Bachelor Thesis Project- II

on

Classification and Prediction of Bird Species using Bird Song Recognition

Prepared By:- Soumya Ranjan Patra
Roll Number:- 18ME3AI30
Mechanical Engineering Department

under the guidance of

Dr. Akhilesh Kumar
Associate Professor
Department of Industrial and Systems Engineering

Indian Institute of Technology, Kharagpur
Spring Semester
(April,2022)

Contents

1. Introduction

Bird vocalizations are frequently used to conduct population surveys and distinguish different species of birds. The call rate is a good indicator of whether a bird population is steady, growing, or declining over time. It is also useful for identifying the species of a bird which is also the prime objective of this project. Acoustic Recorders offer long-term soundscape recording in a given location. Despite the fact that the recordings are automated, the analysis is still largely manual. While there is a large amount of research into the general problem of birdsong recognition, methods that are accurate in the presence of noise and that can detect birds that are far away from the microphone as reliably as expert humans are still not that good.

Manual recording analysis is a laborious and time-consuming task. It demands a high level of expertise, is not scalable, and is subject to observer bias. This bias is critical for rare birds, because their calls are likely to be misclassified if they aren't expected in a particular location, but conversely over-detected in regions where they are expected. Furthermore, because the populations are small, many of the birds are unlikely to be close to the microphone, therefore, making accurate identification of the calls when they are considerably damaged by sound attenuation is essential.

There are already many current projects to extensively monitor bird's species by continuously recording natural bird songs over long periods. There are a lot of methods used to identify bird species using bird song. For example, researchers have used neural network, deep convolutional neural network, ensemble approach, HMM models, etc.

2. Problem Statement

In this project I am using different models to predict bird species using bird call recognition. However, depending upon the computation power I have used various sizes of data set for different models.

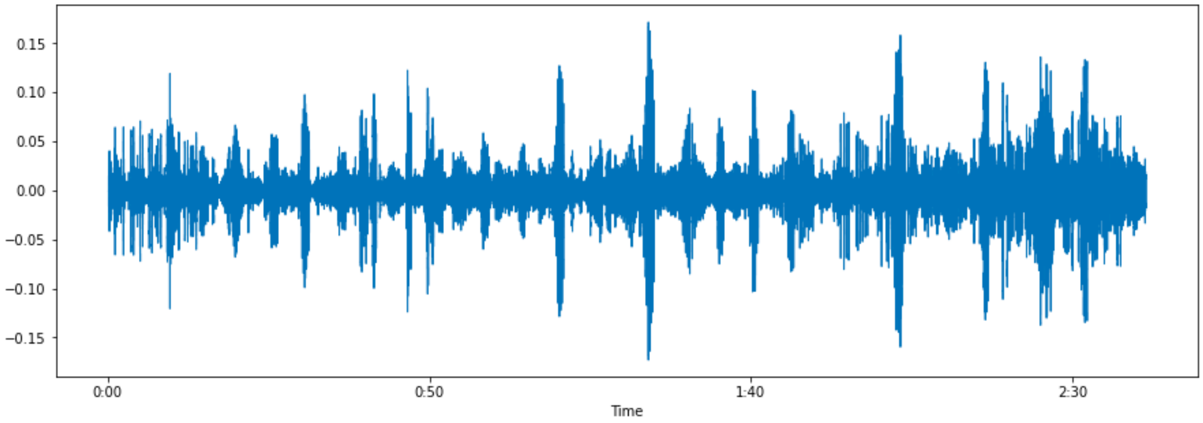
3. Data set

The data set consists of short recordings of individual bird calls downloaded from xenocanto.org. I used more than 4GB of data for training various models. The dataset had sound ranging from 5 seconds to 180 seconds. I used 20% of the data set for testing for each model. Most of the recordings were rated greater than 3.0.

4. Audio Processing

An audio signal consists of vital information. The capturing of such time-varying characteristics would help us distinguish between various types of audio. The recordings are in '.mp3' format. Using a sampling technique, these are converted into a one-dimensional array of digital values. This conversion is done using Analog Digital Conversion, which consists of sampling, quantisation, and encoding. These digital values represent the Amplitude frequency at that given instance. Here the sampling rate is fixed to 44,100 for all the recordings. I used the inbuilt **librosa** library in python for feature extraction.

