WALC 2023
Applied AI

Sound Classification Intro & Hands-On

Prof. Marcelo J. Rovai
rovai@unifei.edu.br

UNIFEI - Federal University of Itajuba, Brazil
TinyML4D Academic Network Co-Chair

Vision

Sound

Vibration

KWS (KeyWord Spotting) Introduction

Personal Assistant

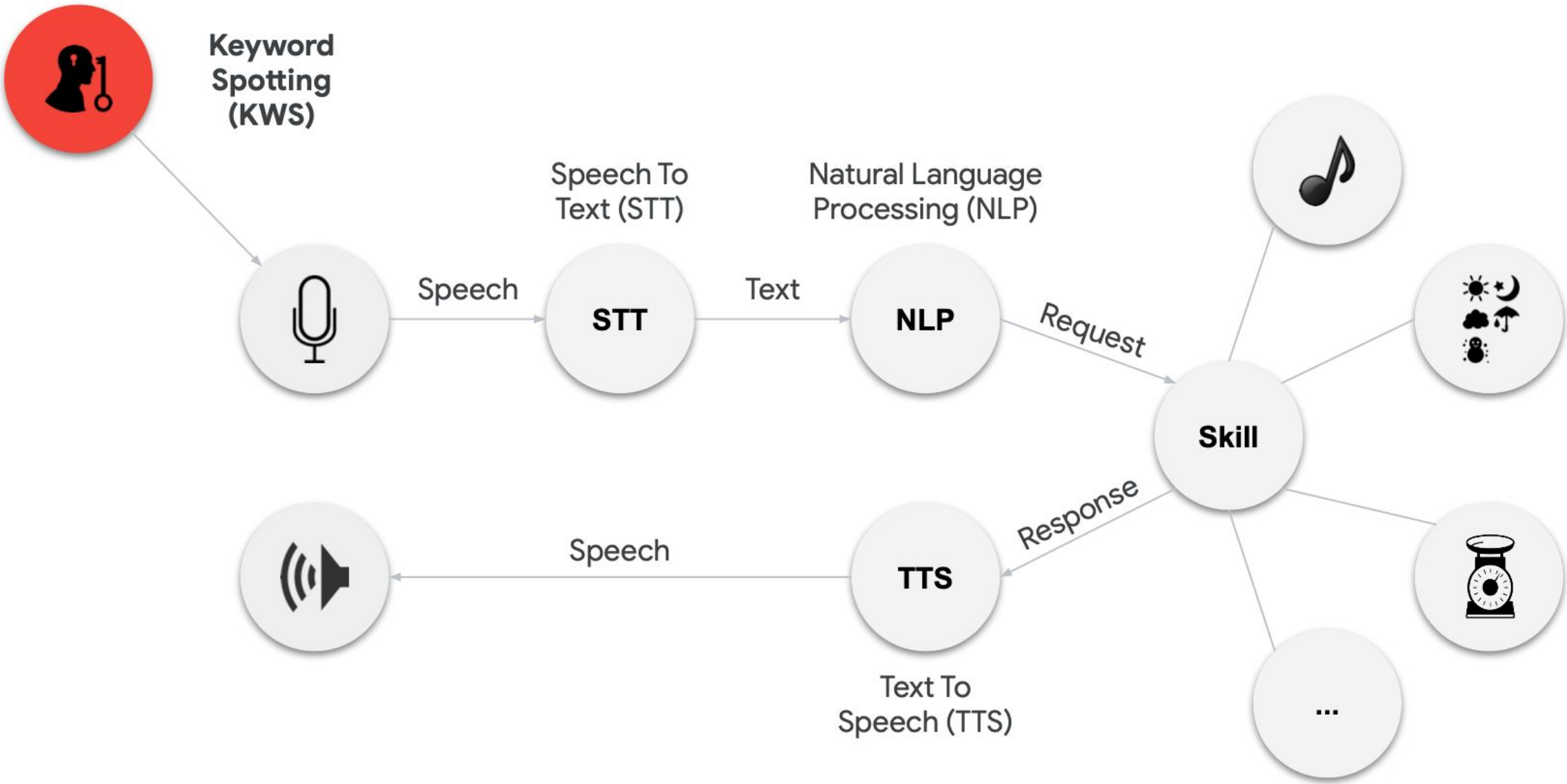
Keyword Spotting v. General Speech Recognition

