How to Use a Model for Inference

EDGE IMPULSE

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

GETTING STARTED

Try Enterprise Free

Get access to high job limits and training on GPUs.

Start free trial

Project info | Keys | Export | Jobs

Fabio / my-smartphone-motion-project

This is your Edge Impulse project. From here you acquire new training data, design impulses and train models.

ACCELEROMETER | + New tag

Getting started

Start building your dataset or validate your model's on-device performance:

Add existing data

Collect new data

Upload your model

Start with a tutorial

Not sure where to start? Follow a tutorial to build your first model in just minutes!

Motion: Gesture recognition

Images: Object detection

Audio: Audio classification

Sharing

Your project is private.

Make this project public

Run this model

Scan QR code or launch in browser

Launch in browser

Download block output

TITLE	TYPE	SIZE	
Spectral features training data	NPY file	6161 windows	
Spectral features training labels	NPY file	6161 windows	
Classifier model	TensorFlow Lite (float32)	5 KB	
Classifier model	TensorFlow Lite (int8 quantized)	3 KB	
Classifier model	TensorFlow SavedModel	11 KB	
Classifier model	Keras h5 model	5 KB	

Performance settings

Use GPU for training

Enterprise performance

Job limit in minutes: 20

Train job memory (MB): 8192

DSP file size limit (MB): 4096

Administrative zone

Summary

DEVICES CONNECTED
1

DATA COLLECTED
13m 20s

Project info

Project ID	265765
Labeling method	One label per data item
Target device	Cortex-M4F 80MHz

Download block output

TITLE	TYPE	SIZE	
Spectral features training data	NPY file	6161 windows	
Spectral features training labels	NPY file	6161 windows	
Classifier model	TensorFlow Lite (float32)	5 KB	
Classifier model	TensorFlow Lite (int8 quantized)	3 KB	
Classifier model	TensorFlow SavedModel	11 KB	
Classifier model	Keras h5 model	5 KB	

TensorFlow Lite (int8 quantized)