Waveform of a sample bird song of species "lesning"



- Mel Scale:-

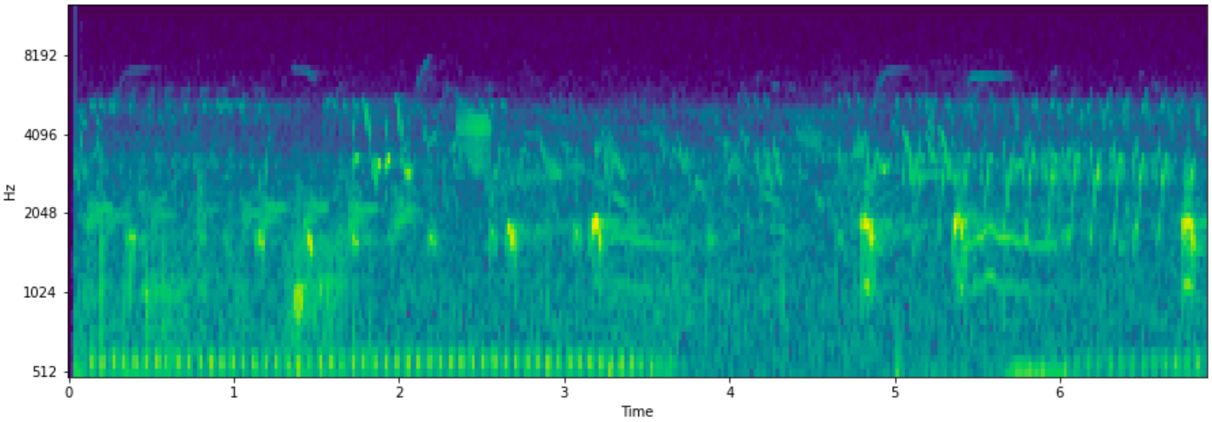
The Mel Scale is a logarithmic transformation of a signal's frequency. The transformation from the Hertz scale to the Mel Scale can be done using the following formula:

$$m = 1127 \cdot \log\left(1 + \frac{f}{700}\right)$$

- Mel-spectrogram:-

Spectrogram can be visualized as a bunch of FFTs stacked on top of each other. It is a way to visually represent a signal's loudness, or amplitude, as it varies over time at different frequencies. A **mel-spectrogram** is a spectrogram where the frequencies are converted to the mel scale.

Mel-spectogram of a sample bird call of species "lesning"



- Mel Frequency Cepstral Coefficients (MFCCs):-

MFCC is a compact representation of the spectrum, a primary feature in research areas that includes audio signals ranging from detecting cough sounds to automatic speech recognition. There are certain alterations in frequency in a bird sound. which are captured using MFCC and is a vital component in detecting features in audio recordings.

$$Magnitude(dB) = 10log\left(\frac{Spectral\ Power}{Spectral\ Power_0}\right)$$

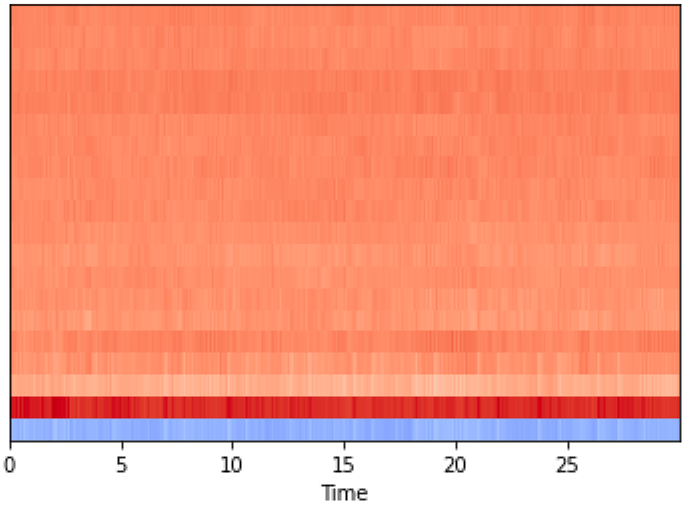
MFCC represents the sound spectrum by converting the audio signal via a sequence of steps.

