Automatic Capture and Classification of Frog Calls

ELIZA G. FORAN*, Department of Biology Indiana University
TENECIOUS A. UNDERWOOD, Department of Computer Science Kentucky State University
WINONA SNAPP-CHILDS, Pervasive Technology Institute Indiana University
SHERI A. SANDERS†, National Center for Genome Analysis Support Indiana University

Global frog populations are threatened by an increasing number of environmental threats such as habitat loss, disease, and pollution. Traditionally, in-person acoustic surveys of frogs have measured population loss and conservation outcomes among these visually cryptic species. However, these methods rely heavily on trained individuals and time-consuming field work. We propose an end-to-end workflow for the automatic recording, presence-absence identification, and web page visualization of frog calls by their species. The workflow encompasses recording of frog calls via custom Raspberry Pis, data-pushing to Jetstream cloud computer, and species classification by three different machine learning models: Random Forest, Convolutional Neural Network, and Recursive Neural Network.

CCS Concepts: • **Hardware** → Communication hardware, interfaces and storage; • **Computing methodologies** → **Machine learning approaches**; • **Information systems** → *Database design and models.*

Additional Key Words and Phrases: bioacoustics, frog, survey, machine learning, Random Forest, Neural Network

ACM Reference Format:
Eliza G. Foran, Tenecious A. Underwood, Winona Snapp-Childs, and Sheri A. Sanders. 2018. Automatic Capture and Classification of Frog Calls. In . ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

In the past few decades, habitat loss, disease, and pollution threaten frog populations globally. Frogs are sensitive to these environmental changes, where they serve as bioindicators of ecosystem health [11]. Frog call surveys inform conservation efforts and outcomes, but these sound-based surveys are limited by labor and manual identification [14].

We have created an end-to-end acoustic workflow for the automatic capture, analysis, and presence-absence identification of nine frog species. Calls are captured from the field with custom Raspberry Pi recorders [9], organized into databases, and classified by species. Metadata is available on a web portal as raw sound files, spectrograms, and species predictions. Species predictions may come from three different machine learning models (Random Forest and two Neural Networks). The Random Forest was faster in predicting, more accurate, and could predict more species than the Neural Networks.

*Foran and Underwood contributed equally to this research.
†Corresponding Author: Sheri A. Sanders, PhD, E-mail: ssanders@iu.edu

SC2020, November 15–20, 2020, Atlanta, GA

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

2 METHODS

2.1 Data Management

Raspberry Pis capture frog calls at scheduled times with USB microphones and flash drives. Data is appended to a database which merges into a master database on a Drupal [1] web portal hosted on Jetstream Cloud [3, 12]. While recordings can be used, 5-minute Waveform Audio Files (WAV) files from the Great Lakes Inventory and Monitoring Network served as the training and testing data for all models.

The autodetec() function in the warbleR [4] package was optimized to isolate potential calls, or signals, based on predetermined parameters, primarily duration and frequency. These parameters exclude most non-frog calls, such as birds. By design, it captures overlapping signals, especially if signal length or mean frequencies are very different. A table of 26 default signal features was generated with the specan() function. Of the 26 features provided by warbleR pre-processing, only 14 were subset, based on results from the DecisionTreeClassifier's feature_importance_ attribute plots [5]. Month of calling was also added as a feature. Features were removed in order of least importance pending stable ten-fold cross validation accuracy [10]. This output was fed into the Random Forest model, while signal-cropped WAV files to were passed to the Neural Network models.

Table 1. Summary of training data

Code	Species	Number of Calls
ANAM	American toad	88
LIPI	Northern leopard frog	37
LICL	Green frog	455
HYVE	Gray treefrogs	9830
PSMA	Chorus frog	178
PSCR	Spring peeper	10,000
LISY	Wood frog	363
ACBL*	Blanchard's cricket frog	716

*Neural networks do not include ACBL (Blanchard's cricket frog)

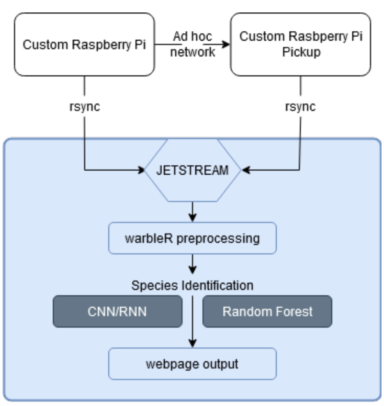

Fig. 1. Workflow from custom Raspberry Pi's to Jetstream.

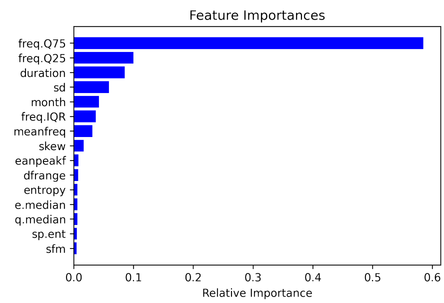

Fig. 2. Attribute plot of warbleR specan() features by gini importance.

2.2 Random Forest Model

The Random Forest model comprises 100 decision trees, truncated to a maximum depth of 10. Tree truncation and sklearn's OneVsRest Classification method were used to quantify probabilities of call identification where 80+% positive probability are kept, while sub-80% are classified as "unidentified". The 80% positive probability threshold is consistent with the accuracy required to survey frogs for FrogWatch USA community science organization [6].

We implemented sklearn's class_weight hyperparameter where weights are recalculated as inversely proportional to class frequencies to address training sample imbalance (Table 1) [7]. The Random Forest model's hyperparameters (Table 2) were optimized to achieve the most accurate class probabilities without sacrificing accuracy in 10-fold cross validation [10]. Confusion matrices were used to identify which species were being misidentified [13], and these species predictions were further investigated with class probabilities.

