Project 15
Group 1

Objective - Goal

- The goal of our project is to use cloud based Azure IoT Edge and ML/AI Cloud based technology to train, test and deploy AI models to an IoT Edge device
- The device is capable of detecting both elephant and gunshot sounds at a defined locations with the specific purpose of catching individuals who are killing elephants for profit

Background - Project Preamble - Group

Geographic Extent: latitude 1.89020 to 2.32474, longitude 16.44315 to 16.74016 (WGS84)

Site	Habitat (at recorder site)	Protection
nn01a	Gilbertiodendron (mono-dominant); very tall canopy, sparse understory	national park
nn01c	mixed forest-closed	national park
nn01d	mixed forest-closed	national park
nn01e	mixed forest-closed	national park
nn01f	Gilbertiodendron (mono-dominant); very tall canopy, sparse understory	national park
nn01g	Gilbertiodendron (mono-dominant); very tall canopy, sparse understory	national park
nn02a	Gilbertiodendron (mono-dominant); very tall canopy, sparse understory	national park
nn02b	mixed forest-closed	national park
nn02c	Gilbertiodendron (mono-dominant); very tall canopy, sparse understory	national park
nn02d	mixed forest-closed	national park
nn02e	mixed forest-closed	national park
nn02f	mixed forest-closed	national park
nn02g	mixed forest-open	national park
nn03a	mixed forest-closed	national park
nn03b	Gilbertiodendron (mono-dominant); very tall canopy, sparse understory	national park
nn03c	Gilbertiodendron (mono-dominant); very tall canopy, sparse understory	national park
nn03d	mixed forest-closed	national park
nn03e	Gilbertiodendron (mono-dominant); very tall canopy, sparse understory	national park
nn03f	swamp forest	national park
nn03g	mixed forest-closed	national park
nn04a	foret mixte	national park
nn04b	mixed forest-closed	national park
nn04c	mixed forest-closed	national park
nn04d	Gilbertiodendron (mono-dominant); very tall canopy, sparse understory	logging – active 2017-2018
nn04e	mixed forest-open	logging – active 2017-2018
nn04f	mixed forest-closed	logging – active 2017-2018
nn05a	mixed forest-closed	national park
nn05b	mixed forest-closed	national park
nn05c	mixed forest-closed	logging – active 2017-2018
nn05d	mixed forest-closed	logging – active 2017-2018
nn05e	swamp forest	logging – active 2017-2018
nn05f	mixed forest-closed	logging – active 2017-2018

Challenges to forensic analysis of gunshots … due to broadcast and noisy nature of acoustic channel.

Gunshots are a result of multiple acoustic events, namely:

- Muzzle blast (explosion inside barrel)
- Ballistic shockwave (supersonic projectile)

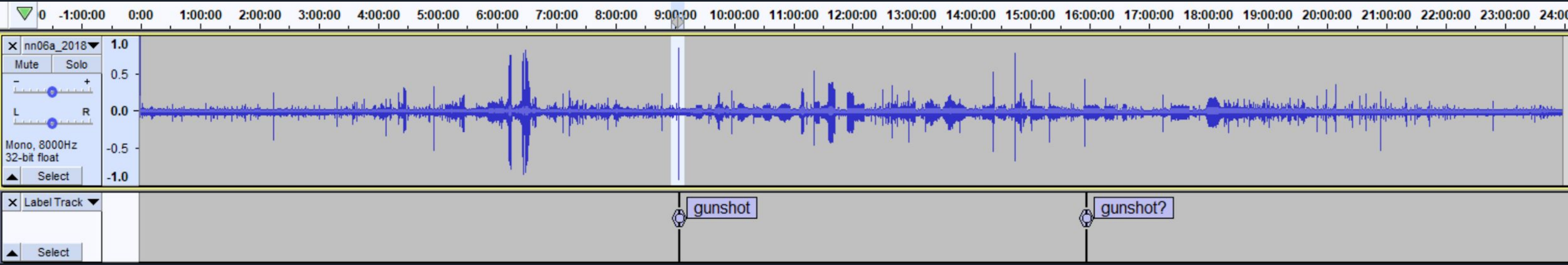
Most ad-hoc models do not address 'Room Reverb', 'Concert Reverb', and doppler effect.

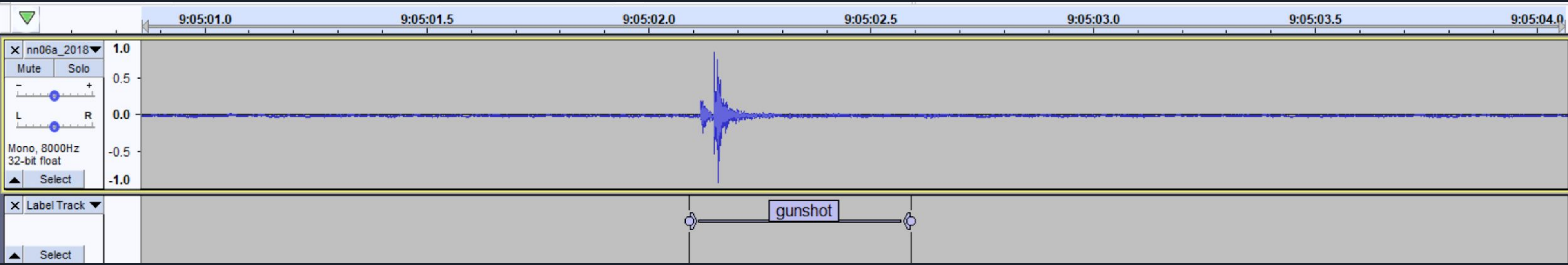
Raponi et al. Sounds of Guns:

Analysis of problem

First task to analyse data (ELP audios). Some questions to ask:

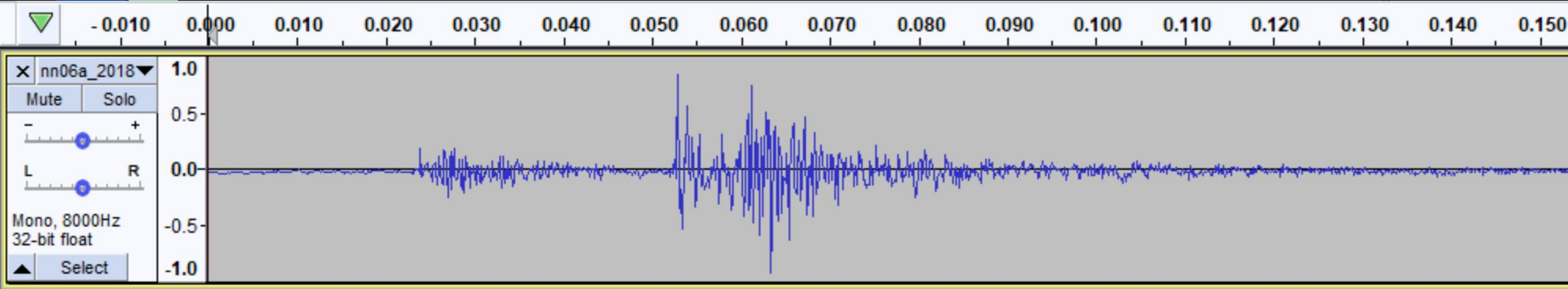
- What does a gunshot sound 'look' like in the 24 hr audio data gathered.
- Does it look similar to literature search of 'domesticated' gunsounds.

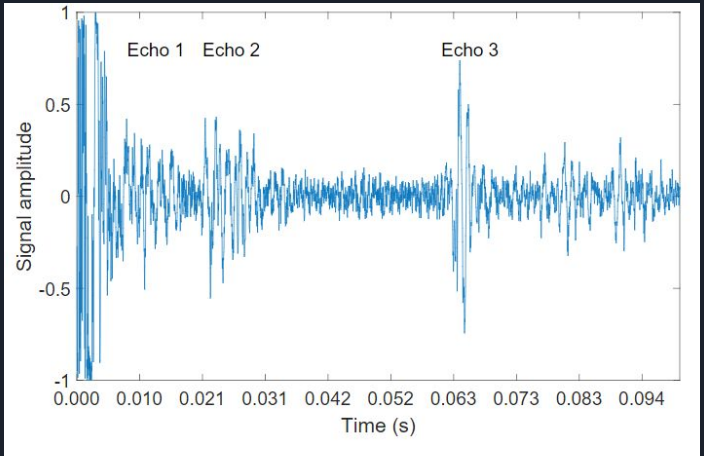


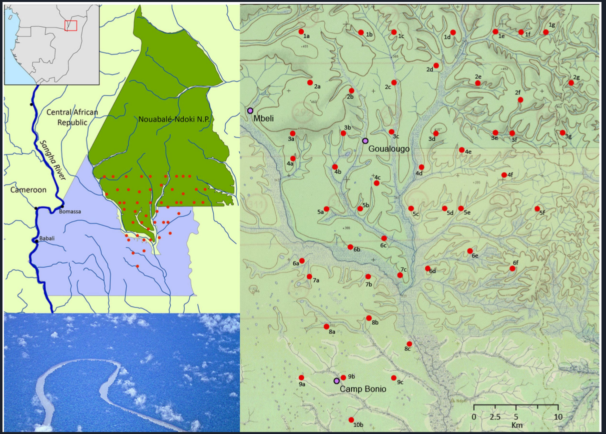


Data preview

Does it look similar to literature search of ‘domesticated’ gunsounds.



