Frog and Bird Song Classification

Team Members: Joseph Melby, Gautam Prakriya, and Florian Stecker

Project Goals: Classify frog and bird species in audio recordings from Puerto Rico.

Societal Impact: Classifying bird, frog, and other animal species by their communication techniques is of interest to a wide variety of stakeholders, including conservation-focused organizations and governments, endangered and threatened species advocates, and even hobbyists throughout the general public. Our motivation for this project aligns with the mission of these stakeholders: the more we know about the communication dynamics of these species of rainforest dwellers, the better we can act to protect their homes. This project gave us a deeper appreciation of the important efforts of conservationists and scientists dedicated to understanding these frog and bird species.

Description of Dataset: The dataset came from a Kaggle competition called “Rainforest Connection Species Audio Detection” and consisted of 4727 minutes of sound recordings from various locations in the Puerto Rican rainforest, together with 1216 labels identifying subsections of 0.2 - 8s long as the song of one of 24 selected species. Of the 24 species, 11 are frogs and 13 are birds, and many of them are endemic to Puerto Rico.

Model 1: For each of the 1216 labels we extracted a 2 second sound clip from the recordings; if the song was longer than 2 seconds, we chose a random 2 second subsegment, if it was shorter than 2 seconds, we added a random amount of context before and after to pad it to 2 seconds. Then we converted the clips into 64x44 pixel log-mel-spectrograms, flattened them, and ran various classification algorithms on the resulting 2816-dimensional vectors. We found the best results using a logistic regression classifier. In particular, we were able to tune our model using various regularization techniques to tame its tendency to overfit. Given this, the model was able to identify the correct species in 68% of the test cases. We also tested SVM and KNN models to a lesser degree of accuracy.

Model 2: In order to improve our results, we explored existing models for these datasets, namely using the BirdNET model. We extracted 3 second sound clips in the same manner as above, and transformed these into 1024-dimensional feature vectors using the pretrained *BirdNET* neural network. The BirdNET model applies a convolutional neural network to mel-spectrograms in order to classify a recording into one of over 3300 bird species. It has been trained on over 30000 hours of sound data and is freely available online. To adapt the model to the Puerto Rican species in our dataset (which are not present in the original BirdNET training data), we cut of the final layer of the neural network to obtain 1024-dimensional feature embeddings, and then train a custom classifier on these embeddings. Again, a regularized logistic regression model gives the best results, identifying the correct species in 82.5% of test cases.

Takeaways: It is clear that this project and data merit far more time and exploration than was feasible in the time we had, but we are motivated to continue to learn what we can from the process.This project gave great insight into the challenges faced by conservation-minded parties to understand audible animal communication, especially the data collection challenges therein. With time, we would have loved to explore many different data preprocessing avenues, including noise removal and data augmentation techniques, more sophisticated image classification techniques, and applications of our methods to other animal sound datasets.