STM JOURNALS

Journal of
Aerospace Engineering & Technology

ISSN: 2231-038X (Online)
ISSN: 2348-7887 (Print)
Volume 14, Issue 2, 2024
May–August
DOI (Journal): 10.37591/JoAET

https://journals.stmjournals.com/joaet

Review JoAET

Avian Echoes: Convolutional Neural Network for Bird Vocalization Detection

Ved Chaudhari[1,*], Siddharth Chhajed[2], Piyush Jain[3], Adarsh Kamble[4], Dnyaneshwar Kapse[5]

Abstract

Bird species identification is a complex task within ornithology that demands advanced technological solutions. This research presents an approach leveraging Convolutional Neural Networks (CNNs) for bird species recognition based on identification of bird sound, each employing unique datasets and methodologies. The objective involves a two-stage identification process, beginning with the construction of an ideal dataset. The crucial step involves converting 1D audio waveforms to 2D spectrograms, enhancing CNNs' ability to analyze temporal and frequency features simultaneously. Spectrograms are generated for each sound clip to capture essential features, contributing to advancements in accurate and effective automatic bird species classification in ornithology. In the next step, a neural network, specifically a Convolutional Neural Network (CNN), processed spectrograms as input. CNN analyzed these features, conducting classification on the sound clip and accurately recognizing the bird species associated with the input audio. This underscores CNN's adeptness in discerning intricate patterns within spectrograms. Our website analyzes user-input audio recordings to predict bird species. Results, displaying identified bird species names and corresponding spectrograms, demonstrate the practical application of the automated recognition system.

General Terms: Model Prediction, Pattern Recognition, Feature Extraction.

Keywords: Ornithology, CNN, Spectogram, Bird species recognition, Machine learning.

INTRODUCTION

The world's rich biodiversity, encompassing diverse living creatures, includes birds that have captivated human interest for centuries due to their distinctive features and ecological significance. Unfortunately, many bird species confront existential threats from human activities like deforestation, climate change, and pollution, jeopardizing their habitats and food sources. Safeguarding endangered bird species is now a global imperative, presenting a challenge in accurate identification. Traditional methods relying on visual observations and expert knowledge prove time-consuming and limited by the availability of ornithology experts. In response to the urgent need for innovative monitoring amid environmental changes and declining biodiversity, machine learning in bird audio classification emerges as a promising solution. This approach, leveraging birdsongs as natural identifiers, provides real-time ecological surveillance, offering non-intrusive insights into avian presence and behavior, addressing crucial gaps in conventional monitoring methods. The goal is to enhance our understanding of avian dynamics, facilitating proactive conservation efforts amidst ongoing environmental challenges. The automated identification of bird calls from continuous environmental recordings is a

***Author for Correspondence**
Ved Chaudhari
E-mail: vedc2853@gmail.com

[1,2,4]Student, Department of Computer Engineering, Rajiv Gandhi Institute of Technology, Kharghar, Navi Mumbai, Maharashtra, India
[3]Professor, Department of Computer Engineering, Rajiv Gandhi Institute of Technology, Kharghar, Navi Mumbai, Maharashtra, India
[5]Assistant Professor, Department of Computer Engineering, Rajiv Gandhi Institute of Technology, Kharghar, Navi Mumbai, Maharashtra, India

Receiving Date: July 29, 2024
Accepted Date: July 31, 2024
Published Date: August 04, 2024

Citation: Ved Chaudhari, Siddharth Chhajed, Piyush Jain, Adarsh Kamble, Dnyaneshwar Kapse. AvianEchoes: Convolutional Neural Network for Bird Vocalization Detection Journal of Aerospace Engineering & Technology. 2024; 14(2): 26–37p.

crucial addition to research methodologies in ornithology and biology. Often, these recordings are truncated, making manual conventional methods unreliable. Manual inspections of spectrograms are error-prone and involve multiple experts, necessitating the development of automated techniques. The significance of such systems extends beyond scientific research, presenting substantial commercial potential due to the popularity of bird watching as a hobby in many countries. International programs actively support advancements in bioacoustics signal processing and pattern recognition, contributing to the growing interest and progress in this field. This research proposed a technique which involves the use of sound processing and convolutional neural networks to automate the entire process of bird sound identification. Convolutional Neural Networks (CNNs) play a crucial role in analyzing spectrograms, which represent audio data by showing frequency distribution over time. The audio is first transformed into a spectrogram, and CNNs extract features hierarchically through convolutional and pooling layers. These networks are trained on labeled bird audio datasets, optimizing filter weights to reduce prediction errors. Flattening and fully connected layers capture global relationships, culminating in a final output layer that predicts species probabilities.

A user-friendly Graphical User Interface (GUI) has been meticulously crafted for our system.Users simply need to upload an audio recording of the bird, and our model will conduct a comprehensive analysis by testing the audio against the pre-existing training dataset.

LITERATURE REVIEW

[1] “In a study by Incze, Szilagyi, Farkas, and Sulyok in 2018 they explored avian audio classification using machine learning techniques like CNNs and transfer learning, alongside spectrograms. However, the approach faces hurdles with increased bird species and struggles with noise in environments due to color map sensitivity and limited noise reduction capabilities, necessitating crucial improvements for practical use.”

