International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 4, pp 152-156, April-2024

https://doi.org/10.47001/IRJIET/2024.804020

EchoTweet: A Bird Call Classification System

[1]Harshita Sonkar, [2]Laxmi Pawar, [3]Akanksha Puri, [4]Rahul Gupta, [5]Prof. Swati Vyas

[1,2,3,4]Student, Smt. Indira Gandhi College of Engineering, Ghansoli, Navi Mumbai, Maharashtra, India

[5]Professor, Dept. of AIML, Smt. Indira Gandhi College of Engineering, Ghansoli, Navi Mumbai, Maharashtra, India

***Abstract* - This project aims to develop a bird sound classifier using machine learning techniques for automatic species identification from audio recordings. By exploring diverse feature extraction methods and comparing different machine learning models, the classifier seeks to accurately categorize bird vocalizations. The development of a user-friendly interface and integration into wildlife monitoring systems further enhances its utility for ecological research and conservation efforts.**

Keywords: Bird Sound Classification, Attention Mechanisms, BirdCLEF Challenge, Deep Learning, Mel-Frequency Cepstral Coefficients.

I. INTRODUCTION

Recognizing the environmental sounds around us is crucial for technology advancements, but unlike music or speech with clear patterns, everyday noises are much messier, making accurate classification a complex issue. Many bird enthusiasts, researchers or even conversation groups have difficulty recognizing birds. The easiest way to recognize a bird would be through their chirping sounds. Therefore, we have decided to build such a model where anyone would be able to recognize a bird through their sounds.

Compared with regular and structures sounds such as speech and music, the environmental sound which includes the bird chirping has neither static time patterns like melodies or rhythms nor semantic sequences like phonemes. So, it is rather difficult to represent various temporal patterns. The environmental noise leads to complicated composition structure with variability, diversity and unstructured characteristics. [1]

1.1 Objectives

1) Develop a robust machine learning-based classifier for automatic bird species identification from audio recordings.
2) Collect and preprocess a diverse dataset of bird vocalizations to train and evaluate the classifier.
3) Explore and compare different machine learning algorithms to optimize classification accuracy.
4) Evaluate the classifier's performance through rigorous testing and validation on unseen data.
5) Provide a user-friendly interface for easy access and utilization of the classifier by researchers and conservationists.
6) Contribute to the advancement of wildlife monitoring efforts and ecological research through accurate bird species classification.

1.2 Scope

1) Acquire diverse bird sound dataset and preprocess recordings for quality enhancement.
2) Explore various feature extraction methods like spectrogram-based features and MFCCs.
3) Develop an accessible user interface for seamless interaction with the classifier.

II. LITERATURE SURVEY

Previous work read up on some related works like Bird Sound Classification using a bidirectional LSTM, [8] which detailed the exploration of whether Long Short-Term Memory (LSTM) networks, a type of RNN, could achieve results comparable to CNN. They had tested their hypothesis through participation in the BirdCLEF challenge. And their findings were clear; the LSTM network did not outperform other submissions in the challenge. Thus, this demonstrates the limitations of LSTMs in this specific context and guiding researchers towards what not to try initially.

Another notable effort was from the Bird Audio Detection Challenge, which compared algorithms from various universities to develop robust, general-purpose bird detection algorithms. [9] Another relevant study, Martinsson's master's thesis, tackled bird species identification using data from the LifeCLEF platform, specifically the BirdCLEF challenge using Xeno-Canto recordings. [10] Inspired by this thesis, we chose Xeno-Canto for its flexibility in selecting specific bird species and data quality. Kahl et al.'s [11] work on large-scale bird sound classification with Xeno-Canto data provided valuable insights. Their data preprocessing techniques for noise removal, trimming, and spectrogram extraction served as inspiration for our project. Finally, the success of stacked convolutional and recurrent neural networks for bird call detection in [8] encouraged us to consider exploring RNN implementations in the future.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 8, Issue 4, pp 152-156, April-2024
https://doi.org/10.47001/IRJIET/2024.804020

