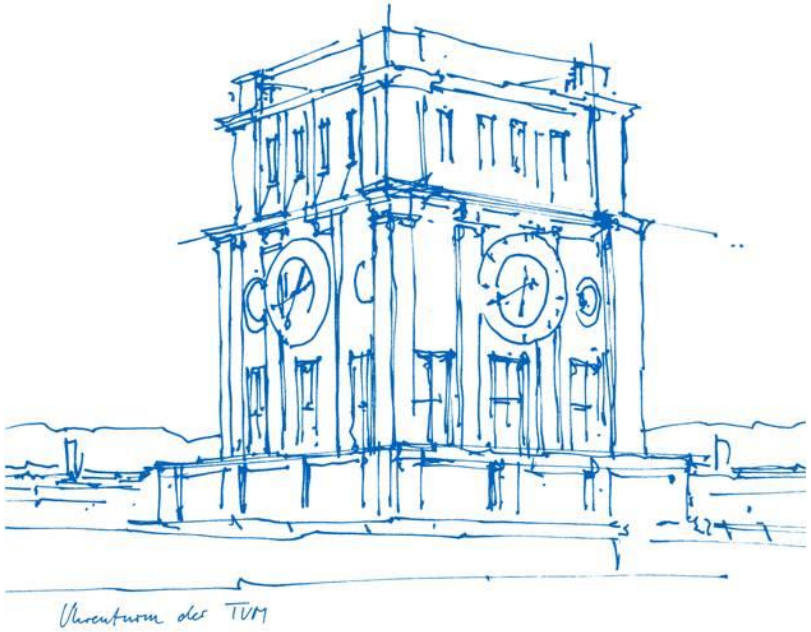
BirdNet

Ahmed Kaddah, Shao Jie Hu Chen, Marlon Müller

Edge Computing and the Internet of Things

Technische Universität München

Munich, 02.02.2024



Introduction

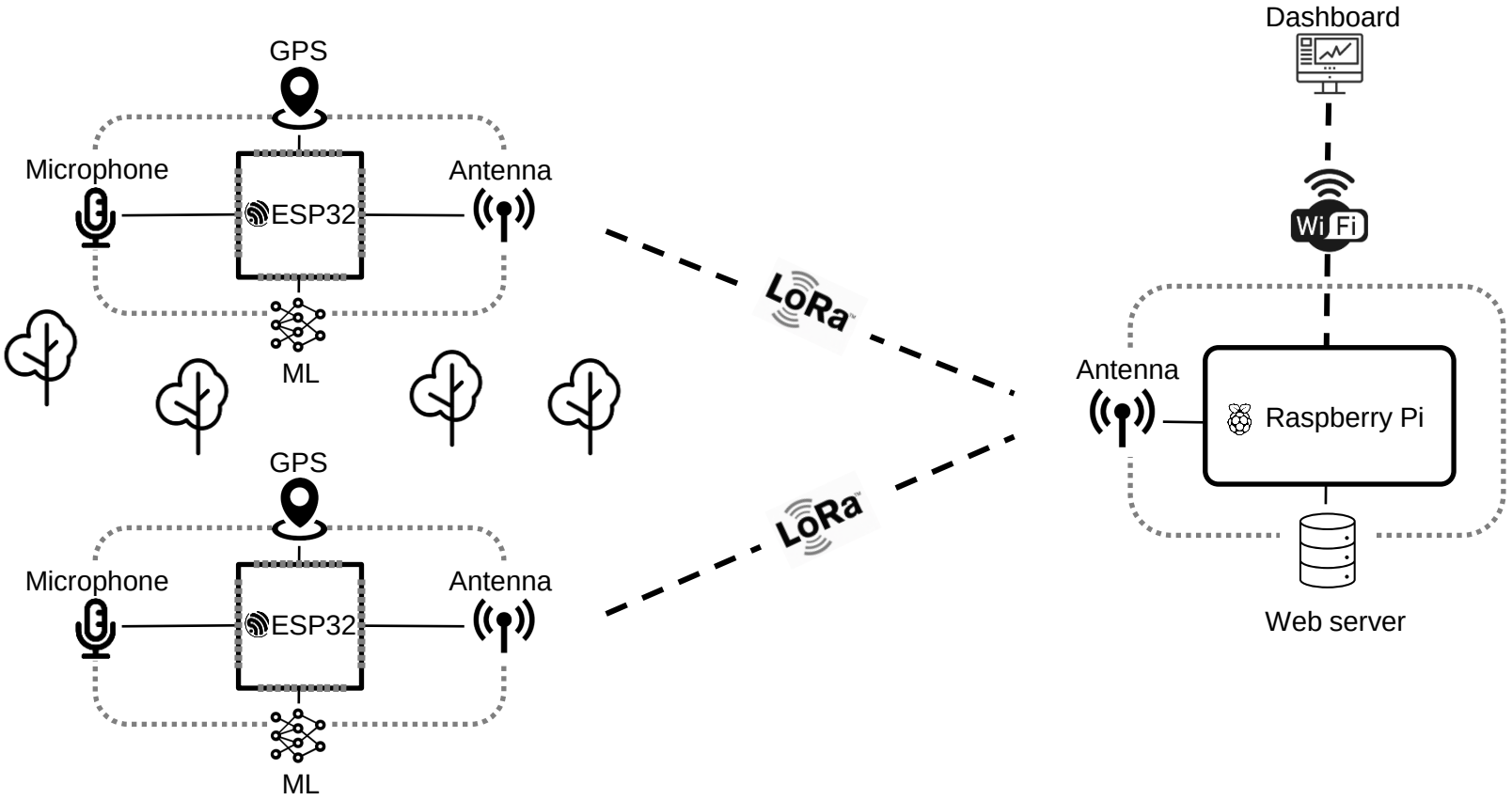
Motivation

- Casual interest: tourism, hiking …
- Reduces human intervention
- Enables real-time monitoring
- Studying ecosystem health
- Combines edge computing and IOT

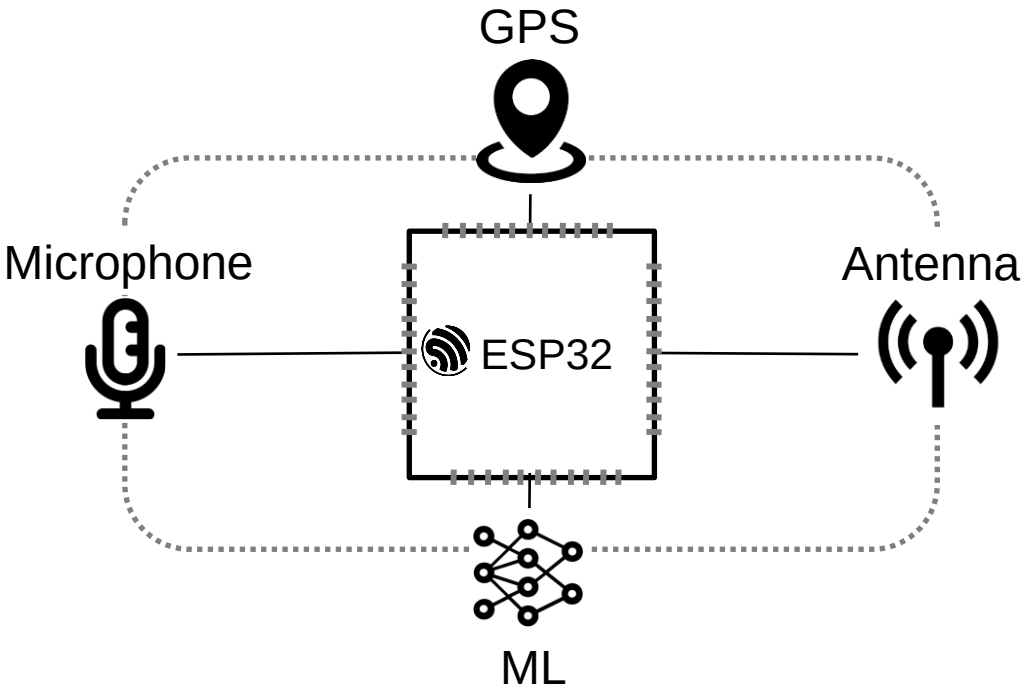
Objective

- ESP32s equipped with microphones
- On-device classification of bird species
- Raspberry Pi base station
- Statistics accessible via dashboard