[2] “In a study Chandu B, Akash Munikoti, Karthik S Murthy, Ganesh Murthy V, and Chaitra Nagaraj. in 2020 they proposed a bird species recognition system using transfer learning with an AlexNet model, reaching 97% accuracy, but facing issues with noise sensitivity and limited species coverage, prompting improvements for wider real-time applicability.”

[3] “In 2021, Shekar and Lin proposed a bird audio detection system using ResNet34 with transfer learning, classifying 10,000 ten-second audio clips from UK locations via MelSpectrogram representations. However, solely relying on the Warblr dataset might restrict bird species diversity representation, hinting at the need for broader species and non-bird audio augmentation for improved real-world adaptability”.

[4] ” In Mario Lasseck's 2018 study, a CNN-based approach was proposed for bird detection in audio recordings, emphasizing data augmentation for performance improvement. However, its focus on binary bird presence detection may hinder species-specific data provision, with challenges in generalizing across different recording conditions and habitats, prompting the need for further research.”

[5] “Elias Sprengel, Martin Jaggi, Yannic Kilcher, and Thomas Hofmann, winners of the 2016 BirdCLEF challenge, achieved single labeling accuracy of 68.6% and multiple labeling accuracy of 55.5% using a CNN architecture with five convolutional layers and one dense layer, utilizing spectrograms with preprocessing to isolate bird sounds from noise..”

PROPOSED SYSTEM

Analysis

Data Selection

Our first step involves carefully selecting bird audio recordings for our dataset. A specific criteria is applied, including minimum count thresholds and quality ratings, to ensure the data's reliability and relevance and high-quality recordings meeting our criteria proceed to further processing.

Preprocessing

In the preprocessing stage, the chosen audio signals undergo essential transformations to prime them for CNN input. Initially presented in a 1D format, these signals are morphed into fixed-length waveforms [6]. This conversion serves as a crucial step, enabling the transition from a 1D signal to a 2D representation, thereby rendering them compatible with CNN processing.

Feature Extraction

Feature extraction is performed by applying a set of filters to the input audio signal through convolutional layers. These layers produce feature maps that capture important patterns and characteristics within the audio signals [7].

Spectogram Conversion

The audio signals are converted into spectrograms, which provide a visual representation of the audio signal's frequency content over time. These spectrograms serve as input to the CNN model for further processing.

CNN Model

The CNN model uses convolutional blocks with ReLU activation, batch normalization, and max-pooling. It progressively increases output filters for feature extraction. Global average pooling reduces dimensions, followed by two dense classification layers. The final layer employs softmax activation for multi-label classification, assigning class probabilities.

Prediction

During the prediction phase, the CNN model assigns a class label to the audio signal based on the highest probability among the predicted classes [8]. The network leverages the learned patterns and features extracted from the spectrograms to make accurate predictions regarding the bird species present in the audio recordings.

METHODOLOGY

Preprocessing Process

Short-Time Fourier Transform (STFT)

STFT breaks down audio signals into short segments and applies Fourier Transform, visualized in a spectrogram [9–12]. Librosa is used to choose parameters like window size. Relevant features, including dominant frequencies, are extracted and utilized to train CNN models, refined based on evaluation results, and deployed for classifying bird sounds in new audio data.

Magnitude Spectogram

The magnitude spectrogram, derived from the Short-Time Fourier Transform, depicts frequency intensity variations in an audio signal over time, pivotal for bird audio classification. Its segmentation of audio into short intervals offers a time-frequency view, aiding machine learning models in identifying distinct bird call patterns. Model Architecture shown in Figure 1

Mel-Filter Bank

Mel-Filter bank partitions the audio spectrum into mel-frequency bands, accentuating frequencies pertinent to human hearing and providing crucial features for bird audio classification. Applied to the STFT, it produces mel-frequency coefficients, capturing vital frequency components and enhancing machine learning models' capability to differentiate and classify bird species based on distinct spectral attributes.

Design Details

Frontend Interaction

The user interacts with the frontend of the application. Users upload bird sound audio files, which are checked for proper accepted formats such as .mp3, .wav, etc. Once validated, the audio file is sent to the backend for further processing.

