ESP32 - KeyWord Spotting (KWS)

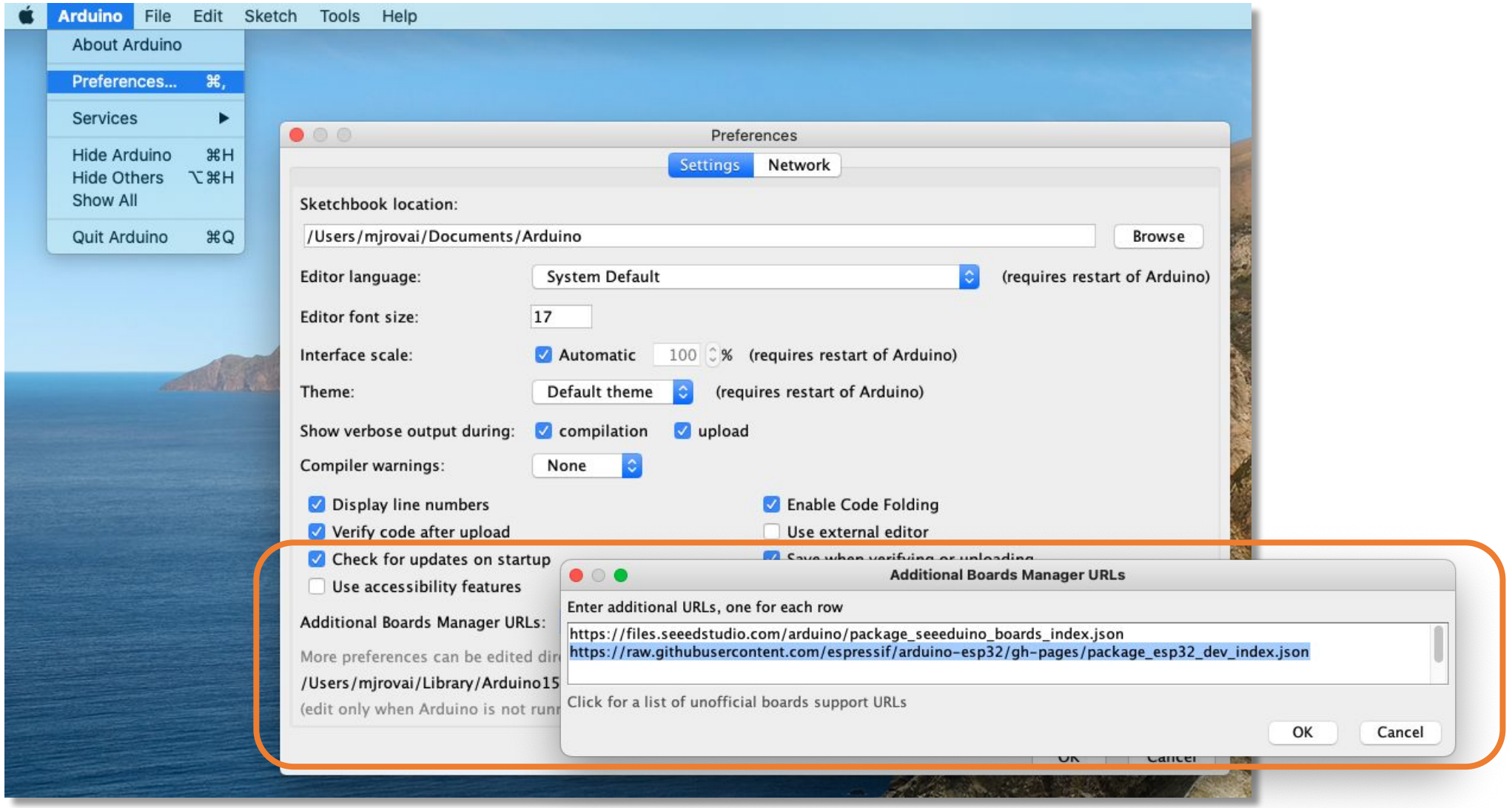
Prof. Marcelo José Rovai
UNIFEI - Federal University of Itajubá, Brazil
TinyML4D Academic Network Co-Chair

UNIFEI

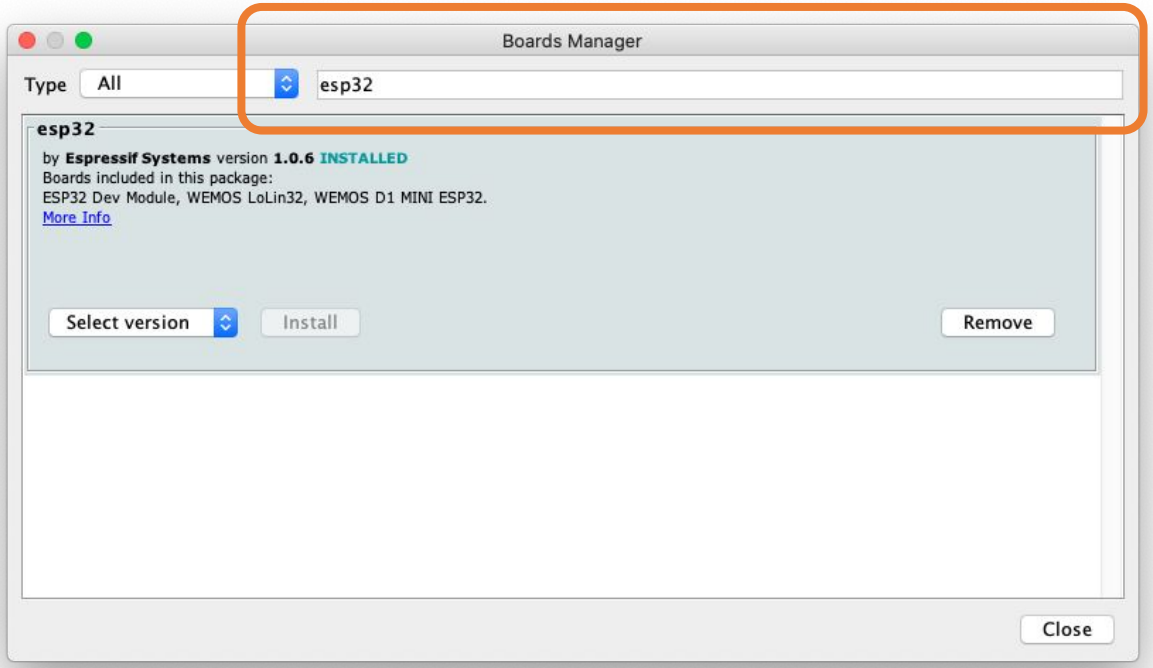
Code available at:

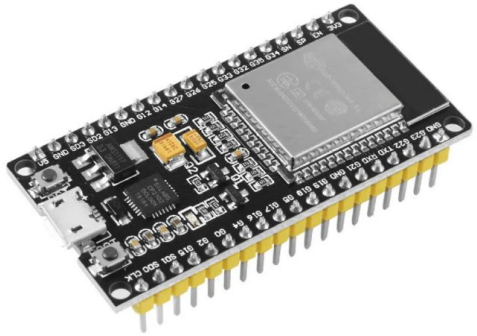
https://github.com/Mjrovai/ESP32-TinyML/tree/main/ESP2-KWS

On Arduino IDE, navigate to **File > Preferences**, and fill in **"Additional Boards Manager URLs"** with the URL below:
https://raw.githubusercontent.com/espressif/arduino-esp32/gh-pages/package_esp32_dev_index.json



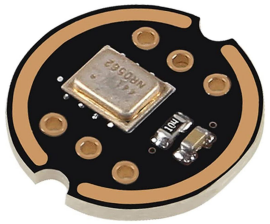
Install HW Libraries

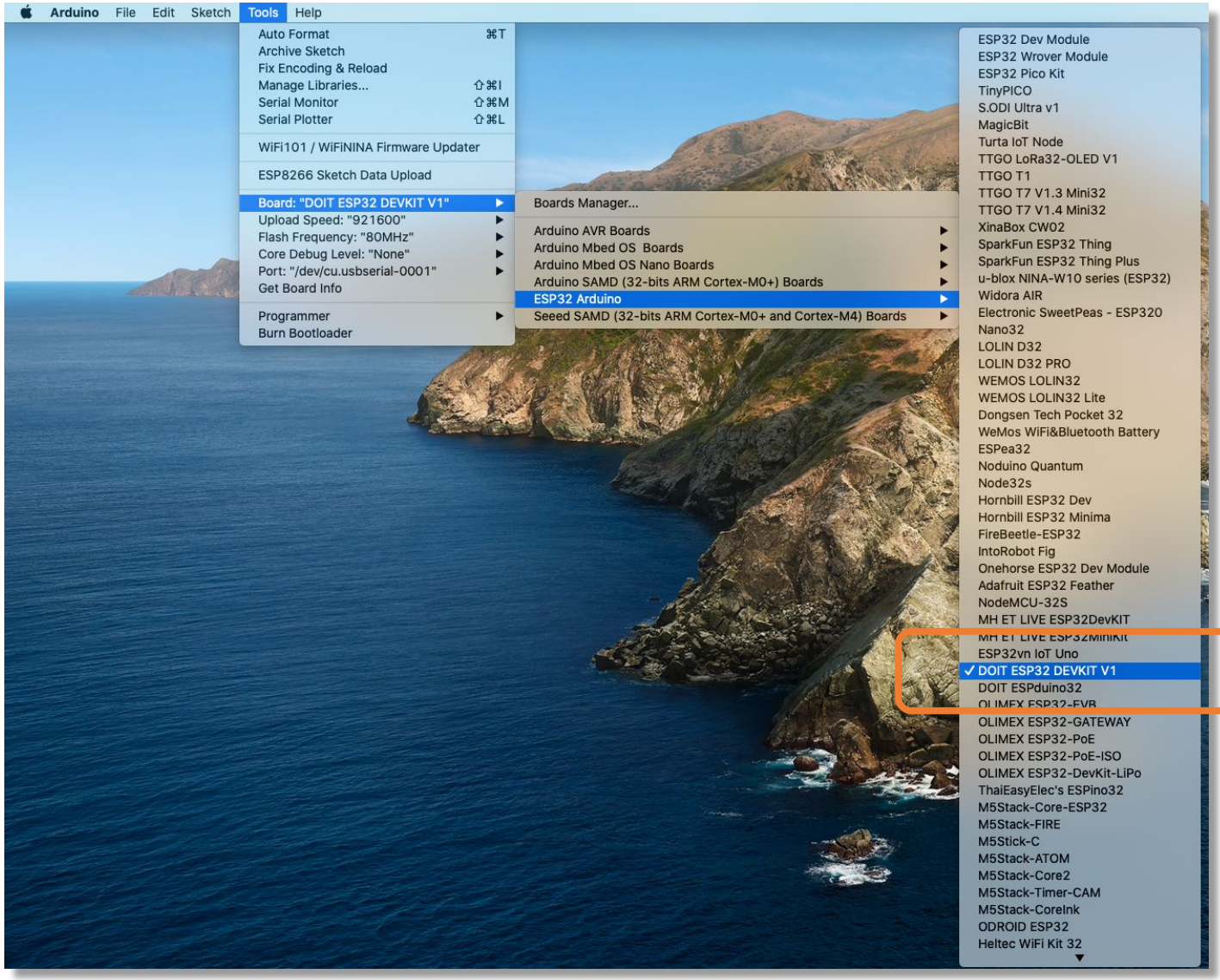


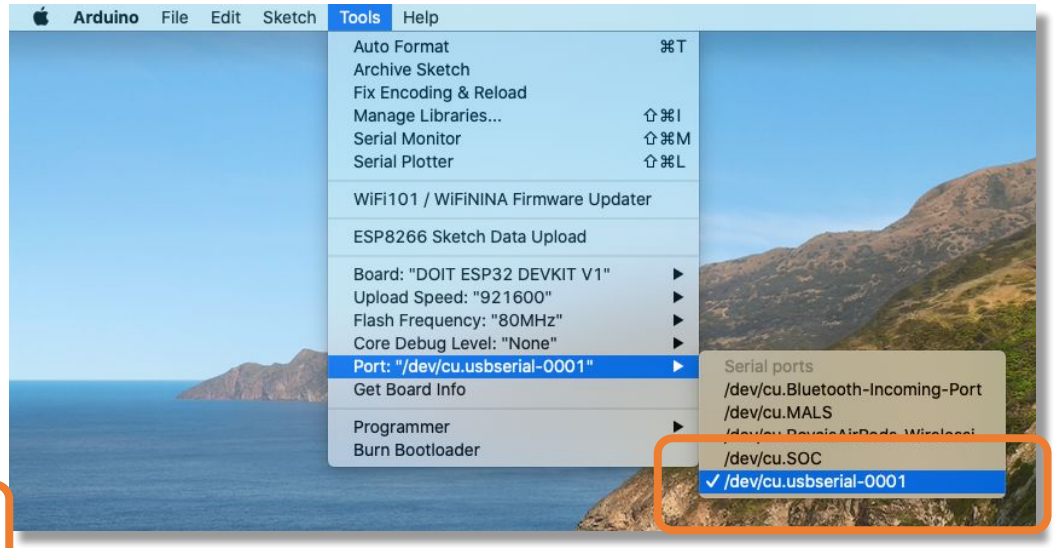


INMP441 Omnidirectional Microphone Module MEMS High Precision Low Power I2S Interface

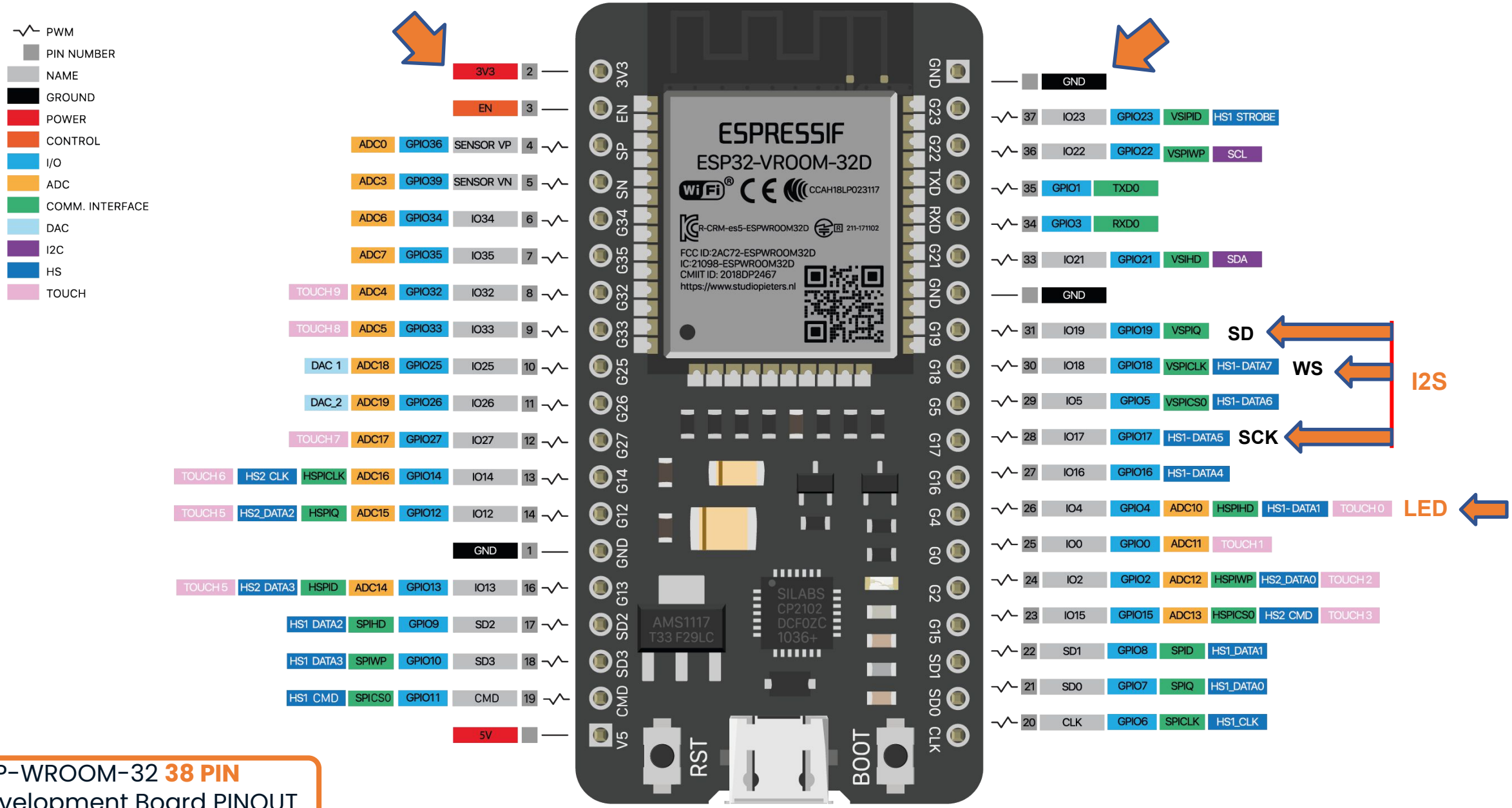
Library ("driver/i2s.h"), is part of ESP32 basic code







