



# Bird Species Identification From Audio Using Deep Learning

<sup>1</sup> Mrs. K. Sai Jyostna, <sup>2</sup> Alamuri Yamini, <sup>3</sup> Garikimukku Vamsi,

<sup>4</sup> Ballamudi Divya Sambhavi, <sup>5</sup> Sanagala Naga Charan Teja

<sup>1</sup> Assistant Professor, <sup>2</sup> Student, <sup>3</sup> Student, <sup>4</sup> Student, <sup>5</sup> Student

<sup>1,2,3,4,5</sup> Computer Science and Engineering,

<sup>1</sup> Dhaneekula Institute of Engineering & Technology, Vijayawada, India

**Abstract:** Bird species classification using audio signals is a growing field in bioacoustics, with significant potential for improving biodiversity monitoring and conservation efforts. This project aims to develop an automated system that classifies bird species based on their vocalizations using deep learning models like OpenL3 and YAMNet. OpenL3, trained to extract high-level audio features, and YAMNet, a lightweight audio event classifier, are finetuned to detect bird sounds from spectrogram representations. These models effectively extract high-level audio features and capture intricate patterns in bird calls. Bird audio is preprocessed into spectrograms to enhance the models ability to detect temporal and spectral features. This automated approach minimizes the need for manual species identification, which can be time consuming and prone to human error. The solution is intended to support ecological research by providing a more efficient and accurate method for recognizing and cataloging bird species, contributing to conservation efforts and a better understanding of biodiversity.

**Index Terms** - Deep learning, OpenL3, YAMNet, Bird species classification, spectrogram, Audio signals, Bioacoustics, Bird vocalizations, Bird sounds.

## I. INTRODUCTION

The introduction of artificial intelligence (AI) in bioacoustics has revolutionized wildlife monitoring, biodiversity conservation, and ecological research. Conventional bird species identification is based on human observation and expert opinion, while machine learning models provide a scalable and efficient solution through audio signal-based automated classification. This research investigates deep learning methods for bird species classification, resulting in the creation of a Django web application that can identify species from recorded calls. In contrast to conventional models that are challenged by environmental noise and variations in vocalization, this application guarantees correct species identification using sophisticated audio feature extraction and deep learning methods. The general goal is to evaluate the performance of deep learning models, namely OpenL3 for feature extraction and Convolutional Neural Networks (CNNs) for classification, in learning and discriminating complex patterns of bird calls. As bird calls exhibit sequential patterns and acoustic dependencies, deep learning models are well-equipped to identify complex relationships between frequencies, timbres, and temporal patterns. Based on a carefully curated dataset from Xeno-canto, the system trains essential acoustic features that distinguish one species from another to provide accurate classification. The system allows users to upload bird recording audio, which is processed using a pipeline that includes spectrogram conversion, OpenL3 feature extraction, and CNN-based classification. With an interactive web interface powered by Django, the platform offers accessibility to researchers, conservationists, and citizen scientists without any need for specialized technical expertise. A number of features set this work apart from other projects. For one, it is audio-based classification as opposed to traditional image-based techniques, closing an important gap in bioacoustics research. Two, Django integration provides for an easy-to-use, real-time classification process. As opposed to independent bioacoustic analysis software which demands intensive

manual processing and skilled intervention, this web application democratizes the use of AI-backed species identification. Scholarly, this study adds to the current literature on AI usage in biodiversity monitoring through the combination of machine learning, acoustic analysis, and web-based implementation into a single system. The research includes AI usage in bioacoustics, a survey of current deep learning techniques for bird classification, comprehensive descriptions of OpenL3 and CNN architectures, dataset preprocessing methods, and Django deployment. It ends with model performance evaluation, limitations, and possible enhancements. By improving AI-based biodiversity monitoring, the research points out the contribution of deep learning towards conservation. The system lays a basis for subsequent AI-aided bioacoustic analysis and places technology at the forefront in ecological studies and wildlife conservation.

## II. LITERATURE SURVEY

Species classification in birds was historically based on professional auditory analysis, requiring great expertise in recognizing vocal differences (Kroodsma & Miller, 1996; Marler & Slabbekoorn, 2004). Computational approaches in the past employed hand-designed acoustic features such as MFCCs and STFT but had difficulty with noise, variability in calls, and high preprocessing requirements (Kogan & Margoliash, 1998; Laiolo, 2010).