- **Keyword spotting** is one of the most successful examples of **TinyML**
  - Low-power, continuous, on-device
  - Common Voice SWTS* expands keyword spotting to more languages

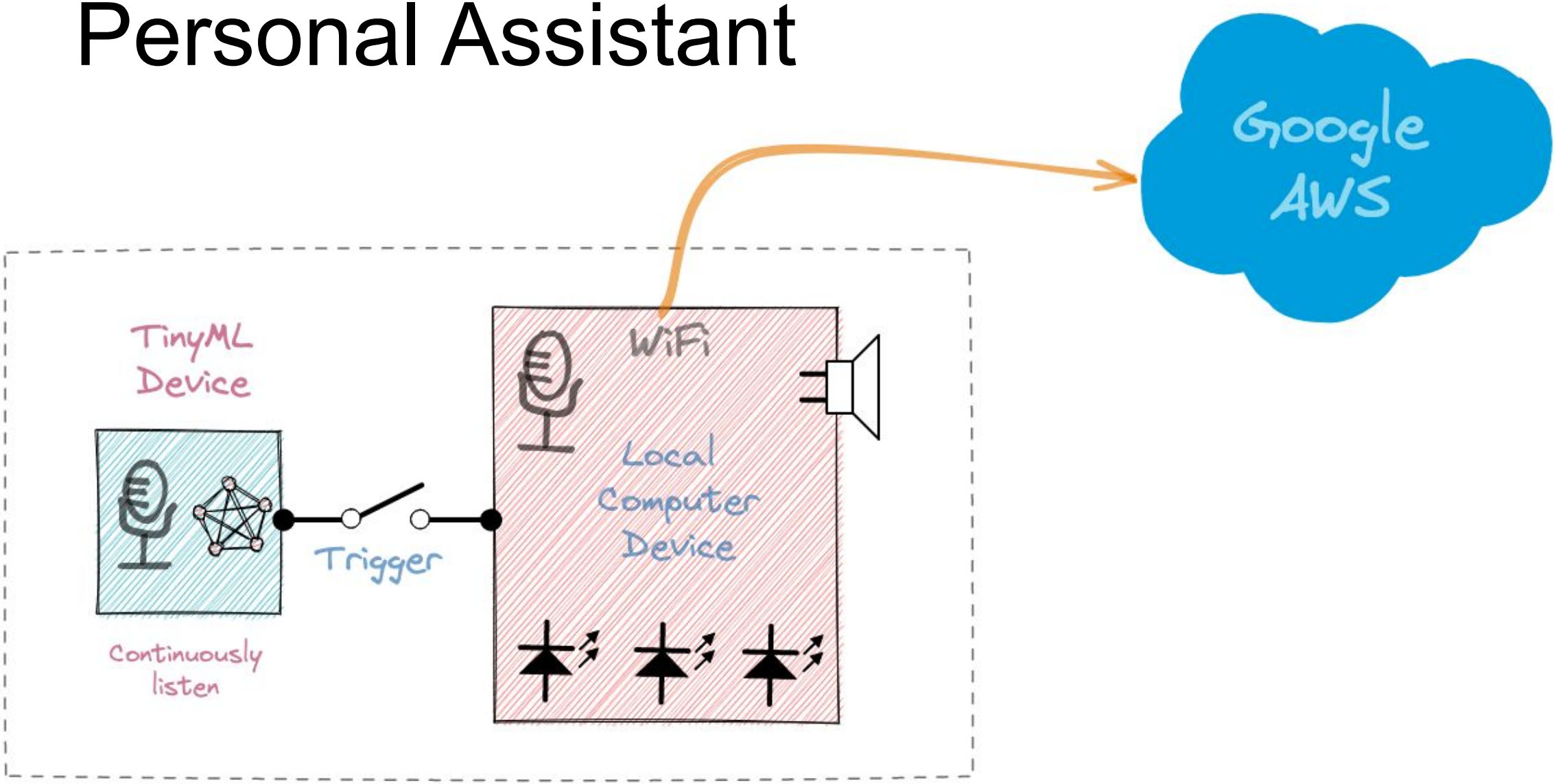
    * **S**ingle **W**ord **T**arget **S**egment

- **General ASR*** still requires **larger, power-hungry models**
  - But it can run on mobile devices (offline dictation on smartphones)

