Identification of bird species using audio classification techniques

SE333 MINOR PROJECT

Aryaman Jain 2K21/SE/38

Problem Statement

In this project we attempt to create an audio classification model to detect which species a bird belongs to by processing its sound. This tool can be used by avid bird spotters to detect a bird from its sound as well as by bird parks and sanctuaries to detect which bird is present where in real time.

Scope of the Project

Development of a Robust Classifier

To create a machine learning model capable of accurately classifying bird sounds from audio clips. The target is to achieve an accuracy rate near or above 90%.

Handling various bird Species

To ensure the model is versatile enough to classify sounds from a range of species.

Ecological and business objectives of Model Implementation

Through this project we hope to make an application to assist in two tasks-

- Being a reliable companion of bird watchers to help them classify birds from their sounds and also read more about birds in the application.
- For bird parks and sanctuaries, we hope to implement a bird geolocation system by installing clusters of microphones which are connected to our model in bird parks to provide real time information of presence of varied species of birds in different parts of the park to enhance visitor experience as well as administration efficiency of forest officials.

Description of the Dataset Used

1 Source and Composition

The "Bird Song Dataset" sourced from Kaggle was considered for the study

It has been compiled from xeno-canto.org, and contains recordings from 5 bird species.

2 Format and Quality

Audio files are provided in .wav with varying lengths and quality. This variation mimics real-world conditions, adding to the dataset's practical value.

3 Metadata

Comes with 1 .csv file containing metadata for over 9107 .wav recordings.

The dataset contains 18 columns, but only 2 classes are relevant for us - species and filename (.wav)

Variables Used

1. ### Initial Variables

   Species and filename were the initial variables from the dataset.

2. ### Feature Extraction

   Conversion of audio files into numerical format and extraction of MFCCs.

3. ### Technologies Used

   Librosa, TensorFlow, Keras, Numpy, Pandas, and Matplotlib were utilized.

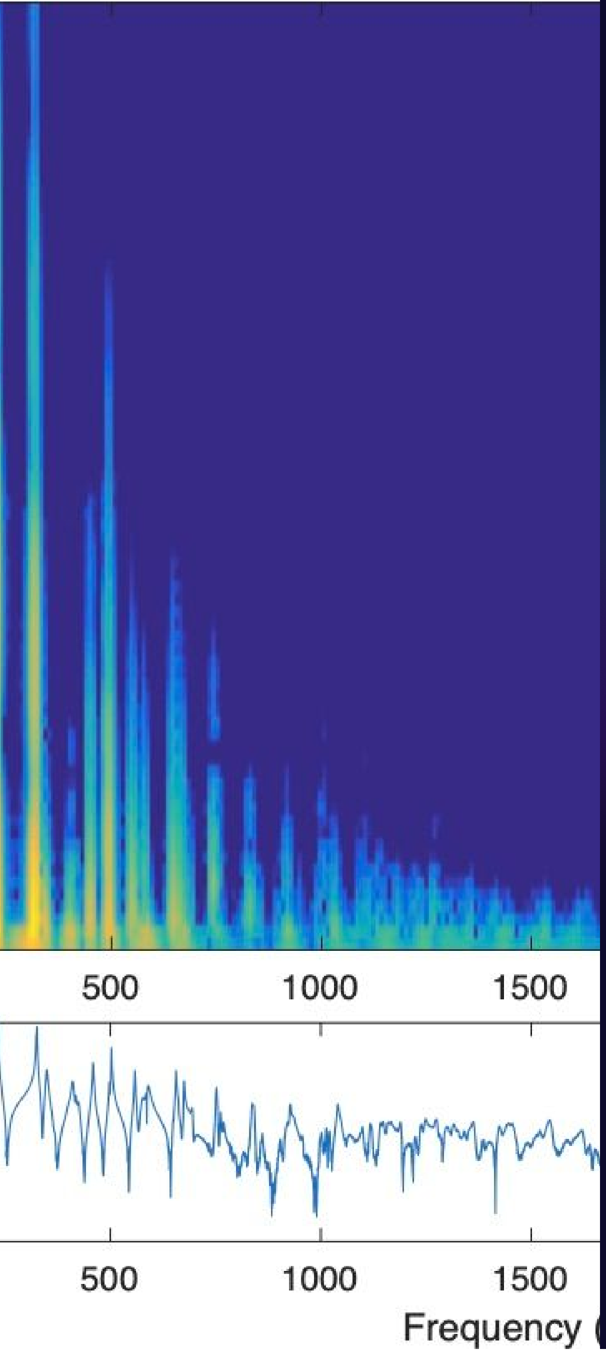
Exploratory Data Analysis (EDA) And Data Processing