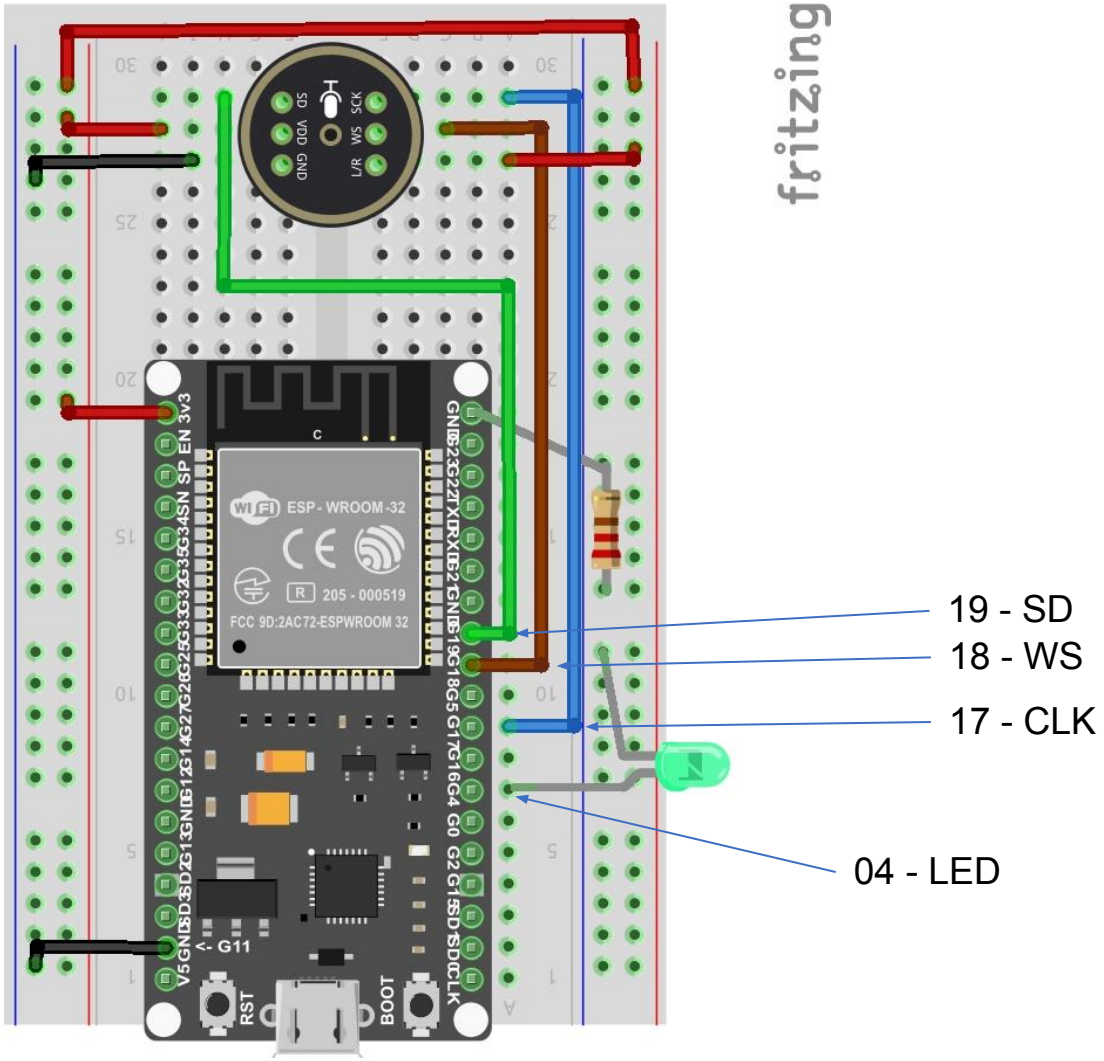
Select Board and Port

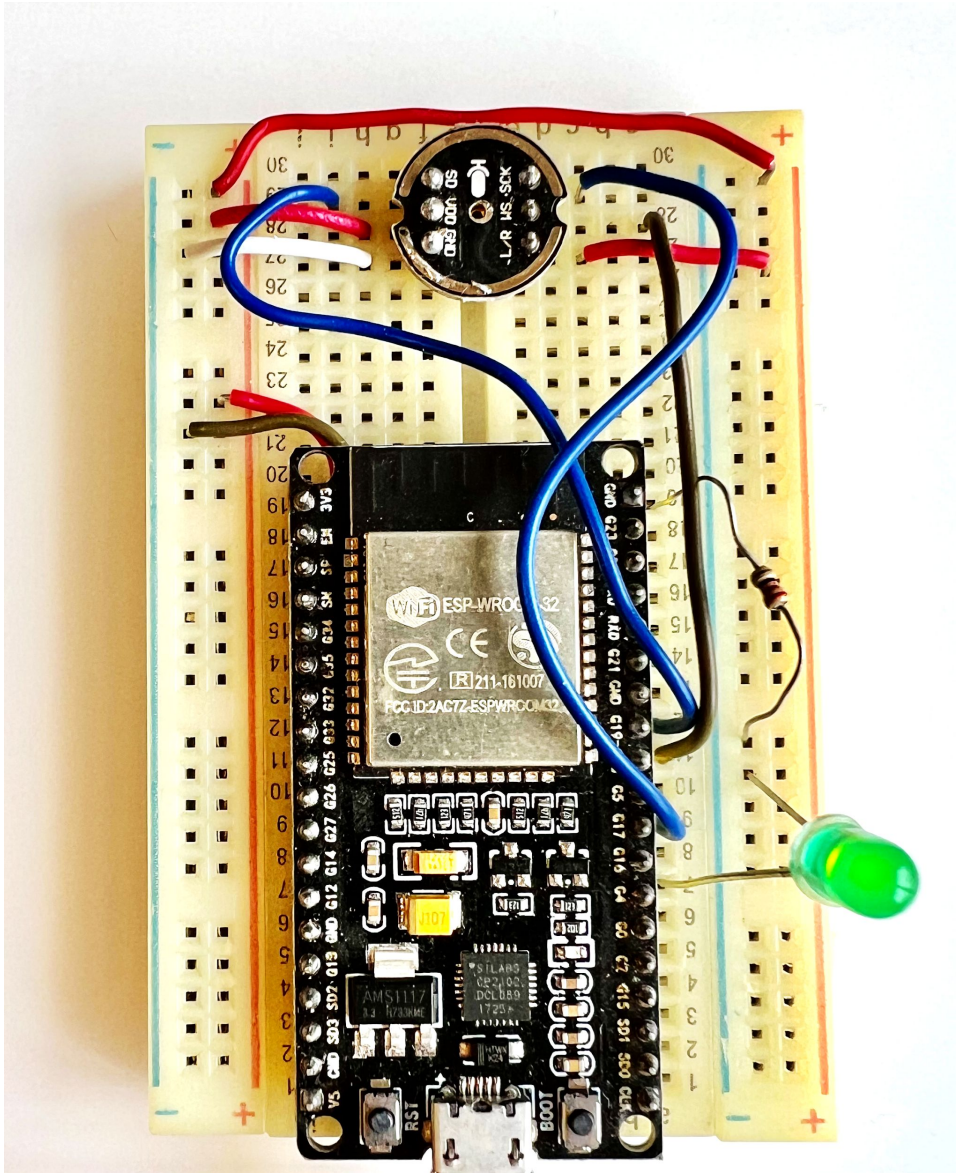


ESP-WROOM-32 **38 PIN**
Development Board PINOUT

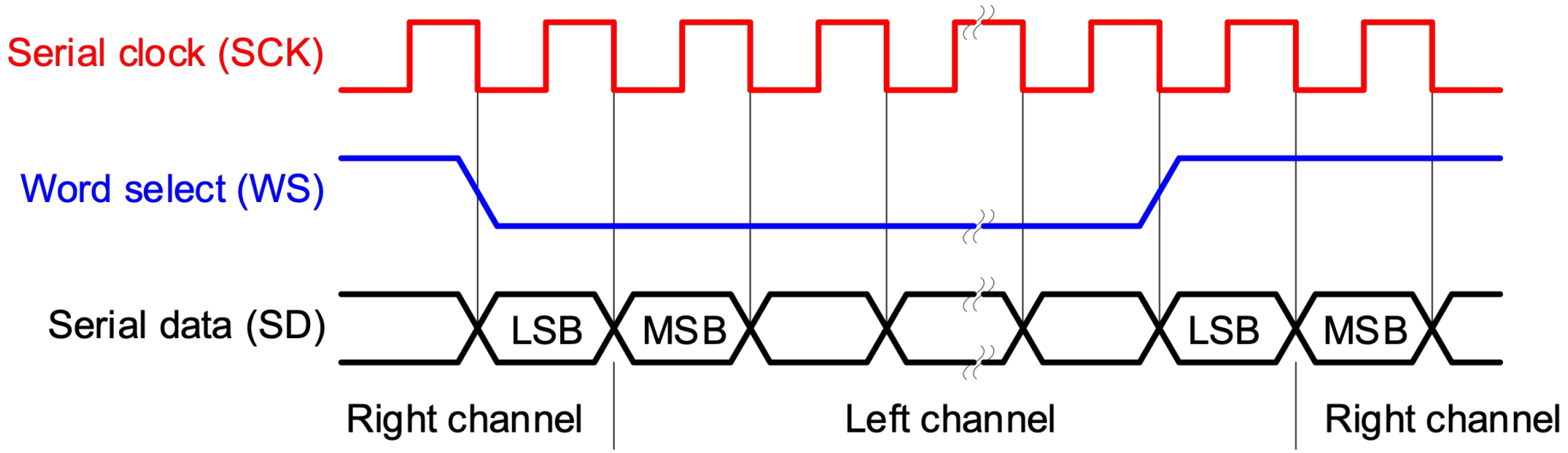
https://www.studiopieters.nl/esp32-pinout/

Connect HW

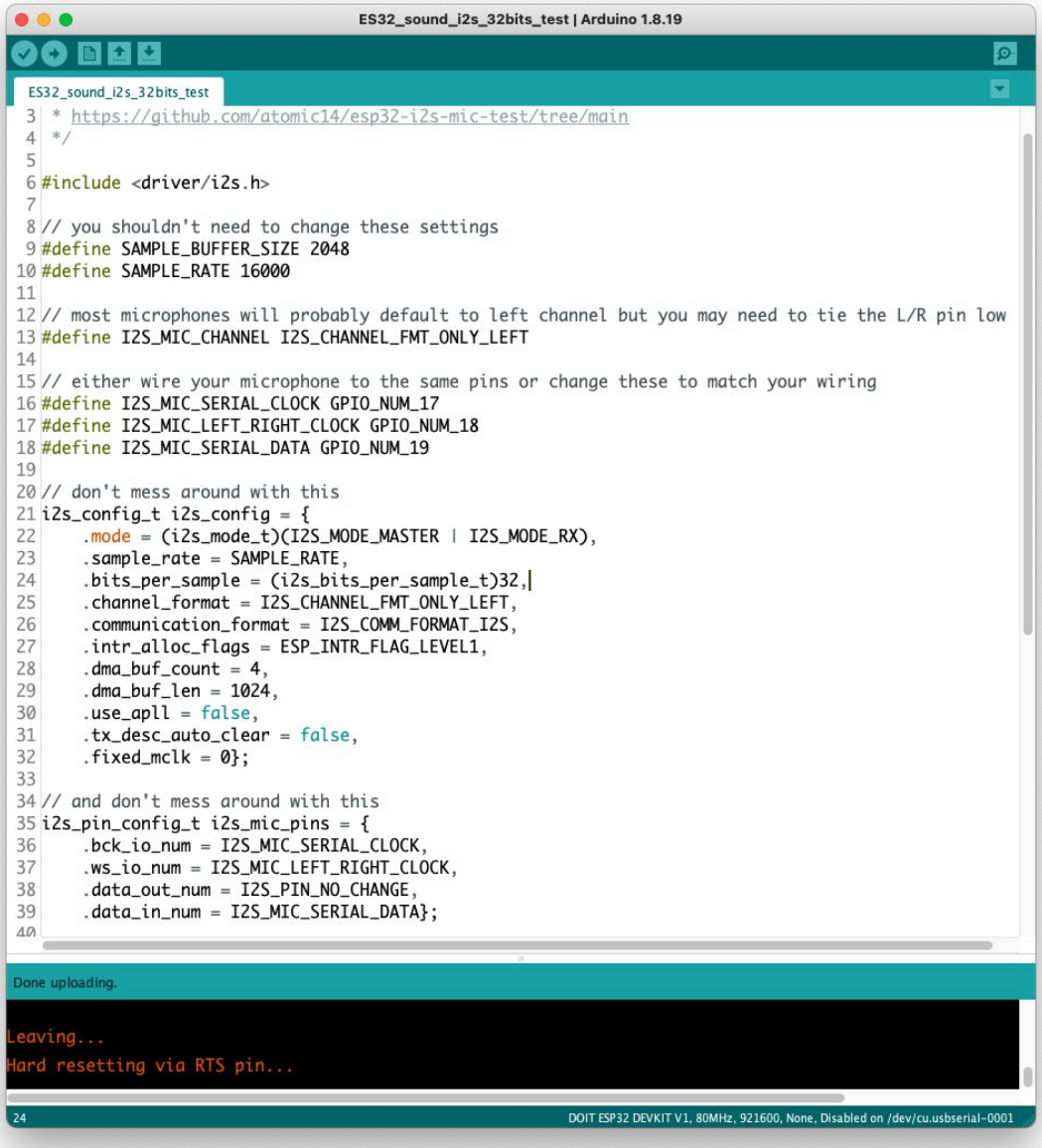

fritzing
SD
VDD
GND
L/R
WS
SCK
ESP - WROOM-32
205 - 000519
FCC 9D:2AC72-ESPWROOM 32
<- G11
RST
BOOT
19 - SD
18 - WS
17 - CLK
04 - LED

