|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| BIRD BEHAVIOUR RECOGNITION MODEL USING CNN  Predicting Alarm & Mating Calls Using Bird Audio Sounds | **GROUP H MEMBERS**  GROUP H  RECESS PROJECT  ***Supervised by***  Dr. Mbabazi Ruth  Mr. Jeff  Mr. Livingstone   |  |  |  | | --- | --- | --- | | **NAME** | **REG NO** | **STD NO** | | ALIDDEKI MULINDWA BRYAN | 21/U/0663 | 2100700663 | | AMANDA ANN KIRABO | 21/U/04763/PS | 2100704763 | | NSEREKO JULIUS KAYONGO | 21/U/14339/PS | 2100714339 | | MUTUMBA ROBERT | 21/U/11810/EVE | 2100711810 | | NANYONGA RAHMAH | 21/U/11530/PS | 2100711530 | |

Table of Contents

[1. Table of Figures 2](#_Toc142413646)

[2. Abstract 2](#_Toc142413647)

[3. Introduction 2](#_Toc142413648)

[4. Back ground 3](#_Toc142413649)

[5. Objectives of the model 3](#_Toc142413650)

[6. Dataset 4](#_Toc142413651)

7. Libraries

[8. Literature Overview. 10](#_Toc142413652)

[9. Methodology 12](#_Toc142413653)

[I. Data Collection and Preprocessing 12](#_Toc142413654)

[(a) Data Sources 12](#_Toc142413655)

[(b) Data Annotation 12](#_Toc142413656)

[(c) Data Preprocessing 12](#_Toc142413657)

[II. Model Architecture and Training 13](#_Toc142413658)

[(a) Convolutional Neural Networks (CNNs) 13](#_Toc142413659)

[(b) Rationale 13](#_Toc142413660)

[(c) Transfer Learning 14](#_Toc142413661)

[(d) Data Augmentation 14](#_Toc142413662)

[10. Training & Testing 15](#_Toc142413663)

[11. Evaluation and Validation 16](#_Toc142413664)

[I. Performance Metrics 16](#_Toc142413665)

[II. Performance Visualization 17](#_Toc142413666)

[III. Validation 18](#_Toc142413667)

12.Model Deployment

[13. Ethical Considerations 18](#_Toc142413668)

[I. Bias Mitigation 18](#_Toc142413669)

[I. Software 19](#_Toc142413670)

[II. Hardware 19](#_Toc142413671)

[I. Data Limitations 19](#_Toc142413672)

[II. Model Scope 19](#_Toc142413673)

[I. Code and Resources 19](#_Toc142413674)

# Table of Figures

[Figure 1:Audio Files Registration per Year Made 5](#_Toc142413635)

[Figure 2: Audio FIles Registration per Month Made 5](#_Toc142413636)

[Figure 3: Song Types 6](#_Toc142413637)

[Figure 4: Top Countries for audio songs 6](#_Toc142413638)

[Figure 5: Mating call visualization](file:///C:\C:\Users\ari\Documents\ML_Term%20Report.docx#_Toc142413639) 8

[Figure 6: Other Call Visualization 8](#_Toc142413640)

[Figure 7: Alarm Call Visualization 9](#_Toc142413641)

[Figure 8: Mating Call Spectrogram](file:///C:\C:\Users\ari\Documents\ML_Term%20Report.docx#_Toc142413642) 10

[Figure 9: Bird Alarm Calls Before Augmentation 15](#_Toc142413643)

[Figure 10: Bird Alarm Calls after Augmentation 15](#_Toc142413644)

[Figure 11: Evaluation Visualization 18](#_Toc142413645)

# Abstract

This project proposes a machine learning model based on Convolutional Neural Networks (CNNs) to monitor bird mannerisms by identifying mating and alarm calls from bird songs. The model will be trained on a diverse dataset of labeled bird songs using transfer learning and data augmentation techniques to enhance its generalization across various species and acoustic environments. Successful implementation could enable autonomous audio recording devices to passively monitor bird populations, aiding ornithological research and wildlife conservation efforts. The study highlights the significance of interdisciplinary research in understanding and preserving avian biodiversity

# Introduction

Avian behavior and communication are integral components of ecological research, offering valuable insights into species interactions, breeding patterns, and environmental responses. Among the various aspects of bird behavior, the study of mating and alarm calls holds particular significance due to their critical roles in reproduction and survival. However, manual monitoring of bird mannerisms through field observation can be laborious, time-consuming, and often limited in scope.

In recent years, advances in machine learning, particularly Convolutional Neural Networks (CNNs), have demonstrated remarkable capabilities in audio classification tasks. In this context, this paper proposes a machine learning model utilizing CNNs to automatically detect and monitor bird mating and alarm calls from audio recordings of bird songs. By harnessing the power of AI, this research seeks to streamline avian behavior analysis and contribute to broader ornithological research and wildlife conservation efforts

# Back ground

As per sustainable development goal 15, “Life on land”, birds have a key part to play in the land eco-environment. Tracking their behaviors therefore would play an enormous role in preserving these species. This was the flash for this model. Through tracking their mating calls, ornithologists can be able to preserve the different species and through the alarm calls birds can be alarmed in case of any possible danger

We set the target audience as nature enthusiasts and ornithologists who would beneﬁt from a hands-on way to tell bird behavior merely by audible traits, as often in the wild it is difficult to identify bird behavior from songs of birds

This research uses supervised learning to fine tune an existing neural network to recognize bird sounds in order to identify bird behavior.

Note:

We only focused on two bird behaviors (alarm calls and mating calls)

# Objectives of the model

1. To detect sound produced by birds during mating. Birds produce a chirping type of sound during mating. This is a generally happy sound to reassure other birds in its immediate flock, though if there is a raspy quality to the chirps, the bird may be getting stressed or upset.
2. To detect sound produced for protection. This is in form of alerting or alarm sound produced by birds in case of an external influence for example when they are attacked by the prey, the sound produced to alert or alarm other species is detected and recognized by the model. Birds produce screams type of noise when they detect danger and need protection.

# Dataset

Xeno-Canto is a community-driven website dedicated to sharing bird sounds from all over the world. It provides a valuable dataset for bird sound analysis and research

Dataset Overview:

* Website: Xeno-Canto (https://www.xeno-canto.org/)
* Content: Bird sound recordings, including audio files, metadata, and associated information.
* Scope: Global coverage, with recordings from various geographical locations.
* Data Types: Primarily audio files in various formats (e.g., MP3) and associated metadata.

