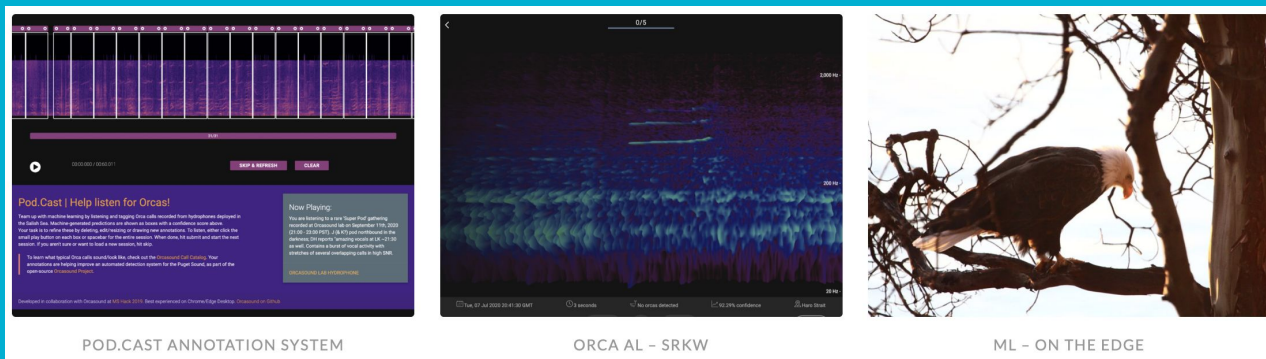


Orcasound & OrcaAL

Orcasound's open approach to orca detection
OrcaAL = Orca Active Learning

Presentation by Scott Veirs and Kunal Mehta
Meridian "Winter Webinar on Tuesday, December 15, 2020

Orcasound & OrcaAL

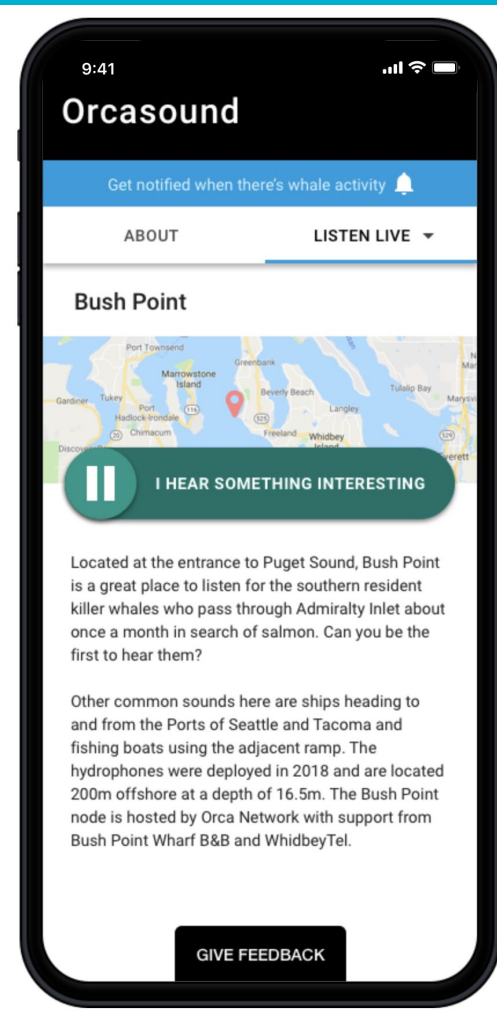
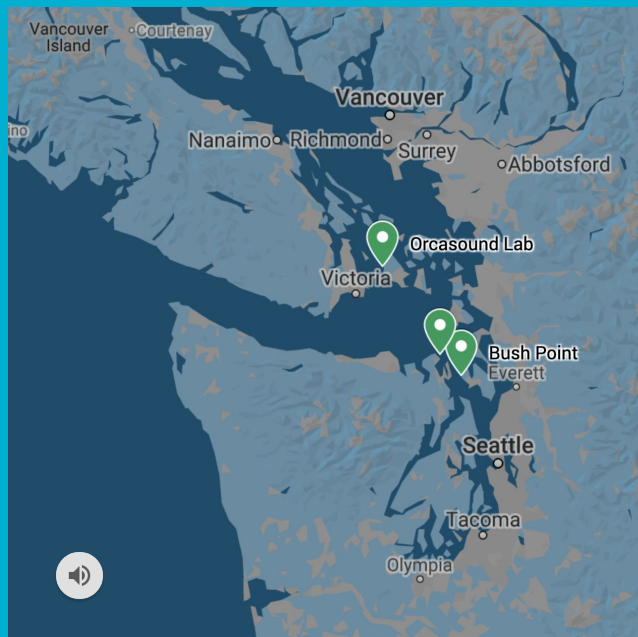


AI for orcas (#ai4orcas) -- ai4orcas.net
towards open bioacoustic data science...

ORCASOUND

A hydrophone network (WA, USA) & open-source software community

- 3 cabled nearshore sites streaming 24/7 in 2020
- Citizen scientists detect **orca sounds** in real-time via web app -- live.orcasound.net
- How can AI & human listeners work synergistically to conserve orcas & advance marine bioacoustics?



