# Classification of Bat Echolocation Calls

Matt Haberkorn
Phoenix College Bioscience Dept.
matt.haberkorn@phoenixcolleg.edu
matt.a.haberkorn@gmail.com

# Problem

- Relatively little is known about bats
- Difficult to observe and handle
- Essential but poorly understood part of ecosystem
- Land manages have no idea what bats are present

# Who may be interested in solutions?

- Land managers
- Researchers
- Citizen scientists
- Wildlife specialists
- City parks
- Conservation agencies

# Solution and problem...

Solution: Ultrasonic recorders

- Passively record echolocation calls of bats
- No direct observation necessary
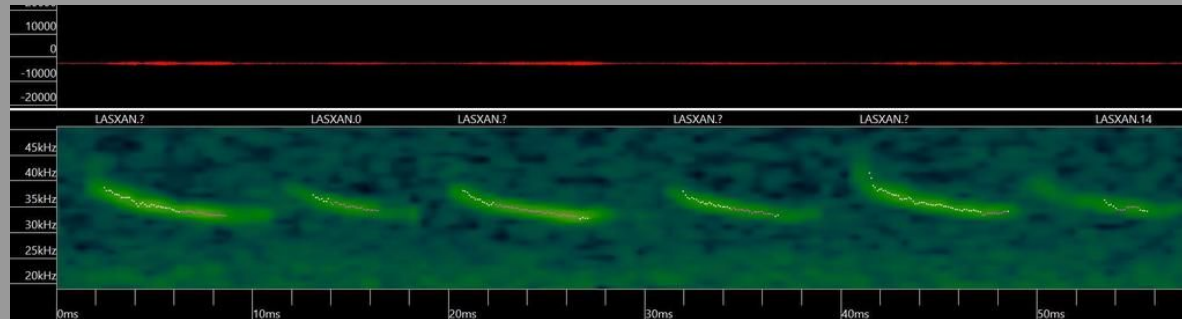- Collect large amounts of data

Problem…

- Large amount of data is difficult to process
- Ultrasonic sound requires software to analyze

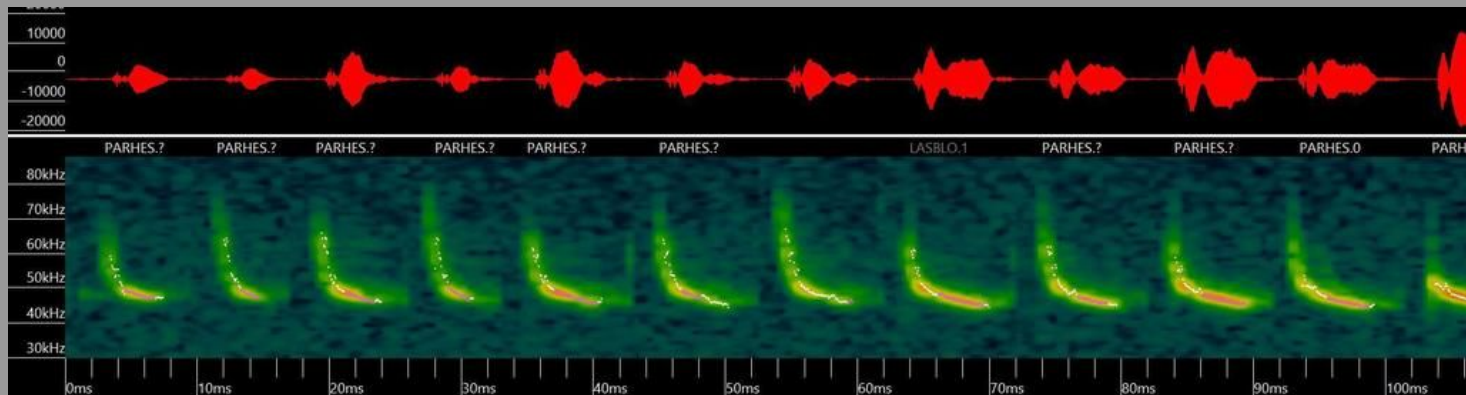# Solution to processing large amounts of data