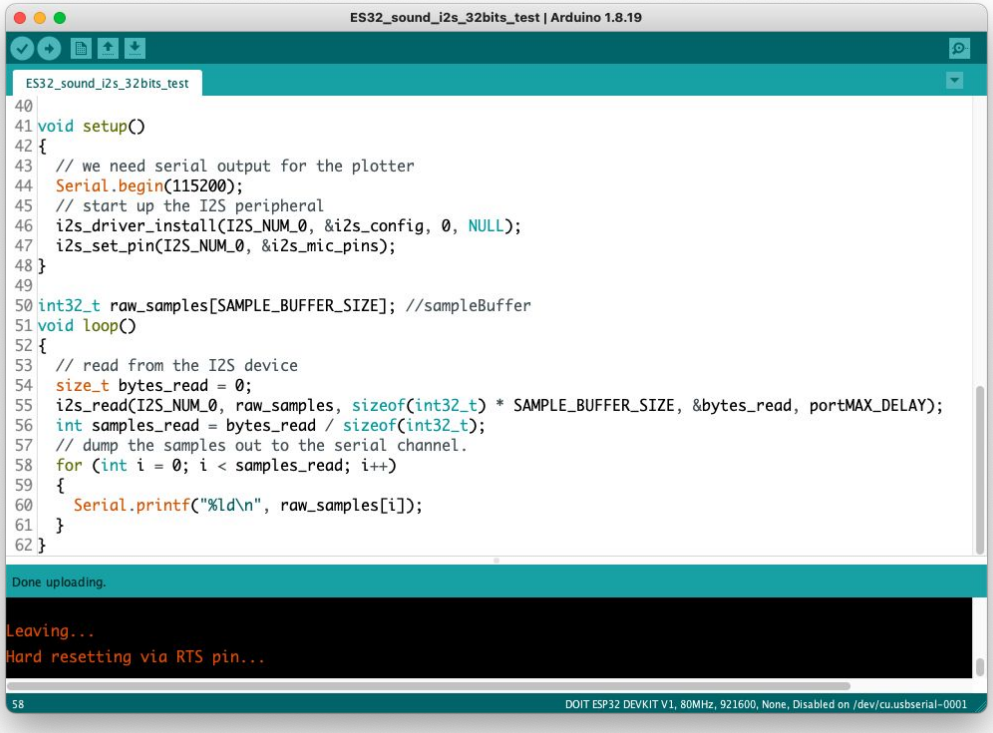


I2S Signals

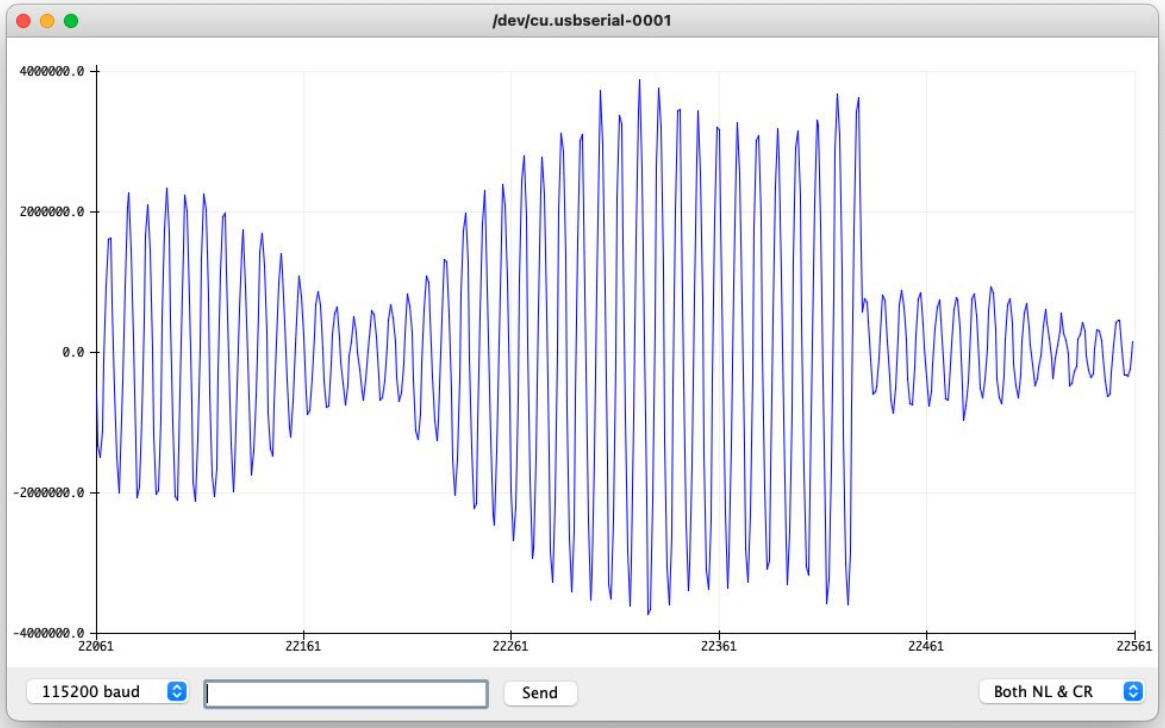

Serial clock (SCK)
Word select (WS)
Serial data (SD)
LSB
MSB
LSB
MSB
Right channel
Left channel
Right channel

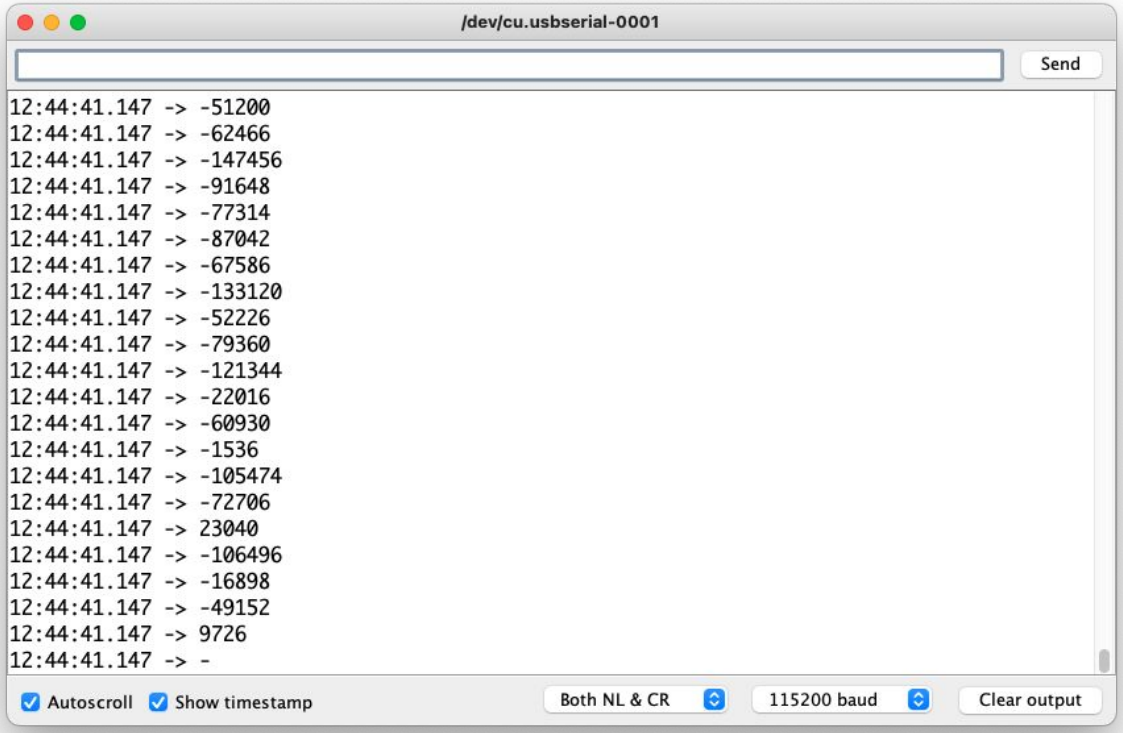
Test ESP32 + Mic

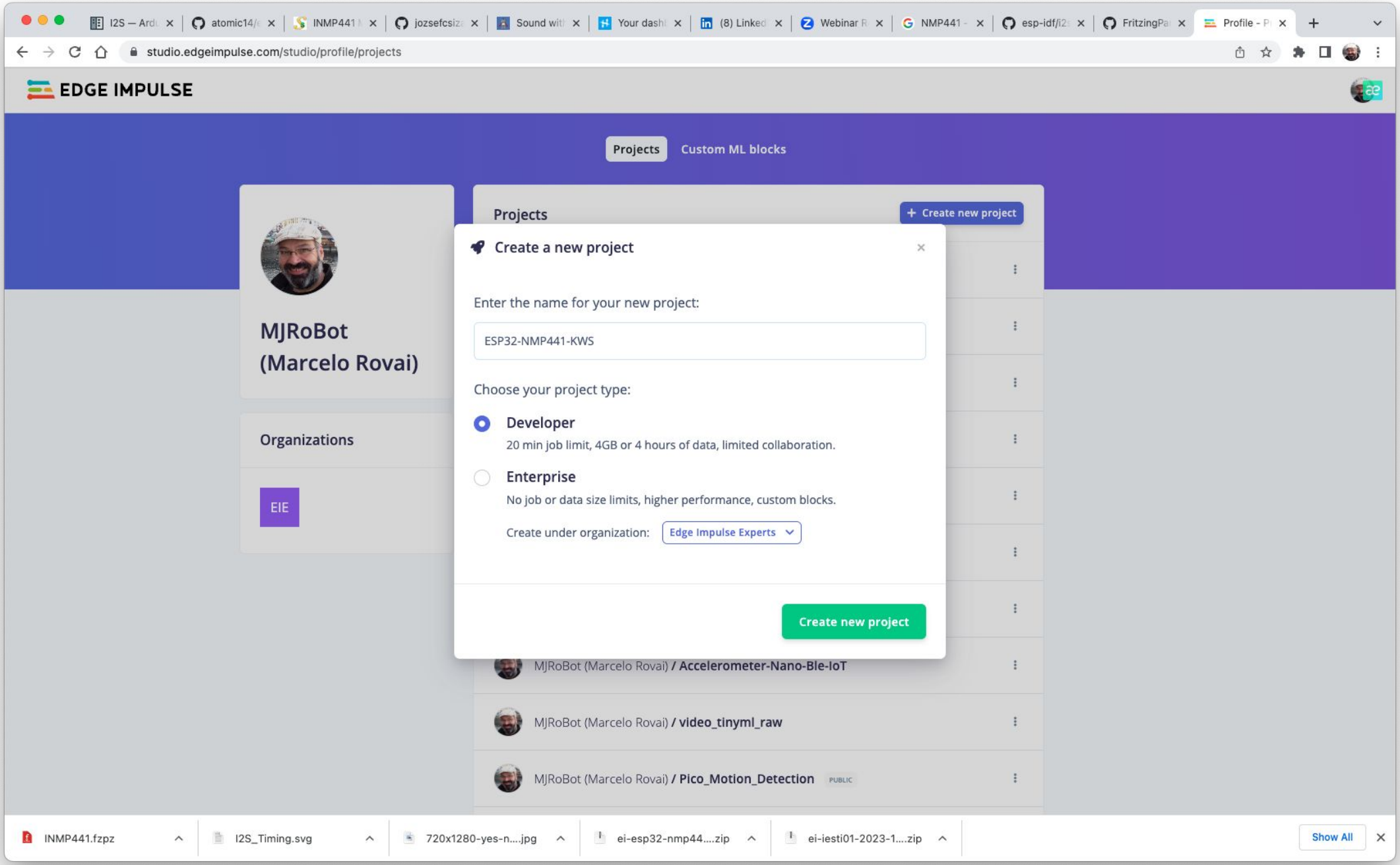




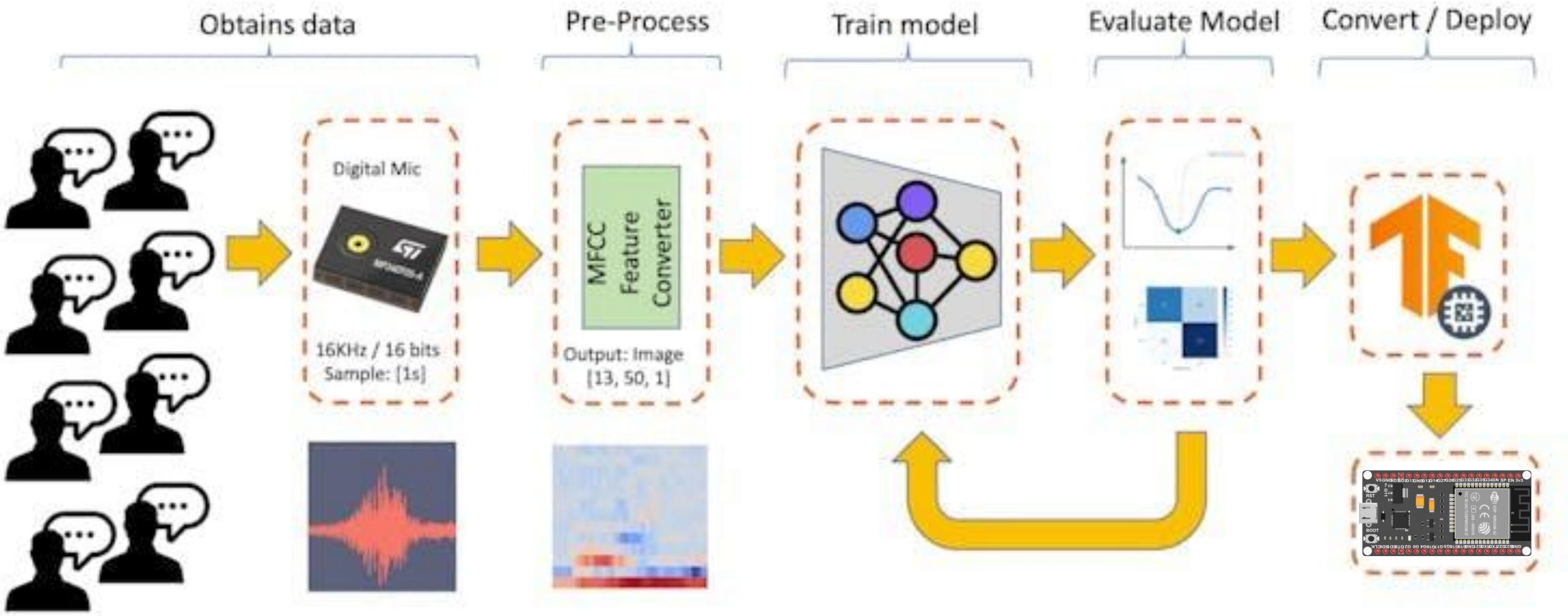
Test ESP32 + Mic

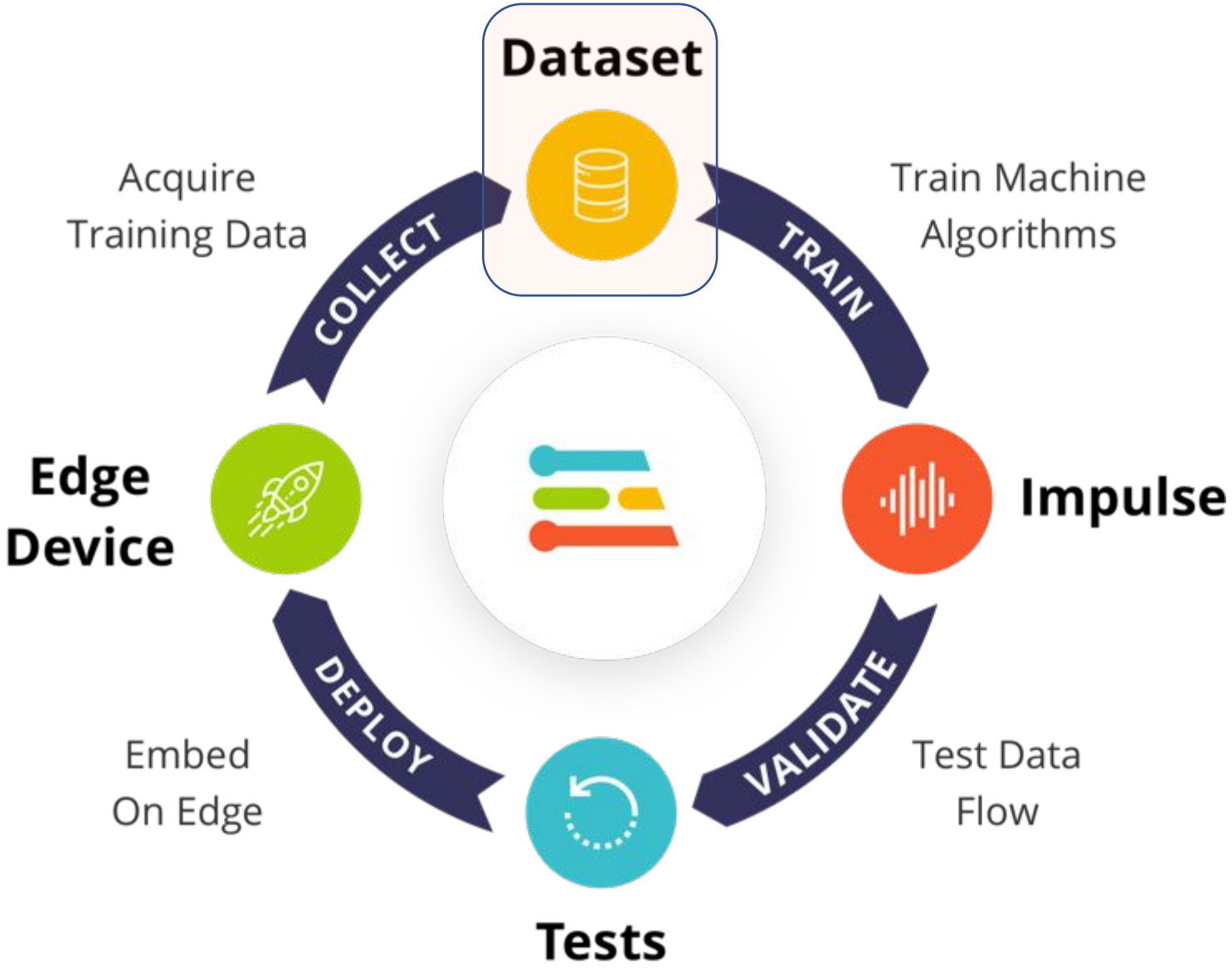





studio.edgeimpulse.com/studio/profile/projects
EDGE IMPULSE
Projects
Custom ML blocks
MJRoBot
(Marcelo Rovai)
Organizations
EIE
Projects
+ Create new project
Create a new project
Enter the name for your new project:
ESP32-NMP441-KWS
Choose your project type:
Developer
20 min job limit, 4GB or 4 hours of data, limited collaboration.
Enterprise
No job or data size limits, higher performance, custom blocks.
Create under organization:
Edge Impulse Experts
Create new project
MJRoBot (Marcelo Rovai) / Accelerometer-Nano-Ble-IoT
MJRoBot (Marcelo Rovai) / video_tinyml_raw
MJRoBot (Marcelo Rovai) / Pico_Motion_Detection
PUBLIC
INMP441.fzpz
I2S_Timing.svg
720x1280-yes-n....jpg
ei-esp32-nmp44....zip
ei-iesti01-2023-1....zip
Show All