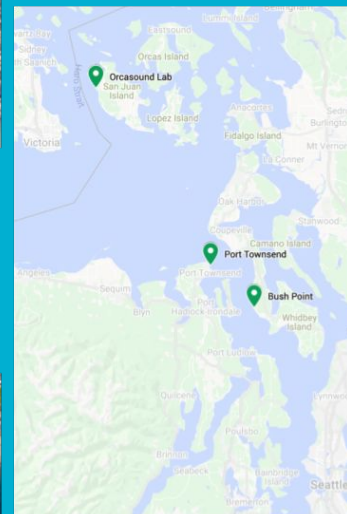
Orcasound = a hydrophone network

1970: Lime Kiln / Seasound

2002: Orcasound Lab

2008-2012 NOAA funding (expansion to 5 nodes)

2013-2015 Philanthropy only (decline to 2 nodes)



2016-2020

Cooperative agreement (now 14 NGOs)

2017 crowd-funding (\$20k Kickstarter)

2018+ open-sourcing, crowd-sourcing, open data...

Orcasound = open-source soft/hard-ware

<2002-17

- humans listening via mp3 streams
- Val building custom software, alone
- Scott building web static web sites, Google sheets, manual Twitter/email notifications...

2017-18: software + hardware (Kickstarter for v1 web app), live.orcasound.net launched Nov '18

2019-20: v2 UI beta-testing in Nov '19, launched in May, 2020

2021: v3 UI + a proliferation of related projects
(17 [Orcasound Github repositories](#))



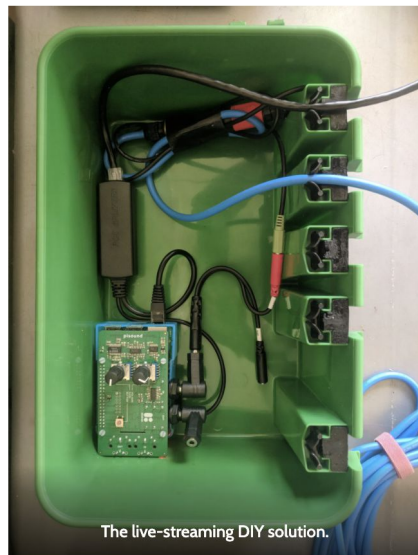
Val hacking node code.



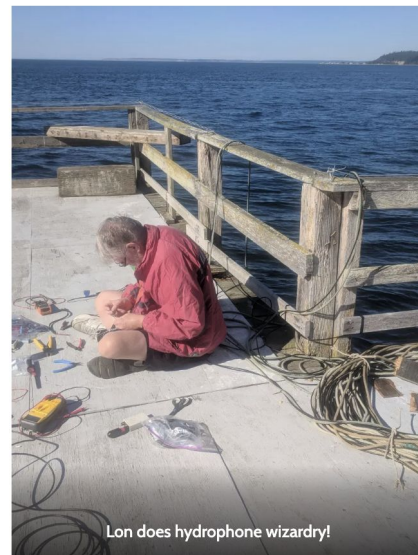
Binaural hydrophone stand.



DemocracyLab hackathon.



The live-streaming DIY solution.



Lon does hydrophone wizardry!

Orcasound = open raw data

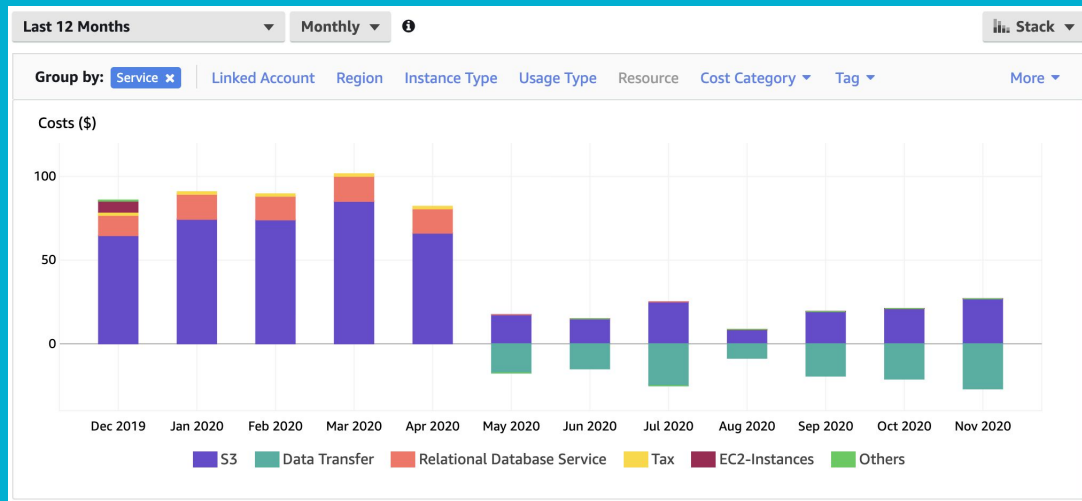
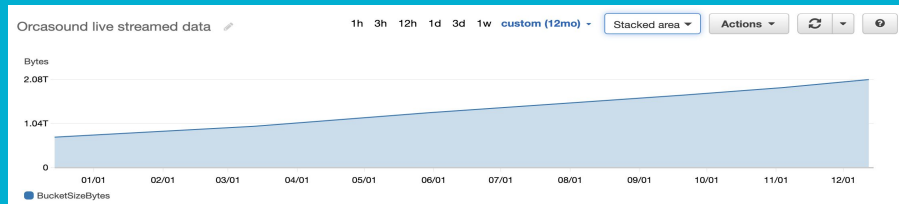
```
aws --no-sign-request s3 sync streaming-orcasound-net
```

