Project 15

Counting Elephants by Base Frequency Spectrogram Analysis in Real Time

University of Oxford

Artificial Intelligence: Cloud and Edge Implementations

Cohort 2020-2021, Group 4

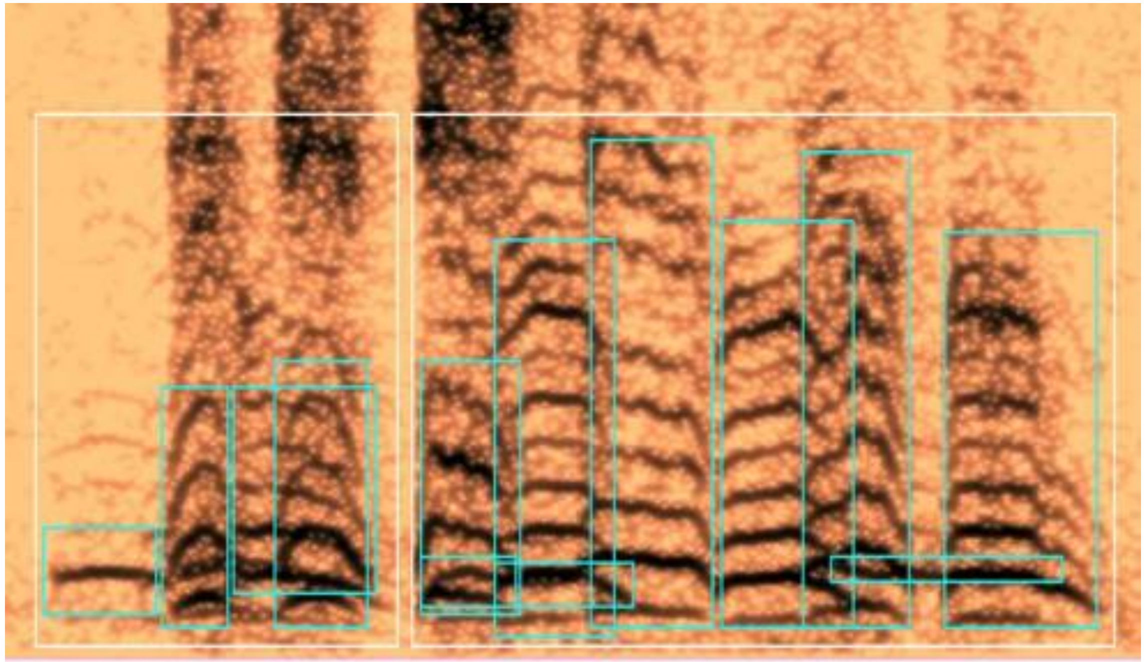
In cooperation with Cornell University and Microsoft

Group 4 Team

- Our team is composed of:
  - Abhishekh Baskaran - https://www.linkedin.com/in/abhishekh-jay-baskaran-566286bb/
  - Bas Geerdink - https://www.linkedin.com/in/geerdink/
  - Chandan Konuri - https://www.linkedin.com/in/konuri/
  - Henrietta Ridley - https://www.linkedin.com/in/henriettaridley/
  - Jay Padmanabhan - https://www.linkedin.com/in/jay-padmanabhan-95274b/
  - Paulo Henrique De Campos Filho - https://www.linkedin.com/in/paulohcampos/
  - Vishweshwar Manthani - https://www.linkedin.com/in/vishweshwar/

Goal

- The goal of the project was to **count the number of elephants** in a sound file
- To do so, we detected whether rumbles are belonging to the same elephant or not



white boxes might be what our detector boxes - but blue boxes are the actual number of differenct rumbles in the spectrogram