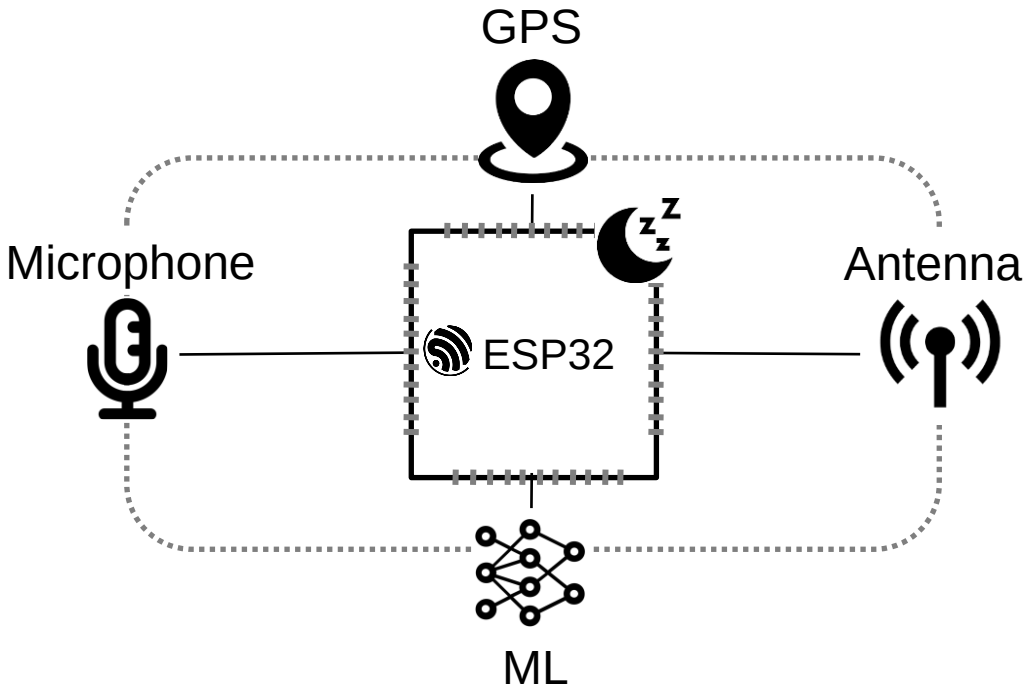
Architecture



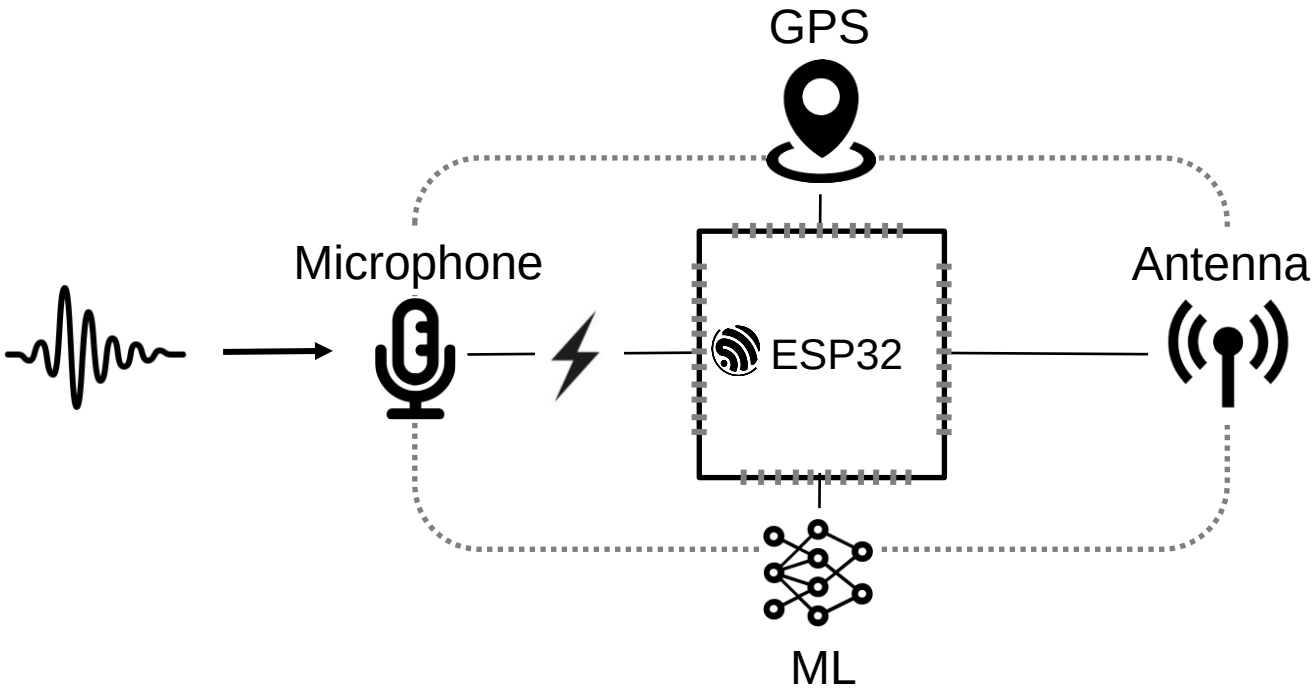
ESP32-S3



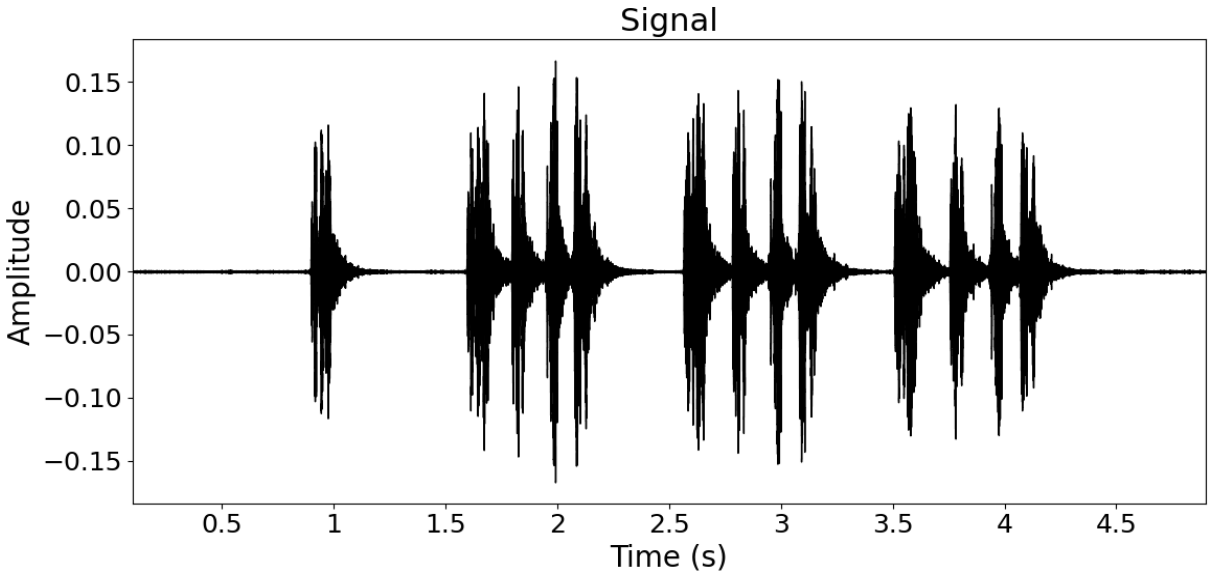
Deep sleep



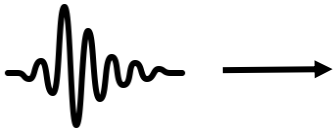
Microphone interrupt

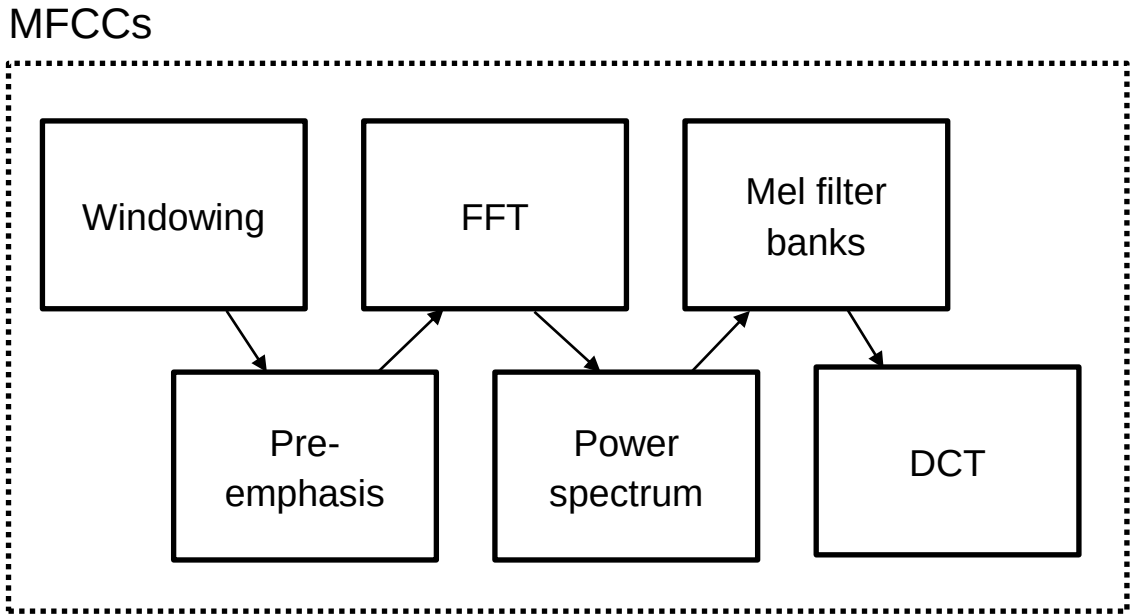


Raw audio

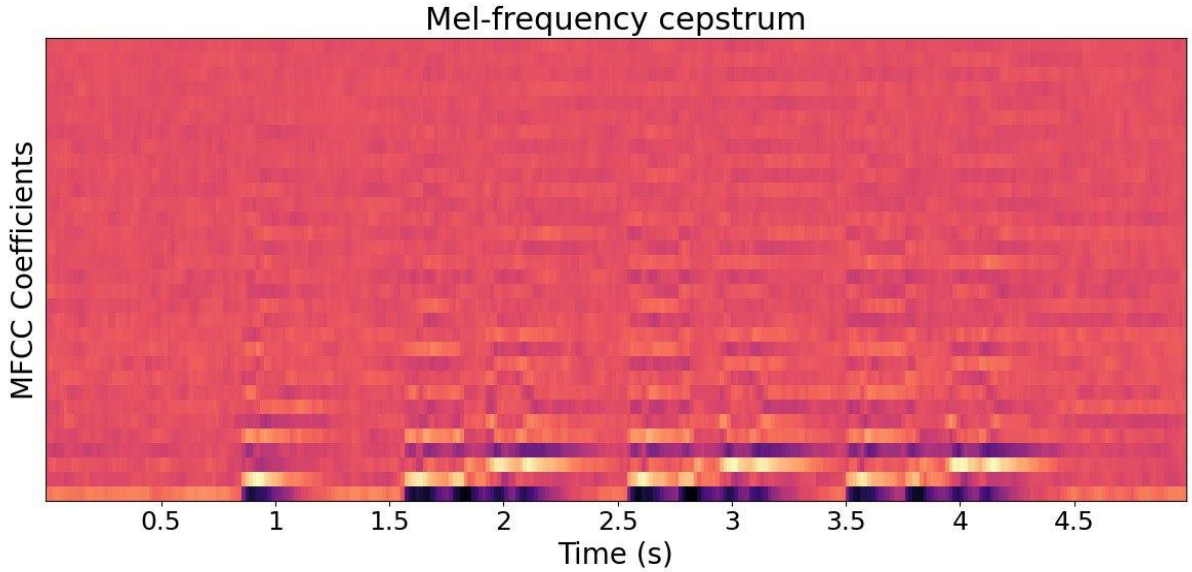

Signal
Amplitude
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
0.5
1
1.5
2
2.5
3
3.5
4
4.5
Time (s)