Data volume

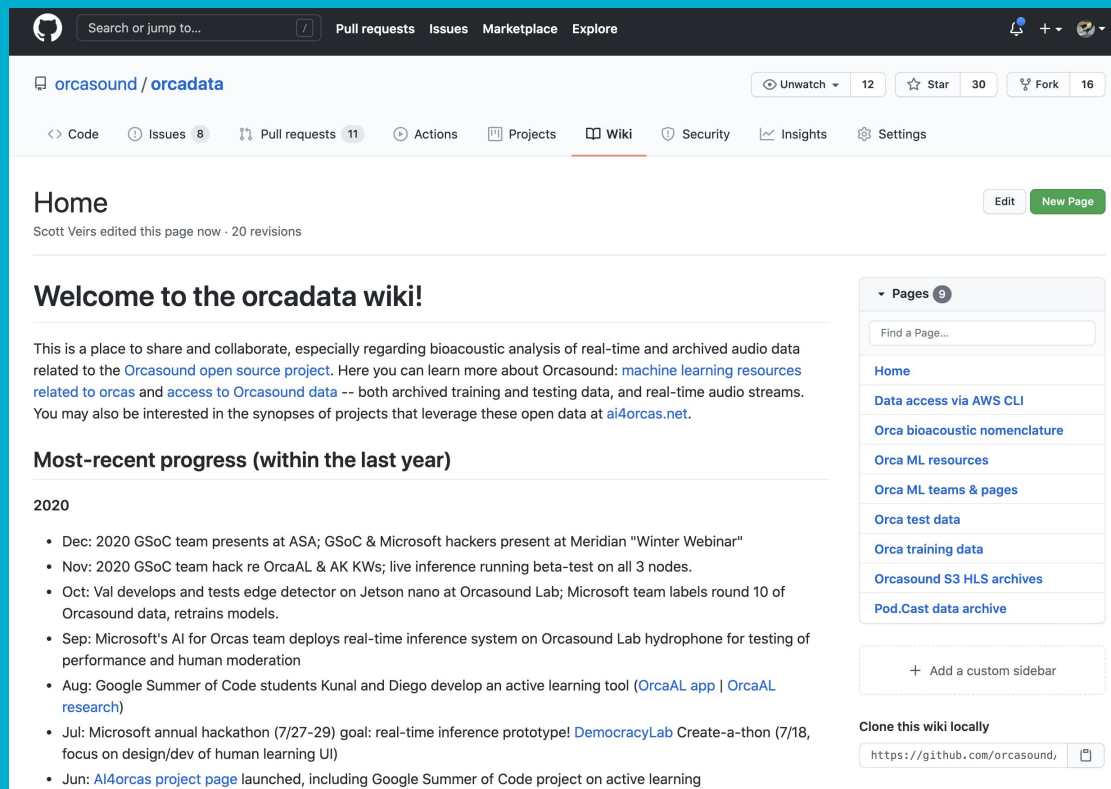
- 2.3 TB (HLS audio only)
- 0.4 TB/yr/node
- Increasing soon! (FLAC, stereo)

Data costs

- <\$100/mo w/3 nodes
- ~30\$/mo/node, but scaleable
- Free with credits!



Orcasound = open labeled data



The screenshot shows the GitHub Wiki page for the 'orcasound/orcadata' repository. The page has a dark header with the GitHub logo, a search bar, and navigation links for Pull requests, Issues, Marketplace, and Explore. Below the header, the repository name 'orcasound / orcadata' is displayed with statistics for Unwatch (12), Star (30), Fork (16), and a 'New Page' button. The main content area is titled 'Home' and includes a welcome message, a paragraph about the project's purpose, and a list of recent progress updates for the year 2020. A sidebar on the right lists various wiki pages under the heading 'Pages 9'.

Home Edit New Page

Scott Veirs edited this page now · 20 revisions

Welcome to the orcadata wiki!

This is a place to share and collaborate, especially regarding bioacoustic analysis of real-time and archived audio data related to the [Orcasound open source project](#). Here you can learn more about Orcasound: [machine learning resources related to orcas](#) and [access to Orcasound data](#) -- both archived training and testing data, and real-time audio streams. You may also be interested in the synopses of projects that leverage these open data at [ai4orcas.net](#).