Bat echolocation call classifying software

- Sonobat $1000 annually
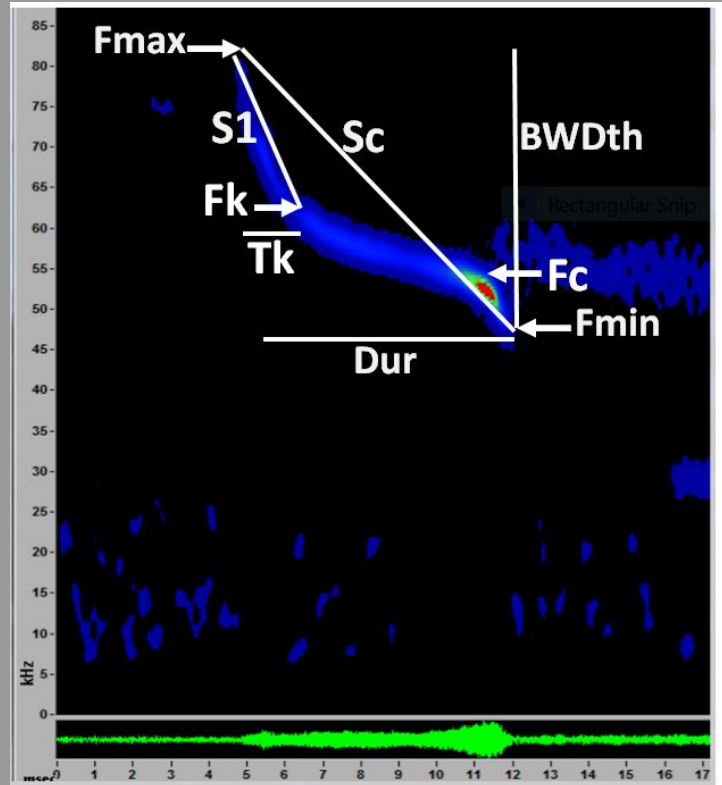- Kaleidoscope Pro $400 annually
- Limited accuracy of call classification

# Echolocation classifiers

- Every bat species produces a unique call
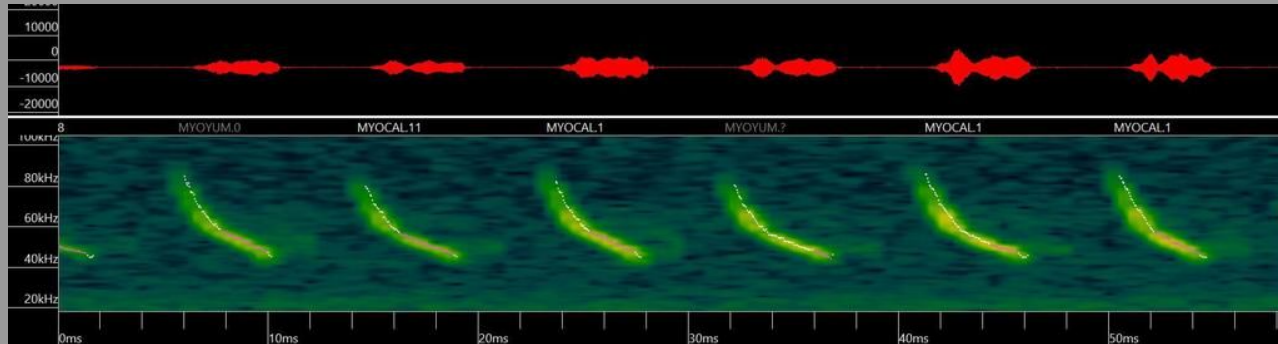- Features of call are derived from sonograms to ID species

# Bat call features

# Project goals

- Assess accuracy of Kaleidoscope Pro Software from Wildlife Acoustics
- Develop machine learning models for bat echolocation call classification

## Project data

- Collected by author between 2018 and 2020
- Sonoran Desert of Central Arizona
- All calls truthed manually
- 5849 calls
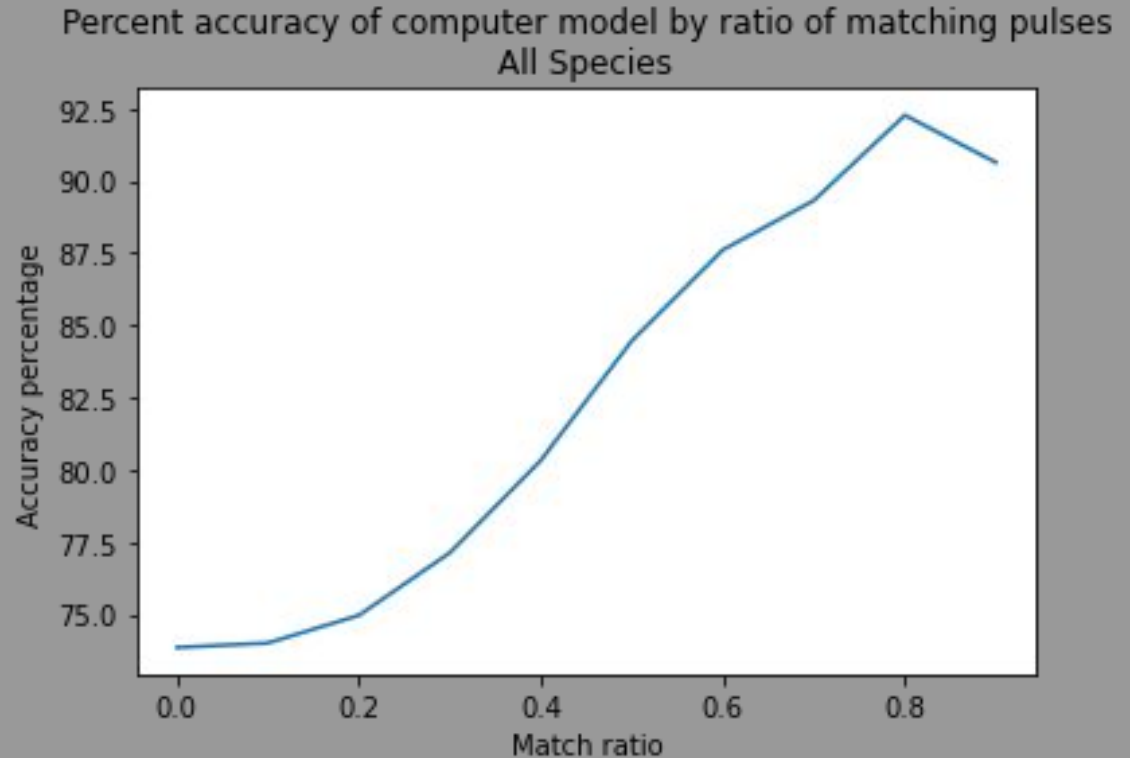- 17 bat species and no ID recordings
- 10 features