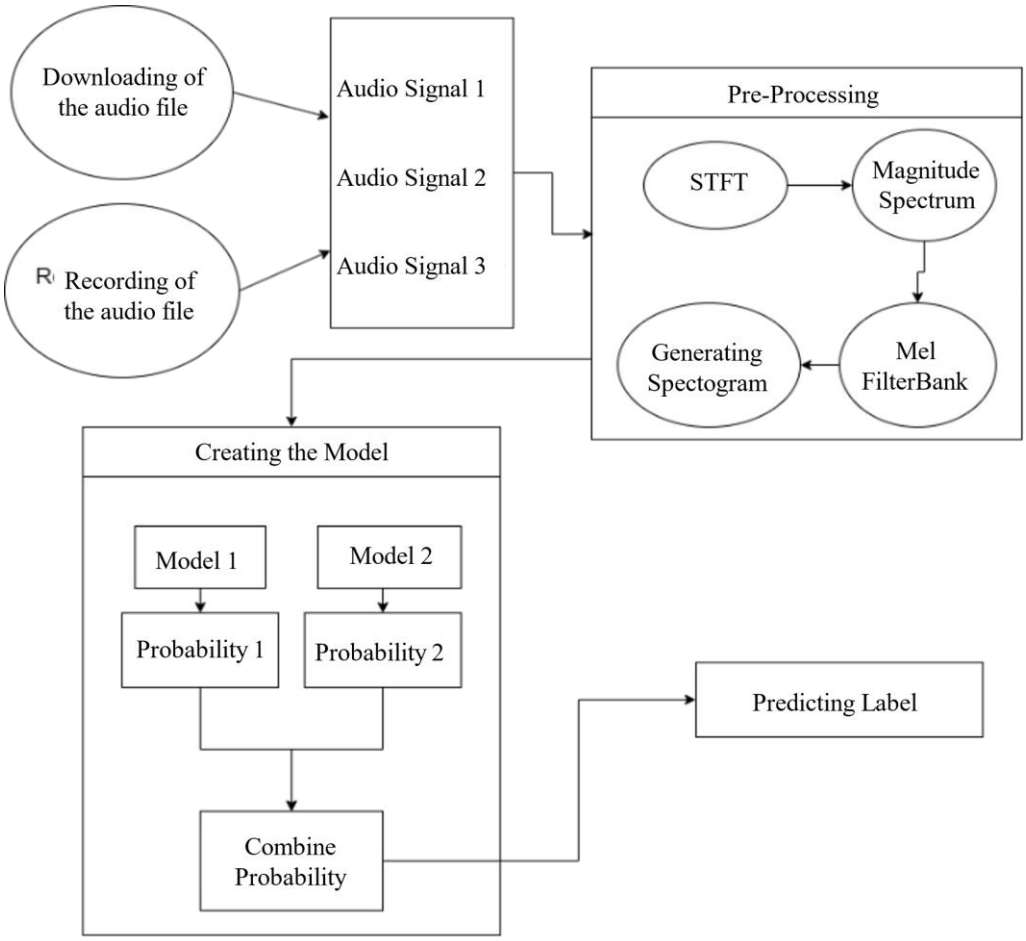


Figure 1. Model Architecture.

Backend Processing

Upon receiving the audio file, the backend divide it into segment of 5 sec each generating spectrograms. Spectrograms are generated from the audio segments and forwarded for subsequent analysis using CNN models. Model Workflow shown in Figure 2

CNN Model Analysis

Two CNN-based models are utilized for analyzing the spectrograms. Predictions from both models are combined to enhance result accuracy.

Model 1:

1. Convolutional Layer 1
   - *Input:* Spectrogram image of size 48x128
   - *Number of filters:* 16
   - *Size of each filter:* (3,3)
   - *Output size after convolution:* 46x126
   - Output size after max pooling: 23x63 with 16 feature maps

2. Convolutional Layer 2
   - *Input size:* (23,63,16)
   - *Input:* Spectrogram image of size 23,63, 16 feature maps
   - *Number of filters:* 32
   - *Size of each filter:* (3,3)

- *Output size after convolution:* 21x61
- *Output size after max pooling:* 10x30 with 32 feature maps