Most-recent progress (within the last year)

2020

- Dec: 2020 GSoC team presents at ASA; GSoC & Microsoft hackers present at Meridian "Winter Webinar"
- Nov: 2020 GSoC team hack re OrcaAL & AK KWs; live inference running beta-test on all 3 nodes.
- Oct: Val develops and tests edge detector on Jetson nano at Orcasound Lab; Microsoft team labels round 10 of Orcasound data, retrains models.
- Sep: Microsoft's AI for Orcas team deploys real-time inference system on Orcasound Lab hydrophone for testing of performance and human moderation
- Aug: Google Summer of Code students Kunal and Diego develop an active learning tool ([OrcaAL app](#) | [OrcaAL research](#))
- Jul: Microsoft annual hackathon (7/27-29) goal: real-time inference prototype! [DemocracyLab](#) Create-a-thon (7/18, focus on design/dev of human learning UI)
- Jun: [AI4orcas project page](#) launched, including Google Summer of Code project on active learning


Pages 9

Find a Page...

- [Home](#)
- [Data access via AWS CLI](#)
- [Orca bioacoustic nomenclature](#)
- [Orca ML resources](#)
- [Orca ML teams & pages](#)
- [Orca test data](#)
- [Orca training data](#)
- [Orcasound S3 HLS archives](#)
- [Pod.Cast data archive](#)

+ Add a custom sidebar

Clone this wiki locally

<https://github.com/orcasound/> 

Grand challenges (human-machine synergy)

Where are they now?

Applied conservation problems

Automated (vs human-moderated) real-time notification of SRKW presence

1. Notify as soon as possible? only if you're sure it's a SRKW?
2. What is user's false positive tolerance?
3. Scott's dilemma: how to evoke empathy via a live "concert" of singing orcas without being annoying
 - a. How long will it last?
 - b. What will be the signal:noise?
 - c. How critical is the call to action for the orcas?

What are they saying?

Basic biological problems

Bioacoustics:

- marine mammal communication systems
- biosonar
- marine acoustic ecology
- soundscape analysis

Active Listening & Learning for Orca Sound Detection

Kunal Mehta, Diego Rodriguez,
Valentina Staneva, Jesse Lopez,
Abhishek Singh, Val Veirs, & Scott
Veirs

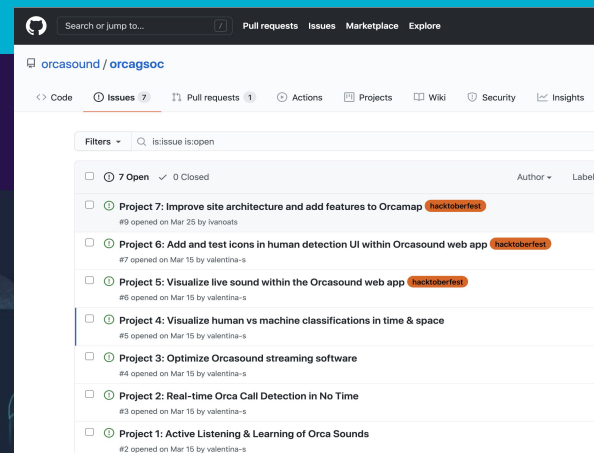
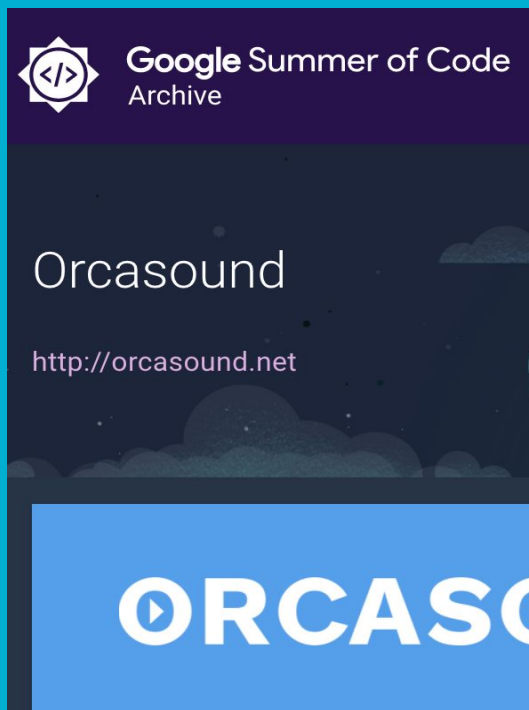
Presented at Acoustical Society of America

Thursday, December 10, 2020



Orcasound & Google Summer of Code

- GSoC: supports students & open-source software projects, like Orcasound
- 2019: OrcaCNN for Alaskan killer whale calls
- 2020: Build a tool to speed up the labeling of calls from the Southern Resident Killer Whales (SRKWs)



Orca call detection: open-source efforts

Open labeling tools for marine bioacoustics:

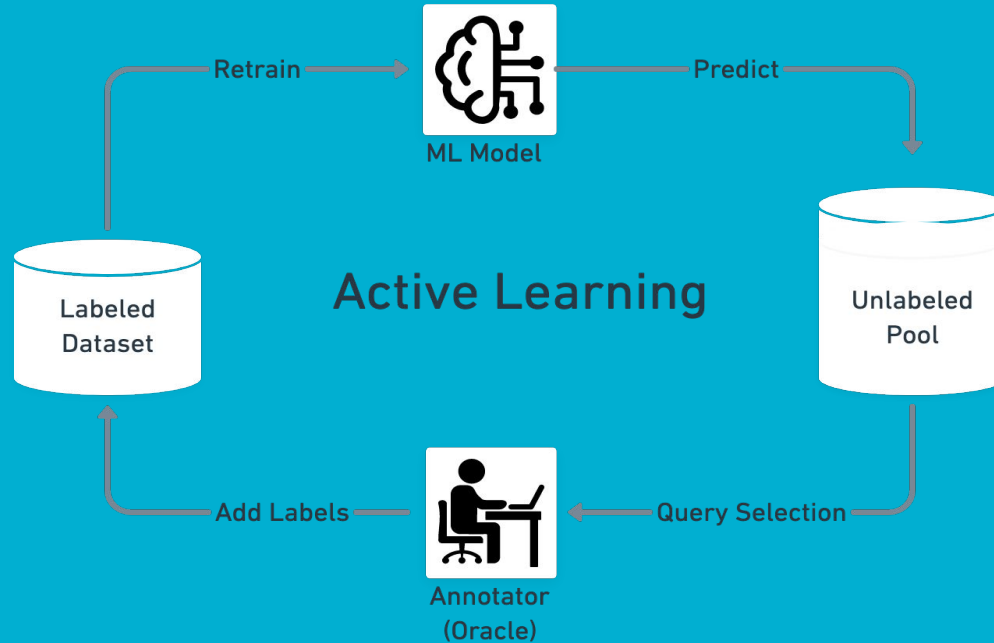
- [PAMGuard](#)
- [APLOSE](#)
- [Whaledr App](#)
- [Whale-FM](#)

Emerging open orca call machine learning models:

- [Orca-SPOT](#) (NRKWs)
- [OrcaCNN](#) (Alaskan KWs)
- [Orcasound models](#) (SRKWs)
 - University of Washington efforts
 - Microsoft hackathons
 - Other independent contributors
- More coming... (SRKWs)

Challenge: integrate efforts of human annotators & machine learning developers

Active Learning



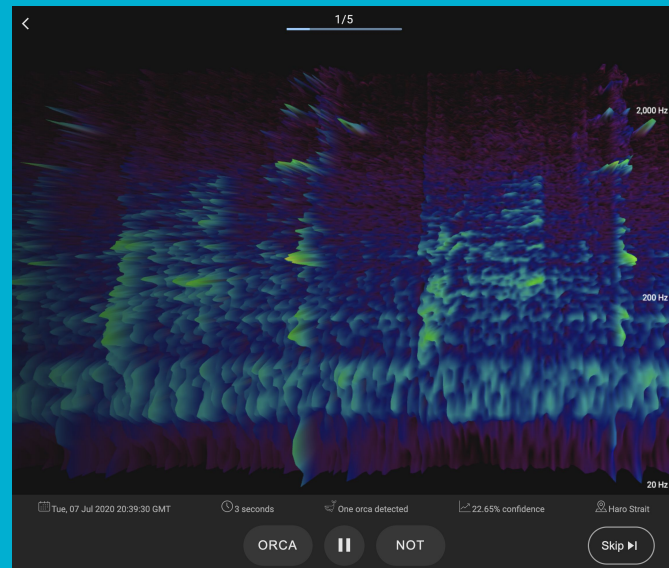
By labeling only a selective subset of samples we can save time, efforts, money, etc.

Orca active learning (OrcaAL)