## Project data

- Highly unbalanced data set
- Three most common species 54% of data
  - Mexican Free-Tailed Bat
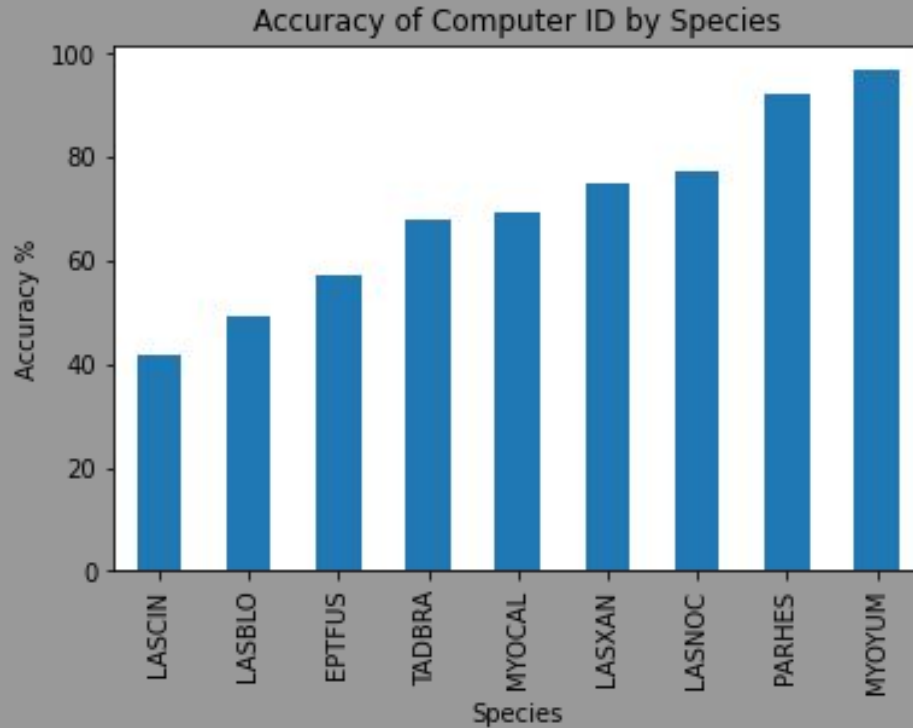  - Yuma Myotis
  - Canyon Bat

# Kaleidoscope Pro software accuracy

- Accuracy increases with match ratio
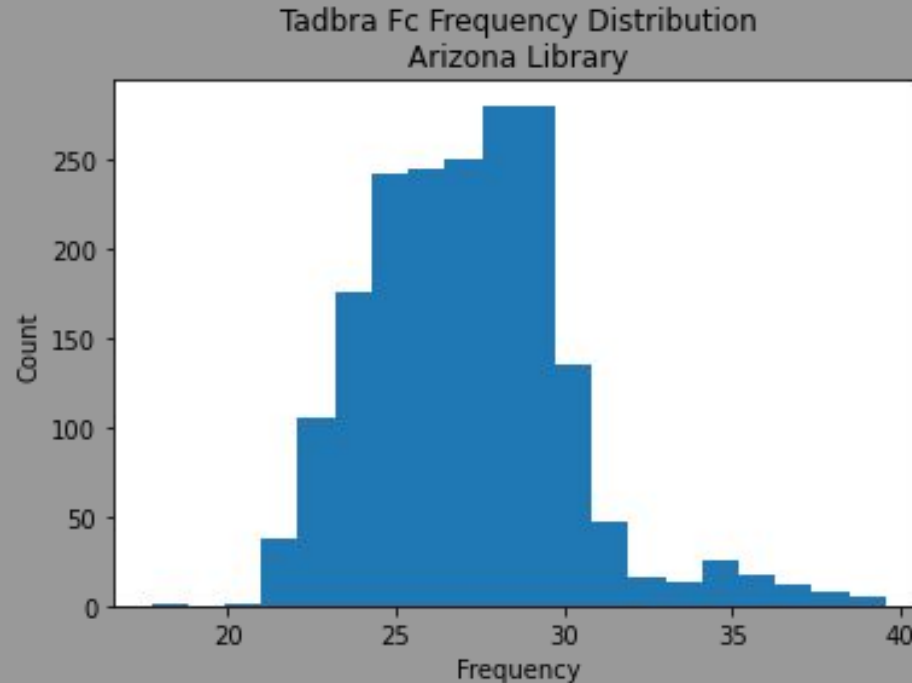- Significant room for improvement



