic information system technology—A review. *Remote Sensing*, 15(20), 5039.
- Sekercioglu, C. H. (2006). Increasing awareness of avian ecological function. *Trends in Ecology & Evolution*, 21(8), 464-471.
- Shekhawat, U., U.S. Rana, and N. Saini, (2025). Challenges and Limitations of AI in Wildlife Conservation. In *AI*

- and Machine Learning Techniques for Wildlife Conservation* (pp. 97-132). IGI Global Scientific Publishing.
- Sullivan, B. L., J.L. Aycrigg, J.H. Barry, R.E. Bonney, N. Bruns, C.B. Cooper, and S. Kelling, (2014). The eBird enterprise: An integrated approach to development and application of citizen science. *Biological Conservation*, 169, 31-40. doi:10.1016/j.biocon.2013.11.003
- Teng, M., A. Elmustafa, B. Akera, Y. Bengio, H. Radi, H. Larochelle, and D. Rolnick, D. (2023). Satbird: A dataset for bird species distribution modeling using remote sensing and citizen science data. *Advances in Neural Information Processing Systems*, 36, 75925-75950.
- Truong, M. X. A., and R. Van der Wal, (2024). Exploring the landscape of automated species identification apps: Development, promise, and user appraisal. *BioScience*, 74(9), 601-613.
- Van Horn, G., S. Branson, R. Farrell, S. Haber, J. Barry, P. Ipeirotis, and S. Belongie, (2015). Building a bird recognition app and large scale dataset with citizen scientists: The fine print in fine-grained dataset collection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 595-604).
- Van Horn, G., O. Mac Aodha, Y. Song, Y. Cui, C. Sun, A. Shepard, and S. Belongie, (2018). The iNaturalist species classification and detection dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8769-8778).
- Wimmer, J., M. Towsey, P. Roe and I. Williamson, (2013). Sampling environmental acoustic recordings to determine bird species richness. *Ecological Applications*, 23(6), 1419-1428.
- Xie, J., Y. Zhong, J. Zhang, S. Liu, C. Ding, and A. Triantafyllopoulos, (2023). A review of automatic recognition technology for bird vocalizations in the deep learning era. *Ecological Informatics*, 73, 101927. <https://doi.org/10.1016/j.ecoinf.2022.101927>
- Zeeman, S. (2023). *Why are bird surveys so important?* Carbon Rewild. <https://carbonrewild.com/why-are-bird-surveys-so-important/#:~:text=It%20can%20provide%20valuable%20insights,disturbing%20them%20or%20their%20habitats>

Received Date : 21 April 2025

Accepted Date : 08 June 2025