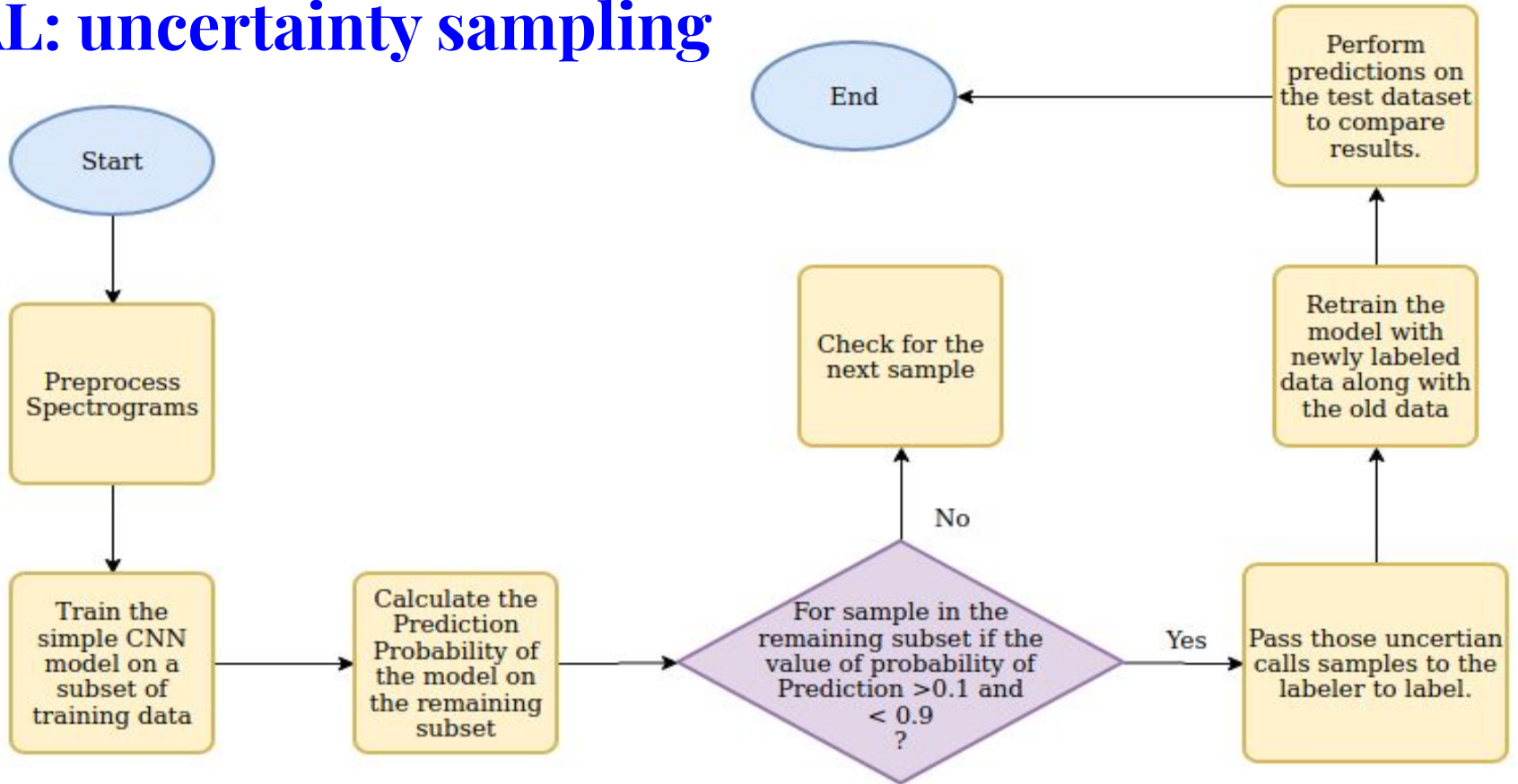
- Train a **deep learning** algorithm on an initial small labeled dataset
- Use “**uncertainty sampling**” strategy to label only those samples for which the model is most uncertain (i.e. samples with confidence near 0.5)
- Integrate model training, sample selection, and annotation in an AL tool.

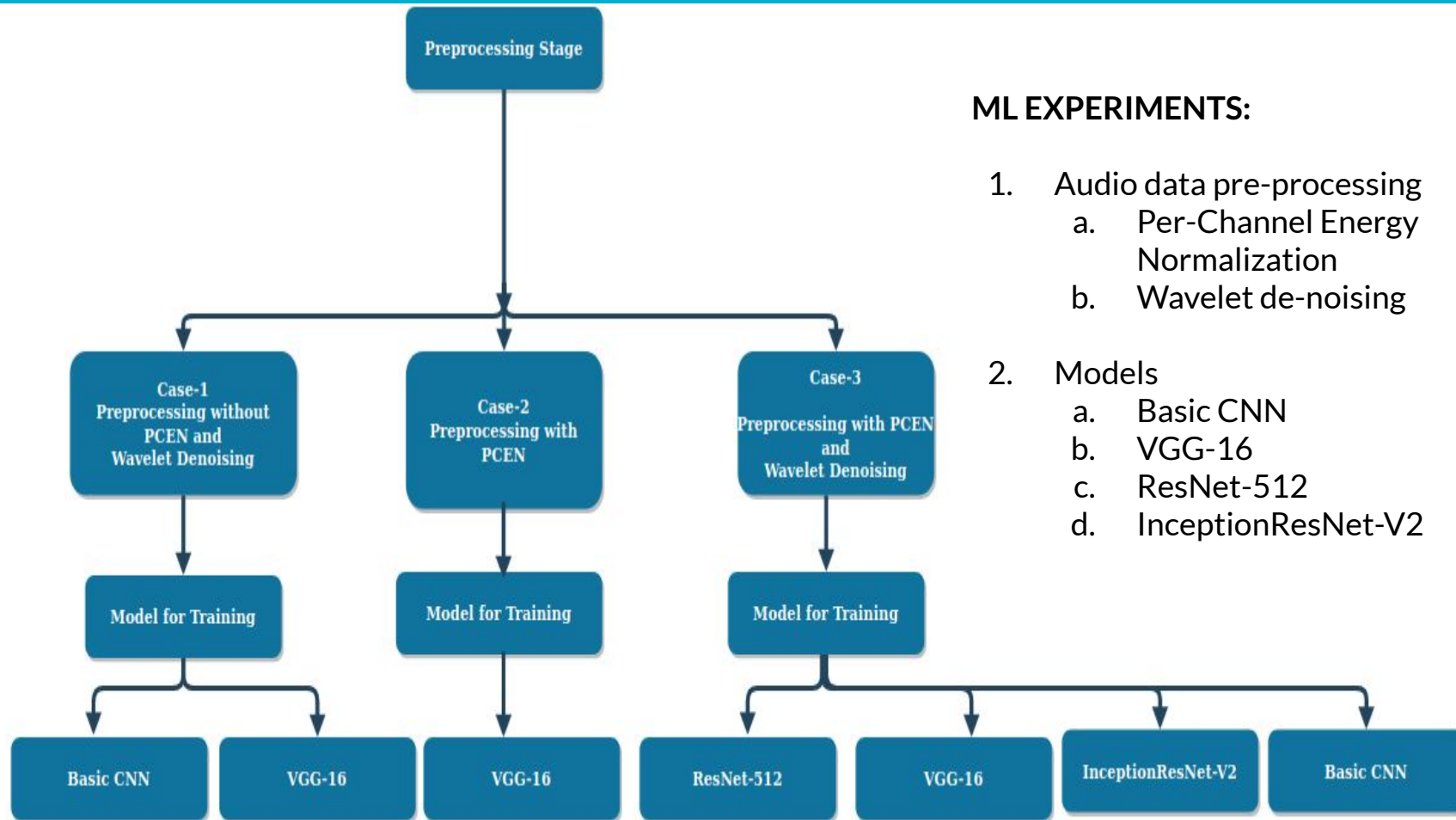
Motivating question: How much does labelling a subset of the samples increase accuracy?