Percent accuracy of computer model by ratio of matching pulses
All Species

# Kaleidoscope Pro software accuracy

## Accuracy varied by species



Accuracy of Computer ID by Species

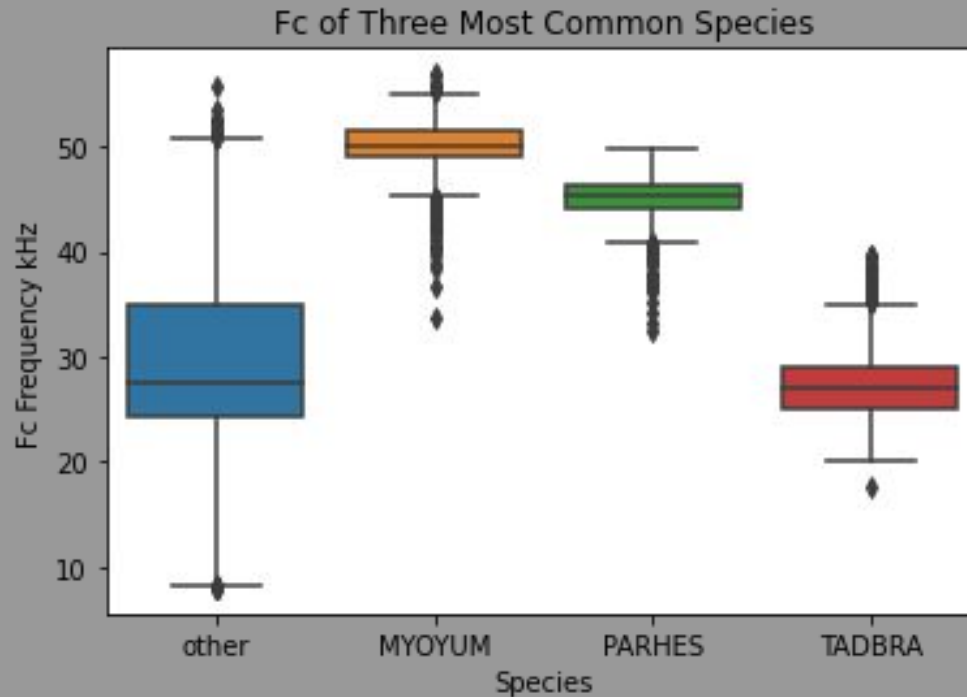# Exploratory data analysis

- ● Non-normal distributions

# Exploratory data analysis

- Features varied between species



Fc of Three Most Common Species

# Exploratory data analysis

- Random Forest feature importance
- High multicollinearity between Fc, Fmin, Fmax, and Fk
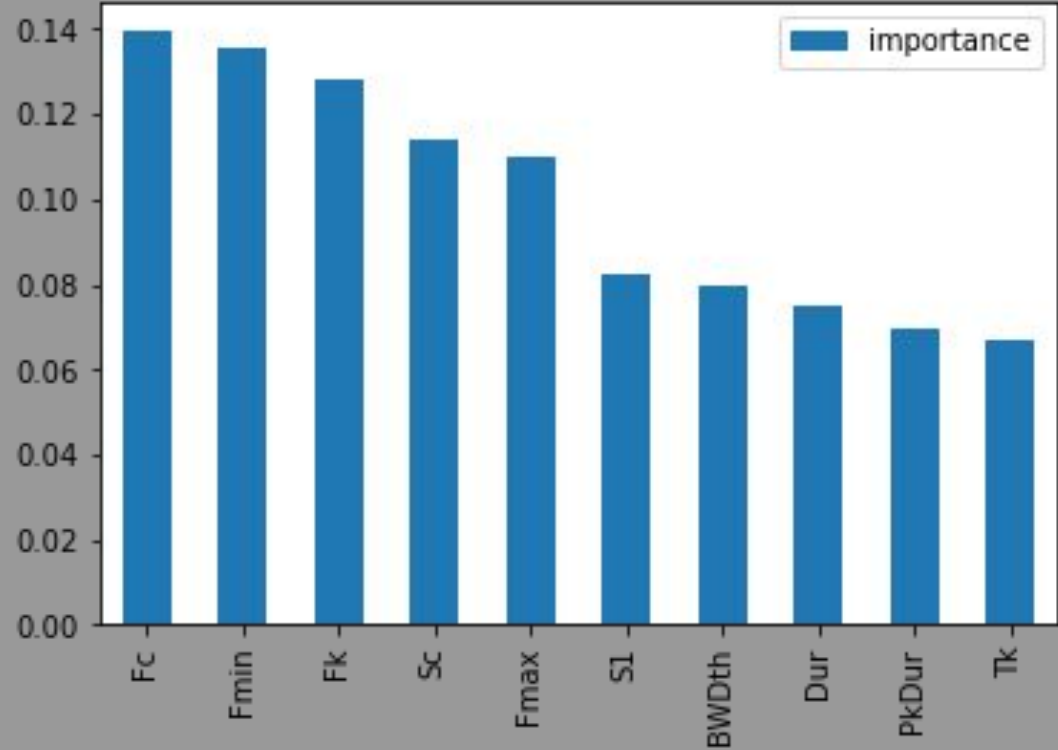
# Data and modeling