Performance settings

Use GPU for training

Enterprise performance

Job limit in minutes: 20

Train job memory (MB): 8192

DSP file size limit (MB): 4096

Administrative zone

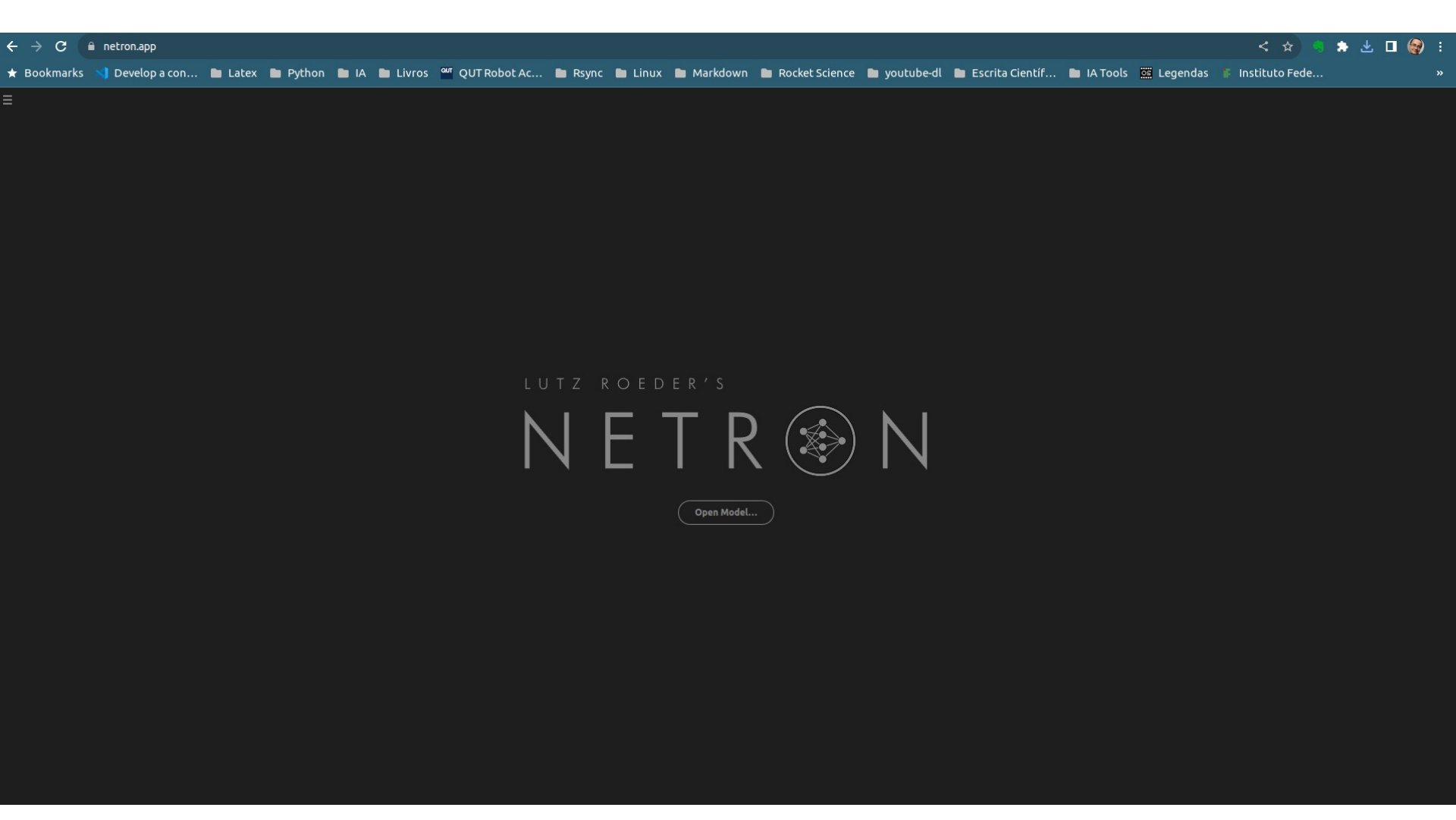
Summary

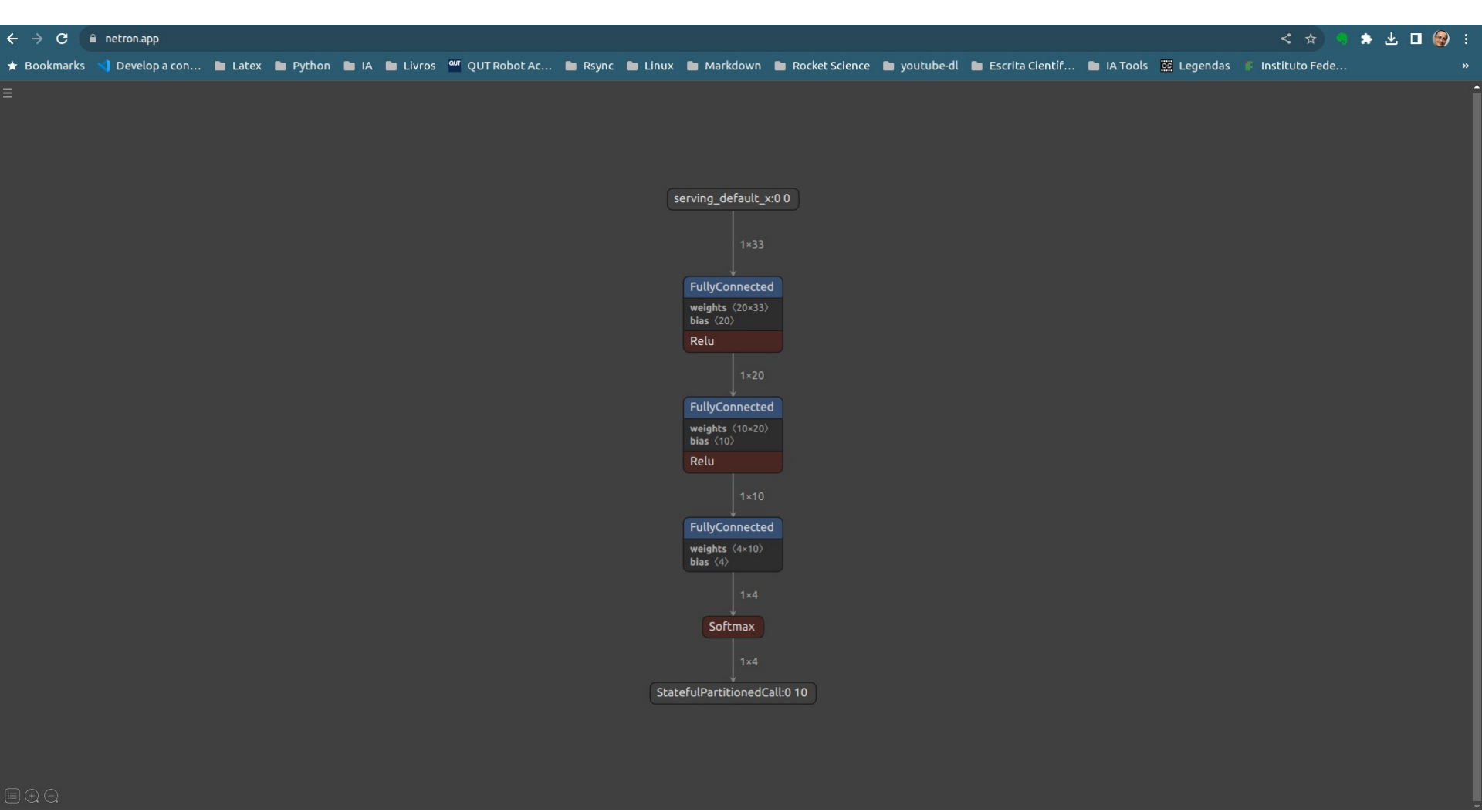
DEVICES CONNECTED
1

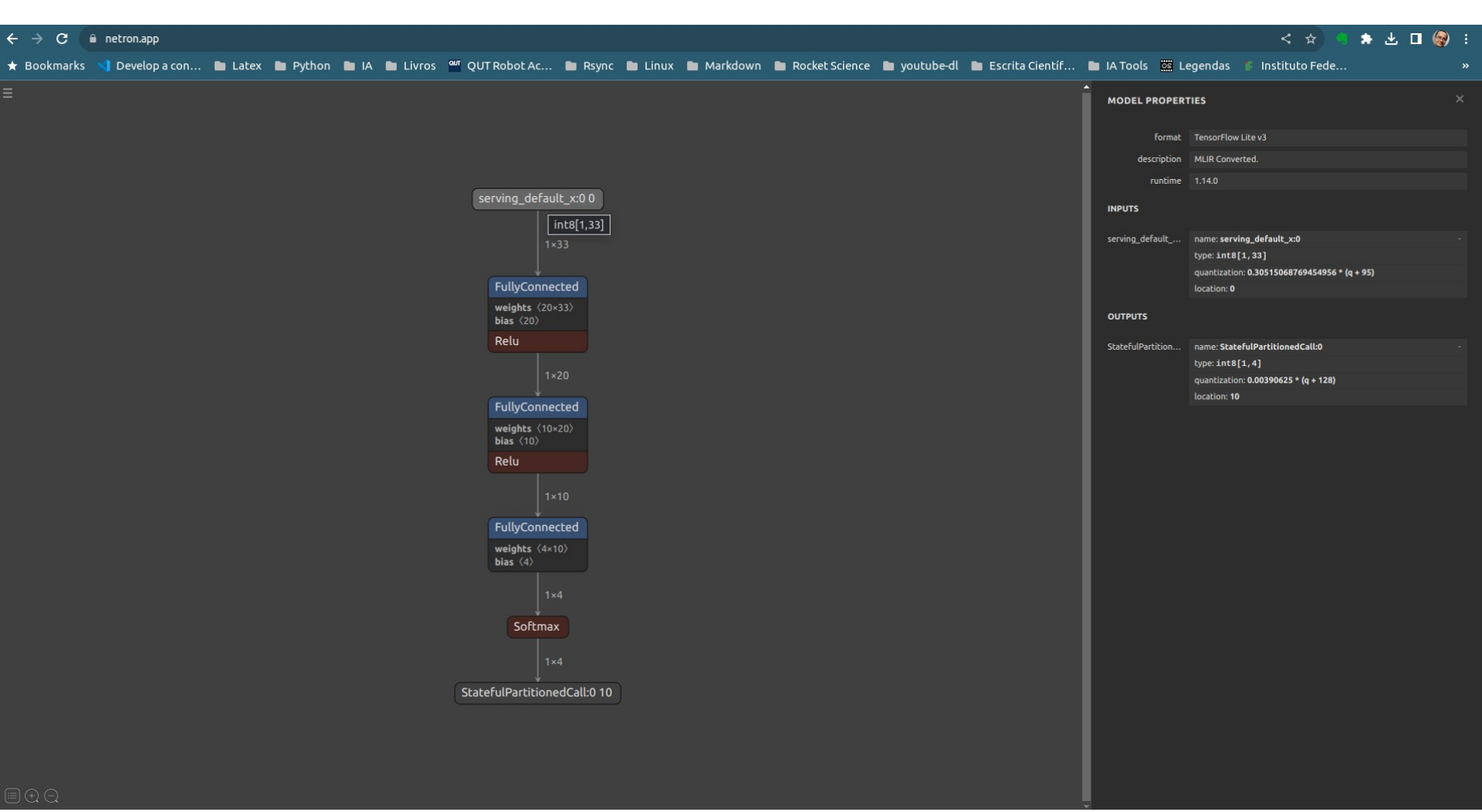
DATA COLLECTED
13m 20s

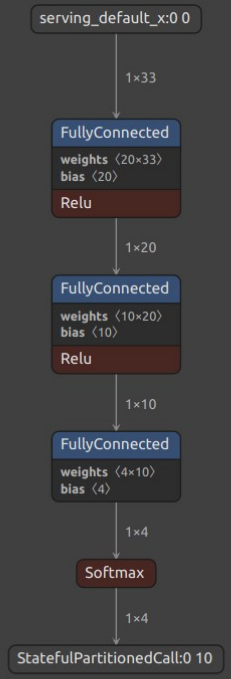
Project info

Project ID	265765
Labeling method	One label per data item
Target device	Cortex-M4F 80MHz


netron.app
Bookmarks
Develop a con...
Latex
Python
IA
Livros
QUT Robot Ac...
Rsync
Linux
Markdown
Rocket Science
youtube-dl
Escrita Científ...
IA Tools
Legendas
Instituto Fede...
LUTZ ROEDER'S
NETRON
Open Model...


netron.app
Bookmarks
Develop a con...
Latex
Python
IA
Livros
QUT Robot Ac...
Rsync
Linux
Markdown
Rocket Science
youtube-dl
Escrita Científ...
IA Tools
Legendas
Instituto Fede...
serving_default_x:0 0
1×33
