Explainable Audio-Based Machine Learning for Bird Species Identification in Child-Friendly Educational Systems

D.M.S.Y.Jayasundara [1], I.W.S.C.Rajaguru [2], M.A.D.H.Wijewardhana [3], M.M.S.Hansaja [4]

Ms. Buddhima Lakchani[5], Mr. Chameera De Silva[6]

[1,2,3] Department of Data Science, [4]Department of Software Engineering, [6] Department of Data Science

Sri Lanka Technological Campus (Pvt) Ltd, Ingiriya Road, Meepe, Sri Lanka

1. Introduction

Identifying bird species is crucial for environmental education and biodiversity awareness, especially in Sri Lanka where the diversity of birds is very high. The traditional method of identification is based on visual observation, but many bird species can be recognized better by their voices. Hence, sound-based learning is an engaging and easy way to introduce children to birds and nature [1], [2]. The latest developments in audio-based machine learning allow for the automatic classification of bird sounds based on features like MFCCs and Mel-spectrograms using CNN-based models [3], [4]On the other hand, the majority of existing systems lack explainability, hence making them less suitable for child-centered learning settings [5]

1.1. Summary of literature

Previous research on classifying bird sounds mostly focused on machine learning and deep learning techniques that used audio features like MFCCs and Mel-spectrograms followed by various types of classifiers such as CNNs and lightweight models [1], [4]. These methods not only reached high accuracy but also frequently relied on public datasets like Xeno-canto. However, the majority of such systems still act as black-box models that lack interpretability and therefore are usually not considered as being very relevant to bird species in Sri Lanka or educational applications [5], [6]. This situation points to a need for explainable and locally relevant bird sound classification systems targeted at learning that are not available yet.

1.2. Problem definition

While current bird sound classification systems are indeed precise, their black-box characteristics render them inappropriate for children's education as they do not provide clear and simple explanations [5]. Moreover, not giving enough attention to the most common Sri Lankan birds lessens their local educational importance [7] , [6]. Consequently, there is a demand for a bird sound identification system that is not only explainable and child-friendly but also accurate predictions with simple visual explanations appropriate for young Sri Lankan learners.

2. Goal of the project

The goal of this project is to develop a child-friendly educational system that can identify bird species from their sounds [1], [2], [3]. While there are many aspects to bird identification, this project will focus specifically on common local bird species and their audio calls. The system will also provide simple explanations of its predictions so that children can engage with and learn from the results, rather than just seeing the answer. [5], [8],[13]

3. Aims and objectives

The aim of this project is to develop a child-friendly system that identifies bird species using bird sounds and supports learning.

- To collect and preprocess bird sound recordings from public sources and local environments
- To design and train a deep learning model to identify bird species using audio features.
- To develop and evaluate a user-friendly mobile application interface that allows children to record or upload sounds and learn about birds.

4. Proposed methodology

Bird audio data were collected from Xeno-canto [7] Sri Lankan Nature Sounds [9] and field recordings made in the Kegalle region. The audio tracks were processed by removing background noise, adjusting the volume, and breaking longer clips into shorter segments [10]. From these recordings, spectrograms and Mel-frequency cepstral coefficients were extracted so the model could recognize important sound patterns. A deep learning model, such as a Convolutional Neural Network, was trained to identify bird species [11],To make the system more educational, a feature was added to explain how predictions were made, showing which parts of the bird sounds influenced the results.