Machine learning brought in automation, with SVM and k-NN models labeling bird songs using extracted features (Briggs et al., 2012; Stowell & Plumbley, 2014). But these approaches were not scalable and had issues with big and diverse datasets (Somervuo et al., 2006).

CNNs were especially instrumental in transforming bird call classification through deep learning, with features being automatically extracted from spectrograms to enhance accuracy even in noisy conditions (Salamon et al., 2017; Stowell et al., 2019). OpenL3 took classification further by deep audio embeddings learning, which surpassed MFCC-based methods in challenging acoustic conditions (Cramer et al., 2019; Goëau et al., 2021).

Recent developments consolidate AI into web applications, providing species identification to researchers and conservationists. Systems based on Django allow for real-time classification using easy-to-use interfaces, with cloud-based deployment maximizing massive data processing (Wimmer et al., 2018; Kahl et al., 2021).

This study extends these advances by developing a Django web application for classifying bird species. It analyzes user-provided recordings, extracts features with OpenL3, and uses a CNN for prediction. Through the promotion of AI-powered bioacoustic analysis, this research supports biodiversity monitoring and conservation through an accessible and scalable species identification method (Hadjipantelis et al., 2022).

## III. METHODOLOGY

### 1. Data Collection

Bird species audio recordings are gathered from Xeno-canto, a public repository of bird call recordings. Recordings of ten species are utilized, which allow for diversity in the acoustic features. Audio recordings are of different lengths, quality, and noise, making preprocessing and classification more challenging. The metadata for each recording, such as location, recording date, and weather conditions, are also considered to enhance the model's ability to generalize across different acoustic conditions.

### 2. Data Preprocessing

- Preprocessing steps are carried out prior to feature extraction to enhance the quality of the dataset. They are:
- Noise Reduction: Removing background noise by spectral subtraction techniques and adaptive filtering.
- Resampling: Bringing all the recordings to a single sample rate (say, 16kHz) for consistency.
- Segmentation: Breaking down long audio files into short, informative segments to extract important parts of bird calls.
- Silence Deletion: Deleting silent sections to improve feature extraction efficiency.
- Normalization: Amplitude scaling of signals for standardization in different recordings.

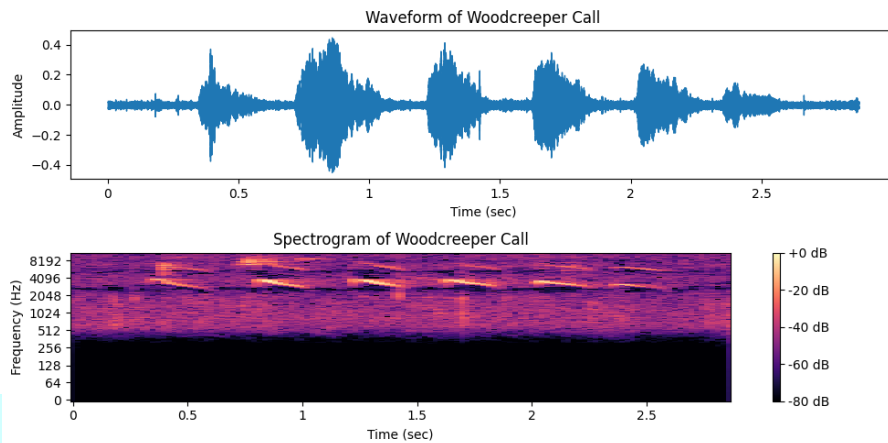
- **Data Augmentation:** Applying techniques such as time stretching, pitch shifting, and background noise addition to improve model robustness and prevent overfitting.

### 3. Feature Extraction with OpenL3 and YAMNet

Two deep learning models are employed in order to extract relevant audio features:

3.1. OpenL3: Transforms audio signals into representations in the form of spectrograms and outputs deep embeddings. It is

trained on a wide variety of acoustic data and thus is able to capture both temporal and spectral information.



3.2. YAMNet: A deep learning model that was trained on a large-scale dataset for sound event detection. It classifies the environmental sounds and gives embeddings that augment OpenL3's feature extraction.

The system gets to enjoy an augmented feature set when it uses both OpenL3 and YAMNet embeddings, enhancing classification accuracy. The features are then input into the classification model.

### 4. Classification Model

A Convolutional Neural Network (CNN) is trained to process the extracted features. The architecture involves several convolutional layers followed by fully connected layers to classify input spectrograms into the corresponding bird species.