## Data used
- Raw and robust scaled
- High multicollinear features retained and removed
- 3 species vs all other species and all species data

## Models developed
- Random Forest
- KNN
- XGBoost
- Logistic Regression

# Best performing models

| Model | All Species Score | Three Species Score | Data Scaled | Multicollinear Data | Best Hyperparameters |
|---|---|---|---|---|---|
| Random Forest | 0.6629 | 0.7613 | No | Not removed | Criterion: gini<br>Max depth:11<br>Max features: auto<br>N estimators: 220 |
| Knn | 0.6426 | 0.7421 | Yes | Removed | Distance: Manhattan<br>Weight: distance<br>N neighbors: 19 (all species)<br>            16 (3 species) |
| XGBoost | 0.6507 | 0.7579 | No | Not removed | Learning rate: 0.05<br>Max features: 3 (all species)<br>            2 (3 species)<br>N estimators: 20 |
| Multinomial Logistic Regression | 0.5962 | 0.6695 | Yes | Removed | C: 0.001<br>Penalty: Ridge reg. (all species)<br>        None (3 species) |

# Random Forest model vs. Kaleidoscope Pro

- RF model consistently under performed Kaleidoscope Pro overall



Model F1 By Match Ratio
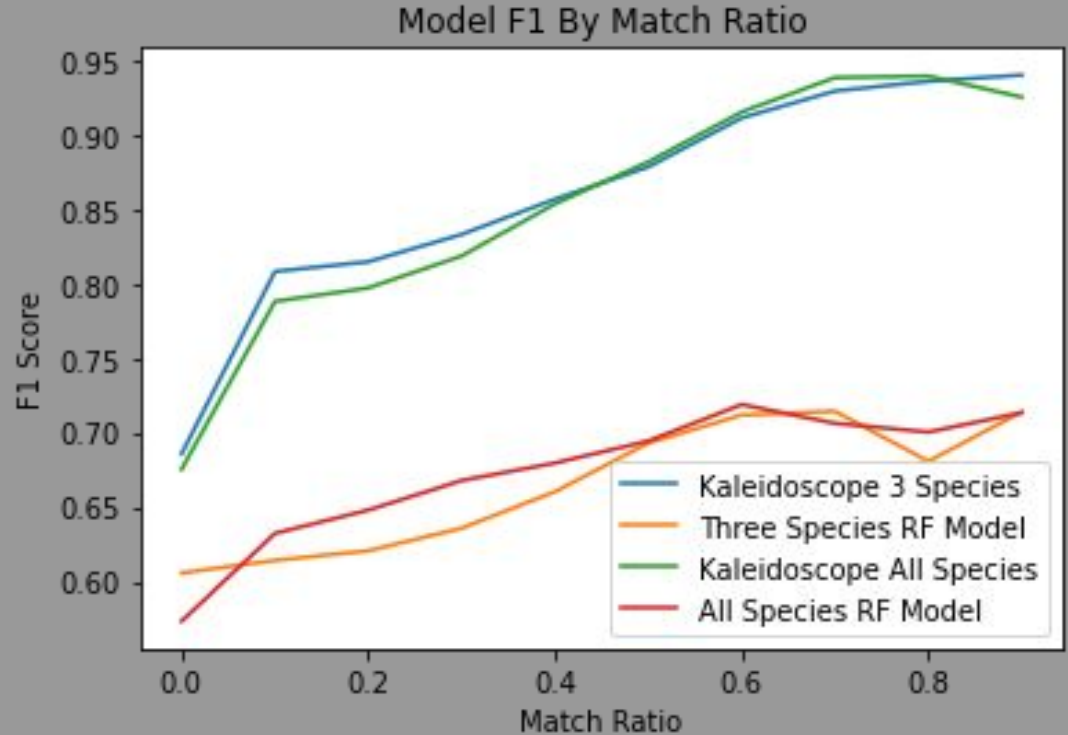
# One vs. all Random Forest model