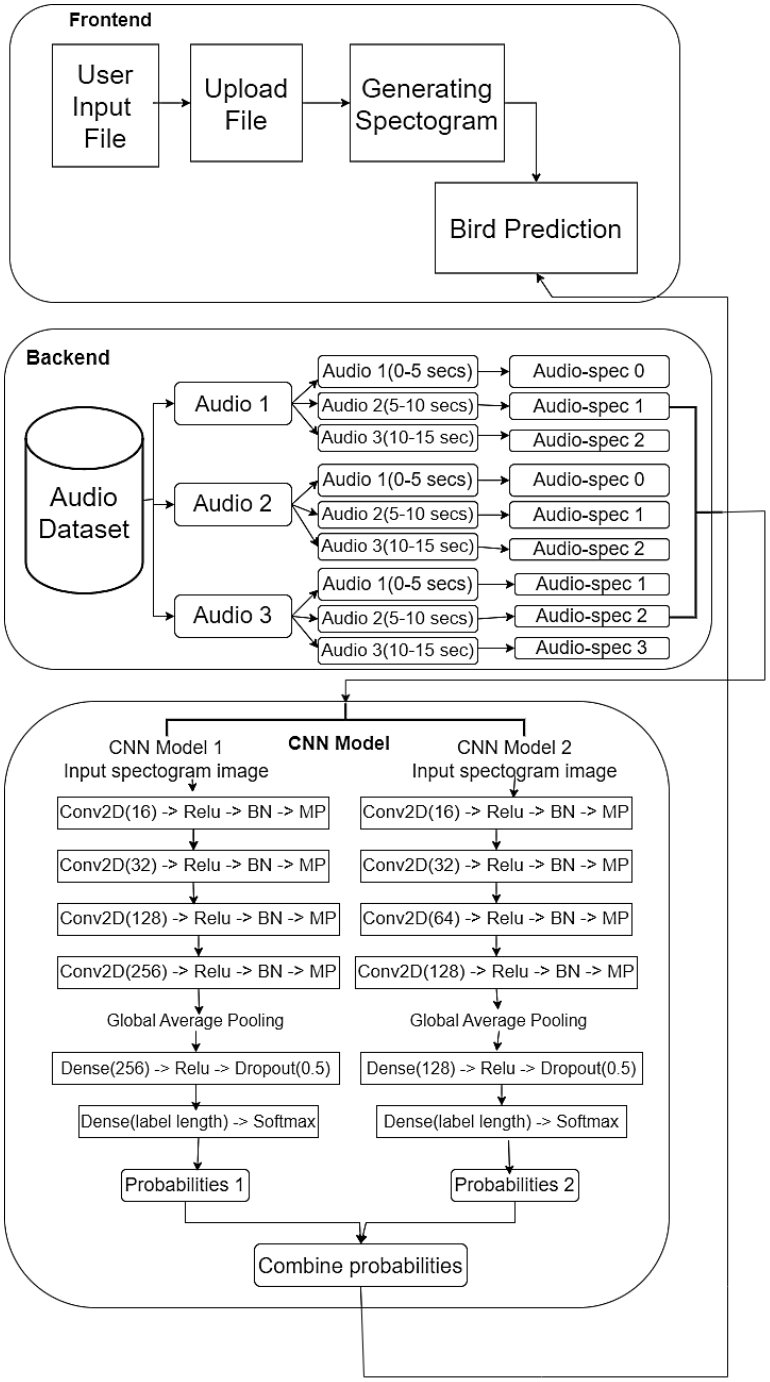


Figure 2. Model Workflow.

3. Convolutional Layer 3
   - *Input:* Spectrogram image of size 10,30, 32 feature maps
   - *Number of filters:* 128
   - *Size of each filter:* (3,3)
   - *Output size after convolution:* 8x28
   - *Output size after max pooling:* 4x14 with 128 feature maps

4. Conv2D (256) -> Rectified Linear Unit -> Max Pooling
   - *Input:* Spectrogram image of size 4,14,128 feature maps
   - *Number of filters:* 256
   - *Size of each filter:* (3,3)
   - *Output size after convolution:* 2x12
   - *Output size after max pooling:* 1x6 with 256 feature maps

5. Global Average Pooling
   - *Input:* 1x6 images with 256 feature maps
   - *Output:* 1x1 image with 256 features (1D array)

6 Dense (256) -> Rectified Linear Unit -> Dropout (0.5)
   - Adds weights and biases to previous input
   - *Input:* 1D array of size 256
   - *Output:* 1D array of size 256
   - *Dropout rate:* 0.5

7. Dense (label length) -> Softmax
   - Label length = 40
   - Fully connected layer mapping to the output labels
   - *Input:* 1D array
   - *Output:* Softmax probabilities over the 40 classes

Model 2:

1. Convolution Layer 1
   - *Input:* Spectrogram image of size 48x128
   - *Number of filters:* 16
   - *Size of each filter:* (3,3)
   - *Output size after convolution:* 46x126
   - *Output size after max pooling:* 23x63 with 16 feature maps

2. Convolution layer 2
   - *Input:* Spectrogram image of size 23,63, 16 feature maps
   - *Number of filters:* 32
   - *Size of each filter:* (3,3)
   - *Output size after convolution:* 21x61
   - *Max pooling:* Reduces output image size to half
   - *Output size after max pooling:* 10x30 with 32 feature maps

3. Convolution layer 3
   - *Input:* Spectrogram image of size 10,30 , 32 feature maps
   - *Number of filters:* 64
   - *Size of each filter:* (3,3)
   - *Output size after convolution:* 8x28
   - *Output size after max pooling:* 4x14 with 64 feature maps

4. Convolution layer 4
   - *Input:* Spectrogram image of size 4,14, 64 feature maps
   - *Number of filters:* 128
   - *Size of each filter:* (3,3)

- *Output size after convolution:* 2x12
- *Output size after max pooling:* 1x6 with 128 feature maps

5. Global Average Pooling
   - *Input:* 1x6 images with 128 feature maps
   - *Output:* 1x1 image with 128 features(1D array)

6. Dense (128) -> Rectified Linear Unit -> Dropout (0.5)
   - Adds weights and biases to previous input
   - *Input:* 1D array of length 128
   - *Output:* 1D array of size 128
   - Dropout rate : 0.5

7. Dense (label length) -> Softmax
   - Label length = 40
   - Fully connected layer mapping to the output labels
   - *Input:* 1D array
   - *Output:* Softmax probabilities over the 40 classes

Result Aggregation

Ensemble Prediction: After obtaining predictions from both Model 1 and Model 2, a weighted sum of their predictions is computed. This is done by taking a linear combination of the prediction probabilities from both models, each multiplied by a weight of 0.5. For each spectrogram chunk, the probabilities of all bird species are calculated by averaging the corresponding probabilities from both models. Model 1 accuracy & loss curve shown in Figure 3

Species Identification: Identify the species with the highest probability score from the ensemble prediction. Determine the index of the maximum probability score in the combined probability array to find the predicted label. User Interface shown in Figure 5

Mapping to Bird Name: The index of the highest probability score corresponds to the position of the species label in the LABELS array. The bird name corresponding to this index is extracted from the LABELS array.

DATASET DESCRIPTION

If you're looking for bird audio datasets, there are several high-quality resources available for different purposes such as research, machine learning, and conservation. Here are some notable ones. Details of bird audio dataset Shown in Table 1

Training & Testing size

- Total audio file 10955, 40 classes, total number of spectogram 30322
- Training audio files 80% = 8764
- Total number of spectogram 80% = 24258
- Testing audio files 20% = 2191
- Total number of spectograms 20% = 6064

RESULTS

Performance of Algorithms on Training Data

Model 1

The model achieved 95.35% training accuracy and 72.74% testing accuracy, with F1 scores of 0.9508 and 0.7239 respectively. Model 1 Training & testing accuracy shown in Table 2

Table 1. Details of bird audio dataset.