FullyConnected
weights ⟨20×33⟩
bias ⟨20⟩
Relu
1×20
FullyConnected
weights ⟨10×20⟩
bias ⟨10⟩
Relu
1×10
FullyConnected
weights ⟨4×10⟩
bias ⟨4⟩
1×4
Softmax
1×4
StatefulPartitionedCall:0 10


netron.app
Bookmarks
Develop a con…
Latex
Python
IA
Livros
QUT Robot Ac…
Rsync
Linux
Markdown
Rocket Science
youtube-dl
Escrita Científ…
IA Tools
Legendas
Instituto Fede…
serving_default_x:0 0
int8[1,33]
1×33
FullyConnected
weights ⟨20×33⟩
bias ⟨20⟩
Relu
1×20
FullyConnected
weights ⟨10×20⟩
bias ⟨10⟩
Relu
1×10
FullyConnected
weights ⟨4×10⟩
bias ⟨4⟩
1×4
Softmax
1×4
StatefulPartitionedCall:0 10
MODEL PROPERTIES
format
TensorFlow Lite v3
description
MLIR Converted.
runtime
1.14.0
INPUTS
serving_default_…
name: serving_default_x:0
type: int8[1,33]
quantization: 0.30515068769454956 * (q + 95)
location: 0
OUTPUTS
StatefulPartition…
name: StatefulPartitionedCall:0
type: int8[1,4]
quantization: 0.00390625 * (q + 128)
location: 10