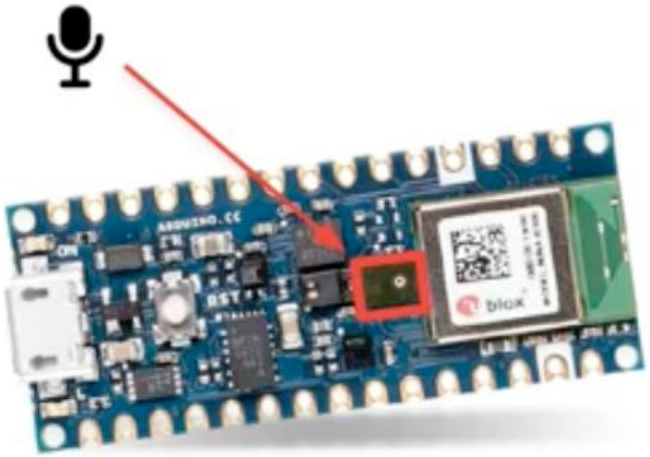
    * Automatic Speech Recognition


Keyword Spotting (KWS)
Speech To Text (STT)
Natural Language Processing (NLP)
Speech
STT
Text
NLP
Request
Skill
Response
Speech
TTS
Text To Speech (TTS)
...

Personal Assistant



“Cascade” Detection: multi-stage model



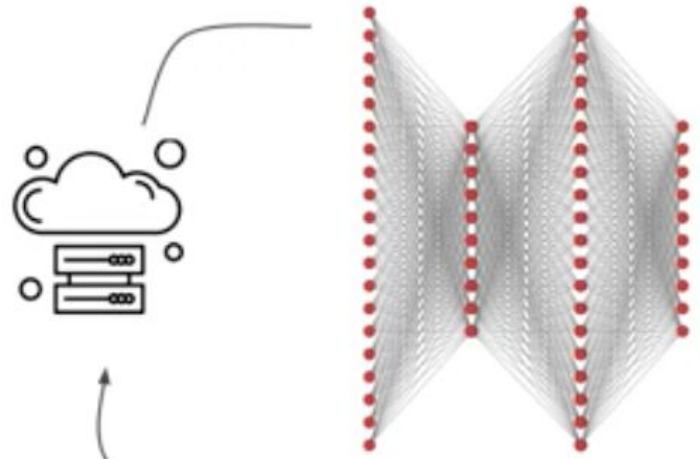
1 Continuously listen on the microcontroller

2 Process the data with **TinyML** at the edge

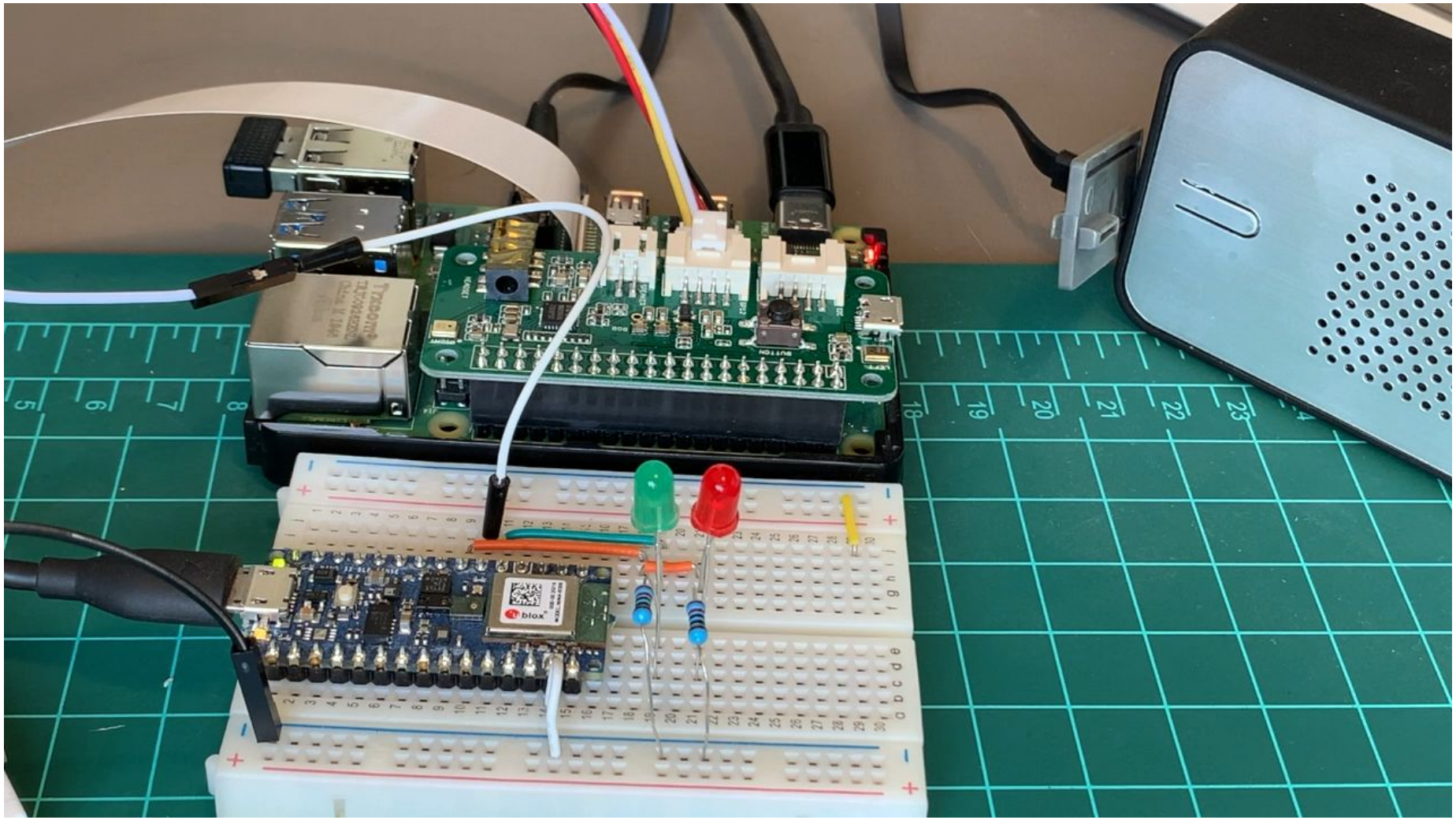
3 Process on a secondary larger model on a larger local device