Preprocessing

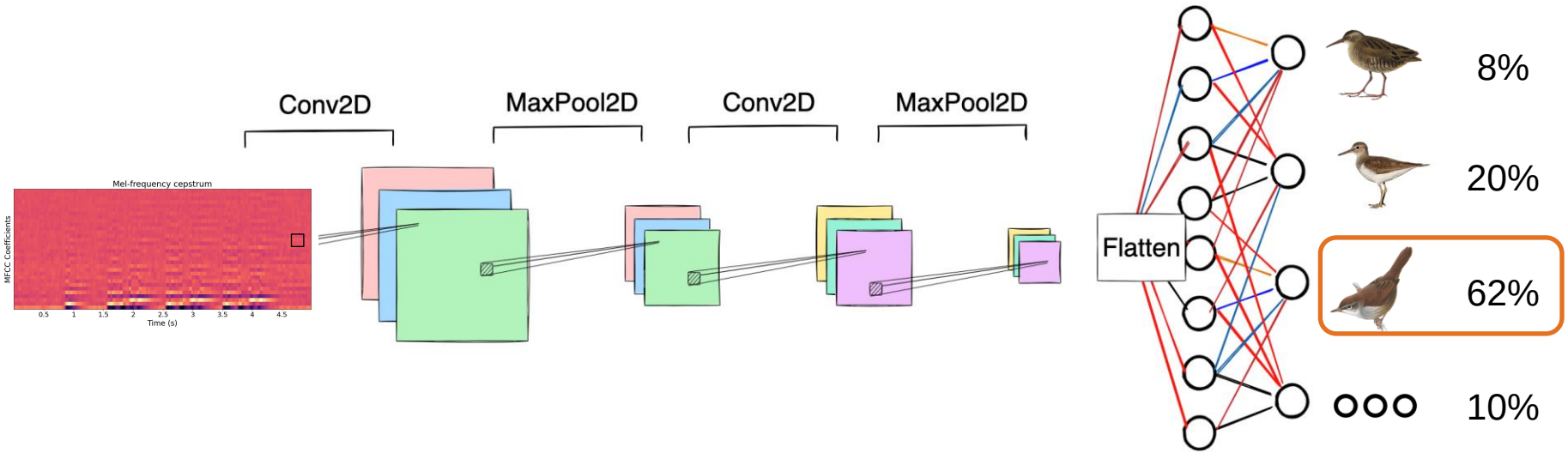




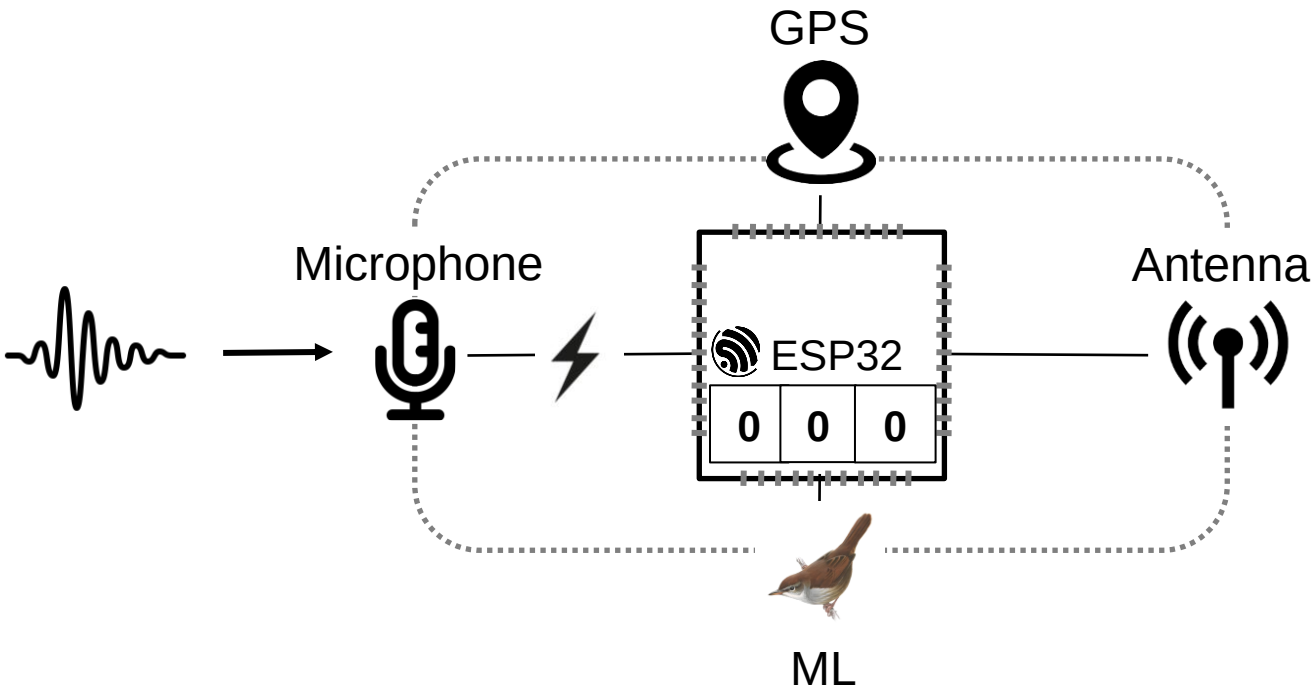
Preprocessing



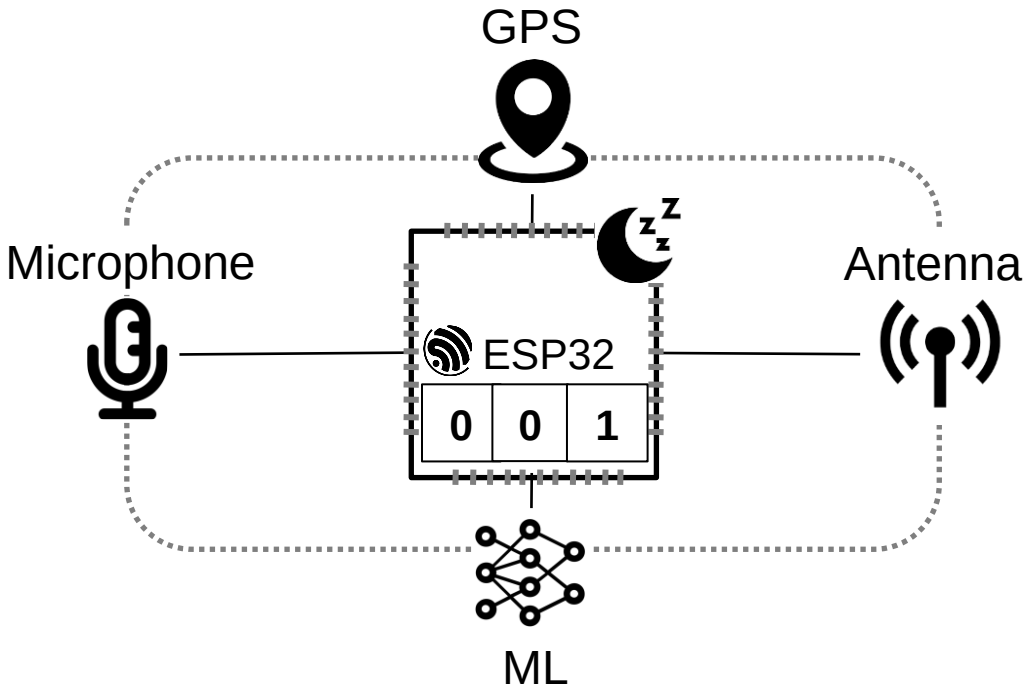
Convolutional Neural Network



RTC memory

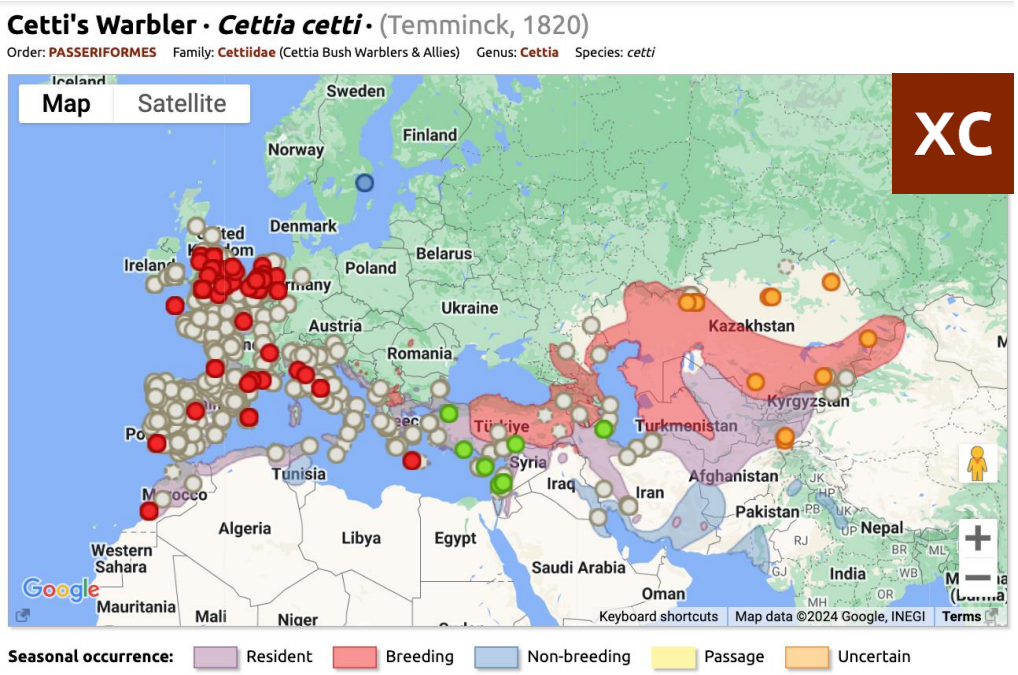


RTC memory



Xeno-canto

"Xeno-canto is a website dedicated to sharing wildlife sounds from all over the world"

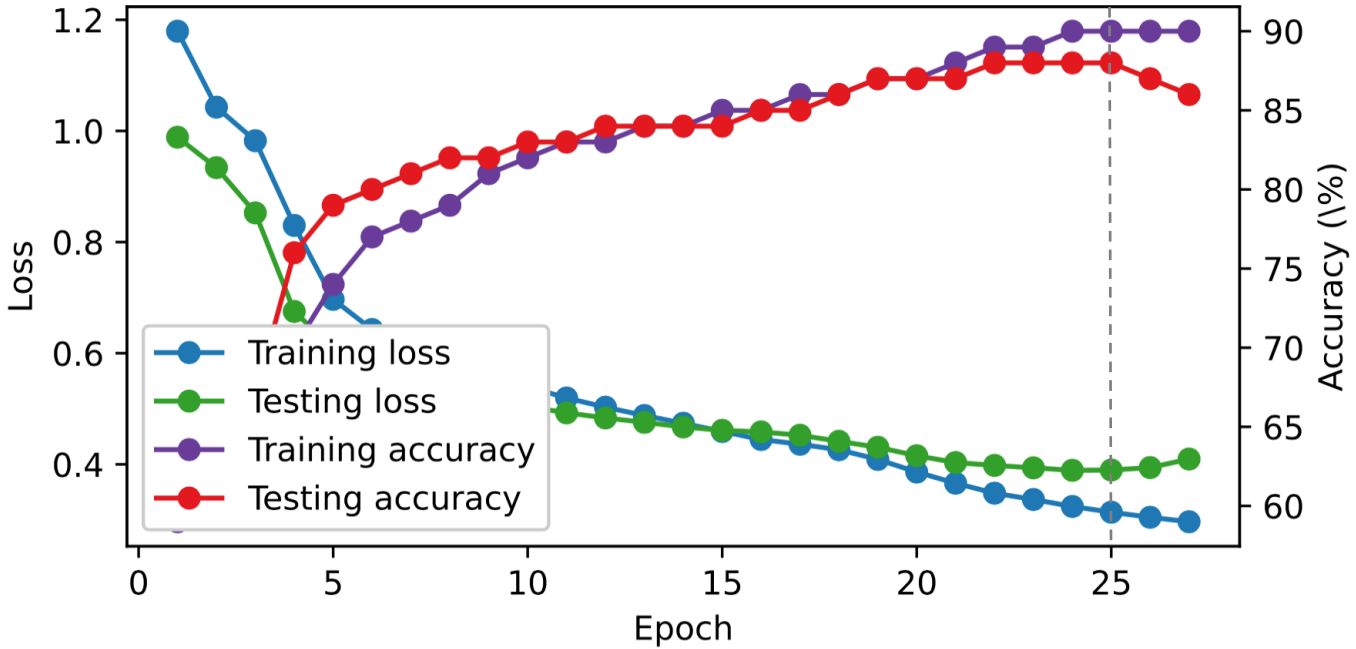


ESC-50

"Dataset for environmental sound classification"




