

AMMonitor: A case study in automated detection for invasive Cuban Treefrog with a two-stage machine learning approach

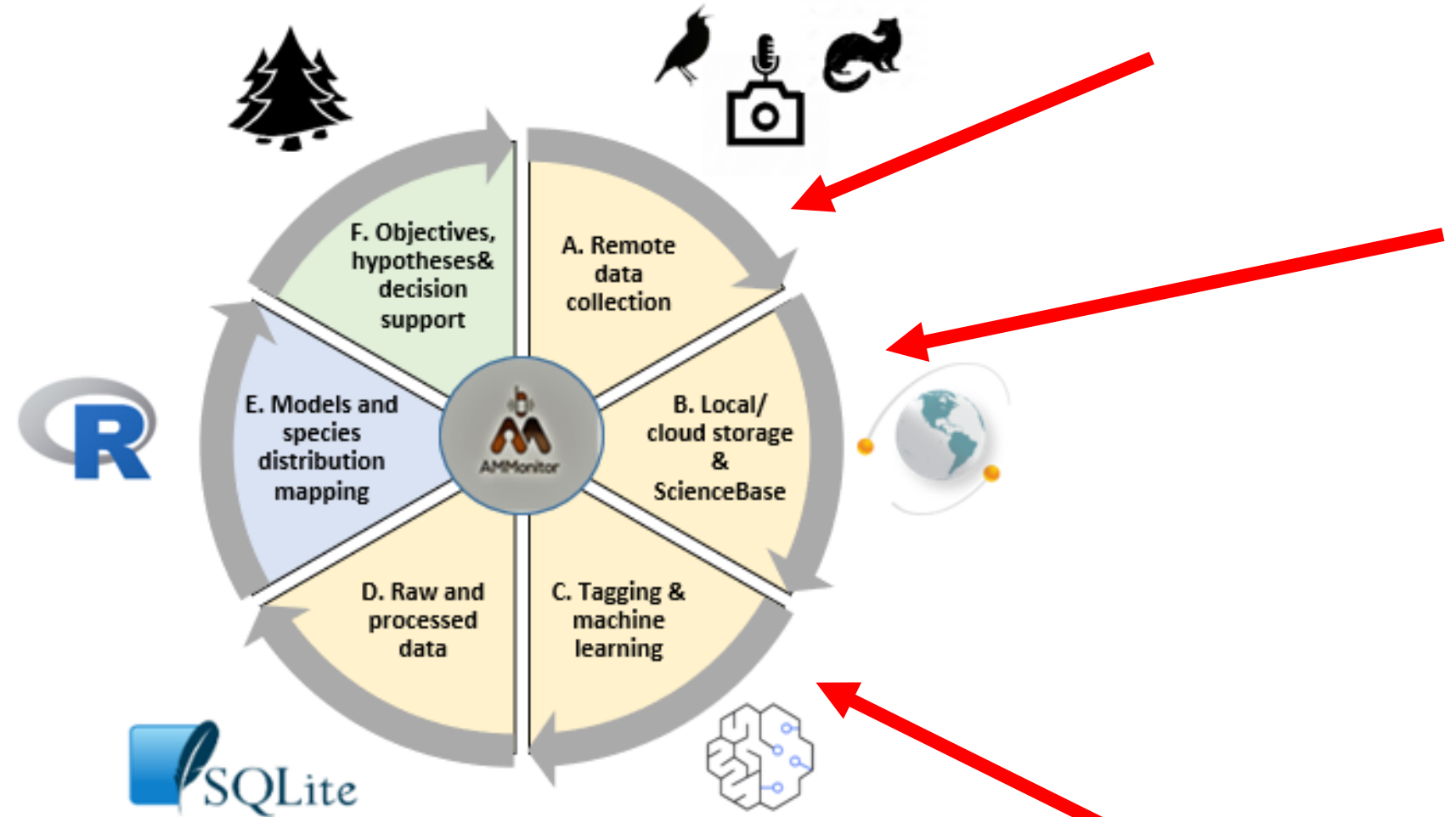
Kate Huber, PhD Student

Affiliation: Vermont Cooperative Fish and Wildlife Research Unit

Co-Authors: J. Hardin Waddle, Brad M. Glorioso, Therese M. Donovan



AMMonitor for Wildlife Monitoring



Clarfeld, L., C. Tang, K. Huber, C. Balantic, K. Dillon, and T. Donovan.
AMMonitor: Remote monitoring of biodiversity in an adaptive framework.
Version 2.2.0: U.S. Geological Survey software release. Reston, VA.
<https://doi.org/10.5066/P13MRDRV>

Collaborators

- **Dr. Hardin Waddle**
- USGS Wetland and Aquatic Research Center (Gainesville, FL)
- Part of USGS ARMI Program (Amphibian Research and Monitoring Initiative)