Sound Wave Exploration

Exploration of sounds in the dataset for format, size, and other relevant characteristics. Used matplotlib to see mono and sterio channels

Sample Rate Extraction

Using Librosa, Data and sample rate of an audio file can be extracted. Sample rate is the number of samples per second. Data is a one-dimensional array that represents the amplitude of the audio signal at each point in time.



Feature Extraction and Engineering

1

Data and Sample_rate

Numerial representation of the sound. We use Librosa's MFCC function which generates an MFCC from this audio data.

2

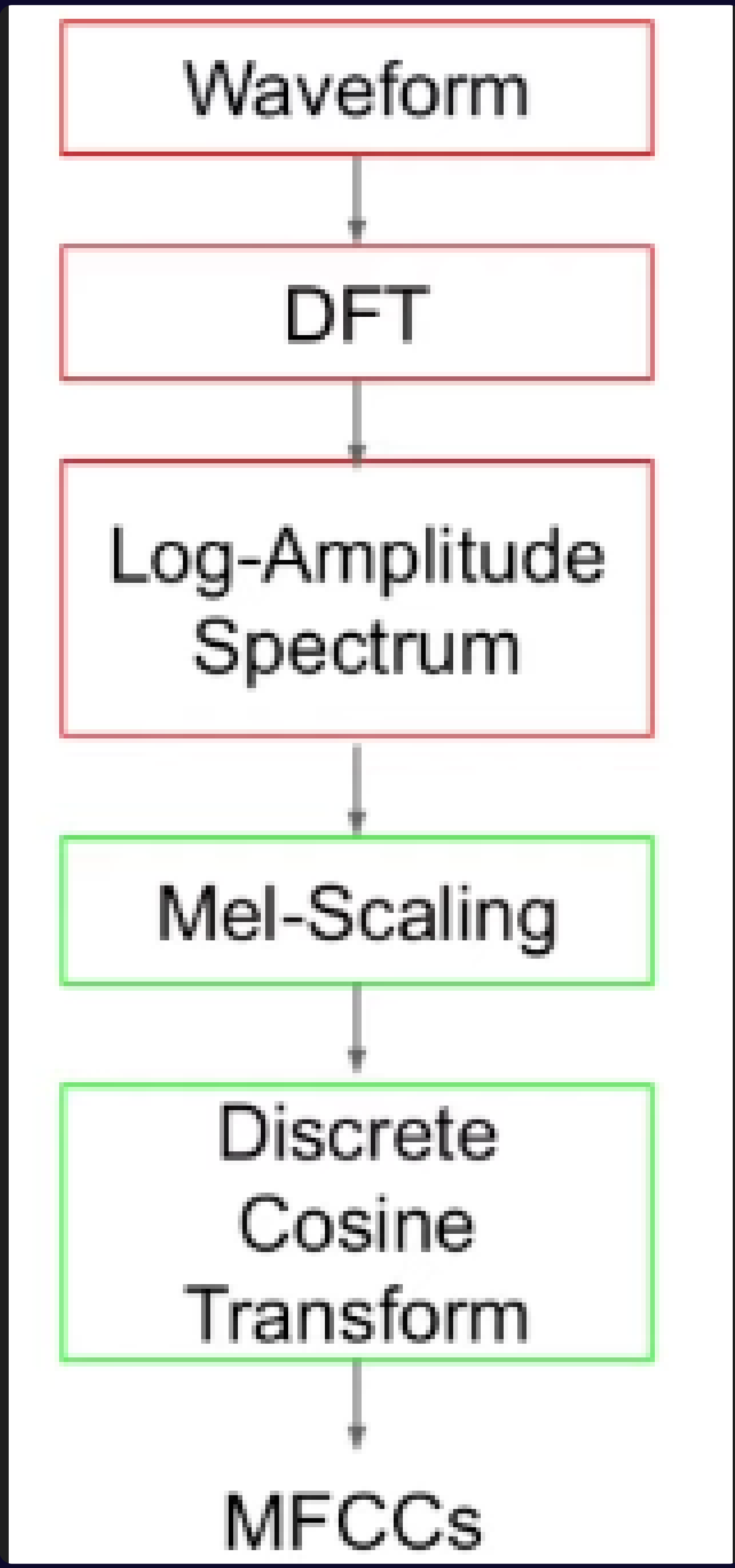
Mel-Frequency Cepstral Coefficients (MFCC)

Creation of an image representation of each audio sample to identify features for classification.