KeyWord Spotting (KWS) - **Inference**





- Pre-Processing Data
- Design a Model
- Train a Model

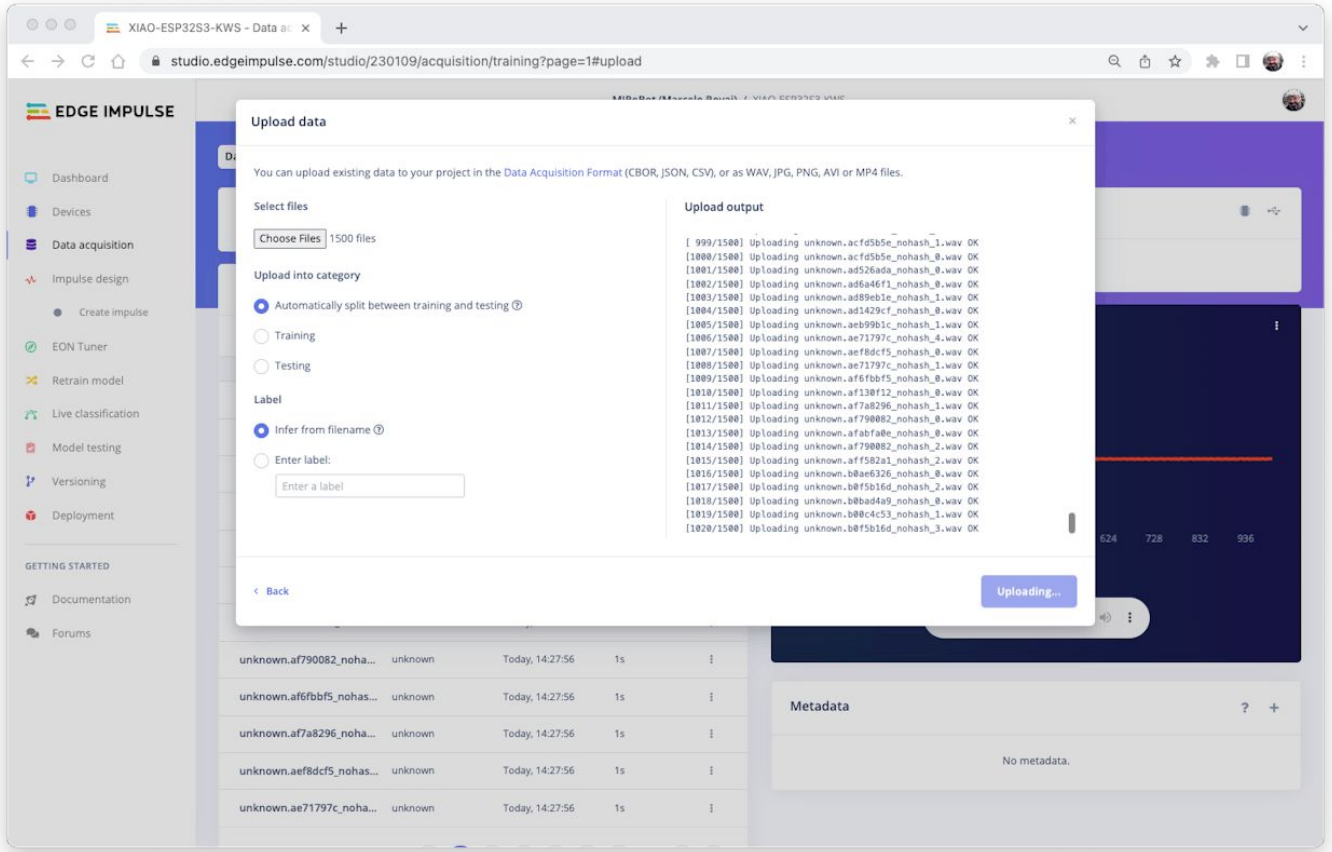
Dataset – Speech Commands

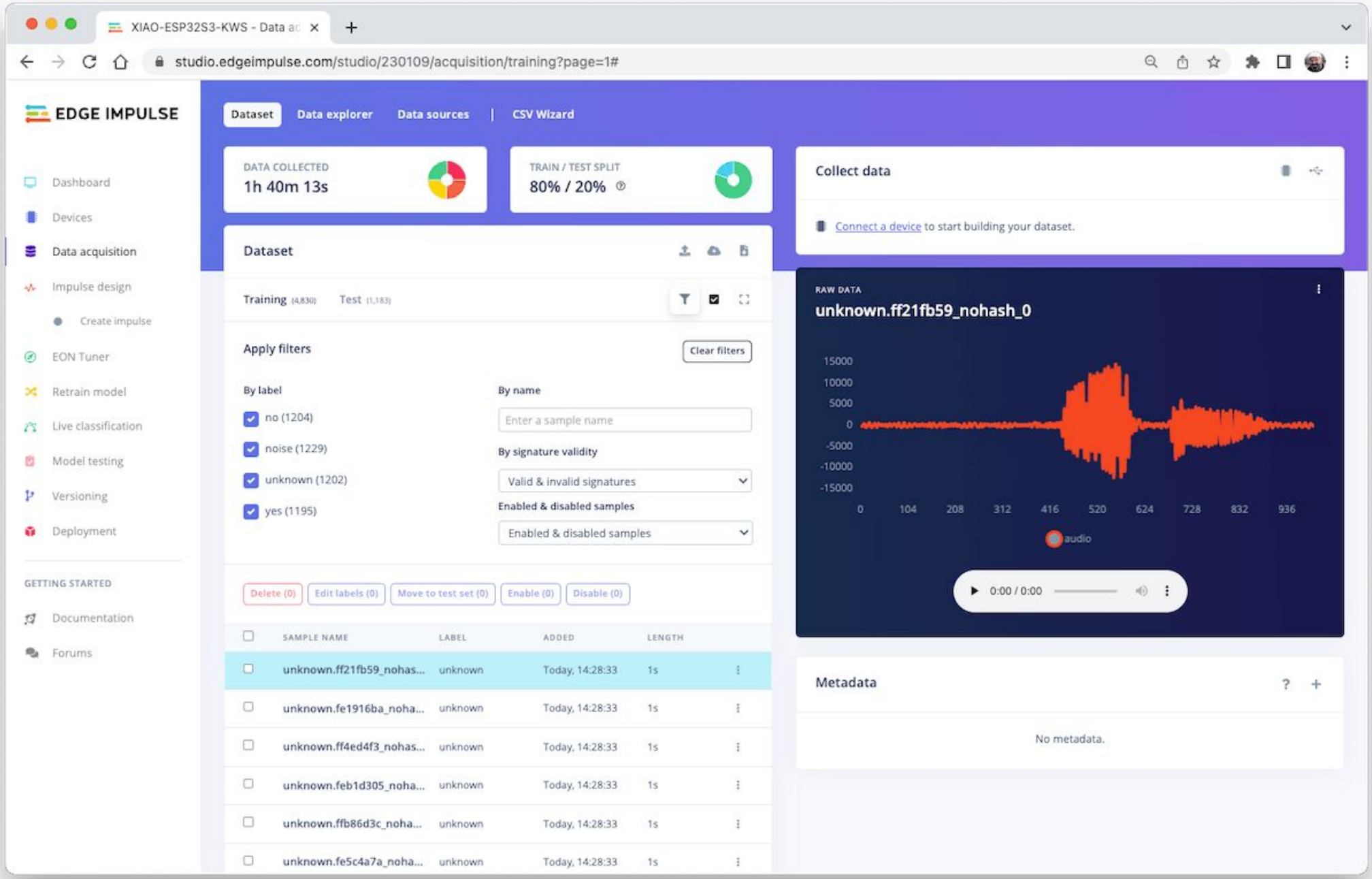
Pete Warden's "Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition".

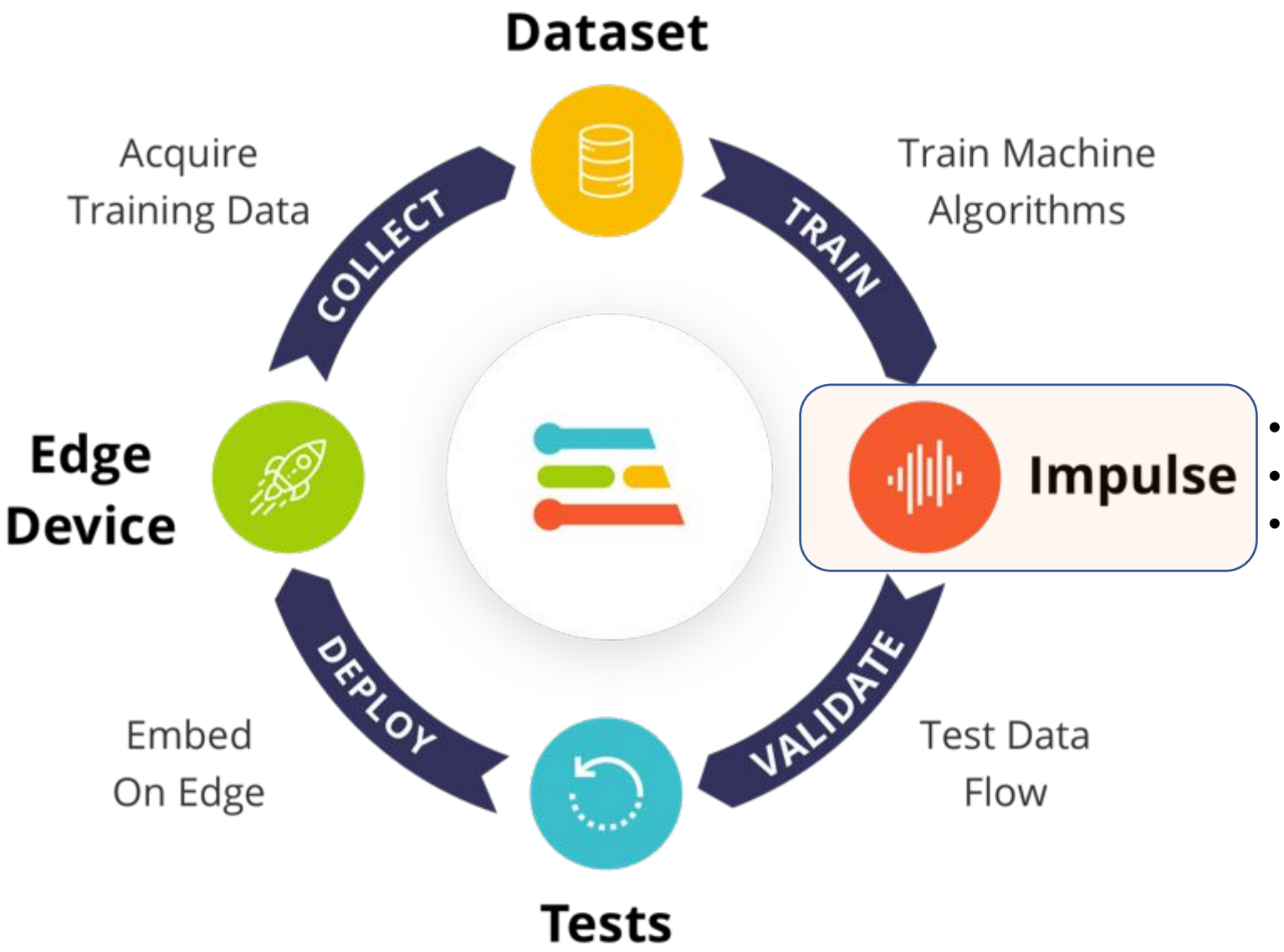
- 35 keywords (with +1, 000 samples each), such as yes, no, stop, and go.
- In some words, we can get 1,500 samples, such as yes and no.

Download a small portion of the dataset from Edge Studio (Keyword spotting pre-built dataset), which includes samples from the four classes we will use in this project: **yes, no, noise, and background**. For this, follow the steps below:

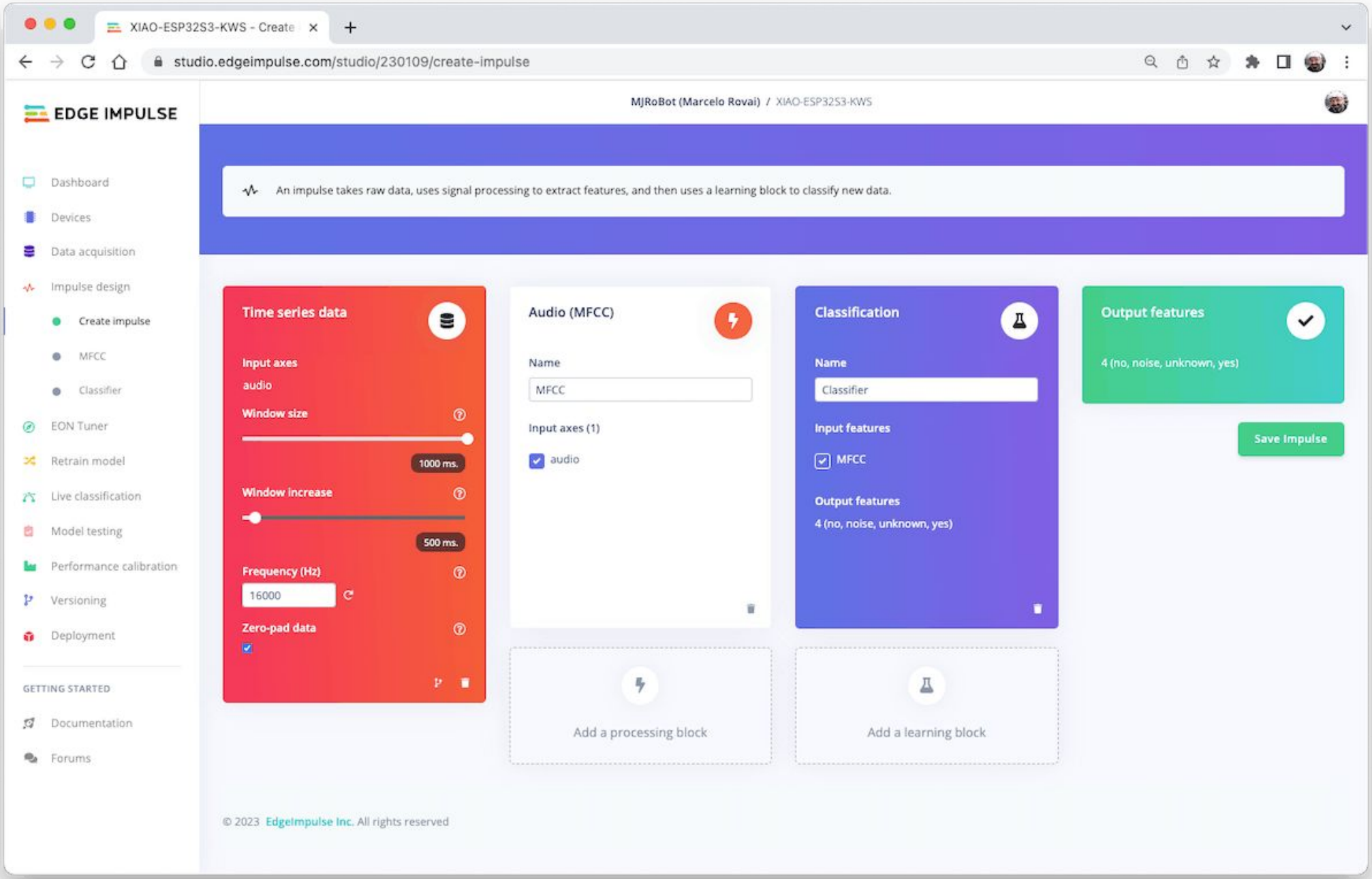
- Download the keywords dataset.
- Unzip the file in a location of your choice.
- Upload to the Studio