MODEL PROPERTIES

format	TensorFlow Lite v3
description	MLIR Converted.
runtime	1.14.0

INPUTS

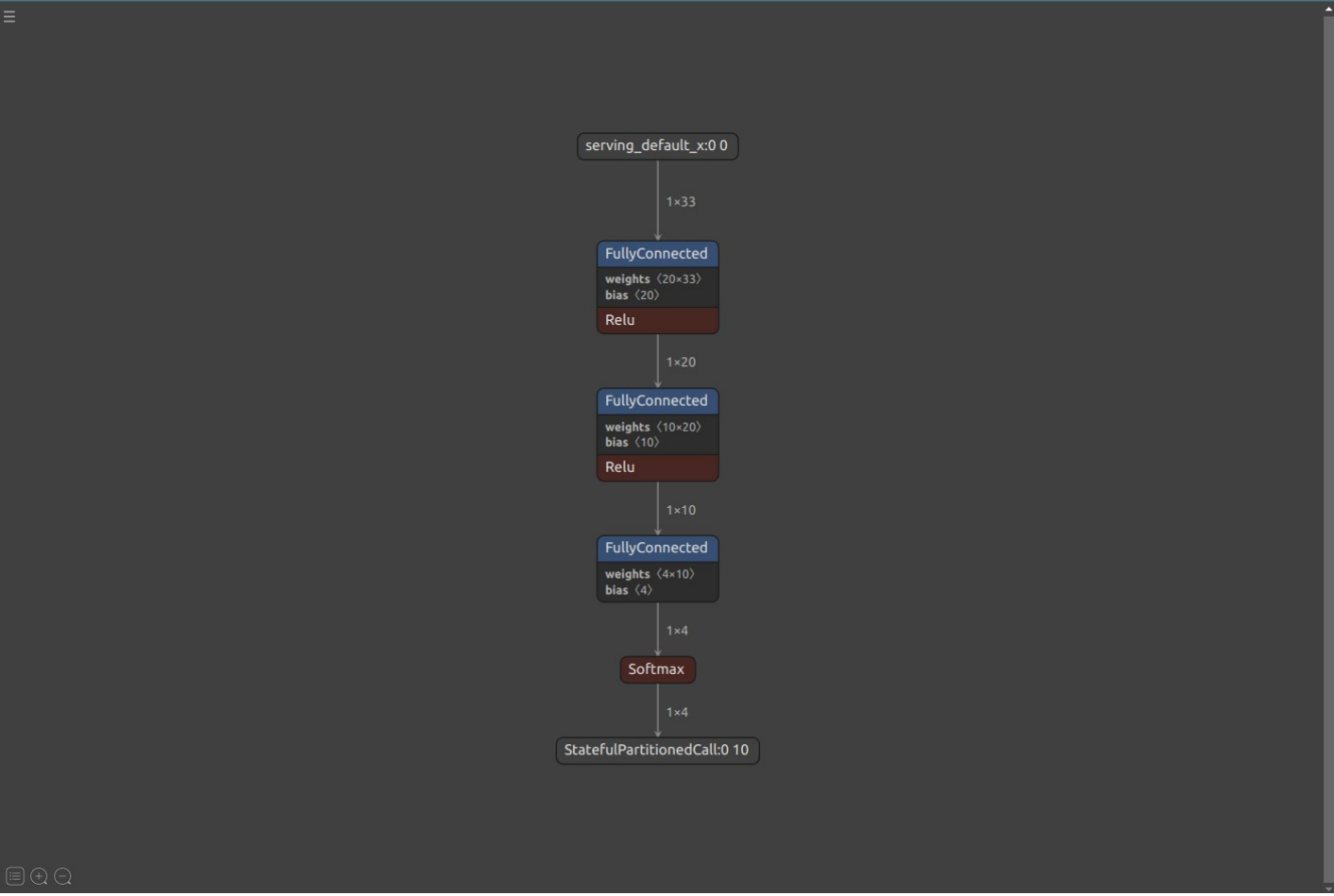
serving_default_…

- name: **serving_default_x:0**
- type: **int8[1, 33]**
- quantization: **0.30515068769454956 * (q + 95)**
- location: **0**

OUTPUTS

StatefulPartition…

- name: **StatefulPartitionedCall:0**
- type: **int8[1, 4]**
- quantization: **0.00390625 * (q + 128)**
- location: **10**



MODEL PROPERTIES

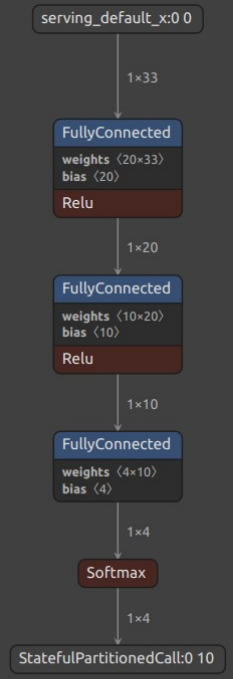
format	TensorFlow Lite v3
description	MLIR Converted.
runtime	1.14.0