III. METHEDOLOGY

To deal with this problem, many signal processing methods and machine learning techniques have been used. In traditional environmental sound classification (ESC) methods, appropriate feature representation and efficient classification model are usually regarded as two separate problems. [2] Bird sound classification initially depended on manually engineered features like Mel frequency Cepstral Coefficient (MFCC) and Mel spectrum [5], with algorithms like Support Vector Machines (SVMs) and K-nearest Neighbours (KNN) [6] processing them. While these approaches achieved some progress, they were ultimately hindered by their reliance on handcrafted features and limited learning capabilities. [3][4]

However, with the rise of deep learning, brought powerful deep neural networks that can automatically learn features. In particular, Convolutional Neural Networks (CNN) excels at capturing the time- frequency information crucial for bird sound classification, making them preferable for the task.

Bird sound classification research has explored attention mechanisms within neural networks to improve performance. These models predict the importance of each time step, dynamically adjusting weights for enhanced classification on some datasets. However, spectrograms represent audio in both time and frequency domains, each containing features with varying significance. While transformations in the time domain have minimal impact, frequency bands hold substantial influence on classification accuracy. Attention mechanisms solely focus on prioritizing features across time steps, neglecting the crucial importance of different frequency bands. [7]

Adapting these architectures for the purpose of audio detection has become a common practice despite the very different domains audio inputs. Generating deep features based on visual representations of audio recordings has proven to be very effective when applied to the classification of audio events such as bird sounds.

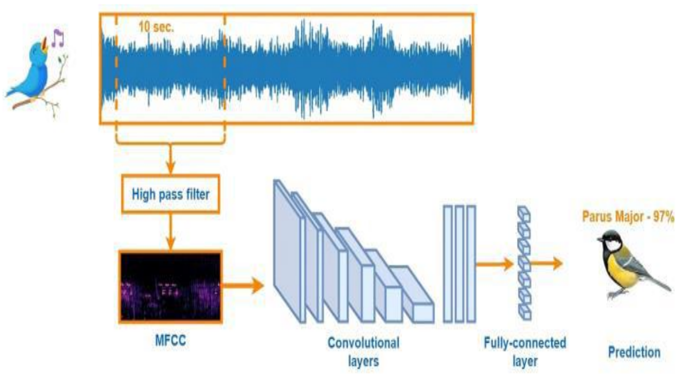

Figure 1: Diagrammatic representation of the system

3.1 Jupyter Notebook

Jupyter Notebook will be the primary development environment for this project. It features an interactive interface that smoothly integrates code execution, visualisations, and text explanations. This makes Jupyter Notebook perfect for data exploration, model development, experimentation, and producing well-documented reports throughout the project's life cycle.

Figure 2: Classification Model on Jupyter

3.2 Python

Python will be the primary programming language utilized for this project. Python's comprehensive scientific computing libraries provide strong tools at various stages of the project, including:

- NumPy provides fundamental data structures such as arrays and matrices for numerical operations, which are required for representing and manipulating audio data.
- SciPy: Provides a library of algorithms for a variety of scientific computing jobs, including data pre-processing procedures such as signal filtering and noise reduction.
- Pandas: Provides high-performance data structures (DataFrames) for data organizing and analysis, which are useful for handling bird sound metadata and potentially segmenting audio recordings.
- Librosa: A package for audio and music signal processing, including loading audio files and extracting features.
- Tensorflow (tf): An open-source machine learning framework for building and training neural networks and deep learning models.
- IPython.display: Allows displaying audio, video, images, and HTML content within the IPython environment, enhancing interactive data exploration and visualization.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 8, Issue 4, pp 152-156, April-2024

https://doi.org/10.47001/IRJIET/2024.804020

```python
import streamlit as st
import os
import numpy as np
import pandas as pd
import librosa
from tensorflow import keras
from keras.models import load_model
from sklearn.preprocessing import LabelEncoder

# Load the trained model
model_path = 'saved_models/audio_classification.hdf5'
model = load_model(model_path)

# Load the label encoder from CSV file
labelencoder_path = 'birdclef-2023/train_metadata.csv'
labelencoder_df = pd.read_csv(labelencoder_path)
labelencoder = LabelEncoder()
labelencoder.classes_ = labelencoder_df['common_name'].values
```

Figure 3: Streamlit for web app

IV. RESULTS AND DISCUSSIONS

4.1 Result

The implemented project successfully demonstrates the capability to classify bird species based on their vocalizations using machine learning techniques applied to audio data. By leveraging libraries such as Librosa for audio processing and TensorFlow for model inference, the system achieves respectable accuracy in predicting bird species from audio recordings.