$$Short\ Term\ Fourier\ Transform(\tau, f)\ =\ \int x(t)g(t\ -\ \tau)e^{-j2\pi ft}dt$$

$$Inverse\ Discrete\ Fourier\ Transform \rightarrow x(n)\ =\frac{1}{N}\sum_{k=0}^{N-1}X(k)e^{j\frac{2\pi kn}{N}}$$

$$Discrete\ Fourier\ Transform\ \rightarrow X(k)\ =\sum_{n=0}^{N-1}x(n)e^{-j\frac{2\pi kn}{N}}$$

MFCC of a sample bird call of species “lesning”



The MFCCs is a bit more decorrelarated, which can be beneficial with linear models like Gaussian Mixture Models. With lots of data and strong classifiers like Convolutional Neural Networks, mel-spectrogram can often perform better. The same thing I am doing in this project.

5. Convolutional Neural Network

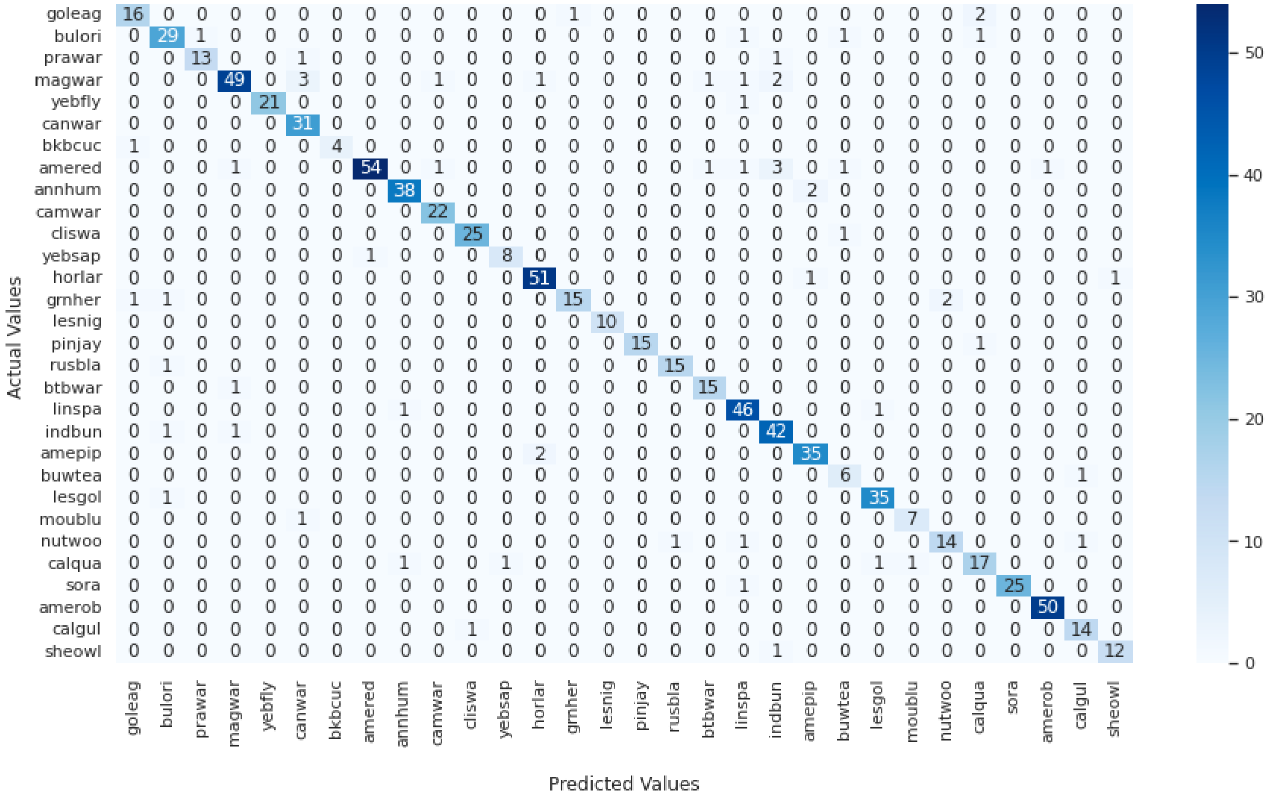
There are 4 convolutional layers. Each layer has a convolution layer of size (3,3) followed by a RELU activation, Batch Normalization and then MaxPooling in 2D of size (2,2). Further there is a global average pooling followed by 2 dense layers. At last there is an output layer with SoftMax activation. The model had 2,09,778 parameters. Batch size was 32.