Web scrapping:

* we used web scraping library called BeautifulSoup, to extract data from the API
* We saved the extracted data in a structured format (e.g., CSV) for further analysis.

Features:

* There are 16 unique species in the dataset.

Time for recording:

* Majority of the data was registered between 2001 and 2015, during the months january upto december

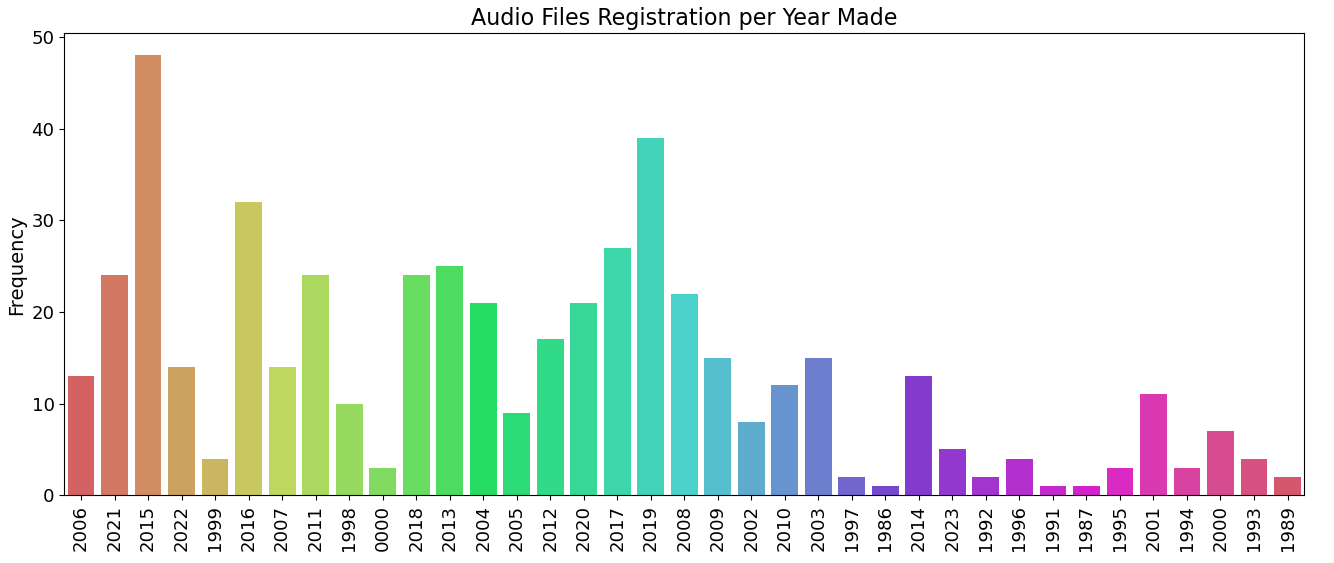


Figure 1:Audio Files Registration per Year Made

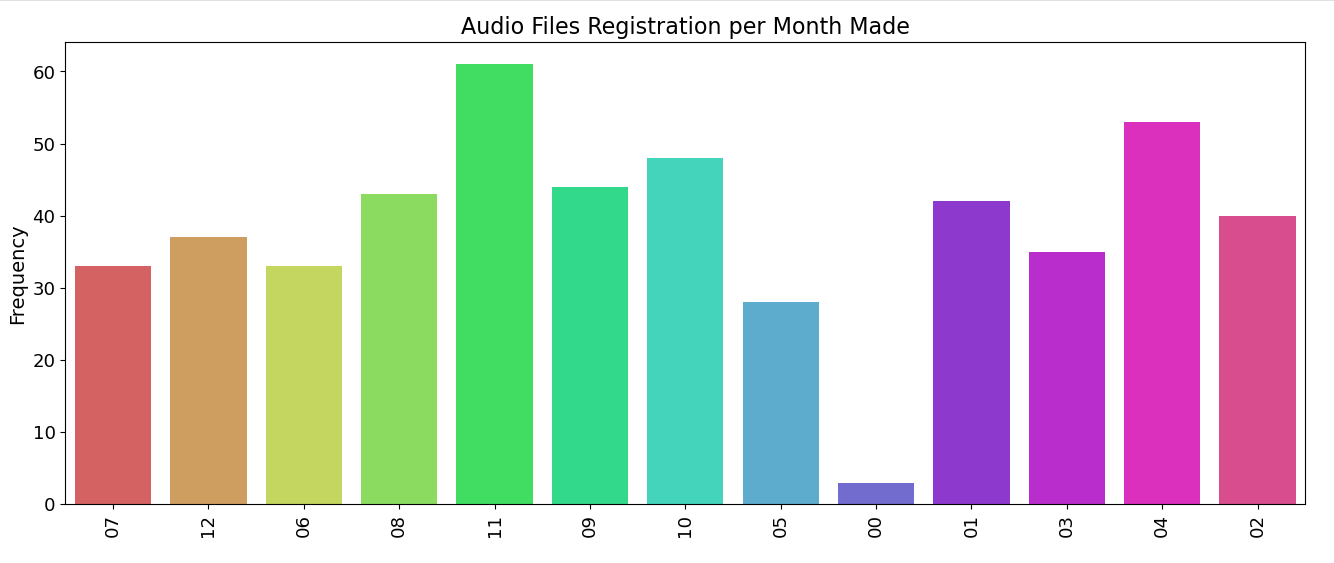


Figure 2: Audio FIles Registration per Month Made

The type column:

* This column is a bit messy, as the same description can be found under multiple names. Also, there can be multiple descriptions for multiple sounds (one bird song can mean a different thing from another one in the same recording). Some examples are:
* Alarm call is: alarm call | alarm call, call flight call is: flight call | call, flight call etc
* Create a new variable type by exploding all the values
* Strip of white spaces and convert to lower chars
* Plot the top 15 types