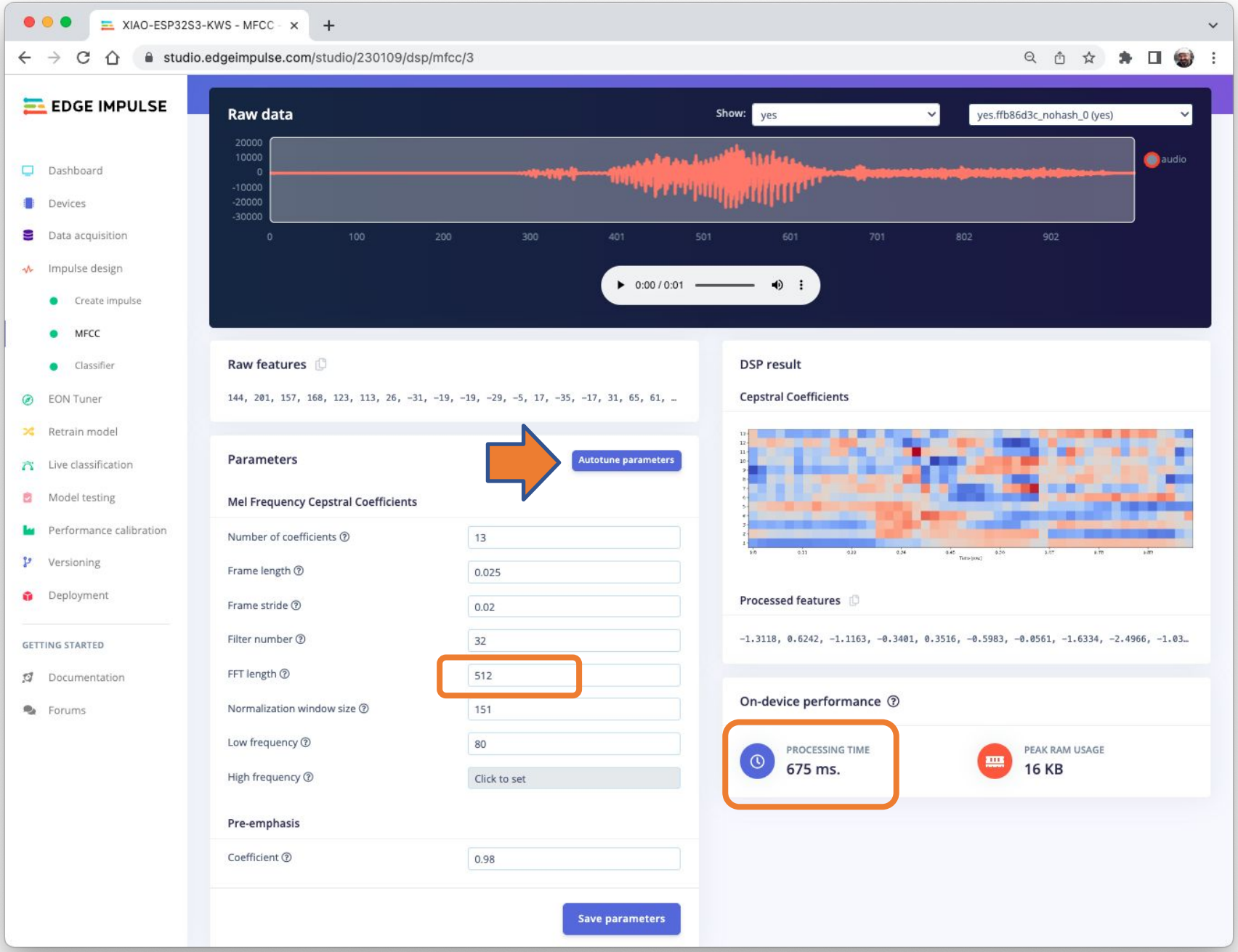



XIAO-ESP32S3-KWS - Data ac
studio.edgeimpulse.com/studio/230109/acquisition/training?page=1#
EDGE IMPULSE
Dashboard
Devices
Data acquisition
Impulse design
Create impulse
EON Tuner
Retrain model
Live classification
Model testing
Versioning
Deployment
GETTING STARTED
Documentation
Forums
Dataset
Data explorer
Data sources
CSV Wizard
DATA COLLECTED
1h 40m 13s
TRAIN / TEST SPLIT
80% / 20%
Collect data
Connect a device to start building your dataset.
Dataset
Training (4,830)
Test (1,183)
Apply filters
Clear filters
By label
no (1204)
noise (1229)
unknown (1202)
yes (1195)
By name
Enter a sample name
By signature validity
Valid & invalid signatures
Enabled & disabled samples
Enabled & disabled samples
Delete (0)
Edit labels (0)
Move to test set (0)
Enable (0)
Disable (0)
SAMPLE NAME
LABEL
ADDED
LENGTH
unknown.ff21fb59_nohas...
unknown
Today, 14:28:33
1s
unknown.fe1916ba_noha...
unknown.ff4ed4f3_nohas...
unknown.feb1d305_noha...
unknown.ffb86d3c_noha...
unknown.fe5c4a7a_noha...
RAW DATA
unknown.ff21fb59_nohash_0
15000
10000
5000
0
-5000
-10000
-15000
0
104
208
312
416
520
624
728
832
936
audio
0:00 / 0:00
Metadata
No metadata.

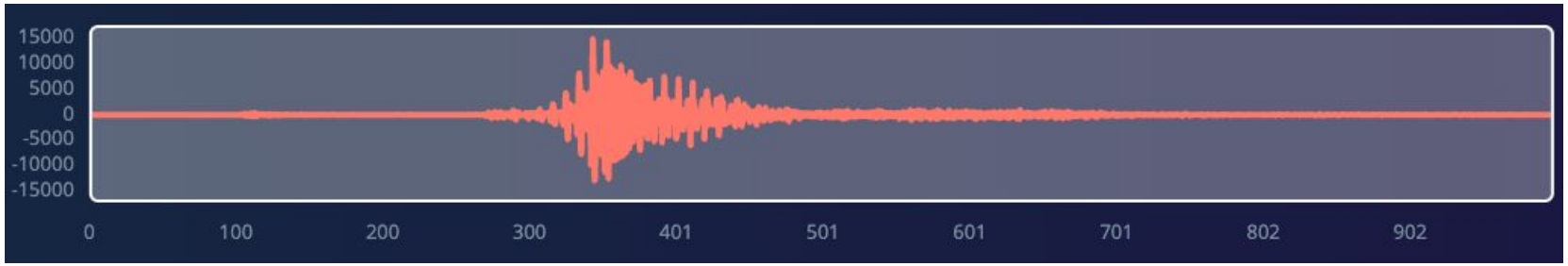


- Pre-Processing Data
- Design a Model
- Train a Model

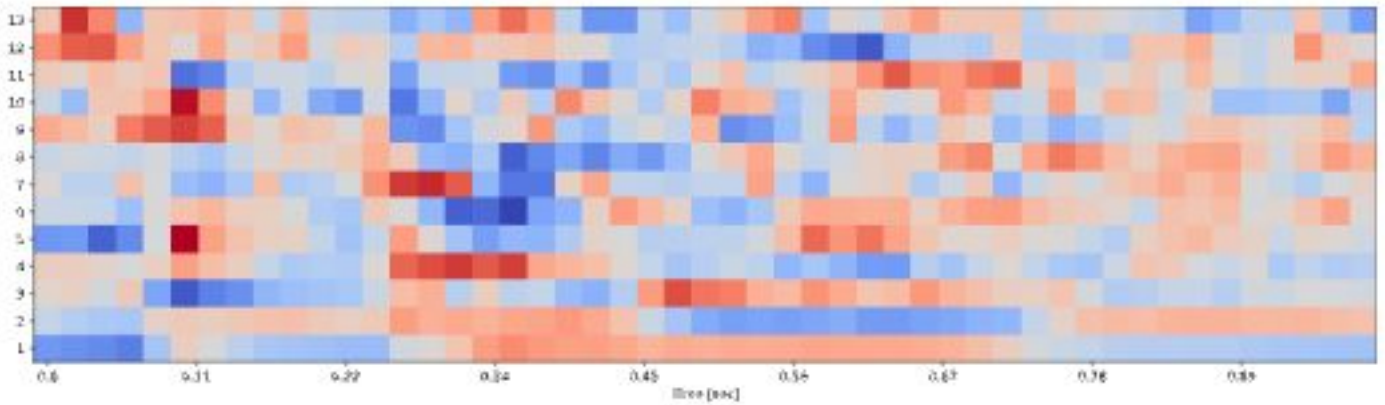

XIAO-ESP32S3-KWS - Create
studio.edgeimpulse.com/studio/230109/create-impulse
EDGE IMPULSE
MJRoBot (Marcelo Rovai) / XIAO-ESP32S3-KWS
Dashboard
Devices
Data acquisition
Impulse design
Create impulse
MFCC
Classifier
EON Tuner
Retrain model
Live classification
Model testing
Performance calibration
Versioning
Deployment
GETTING STARTED
Documentation
Forums
An impulse takes raw data, uses signal processing to extract features, and then uses a learning block to classify new data.
Time series data
Input axes
audio
Window size
1000 ms.
Window increase
500 ms.
Frequency (Hz)
16000
Zero-pad data
Audio (MFCC)
Name
MFCC
Input axes (1)
audio
Classification
Name
Classifier
Input features
MFCC
Output features
4 (no, noise, unknown, yes)
Output features
4 (no, noise, unknown, yes)
Save Impulse
Add a processing block
Add a learning block
© 2023 EdgeImpulse Inc. All rights reserved


XIAO-ESP32S3-KWS - MFCC
studio.edgeimpulse.com/studio/230109/dsp/mfcc/3
EDGE IMPULSE
Dashboard
Devices
Data acquisition
Impulse design
Create impulse
MFCC
Classifier
EON Tuner
Retrain model
Live classification
Model testing
Performance calibration
Versioning
Deployment
GETTING STARTED
Documentation
Forums
Raw data
Show:
yes
yes.ffb86d3c_nohash_0 (yes)
audio
0:00 / 0:01
Raw features
144, 201, 157, 168, 123, 113, 26, -31, -19, -19, -29, -5, 17, -35, -17, 31, 65, 61, …
Parameters
Autotune parameters
Mel Frequency Cepstral Coefficients
Number of coefficients 13
Frame length 0.025
Frame stride 0.02
Filter number 32
FFT length 512
Normalization window size 151
Low frequency 80
High frequency Click to set
Pre-emphasis
Coefficient 0.98
Save parameters
DSP result
Cepstral Coefficients
Processed features
-1.3118, 0.6242, -1.1163, -0.3401, 0.3516, -0.5983, -0.0561, -1.6334, -2.4966, -1.03…
On-device performance
PROCESSING TIME
675 ms.
PEAK RAM USAGE
16 KB

Raw data -> 16,000 features



Processed features -> 637 features (13 x 49)



13

49

XIAO-ESP32S3-KWS - MFCC

studio.edgeimpulse.com/studio/230109/dsp/mfcc/3/generate-features

EDGE IMPULSE

MJRoBot (Marcelo Rovai) / XIAO-ESP32S3-KWS

- Dashboard
- Devices
- Data acquisition
- Impulse design
  - Create impulse
  - MFCC
  - Classifier
- EON Tuner
- Retrain model
- Live classification
- Model testing
- Performance calibration
- Versioning
- Deployment

GETTING STARTED

- Documentation
- Forums

#1 Click to set a description for this version

Parameters | Generate features

Training set

Data in training set	1h 20m 30s
Classes	4 (no, noise, unknown, yes)
Training windows	4,830

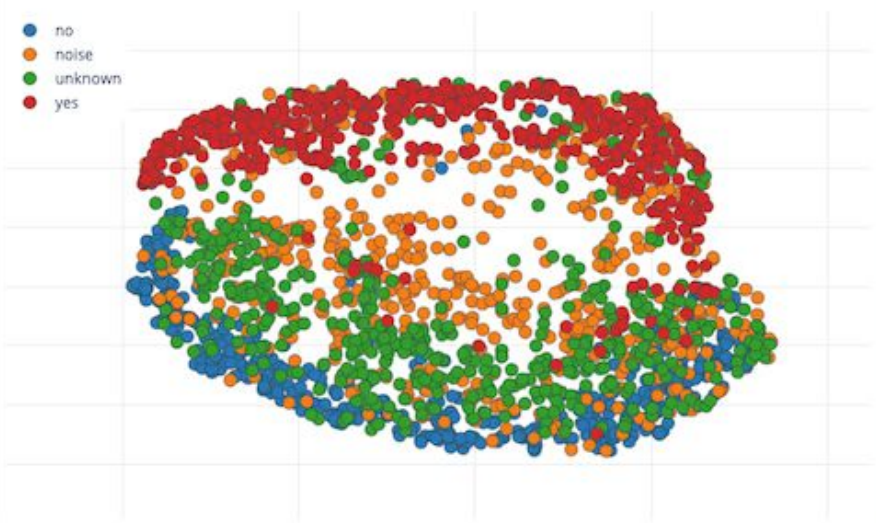
Generate features

Feature generation output

(0)

```
        completed  100  /  500 epochs
Still running...
        completed  150  /  500 epochs
        completed  200  /  500 epochs
        completed  250  /  500 epochs
Still running...
        completed  300  /  500 epochs
        completed  350  /  500 epochs
        completed  400  /  500 epochs
Still running...
        completed  450  /  500 epochs
Mon May 22 18:41:49 2023 Finished embedding
Reducing dimensions for visualizations OK

Job completed
```

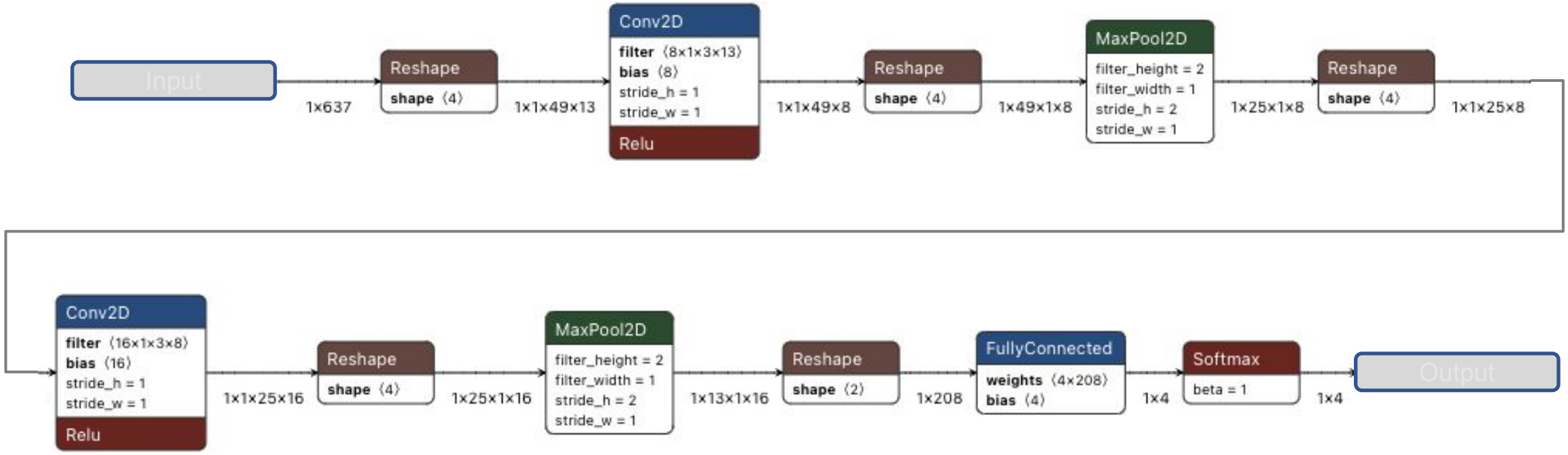
Feature explorer



On-device performance

PROCESSING TIME
675 ms.