Table 2. Random Forest Hyperparameters

Hyperparameter	Value
criterion	gini
max_features	n_features
class_weight	balanced
n_estimators	100
max_depth	10
min_samples_leaf	1 height

2.3 Deep Learning: CNN/RNN Models

A Convolutional Neural Network (CNN) is a specific type of deep learning that uses fully connected (dense) layers while a Recurrent Neural Network (RNN) is a sequential Neural Network model that looks at information as ordered [15]. Raw training data was subjected to a series of transformations: Fast-Fourier Transform, Short-time Fourier Transform/Hanning window, Mel banks, Mel Frequency Cepstral Coefficients (Discrete Cosine Transform)[2]. This process creates a uniquely identifiable sound profile of each species. Transformations did not allow very short signals to be processed, leading to the exclusion of the Blanchard's cricket frog from analysis. Efficiency and effectiveness of the two Neural Networks were optimized using 10-fold cross validation and confusion matrices as in the Random Forest. Optimization included batch normalization to improve convergence and generalization in training models [2, 8], plus increasing the number of extracted samples from WAV files.

3 RESULTS

The Random Forest was 97.80% accurate, the CNN was 97.85% accurate, and the RNN was 97.38% accurate in 10-fold cross validation (Table 3). While the overall accuracy for all models was comparable, the Random Forest excels in minimizing prediction time and accepting very short calls. The Random Forest model is 0.3596s/call faster than the Neural Networks in predictions. In multi-species trials, a novel species was introduced and labeled as false positives by the RNN and Random Forest, however false positives arose for four known species.

Table 3. Comparing Classification Models

	Random Forest*	**CNN**	**RNN**
Time to Build	12min 26s	50m 15s	51m 19s
Time to Predict	0.0004s/call	0.36s/call	0.37s/call
CV Accuracy	97.80%	97.85%	97.38%
Multi-Call Trial Accuracy	2 true pos, 5 false pos	0 true pos, 4 false pos	2 true pos, 5 false pos

*Random Forest built and tested with nine species, as opposed to eight in NNs

4 CONCLUSION

For final implementation, we recommend the Random Forest model because it detects all species, has the fastest prediction time, and is simple to rebuild and optimize. The RNN is superior to the CNN because its LSTM layers allow the model to learn at a faster rate and better distinguish multiple calls. In the future, false positives may be reduced by increasing the threshold for "unidentified" calls. The Neural Networks may be improved by adjusting the sampling rate to handle shorter calls. Adding the temperature of the environment as a signal feature may improve accuracy of the Random Forest.

ACKNOWLEDGMENTS

We thank Dr. Harmony Jankowski (Pervasive Technology Institute) for helping us in the writing process. We appreciate and thank Dr. Thomas Parr (National Park Service) for providing the audio files for our training sets from the Great Lakes Inventory and Monitoring Network.

REFERENCES

[1] [n.d.]. Drupal - Open Source CMS. https://www.drupal.org/

[2] Seth Adams. 2018. *DSP Background - Deep Learning for Audio Classification p.1.* Youtube. https://www.youtube.com/watch?v=Z7YM-HAz-IY&list=PLhA3b2k8R3t2Ng1WW_7MiXeh1pfQJQi_P&index=2&t=0s [Online; accessed <7/17/2019>].

[3] Jazzly Anderson, Thomas Slayton, Emmanuel Guido, Thomas Doak, and Sheri Sanders. 2018. Harvesting Field Station Data; Raspberry Pi Sensors to Jetstream Databases. (2018).

[4] M. Araya-Salas and G. Smith-Vidaurre. 2017. warbleR: an r package to streamline analysis of animal acoustic signals. http://dx.doi.org/10.1111/2041-210X.12624

[5] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.

[6] Debi Colbert. 2016. FrogWatch USA. (2016).

[7] Gary King and Langche Zeng. 2001. Logistic regression in rare events data. *Political analysis* 9, 2 (2001), 137–163.

[8] Ping Luo, Xinjiang Wang, Wenqi Shao, and Zhanglin Peng. 2018. Towards understanding regularization in batch normalization. *arXiv preprint arXiv:1809.00846* (2018).

[9] Raspberry Pi. 2015. Raspberry pi 3 model b. *online].(https://www. raspberrypi. org* (2015).

[10] J. D. Rodriguez, A. Perez, and J. A. Lozano. 2010. Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 3 (2010), 569–575.

[11] Edina Simon, Miklós Puky, Mihály Braun, and Béla Tóthmérész. 2011. Frogs and toads as biological indicators in environmental assessment. In *Frogs: biology, ecology and uses.* Nova Science Publishers, Inc., 141–150.

[12] John Towns, Timothy Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, Scott Lathrop, Dave Lifka, Gregory D Peterson, et al. 2014. XSEDE: accelerating scientific discovery. *Computing in science & engineering* 16, 5 (2014), 62–74.

[13] Sofia Visa, Brian Ramsay, Anca L Ralescu, and Esther Van Der Knaap. 2011. Confusion Matrix-based Feature Selection. *MAICS* 710 (2011), 120–127.

[14] Emma M Williams, Colin FJ O'Donnell, and Doug P Armstrong. 2018. Cost-benefit analysis of acoustic recorders as a solution to sampling challenges experienced monitoring cryptic species. *Ecology and evolution* 8, 13 (2018), 6839–6848.

[15] Wojciech Zaremba, Ilya Sutskever, and Oriol Vinyals. 2014. Recurrent neural network regularization. *arXiv preprint arXiv:1409.2329* (2014).