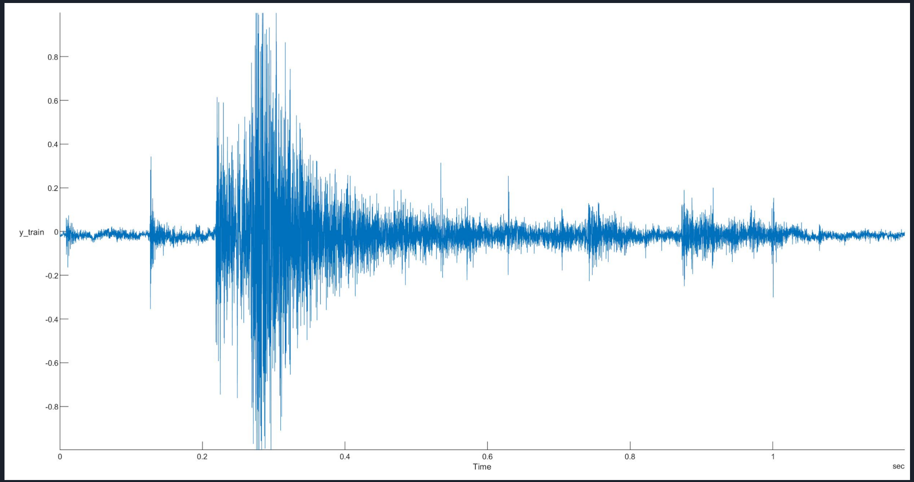


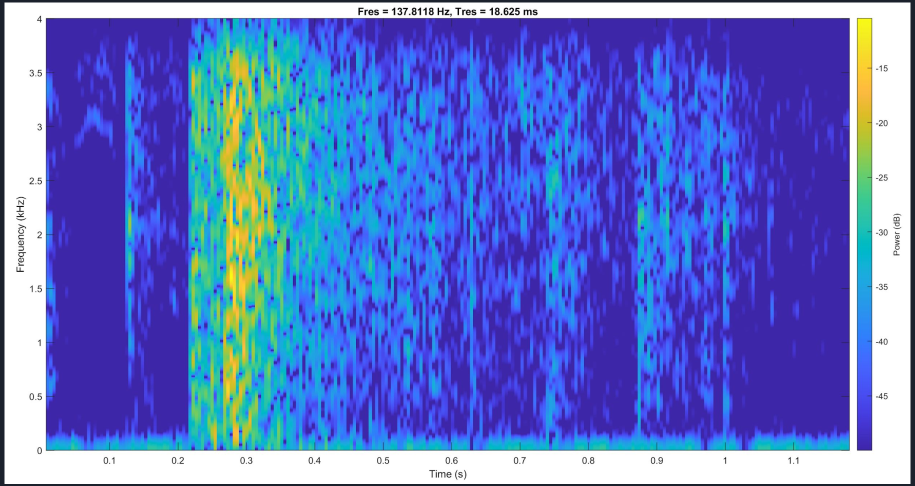


Data preview

From power spectrogram we can see that a muzzle blast has a large frequency range and differs between gun types. This may be a challenge to filter data.

The question now is how best to ‘harvest’ this data in the large ELP audios.

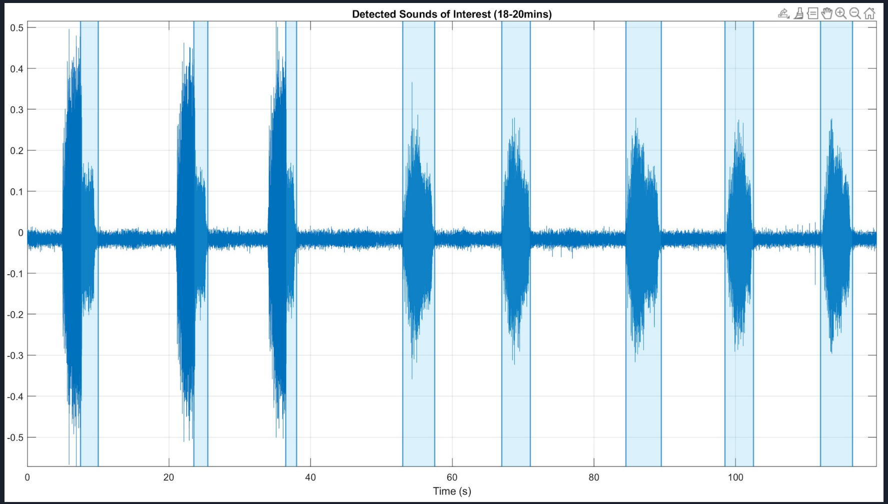


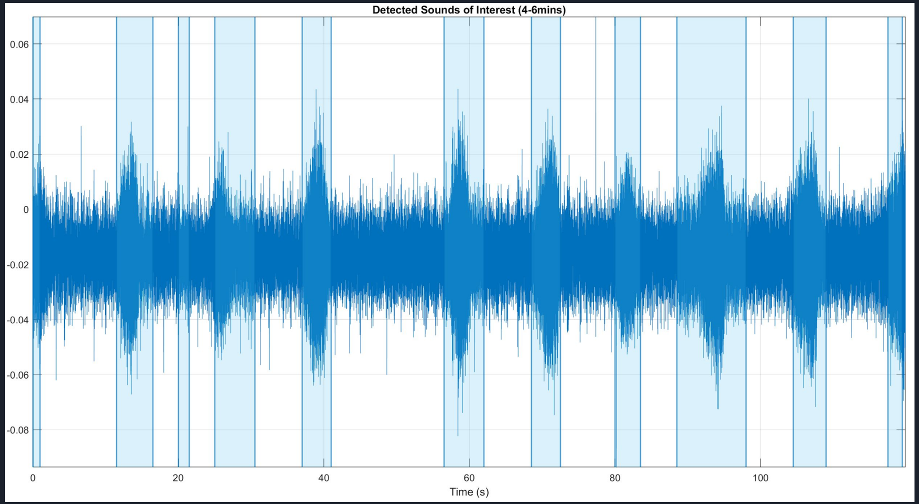


Data preview leading to a MATLAB data processing approach

MATLAB function to the ‘detectSpeech’ to find audio snippets of interest. This will be stored as a datastore.

This method is to only analyse / compare the audios against a trained audio of a gunshot. This is lighten the load as the entire audio ELP data will not have to be analysed but only snippets of interest.





Observation

Observation 1: Acoustic signature of guns are distinctive based on literature, unique to even gun type. This can be used to help label audio snippets for gun sounds.

Observation 2: However echo characteristics differ than those in the wild as gun shots are not cultivated in audio-rooms or urban environments (3ms vs 60ms).

Observation 3: Frequency range of gunshots are larger (100-4000Hz), bandwidth filters should be explored. (0-24KHz).

Observation 4: From literature search, model may require human-assisted supervised learning from ELP.

Observation 5: Validity of using purely cultivated gunshot sounds to detect gunshot sounds in wild (ELP data) is questioned?

Background - Project Preamble - Group

The task is to create and implement that allows detecting gunshots. The problem is that there are few data points and also these are very rare sounds. We were able to identify two alternatives to solve this problem.

Create a shot classification model using:

Use the actual recordings and then identify gunshots.

- Computationally expensive when processing and analyzing all audios files (24 hours of audio in each recording)

Use an external database with gunshot sounds

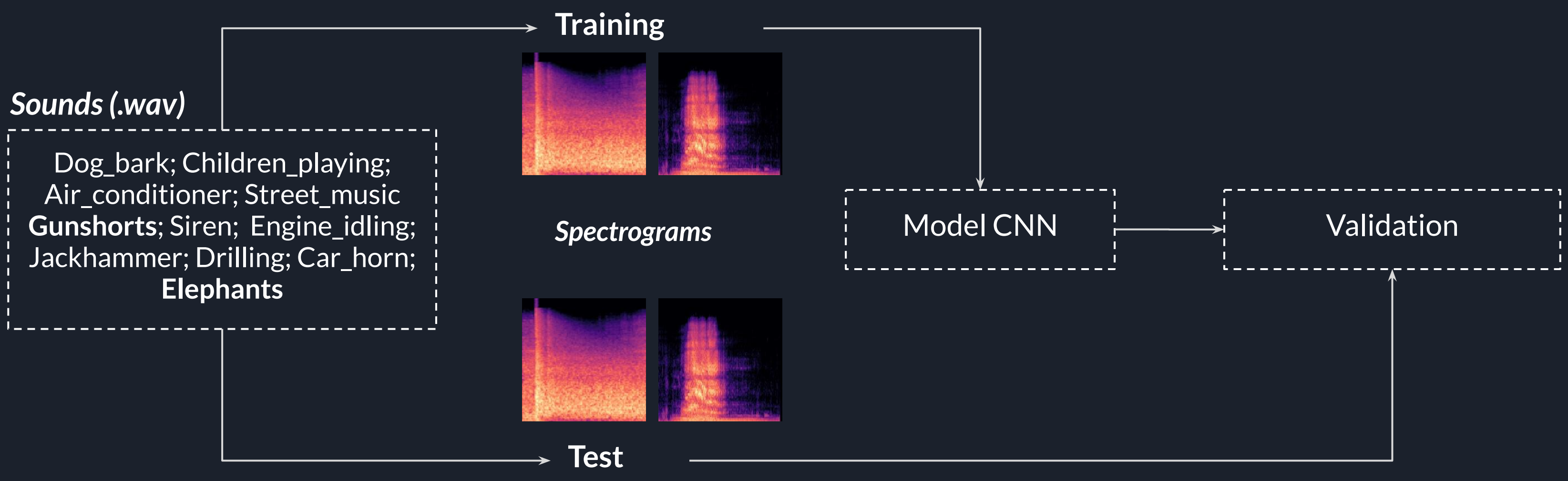
- Shots from external databases do not have ambient sound. There is a risk that the model cannot identify actual gunshots.