3

Storage of MFCCs

Storage of MFCCs in a separate dataframe for later use.



MFCCs are a type of audio features.

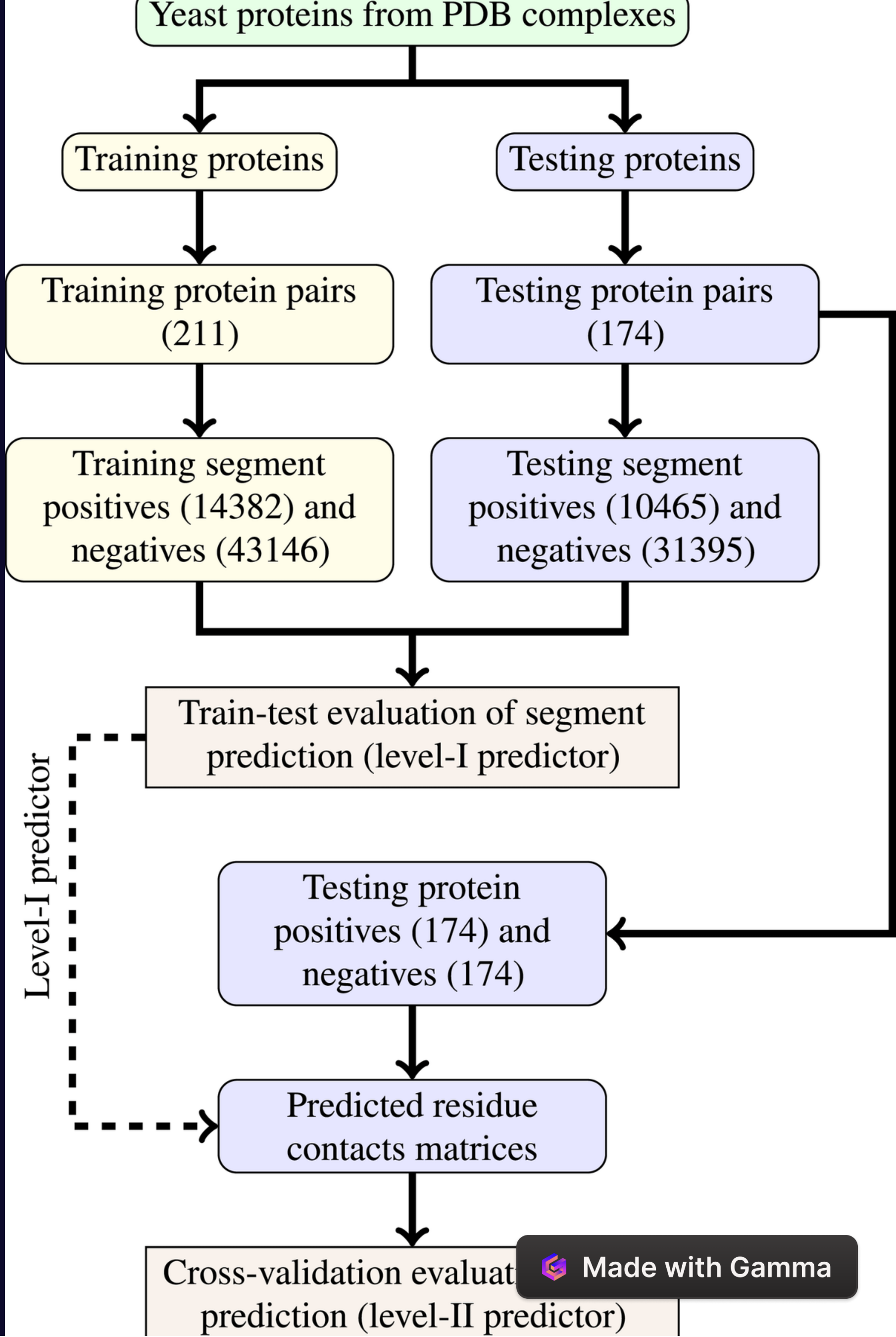
A waveform in numerical form (with sample_rate and data) has to undergo multiple fourier transformations and mel scaling to become MFCCs.