- Feature Extraction:-

I have taken only the first 15 seconds of audio into consideration. That audio signal was divided into frames of 5 seconds and then features were extracted as a mel spectrogram and then stored as a '.png' file. Further I converted those images as numpy arrays. I normalized the data and excluded samples with 'nan'.

- Results:-

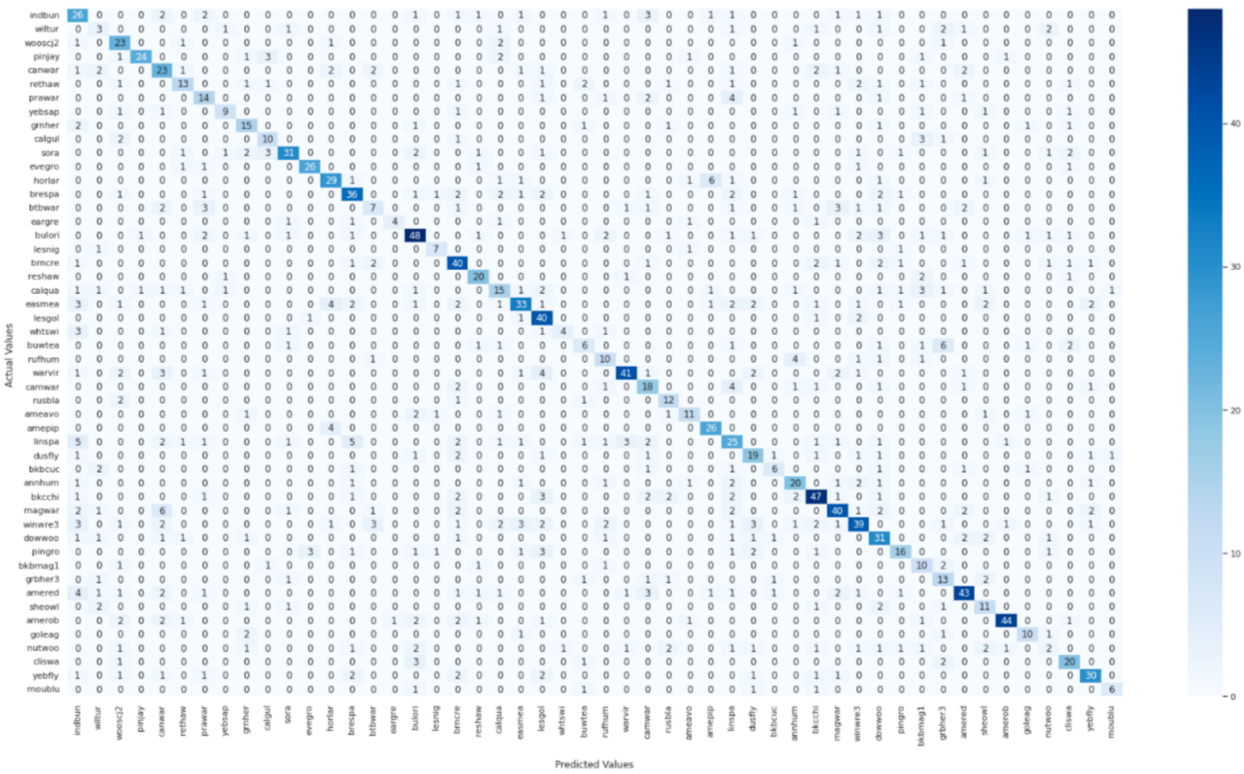
A. For **30** birds with 'rating' >=**4.0**.
It consisted of 1,562 audio files and 3,971 mel spectrograms.