Deploy the model in a cloud environment

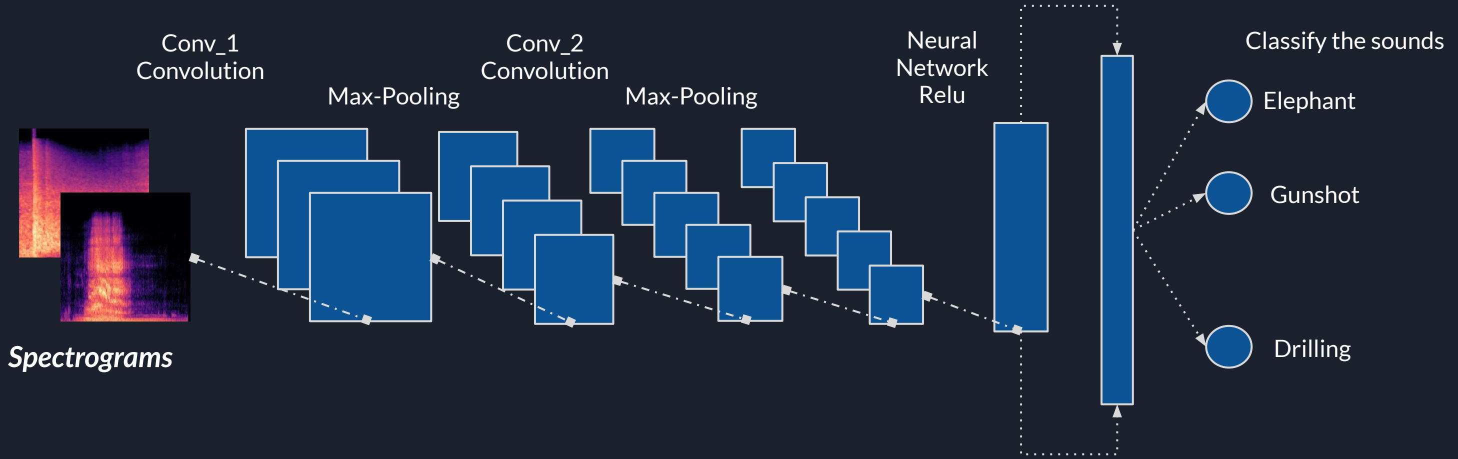
Edge device

Data structure and model flow

We decided to use an external database that has different types of sounds (10 classes) including gunshots. Additionally, we also add elephant sounds.



Model structure



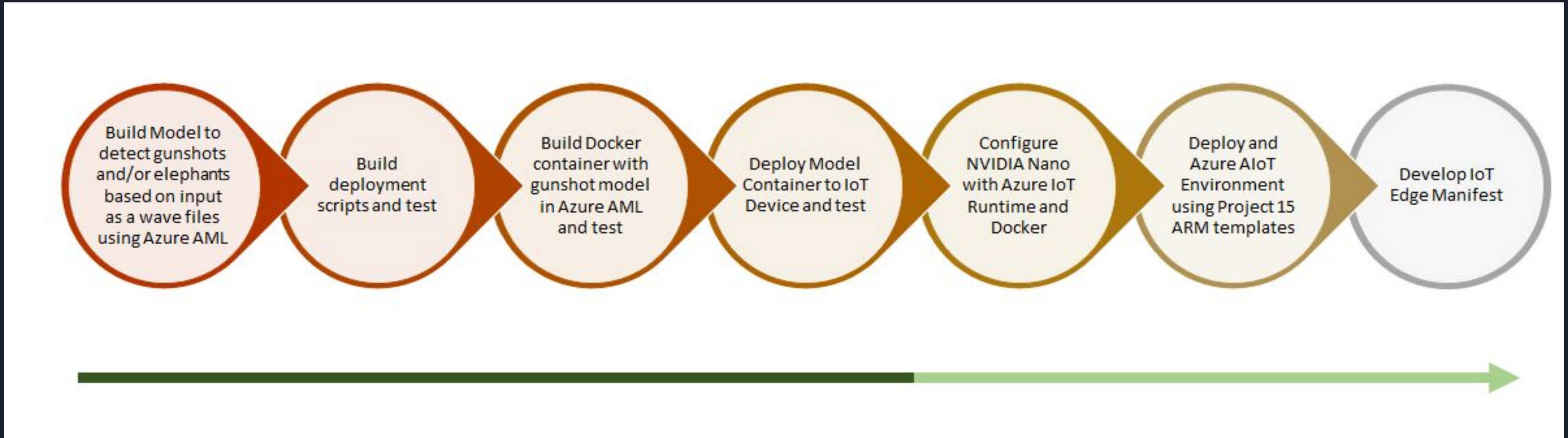
Assumptions - Group

- During the model training we are making the assumption that we can use a large corpus of urban gunshot wav files to train a model to detect gunshot sounds that would occur in areas that are very different than those in urban areas

- We are assuming that the IoT edge device will have connectivity either with a cellular connection or by a satellite connect neither of which would be continuous i.e. the edge device could experience periods without connectivity.

- The edge device will have access either via the cell or satellite connection to lat/long coordinates or would have them as a device twin property updated from the hub once the device was in position (the latter being the least complicated configuration)

Project 15 | End to end workflow



Build Models to detect gunshots or elephant sounds using wav files

Methodology

Model Step by Step Process:

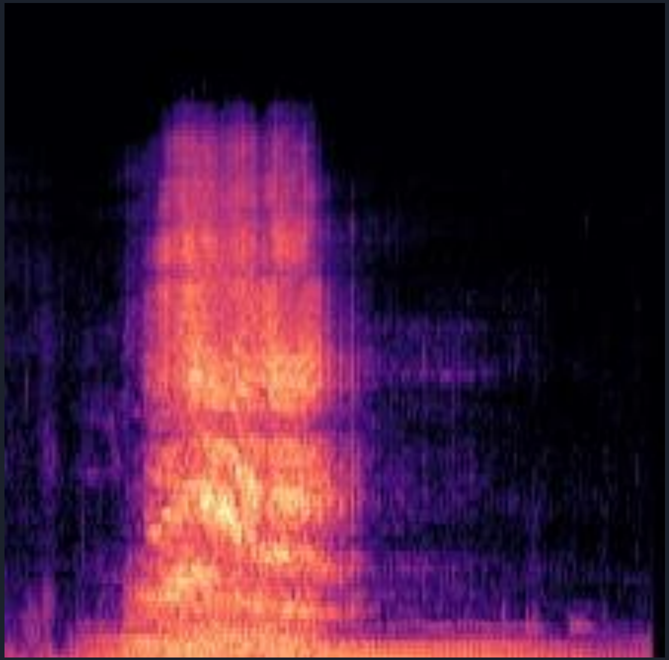
1. Convert the train/test audio files into train/test spectrogram image files
2. Convert image pixels to NumPy arrays and then normalize the arrays
3. Create the Convolutional Base
4. Compile and Train the model
5. Evaluate the model

Discussion

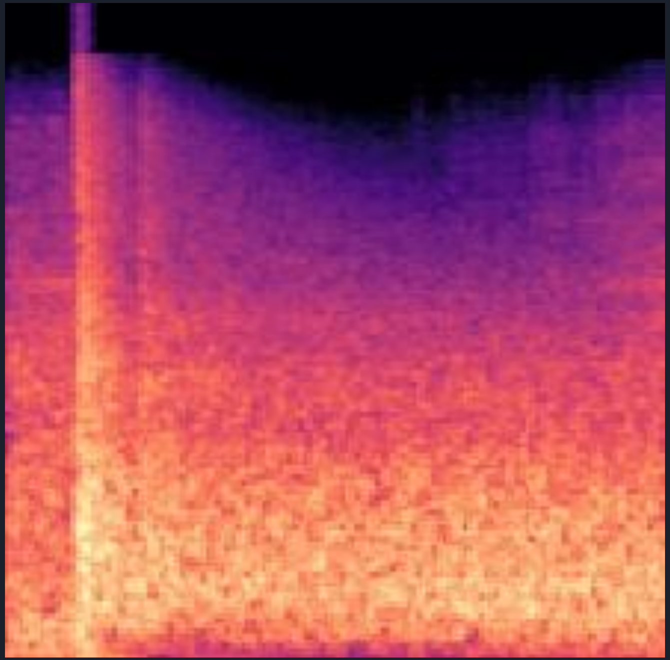
Approach:

- Audio files ----> Spectrogram images ----> pixel arrays ----> Normalization ----> Train/Evaluate model
- Spectrograms:
  - Visualize frequencies present in audio recordings
  - Frequency patterns vary for each class of sound
  - Classify images based on the frequencies