INPUTS

serving_default_...	name: **serving_default_x:0** type: **int8[1,33]** quantization: **0.30515068769454956 * (q + 95)** location: **0**

OUTPUTS

StatefulPartition...	name: **StatefulPartitionedCall:0** type: **int8[1,4]** quantization: **0.00390625 * (q + 128)** location: **10**



MODEL PROPERTIES

format	TensorFlow Lite v3
description	MLIR Converted.
runtime	1.14.0

INPUTS

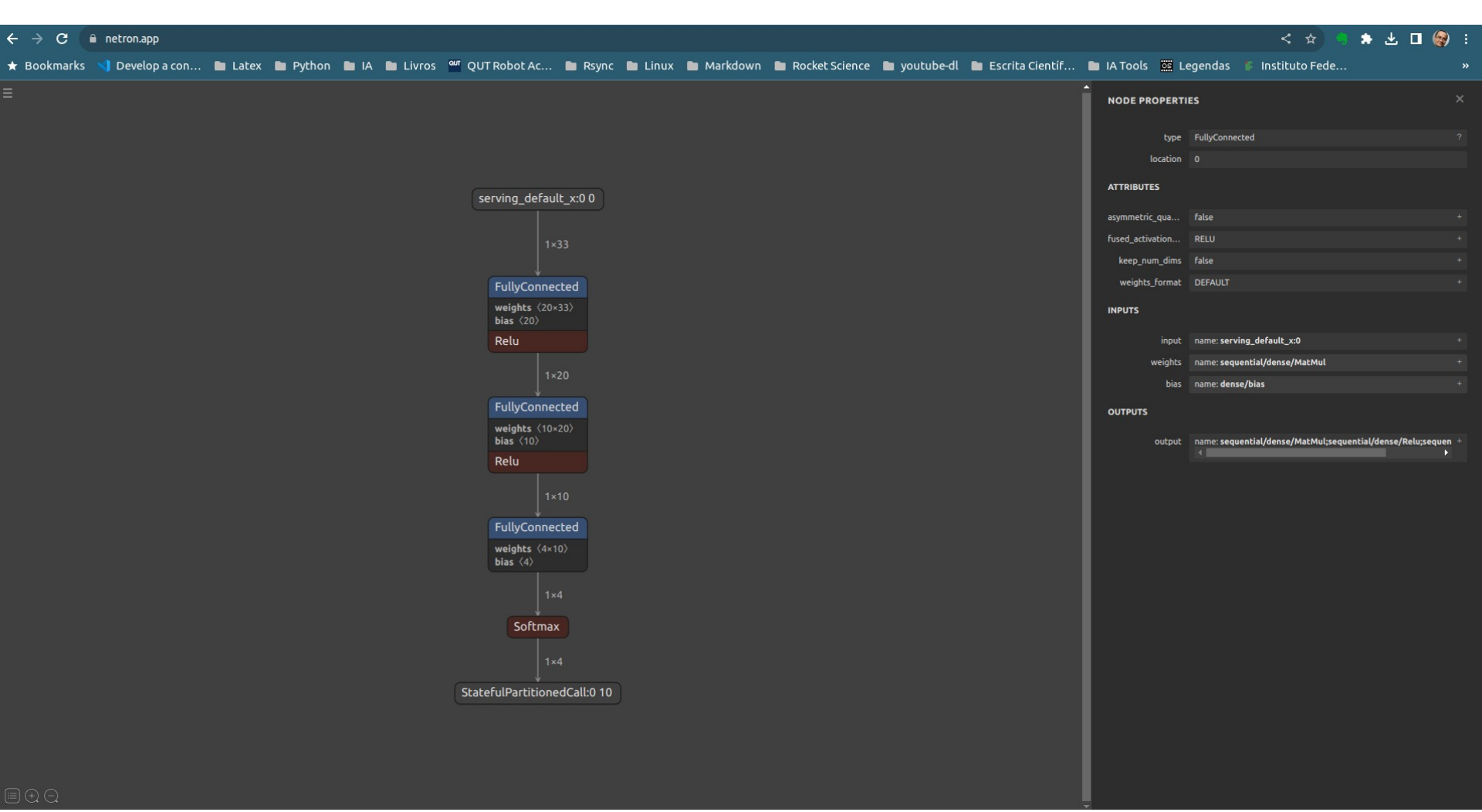
serving_default_...

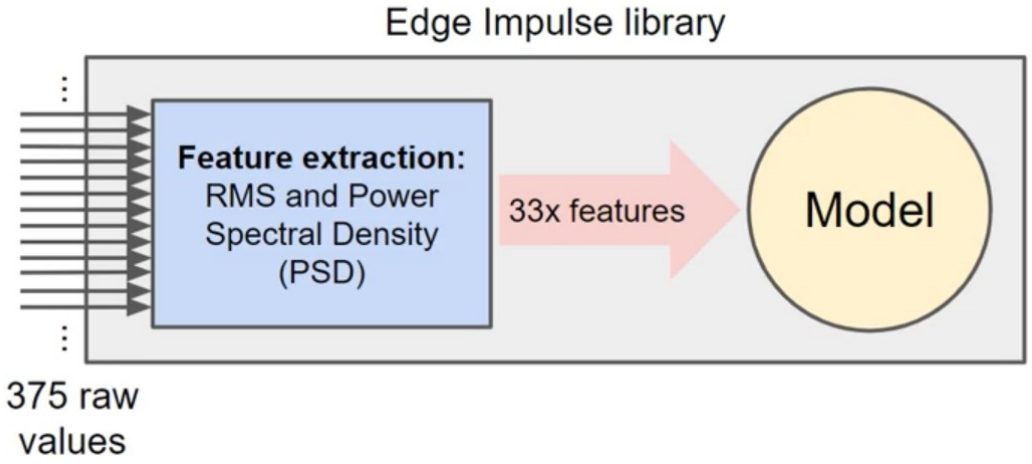
name: **serving_default_x:0**
type: **int8[1,33]**
quantization: **0.30515068769454956 * (q + 95)**
location: **0**