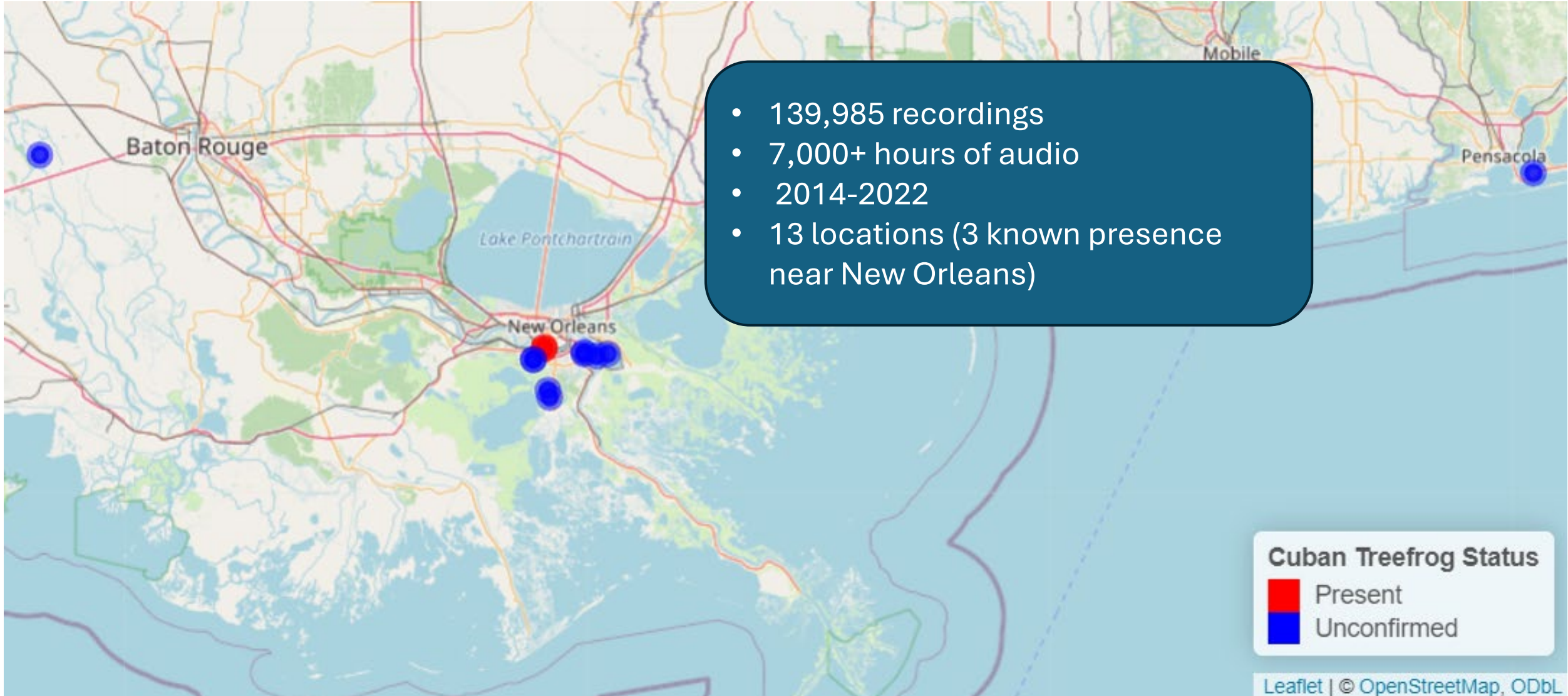
Cuban Treefrog

- Problematic species invasive to Florida, Georgia, and Louisiana
- Predatory and competitive to native treefrogs
- Early detection is crucial to prevent new establishments
- Emits a distinctive call