Elephant sound

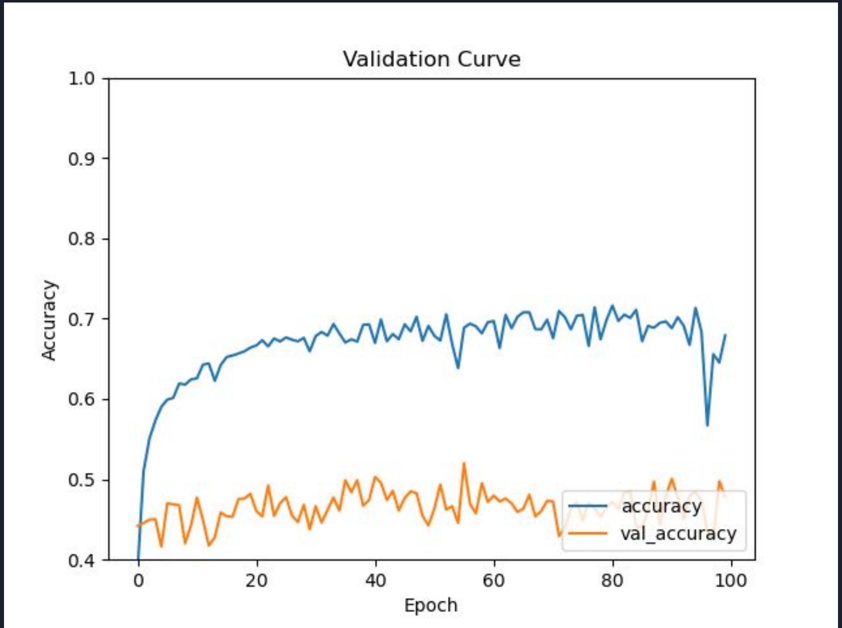


Gunshot sound

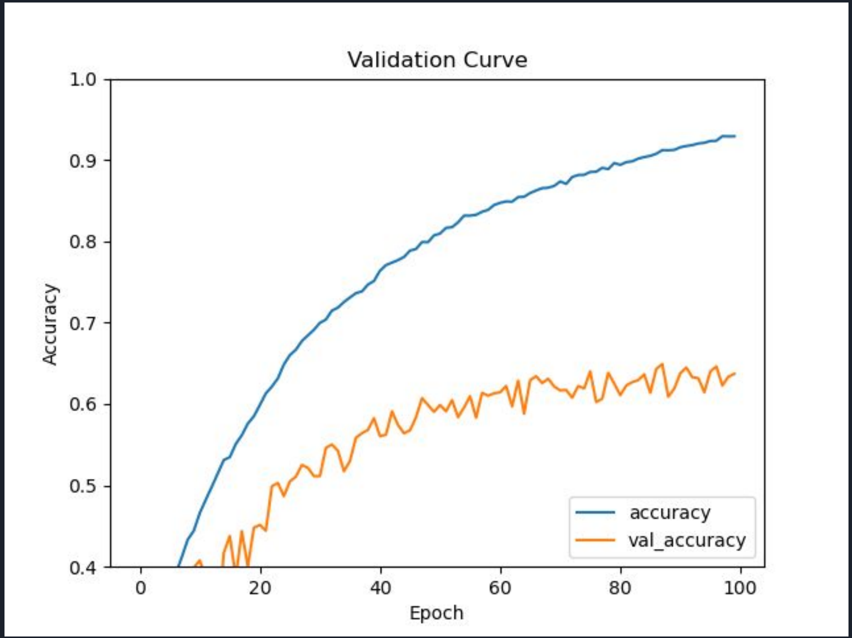


Discussion - Hyperparameter tuning

Adam Optimizer



SGD Optimizer



Adam

================================

------MODEL PREDICTIONS---------

================================

	Filename	Actual_Result	Model_Predicted_Result
0	135526-6-0-0.wav	gun_shot	gun_shot
1	135526-6-1-0.wav	gun_shot	street_music
2	135526-6-2-0.wav	gun_shot	gun_shot
3	135526-6-3-0.wav	gun_shot	gun_shot
4	135526-6-4-0.wav	gun_shot	gun_shot
5	135526-6-5-0.wav	gun_shot	gun_shot
6	135526-6-6-0.wav	gun_shot	gun_shot
7	135526-6-7-0.wav	gun_shot	gun_shot
8	135526-6-8-0.wav	gun_shot	gun_shot
9	135526-6-9-0.wav	gun_shot	street_music
10	203929-7-3-10.wav	jackhammer	jackhammer
11	203929-7-3-2.wav	jackhammer	jackhammer
12	203929-7-3-3.wav	jackhammer	jackhammer
13	203929-7-3-5.wav	jackhammer	jackhammer
14	203929-7-3-9.wav	jackhammer	jackhammer
15	344-3-0-0.wav	dog_bark	dog_bark
16	344-3-1-0.wav	dog_bark	dog_bark
17	344-3-4-0.wav	dog_bark	dog_bark
18	344-3-5-0.wav	dog_bark	dog_bark
19	518-4-0-0.wav	drilling	drilling
20	518-4-0-1.wav	drilling	drilling
21	518-4-0-2.wav	drilling	jackhammer
22	518-4-0-3.wav	drilling	jackhammer
23	Elephant.1.wav	Elephant	street_music
24	Elephant.10.wav	Elephant	air_conditioner
25	Elephant.11.wav	Elephant	Elephant
26	Elephant.12.wav	Elephant	street_music
27	Elephant.13.wav	Elephant	children_playing
28	Elephant.14.wav	Elephant	dog_bark
29	Elephant.15.wav	Elephant	air_conditioner
30	Elephant.16.wav	Elephant	air_conditioner
31	Elephant.17.wav	Elephant	air_conditioner
32	Elephant.18.wav	Elephant	air_conditioner
33	Elephant.19.wav	Elephant	Elephant
34	Elephant.2.wav	Elephant	street_music
35	Elephant.20.wav	Elephant	engine_idling
36	Elephant.27.wav	Elephant	street_music
37	Elephant.3.wav	Elephant	street_music

This model has predicted 21/38 correctly, which is 55.26%

Time of execution: 527.9309256076813

@ 100 Epochs

SGD

================================

------MODEL PREDICTIONS---------

================================

	Filename	Actual_Result	Model_Predicted_Result
0	135526-6-0-0.wav	gun_shot	gun_shot
1	135526-6-1-0.wav	gun_shot	gun_shot
2	135526-6-2-0.wav	gun_shot	gun_shot
3	135526-6-3-0.wav	gun_shot	gun_shot
4	135526-6-4-0.wav	gun_shot	gun_shot
5	135526-6-5-0.wav	gun_shot	gun_shot
6	135526-6-6-0.wav	gun_shot	gun_shot
7	135526-6-7-0.wav	gun_shot	gun_shot
8	135526-6-8-0.wav	gun_shot	gun_shot
9	135526-6-9-0.wav	gun_shot	gun_shot
10	203929-7-3-10.wav	jackhammer	jackhammer
11	203929-7-3-2.wav	jackhammer	jackhammer
12	203929-7-3-3.wav	jackhammer	jackhammer
13	203929-7-3-5.wav	jackhammer	jackhammer
14	203929-7-3-9.wav	jackhammer	jackhammer
15	344-3-0-0.wav	dog_bark	dog_bark
16	344-3-1-0.wav	dog_bark	dog_bark
17	344-3-4-0.wav	dog_bark	dog_bark
18	344-3-5-0.wav	dog_bark	dog_bark
19	518-4-0-0.wav	drilling	drilling
20	518-4-0-1.wav	drilling	drilling
21	518-4-0-2.wav	drilling	drilling
22	518-4-0-3.wav	drilling	drilling
23	Elephant.1.wav	Elephant	Elephant
24	Elephant.10.wav	Elephant	drilling
25	Elephant.11.wav	Elephant	Elephant
26	Elephant.12.wav	Elephant	Elephant
27	Elephant.13.wav	Elephant	Elephant
28	Elephant.14.wav	Elephant	Elephant
29	Elephant.15.wav	Elephant	Elephant
30	Elephant.16.wav	Elephant	Elephant
31	Elephant.17.wav	Elephant	Elephant
32	Elephant.18.wav	Elephant	Elephant
33	Elephant.19.wav	Elephant	Elephant
34	Elephant.2.wav	Elephant	Elephant
35	Elephant.20.wav	Elephant	Elephant
36	Elephant.27.wav	Elephant	Elephant
37	Elephant.3.wav	Elephant	Elephant

This model has predicted 37/38 correctly, which is 97.37%

Time of execution: 550.3758449554443

Epoch size = 100

================================
------MODEL PREDICTIONS---------
================================

	Filename	Actual_Result	Model_Predicted_Result
0	135526-6-0-0.wav	gun_shot	gun_shot
1	135526-6-1-0.wav	gun_shot	gun_shot
2	135526-6-2-0.wav	gun_shot	gun_shot
3	135526-6-3-0.wav	gun_shot	gun_shot
4	135526-6-4-0.wav	gun_shot	gun_shot
5	135526-6-5-0.wav	gun_shot	gun_shot
6	135526-6-6-0.wav	gun_shot	gun_shot
7	135526-6-7-0.wav	gun_shot	gun_shot
8	135526-6-8-0.wav	gun_shot	gun_shot
9	135526-6-9-0.wav	gun_shot	gun_shot
10	203929-7-3-10.wav	jackhammer	jackhammer
11	203929-7-3-2.wav	jackhammer	jackhammer
12	203929-7-3-3.wav	jackhammer	jackhammer
13	203929-7-3-5.wav	jackhammer	jackhammer
14	203929-7-3-9.wav	jackhammer	jackhammer
15	344-3-0-0.wav	dog_bark	dog_bark
16	344-3-1-0.wav	dog_bark	dog_bark
17	344-3-4-0.wav	dog_bark	dog_bark
18	344-3-5-0.wav	dog_bark	dog_bark
19	518-4-0-0.wav	drilling	drilling
20	518-4-0-1.wav	drilling	drilling
21	518-4-0-2.wav	drilling	drilling
22	518-4-0-3.wav	drilling	drilling
23	Elephant.1.wav	Elephant	Elephant
24	Elephant.10.wav	Elephant	drilling
25	Elephant.11.wav	Elephant	Elephant
26	Elephant.12.wav	Elephant	Elephant
27	Elephant.13.wav	Elephant	Elephant
28	Elephant.14.wav	Elephant	Elephant
29	Elephant.15.wav	Elephant	Elephant
30	Elephant.16.wav	Elephant	Elephant
31	Elephant.17.wav	Elephant	Elephant
32	Elephant.18.wav	Elephant	Elephant
33	Elephant.19.wav	Elephant	Elephant
34	Elephant.2.wav	Elephant	Elephant
35	Elephant.20.wav	Elephant	Elephant
36	Elephant.27.wav	Elephant	Elephant
37	Elephant.3.wav	Elephant	Elephant

This model has predicted 37/38 correctly, which is 97.37%
Time of execution: 550.3758449554443

