

Dataset Description

The main aim of this dataset is to automate the detection of the presence of peafowl using their vocalizations.

This dataset contains two types of audio recordings:

- **Positive samples:** recordings with peafowl vocalizations.
- **Negative samples:** recordings without peafowl vocalizations, including similar-sounding species and ecoacoustic events.

The goal is to support research in **peafowl behavior monitoring**, **human-peafowl conflict mitigation**, and **bioacoustics detection systems**.

The vocalizations are from the **Indian Peafowl (*Pavo cristatus*)**, a species native to the Indian subcontinent and especially abundant in India and Sri Lanka. Peafowl belong to the genus *Pavo* in the family *Phasianidae*. "Peafowl" refers to both males (peacocks) and females (peahens).

The Indian Peafowl thrives in dry lowland forests, scrublands, and agricultural areas. Males are renowned for their iridescent blue and green plumage, while females display more subdued brownish feathers for camouflage.

This dataset was curated mainly to cater to mitigation strategies for the **Human-Peafowl Conflict** that exists in these regions. The absence of natural predators has contributed to a significant increase in the peafowl population, exacerbating challenges for farmers. Peafowls are sometimes considered agricultural pests due to their tendency to feed on and damage crops.

Their distinct vocalizations serve as a reliable indicator of their presence.

Folder Details

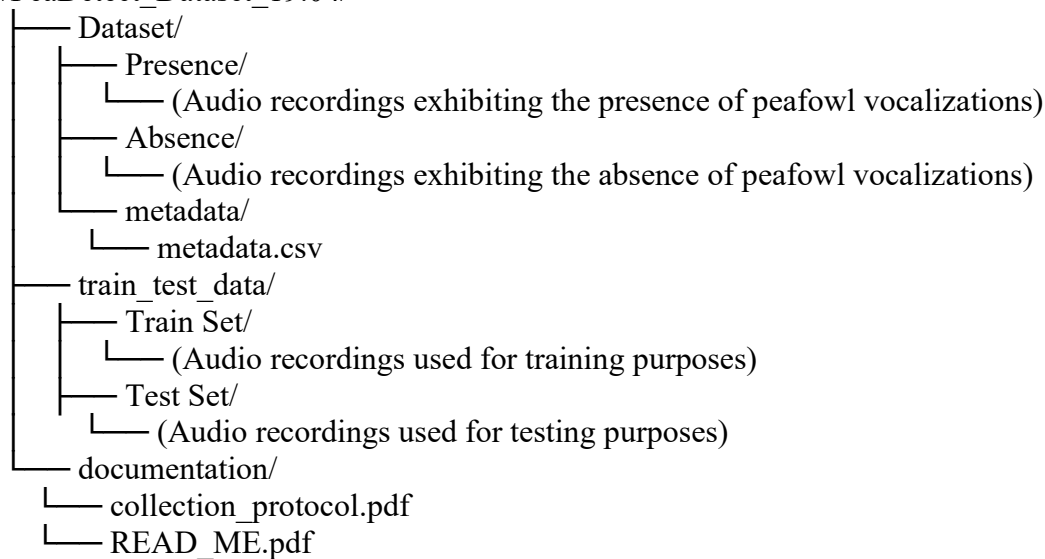
- /Dataset/
 - /Presence/
Contains audio recordings that exhibit the **presence** of peafowl vocalizations.
 - /Absence/
Contains audio recordings that exhibit the **absence** of peafowl vocalizations.
 - /metadata/
 - **metadata.csv** — Metadata file describing each audio recording with the following columns:
 - **Source File Name:** The original identifier of the recording from the source repository.
 - **Dataset File Name:** The standardized name assigned within the PeaDetect dataset.
 - **ID:** A binary indicator where 1 denotes the presence and 0 denotes the absence of peafowl vocalizations.
 - **Source:** The online repository or platform from which the audio recording was collected.
 - **Category:** A descriptive label (e.g., "vehicle horn," "parrot") for absence samples.

- /train_test_data/
 - /Train Set/
 - Audio recordings intended for model training.
 - /Test Set/
 - Audio recordings intended for model testing.

This division enables reproducible experiments and standard evaluation.
- /documentation/
 - **collection_protocol.pdf**
Describes the detailed methodology and protocols followed during dataset curation.

Folder Organization

/PeaDetect_Dataset_19.04/



Instructions for Use

1. **Audio Playback**
Load .wav files using any standard audio library (e.g., librosa in Python) for playback or analysis.
2. **Metadata Matching**
Match each audio file to its metadata using the File_Name field, with their annotation.
3. **Experimentation and Analysis**
The audio recordings are renamed with the suffixes _1 for presence and _0 for absence, in the train and test datasets.
4. **Citation**
Please cite the dataset as:
[Rukshani Puvanendran], "[PeaDetect]," IEEE DataPort, [2025]. DOI:
[<https://10.0.82.235/nqaq-w629>]