Training


1.2
1.0
0.8
0.6
0.4
Loss
90
85
80
75
70
65
60
Accuracy (\%)
0
5
10
15
20
25
Epoch
Training loss
Testing loss
Training accuracy
Testing accuracy

Quantization


```
const static __attribute__((aligned(16))) int16_t _co
    -1782,   1660,   1449,   4278,   -995,   1297,
    -1205,   1112,   1364,  11567,   4035,   1949,
     1322,  -1022,   3667,   3183,  -1290,  -1773,
    -1139,   7841,  -1045,   3881,  -1795,   -251,
     1661,  -1897,   2759,   1978,  -2340,   4560,
```

Memory

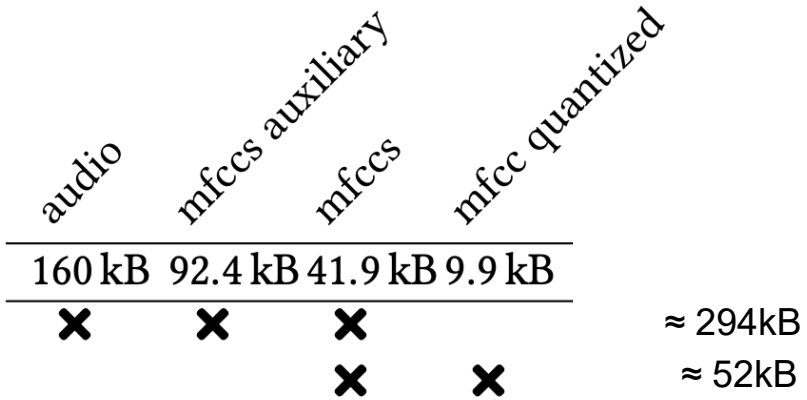
$$2\text{B} \times 5\text{s} \times 16\text{kHz}$$

$$\frac{\text{audio}}{160\,\text{kB}}$$

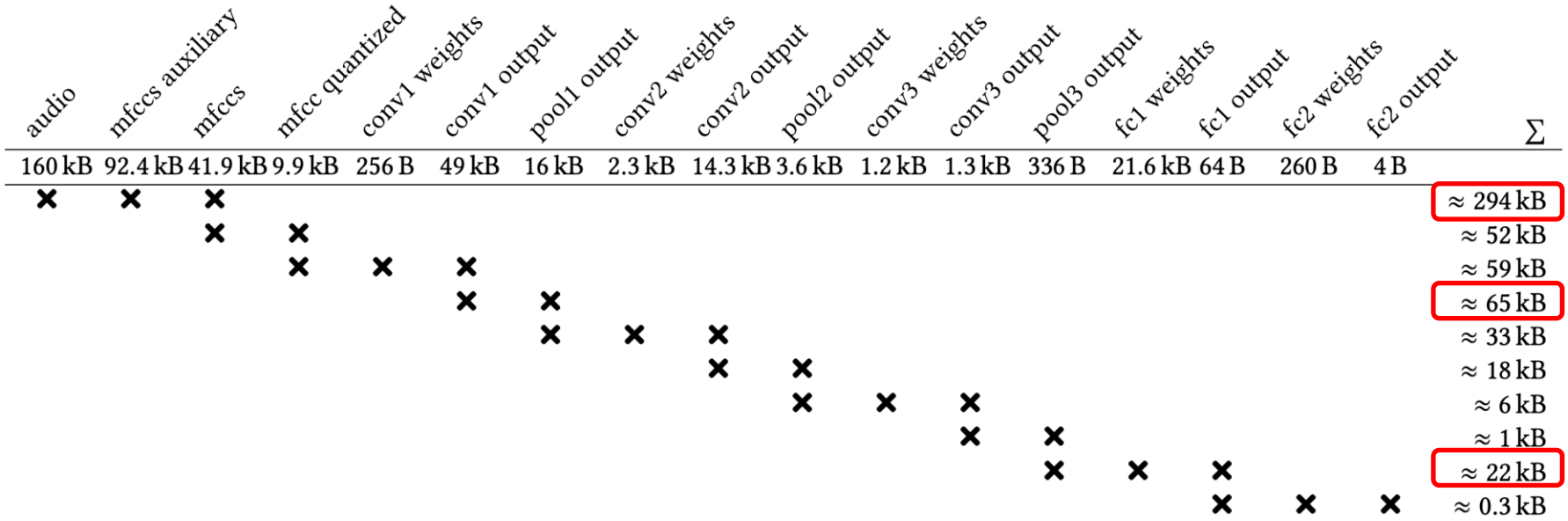
Memory

audio	mfccs auxiliary	mfccs	
160 kB	92.4 kB	41.9 kB	
✗	✗	✗	≈ 294kB

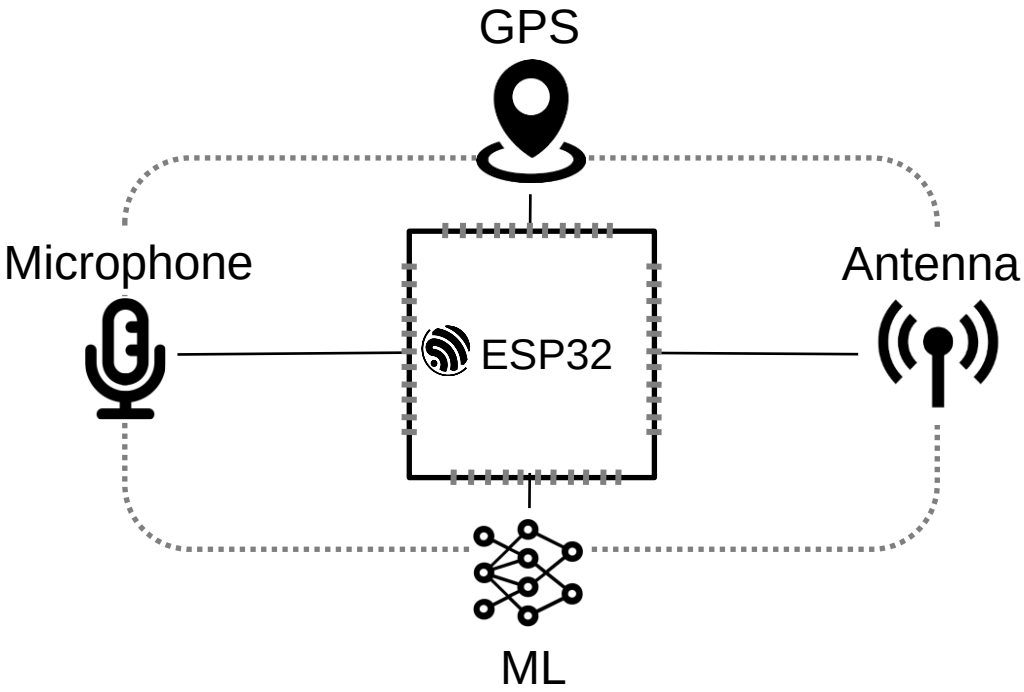
Memory


audio
mfccs auxiliary
mfccs
mfcc quantized
160 kB
92.4 kB
41.9 kB
9.9 kB
≈ 294kB
≈ 52kB

Memory

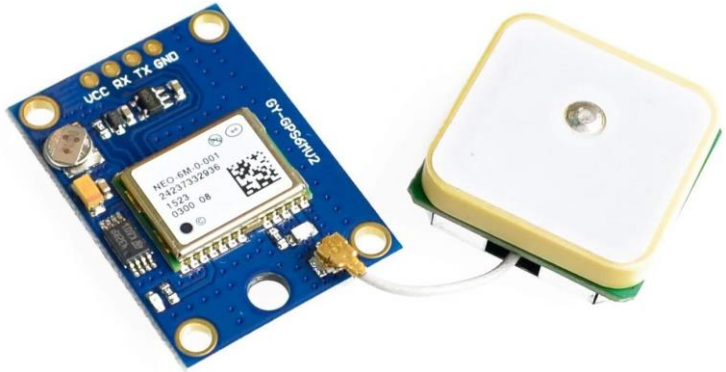

audio
mfccs auxiliary
mfccs
mfcc quantized
conv1 weights
conv1 output
pool1 output
conv2 weights
conv2 output
pool2 output
conv3 weights
conv3 output
pool3 output
fc1 weights
fc1 output
fc2 weights
fc2 output
Σ
160 kB
92.4 kB
41.9 kB
9.9 kB
256 B
49 kB
16 kB
2.3 kB
14.3 kB
3.6 kB
1.2 kB
1.3 kB
336 B
21.6 kB
64 B
260 B
4 B
≈ 294 kB
≈ 52 kB
≈ 59 kB
≈ 65 kB
≈ 33 kB
≈ 18 kB
≈ 6 kB
≈ 1 kB
≈ 22 kB
≈ 0.3 kB