OUTPUTS

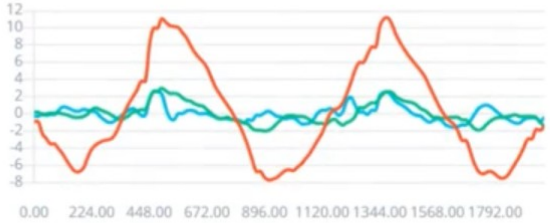
StatefulPartition...

name: **StatefulPartitionedCall:0**
type: **int8[1,4]**
quantization: **0.00390625 * (q + 128)**
location: **10**


netron.app
Bookmarks
Develop a con...
Latex
Python
IA
Livros
QUT Robot Ac...
Rsync
Linux
Markdown
Rocket Science
youtube-dl
Escrita Científ...
IA Tools
Legendas
Instituto Fede...
serving_default_x:0 0
1×33
FullyConnected
weights ⟨20×33⟩
bias ⟨20⟩
Relu
1×20
FullyConnected
weights ⟨10×20⟩
bias ⟨10⟩
Relu
1×10
FullyConnected
weights ⟨4×10⟩
bias ⟨4⟩
1×4
Softmax
1×4
StatefulPartitionedCall:0 10
NODE PROPERTIES
type
FullyConnected
location
0
ATTRIBUTES
asymmetric_qua...
false
fused_activation...
RELU
keep_num_dims
false
weights_format
DEFAULT
INPUTS
input
name: serving_default_x:0
weights
name: sequential/dense/MatMul
bias
name: dense/bias
OUTPUTS
output
name: sequential/dense/MatMul;sequential/dense/Relu;sequen

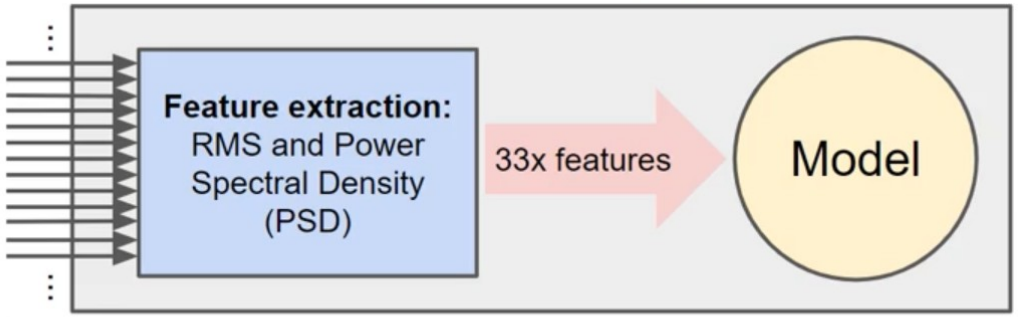

Edge Impulse library
Feature extraction:
RMS and Power
Spectral Density
(PSD)
33x features
Model
375 raw
values

Sample accelerometer for 2 seconds



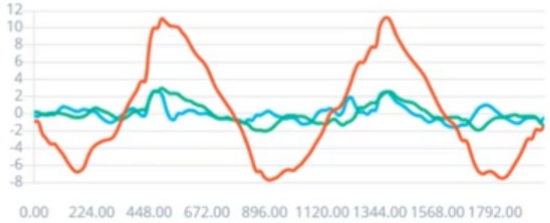
62.5 Hz sampling for 2 seconds with 3 axes = 375 values