Literature

- Poole, Joyce H. (1999). Signals and assessment in African elephants: evidence from playback experiments. *Animal Behaviour*, 58(1), 185-193
- Jarne, Cecilia (2019). A method for estimation of fundamental frequency for tonal sounds inspired on bird song studies. *MethodX*, 6, 124-131
- Stoeger, Angela S. et al (2012). Visualizing Sound Emission of Elephant Vocalizations: Evidence for Two Rumble Production Types.
- O'Connell-Rodwell, C.E. et al (2000). Seismic properties of Asian elephant (Elephas maximus) vocalizations and locomotion. *Journal of the Acoustic Society of America*, 108(6), 3066-3072
- Heffner, R. S., & Heffner, H. E. (1982). Hearing in the elephant (Elephas maximus): Absolute sensitivity, frequency discrimination, and sound localization. *Journal of Comparative and Physiological Psychology, 96*(6), 926–944
- Elephant Listening Project, Cornell University: https://elephantlisteningproject.org/
- Project 15, Microsoft: https://microsoft.github.io/project15/

Introduction

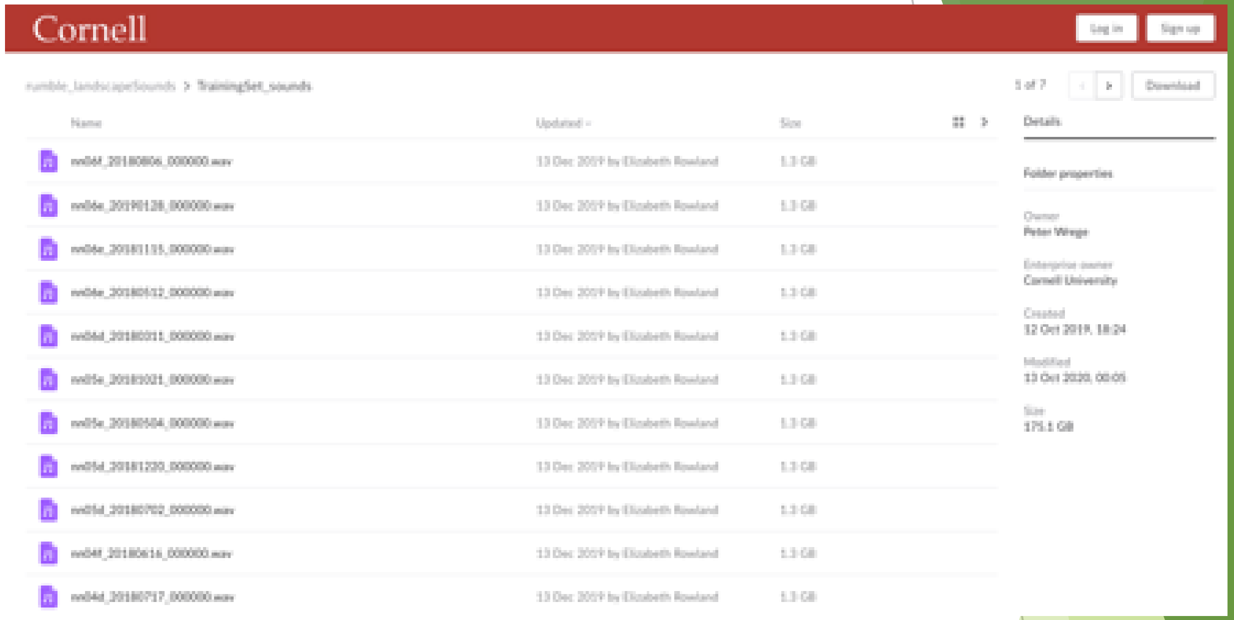
- Sound files can be analysed by transforming them into a 2D image: a *spectrogram* of time (seconds) vs frequency (Hertz). The third dimension is sound intensity (decibel), which can be shown as a colour or grayscale.
- Elephants produce *rumbles* to communicate with a typical frequency of 10 – 50 Hz and lasting 2 - 6 seconds
- One elephant rumble will have many *harmonics*, which are sound waves of increasing frequency.
- An elephant can be identified by its *base frequency*. If there are two slightly overlapping or separated rumbles with a different base frequency, they probably belong to separate animals.