Attributes	Description
Primary Label	Bird species category
Type	Sound type (call or flight)
Latitude	North-south coordinate of recording
Longitude	East-west coordinate of recording
Scientific Name	Formal taxonomic name of bird species
Common Name	Informal name of bird species
Author	Original recorder of the audio
Rating	Sound quality (1-5 scale)
URL	Web location of audio file
Filename	Name of audio file

Table 2. Model 1 Training & testing accuracy.

Model 1	Accuracy (%)	F1-Score	Loss
Training	95.35	0.9508	0.3179
Testing	72.74	0.7239	1.2800

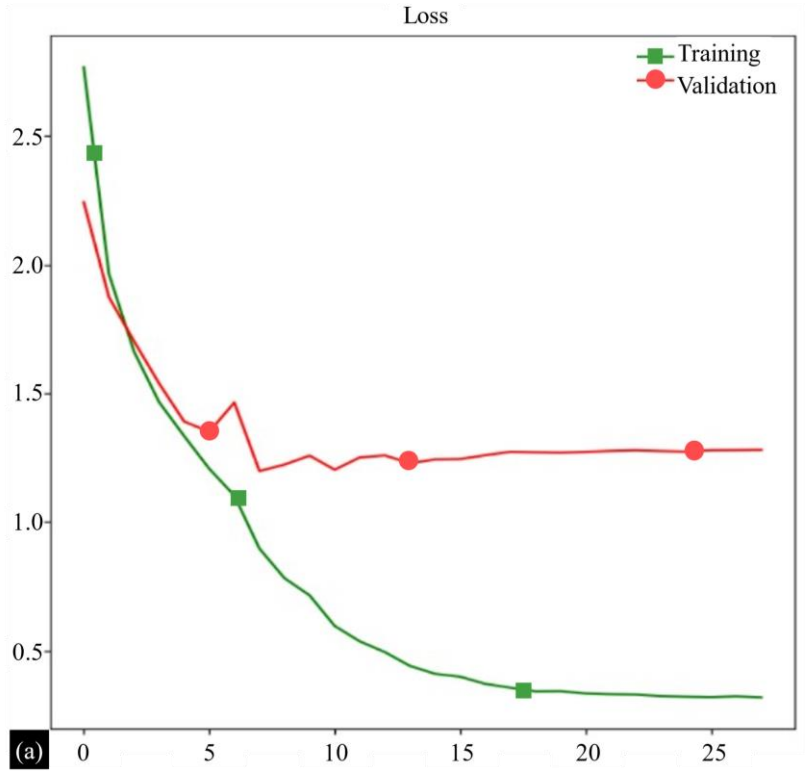
Model 2

The model achieved 88.69% training accuracy and 70.00% testing accuracy, with F1 scores of 0.8816 and 0.6934 respectively. Model 2 accuracy & loss curve shown in Figure 4 and discussed in Table 3.

Table 3. Model 2 Training & testing accuracy.

Model 2	Accuracy (%)	F1-Score	Loss
Training	88.69	0.8816	0.5260
Testing	70.00	0.6934	1.4119

GUI Results:



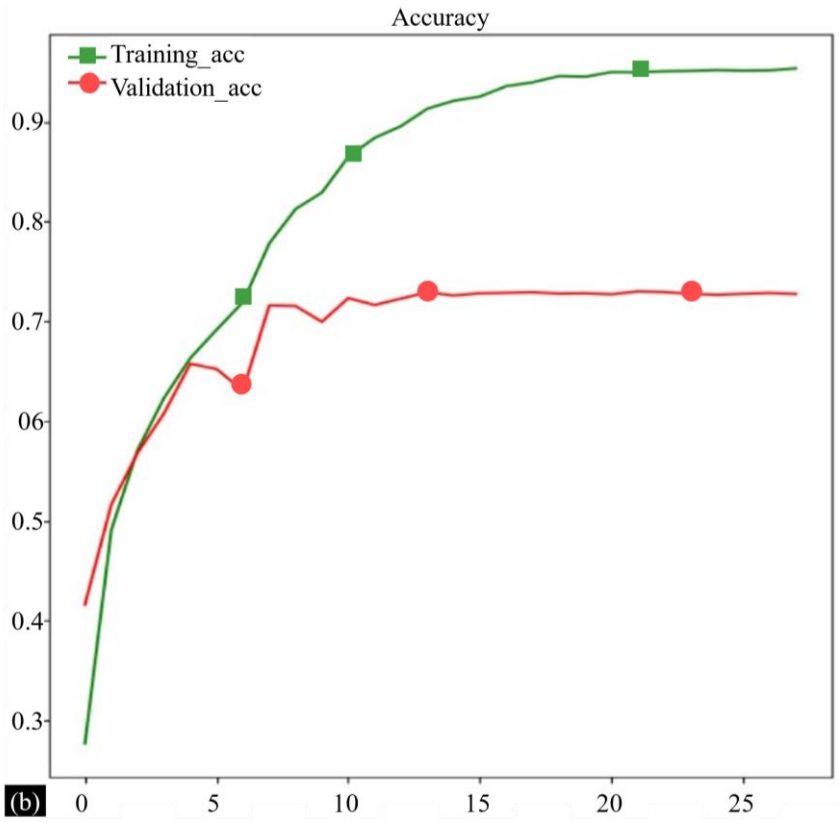
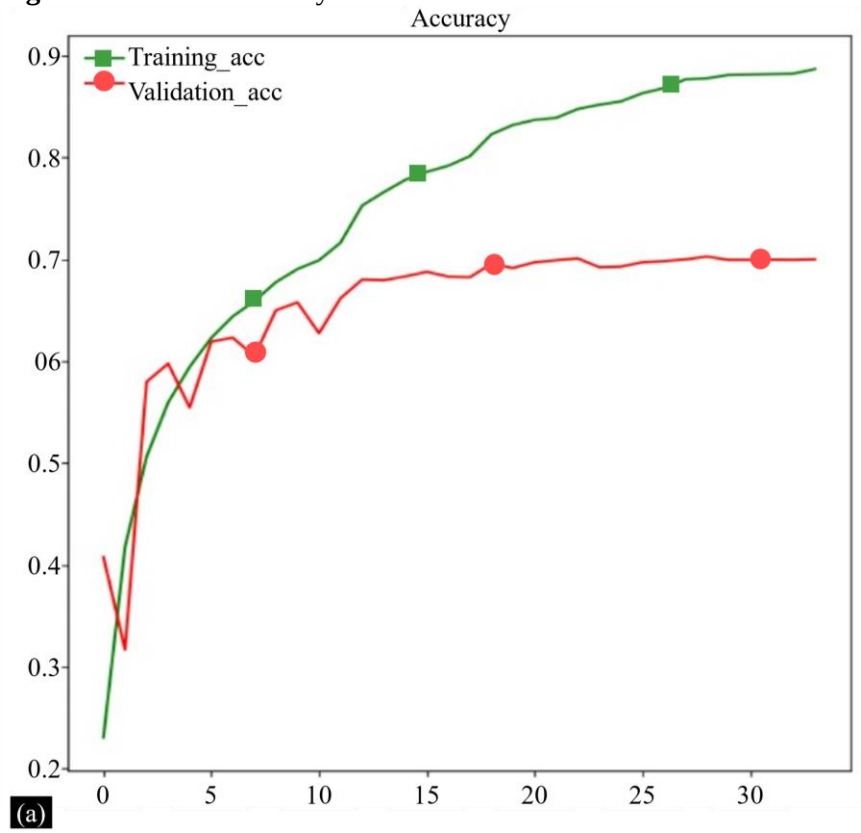


Figure 3. Model 1 accuracy & loss curve.



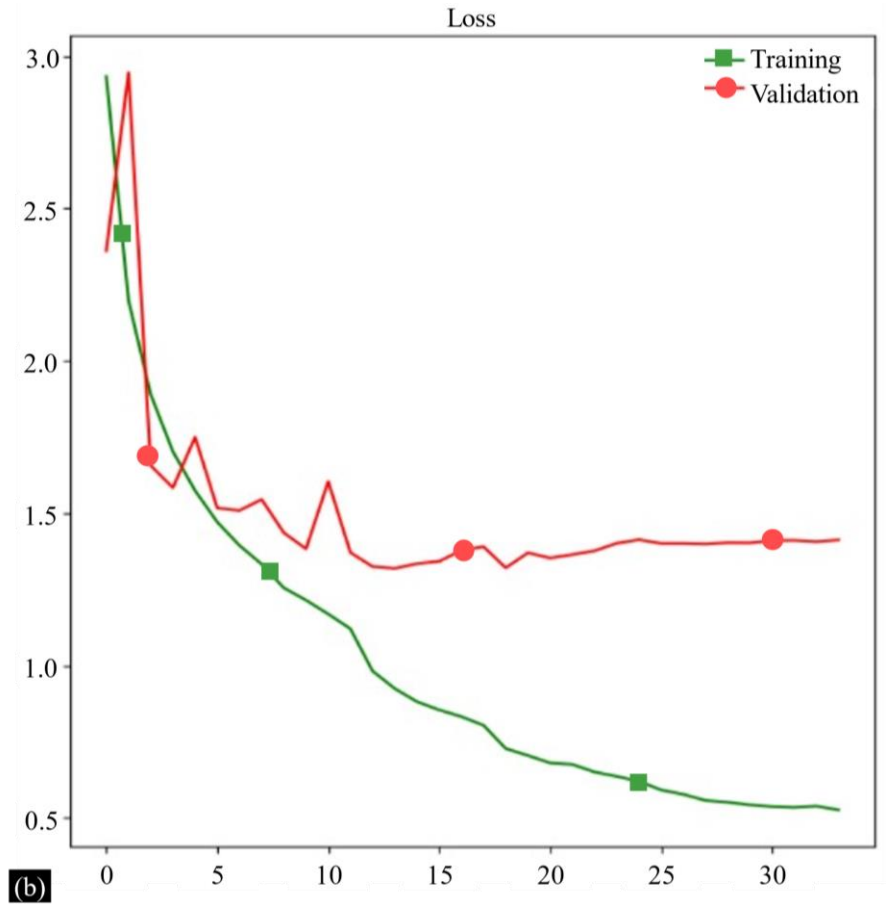


Figure 4. Model 2 accuracy & loss curve.