Epoch size = 50

================================
------MODEL PREDICTIONS---------
================================

	Filename	Actual_Result	Model_Predicted_Result
0	135526-6-0-0.wav	gun_shot	gun_shot
1	135526-6-1-0.wav	gun_shot	gun_shot
2	135526-6-2-0.wav	gun_shot	gun_shot
3	135526-6-3-0.wav	gun_shot	gun_shot
4	135526-6-4-0.wav	gun_shot	gun_shot
5	135526-6-5-0.wav	gun_shot	gun_shot
6	135526-6-6-0.wav	gun_shot	gun_shot
7	135526-6-7-0.wav	gun_shot	gun_shot
8	135526-6-8-0.wav	gun_shot	gun_shot
9	135526-6-9-0.wav	gun_shot	gun_shot
10	203929-7-3-10.wav	jackhammer	jackhammer
11	203929-7-3-2.wav	jackhammer	jackhammer
12	203929-7-3-3.wav	jackhammer	jackhammer
13	203929-7-3-5.wav	jackhammer	jackhammer
14	203929-7-3-9.wav	jackhammer	jackhammer
15	344-3-0-0.wav	dog_bark	dog_bark
16	344-3-1-0.wav	dog_bark	dog_bark
17	344-3-4-0.wav	dog_bark	dog_bark
18	344-3-5-0.wav	dog_bark	dog_bark
19	518-4-0-0.wav	drilling	jackhammer
20	518-4-0-1.wav	drilling	engine_idling
21	518-4-0-2.wav	drilling	engine_idling
22	518-4-0-3.wav	drilling	jackhammer
23	Elephant.1.wav	Elephant	Elephant
24	Elephant.10.wav	Elephant	air_conditioned
25	Elephant.11.wav	Elephant	Elephant
26	Elephant.12.wav	Elephant	Elephant
27	Elephant.13.wav	Elephant	dog_bark
28	Elephant.14.wav	Elephant	Elephant
29	Elephant.15.wav	Elephant	Elephant
30	Elephant.16.wav	Elephant	Elephant
31	Elephant.17.wav	Elephant	Elephant
32	Elephant.18.wav	Elephant	Elephant
33	Elephant.19.wav	Elephant	Elephant
34	Elephant.2.wav	Elephant	Elephant
35	Elephant.20.wav	Elephant	dog_bark
36	Elephant.27.wav	Elephant	Elephant
37	Elephant.3.wav	Elephant	Elephant

This model has predicted 31/38 correctly, which is 81.58%
Time of execution: 252.86861968040466

Epoch size = 10

================================
------MODEL PREDICTIONS---------
================================

	Filename	Actual_Result	Model_Predicted_Result
0	135526-6-0-0.wav	gun_shot	gun_shot
1	135526-6-1-0.wav	gun_shot	street_music
2	135526-6-2-0.wav	gun_shot	jackhammer
3	135526-6-3-0.wav	gun_shot	gun_shot
4	135526-6-4-0.wav	gun_shot	gun_shot
5	135526-6-5-0.wav	gun_shot	jackhammer
6	135526-6-6-0.wav	gun_shot	gun_shot
7	135526-6-7-0.wav	gun_shot	street_music
8	135526-6-8-0.wav	gun_shot	street_music
9	135526-6-9-0.wav	gun_shot	street_music
10	203929-7-3-10.wav	jackhammer	jackhammer
11	203929-7-3-2.wav	jackhammer	jackhammer
12	203929-7-3-3.wav	jackhammer	jackhammer
13	203929-7-3-5.wav	jackhammer	jackhammer
14	203929-7-3-9.wav	jackhammer	jackhammer
15	344-3-0-0.wav	dog_bark	dog_bark
16	344-3-1-0.wav	dog_bark	dog_bark
17	344-3-4-0.wav	dog_bark	dog_bark
18	344-3-5-0.wav	dog_bark	dog_bark
19	518-4-0-0.wav	drilling	jackhammer
20	518-4-0-1.wav	drilling	jackhammer
21	518-4-0-2.wav	drilling	jackhammer
22	518-4-0-3.wav	drilling	jackhammer
23	Elephant.1.wav	Elephant	children_playing
24	Elephant.10.wav	Elephant	air_conditioner
25	Elephant.11.wav	Elephant	dog_bark
26	Elephant.12.wav	Elephant	dog_bark
27	Elephant.13.wav	Elephant	dog_bark
28	Elephant.14.wav	Elephant	dog_bark
29	Elephant.15.wav	Elephant	dog_bark
30	Elephant.16.wav	Elephant	children_playing
31	Elephant.17.wav	Elephant	engine_idling
32	Elephant.18.wav	Elephant	engine_idling
33	Elephant.19.wav	Elephant	dog_bark
34	Elephant.2.wav	Elephant	drilling
35	Elephant.20.wav	Elephant	dog_bark
36	Elephant.27.wav	Elephant	drilling
37	Elephant.3.wav	Elephant	drilling

This model has predicted 13/38 correctly, which is 34.21%
Time of execution: 73.53798031806946

Model optimization

- Dropout is not necessary.

- SGD has better generalization performance than ADAM. *SGD usually improves model performance slowly but could achieve higher test performance."*

- The sounds are quite different so the model is quite general. If the sounds were more identical to each other, ADAM would be the best option. Since it could quickly get out of the local minimums.

- The model learns fast. It is not necessary to use a small learning rate.

Optimizer	Epochs	Dropout	Accuracy
SGD	100	No	97%
SGD	200	No	100%
SGD	50	No	82%
SGD	50	Yes	74%
SGD	100	Yes	76%

Source: https://proceedings.neurips.cc//paper/2020/file/f3f27a324736617f20abbf2ffd806f6d-Paper.pdf

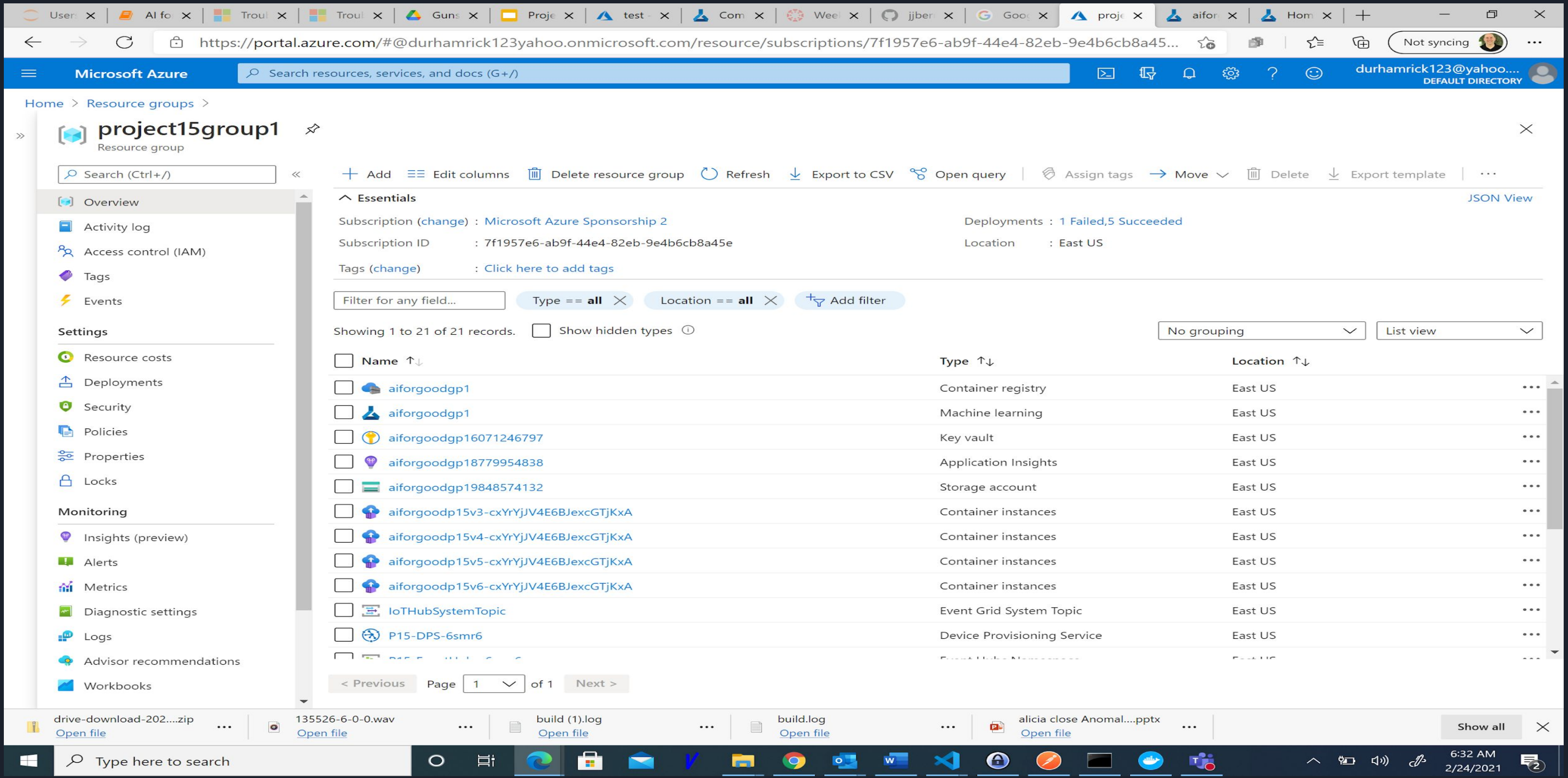
Discussion

Challenges:

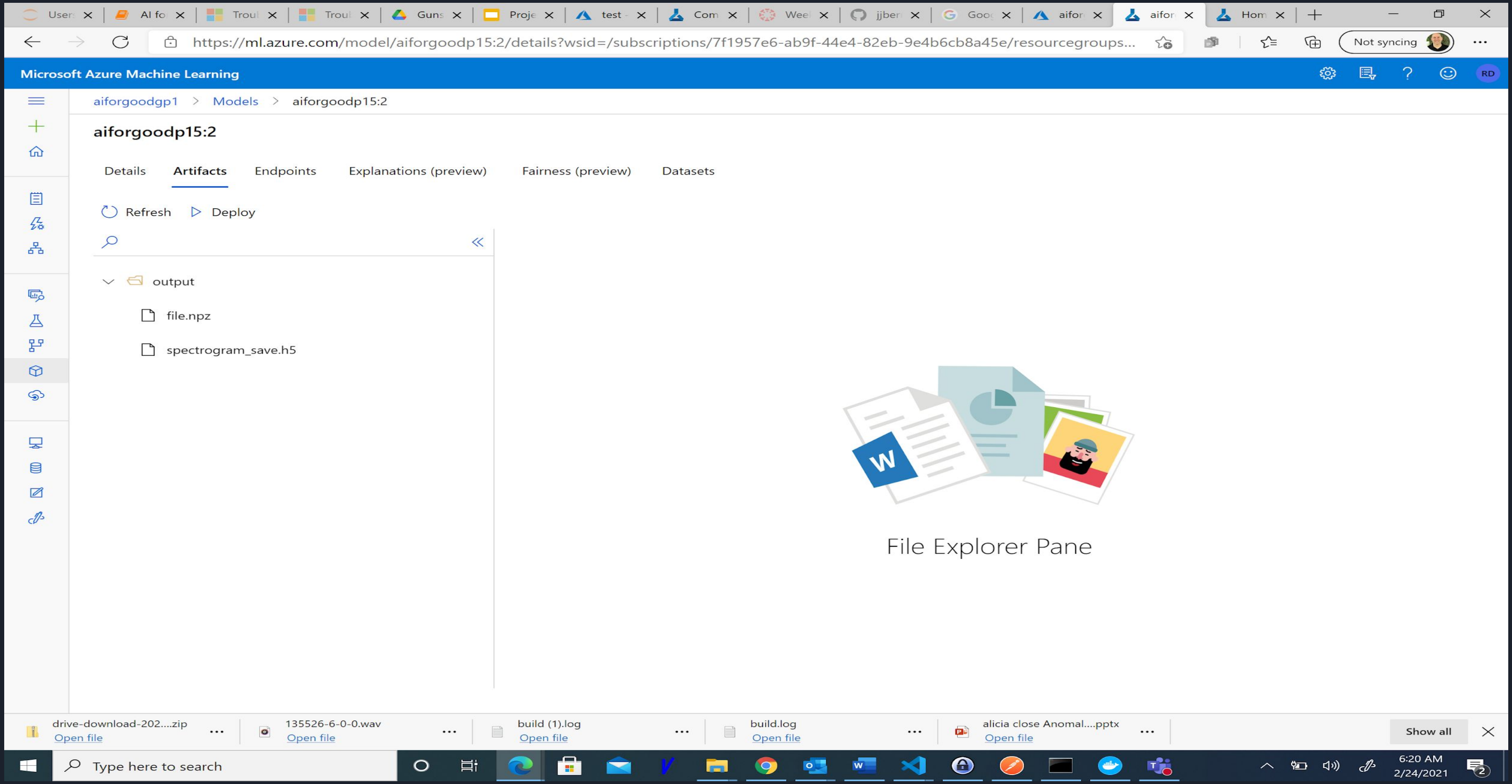
- Imbalance in dataset. Owing to smaller sample size of elephant audio files
- File size of dataset too large

Build deployment environment, scripts and edge devices

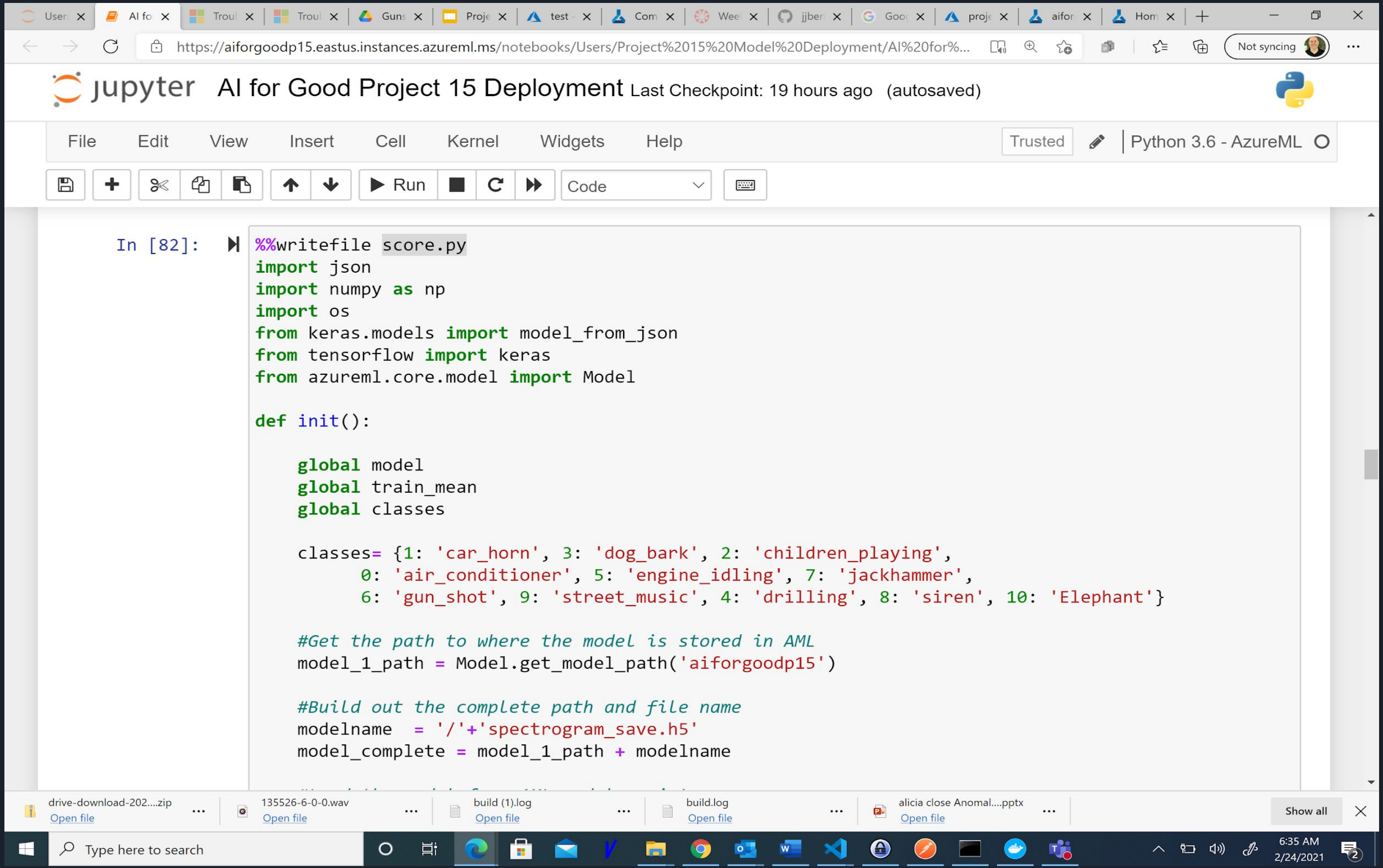
Creation of the Azure ML/AI deployment environment to support Project 15



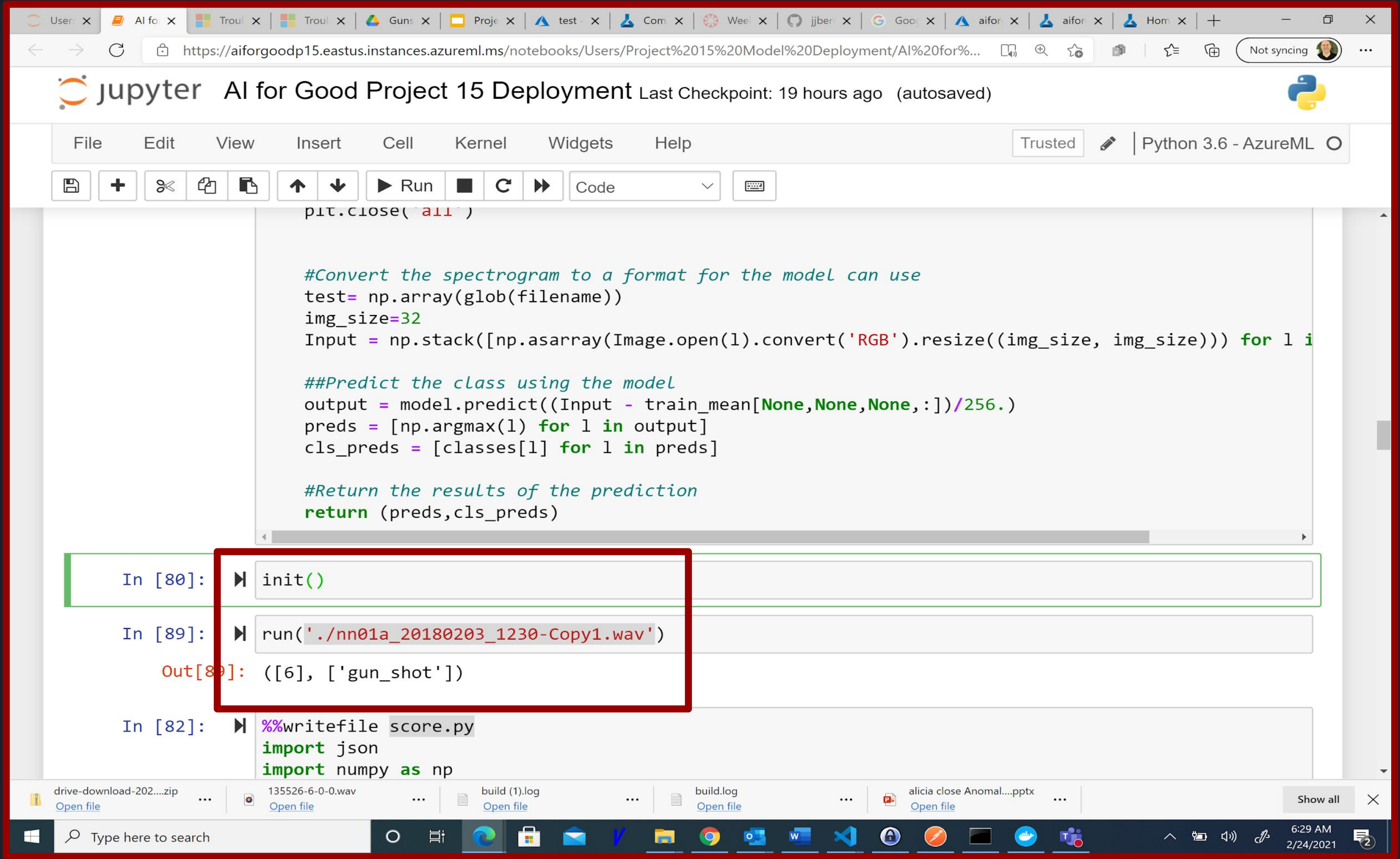
(Creation of Azure Azure Machine Learning Environment and Registration of the model)



