

AUDIO BIRD SONG PREDICTION MODEL WITH CONVOLUTIONAL NEURAL NETWORKS

“Application of machine learning to predict bird songs according to species”

**KATENDE CHRIS MARVIN 2100705689 21/U/05689/EVE**

**MUGENYI MARTIN 2100710044 21/U/10044/EVE**

**KYOMUHANGI FATUMAH 2100707188 21/U/07188/EVE**

**NAKAYE BABIRYE MARBLE 2100704986 21/U/04989/EVE**

**KIMULI ARNOLD MULIISA 2100704897 21/U/04897/EVE**

GROUP G

# 

# INTRODUCTION

A machine learning model to predict the bird species according to the kind of songs they make as one can say “ the shazam for bird songs”. This dataset was scrapped from the [https://xeno-canto.org](https://xeno-canto.org/) website that talks about sharing wildlife sounds all over the world, a csv file was obtained and over 1000 bird songs downloaded from the different API’s provided like

1. [**https://xeno-canto.org/api/2/recordings?query=cnt:brazil**](https://xeno-canto.org/api/2/recordings?query=cnt:brazil)
2. [**https://xeno-canto.org/api/2/recordings?query=troglodytes+troglodytes**](https://xeno-canto.org/api/2/recordings?query=troglodytes+troglodytes)
3. [**https://xeno-canto.org/api/2/recordings?query=bearded+bellbird+q:A**](https://xeno-canto.org/api/2/recordings?query=bearded+bellbird+q:A)

Python libraries like Beautifulsoup was used to in the data collection to scrap the website to obtain the data used in implementing the model which gives an 80% train accuracy and a 78% test accuracy.

# HYPOTHESIS

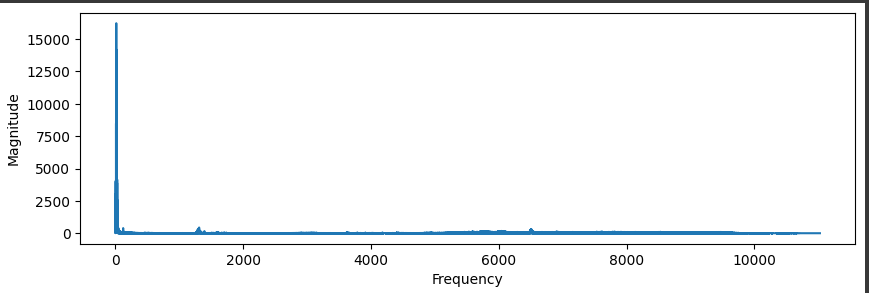
For visualization 3 plots were made like the;

1. Fast Fourier Transform Synthesis(Ffts)
2. Waveplot
3. Short time Fourier transform(stft)
4. Mel Frequency Cepstral Coefficients (mfccs)

# Ffts

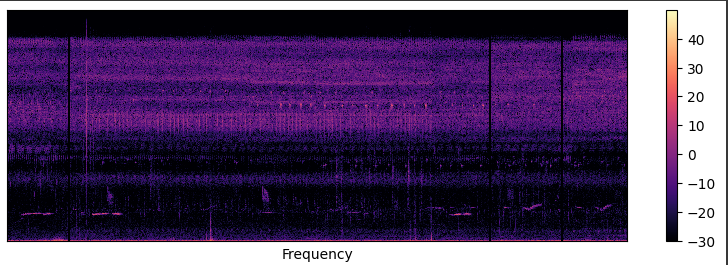
The FFTS algorithm focuses on efficiently and accurately reconstructing a time-domain signal from its frequency-domain representation, which can be useful in various applications such as audio synthesis, data compression, and signal reconstruction. The basic idea behind FFTS is to take advantage of the properties of the FFT to perform the synthesis process more efficiently. [More..](https://www.nti-audio.com/en/support/know-how/fast-fourier-transform-fft)

An audio was processed with the librosa library and with [this code](https://github.com/Martinkalz26/bird_song_classification/commit/6db04f453a5e7acdb188c8b555046e97c2f1897c), it appears to be processing a time-domain signal, computing its FFT, and then plotting the magnitude spectrum

Visualization of ffts

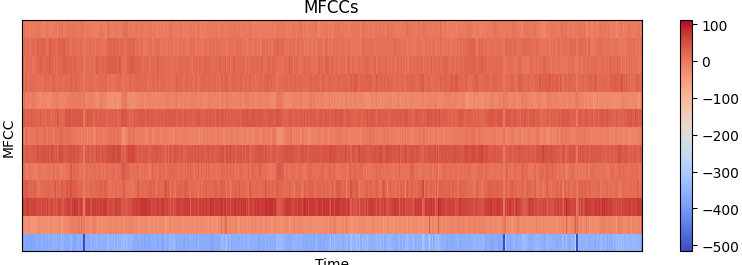
STFTs

The basic idea behind the STFT is to break down a signal into its frequency components over short time intervals. This is in contrast to the standard Fourier Transform, which gives you the frequency content of an entire signal without considering how it changes over time. And it was expressed with [this code](https://github.com/Martinkalz26/bird_song_classification/commit/6db04f453a5e7acdb188c8b555046e97c2f1897c).