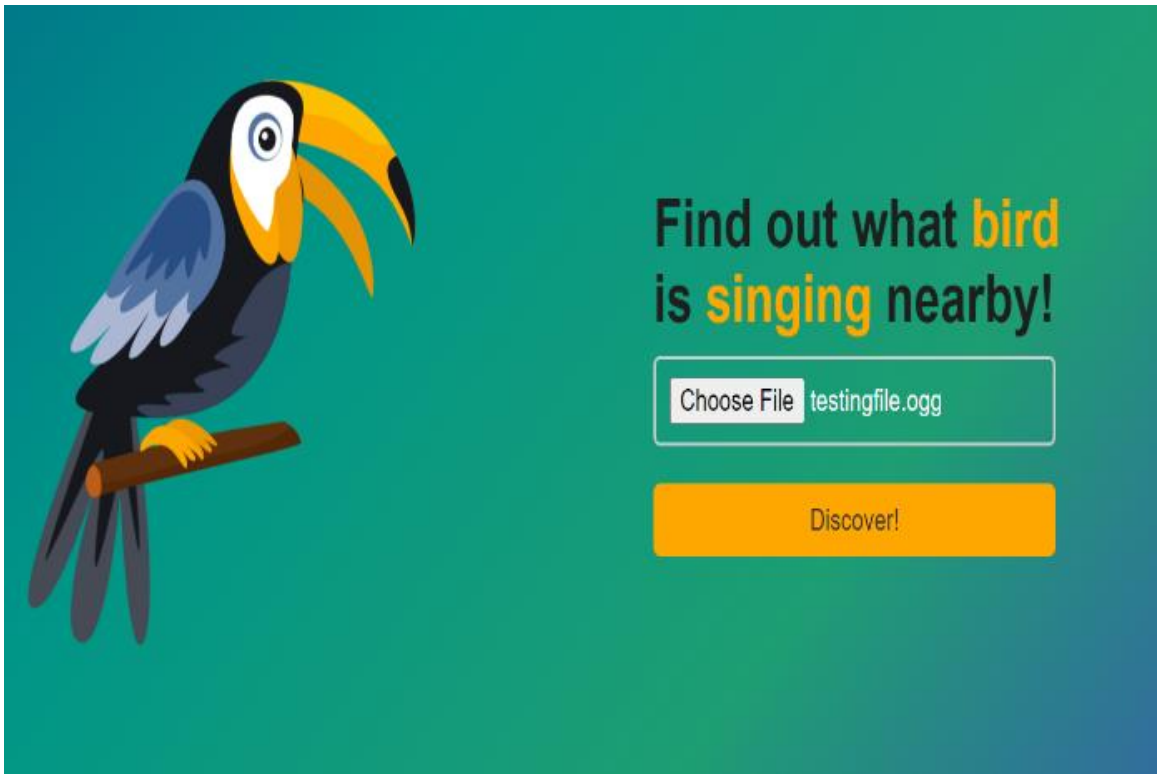


Figure 5. User Interface.