5. Resource requirements

To implement the proposed explainable bird species identification system based on audio, specific hardware, software, and dataset resources are necessary for audio preprocessing, machine learning model training, explainability features, and developing a child-friendly interface. [1], [5]

1. Hardware Requirements
- Minimum requirements for a desktop or laptop include an Intel Core i5 or AMD Ryzen 5 processor, at least 8 GB of RAM, and sufficient storage for audio datasets and trained models.
- A microphone or mobile device is required to record bird sounds, while an internet connection is required for downloading datasets and using cloud services. [2], [3]

2. Software Requirements
- The project utilizes various technologies for audio processing, feature extraction, and machine learning. It runs on operating systems like Windows, Linux, or macOS and employs Python as the programming language. [2], [11]
- Key libraries include Librosa for audio loading and preprocessing, with MFCC techniques and the openSMILE toolkit for feature extraction. [4] , [8] [10]

- Machine learning frameworks such as Convolutional Neural Networks (CNN) and Extreme Learning Machines (ELM) are used for analysis [1] , [3] , [11]. Explainable AI is supported through Grad-CAM for visualization CNN predictions. [5]
- The frontend is developed using the Flutter framework, while Firebase functions as the backend for audio handling, data storage, and model inference.

3. Dataset Requirements
Dataset of bird sounds that can be found online include Xeno-canto and Sri Lanka [4]Nature Sounds [7], [6] , [9], [10] Bird sound recordings taken in the field were obtained from the Kegalle region. Only common birds of Sri Lanka are present in these datasets and it is suitable for children's educational applications [6]., [7]

6. Timeplan
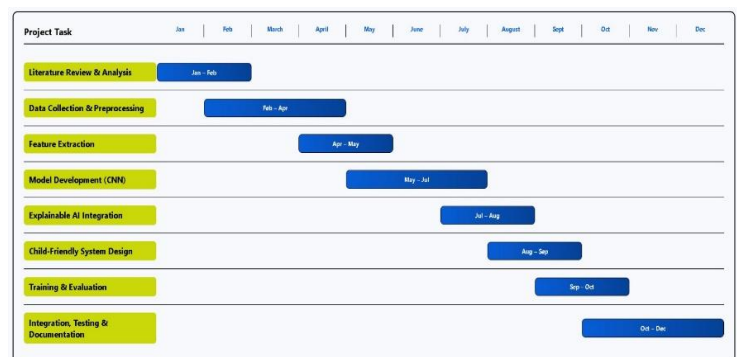
7. Conclusion

This project proposes a child-friendly machine learning system that distinguishes birds by their sounds. It combines audio pre-processing, feature extraction, CNN-based classification, and interpretable AI methods for accurate and understandable predictions [3] [5]. By utilizing regionally relevant bird sound datasets and a user-friendly interface, the system aims to enhance environmental awareness and education about biodiversity while laying a foundation for future developments. [6], [7]

8. References

[1] A. E. Mehyadin et al., "Birds Sound Classification Based on Machine Learning[Link](#)

[2] Y. Jadhav, V. Patil, and D. Parasar, "Machine Learning Approach to Classify Birds on the Basis of Their Sound," 2020.

[LInk](#)

[3] E. Sprengel et al., "Audio Based Bird Species Identification using Deep Learning Techniques," 2016.

[4] S. Carvalho and E. F. Gomes, "Automatic Classification of Bird Sounds Using MFCC and Mel Spectrogram Features," 2022.

[5] R. Heinrich et al., "An Interpretable Deep Learning Model for Bird Sound Classification," 2024.

[6] S. Gunawardena et al., "Sri Lankan nature sound recordings for biodiversity monitoring," Journal of Ecoacoustics, vol. 5, no. 1, pp. 1–10, 2021.

[7] Xeno-canto Foundation, "Xeno-canto: Sharing bird sounds from around the world," 2023. [Online]. Available: https://www.xeno-canto.org

[8] B. Logan, "Mel Frequency Cepstral Coefficients for music modeling," in Proc. Int. Symp. Music Information Retrieval (ISMIR), 2000.

[9] Sri Lankan Nature Sounds, *Sri Lankan Nature Sounds Database*, 2023. [Online]. Available: https://www.slnaturesounds.lk

[10] T. Kinnunen and H. Li, "An overview of text-independent speaker recognition: From features to supervectors," *Speech Communication*, vol. 52, no. 1, pp. 12–40, Jan. 2010.

[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2012, pp. 1097–1105.