| Single Species Random Forest Model | ROC-AUC Score | Accuracy Score | Random Forest Hyperparameters |
|---|---|---|---|
| TADBRA | 0.854 | 0.797 | Criterion: gini<br>Max depth:11<br>Max features: auto<br>N estimators: 400 |
| MYOYUM | 0.992 | 0.965 | Criterion: gini<br>Max depth:10<br>Max features: auto<br>N estimators: 100 |
| PARHES | 0.998 | 0.985 | Criterion: gini<br>Max depth:6<br>Max features: auto<br>N estimators: 600 |

# One vs. all species RF F1 thresholding



Random Forest TADBRA F1 Thresholding
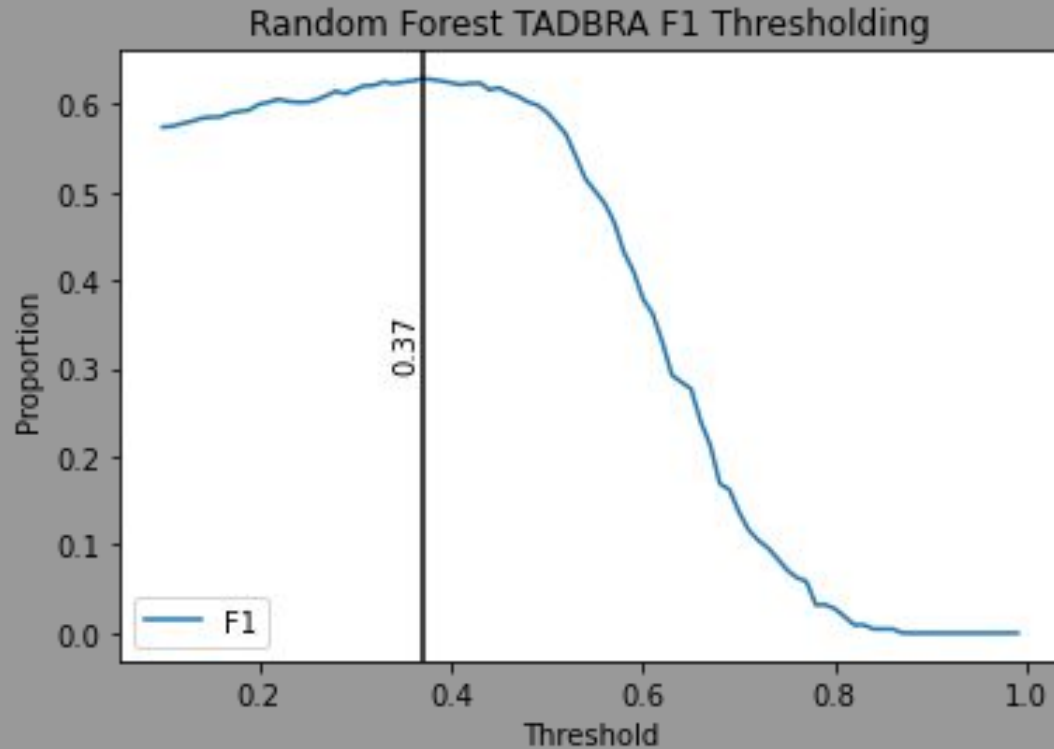
# One vs. all species RF thresholding

| Random Forest Model | Default F1 Score | Optimized F1 Score | Optimized Threshold |
|---|---|---|---|
| TADBRA | 0.591 | 0.629 | 0.37 |
| MYOYUM | 0.886 | 0.891 | 0.37 |
| PARHES | 0.918 | 0.922 | 0.46 |