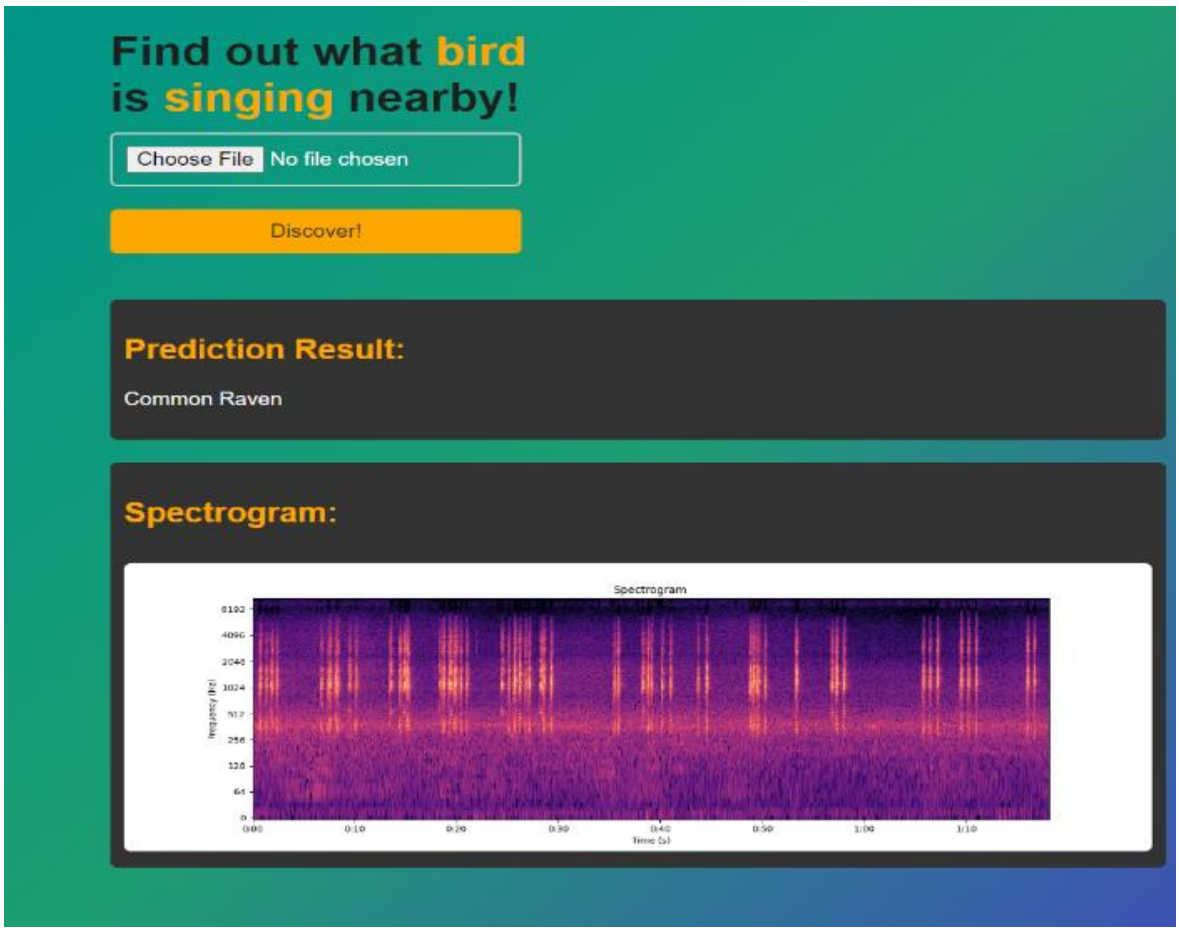


Figure 6. Result.

CONCLUSION

In conclusion, the bird audio recognition system using CNNs represents a significant advancement in avian biodiversity monitoring. Leveraging machine learning and signal processing, it enables accurate species identification from audio recordings. With a user-friendly interface and real-time processing, it aids researchers, citizen scientists, and conservationists in monitoring and conserving bird populations efficiently. Through collaboration and ongoing refinement, it contributes to global biodiversity preservation. Gui result shown result shown in Figure 6

Acknowledgement

We would like to express our sincere gratitude to the following individuals for their valuable contributions to this research: *Ved Chaudhari, Siddharth Chajjed, Piyush Jain, Adarsh Kamble* for their insightful ideas and meticulous data analysis; expertise in experimental design and methodology;diligent literature review and writing contributions and technical support and feedback throughout this project. Their dedication and collaboration have greatly enriched this work

REFERENCES

1. Á. Incze, H. -B. Jancsó, Z. Szilágyi, A. Farkas and C. Sulyok, "Bird Sound Recognition Using a Convolutional Neural Network," 2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY), Subotica, Serbia, 2018, pp. 000295-000300, doi: 10.1109/SISY.2018.8524677.
2. B. Chandu, A. Munikoti, K. S. Murthy, G. Murthy V. and C. Nagaraj, "Automated Bird Species Identification using Audio Signal Processing and Neural Networks," 2020 International Conference

on Artificial Intelligenceand Signal Processing (AISP), Amaravati,India, 2020, pp. 1-5, doi: 10.1109/AISP48273.2020.9073584.
3. G. C. Shekar, C. B. L., "Birds Audio Detection using Convolutional Neural Network and Transfer Learning," in IEEE Transactions on Audio and Speech Processing, vol. 08, no. 09, pp. 2021, Sep. 2021.
4. M. Lasseck, "Acoustic Bird Detection with Deep Convolutional Neural Networks," in Detection and Classification of Acoustic Scenes and Events 2018 Challenge, Technical Report, 2018.
5. E. Sprengel, M. Jaggi, Y. Kilcher, and T. Hofmann, “Audio based bird species identification using deep learning techniques,” Working notes of CLEF, 2016.
6. M. Lasseck, "Audio-based Bird Species Identification with Deep Convolutional Neural Networks," in Proceedings of the Working Notes of CLEF 2018, Museum für Naturkunde – Leibniz Institute for Research on Evolution and Biodiversity.
7. S. Kahl, T. Wilhelm-Stein, H. Hussein, H. Klinck, D.Kowerko, M. Ritter, and M. Eibl, “Large-scale bird sound classification using convolutional neural networks,” Working notes of CLEF, 2017.
8. Fazekas B, Schindler A, Lidy T (2017) A Multi -modal Deep Neural Network approach to Bird - song Identification. In: Working Notes of CLEF 2017.
9. Ruff ZJ, Lesmeister DB, Duchac LS, Padmaraju BK, Sullivan CM. Automated identification of avian vocalizations with deep convolutional neural networks. Remote Sensing in Ecology and Conservation. 2020 Mar;6(1):79-92.
10. LeBien J, Zhong M, Campos-Cerqueira M, Velev JP, Dodhia R, Ferres JL, Aide TM. A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network. Ecological Informatics. 2020 Sep 1; 59:101113.
11. Jasim HA, Ahmed SR, Ibrahim AA, Duru AD. Classify bird species audio by augment convolutional neural network. In2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) 2022 Jun 9 (pp. 1-6). IEEE.
12. Grill T, Schlüter J. Two convolutional neural networks for bird detection in audio signals. In2017 25th European Signal Processing Conference (EUSIPCO) 2017 Aug 28 (pp. 1764-1768). IEEE.

