DEVELOPING A CONVOLUTIONAL NEURAL NETWORK FOR BIRD CALL IDENTIFICATION: A CASE STUDY AT INTAKA ISLAND

by

Maminiaina Rakotovao[1, 2]

1. African Institute for Mathematical Science, South Africa
2. Stellenbosch University, South Africa

INTRODUCTION

Machine learning, a subset of artificial intelligence, has revolutionized various research domains by enabling computers to learn tasks without explicit programming. Its applications span from genomic sequence analysis to ecology and conservation biology.

With the exponential growth of data, machine learning has become increasingly indispensable in analyzing complex datasets that were previously infeasible to process due to computational limitations. In ecological research, the study of birds has garnered significant attention as scientists seek to unravel their communication patterns, behaviors, and ecological roles.

Problem Statement: Despite the abundance of data collected in ecological studies, analyzing bird sounds presents considerable challenges such as background noise, species variability, and the necessity for high-quality recording equipment. However, advancements in technology and machine learning algorithms offer promising avenues to address these challenges and extract meaningful insights from audio data.

The primary challenge addressed in this project is the development of accurate and efficient methods to analyze bird sounds and detect the presence of bird calls within audio recordings.

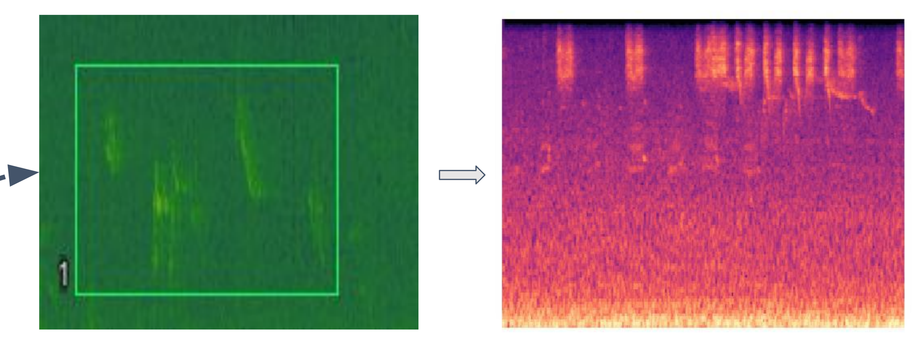
Objectives: The main objective of this project is to leverage machine learning techniques, specifically convolutional neural networks (CNN), to analyze bird sounds and identify the presence of bird calls in audio files.

By achieving this objective, we aim to contribute to the understanding of local ecosystems and wildlife populations, thereby enhancing knowledge of biodiversity dynamics. Additionally, our project aims to advance ecological science by providing insights into bird behaviors and their ecological significance.

MATERIALS AND METHODS

Data: Recorded at different places at Intaka Island using a Raspberry Pi, an audiomoth and handle recorder. It is configured to store data on a server. The data can be access remotely using Filezilla.

Annotations: Sonic visualizer was used. For bird call, we set the label as 1 and 0 for no bird call.



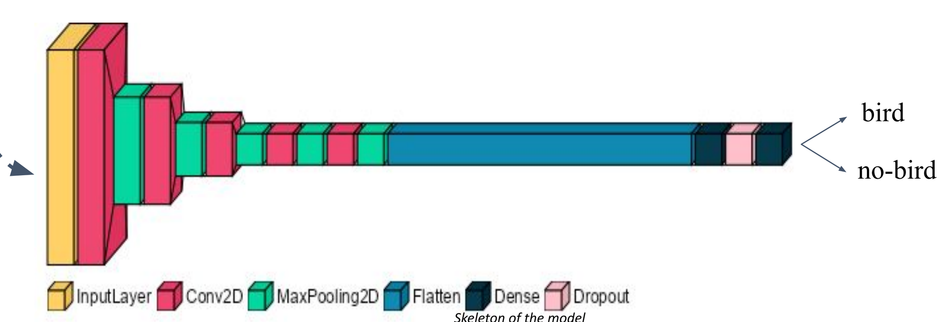
Annotation with sonic visualizer

Spectrogram of a segmented audio

Preprocessing: Split audio into 4-seconds segments and transform them into spectrograms. (Hann window = 1024, hop length = 256, number of mells = 128)

Data augmentation: To address the issue of unbalanced data between the two classes, we implemented a random flip of the spectrograms.

Training data: 7000 spectrograms (3500 for bird call and 3500 for no bird call, after data augmentation)
Validation data: 3000 spectrograms
Testing data: 2000 spectrograms



Skeleton of the model

Model: We created a model that takes the spectrogram as inputs. It uses ReLU in convolutional and fully connected layer. The final layer employs Softmax activation for output.

RESULTS & DISCUSSION

The model demonstrated convergence with a high average accuracy of 93.13% on the test dataset.
It can predict bird call with 93% certainty and no bird call with 93% certainty.

A balanced precision of 93% and f1-score of 93% suggests that a simple model can effectively capture relevant features from spectrogram.

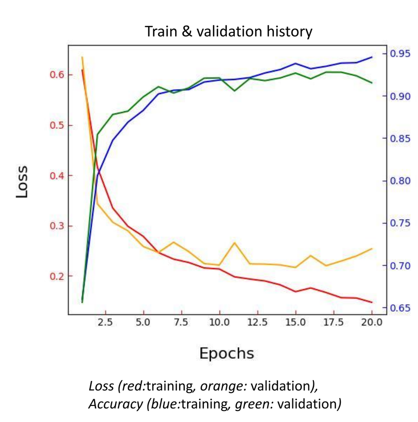
Based on those values, the model shows a good performance and generalize well with audio data taken from real world environment.

	precision	*recall*	*f1-score*	*support*
0.0	0.93	0.93	0.93	1463
1.0	0.93	0.94	0.93	1537
accuracy			0.93	3000
macro avg	0.93	0.93	0.93	3000
weighted avg	0.93	0.93	0.93	3000

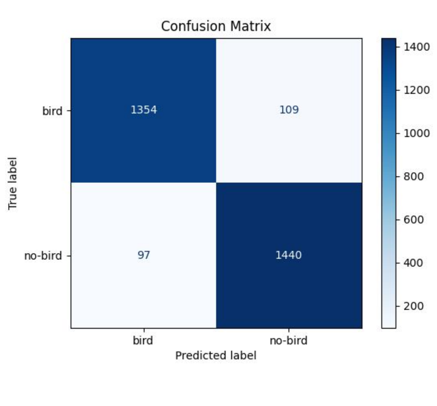
However, the model robustness is limited by computational constraints in augmenting data extensively and the challenges in data quality.

Transfer learning could optimize performance by fine-tuning the model's parameters on a new dataset.

Future work will focus on expanding the model's capabilities to classify bird species accurately and integrate temporal and geographical data to deepen our understanding of avian ecosystems and improve conservation practices.



Loss (red:training, orange: validation), Accuracy (blue:training, green: validation)



CONCLUSION

Our study shows that using data augmentation and training models helps us accurately find bird calls. We reached an accuracy of 93.13%, showing that machine learning can be helpful in ecological research.

Looking ahead, our results set the stage for more research into better ways to monitor birds. We need to deal with computer limits and data problems.

We machine learning, we can make better models that give us important information about birds. This is crucial for protecting wildlife and keeping our planet healthy.

REFERENCES

Jeantet, L., Dufourq, E., 2023. Improving deep learning acoustic classifiers with contextual information for wildlife monitoring. Ecological Informatics, 77, 102256.

https://github.com/ramaminiaina/DL-for-Ecology