Edge Impulse library



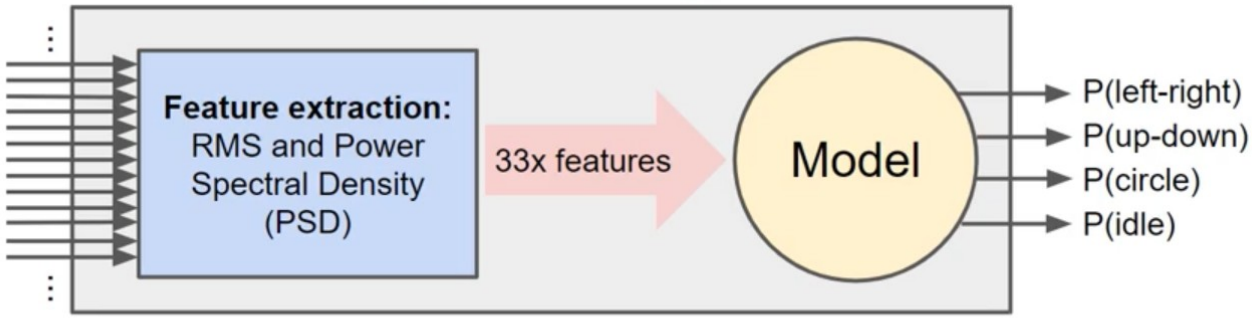
375 raw values

Sample accelerometer for 2 seconds



62.5 Hz sampling for 2 seconds with 3 axes = 375 values

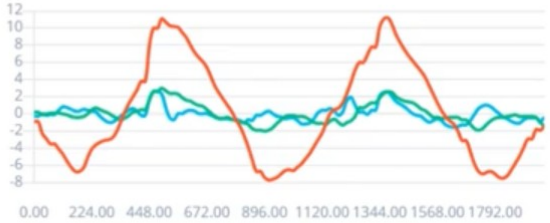
Edge Impulse library



375 raw values

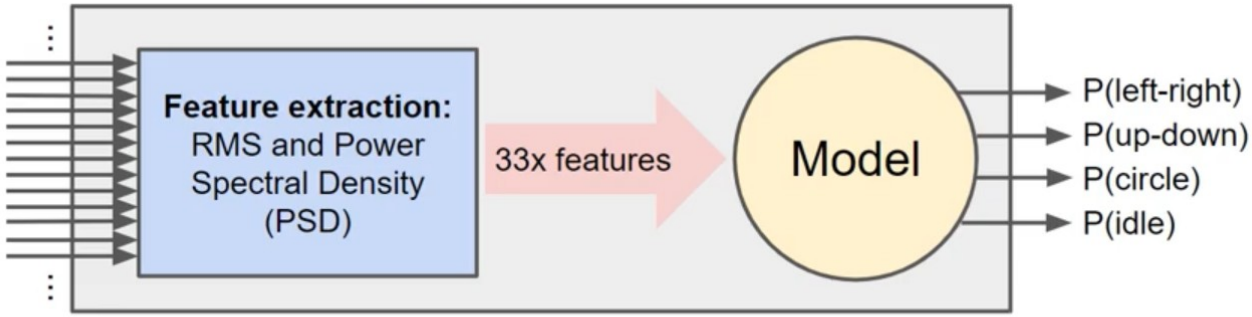
P(left-right) = 0.9143
P(up-down) = 0.0032
P(circle) = 0.0581
P(idle) = 0.0244

Sample accelerometer for 2 seconds



62.5 Hz sampling for 2 seconds with 3 axes = 375 values

Edge Impulse library



375 raw values

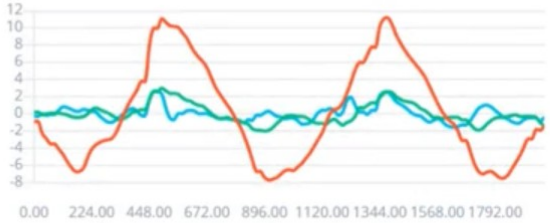
P(left-right)	= 0.9143
P(up-down)	= 0.0032
P(circle)	= 0.0581
P(idle)	= 0.0244

```
if (p_left_right > 0.5) {
    // Do stuff
}
```

```
if (p_left_right > 0.5) {
    // Do stuff
}

if (p_up_down > 0.5) {
    // Do some things
}
```

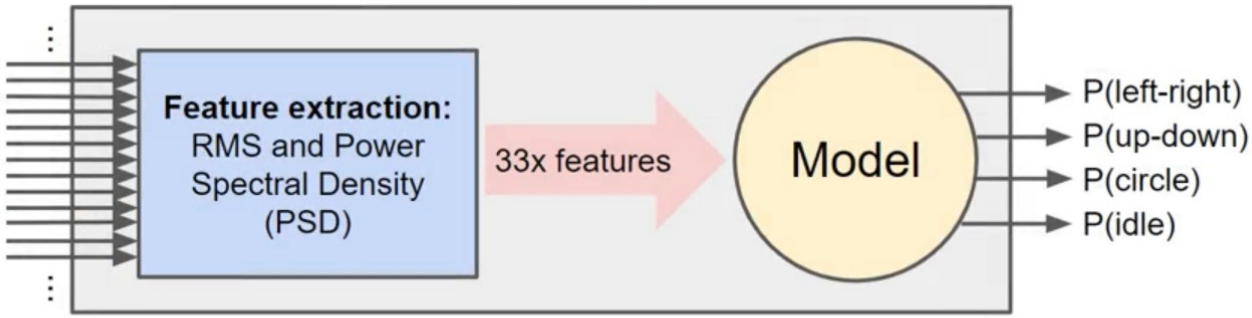
```
if (p_left_right > 0.5) {
    // Do stuff
}

if (p_up_down > 0.5) {
    // Do some things
}

if (p_circle > 0.8) {
    // And now for something completely different
}
```