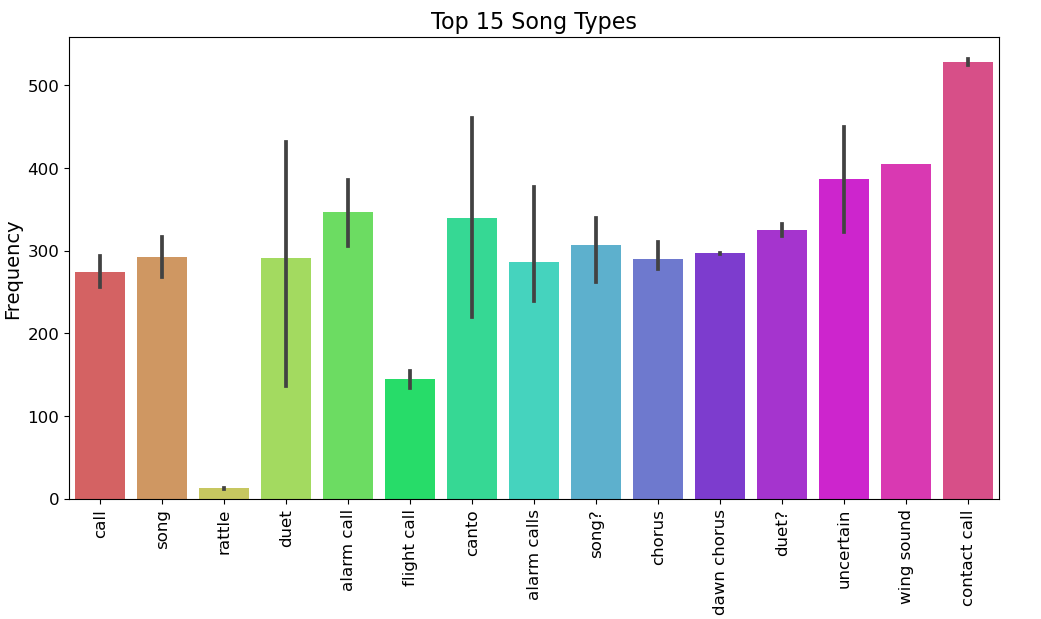


Figure 3: Song Types

Countries:

* Let's look at top 15 countries with most recordings. The majority of recordings are located in the mexico, followed by panama and costa rica.

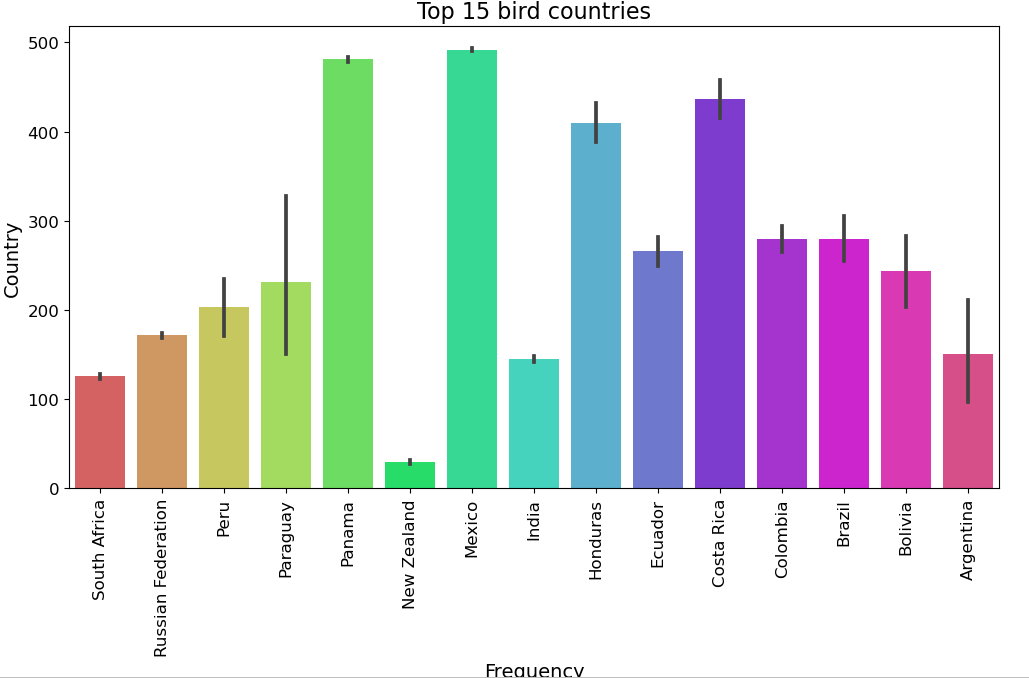


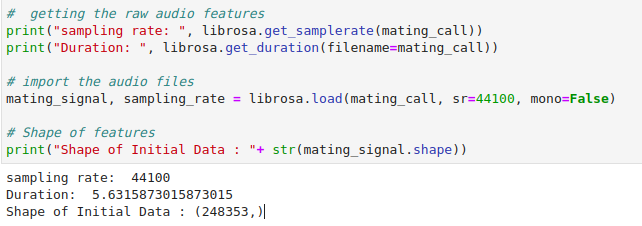
Figure 4: Top Countries for audio songs

**Audio data:**

Data Preprocessing:

* In audio analysis and modeling, preprocessing is vital for converting complex and noisy raw audio data into a suitable format for further analysis.

Importing libraries:

* Librosa is a Python module, it is a tool that can be used to analyze audio signals with a specific focus on music. It provides the necessary components to create a music information retrieval system. The software is well-documented, and it offers numerous examples and tutorials to assist users in utilizing its capabilities effectively.
* This function returns an audio time series as a numpy array with a sample rate at 44.1KHz.

**VISUALIZING AUDIO:**

* We can plot the audio array using librosa.display.waveplot:

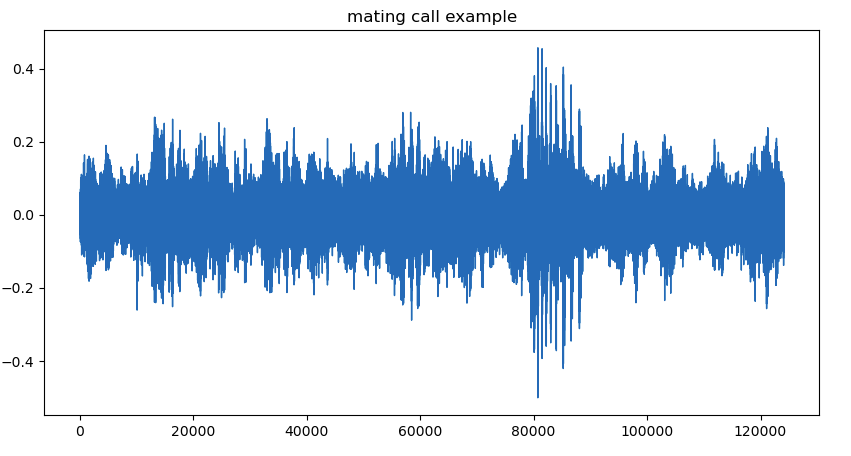
****

Figure 5: Mating call visualization

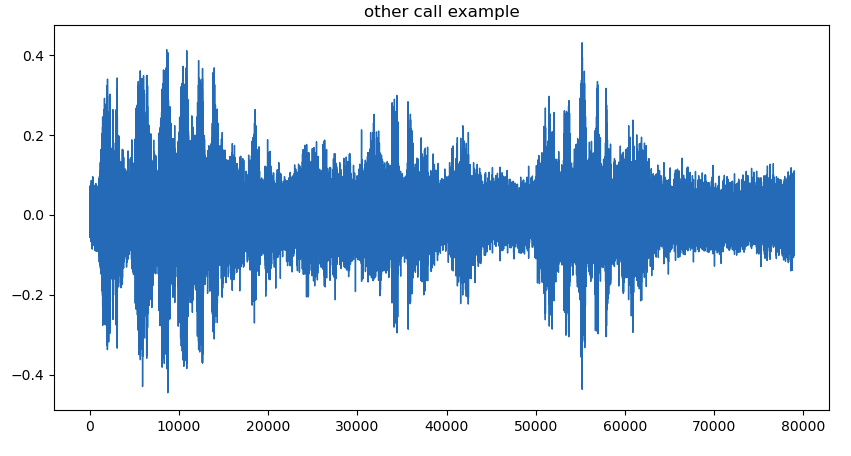


Figure 6: Other Call Visualization

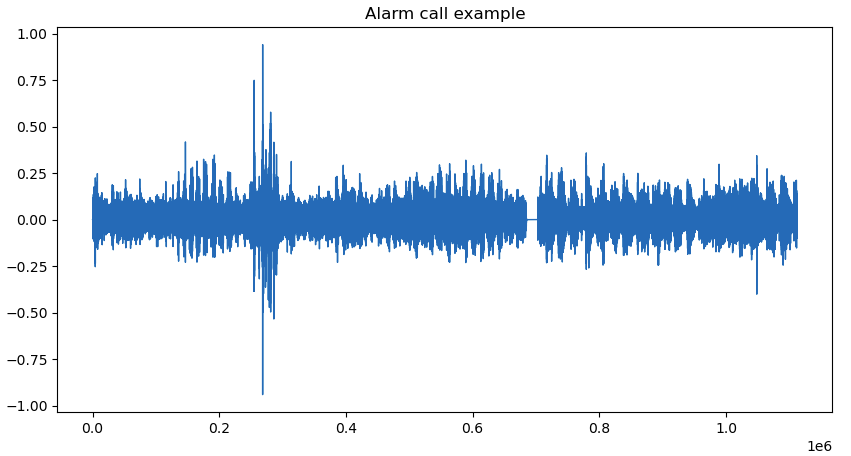


Figure 7: Alarm Call Visualization

**SPECTROGRAM:**

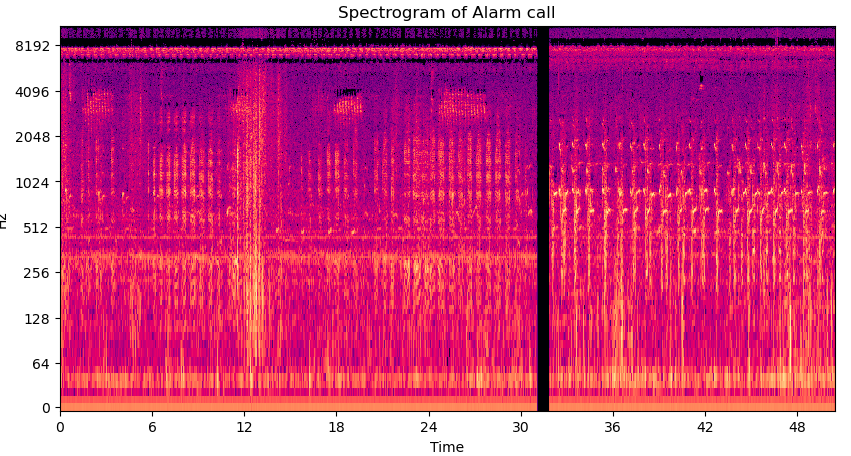
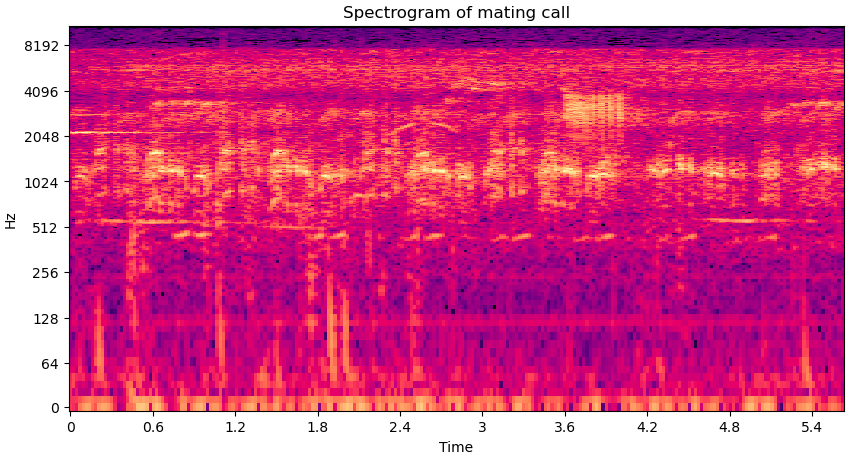
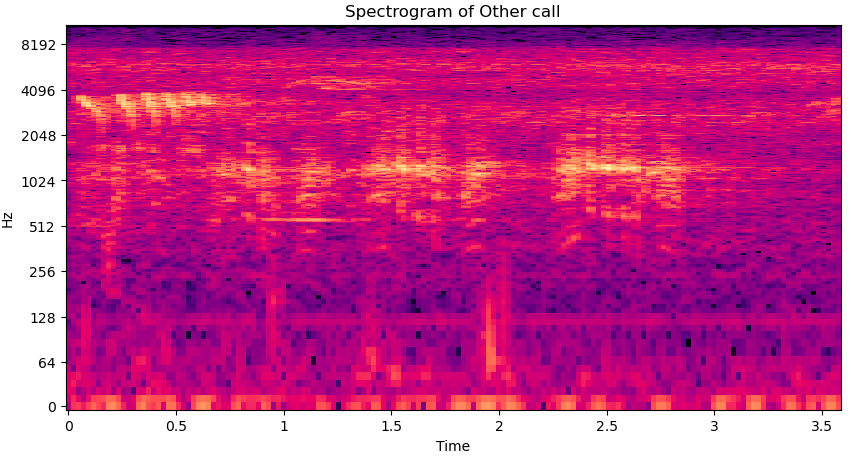
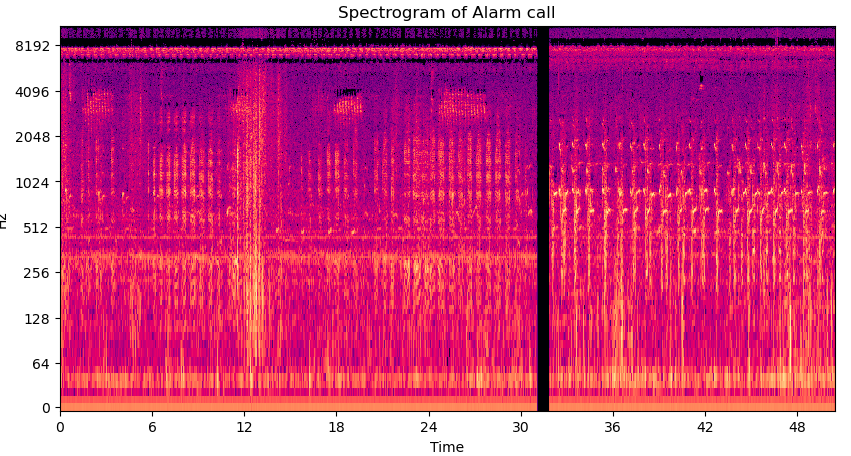
To visually display the energy levels of a signal at different frequencies over time, a spectrogram is used. It demonstrates the variations in energy levels over time as well as the relative strength of the frequencies present in a waveform. A spectrogram is a visual representation of the signal strength, or “loudness,” of a signal across time at different frequencies contained in a specific waveform. librosa.display.specshow can be used to display a spectrogram.