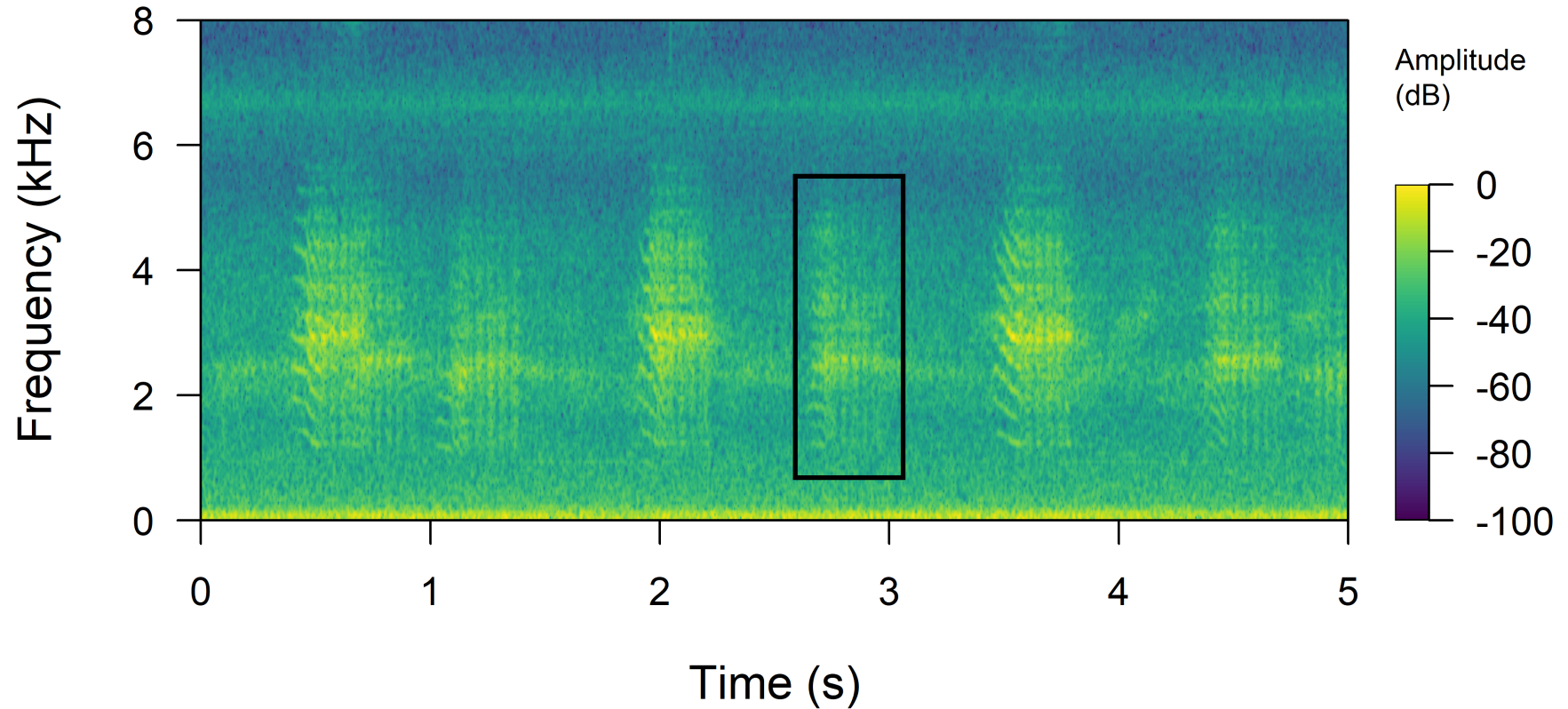
Myers, J., & Myers, A. J. (2020, June 3). *Herping the Cuban treefrog*. Reptiles Magazine.

ARMI Methods: Coastal Florida and Louisiana



Automated Detection

- Ideal for large-scale monitoring programs
 - Rapid media processing
 - Increasingly accurate species labels
- Cuban treefrog call is consistent across:
 - Individuals
 - Locations/
Populations
 - Regions

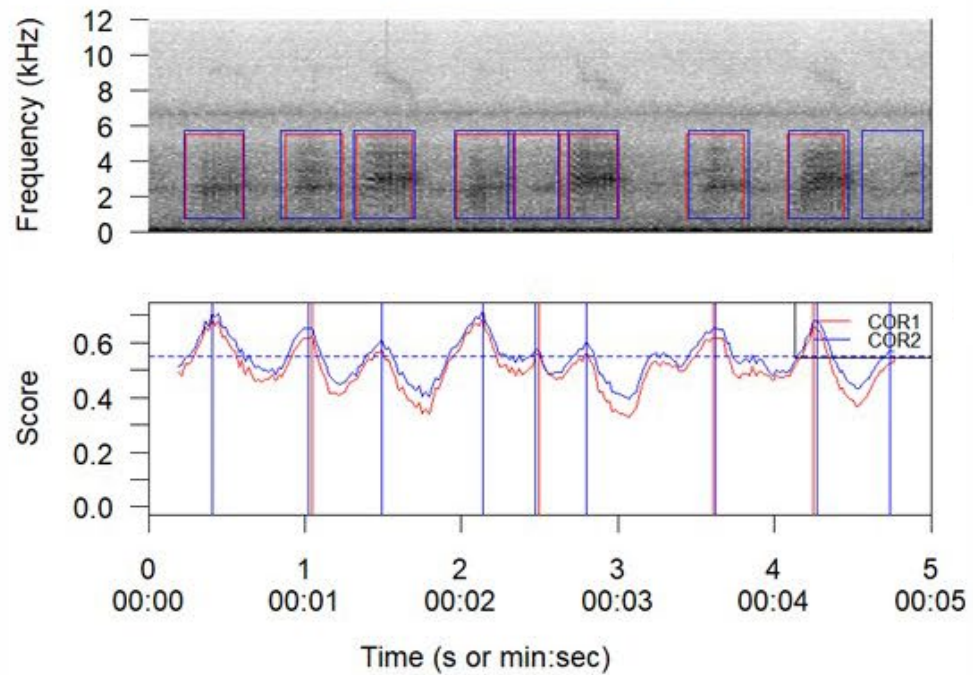


Two-stage approach (within AMMonitor Framework)

- Combines:
 1. Cross-correlation template matching
 2. Supervised machine learning classifiers
- Aim to achieve sensitive species detection with low incidence of false positives