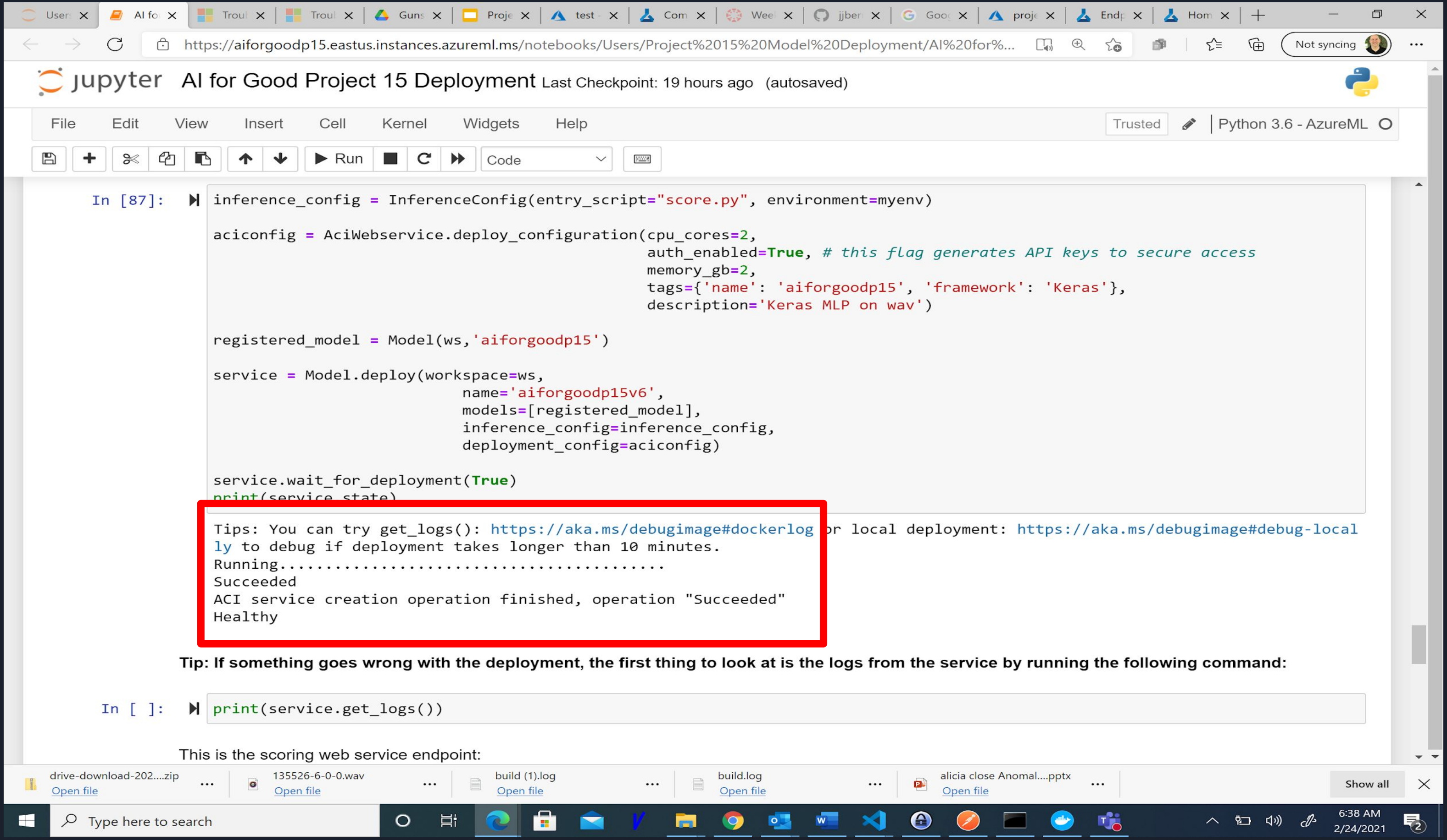
(Creation of Scory.py file needed for deployment)



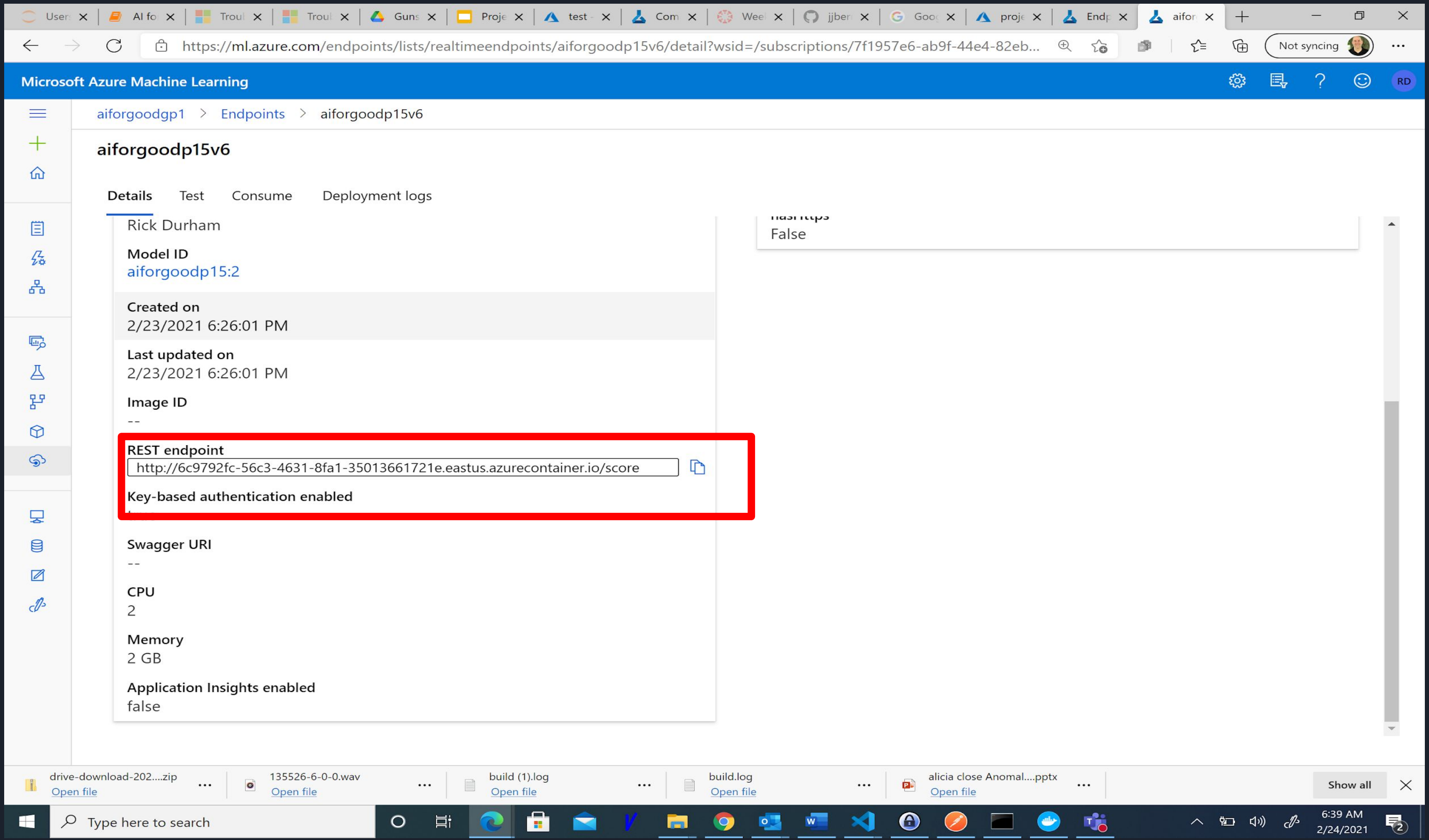
(Create Scory.py model inference results test scoring with real gunshot)



(Deployment of the Model to Azure ACI)