	goleag	bulori	prawar	magwar	yebfly	canwar	bkbcuc	amered	annhum	camwar	cliswa	yebsap	horlar	grnher	lesnig	pinjay	rusbla	btbwar	linspa	indbun	amepip	buwtea	lesgol	moublu	nutwoo	calqua	sora	amerob	calgul	sheowl
goleag	16	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
bulori	0	29	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0
prawar	0	0	13	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
magwar	0	0	0	49	0	3	0	0	0	1	0	0	1	0	0	0	0	1	1	2	0	0	0	0	0	0	0	0	0	0
yebfly	0	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
canwar	0	0	0	0	0	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bkbcuc	1	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
amered	0	0	0	1	0	0	0	54	0	1	0	0	0	0	0	0	0	1	1	3	0	1	0	0	0	0	0	1	0	0
annhum	0	0	0	0	0	0	0	0	38	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
camwar	0	0	0	0	0	0	0	0	0	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cliswa	0	0	0	0	0	0	0	0	0	0	25	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
yebsap	0	0	0	0	0	0	0	1	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
horlar	0	0	0	0	0	0	0	0	0	0	0	0	51	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
grnher	1	1	0	0	0	0	0	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
lesnig	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pinjay	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	1	0	0	0	0
rusbla	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0	0
btbwar	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	0
linspa	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	46	0	0	0	1	0	0	0	0	0	0	0
indbun	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	42	0	0	0	0	0	0	0	0	0	0
amepip	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	35	0	0	0	0	0	0	0	0	0
buwtea	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	1	0
lesgol	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	35	0	0	0	0	0	0	0
moublu	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0
nutwoo	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	14	0	0	0	1	0
calqua	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	17	0	0	0	0
sora	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	25	0	0	0
amerob	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0
calgul	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0
sheowl	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	12

The confusion matrix shows that the model worked pretty fine for the audio set. It has a training accuracy of 99.06 and validation accuracy of 92.70. The above confusion matrix for test data validates that.

B. For **50** birds with rating>=**3.5**
It consisted of 3,266 recordings and 8,457 spectrograms.



It had a train accuracy of 81.98% and test accuracy of 62.66%. Also, the confusion matrix shows pretty good predictions. There are some similarities found in song of some pairs of birds. For example, “hortar” and “amedip” or “grbher3” and “buwtea”. This can be seen in the confusion matrix as the number of false prediction for them is more.

6. Gaussian Hidden Markov Model (HMM):-

- Feature Extraction:-

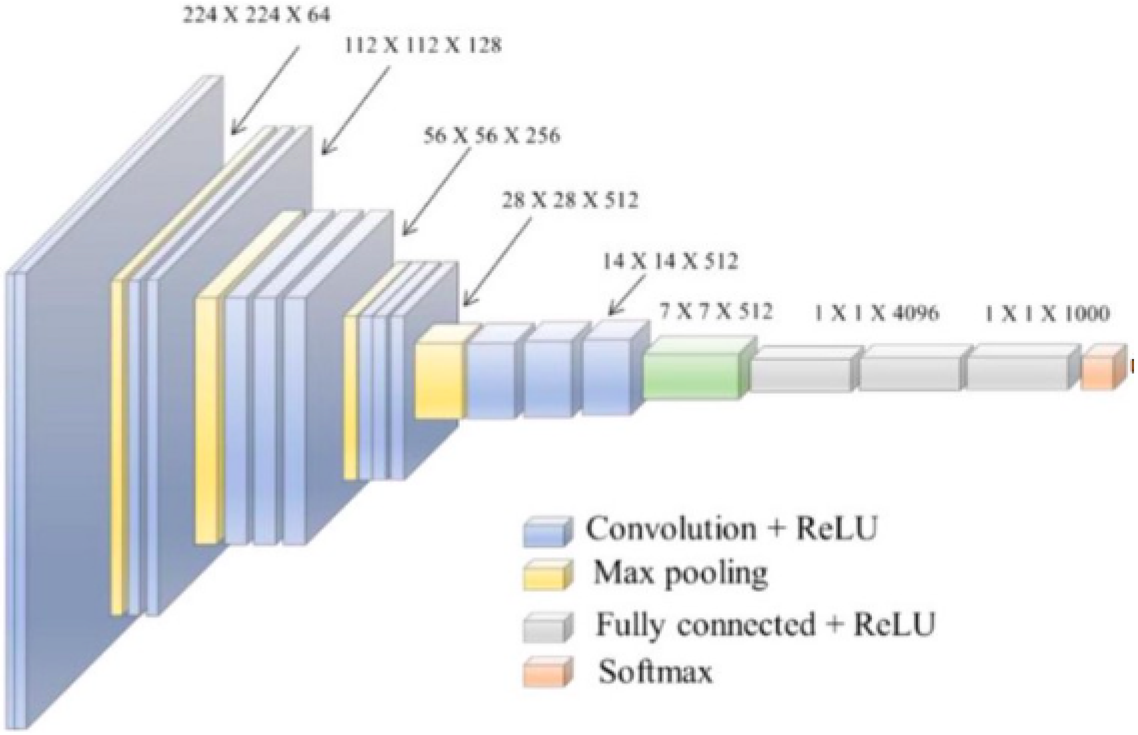
I took the audio recordings for 10 birds with a total of 734 recordings. I digitized the wave signal by converting the recordings into ‘.wav’ format. Used MFCC for extracting 40 features. Removed the samples containing ‘nan’ values. Then for each bird I made an array by appending all the features. I made one hmm model for each of the bird species. After training I used all the 10 models for getting the best score for each of the test data.