ESP32-S3

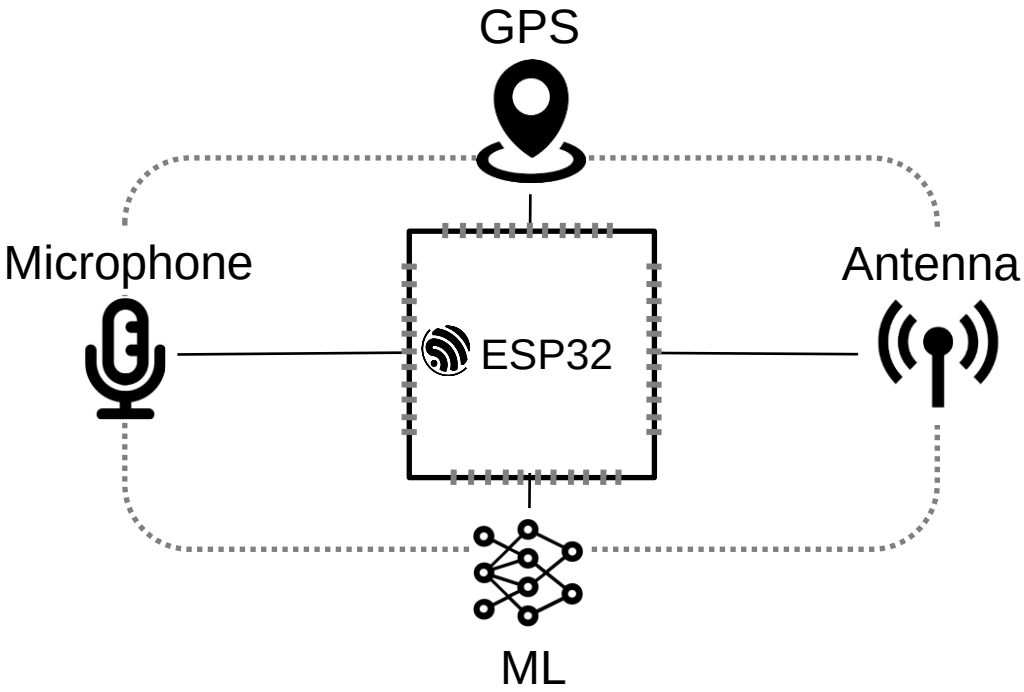


GPS

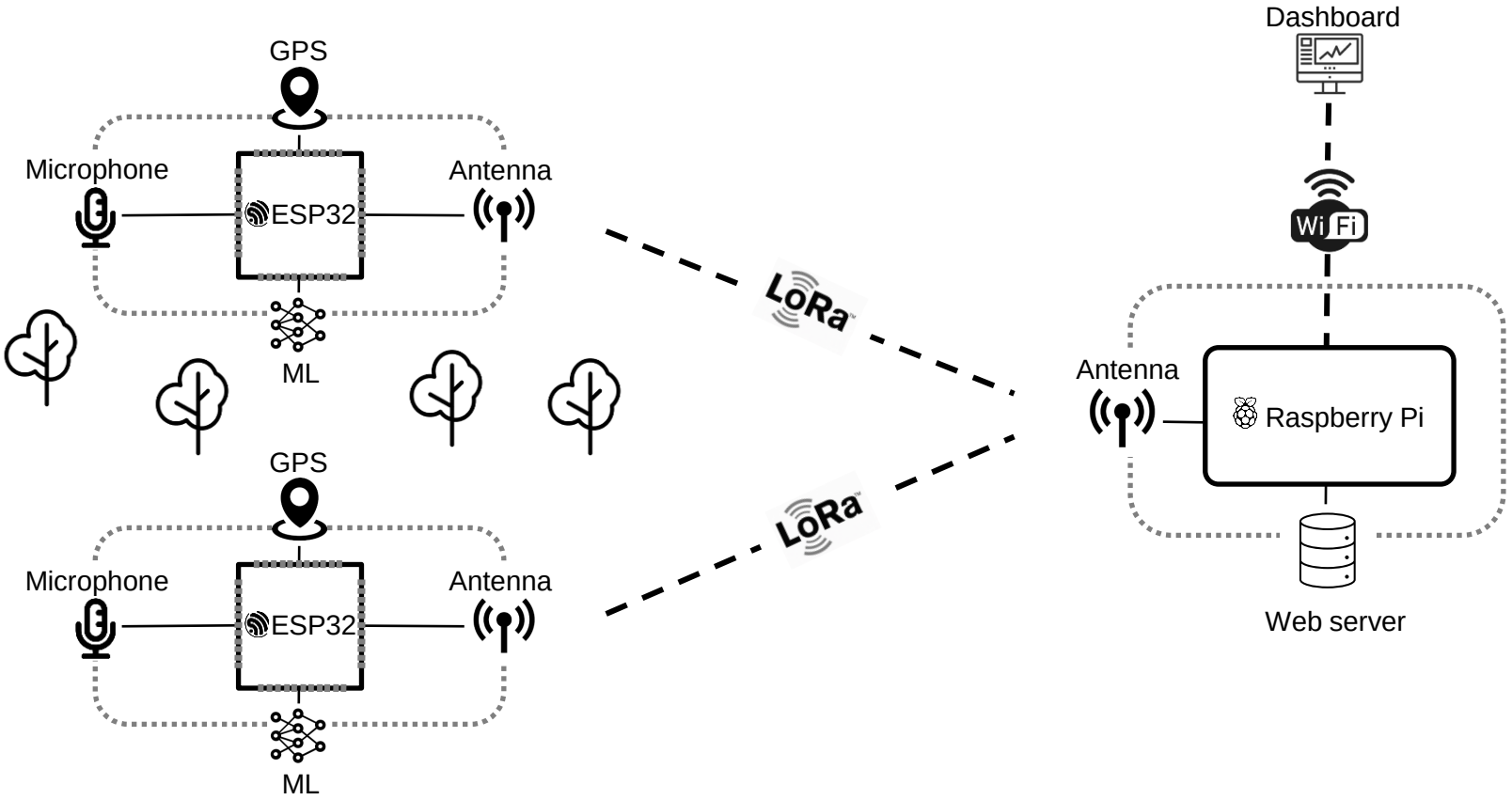
- Up to 2.5m accuracy
- Cold start after deep sleep
  - Ideal conditions ~2 min
  - Non-ideal conditions ~10-20 min
- Timeout after 30 min