STAGE 1- Accumulate Detections with Templates

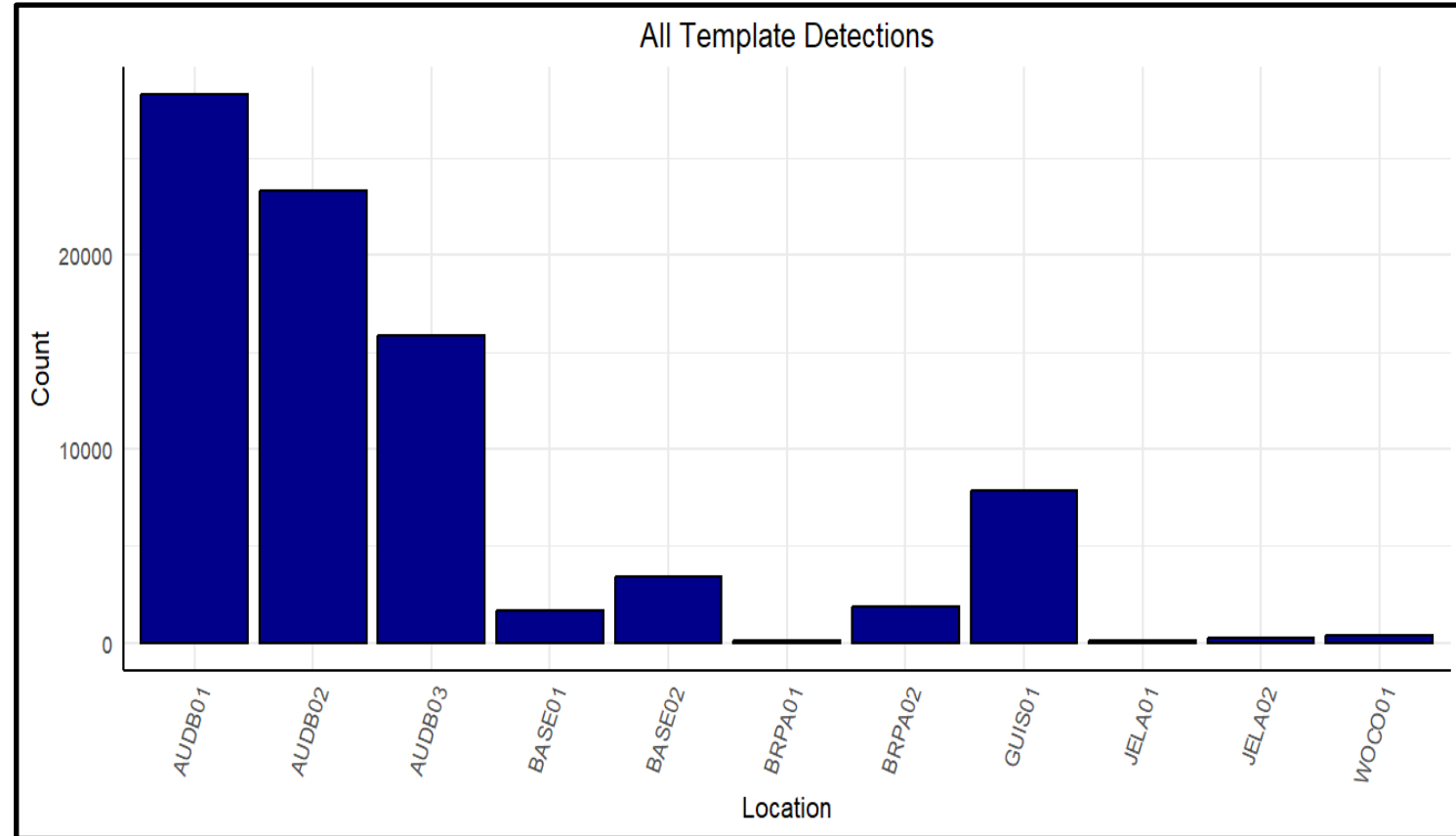
Template Detections



Each value
is a feature

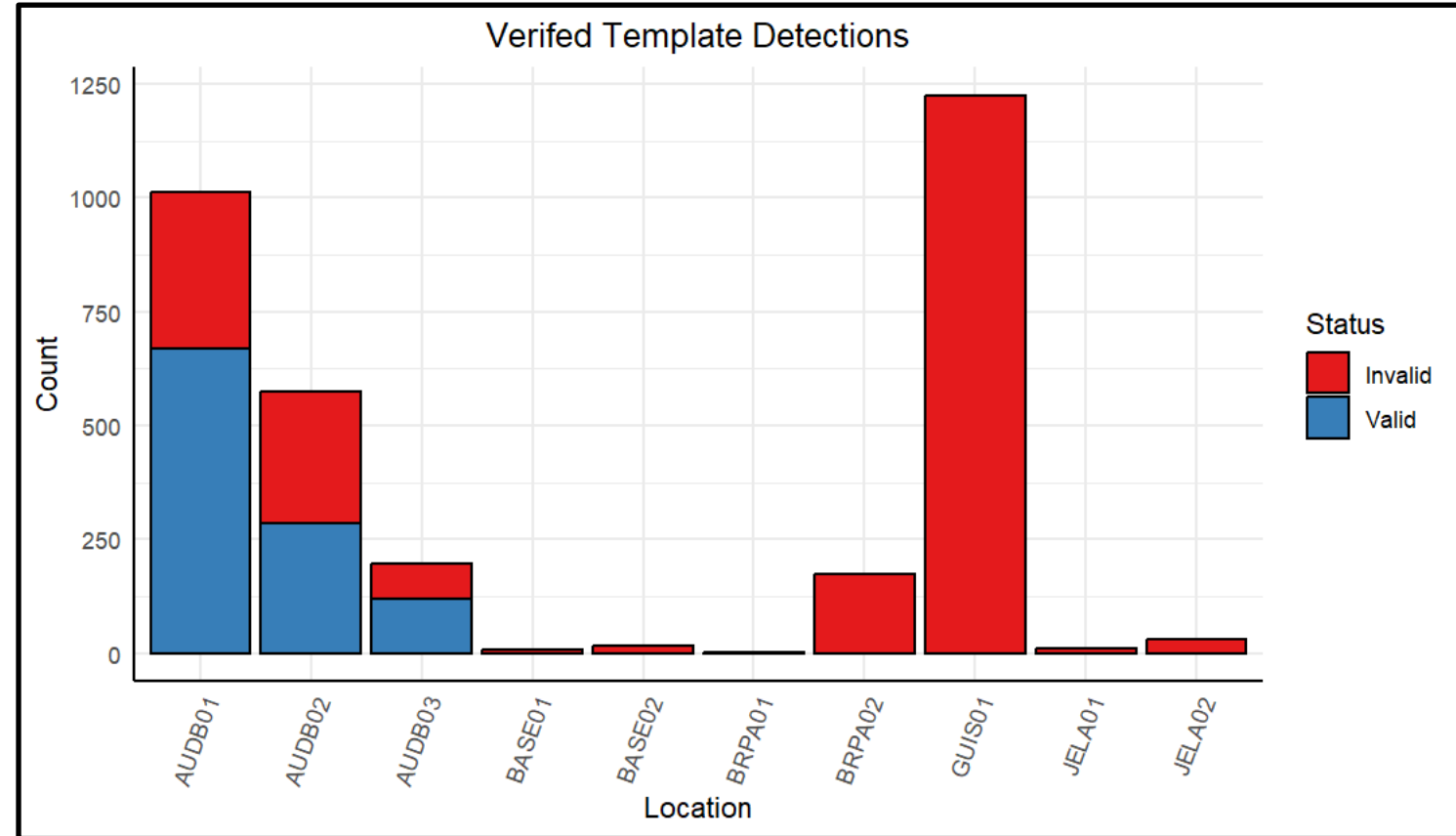
Stage 1: Screen data with Template

- Cross-correlation template with accompanying score cutoff
 - Created using *AMMonitor* Shiny App
- Accumulate initial detections across all audio files
 - *AMMonitor* function `scoresDetect()`
 - **83,191 initial template detections**