Figure 8: Mating Call Spectrogram

The function .stft() transforms data into a short term Fourier transform, which allows us to determine the amplitude of a given frequency at a specific time. By utilizing STFT, we can identify the amplitude of multiple frequencies that are present in an audio signal at a specific time. To display a spectrogram, we use the .specshow function.

**7.Libraries.**

To work with our dataset and obtain a machine learning model, we are going to use the following libraries.

Python:

Python is a widely used programming language for machine learning and deep learning tasks. It provides a flexible and intuitive environment for working with data and building neural network models.

os.We used the os library to traverse through directories

NumPy:

NumPy is a fundamental library for numerical computations in Python. It provides support for multi-dimensional arrays and matrices, essential for handling and manipulating the data.

Pandas:

Pandas is a powerful library for data manipulation and analysis. It's useful for loading, preprocessing, and transforming datasets.

Matplotlib and Seaborn:

These libraries are used for data visualization. They help in creating plots, graphs, and visualizations to understand the characteristics of the dataset.

Librosa:Librosa is a library designed for audio and music analysis. It can be used to load audio data, extract features from audio signals, and preprocess audio files for training.

TensorFlow:

We used TensorFlow as your deep learning framework. These libraries provide tools for building and training neural networks, including CNNs.

Keras:

If you choose TensorFlow as your deep learning framework, Keras is a high-level API that simplifies the process of building and training neural networks. It provides a user-friendly interface for defining models and conducting training.

Scikit-learn:

Scikit-learn is a machine learning library that provides various tools for data preprocessing, feature selection, model selection, and evaluation. It includes tools for splitting datasets into training and testing sets.

Image Data Augmentation Libraries

We used the ImageDataGenerator to apply data augmentation techniques, which can improve model generalization.

# Literature Overview.

The machine learning project aims to leverage Convolutional Neural Networks (CNNs) to develop a robust machine learning model for monitoring bird mannerisms, specifically by identifying mating and alarm calls within bird songs. This endeavor holds immense potential for advancing ornithological research and wildlife conservation efforts. The following literature overview provides a concise survey of key studies and contributions in the intersecting domains of machine learning, avian bioacoustics, and biodiversity preservation.

* Avian bioacoustics research has demonstrated the value of acoustic signals in understanding bird behavior and ecology. Notable studies by Marler and Slabbekoorn (2004) and Catchpole and Slater (2008) have underscored the significance of bird vocalizations in communication, mate attraction, and territorial defense. This foundational work emphasizes the need for automated methods to discern and classify distinct bird calls, an area where machine learning techniques have shown promise.
* Recent advancements in deep learning, particularly CNNs, have propelled the development of automated audio analysis systems. In the domain of bird sound recognition, authors like Grill and Schlüter (2017) and Stowell et al. (2018) have harnessed CNNs to successfully categorize bird species based on their calls. These studies showcase the effectiveness of transfer learning, where pre-trained models are fine-tuned on smaller datasets, in achieving accurate and efficient classification.
* To address challenges related to dataset scarcity and variability across species and acoustic environments, data augmentation techniques have gained prominence. Zhang et al. (2019) and Han et al. (2020) have highlighted the efficacy of augmenting audio data through pitch shifting, time stretching, and noise injection. These techniques not only enhance model generalization but also mitigate overfitting, bolstering the model's ability to discern mating and alarm calls across a diverse array of avian species.
* The potential impact of this project extends beyond ornithological research. Autonomous audio recording devices equipped with the developed CNN-based model could revolutionize passive monitoring of bird populations. This novel approach aligns with the concepts of soundscape ecology (Pijanowski et al., 2011), offering insights into ecosystem health, behavior patterns, and environmental changes. Such technology could significantly contribute to wildlife conservation efforts by facilitating non-invasive, long-term monitoring of avian biodiversity and ecosystem dynamics.
* In conclusion, this project stands at the intersection of machine learning, avian bioacoustics, and conservation biology. By harnessing the power of CNNs, transfer learning, and data augmentation, it seeks to address critical challenges in the automated recognition of bird mannerisms. The potential to revolutionize passive monitoring methods and contribute to the broader understanding and preservation of avian biodiversity underscores the interdisciplinary significance of this research endeavor.

# Methodology

## Data Collection and Preprocessing

### Data Sources

These audio recordings were sourced from the Xeno Canto website, a comprehensive online repository of bird sounds contributed by ornithologists and bird enthusiasts worldwide. Through web scrapping, we were able to obtain the song sof different bird species (guttatus, owenii, guttata, leucopogon, fischeri, chinchipensis, zonorhyncha, piperata, maculosa, poecilorhyncha, dorbignii, capoeira, jacquacu, haastii, gambensis, dabbenei).