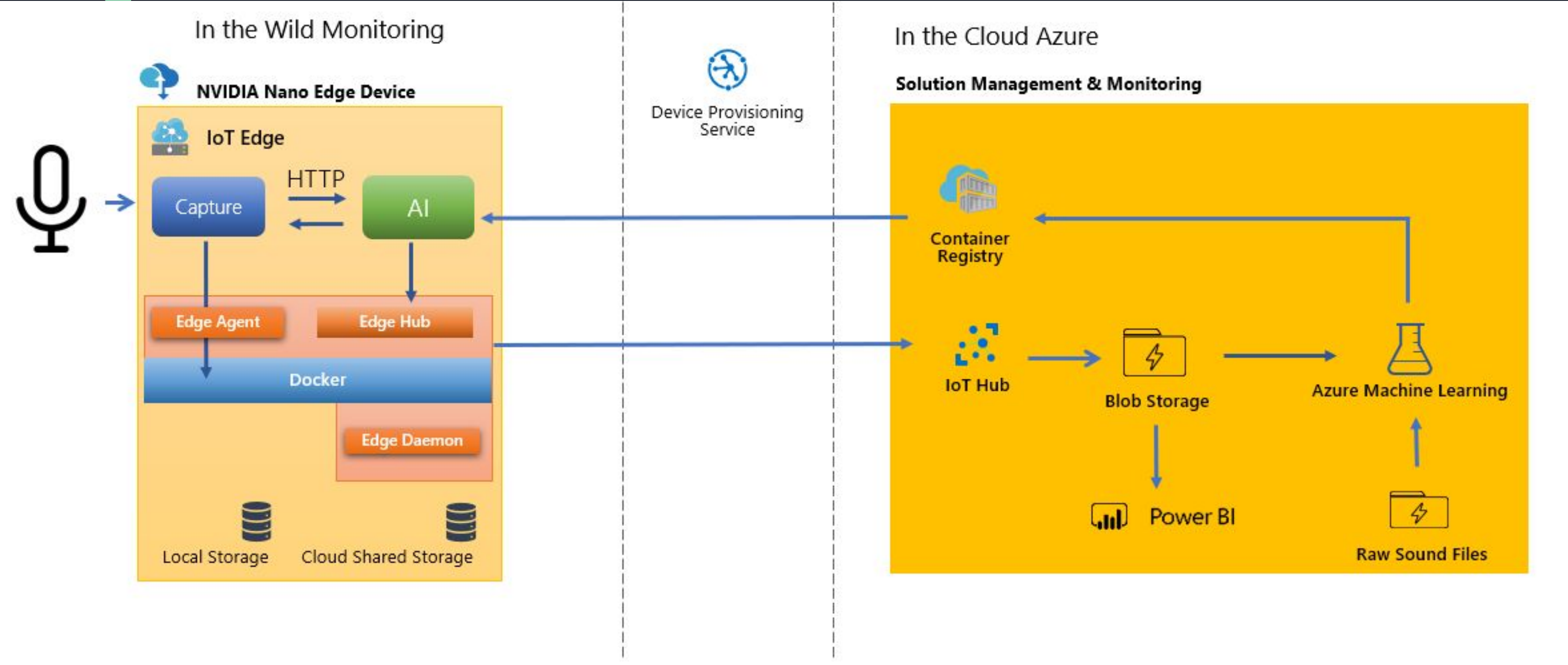


(Deployment of the Model to Azure ACI)

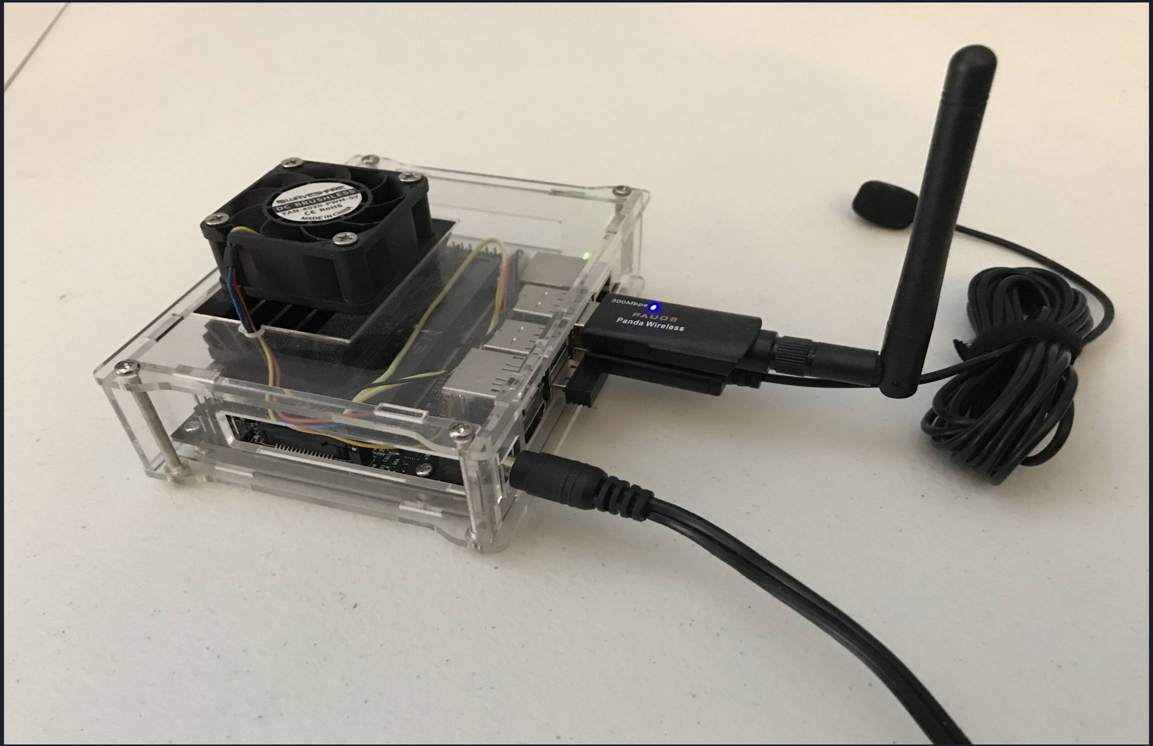


Deploy preparation of IoT Edge Device

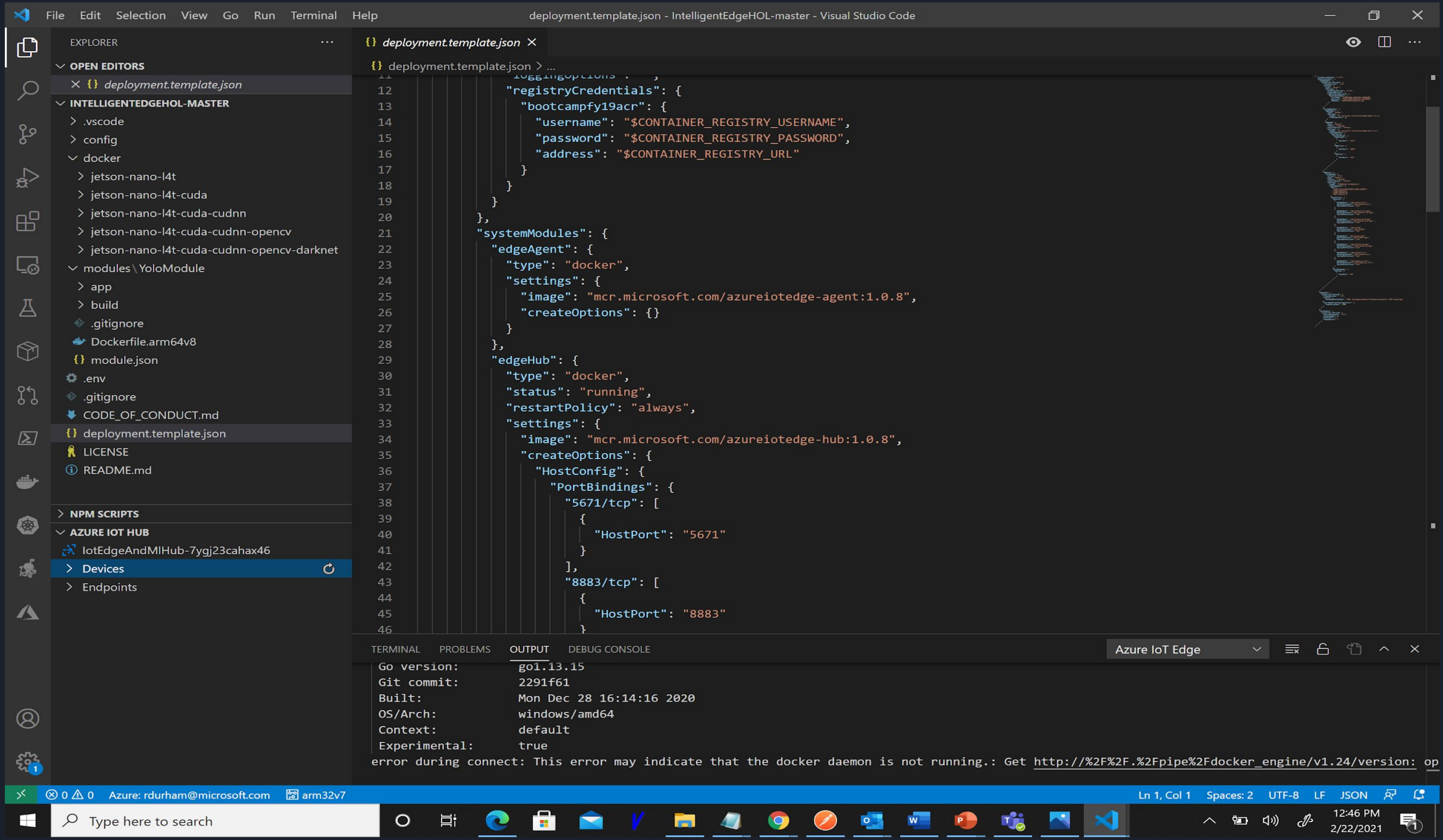
Proposed Cloud/Edge Design Architecture



The following device was created for Project 15 based on the proposed architecture using a NVIDIA (Nano)



The following VSC Deployment Environment code was prepared for the - NVIDIA (Nano)



Conclusion and next steps...

Future

- Gather more data to support improved models in terms of accuracy and samples that more represent the environment where elephants live and are poached
- Complete testing of ACI model deployment using REST calls
- Download deployment Docker container on Edge device and test using real gunshot and elephant sounds
- Classify elephant sounds based on demographics
- Combine gunshots with ambient sounds (Fourier transform)
- Eventually rework the way we inference so that a spectrogram file does not have to be created and ingested by the Keras model. This would be more of an in memory operation.

Reference Documentation

Literature Review:

The following documents were reviewed and respective repos used on the project:

Project 15 from Microsoft | project15

2004.07948.pdf (arxiv.org)

Urban Sound Classification — Part 1: sound wave, digital audio signal | by Ricky Kim | Towards Data ScienceUrban Sound

Urban Sound Classification — Part 2: sample rate conversion, Librosa | by Ricky Kim | Towards Data Science

Classification using Neural Networks | by Shubham Gupta | Towards Data Science

Our Code: https://github.com/Oxford-AI-Edge-2020-2021-Group-1/AI4GoodP15