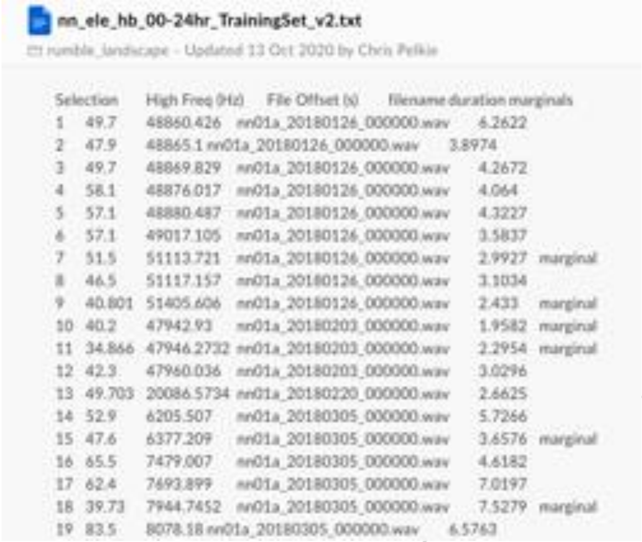
Data

We received a set of sounds files (.wav) and metadata that pointed us to the segments where elephants were likely to produce rumbles.

Challenges:

- Big data set
- Joining the files might be a challenge
- Labels / annotations don't mention the number of elephants





nn_ele_hb_00-24hr_TrainingSet_v2.txt

rumble_landscape - Updated 13 Oct 2020 by Chris Pelkie

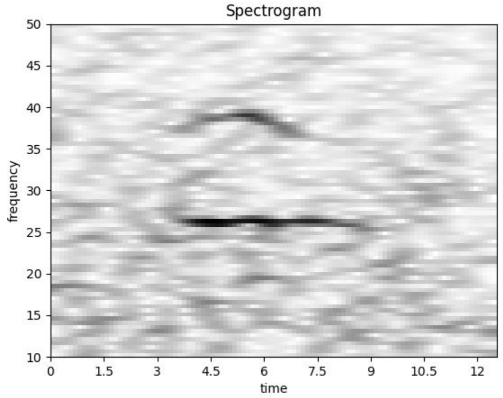
Selection	High Freq (Hz)	File Offset (s)	filename duration marginals		
1	49.7	48860.426	nn01a_20180126_000000.wav	6.2622	
2	47.9	48865.1	nn01a_20180126_000000.wav	3.8974	
3	49.7	48869.829	nn01a_20180126_000000.wav	4.2672	
4	58.1	48876.017	nn01a_20180126_000000.wav	4.064	
5	57.1	48880.487	nn01a_20180126_000000.wav	4.3227	
6	57.1	49017.105	nn01a_20180126_000000.wav	3.5837	
7	51.5	51113.721	nn01a_20180126_000000.wav	2.9927	marginal
8	46.5	51117.157	nn01a_20180126_000000.wav	3.1034	
9	40.801	51405.606	nn01a_20180126_000000.wav	2.433	marginal
10	40.2	47942.93	nn01a_20180203_000000.wav	1.9582	marginal
11	34.866	47946.2732	nn01a_20180203_000000.wav	2.2954	marginal
12	42.3	47960.036	nn01a_20180203_000000.wav	3.0296	
13	49.703	20086.5734	nn01a_20180220_000000.wav	2.6625	
14	52.9	6205.507	nn01a_20180305_000000.wav	5.7266	
15	47.6	6377.209	nn01a_20180305_000000.wav	3.6576	marginal
16	65.5	7479.007	nn01a_20180305_000000.wav	4.6182	
17	62.4	7693.899	nn01a_20180305_000000.wav	7.0197	
18	39.73	7944.7452	nn01a_20180305_000000.wav	7.5279	marginal
19	83.5	8078.18	nn01a_20180305_000000.wav	6.5763	

