BIOACOUSTIC: BIRD AUDIO CLASSIFICATION

A.Devanshu Singh, B.Kishan Singh,
C. Farhan Khan
Department of Computer Science & Engineering,
Under Supervision of Ms. Shaba Irram (Assistant Professor) Integral University, Lucknow, India

Abstract: Bioacoustics play a vital role in biodiversity monitoring and ecological research, serving as indicators of environmental health and biodiversity. Monitoring bird populations is therefore essential for conservation efforts, yet traditional field-based observation methods are often labour-intensive, time-consuming, and limited in scope. The system processes bird vocalizations, extracts acoustic representations, and performs species classification with high accuracy. Experimental evaluation shows reliable performance and scalability for real-world applications.

Keywords: Machine learning, MFCC, CNN, Deep Learning, Audio feature extraction.

I. INTRODUCTION

As indicators of biodiversity, birds are always worthy of being surveyed for biodiversity conservation. It was verified that bird calls are relatively stable as they present observable acoustic features among bird species. Many kinds of research considered identifying bird species or individuals by analysing their calls manually, such as analysing the waveform, spectrogram, Mel−spectrogram and Mel−frequency cepstral coefficients using bird−call clips. Nonetheless, it was time−consuming if the audio data expanded a lot, which is inefficient and biased in subjective ways. Particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have further enhanced classification accuracy by capturing complex temporal and spectral dependencies in bird calls. It ends with model performance evaluation, limitations, and possible enhancements. 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 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

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 labelling 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). Recent developments emphasized deployment and accessibility. Web-based systems, often built with frameworks such as Django, enabled real-time classification through intuitive interfaces, while cloud infrastructures supported large-scale data processing and collaborative research (Wimmer et al., 2018; Kahl et al., 2021). These advances bridged machine learning research with ecological practice, making bioacoustic tools more accessible to conservationists.

Building on these foundations, the present study employs the Librosa library for audio preprocessing, including waveform handling, spectrogram generation, and mel-frequency feature extraction. These features are then integrated into TensorFlow/Keras pipelines, where CNN architectures are trained for bird species classification. By combining efficient audio processing with deep learning frameworks, this research develops a scalable and accessible system for bioacoustic analysis. The integration into a web application further promotes AI-powered biodiversity monitoring, supporting conservation efforts through real-time, researcher-friendly tools (Hadjipantelis et al., 2022).

III. METHODOLOGY

1. Data Collection

The dataset for bird audio classification was compiled from publicly available bioacoustic repositories and field recordings. Sources included open-access platforms such as Xeno-Canto and other libraries, which provide diverse bird vocalizations across species and habitats. Each audio file was stored in standardized formats (e.g., WAV, MP3) to ensure compatibility with downstream processing. And further we converted them into Mel-Frequency Cepstral Coefficients (MFCCs) which are visual like representations of sound. A Convolutional Neural Network (CNN) was then trained on these MFCC images to learn and recognize the unique audio patterns of each bird species. The model was evaluated for accuracy on unseen data to ensure reliable performance.

2. Data Preprocessing

Before extracting features, several preprocessing steps are applied to improve the quality and consistency of the dataset.

- Noise Reduction & Trimming: Silence and irrelevant background sounds were minimized to emphasize bird vocalizations.
- Resampling: All audio files were resampled to a uniform sampling rate (e.g., 16 kHz) for consistency.
- Segmentation: Breaking down long audio files into short to improve training efficiency and model generalization.
- Silence Deletion: Deleting silent sections to improve feature extraction efficiency.
- Normalization: Features were scaled to ensure uniformity across samples.
- Data Augmentation: To improve model generalization and robustness, augmentation techniques were applied such as time shifting, pitch shifting, adding background noise.

3. Feature Extraction

Once the audio recordings were pre-processed, the next step was to extract meaningful features that could represent the acoustic characteristics of bird vocalizations. Using the Librosa library, raw audio signals were converted into spectrograms, which provide a time–frequency representation of the recordings. From these spectrograms, several features were derived to capture both static and dynamic properties of the calls.

4. Classification Model

- Input Layer: Accepts spectrograms or MFCC matrices.
- Convolutional Layers: Multiple 2D convolutional layers with ReLU activation to capture local time–frequency features.
- Pooling Layers: Max-pooling layers to reduce dimensionality while retaining salient acoustic patterns.
- Dropout Layers: Introduced to prevent overfitting and improve generalization.
- Flatten Layer: Converts pooled feature maps into a one-dimensional vector.
- Dense Layers: Fully connected layers with ReLU activation integrate learned features into higher-level representations.
- Output Layer: A softmax activation function produces probability distributions across bird species classes.

5. Model Implementation:

Two architectures were tested:

1. Feed-Forward Neural Network (FFNN) with one hidden layer of 128 neurons and ReLU activation.

2. Convolutional Neural Network (CNN) with multiple convolutional and pooling layers for spectrogram input.

- Both models concluded with a softmax output layer to classify among bird species.

6 Training Procedure:

- Optimizer: Adam optimizer was used for efficient gradient descent.
- Loss Function: Categorical cross-entropy loss was applied to measure prediction error.
- Batch Size & Epochs: Training was conducted in mini-batches, with early stopping to prevent overfitting.
- Validation Split: A portion of the dataset was reserved for validation to monitor generalization.
- Regularization: Dropout layers and learning rate scheduling were employed to improve robustness.

IV RESULTS AND DISCUSSION

The trained Convolutional Neural Network (CNN) model achieved high accuracy in bird call classification, validating the efficacy of deep learning approaches in bioacoustic analysis. By leveraging spectrogram-based representations, the CNN was able to capture subtle acoustic variations across species, outperforming simpler feed-forward baselines.

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.

V. CONCLUSION

This project establishes a functional and effective deep learning pipeline for Bioacoustic bird audio classification By transforming bird calls into spectrograms and using a CNN classifier, the system is highly accurate, contributing to biodiversity conservation and ecological research. The CNN consistently outperformed the feed-forward neural network baseline, validating its ability to capture complex time–frequency structures inherent in bird calls.

VI. ACKNOWLEDGMENT

We would like to express our sincere gratitude to Integral University, Lucknow, India, for providing the academic environment, facilities, and resources necessary to carry out this research. We extend special thanks to the Department of Computer Science and Engineering for their guidance and support throughout the project. We are deeply thankful to our supervisor, whose mentorship, encouragement, and valuable insights in the fields of machine learning and bioacoustics greatly enhanced the quality of this work. 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] Xeno-canto Foundation, “Xeno-canto: Bird sounds from around the world,” Available: https://www.xeno-canto.org.

[2] B. McFee, C. Raffel, D. Liang, D. P. W. Ellis, M. McVicar, E. Battenberg, and O. Nieto, “librosa: Audio and music signal analysis in Python,” in *Proceedings of the 14th Python in Science Conference (SciPy 2015)*, pp. 18–25, 2015.

[3] S. Davis and P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 357–366, Aug. 1980. (MFCCs foundational paper).

[4] J. Salamon and J. P. Bello, “Deep convolutional neural networks and data augmentation for environmental sound classification,” *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279–283, Mar. 2017.

[5] S. Kahl, C. M. Wood, M. Eibl, and H. Klinck, “BirdNET: A deep learning solution for avian diversity monitoring,” *Ecological Informatics*, vol. 61, p. 101236, 2021.