Demo of the new citizen-science labeling tool!



AL: uncertainty sampling





Preprocessing spectrograms

Power Spectral Density Spectrogram
(Viridis)

Test accuracy = 79%

Power Spectral Density Spectrogram
(Gray)

Test accuracy = 75.7%

Mel spectrogram

Test accuracy = 80%

Mel spectrogram, Per-Channel Energy
Normalization & Wavelet Denoising

Test accuracy = 81%

Results: model performance

AL increased accuracy by 1 %
(i.e 83.5% ---> 84.5%)

Dataset sizes:

Training: 1394 samples

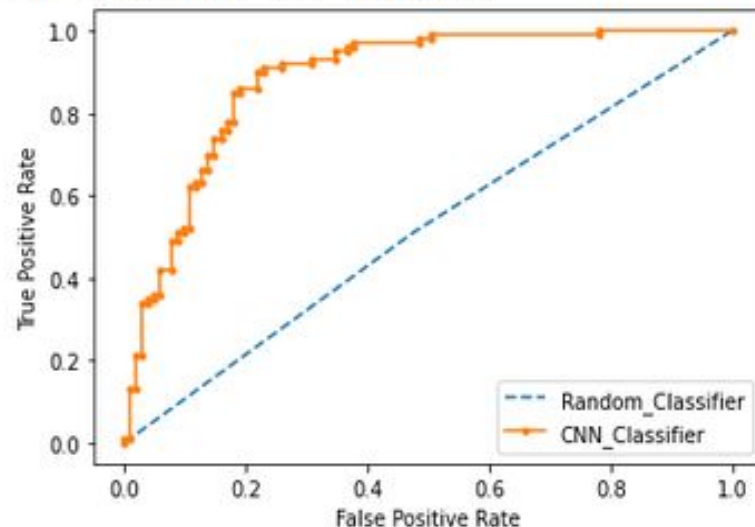
Active learning: 176 samples
(37 are uncertain)

Retraining: 1570 samples

Test dataset: 201 samples

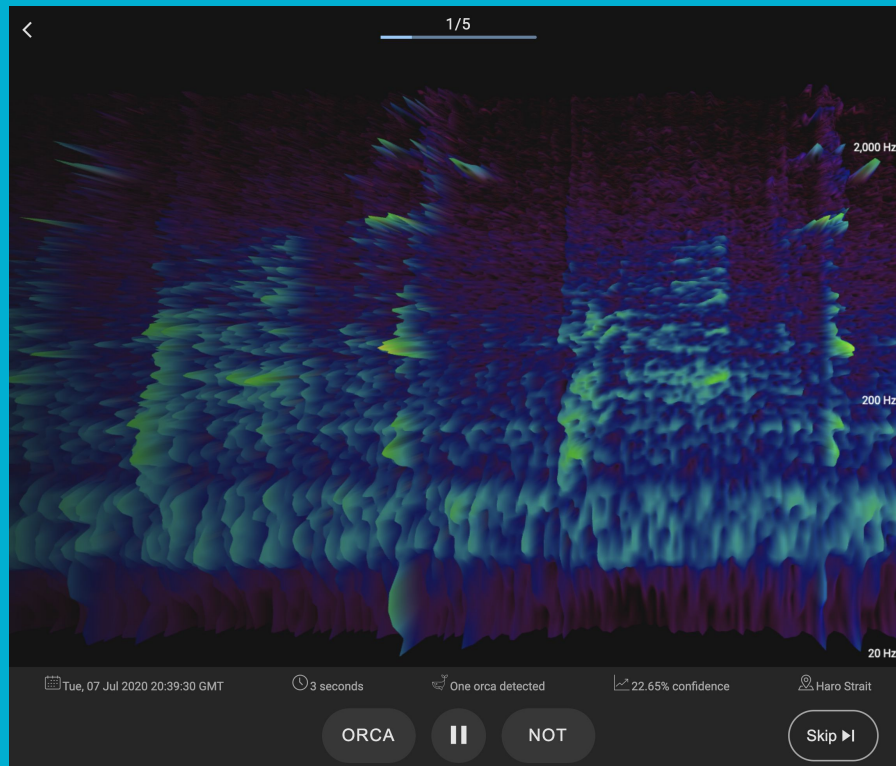
(calls & no-calls equally distributed in initial dataset)