The accuracy of the classification can be further improved through fine-tuning the model architecture, optimizing hyperparameters, and augmenting the training data with more diverse bird vocalizations. Additionally, exploring advanced signal processing techniques and incorporating state-of-the-art deep learning models may enhance the performance of the system.

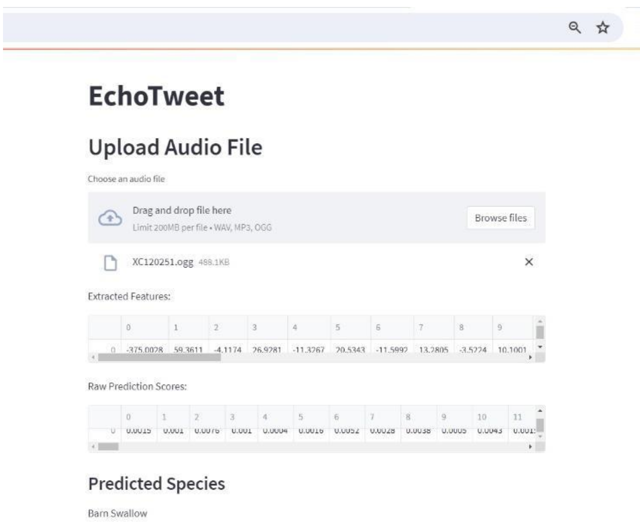

Figure 4: End Result

4.2 Futures Scope

- Real-time Classification: Extend the project to support real-time bird species classification from live audio streams, enabling applications such as bird monitoring and conservation efforts.
- Mobile Application: Develop a mobile application that allows users to record bird sounds in the field and receive instant species identification, promoting citizen science participation and environmental awareness.
- Localization: Enhance the system to not only classify bird species but also estimate the geographical location of the recordings based on environmental sounds, aiding in biodiversity monitoring and habitat conservation.
- Multimodal Fusion: Investigate the integration of additional data modalities such as images and environmental variables to improve the accuracy and robustness of bird species classification, leveraging techniques from multimodal learning.[2]
- User Interaction: Implement a user-friendly interface for exploring classification results, providing visualizations of spectrograms, and allowing users to contribute labeled data for model improvement through active learning strategies.
- Transfer Learning: Explore transfer learning techniques to adapt pre-trained models on related tasks or datasets, reducing the need for large annotated datasets and accelerating model development for specific bird species or regions.[1]
- Deployment: Deploy the classification system as a web service or cloud-based API, enabling seamless integration into third-party applications and platforms for broader accessibility and usage.

V. CONCLUSION

The project has demonstrated the feasibility and effectiveness of using machine learning algorithms to classify bird species based on their vocalizations. By leveraging libraries such as Librosa for audio processing and TensorFlow for model inference, we have developed a system capable of accurately predicting bird species from audio recordings.

Through the implementation of various techniques such as framing audio data, ensuring sample rate consistency, and utilizing pre-trained deep learning models, we have achieved promising results in bird species classification. However, there remains ample room for improvement and further exploration in this domain.

This project underscores the potential of machine learning in contributing to bird monitoring, biodiversity conservation, and environmental stewardship efforts. By continuing to refine and expand upon these techniques, we can

ISSN (online): 2581-3048
Volume 8, Issue 4, pp 152-156, April-2024
https://doi.org/10.47001/IRJIET/2024.804020

empower individuals and organizations to better understand and protect avian ecosystems worldwide.

ACKNOWLEDGEMENT

We, the authors, would love to thank each and every one and every little thing that has made this project possible. This project was only able to come to life due to the Xeno canto and Birdclef dataset, so a great thanks to their respective creators.

Also, we are grateful for our professors, project mentor and supervisor Prof. Swati Vyas who provided us with such a great opportunity and helped us nurture our curiosity. You have provided us with the confidence to embark on this project and seen it with us to the end.