4 Send the data to the cloud when triggered

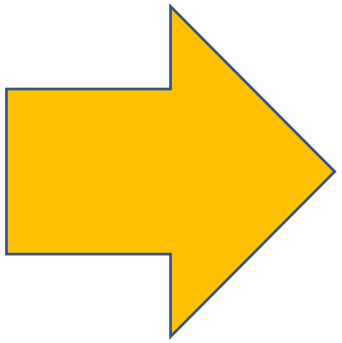
5 Process the full speech data with a large model in the cloud

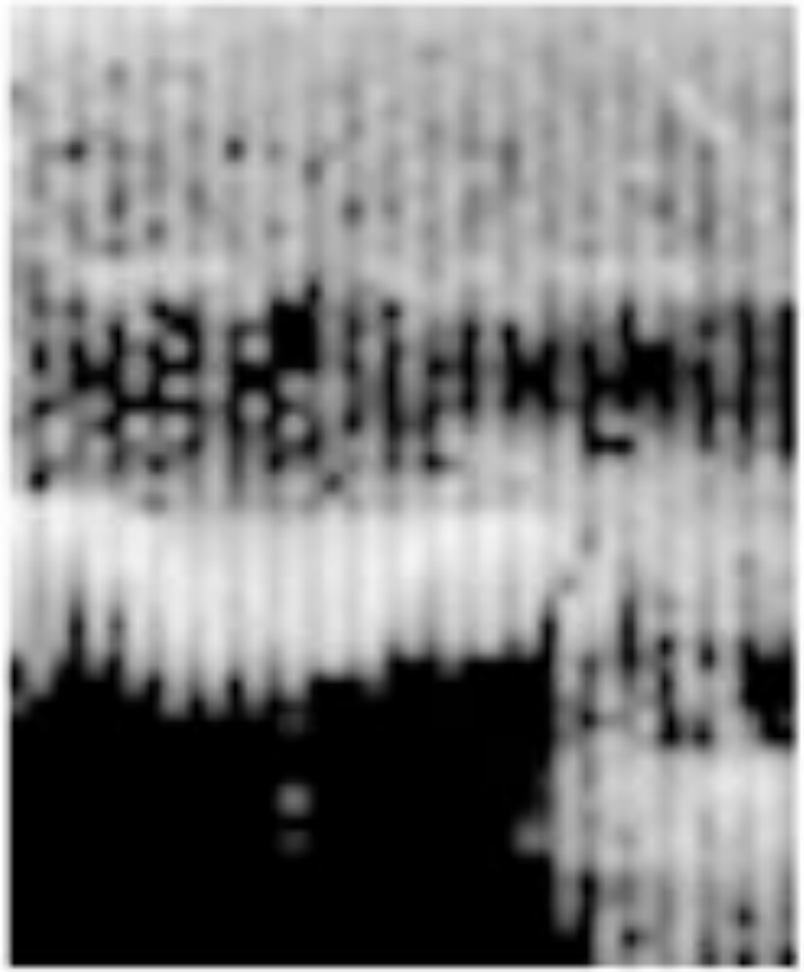


KeyWord Spotting (KWS)