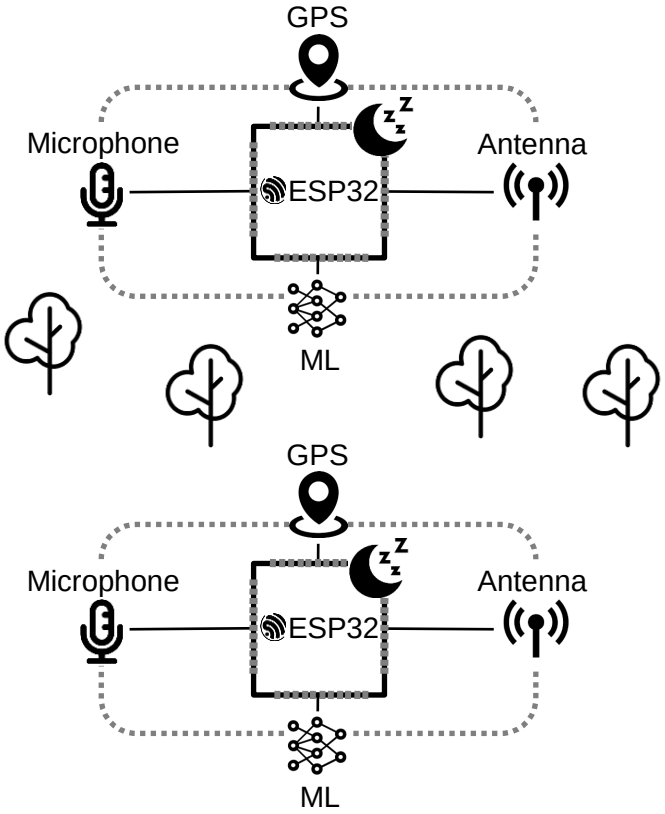
ESP32-S3

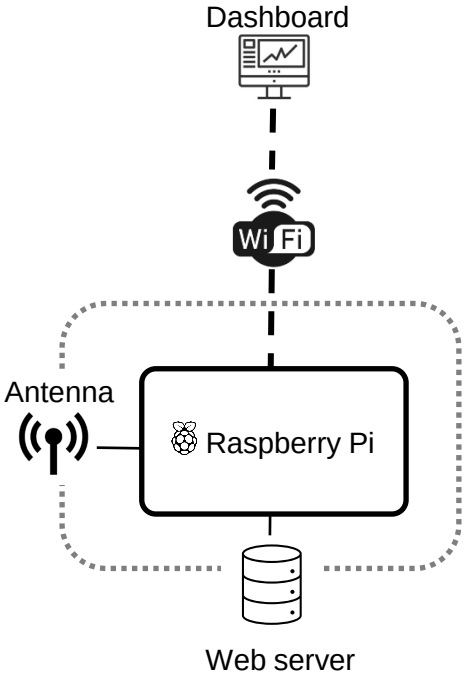


Architecture

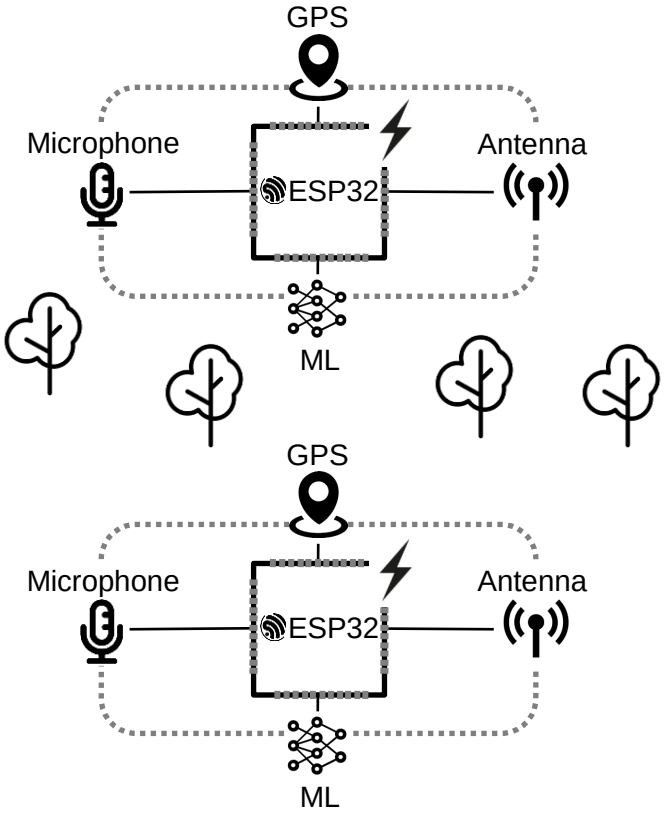


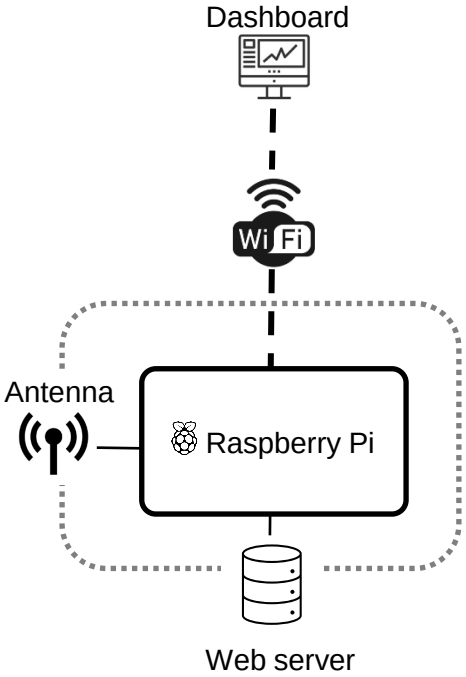
Deep sleep


GPS
Microphone
ESP32
Antenna
ML
GPS
Microphone
ESP32
Antenna
ML

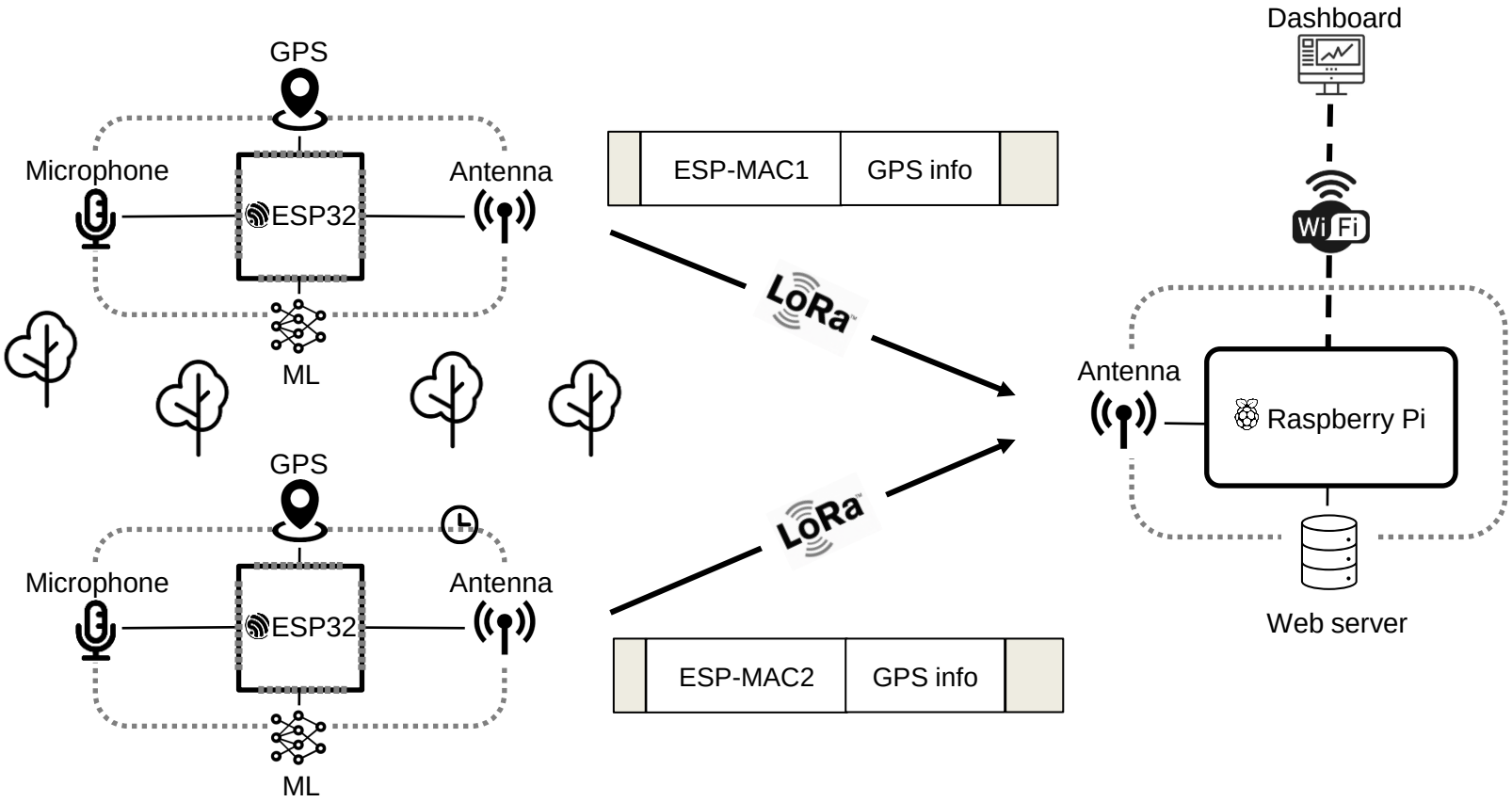

Dashboard
Wi Fi
Antenna
Raspberry Pi
Web server

Timer interrupt 24h

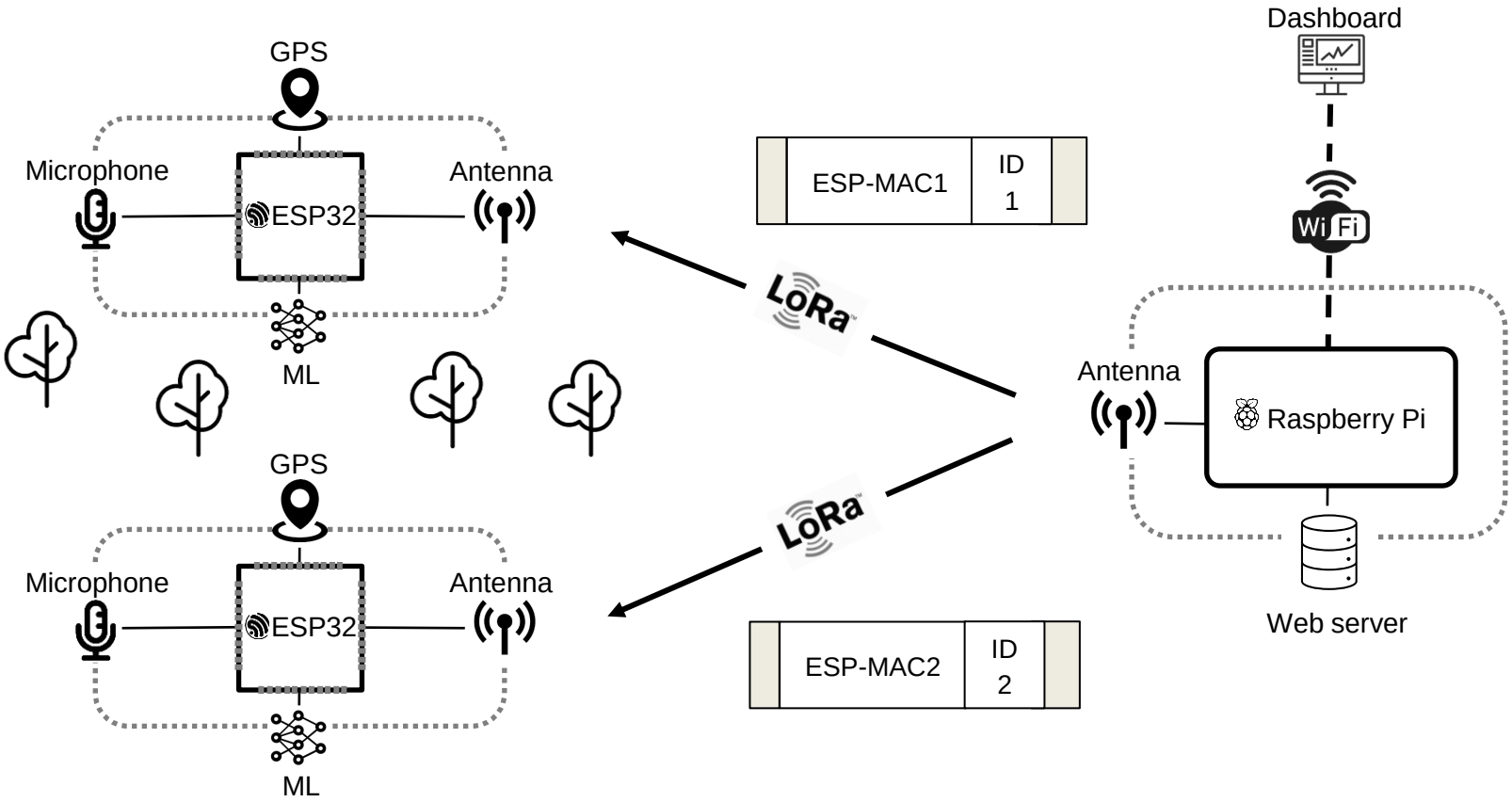




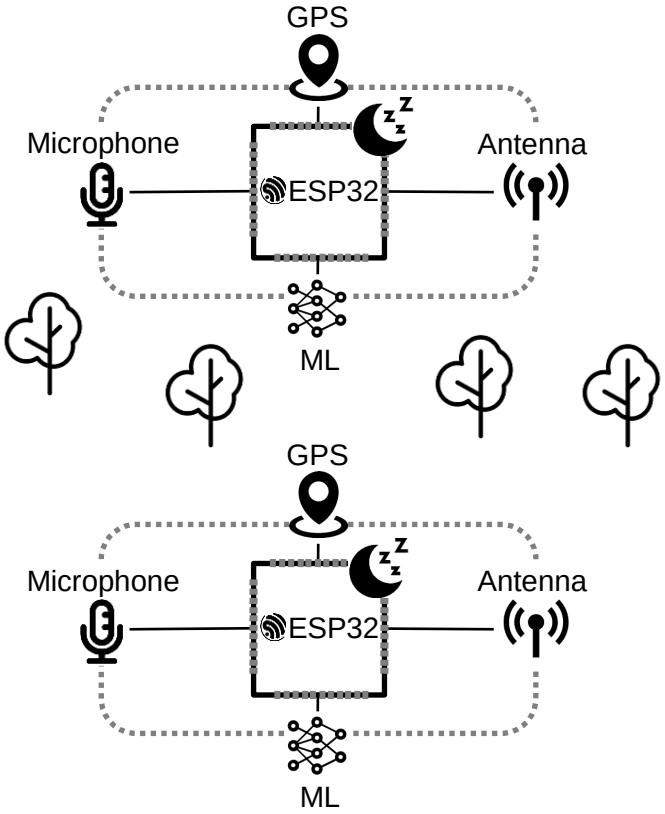
Initialization

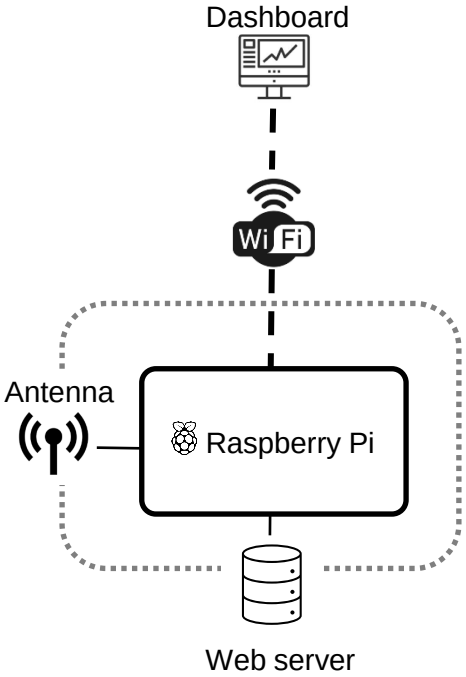


Initialization ACK

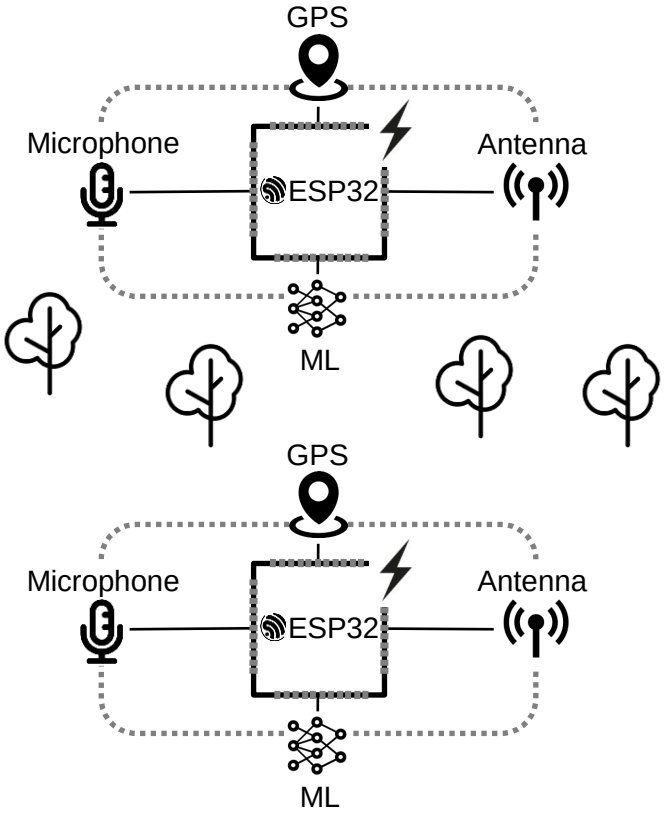


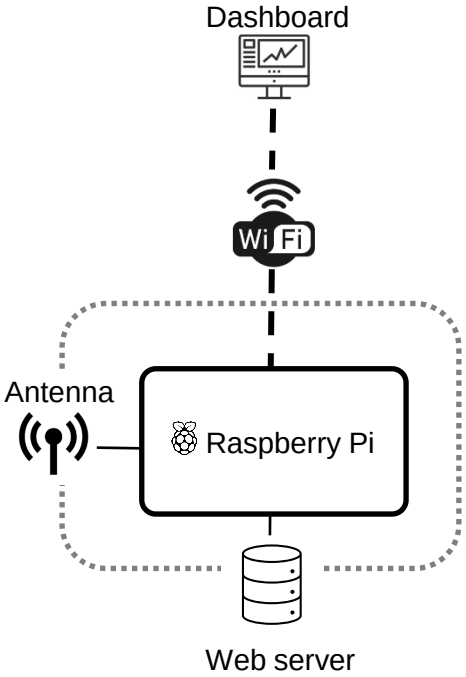
Deep sleep


GPS
Microphone
ESP32
Antenna
ML
GPS
Microphone
ESP32
Antenna
ML