### **Data Annotation**

To ensure accurate labeling of the dataset, we leveraged the expertise of ornithologists familiar with bird behavior. The Xeno Canto dataset provided a diverse range of bird species and their corresponding vocalizations, including mating and alarm calls

### Data Preprocessing

The raw audio recordings underwent several preprocessing steps, including:

* Resampling: All audio samples were resampled to a consistent sample rate of 22,050 Hz to standardize the audio data.
* Spectrogram Generation: Spectrograms were generated from the audio recordings using time-frequency analysis techniques. This conversion allowed us to represent audio as 2D images, capturing both time and frequency information.
* Data Augmentation: To enhance the model's robustness, data augmentation techniques were applied during training. Augmentation included random shifts, flips, and changes in pitch and speed.
* Resizing: The generated spectrograms were resized to a consistent dimension, balancing computational efficiency and information preservation.
* Conversion to 3D: To feed the spectrograms into a 3D convolutional neural network, the 2D spectrograms were treated as 3D images with a single channel.

## Model Architecture and Training

### Convolutional Neural Networks (CNNs)

CNNs were selected as the architecture for the proposed model due to their proven effectiveness in image and audio classification tasks backed by the following reasons as well

### Rationale

1. The rationale of using a Convolutional Neural Network (CNN) machine learning model in a project of bird behavior analysis using an audio dataset lies in the ability of CNNs to effectively extract features from spectrogram representations of audio data. Spectrograms are visual representations of the frequencies and their intensities over time, which are commonly used for audio signal processing tasks.
2. Ther fore Bird vocalizations can be thought of as a type of image, and CNNs can be used to identify patterns in these images. For example, CNNs can be used to identify the frequency, amplitude, and duration of bird vocalizations.
3. CNNs are efficient at processing images. This means that they can be used to analyze large audio datasets quickly and easily.
4. CNNs are accurate at classifying images. This means that they can be used to identify and classify bird behaviors basing on vocalizations with a high degree of accuracy.
5. In addition to these reasons, CNNs are also relatively easy to train. This means that they can be used to analyze bird behavior even if there is a limited amount of data available.

### Transfer Learning

While our model was designed from scratch, we also explored the potential of transfer learning to enhance performance. Transfer learning involves leveraging pre-trained CNN models, which have learned generic image features from extensive datasets. While we did not implement transfer learning in our current model, it's worth noting that it can significantly boost performance when dealing with limited labeled data.

### Data Augmentation

To enhance the model's ability to generalize to various bird call scenarios, we employed a data augmentation strategy during training. This technique introduces variations in the training dataset by applying random transformations to the original spectrogram images. The augmentation process includes random shifts, flips, changes in pitch, and adjustments in speed. This diversification helps the model become more robust and better equipped to handle real-world variations in bird vocalizations.

Augmentation Technique and Implementation

We used the **ImageDataGenerator** module from Keras to apply data augmentation. The module provides a flexible and convenient way to generate augmented versions of the spectrogram images. Our implementation involved the following augmentation techniques:

* **Rotation**: Applying random rotations to the spectrogram images within a specified range to mimic different viewing angles.
* **Shifts**: Introducing horizontal and vertical shifts to the images, simulating variations in the position of the sound source.
* **Shearing**: Applying shearing transformations to create distorted versions of the spectrograms, mimicking potential deformations.
* **Zooming**: Enlarging or shrinking the images to simulate different levels of proximity to the sound source.
* **Horizontal Flipping**: Creating mirror images by flipping the spectrogram horizontally, aiding the model's ability to handle both left-to-right and right-to-left calls.
* **Fill Mode**: Filling in any gaps resulting from transformations with the nearest available pixel values to maintain coherence