Sample accelerometer for 2 seconds

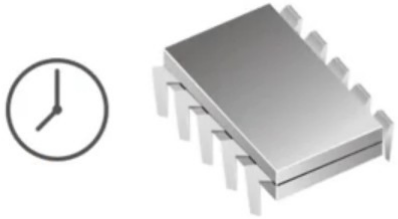


62.5 Hz sampling for 2 seconds with 3 axes = 375 values

Edge Impulse library

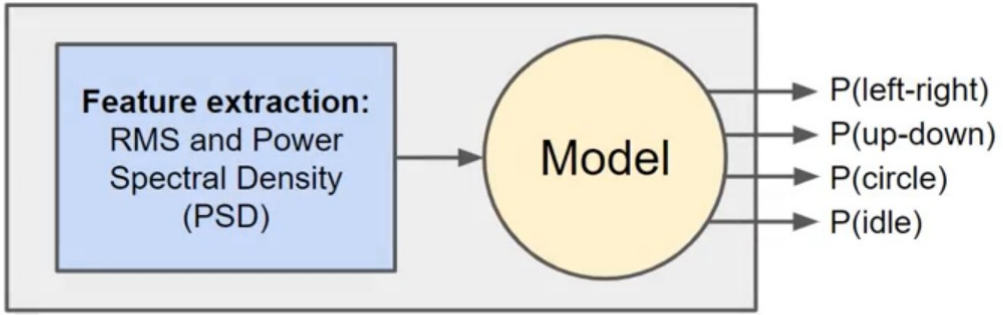


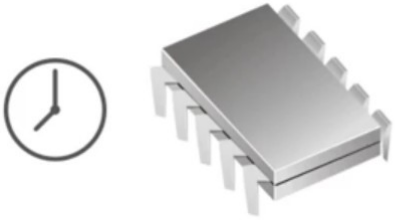
375 raw values



Sample accelerometer every 16 ms (62.5 Hz)

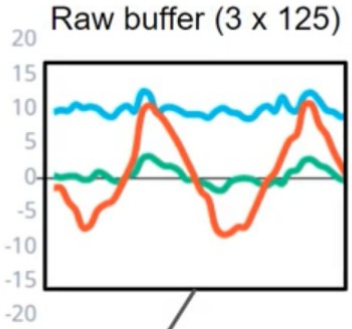
- Timer interrupt
- Real-time operating system (RTOS) task
- Direct memory access (DMA)

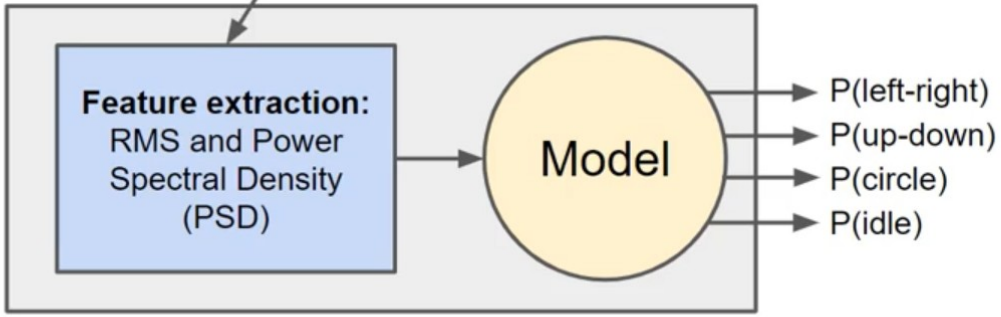


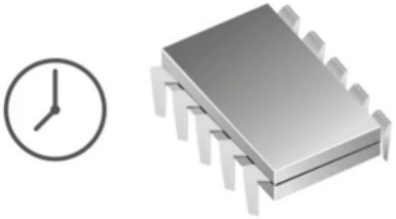


Sample accelerometer every 16 ms (62.5 Hz)

- Timer interrupt
- Real-time operating system (RTOS) task
- Direct memory access (DMA)

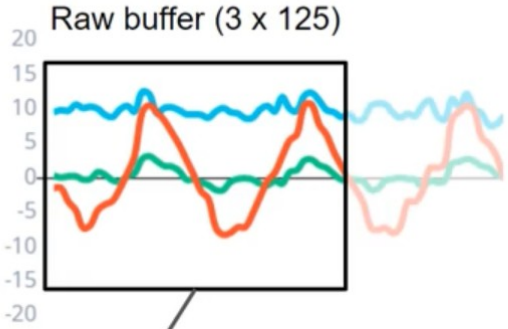


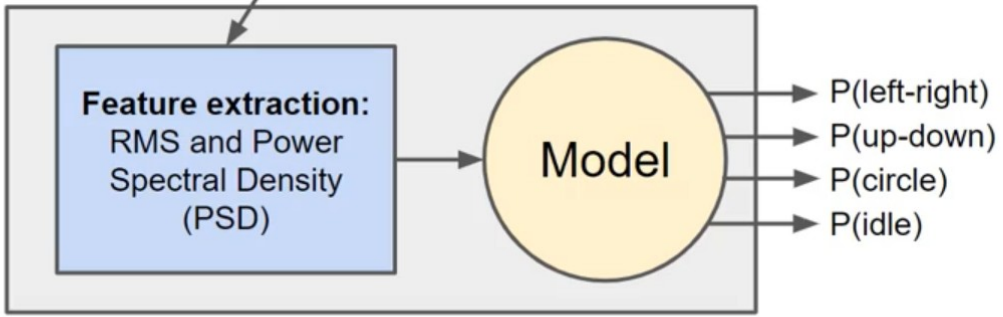


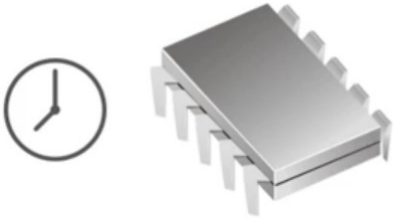


Sample accelerometer every 16 ms (62.5 Hz)

- Timer interrupt
- Real-time operating system (RTOS) task
- Direct memory access (DMA)







Sample accelerometer every 16 ms (62.5 Hz)

- Timer interrupt
- Real-time operating system (RTOS) task
- Direct memory access (DMA)