# Canyon Bat (PARHES) vs. all species



F1 By Match Ratio For PARHES
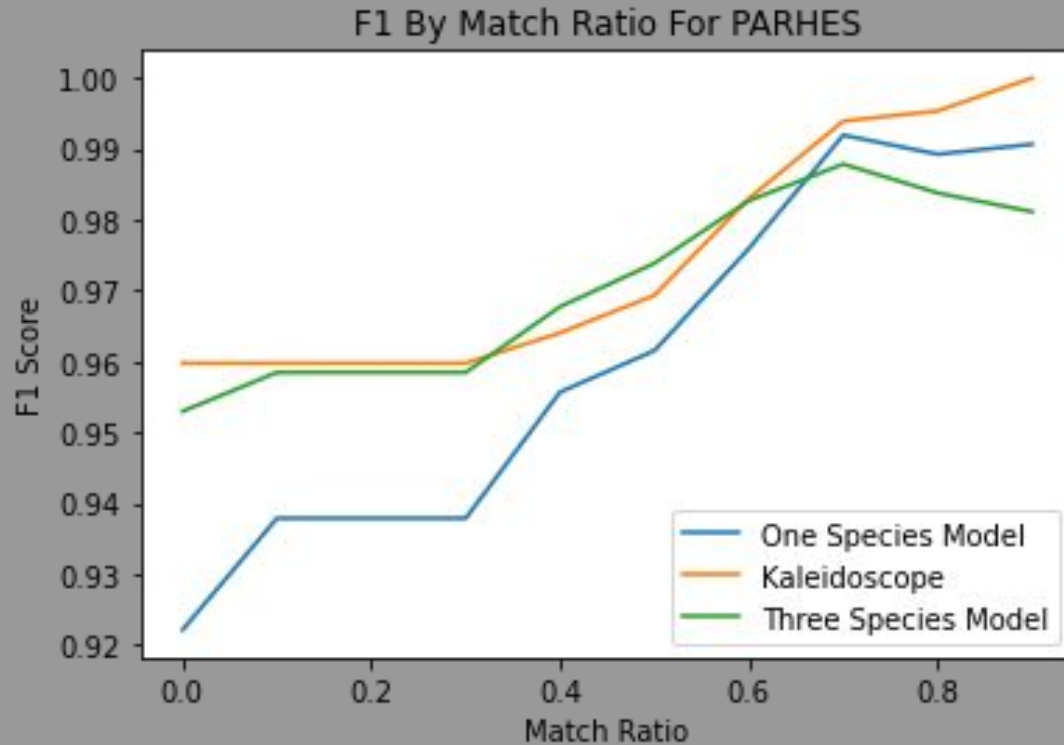
# Yuma Myotis (MYOYUM) vs. all species



F1 By Match Ratio For MYOYUM
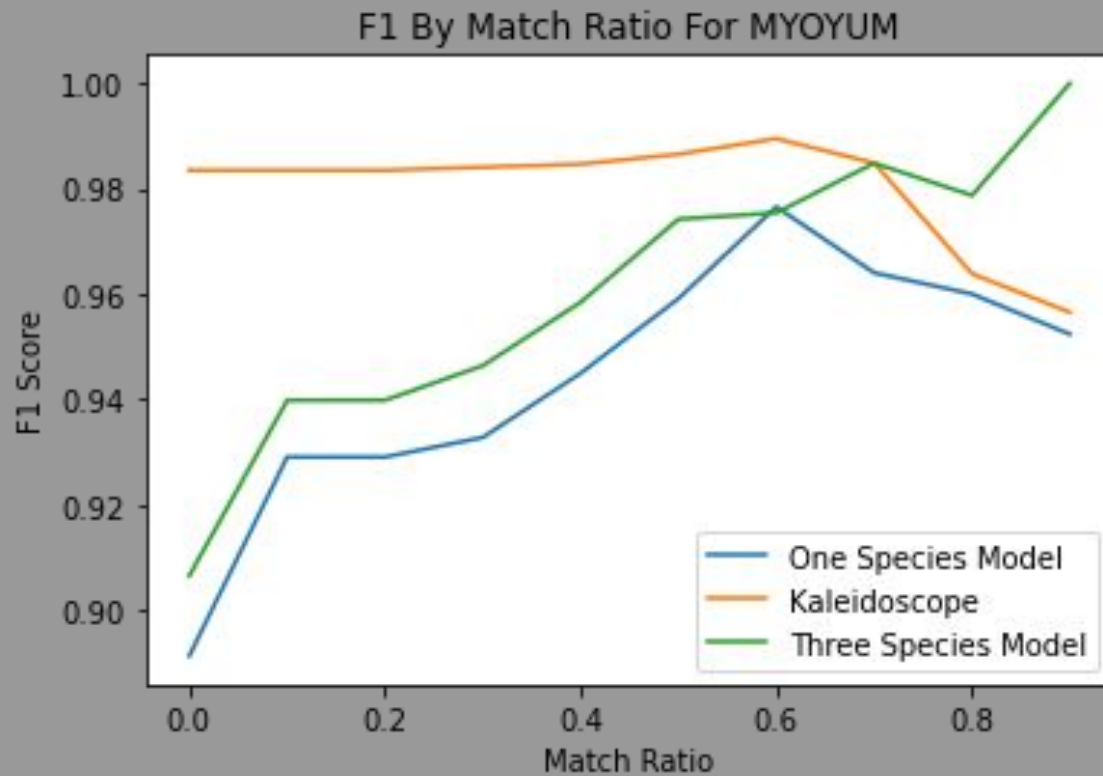
# Mexican Free-Tailed Bat (TADBRA) vs. all species
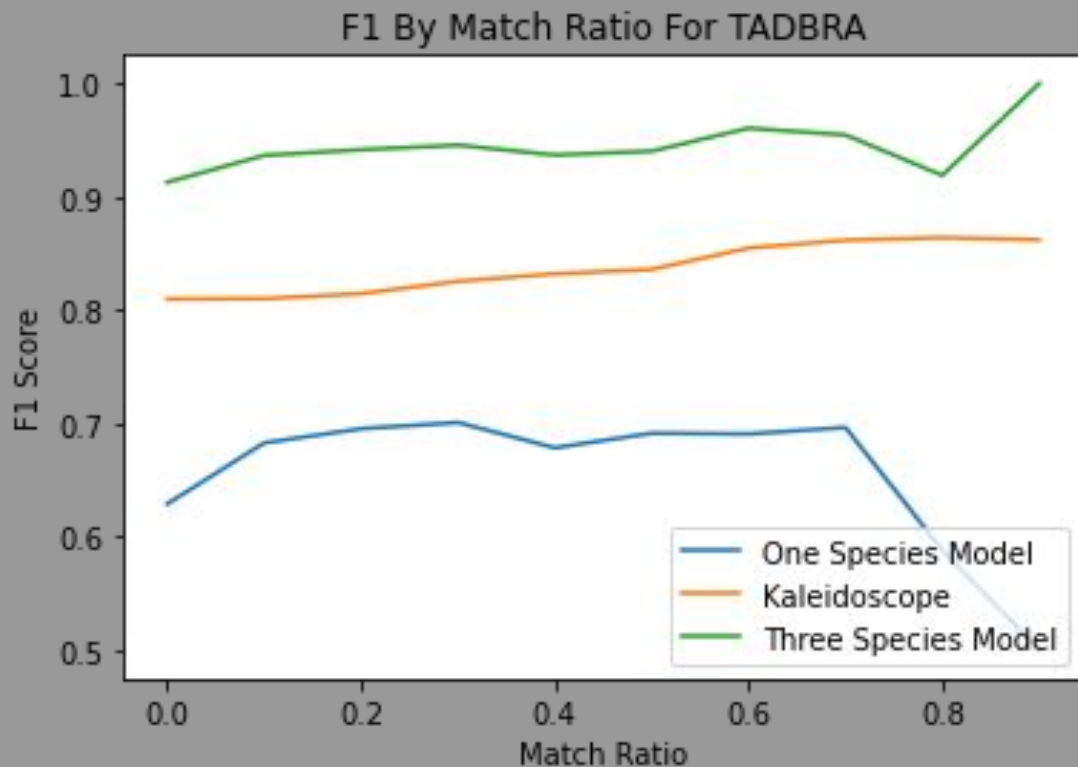
# Conclusion