PEAK RAM USAGE
16 KB


Input
1x637
Reshape
shape ⟨4⟩
1x1x49x13
Conv2D
filter ⟨8x1x3x13⟩
bias ⟨8⟩
stride_h = 1
stride_w = 1
Relu
1x1x49x8
Reshape
shape ⟨4⟩
1x49x1x8
MaxPool2D
filter_height = 2
filter_width = 1
stride_h = 2
stride_w = 1
1x25x1x8
Reshape
shape ⟨4⟩
1x1x25x8
Conv2D
filter ⟨16x1x3x8⟩
bias ⟨16⟩
stride_h = 1
stride_w = 1
Relu
1x1x25x16
Reshape
shape ⟨4⟩
1x25x1x16
MaxPool2D
filter_height = 2
filter_width = 1
stride_h = 2
stride_w = 1
1x13x1x16
Reshape
shape ⟨2⟩
1x208
FullyConnected
weights ⟨4x208⟩
bias ⟨4⟩
1x4
Softmax
beta = 1
1x4
Output

Model

Model version: Quantized (int8)

Last training performance (validation set)

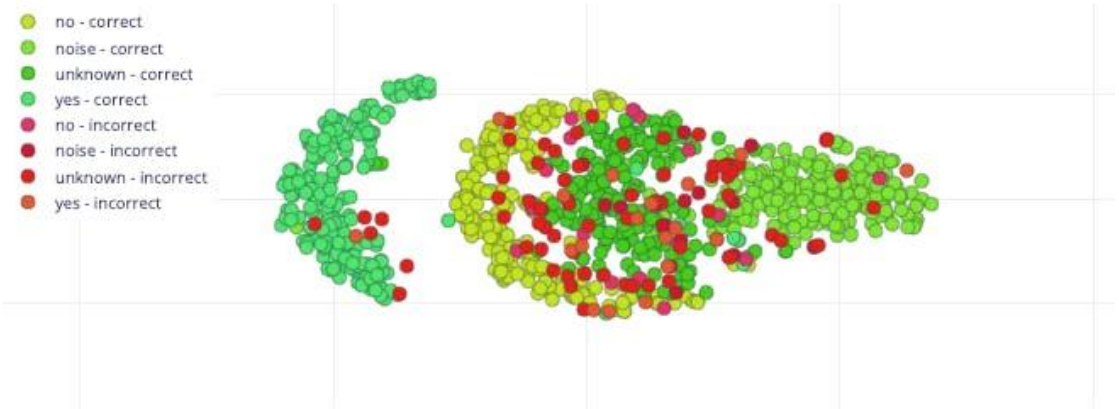
ACCURACY
90.7%

LOSS
0.25

Confusion matrix (validation set)

	NO	NOISE	UNKNOWN	YES
NO	92.2%	0.8%	5.3%	1.6%
NOISE	0.4%	95.2%	4.0%	0.4%
UNKNOWN	10.2%	5.1%	82.0%	2.7%
YES	2.1%	0.4%	3.3%	94.1%
F1 SCORE	0.90	0.94	0.85	0.95

Data explorer (full training set)

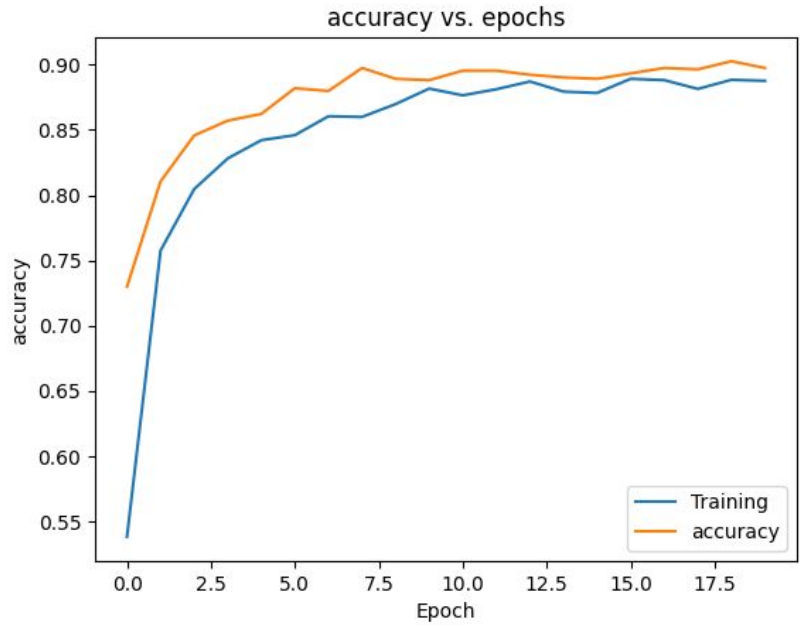


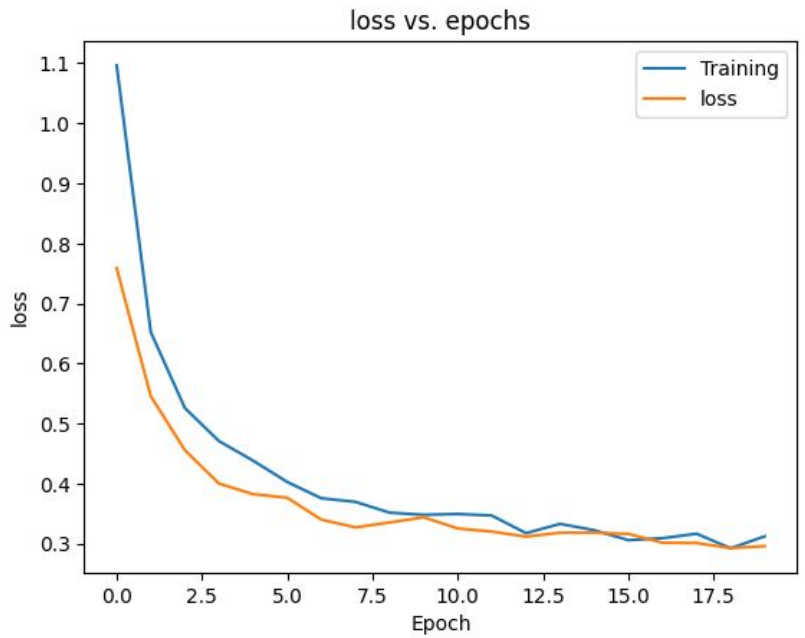
On-device performance

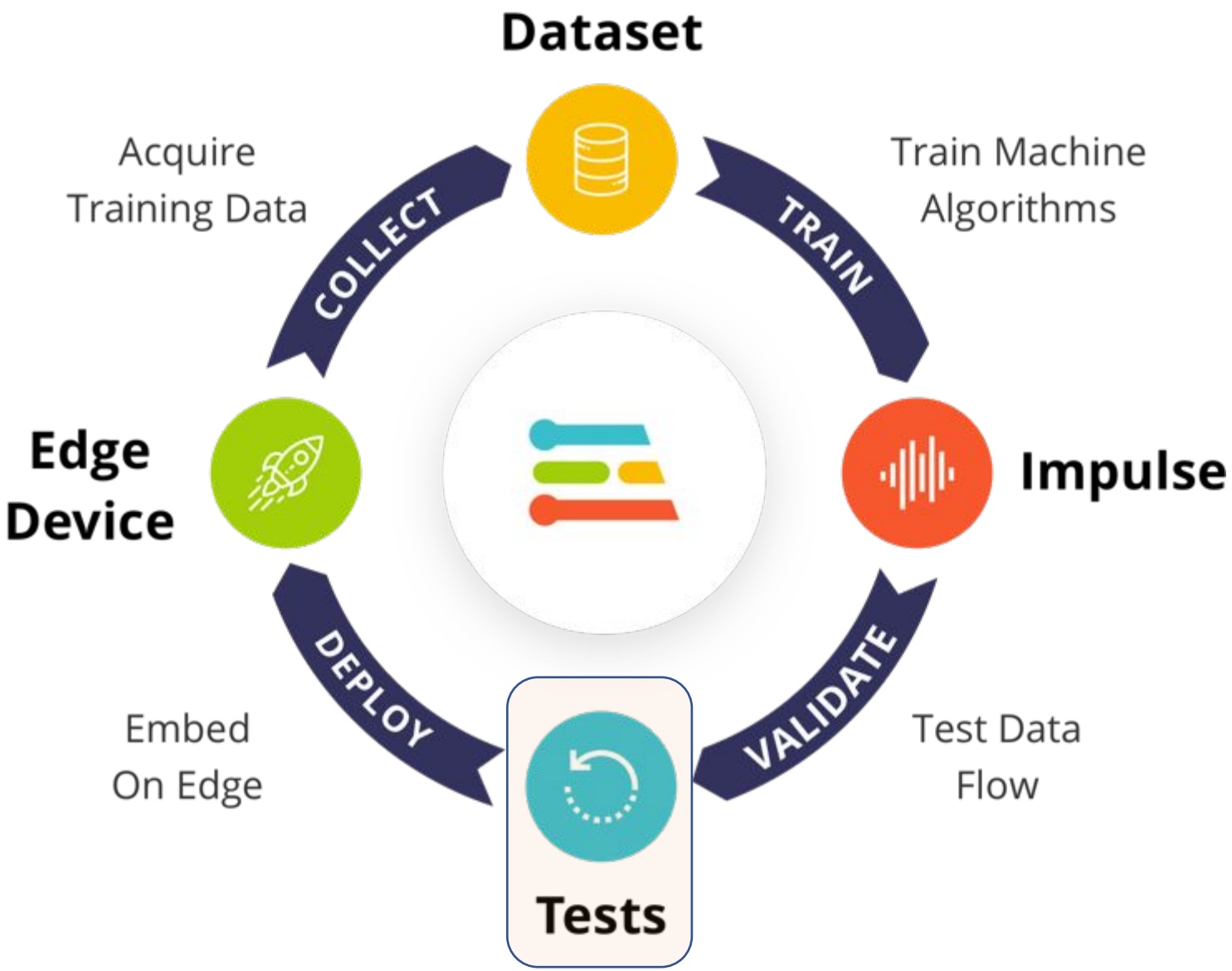
INFERENCING TIME
6 ms.

PEAK RAM USAGE
3.7K

FLASH USAGE
27.1K







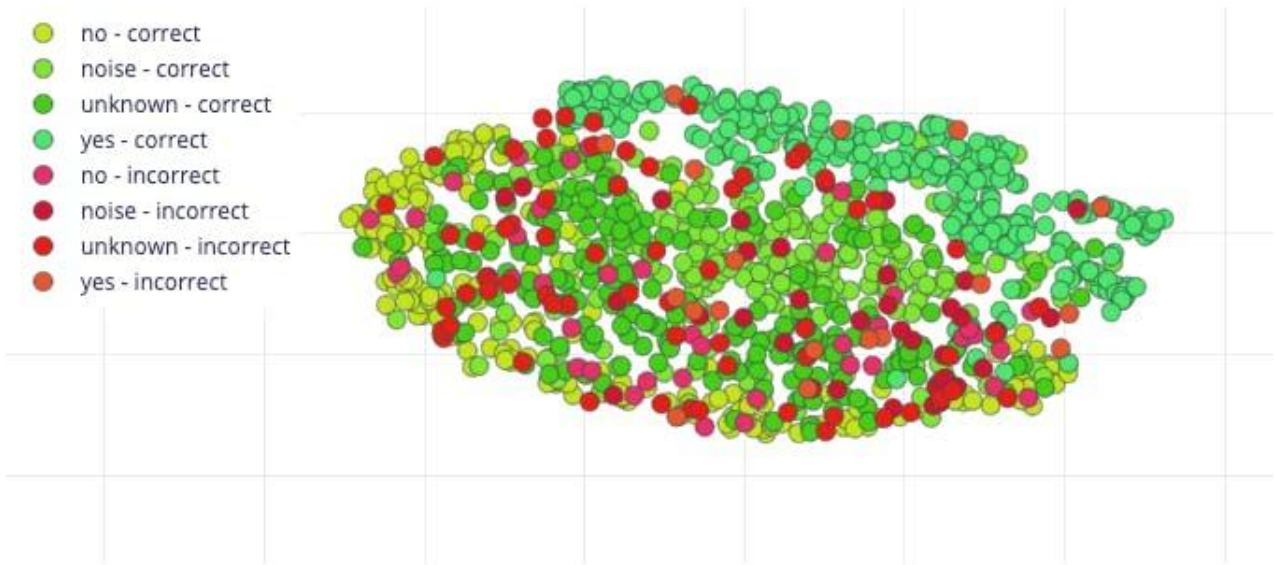
- Pre-Processing Data
- Design a Model
- Train a Model