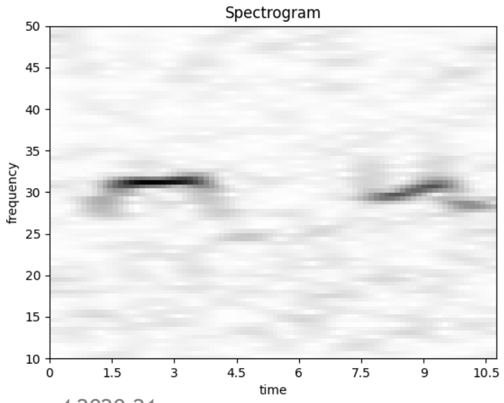
Data Pipeline

1. **Segmenting data:** based the metadata files, we create segments of a few seconds that contain the interesting information
2. **Spectrograms**: each data segment is transformed into a 2D image of time vs frequency (10-50 Hz), using FFT transformation algorithm, lowpass/highpass filters, and frequency filters
3. **Noise reduction:** each spectrogram is reduced of noise and transformed into a simple monochrome (black and white) image
4. **Contours detection**: each monochrome image is evaluated with a contour detection algorithm, to distinguish the separate 'objects' which in our case are the elephant rumbles
5. **Boxing**: for each contour (potential elephant rumble) we calculate the size (height and width) by drawing a box around the contour
6. **Counting**: we compare the boxes that identify the rumbles to each other in each spectrogram. Based on a few business rules, we count the number of unique elephant rumbles in each image

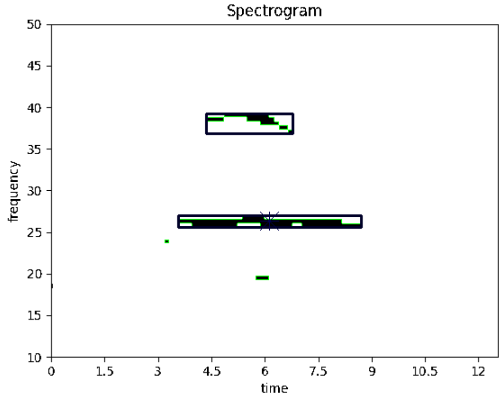
Samples

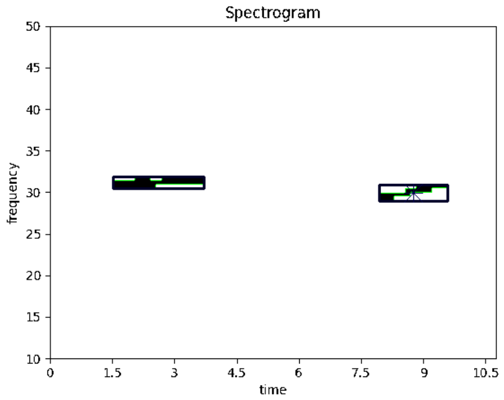
Steps 1, 2





Steps 3, 4, 5





Step 6

"1 elephant"

In this example, we see two harmonics of one elephant rumble

"1 elephant"

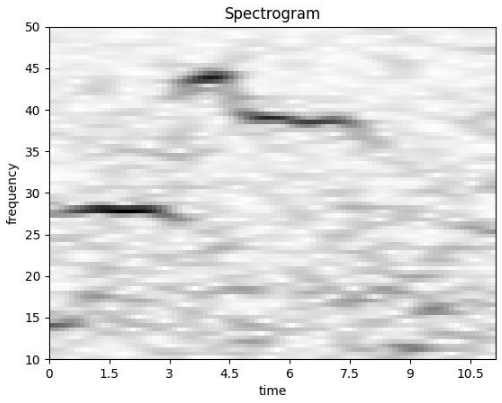
In this example, we see two elephant rumbles with the same base frequencies, a few seconds apart; they probably belong to the same animal