MFCCs

MFCCs are derived from the Short-Time Fourier Transform (STFT) and are particularly effective in capturing the relevant characteristics of human speech and audio signals.MFCCs condense complex spectral information into a relatively small number of coefficients.it was expressed in python with [this code](https://github.com/Martinkalz26/bird_song_classification/commit/6db04f453a5e7acdb188c8b555046e97c2f1897c).



# PYTHON LIBRARIES USED

1. Librosa: It was used to load audios and playback and also extract features.
2. Numpy: Was specifically used to work with arrays
3. Pandas: Used to work with dataframes
4. Matplotlib: Used for plotting/ visualizing
5. Json: Used to serialize our data that is features extracted from the audio dataset.
6. Math: Used for handling math ie use ,ceil method to round to nearest integer
7. Os: Used to navigate directories.
8. Pickle: Used to save the model
9. tensorflow/keras: used in training the model

# PROCEDURE

**DEVELOPING A CNN ARCHITECTURE**

This involves building the model, training, testing and validation. This architecture involves varies layers namely below:

* Input layer; designed to accept the preprocessed MFCC-based feature vectors.
* Pooling layer; max pooling /average pooling to reduce spatial dimer and help in feature down sampling and generalization
* Convolutional layers; captures local patterns in the feature vectors
* Flatten layer; flatten the output from the last pooling layer into a I.D vector to prepare it to the fully connected layers.
* Fully connected layers; add one or more fully connected (dense) layers after the flatten layer to process the higher level features learned by convolutional layers.
* Output layers; compile the model using an appropriate optimizer e.g. RMSprop and a suitable loss fraction for your specific task (e.g categorical cross –entropy for classification ,mean squared error for regression )

# DATA

Data used was downloaded from [xeno-canto website](https://xeno-canto.org/explore/api) then grouped in 16 folders [Audio Dataset](https://drive.google.com/drive/folders/1eEsH1aKTBMOxyXsj-lPHU3jYdtJ5WzWE?usp=sharing) along with a [csv file](https://drive.google.com/file/d/1M_Oxf6_ICeHfbEljD4-AIDvDqH9MoF6f/view?usp=sharing) and it was serialized into a [json file](https://drive.google.com/file/d/1HJfqEPq2c1_XN9fPY9k2JA_XhxsS12z2/view?usp=sharing) that stores a dictionary of keys[mappings,mfccs and labels]

# RESULTS

An eighty percent (80%) train accuracy was achieved with unaugmented data but would have been higher with more data if was provided and also a promising seventy eight percent test accuracy was achieved(78%)

# CONCLUSION

In this document, we've explored the fascinating realm of applying machine learning techniques to the analysis of birds' songs. Birdsongs are rich sources of information, reflecting the diverse behaviors, habitats, and species within avian populations. By harnessing the power of machine learning, we've unlocked new avenues for understanding, classifying, and interpreting these intricate vocalizations.

Throughout this document, we've highlighted several key insights and takeaways:

Data Acquisition and Preprocessing: The foundation of any successful machine learning endeavor is a well-structured dataset. We've emphasized the importance of collecting high-quality audio recordings of birdsongs across varying environments, species, and contexts. Proper preprocessing, including noise reduction, audio segmentation, and feature extraction, forms the crucial initial step in preparing the data for model training.

Feature Engineering: Extracting meaningful features from audio data is a complex task. We've discussed the utilization of techniques like spectrogram generation, Mel-frequency cepstral coefficients (MFCCs), and other time-frequency representations. These engineered features serve as the input to machine learning models, capturing critical information within the songs.

Model Selection: Our exploration into machine learning models encompassed a range of approaches, from traditional methods like Support Vector Machines (SVMs) to modern deep learning techniques like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). The choice of model depends on the complexity of the task, available resources, and desired level of accuracy.

Classification and Segmentation: We delved into the realm of songbird species classification and segmentation. Machine learning models can accurately differentiate between various species based on their distinct vocal patterns. Furthermore, techniques such as Hidden Markov Models (HMMs) and dynamic time warping enable the identification of unique song elements within longer sequences.

Ecosystem Monitoring: The application of machine learning extends beyond species identification. By analyzing the temporal and spatial patterns of birdsongs, we gain insights into avian behavior, migratory patterns, and the impact of environmental changes on bird populations. This has valuable implications for biodiversity conservation and ecosystem health assessment.

Challenges and Future Directions: While machine learning has unlocked remarkable capabilities in birdsong analysis, challenges persist. Addressing issues like limited labeled data, variability in vocalizations, and cross-species generalization remain active areas of research. Future directions include leveraging advances in transfer learning and data augmentation to mitigate these challenges.

In conclusion, the fusion of machine learning and birdsong analysis opens up a world of possibilities for ornithologists, ecologists, and researchers. The ability to accurately classify species, decode communication signals, and monitor ecosystems provides profound insights into avian behavior and the natural world. As technology continues to evolve, the synergy between machine learning and ornithology promises to reshape our understanding of the melodies that fill the air around us.

# FUTURE WORKS

To deploy our model using stramlit app

# 

# REFERENCES

1. [NTi description of ffts](https://www.nti-audio.com/en/)
2. [xeno-canto website](https://xeno-canto.org/explore/api)
3. [Youtube](https://youtu.be/dOG-HxpbMSw)
4. [IEEE](https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9073584)