Train Test Splitting

Creation of training and testing sets for features and labels.

Label Encoding

Encoding of the categorical data is done to convert them into numerical format for use by models.



Model Building

We implemented 3 models for classification and ANN was the clear winner

Support Vector Machine

Accuracy of 65.81%.

CNN

Accuracy of 64.79%.

ANN

Accuracy of 89.95%.

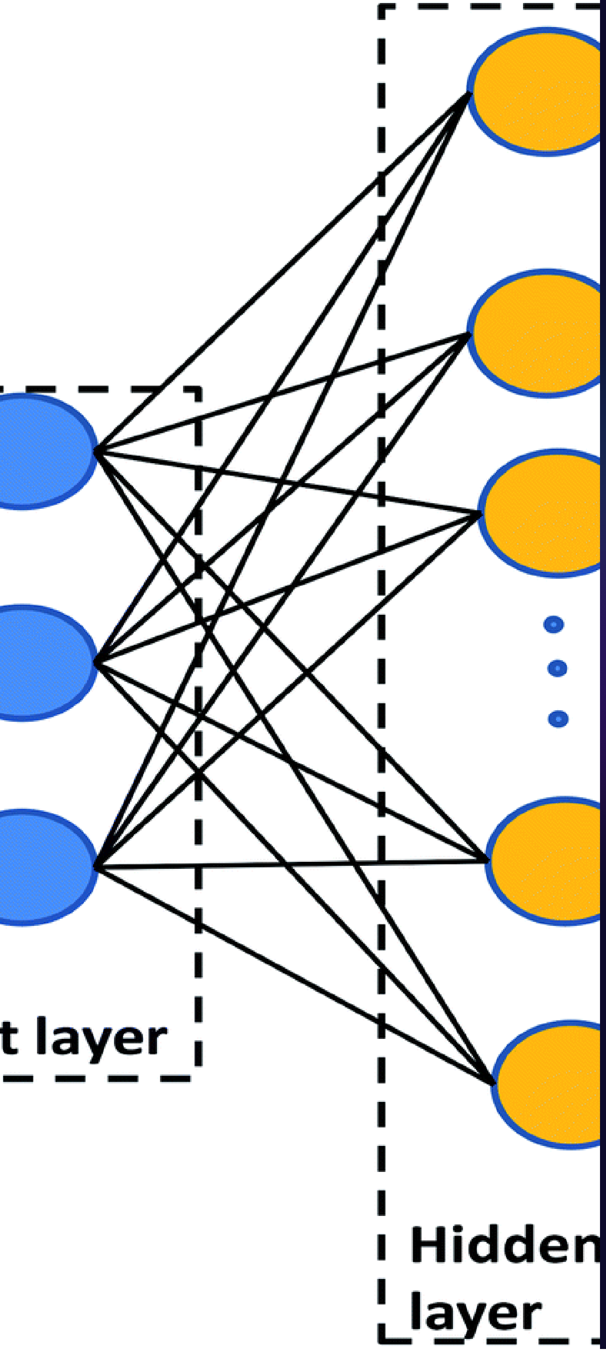
Model Testing and Evaluation

1 Performance Measures

Utilization of confusion matrix, accuracy, recall, and F1 score to evaluate the model.

2 Model Storage

Saved trained CNN models in Hierarchical Data Format (H5) files for easy integration.



Results

1 **Best Performing Model**

ANN showed an accuracy of 89.95% and should be chosen to proceed with the project.

2 **Evaluation on the ANN**

Confusion matrix, precision, recall, and F1-score revealed the model's high performance.

Conclusion and Key Takeaways

1. ANN performs the best out of the 3 implememented models as clearly evaluated by the various performance measures like precision, recall, etc.

2. Data quality and feature engineering are crucial to Model accuracy.

   Collect reliable data and extract key features to improve the model's accuracy.

3. In conclusion, the performance of our classification model can be regarded as highly satisfactory, as evidenced by the impressive accuracy score of **90%**