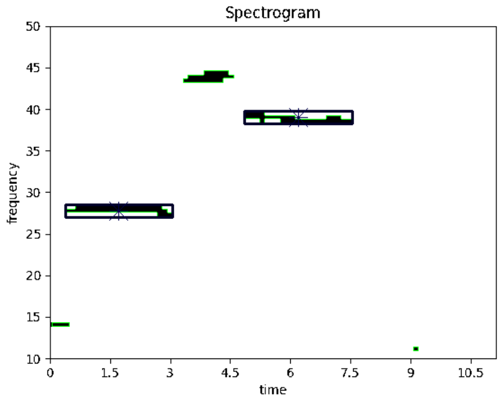
Samples

Steps 1, 2

Steps 3, 4, 5

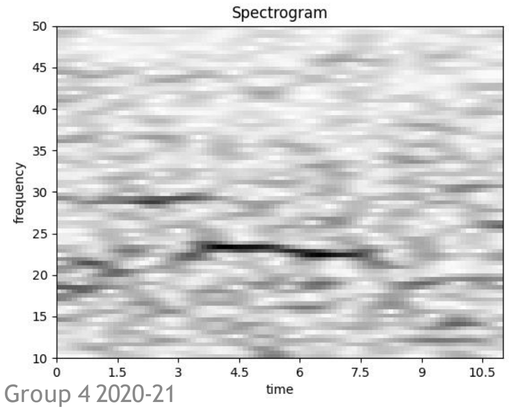
Step 6

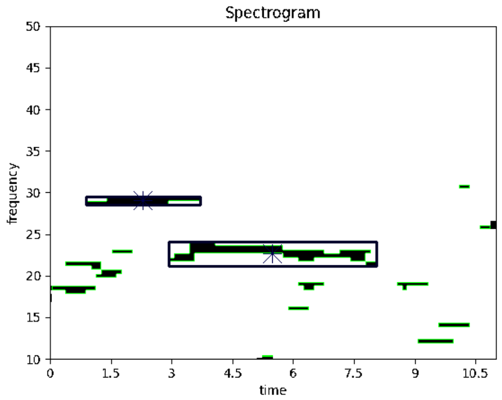




"2 elephants"

In this example, we see two elephant rumbles with different base frequencies; they probably belong to separate animals





"2 elephants"

In this example, we see two elephant rumbles with different base frequencies; they probably belong to separate animals

Source Code

- The source code is made available at: https://github.com/AI-Cloud-and-Edge-Implementations/Project15-G4
- All code is written in Python and runs on premise or in the cloud (Azure)
- We used the following frameworks to process and analyze the data:
  - boto3 for connecting to Amazon AWS
  - Numpy, Pandas, SciPy and MatPlotLib for statistical analysis and visualization
  - Librosa for FFT
  - noisereduce for noice reduction
  - SoundFile
  - OpenCV for contour detection

Results

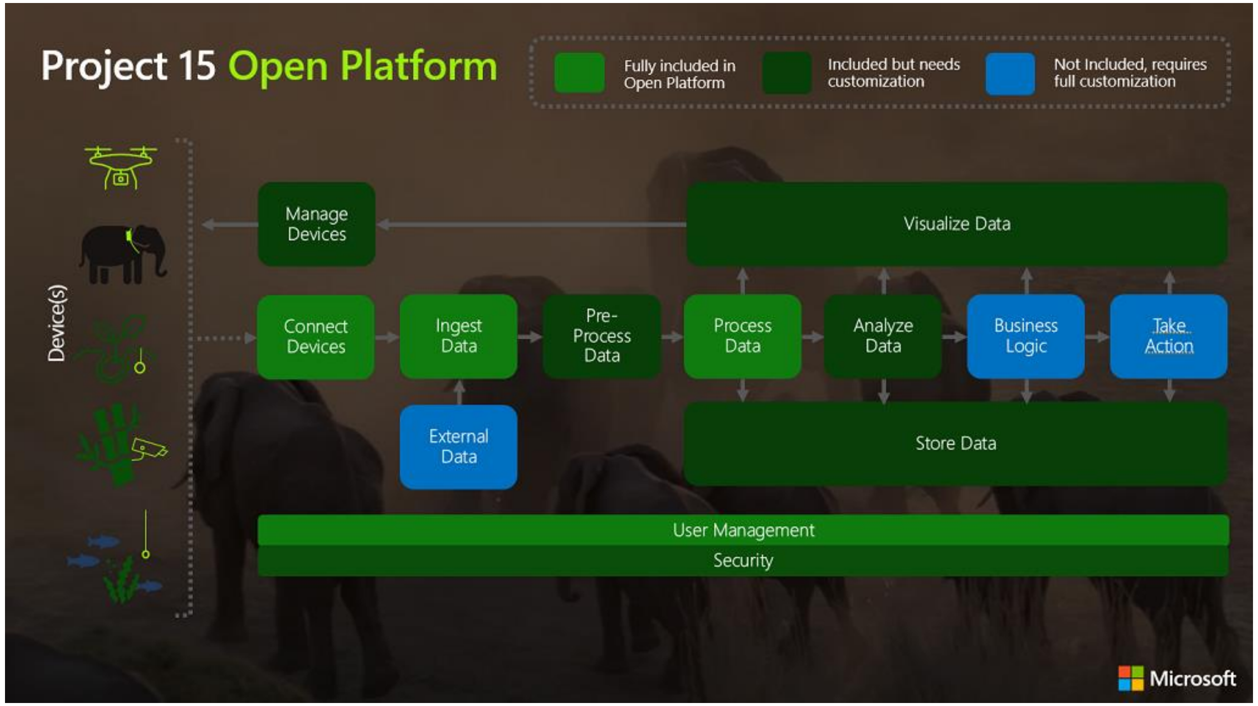
- We analysed 3935 elephant sounds:
  - 112 spectrograms were identified as containing 0 elephants
  - 3277 spectrograms were identified as containing 1 elephant
  - 505 spectrograms were identified as containing 2 elephants
  - 40 spectrograms were identified as containing 3 elephants