Random Classifier: ROC AUC=0.517
CNN Classifier: ROC AUC=0.884

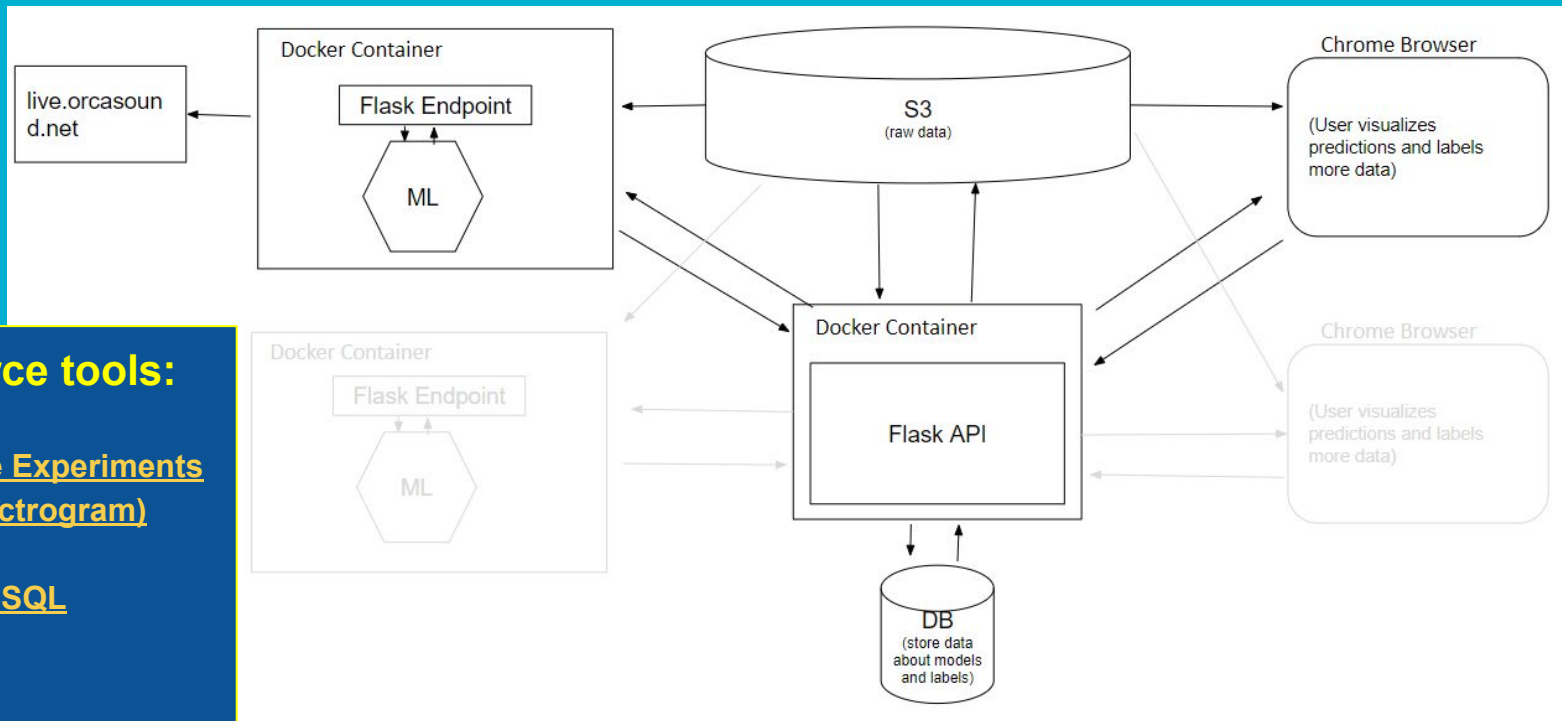


Live demo of OrcaAL

<http://orcaal.ai4orcas.net/>



OrcaAL's architecture



Open-source tools:

- [Chrome Experiments \(3D spectrogram\)](#)
- [Docker](#)
- [PostgreSQL](#)
- [Flask](#)

OrcaAL Acknowledgements & links

Thanks to all our collaborators!

- [Google Summer of Code](#)
- [UW eScience Institute](#)
- [Axiom Data Science](#)
- [Beam Reach](#)
- [North Gulf Oceanic Society](#)

Try the live demo!
orcaal.ai4orcas.net



Orcasound acknowledgements & links

Thanks to all our supporters!

- The Orcasound open-source community's volunteer hackers!
- \$10k [Amazon start-up credits](#) to Beam Reach
- [Microsoft AI for Earth](#) grant of \$15k Azure credits to UW eScience
- Project HALLO (2020-2022)

Join us!
ai4orcas.net