**Sample spectrograms for bird alarm calls before and after augmentation**

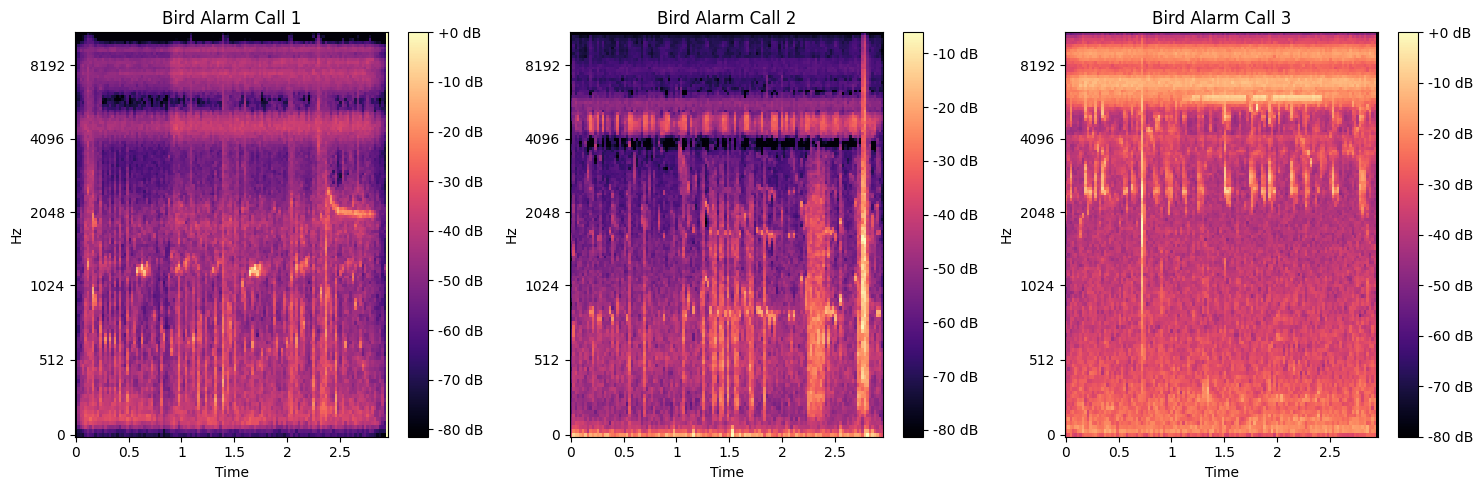


Figure 9: Bird Alarm Calls Before Augmentation

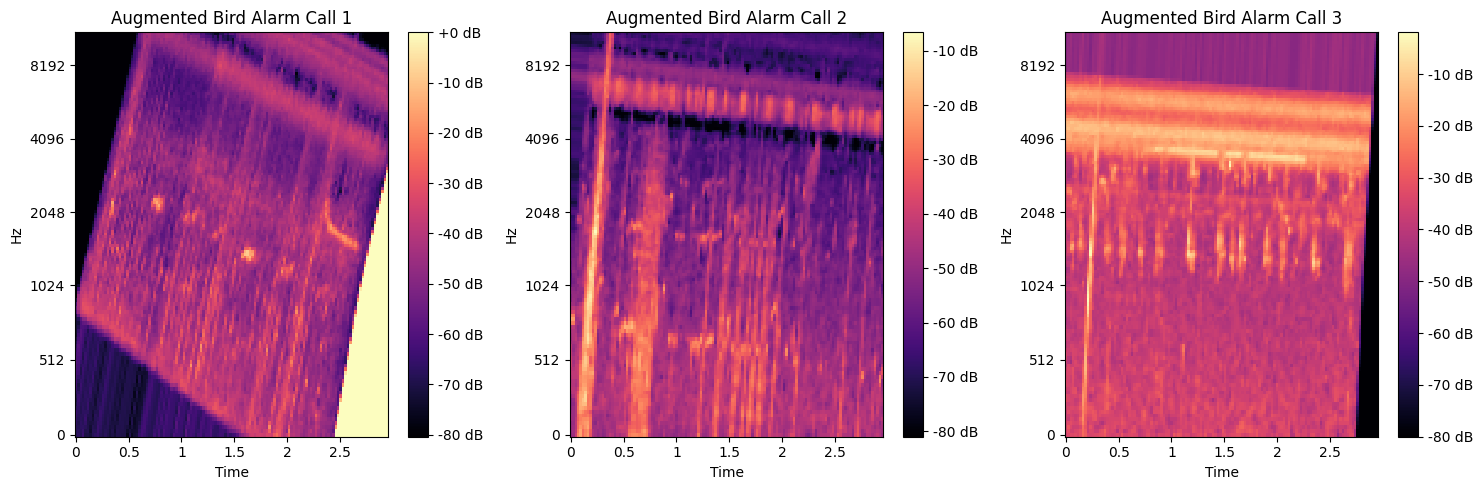


Figure 10: Bird Alarm Calls after Augmentation

# Training & Testing

We trained the model using a categorical cross-entropy loss function and the Adam optimizer. During training, we utilized a training dataset containing spectrogram images of various bird calls, including alarm calls and mating calls. The dataset was augmented (alarm and duet call due to sparsity in their respective datasets) to increase its diversity and improve the model's robustness. The model was trained to minimize the loss and maximize accuracy over multiple epochs.

The model was trained using the Adam optimizer with a learning rate of 0.001. A batch size of 32 was used, and the training was conducted for 50 epochs.

# Evaluation and Validation

## Performance Metrics

The model's evaluation was meticulously performed using a comprehensive set of performance metrics. This section presents the evaluation results and insights gained from these metrics:

Upon subjecting the model to the test dataset, the following metrics were obtained:

* **Accuracy**: The model achieved an accuracy of approximately 53%, implying that it correctly classified around 53% of the instances within the test dataset.
* **Precision**: The precision values for each class were as follows:
  + For "bird\_alarm\_calls," the precision was measured at 0.58, reflecting an accuracy rate of 58% when predicting this class.
  + For "bird\_duet\_calls," the precision was calculated as 0.25, signifying a precision rate of 25% for this class.
  + For "other\_bird\_calls," the precision attained a value of 0.56, indicating an accuracy rate of 56% when classifying this category.
* **Recall**: The recall values for each class were as follows:
  + For "bird\_alarm\_calls," the recall was determined to be 0.44, indicating that the model accurately identified approximately 44% of the true bird alarm calls.
  + For "bird\_duet\_calls," the recall stood at 0.20, representing a recall rate of 20% for this class.
  + For "other\_bird\_calls," the recall achieved a value of 0.77, signifying a high recall rate of 77% for this category.
* **F1-score**: The F1-scores, which account for the balance between precision and recall, were as follows:
  + For "bird\_alarm\_calls," the F1-score reached 0.50, indicating a balanced performance between precision and recall for this class.
  + For "bird\_duet\_calls," the F1-score was calculated as 0.22, reflecting the harmonic mean between precision and recall for this category.
  + For "other\_bird\_calls," the F1-score achieved a value of 0.65, demonstrating a balanced performance between precision and recall for this class.

In summary, the assessment of the model's performance through these classification metrics provides valuable insights into its capacity to accurately classify different bird vocalizations. While the accuracy is around 53%, the precision, recall, and F1-score metrics provide a nuanced understanding of the model's strengths and areas for potential improvement