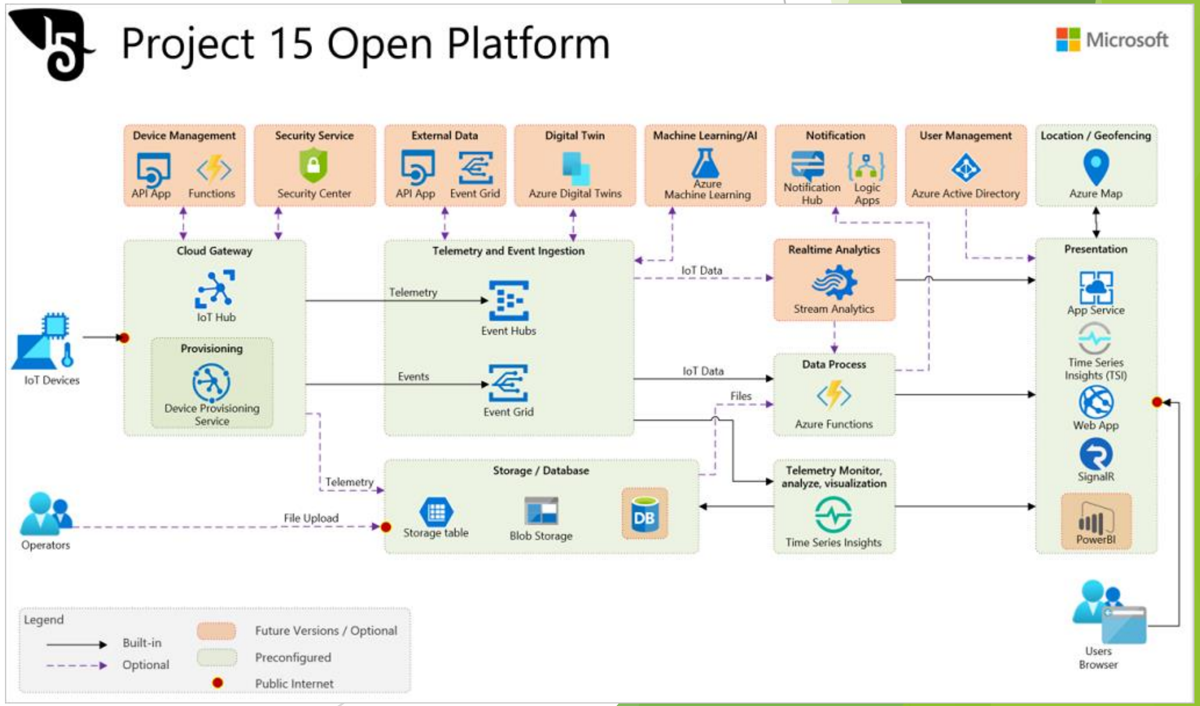
Results of the Boxing algorithm

- The boxing algorithm was evaluated by Liz Rowland of Cornell University
- The reported accuracy of the model is:
  - 97.29 % for the Training dataset (3180 cases)
  - 99.29 % for the Testing dataset (758 cases)