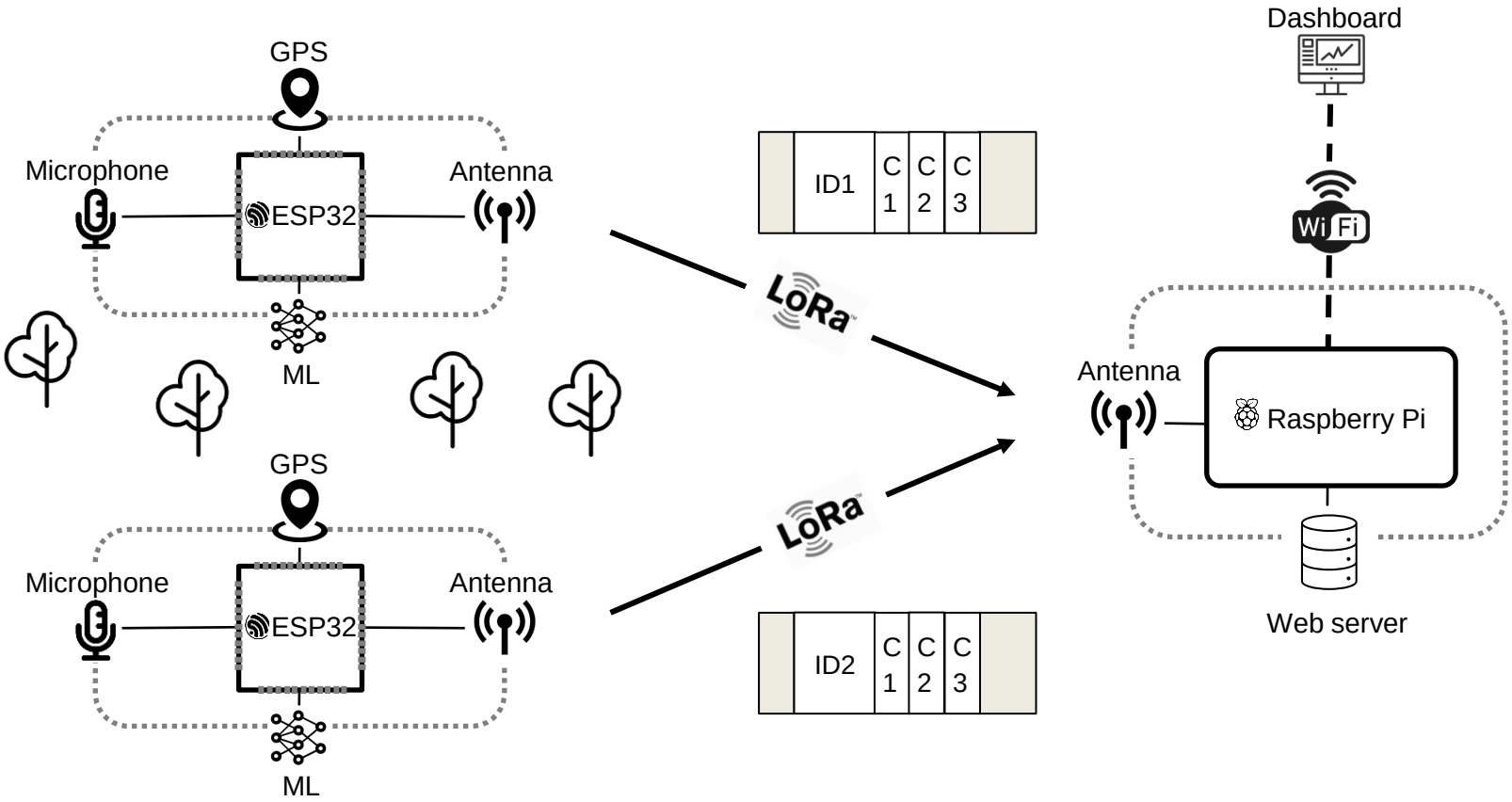

Dashboard
Wi Fi
Antenna
Raspberry Pi
Web server

Timer interrupt 15min

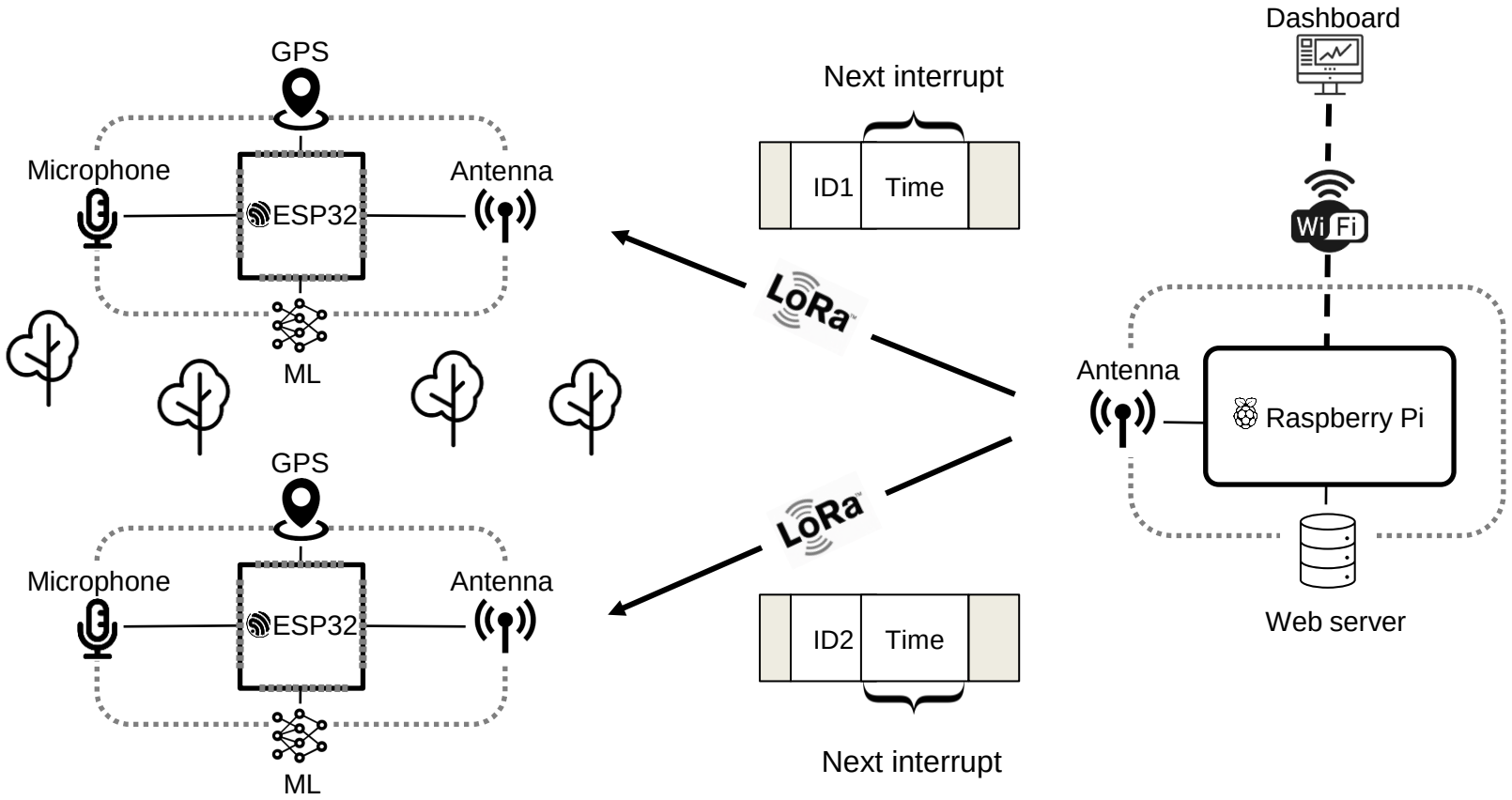




Detection information



Detection information ACK

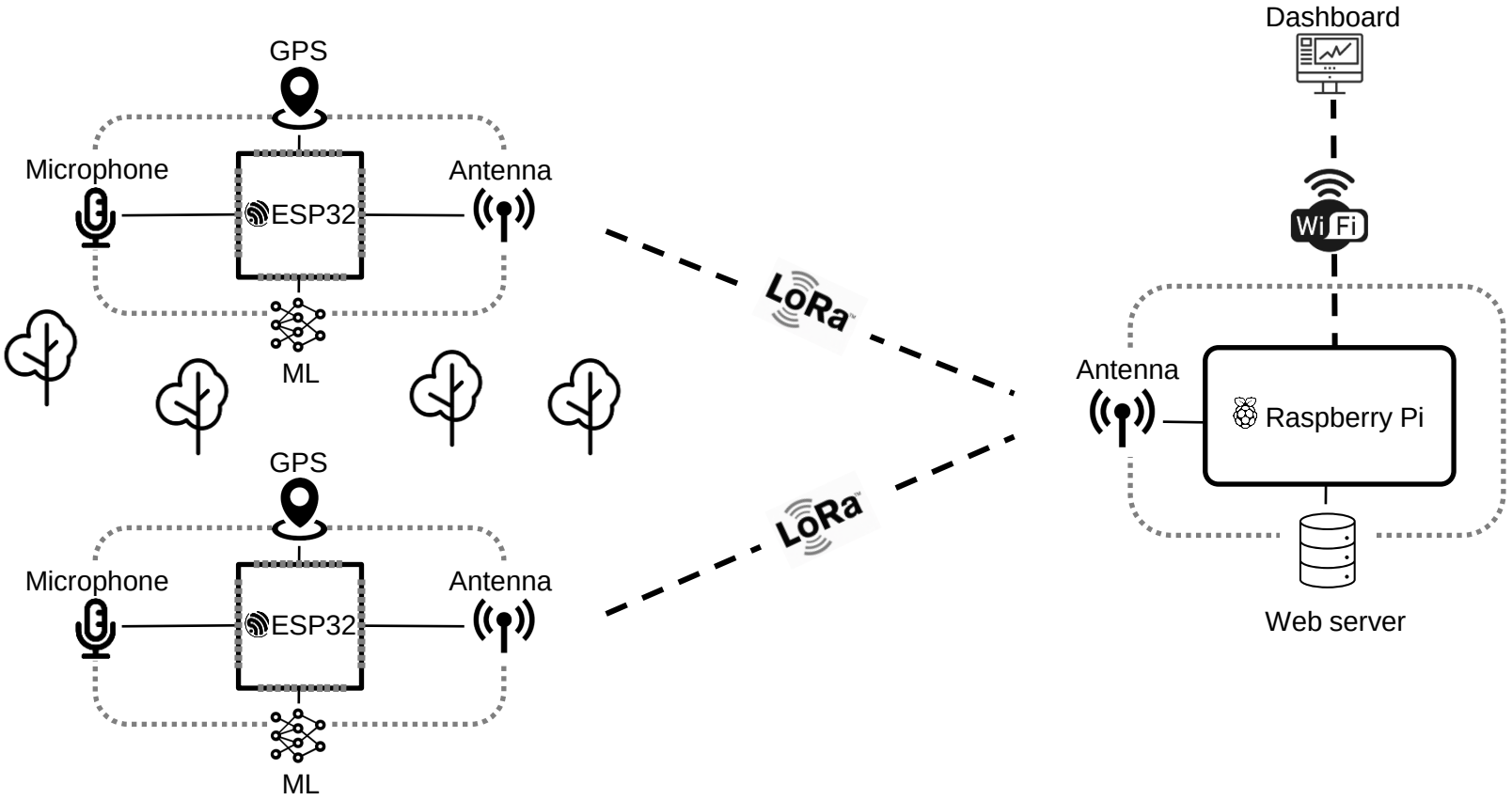


Duty cycle

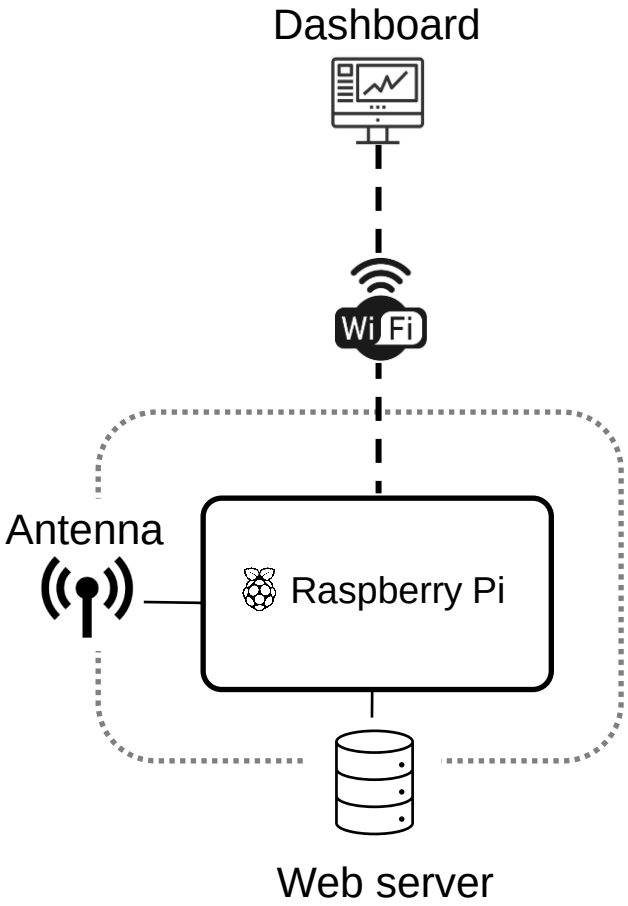
Package type	Payload size (bits)	Time on Air (ms/package)	Max. # packages / 24 h	Max. Time on Air / 24 h (ms)
Init	112	23.17	3	69.51
Init ACK	56	18.05	3	54.15
Detect	19	15.49	288	4,461.12
Detect ACK	16	15.49	288	4,461.12

- Max. Time on Air is ~8.605s / 24 h
- Fair use limit is 30 s / 24 h
- Legal limitation on Time on Air is 864 s / 24 h

Architecture



Raspberry Pi



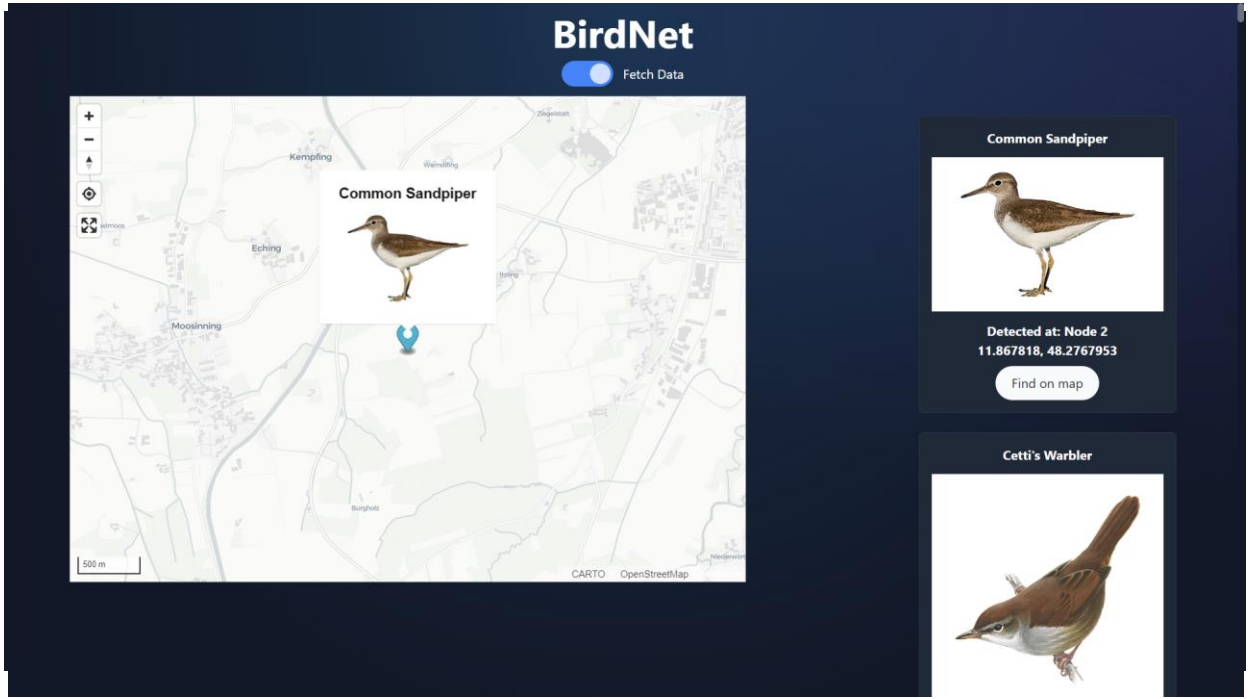