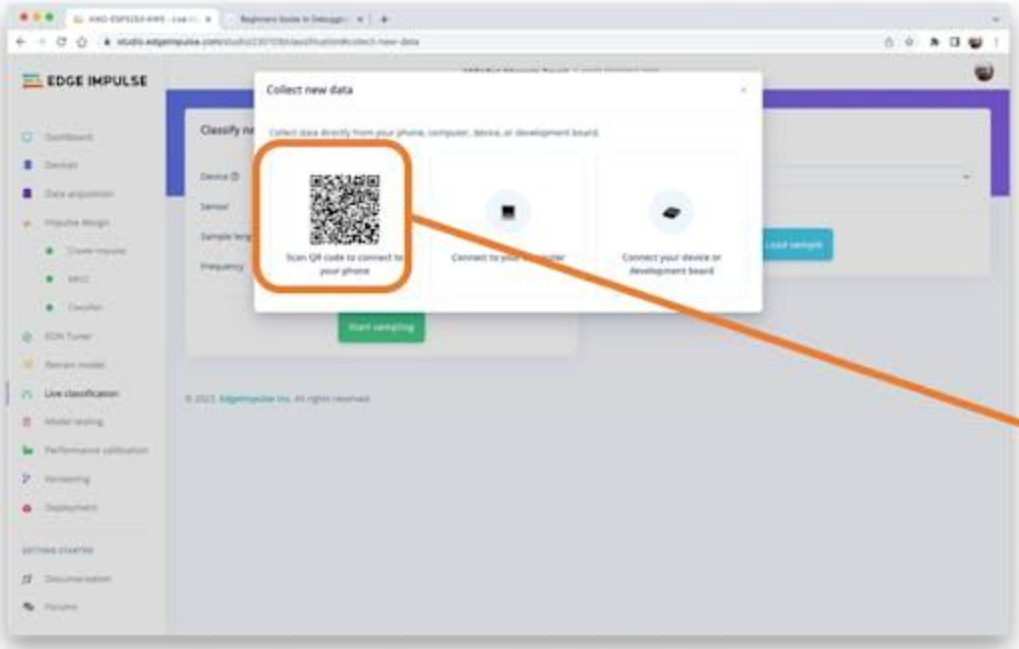
Model testing results

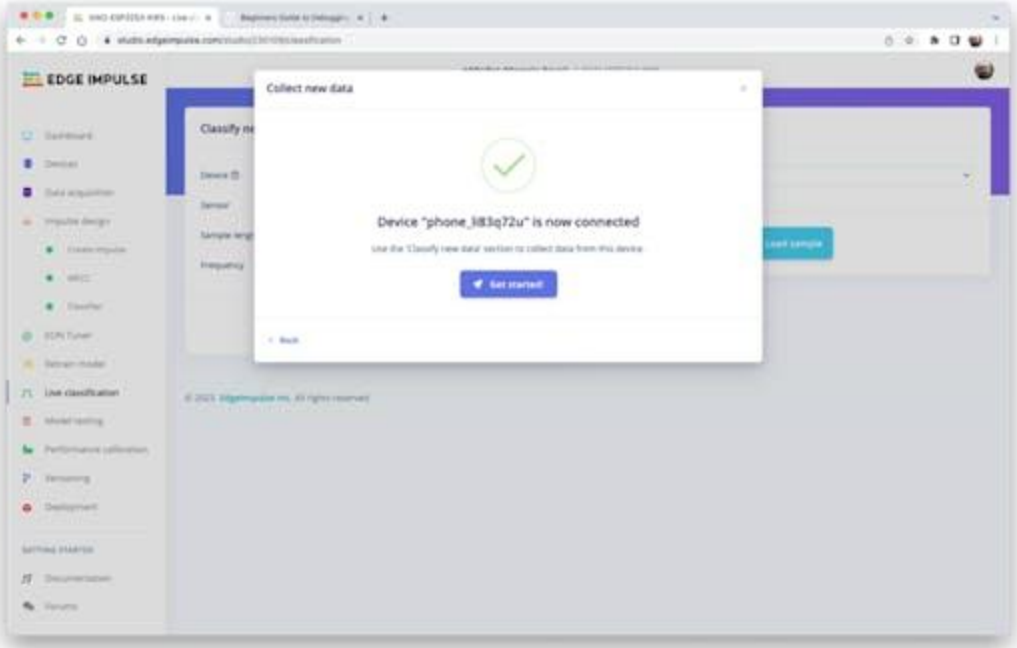
ACCURACY
86.73%

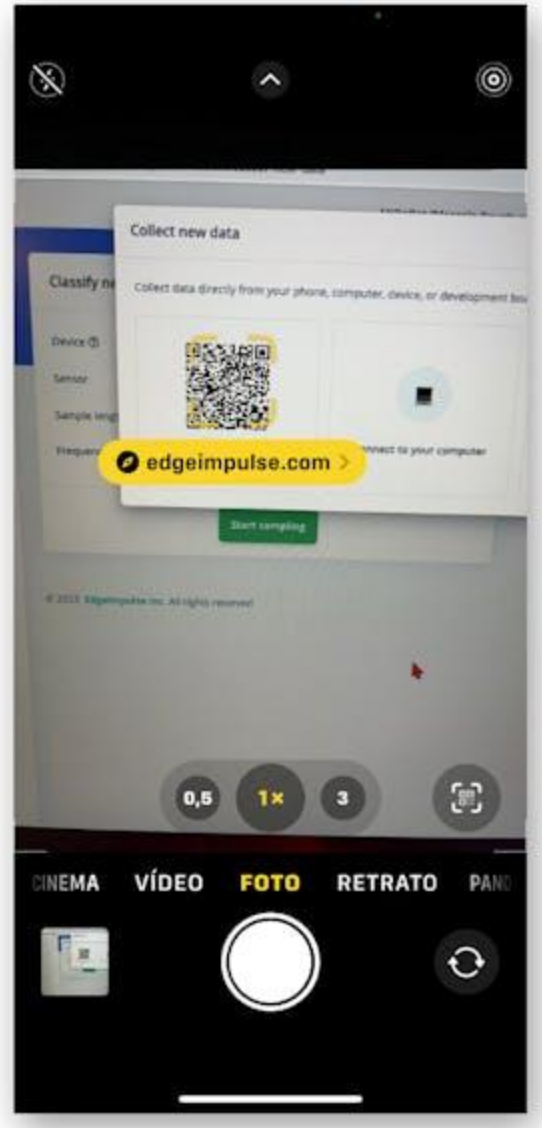
	NO	NOISE	UNKNOWN	YES	UNCERTAIN
NO	86.3%	0.7%	3.9%	1.4%	7.7%
NOISE	0%	88.6%	3.3%	0.7%	7.5%
UNKNOWN	4.4%	2.7%	78.1%	1.7%	13.1%
YES	0.3%	0%	0.7%	93.9%	5.1%
F1 SCORE	0.90	0.92	0.84	0.95	

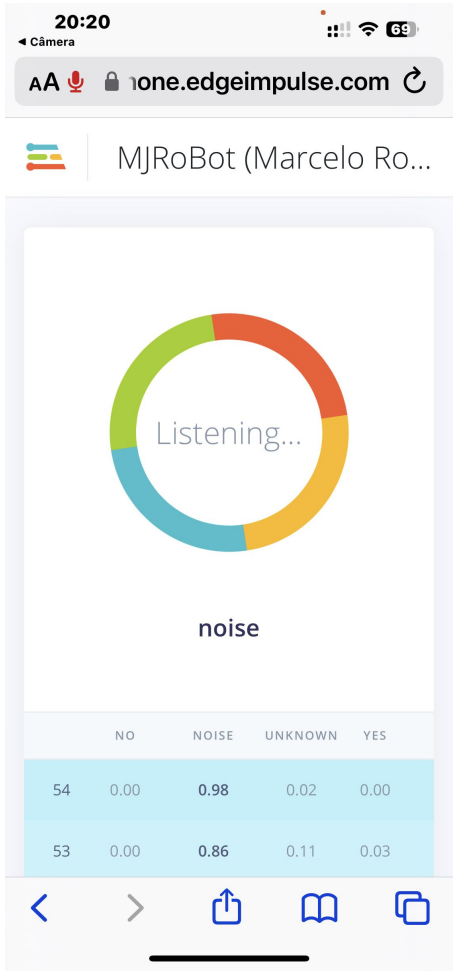
Feature explorer

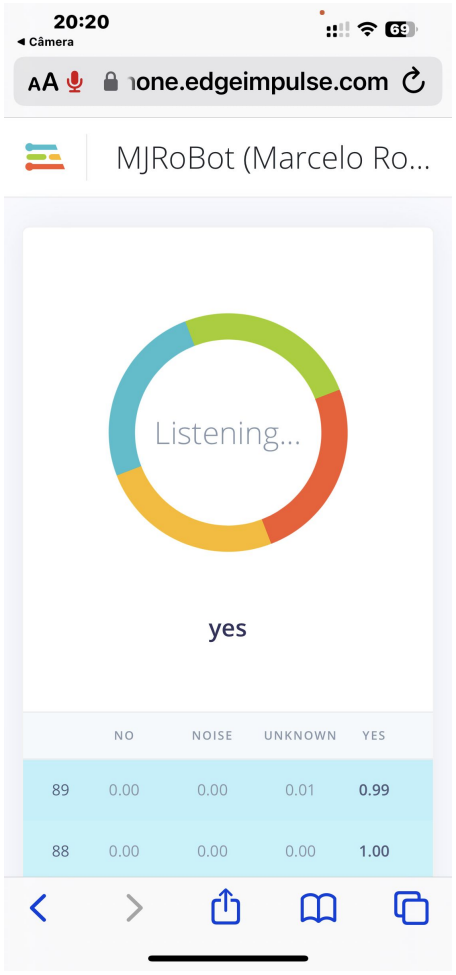


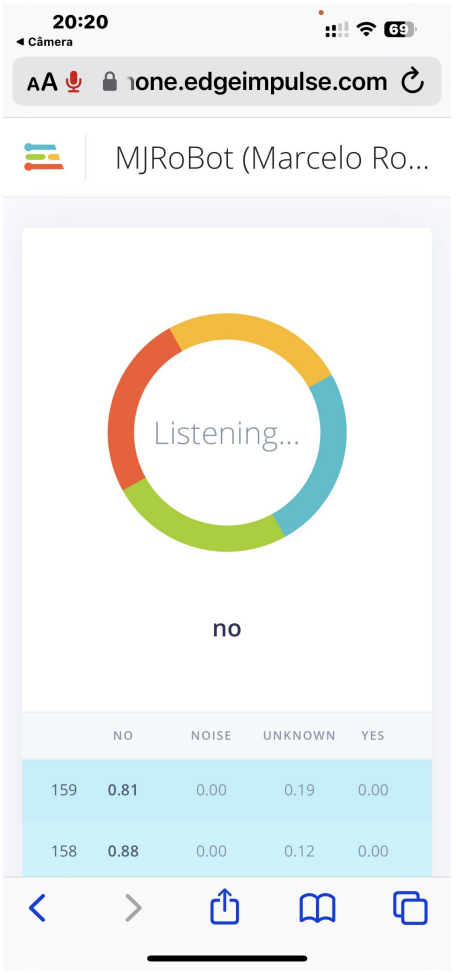

EDGE IMPULSE
Collect new data
Scan QR code to connect to your phone
Connect your device or development board

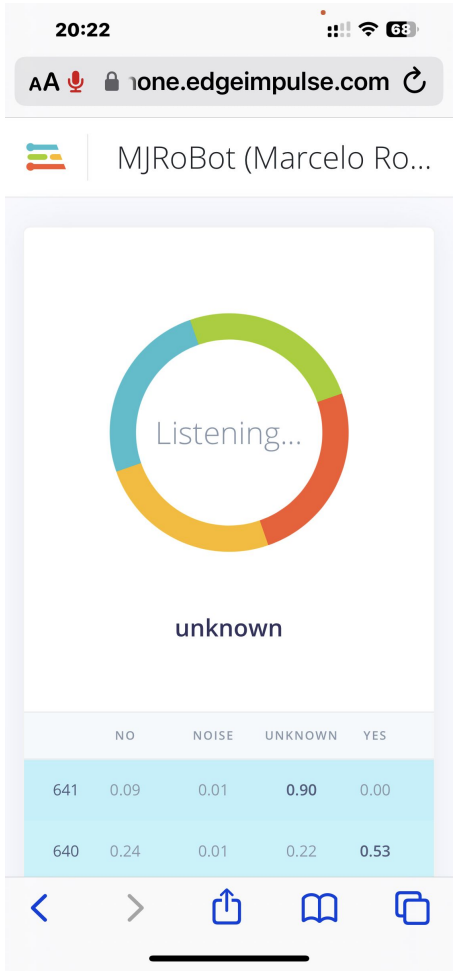

EDGE IMPULSE
Collect new data
Device "phone_l83q72u" is now connected
Get started

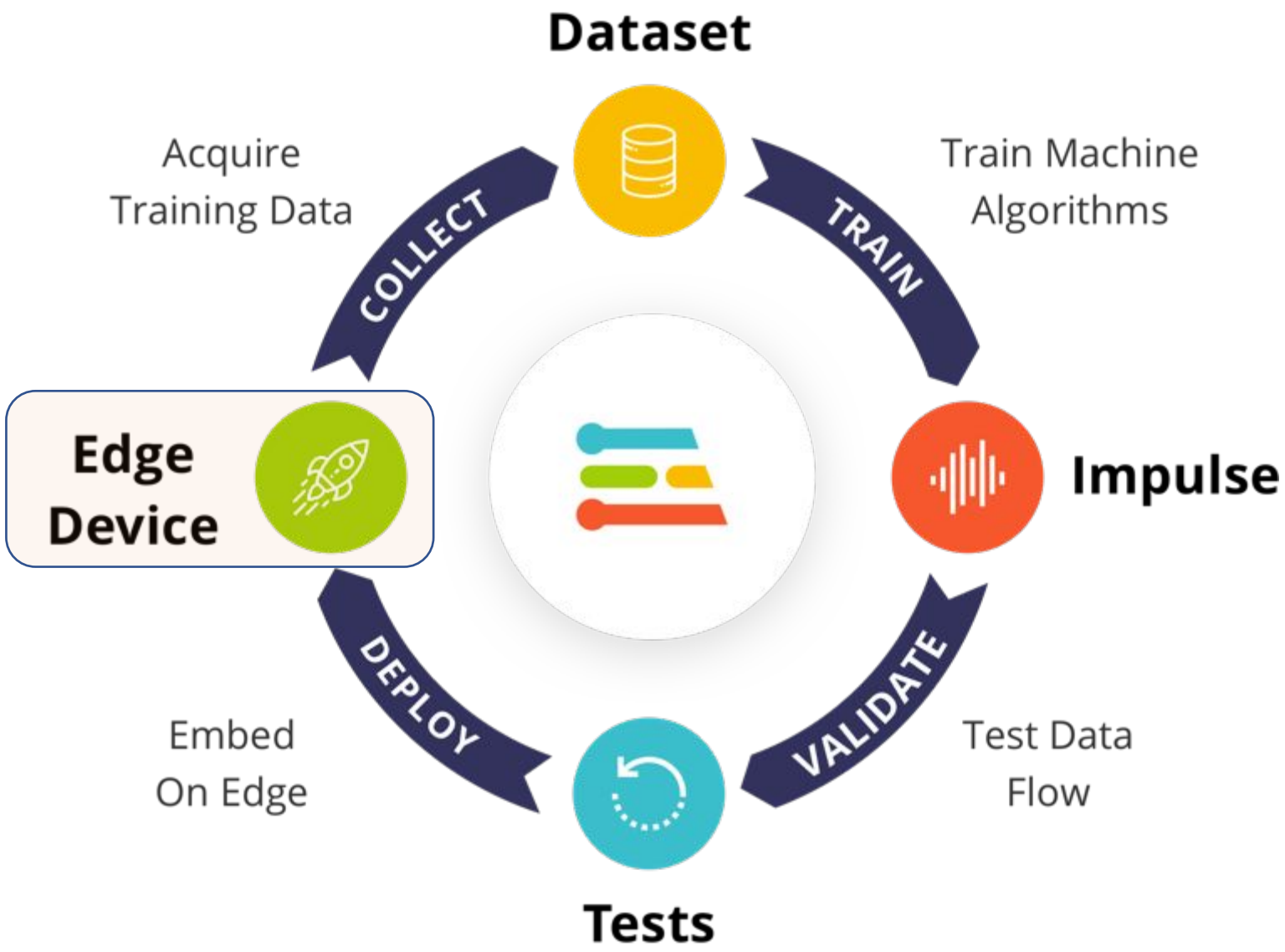

Collect new data
edgeimpulse.com
0,5
1×
3
CINEMA
VÍDEO
FOTO
RETRATO
PANO


20:20
◂ Câmera
AA
one.edgeimpulse.com
MJRoBot (Marcelo Ro...
Listening...
noise
NO NOISE UNKNOWN YES
54 0.00 0.98 0.02 0.00
53 0.00 0.86 0.11 0.03

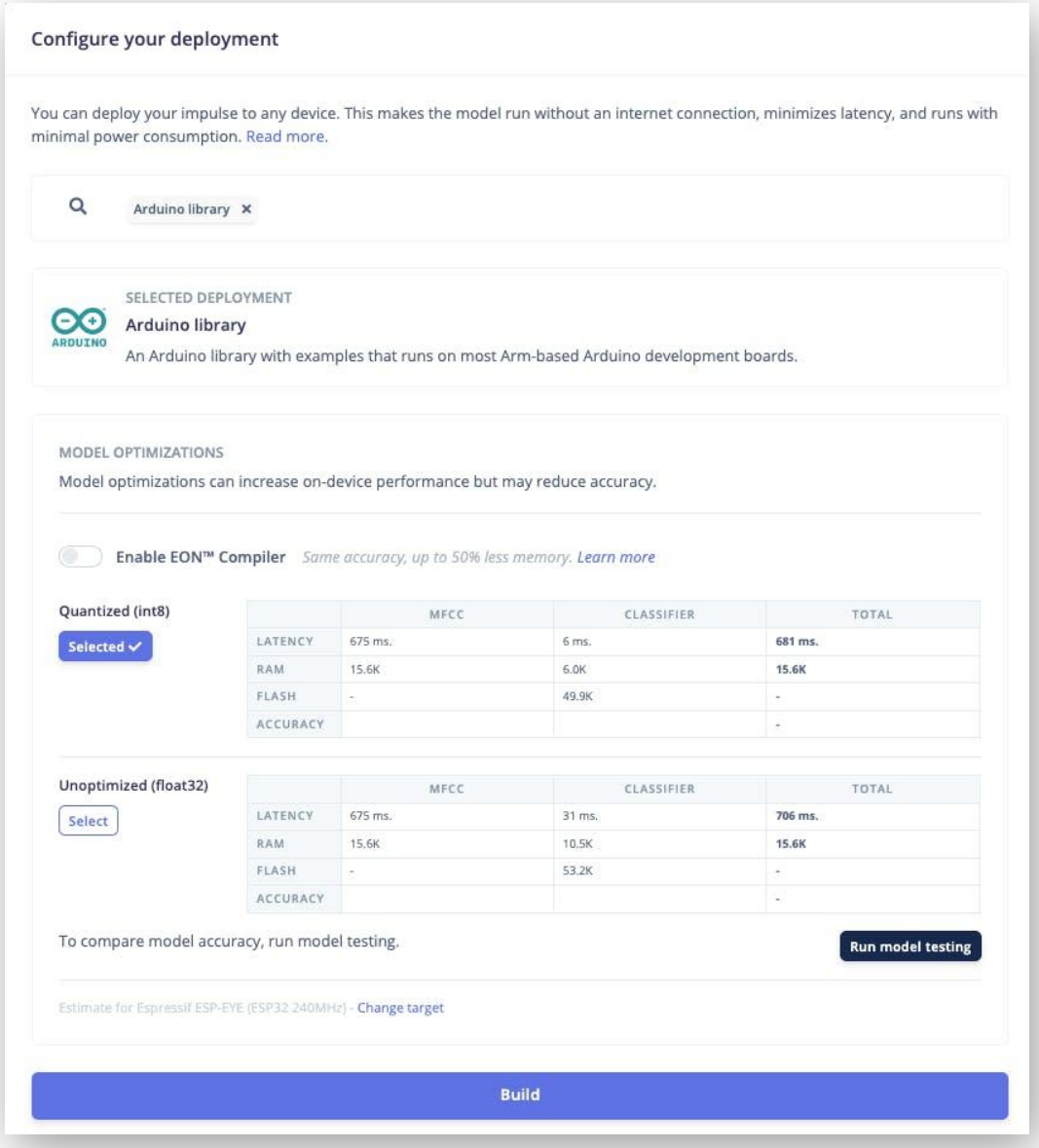

20:20
◂ Câmera
AA
one.edgeimpulse.com
MJRoBot (Marcelo Ro...
Listening...
yes
NO NOISE UNKNOWN YES
89 0.00 0.00 0.01 0.99
88 0.00 0.00 0.00 1.00