- This proves that the model is useful for counting elephants
- In combination with other models (elephant detection), many interesting use case can be built with this model, for example visualizing elephant movements and detecting poaching

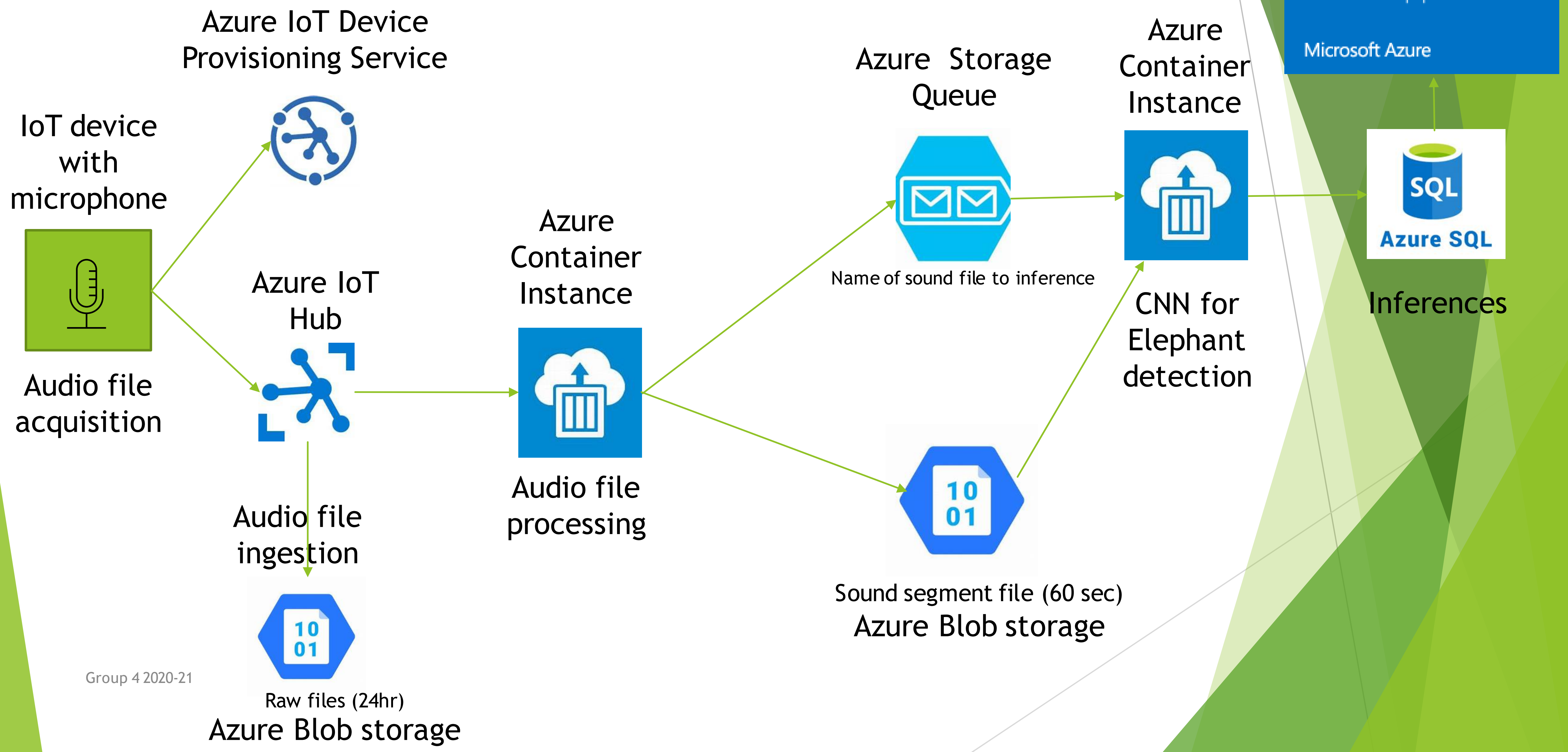
Architecture





Group 4 2020-21

Project 15 Architecture

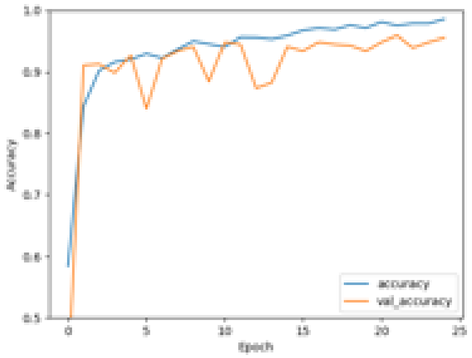


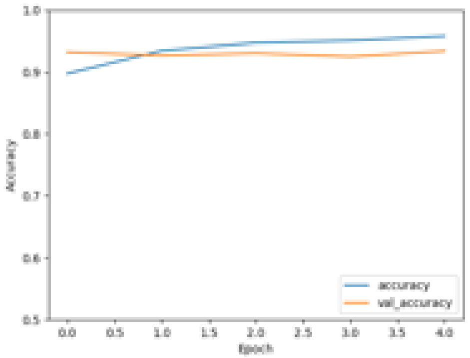
Building ML Models

- **Aim**
  Using the processed spectrogram data as an input to a CNN to automatically categorise how many elephants are present
- **Why are we doing this?**
  - To enable automation the workflow end to end
  - To improve accuracy by reducing human error
  - To save time, enabling researchers to focus their attention on complex problems
- **Our Approach**
  Transfer learning looks to take advantage of models which have been pre-trained on large datasets, then fine tuning to our specific problem. This approach is becoming very popular for several reasons (quicker time to train, better performance, not needing lots of data) and we found it to work well.

Model Summary

- Implemented using keras with a tensorflow backend.
- To evaluate the performance of our models we looked at the following measures of our two most promising architectures:
  - Resnet50
    - accuracy: 0.9620
    - loss: 0.1622
  - VGGNet
    - accuracy: 0.9477
    - loss: 0.3252





These training curves show us the performance of the model over time as it learns the underlying patterns of our data

Model - Resnet50

- Below configuration was found to be optimal while running the classification task on Resnet50
  - Epochs: 25
  - Batch Size: 100
  - Weights = "imagenet"