Web server

nodes	
mac_address	String
local_id	integer
long	float
lat	float
createdAt	timestamp

birds	
_id	integer
name	String
node_local_id	integer
long	float
lat	float
createdAt	timestamp

Dashboard

SVELTE



Demo

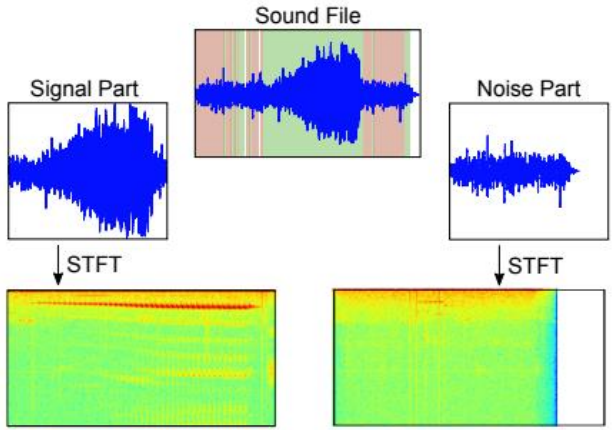
Conclusion

Summary

- (Nearly) end-to-end ML pipeline
- Optimized preprocessing and DL for IoT
- LoRa based networked system
- Initial back- and frontend for end-users

Future Work

- Power measurements
- Distinguish salient audio segments
- Improve circuit & sound processing
- Experiment with MFCC & NN parameters
- **Real-world deployment**



[Sprengel, Elias, et al. "Audio based bird species identification using deep learning techniques]

End