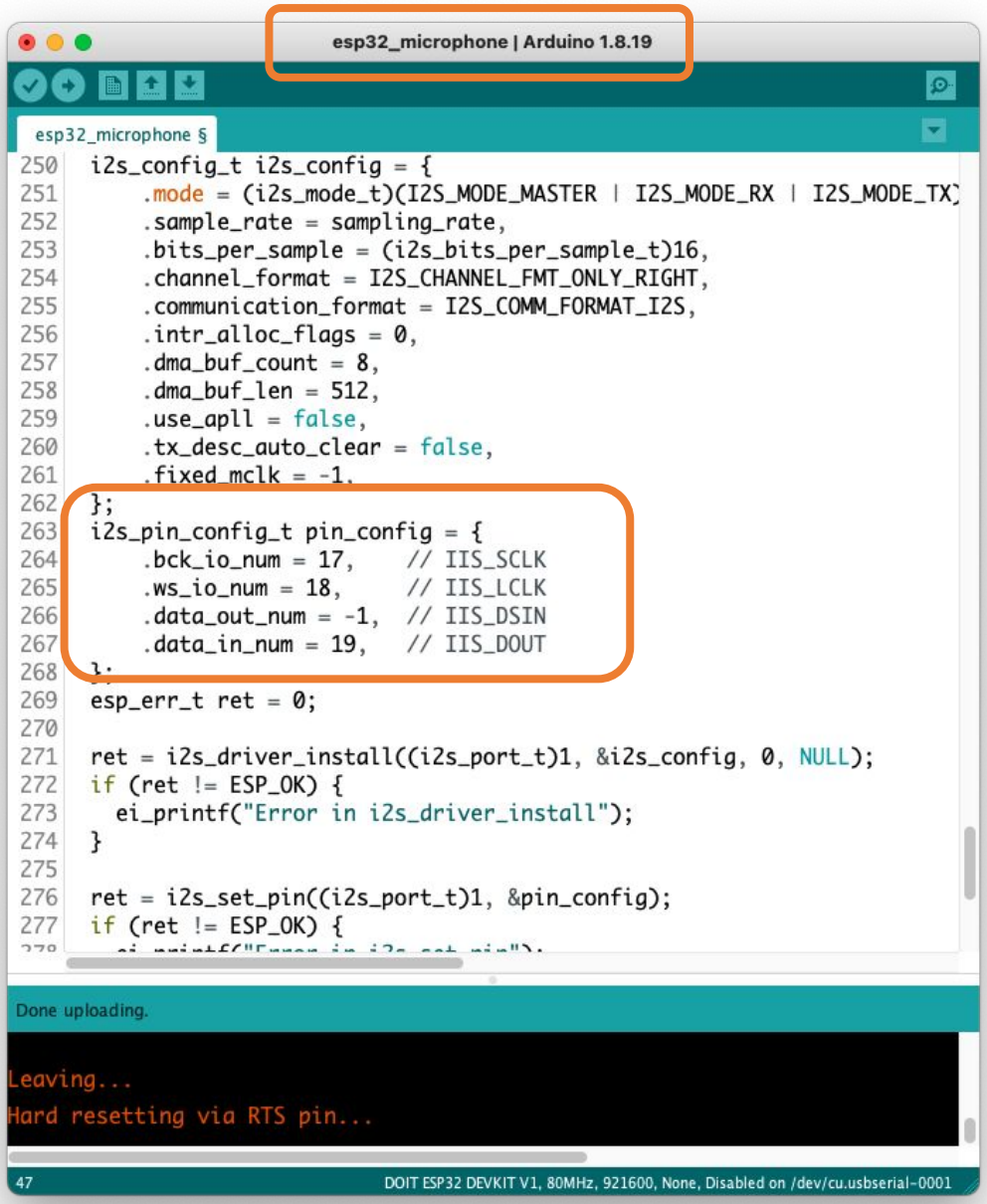

20:20
◂ Câmera
AA
one.edgeimpulse.com
MJRoBot (Marcelo Ro...
Listening...
no
NO NOISE UNKNOWN YES
159 0.81 0.00 0.19 0.00
158 0.88 0.00 0.12 0.00


20:22
AA
one.edgeimpulse.com
MJRoBot (Marcelo Ro...
Listening...
unknown
NO NOISE UNKNOWN YES
641 0.09 0.01 0.90 0.00
640 0.24 0.01 0.22 0.53

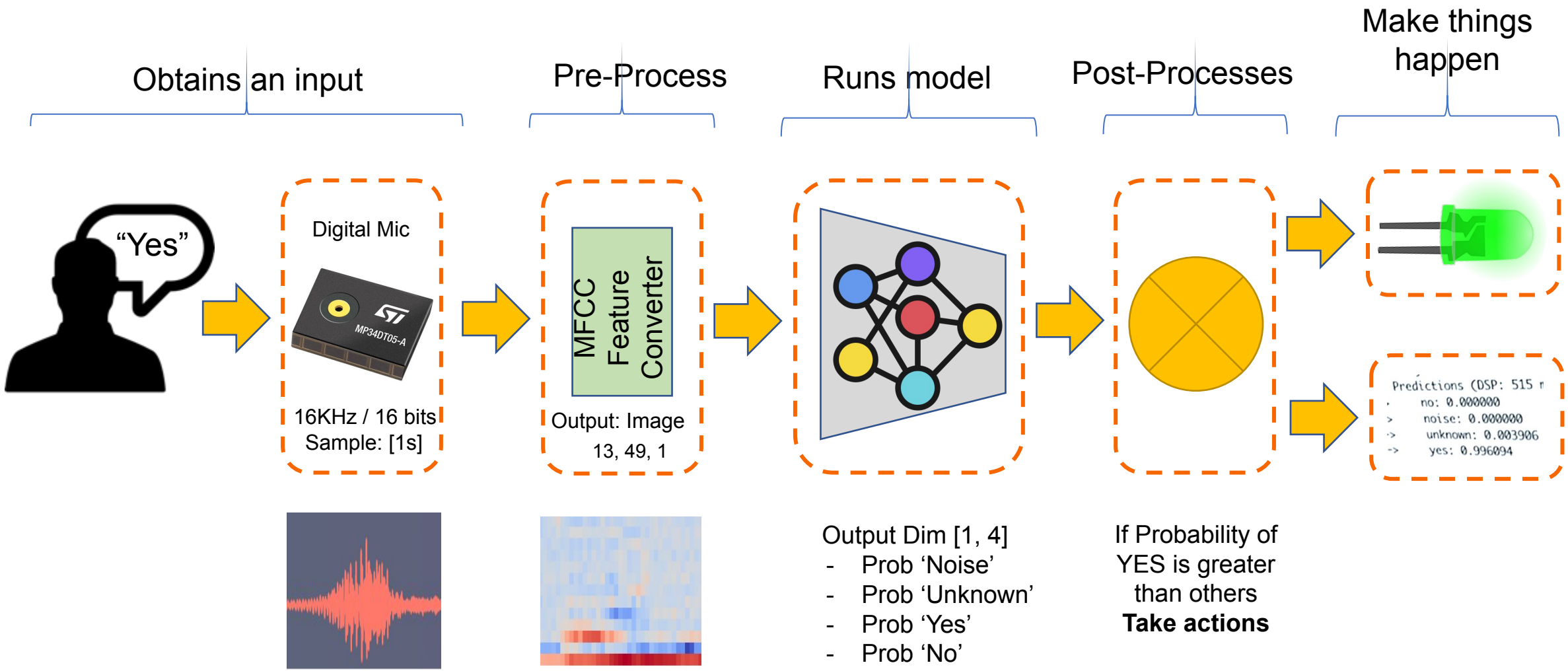


- Pre-Processing Data
- Design a Model
- Train a Model


Configure your deployment
You can deploy your impulse to any device. This makes the model run without an internet connection, minimizes latency, and runs with minimal power consumption. Read more.
Arduino library
SELECTED DEPLOYMENT
Arduino library
An Arduino library with examples that runs on most Arm-based Arduino development boards.
MODEL OPTIMIZATIONS
Model optimizations can increase on-device performance but may reduce accuracy.
Enable EON™ Compiler Same accuracy, up to 50% less memory. Learn more
Quantized (int8)
Selected
MFCC CLASSIFIER TOTAL
LATENCY 675 ms. 6 ms. 681 ms.
RAM 15.6K 6.0K 15.6K
FLASH - 49.9K -
ACCURACY -
Unoptimized (float32)
Select
MFCC CLASSIFIER TOTAL
LATENCY 675 ms. 31 ms. 706 ms.
RAM 15.6K 10.5K 15.6K
FLASH - 53.2K -
ACCURACY -
To compare model accuracy, run model testing.
Run model testing
Estimate for Espressif ESP-EYE (ESP32 240MHz) · Change target
Build


esp32_microphone | Arduino 1.8.19
esp32_microphone §
250 i2s_config_t i2s_config = {
251 .mode = (i2s_mode_t)(I2S_MODE_MASTER | I2S_MODE_RX | I2S_MODE_TX)
252 .sample_rate = sampling_rate,
253 .bits_per_sample = (i2s_bits_per_sample_t)16,
254 .channel_format = I2S_CHANNEL_FMT_ONLY_RIGHT,
255 .communication_format = I2S_COMM_FORMAT_I2S,
256 .intr_alloc_flags = 0,
257 .dma_buf_count = 8,
258 .dma_buf_len = 512,
259 .use_apll = false,
260 .tx_desc_auto_clear = false,
261 .fixed_mclk = -1,
262 };
263 i2s_pin_config_t pin_config = {
264 .bck_io_num = 17, // IIS_SCLK
265 .ws_io_num = 18, // IIS_LCLK
266 .data_out_num = -1, // IIS_DSIN
267 .data_in_num = 19, // IIS_DOUT
268 };
269 esp_err_t ret = 0;
270
271 ret = i2s_driver_install((i2s_port_t)1, &i2s_config, 0, NULL);
272 if (ret != ESP_OK) {
273 ei_printf("Error in i2s_driver_install");
274 }
275
276 ret = i2s_set_pin((i2s_port_t)1, &pin_config);
277 if (ret != ESP_OK) {
Done uploading.
Leaving...
Hard resetting via RTS pin...
47
DOIT ESP32 DEVKIT V1, 80MHz, 921600, None, Disabled on /dev/cu.usbserial-0001

KeyWord Spotting (KWS) - Inference



/dev/cu.usbserial-0001

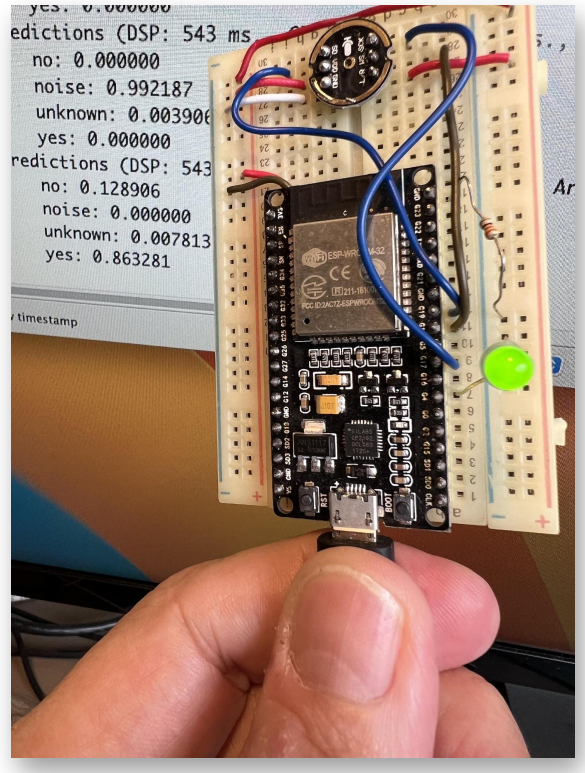
Send

```
13:22:34.389 ->     yes: 0.007813
13:22:35.354 -> Predictions (DSP: 542 ms., Classification: 4 ms., Anomaly: 0 ms.):
13:22:35.354 ->     no: 0.585937
13:22:35.354 ->     noise: 0.000000
13:22:35.354 ->     unknown: 0.011719
13:22:35.354 ->     yes: 0.402344
13:22:36.375 -> Predictions (DSP: 542 ms., Classification: 4 ms., Anomaly: 0 ms.):
13:22:36.409 ->     no: 0.015625
13:22:36.409 ->     noise: 0.921875
13:22:36.409 ->     unknown: 0.054687
13:22:36.409 ->     yes: 0.007813
13:22:37.331 -> Predictions (DSP: 542 ms., Classification: 4 ms., Anomaly: 0 ms.):
13:22:37.364 ->     no: 0.000000
13:22:37.364 ->     noise: 0.996094
13:22:37.364 ->     unknown: 0.003906
13:22:37.364 ->     yes: 0.000000
13:22:38.359 -> Predictions (DSP: 542 ms., Classification: 4 ms., Anomaly: 0 ms.):
13:22:38.359 ->     no: 0.003906
13:22:38.359 ->     noise: 0.000000
13:22:38.359 ->     unknown: 0.003906
13:22:38.359 ->     yes: 0.988281
```

Autoscroll Show timestamp Both NL & CR 115200 baud Clear output

```
/* Includes ---------------------------
#include <ESP32-NMP441-KWS_inferencing.h>

#include "freertos/FreeRTOS.h"
#include "freertos/task.h"

#include "driver/i2s.h"

#define LED_BUILT_IN 4


void setup()
{
    // put your setup code here, to run once:
    Serial.begin(115200);
    // comment out the below line to cancel the wait
    while (!Serial);
    Serial.println("Edge Impulse Inferencing Demo");

    pinMode(LED_BUILT_IN, OUTPUT);
```

```
    int pred_index = 0;     // Initialize pred_index
    float pred_value = 0;   // Initialize pred_value

    // print the predictions
    ei_printf("Predictions ");
    ei_printf("(DSP: %d ms., Classification: %d ms., Anomaly: %d ms.)",
         result.timing.dsp, result.timing.classification, result.timing.anomaly);
    ei_printf(": \n");
    for (size_t ix = 0; ix < EI_CLASSIFIER_LABEL_COUNT; ix++) {
          ei_printf("    %s: ", result.classification[ix].label);
          ei_printf_float(result.classification[ix].value);
          ei_printf("\n");

          if (result.classification[ix].value > pred_value){
             pred_index = ix;
             pred_value = result.classification[ix].value;
          }
    }

    // show the inference result on LED
    if (pred_index == 3){
        digitalWrite(LED_BUILT_IN, HIGH); //Turn on
    }
    else{
       digitalWrite(LED_BUILT_IN, LOW); //Turn off
    }
```



Thanks

UNIFEI