- Kaleidoscope Pro overall outperformed models developed in this project
- 3 species Random Forest model performed best of models developed
- TADBRA ID from 3 species model significantly outperformed Kaleidoscope

# Mexican Free-Tailed Bat (TADBRA) model

- Most common bat species in SW United States
- TADBRA model can be used in conjunction with other models
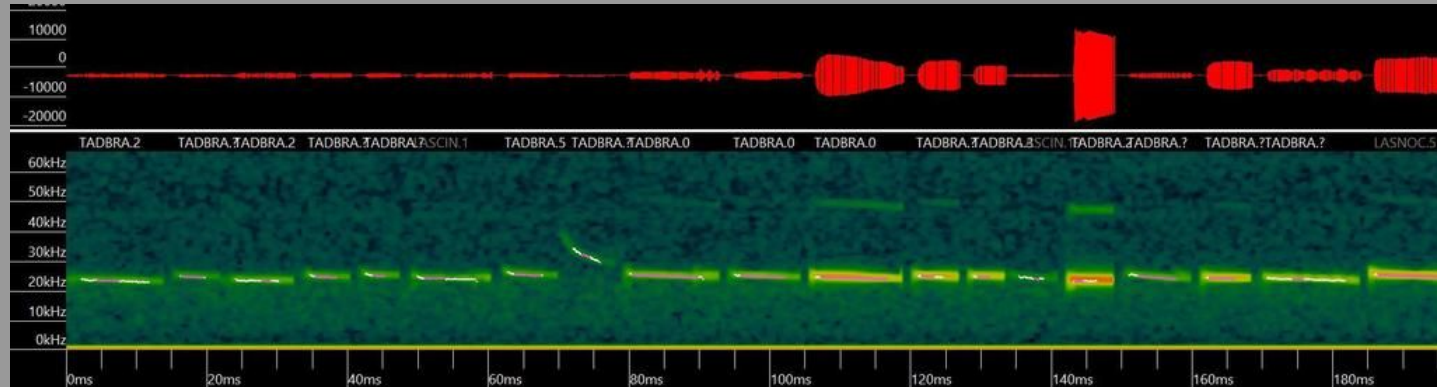
# Future work

- Develop a more balanced dataset
- Remove NoID's from modeling data
- Threshold multispecies models
- Development of different features