Stage 2A: Train Supervised Learning Classifiers

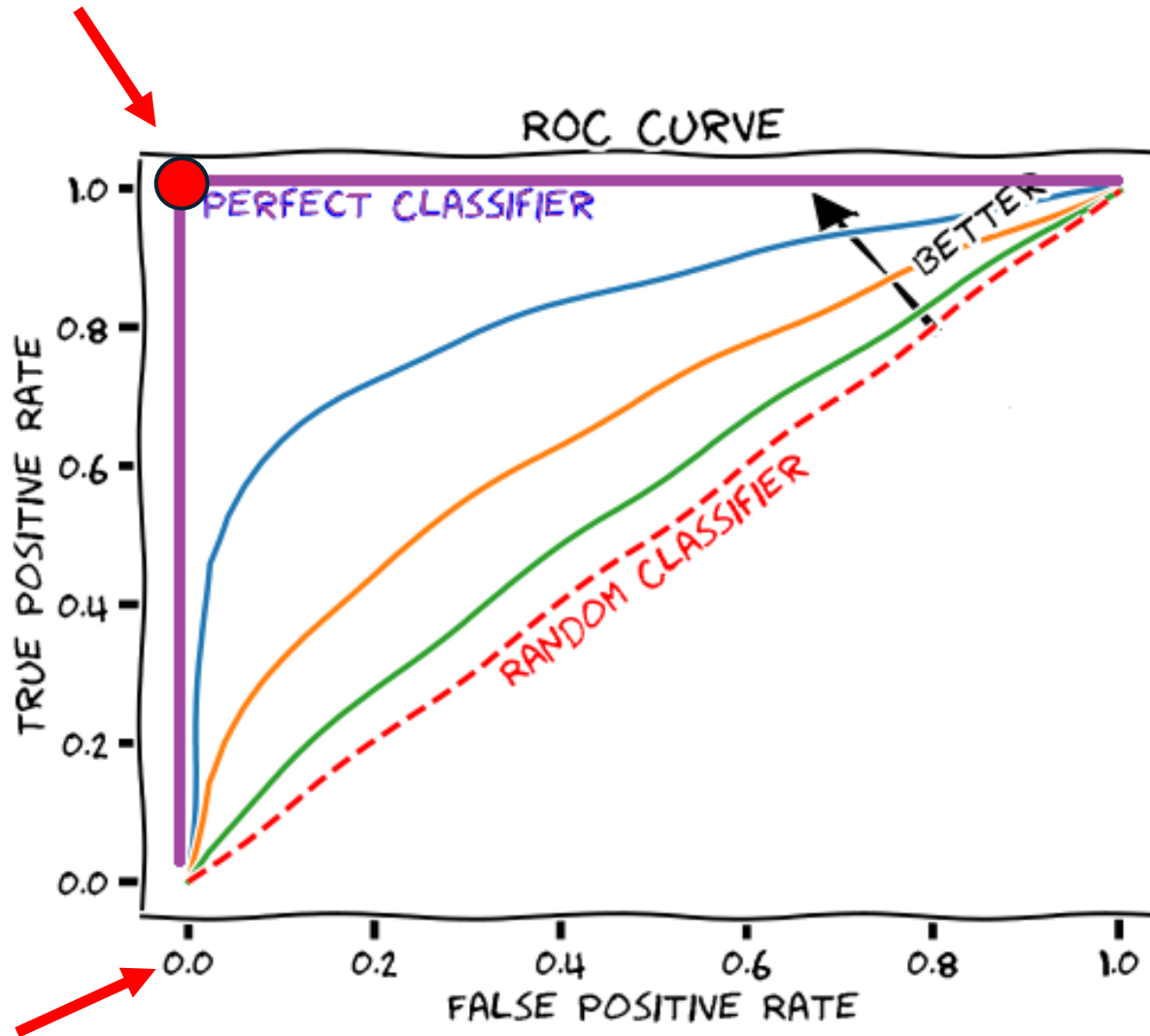
- **Training/Testing Data:**
 - 3200+ verified template detections (“valid” or “invalid” labels)
 - 80/20 training/testing split
- **3 Model Types:**
 - Random Forest (*rf*)
 - Support Vector Machines (Polynomial Kernel) (*svmPoly*)
 - Regularized Logistic Regression (*glmnet*)
- Hyperparameter tuning via tuning grid search



Testing Performance

Receiver Operating Characteristic (ROC) curve:

One method for evaluating model's ability to discern between classes



Calculated from **Confusion Matrix** and each threshold on the axes:

		Reference Classes	
		Valid	Invalid
Predicted Classes	Valid	True Negative	False Negative
	Invalid	False Positive	True Positive

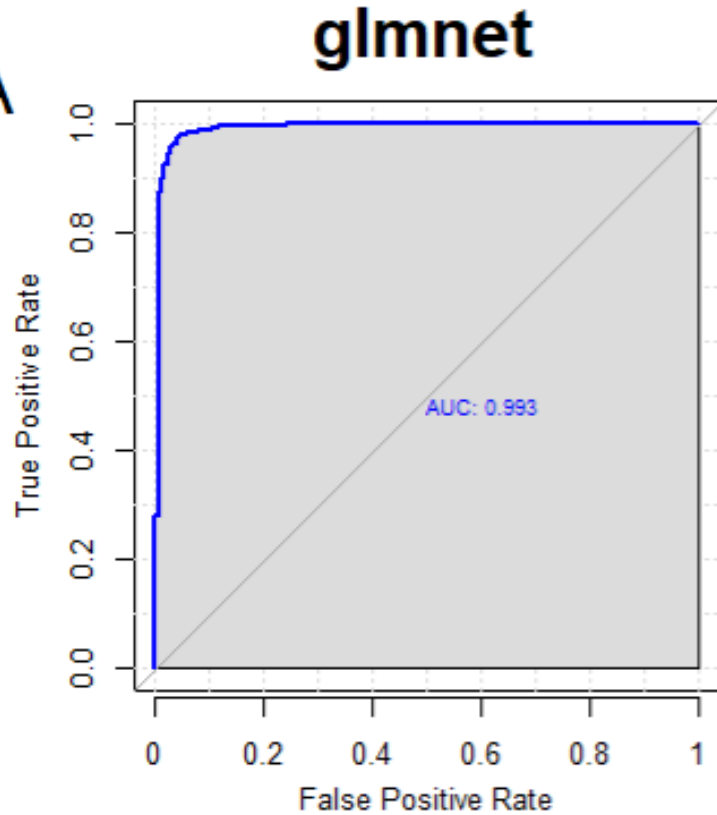
A classifier that **perfectly** identifies all invalid detections will have high TP rate of 1 and will not miss any valid detections (FP rate = 0)

Draelos, A. R., Draelos, P. by R., Draelos, R. (2020, February 2). *Measuring performance: AUC (AUROC)*. Glass Box. <https://glassboxmedicine.com/2019/02/23/measuring-performance-auc-auroc/>

