

## 2014 Bio-Acoustics Data Challenge for the International Community on Machine Learning and Bioacoustics

Christopher W. Clark & Peter J. Dugan  
Bioacoustics Research Program,  
Cornell Laboratory of Ornithology  
Cornell University

159 Sapsucker Woods Road, Ithaca, NY 14850  
phone: 607.254.2408 fax: 607.254.2460 email: [cwc2@cornell.edu](mailto:cwc2@cornell.edu)  
phone: 607.254.1149 fax: 607.254.2460 email: [pjd78@cornell.edu](mailto:pjd78@cornell.edu)

Award Number: N00014-14-1-0356  
<http://www.birds.cornell.edu/brp>

### **LONG-TERM GOALS**

This ONR grant promotes the development and application of advanced machine learning techniques for detection and classification of marine mammal sounds.

### **OBJECTIVES**

The objective is to engage a broad community of data scientists in the development and application of advanced machine learning techniques for detection and classification of marine mammal sounds.

### **APPROACH**

This research project will utilize an existing website, Kaggle.com, to host a bioacoustic detection-classification data challenge. This data challenge will be open to all international parties, allowing for a large quantity of solutions to be tested on a bioacoustic dataset.

Cornell will provide with existing, validated datasets of marine mammal acoustic exemplars for the contest. This requires time to validate and build test sets and training sets, which will be hosted on kaggle.com and will be freely available to the public. Cornell will first validate existing datasets and create the necessary formats for kaggle.com. Kaggle will host the contest, providing access to automated scoring tools and online public leader board.

This approach will be similar to that used in the first open Kaggle competition held in association with the Workshop on Machine Learning for Bioacoustics at the International Conference in Machine Learning, Atlanta, Georgia in 2013 [1, 2, 3].

### **WORK COMPLETED**

No work has yet been completed on this project. We are still in the process of assembling the test and training data sets.

## **RESULTS**

There are no results yet for this project.

## **IMPACT/APPLICATIONS**

The bioacoustics community collects very large amounts of marine acoustic data, which is challenging to analyze for the acoustic occurrence of marine mammals. Novel detection-classification technologies are needed to process these data. The marine mammal bioacoustics community would benefit from engaging and collaborating with the machine learning community.

## **PUBLICATIONS**

None

## **REFERENCES**

[1] P. Dugan, J. Zollweg, H. Glotin, M. Popescu, D. Risch, Y. LeCun and C. Clark (2014), "High Performance Computer Acoustic Data Accelerator (HPC-ADA): A New System for Exploring Marine Mammal Acoustics for Big Data Applications", ICML 2014, Workshop on Machine Learning for Bioacoustics, Beijing, China, *in press*.

[2] M. Pourhomayoun, P.J. Dugan, M.C. Popescu and C.W. Clark (2013), "Bioacoustic Signal Classification Based on Continuous Region Processing, Grid Masking and Artificial Neural Network," *ICML 2013 Workshop on Machine Learning for Bioacoustics*, arXiv preprint arXiv:1305.3635.

[3] M. Pourhomayoun, P.J. Dugan, M.C. Popescu, D. Risch, H. Lewis and C.W. Clark (2013), "Classification for Big Dataset of Bioacoustic Signals Based on Human Scoring System and Artificial Neural Network, *ICML 2013 Workshop on Machine Learning for Bioacoustics* , arXiv preprint arXiv:1305.3633.