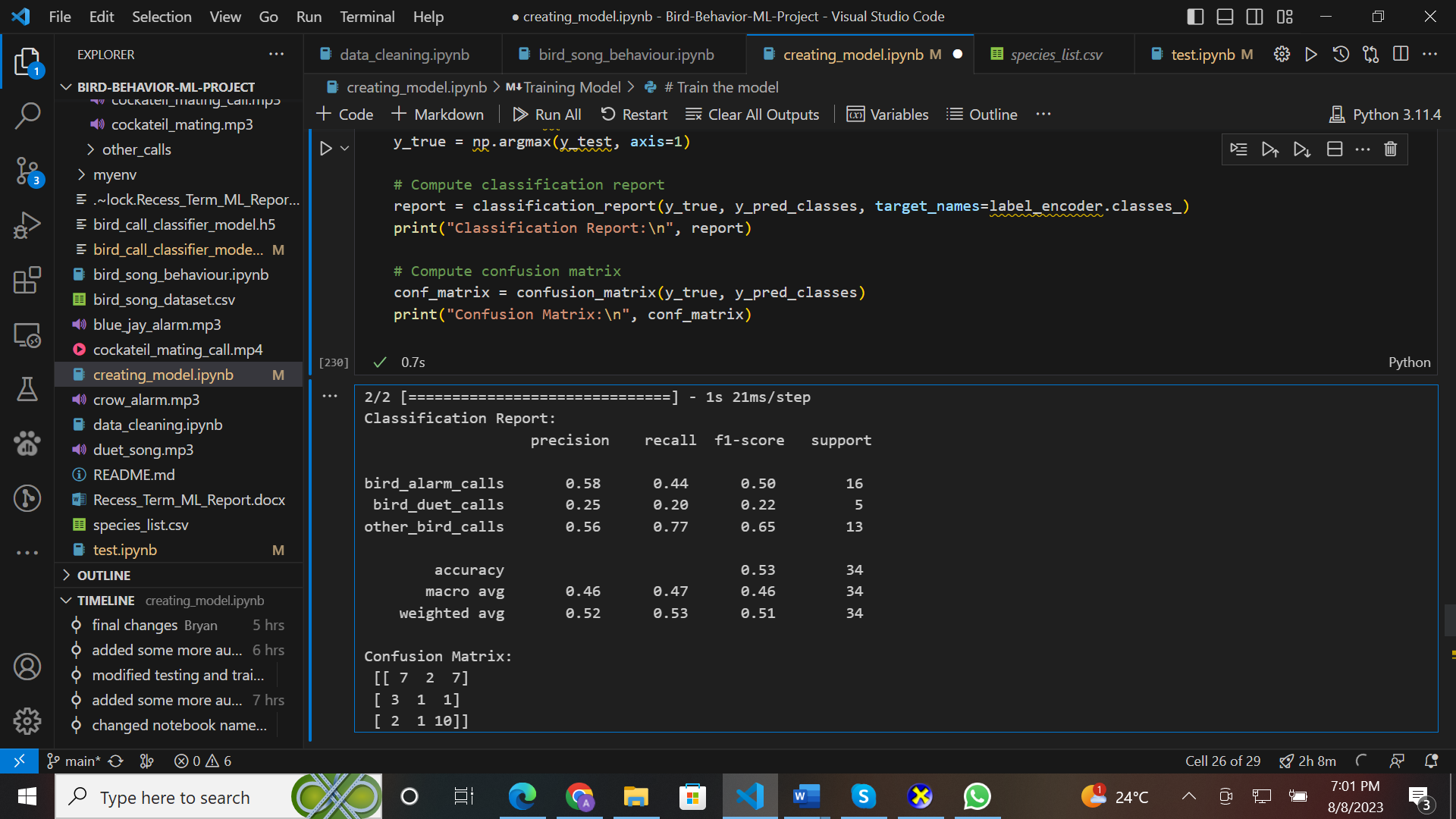


Figure 11: Confusion Matrix

## Performance Visualization

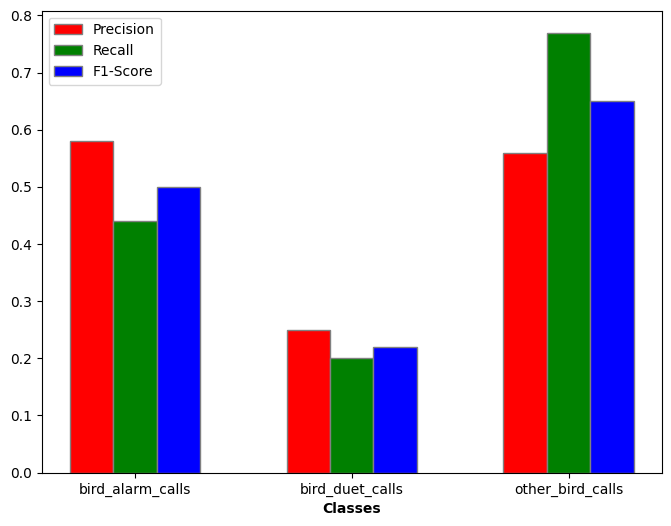


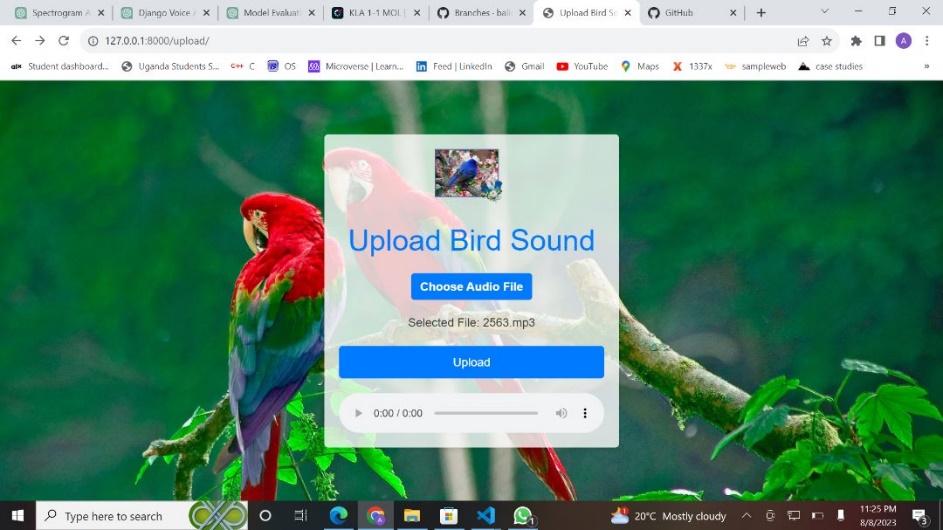
Figure 12: Evaluation Visualization

## Validation

To assess the model's ability to generalize, a validation set (30% of the dataset) was used during training to monitor for overfitting.

**11.Model Deployment.**

Using Django, our model was deployed to create a user Interface where a user is able to upload an audio file and the model detects whether it is an alarm call or a duet call. A user uploads an audio file which is analyzed by the voice analyzer class, passed through the model and returns the predicted result which is displayed by the webapp. The already analyzed sounds are stored in a database and can be viewed by the user if requested



# 12. Ethical Considerations

## Bias Mitigation

Care was taken to ensure that the dataset used for training was diverse and representative of various bird species and acoustic environments. Bias mitigation strategies were implemented to address any potential biases in the data.

**5. Software and Hardware**

## Software

The AI models were implemented using Python and popular deep learning libraries, including TensorFlow and Keras.

## Hardware

The training process utilized a machine with an i5 processor and 250 SSD with shared graphics memory to accelerate model training.

**6. Limitations**

## Data Limitations

The data for the mating calls (duet\_calls) and alarm calls is so limited and this is evident in the accuracy it depicts in evalutation

## Model Scope

The model's effectiveness is constrained to the specific bird behaviors of mating and alarm calls. It may not generalize well to other bird sounds or behaviors.

**7. Reproducibility**

## Code and Resources

The code for the AI models, data preprocessing, and training procedures is available on GitHub at <https://github.com/baliddeki/Bird-Behavior-ML-Project> .Top of Form