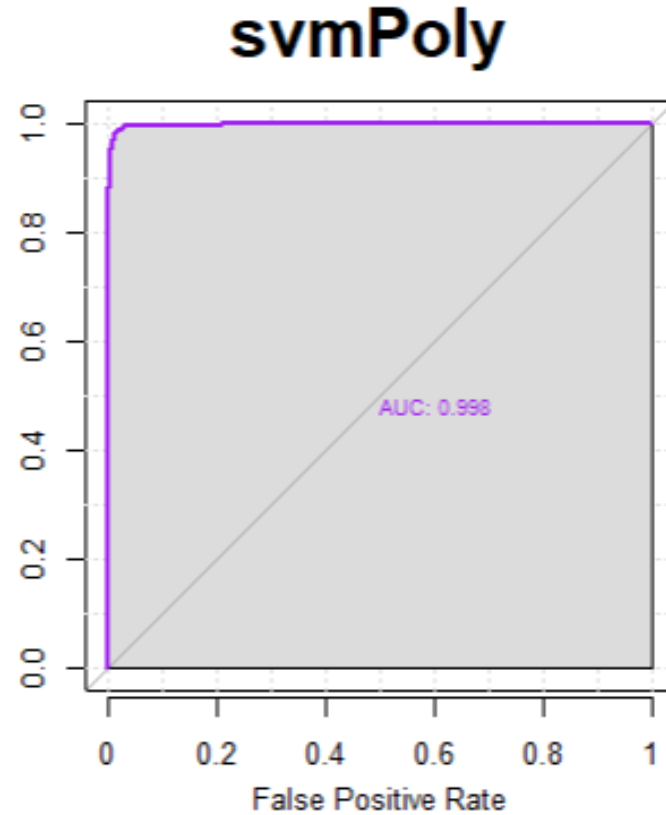
Testing Performance

Best model by all performance metrics

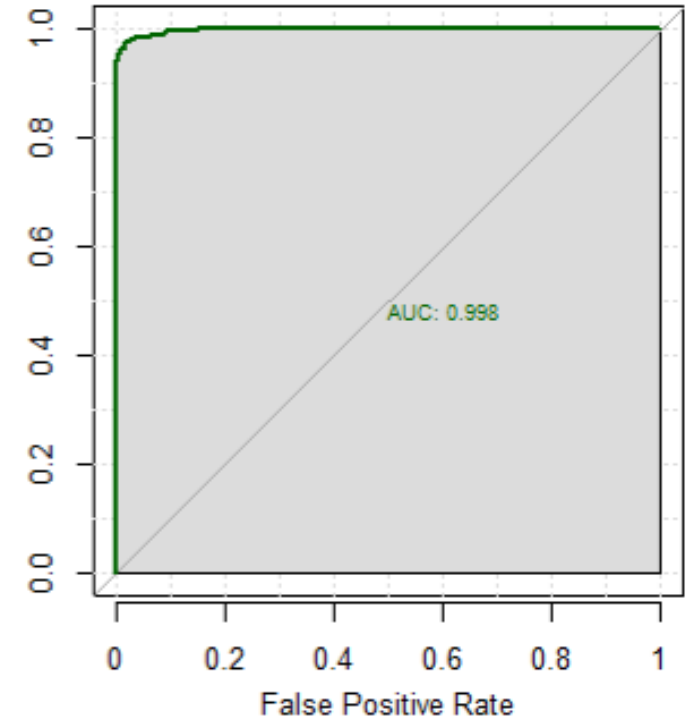
A



B



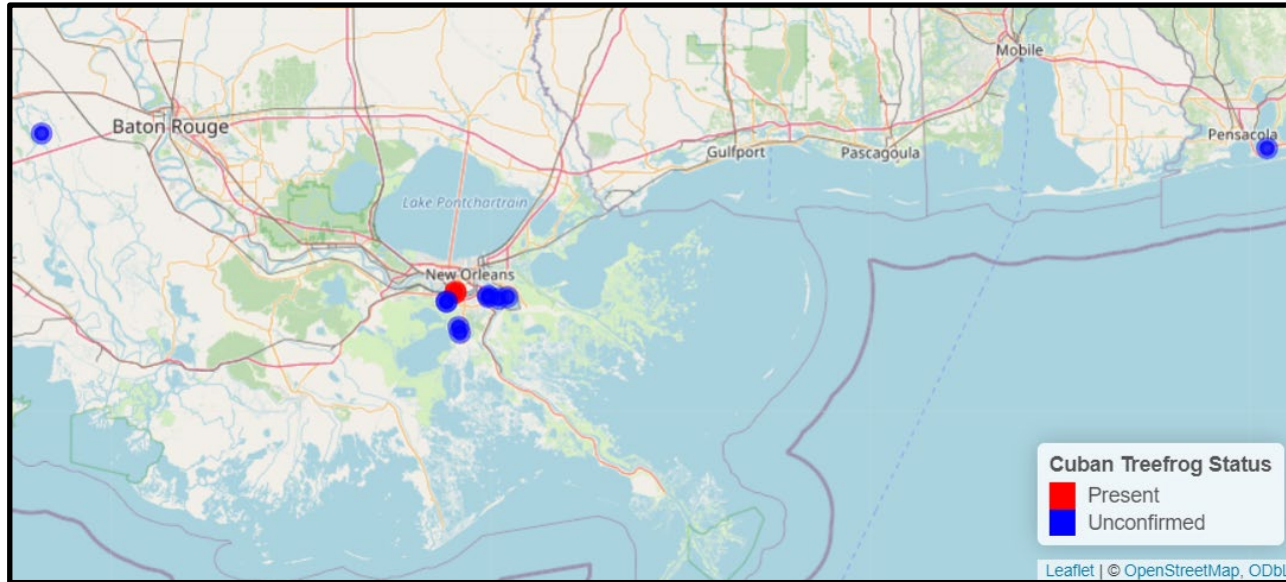
rf



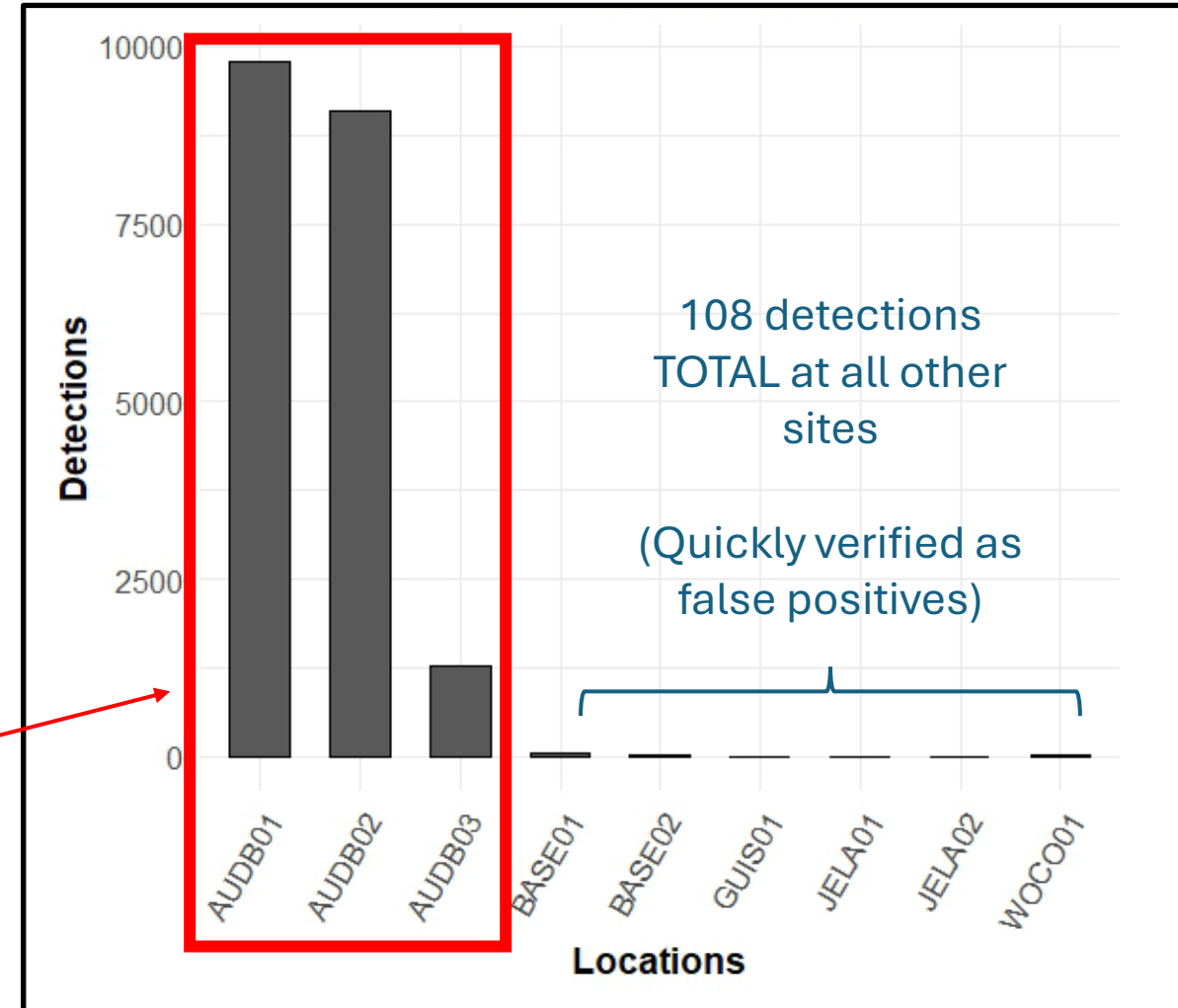
In other words, these machine learning classifiers models are **GREAT** at separating valid and invalid Cuban treefrog detections!

Stage 2B: Run Best Classifier on Remaining Data

**20,271 Final Predicted
Detections**



Only Locations with Confirmed Cuban treefrog presence



Conclusions

- 2-stage approach with *AMMonitor* is an effective method for automated detection for Cuban treefrog
 - Fast method for developing accurate machine learning models
 - Requires less training data than other algorithms
- Process can be applied to any sound-emitting target species, with caveats

Thank you!