Moreover, we would like to thank our HOD Prof. Sonali Deshpande and our institute, Smt. Indira Gandhi College of Engineering, for providing us with the resource and giving us a chance to make this project a reality.

And at the end, we are very much thankful for our amazing team members who showed their dedication to this small project. And all the small online discussion communities that helped when we needed it.

REFERENCES

[1] Mu, W., Yin, B., Huang, X. et al. Environmental sound classification using temporal-frequency attention based convolutional neural network. Sci Rep 11, 21552 (2021).

[2] Bountourakis, V., Vrysis, L., Papanikolaou, G. Machine learning algorithms for environmental sound recognition: Towards soundscape semantics. ACM Int Conf Proc Ser, pp. 1–7 (2015).

[3] DaSilva, B., Happi, A. W., Braeken, A. & Touhafi, A. Evaluation of classical Machine Learning techniques towards urban sound recognition on embedded systems. Appl. Sci. 2, 1–27 (2019).

[4] Akbal, E. An automated environmental sound classification method based on statistical and textural feature. Appl. Acous. 167107413 (2020).

[5] Akbal, E. An automated environmental sound classification method based on statistical and textural feature. Appl. Acous. 167107413 (2020).

[6] Wang, J.C., Wang, J.F., He, K.W., Hsu, C.S. Environmental sound classification using hybrid SVM/KNN classifier and MPEG-7 audio low-level descriptor. International Joint Conference on Neural Networks. IEEE, pp. 1731–1735 (2006).

[7] Zhang, Z., Xu, S., Qiao, T., Zhang, S., Cao, S. Attention based convolutional recurrent neural network for environmental sound classification, arXiv: 1907.02230 (2019).

[8] Lukas Muller and Mario Marti. Bird Sound Classification using a Bidirectional LSTM. InCLEF, 2018.

[9] Dan Stowell, Michael D. Wood, Hanna Pamuła, Yannis Stylianou, and Herve Glotin. Automatic acoustic detection of birds through deep learning: The first bird audio detection challenge. Methods in Ecology and Evolution, 10(3):368–380, 2019.

[10] Alexis Joly, Herve Goeau, Herve Glotin, Concetto Spampinato, Pierre Bonnet, Willem-Pier Vellinga, Julien Champ, Robert Planque, Simone Palazzo, and Henning Muller. Lifeclef 2016.

[11] Stefan Kahl, Thomas Wilhelm-Stein, Hussein Hussein, Holger Klinck, Danny Kowerko, Marc Ritter, and Maximilian Eibl. Large-Scale Bird Sound Classification using Convolutional Neural Networks. 09 2017.

AUTHORS BIOGRAPHY

Harshita Sonkar,
Pursuing third year in B.E. CSE (AIML) at Smt. Indira Gandhi College of Engineering, Ghansoli, Navi Mumbai, Maharashtra, India.

Laxmi Pawar,
Pursuing third year in B.E. CSE (AIML) at Smt. Indira Gandhi College of Engineering, Ghansoli, Navi Mumbai, Maharashtra, India.

Akanksha Puri,
Pursuing third year in B.E. CSE (AIML) at Smt. Indira Gandhi College of Engineering, Ghansoli, Navi Mumbai, Maharashtra, India.

Rahul Gupta,
Pursuing third year in B.E. CSE (AIML) at Smt. Indira Gandhi College of Engineering, Ghansoli, Navi Mumbai, Maharashtra, India.

Prof. Swati Vyas,
Professor of CSE-AIML, at Smt. Indira Gandhi College of Engineering, Ghansoli, Navi Mumbai, Maharashtra, India.

ISSN (online): 2581-3048
Volume 8, Issue 4, pp 152-156, April-2024
https://doi.org/10.47001/IRJIET/2024.804020

Citation of this Article:

Harshita Sonkar, Laxmi Pawar, Akanksha Puri, Rahul Gupta, Prof. Swati Vyas, "EchoTweet: A Bird Call Classification System", Published in *International Research Journal of Innovations in Engineering and Technology* - *IRJIET*, Volume 8, Issue 4, pp 152-156, April 2024. Article DOI https://doi.org/10.47001/IRJIET/2024.804020

****