  - Intermediate dense layers:
    - Nodes: 4 layers of 256,128,64 respectively
    - activation = 'relu'
    - Dropout = 0.5
    - BatchNormalization()
  - Final dense layer:
    - Nodes: 3
    - activation = 'softmax'
  - Optimizer: Adam with a learning rate of 0.001

Further Research

- Machine learning on spectrograms using labelled data
- Automatic classification and better acoustic analysis (https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0048907)

- Further fine-tuning of the boxing algorithm might lead to even better results, e.g.
    - Fixing the time axis in the spectrograms
    - Increasing the frequency range
    - Other (better) noise reduction techniques

Conclusions

- Elephant counting based on base frequency analysis is possible
- The team delivered a ready-to-use software library for counting elephants that with a high accuracy (97% on selected cases)
- The software can be used in the IoT Hub (Project 15) or on-premise
- The application can be integrated into other software
- A machine learning model (VGG or Resnet50) could be used to count the elephants instead of the rule-based boxing algorithm
- Further research is needed to improve the results, for example for broadening to other species

Thanks

- Many thanks to all people who helped with the project, by providing insights, performing reviews, and participating in meetings:
  - Peter Wrege (Cornell University)
  - Liz Rowland (Cornell University)
  - Lee Stott (Microsoft)
  - Sarah Maston (Microsoft)
- Thanks to the organizers of the "Artificial Intelligence – Cloud and Edge Implementations" course:
  - Ajit Jaokar (University of Oxford)
  - Peter Holland (University of Oxford)