The network is arranged as follows:

- **Convolutional Layers:** Extract hierarchical features from spectrograms.
- **Batch Normalization:** Normalize activations for faster convergence.
- **Pooling Layers:** Reduce dimensionality and computational complexity.
- **Fully Connected Layers:** Transform learned features into class probabilities.
- **Softmax Output Layer:** Outputs probability distribution across bird species.

## 5. Implementation and Training

Training is performed using labeled spectrogram data with categorical cross-entropy loss and Adam optimizer. Techniques of data augmentation are used for enhancing generalization and robustness. The training process includes:

- Hyperparameter Tuning: Modifying learning rates, batch sizes, and dropout rates.
- Cross-Validation: Generalizing over various datasets.
- Regularization Techniques: Avoiding overfitting through dropout layers.
- Feature Fusion: Fusing OpenL3 and YAMNet embeddings for enhanced classification.

## IV. RESULTS AND DISCUSSION

The trained CNN model is highly accurate in classification, validating the efficacy of deep learning application in bird call analysis. Comparative analysis with conventional MFCC-based approaches indicates that OpenL3 and YAMNet feature extraction improves the performance of the model. The research also addresses challenges like background noise interference and dataset imbalance, presenting potential solutions to reduce these problems.

In order to assess performance, the following measures are taken into account:

**Accuracy:** Indicates overall correct classification.

**Precision and Recall:** Assess class-wise performance.

**F1-Score:** Scales precision and recall for confident estimation.

**Confusion Matrix:** Offers insights into misclassification behavior.

Results show that OpenL3 and YAMNet-based spectrogram analysis is better compared to traditional feature extraction techniques. Nevertheless, the occurrence of overlapping calls of birds and weather conditions is still a problem.



## V.CONCLUSION

This paper proposes a robust deep learning-based approach to bird species classification from audio signals. OpenL3 and YAMNet surpass conventional MFCC-based methods in feature extraction. By transforming bird calls into spectrograms and using a CNN classifier, the system is highly accurate, contributing to biodiversity conservation and ecological research. In spite of difficulties such as background noise and dataset constraints, the model generalizes effectively to varied bird calls. This work adds to eco-acoustics by showing the capability of deep learning in monitoring wildlife, providing a scalable and automated method for species identification and conservation.

## VI. ACKNOWLEDGMENT

We truly acknowledge our colleagues and faculty members for their ongoing assistance and insightful ideas that have contributed greatly to this study. Their feedback and technical expertise have helped shape this study in a huge way. We further recognize the work of the general academic and research community, whose efforts have established a solid ground for this undertaking. We are also thankful for the debates and collaborative work that have assisted in sharpening various elements of this research. Lastly, we offer our appreciation to the open-access bird sound databases, especially Xeno-canto, for sharing critical information, and to everyone who has funded and spurred us along the way. Without the resource and opportunity that has enabled inquiry and innovation in AI-based bioacoustic analysis, this study would not have been possible.

## VII. REFERENCES

- [1] Badi, A., Ko, K., & Ko, H. (2019). They identified how combining PNCC and strong Mel-log filter bank features can improve bird sound classification. This article published in the Journal of Acoustical Society of Korea offers information about feature extraction methods of bird calls.
- [2] Chen, Y., Dai, X., Liu, M., Chen, D., Yuan, L., & Liu, Z. (2020). This paper presents dynamic convolution, a novel method that uses attention mechanisms on convolution kernels. The results were reported at CVPR 2020, noting its applicability to deep learning tasks.
- [3] Huang, Y., & Basanta, H. (2021). Their research is centered around the identification of endemic bird species by employing deep learning models. This paper, published in IEEE Access, exemplifies the capabilities of AI in identifying bird species through superior neural networks.
- [4] Incze, Á., Jancsó, H., Szilágyi, Z., Farkas, A., & Sulyok, C. (2018). The article discusses a convolutional neural network-based method for bird sound recognition. It was presented at the IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY), highlighting CNN usage in bioacoustics.
- [5] Knight, E. C., Hernandez, S. P., Bayne, E. M., Bulitko, V., & Tucker, B. V. (2020). Their work highlights the fact that preprocessing spectrogram parameters can tremendously enhance bioacoustic classification accuracy with CNNs. This publication in Bioacoustics is beneficial towards maximizing machine learning methods in ecology.