- Results:-

The result was not that good as I could only get an accuracy of 55.02 %. This can be reduced by removing some noises and improvising the data by techniques like cutting it.

7. VGG19 model:-

Below is the architecture of the model.

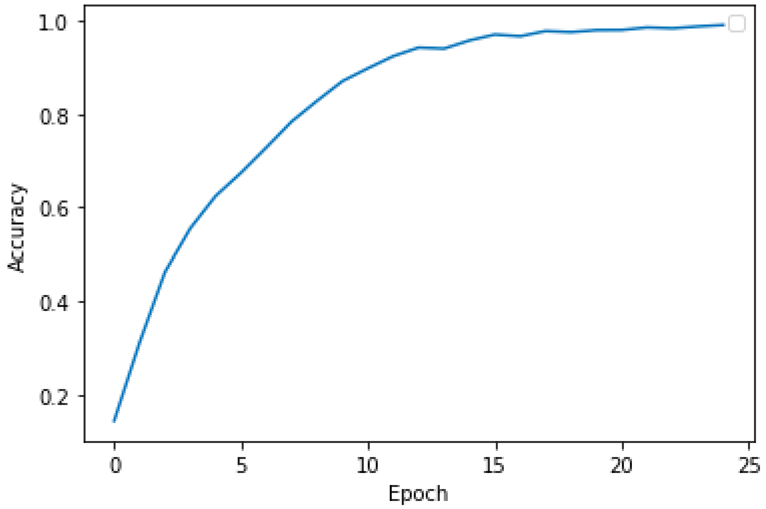


- Feature Extraction:-

I used audio recordings of 10 birds with a total 734 recordings. For all the birds, I removed silence by trimming preceding and succeeding zeros of the feature array. I then took the feature for only 1st five seconds and if the audio length was less than 5 seconds then I padded it with zero. Further features for each song were stored as a '.npy' file. And then used during the training. Batch size was 32.

- Results:-

I got a training accuracy of upto 98.70% as shown below. I could not get test accuracy for the model.



8. Conclusion

In this project I took an audio dataset of bird songs and tried to automatically identify the species of the bird using different machine and deep learning models. I used Convolutional Neural Network, Hidden Markov Model with gaussian emission and VGG 19 models for prediction. The VGG 19 performed the best while the HMM model performed the worst.

During the project I learned about processing audio data and using the features extracted to predict various things from it. I also got to know more about deep learning models. I learned about the HMM model and how to implement it.

Given higher computation power and better systems the models can be expanded for automatic prediction of bird species using pre pre-trained model of bird song on a large scale.