https://mjrobot.org/2021/01/27/building-an-intelligent-voice-assistant-from-scratch/



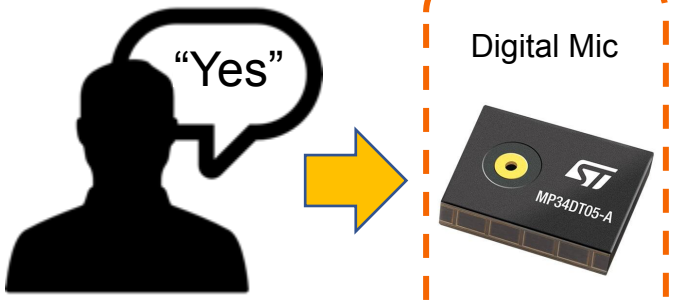


Sound

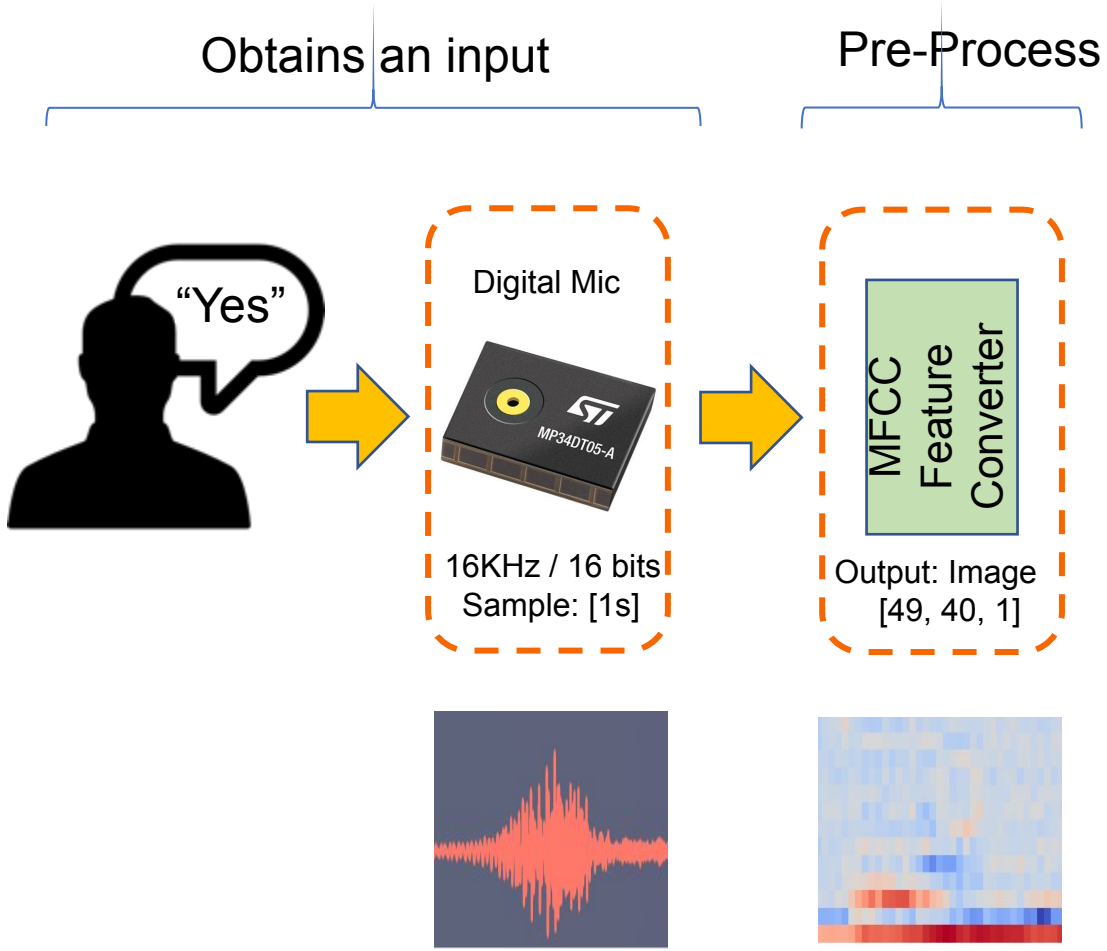
Image

KeyWord Spotting (KWS) - Inference

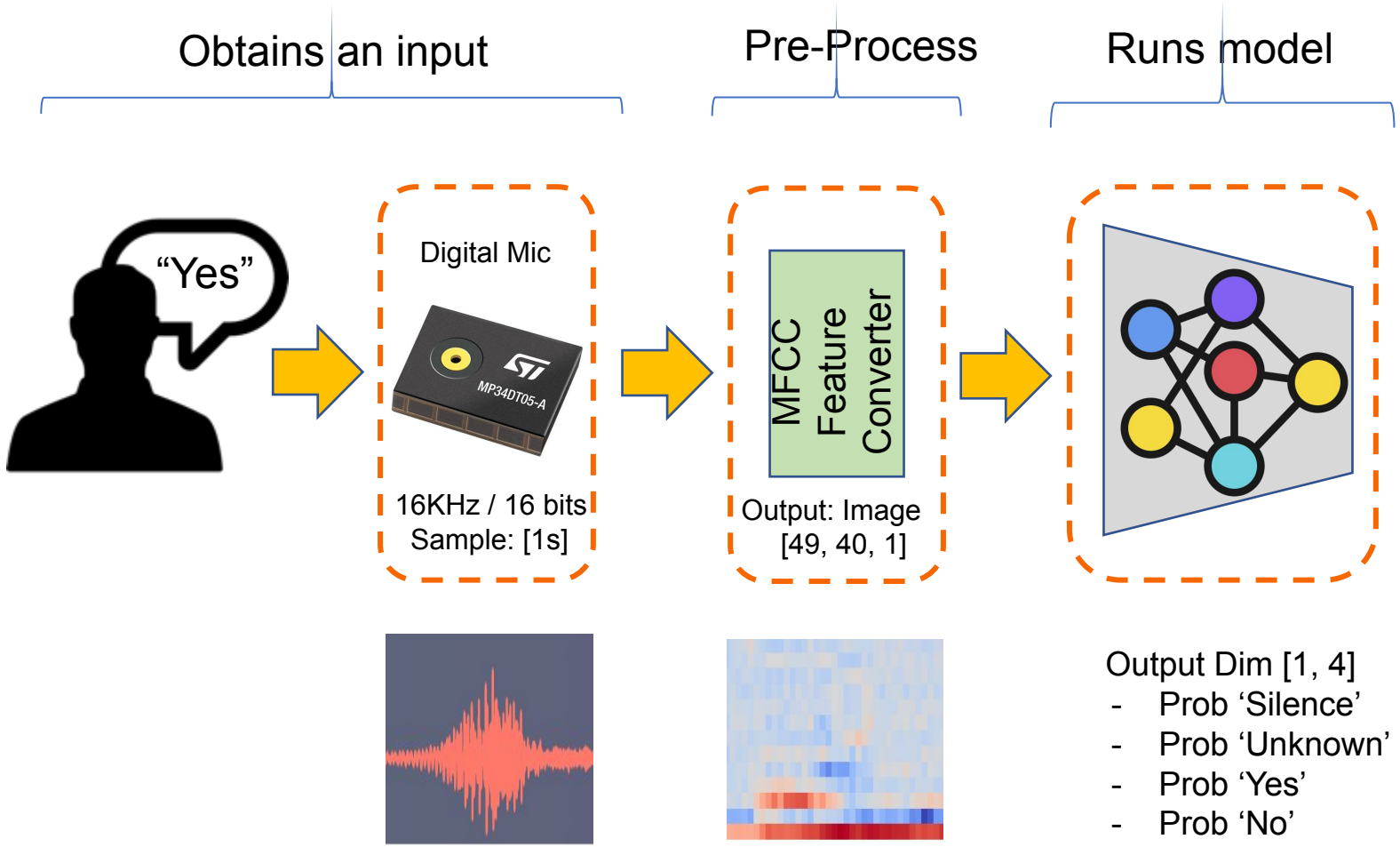
Obtains an input



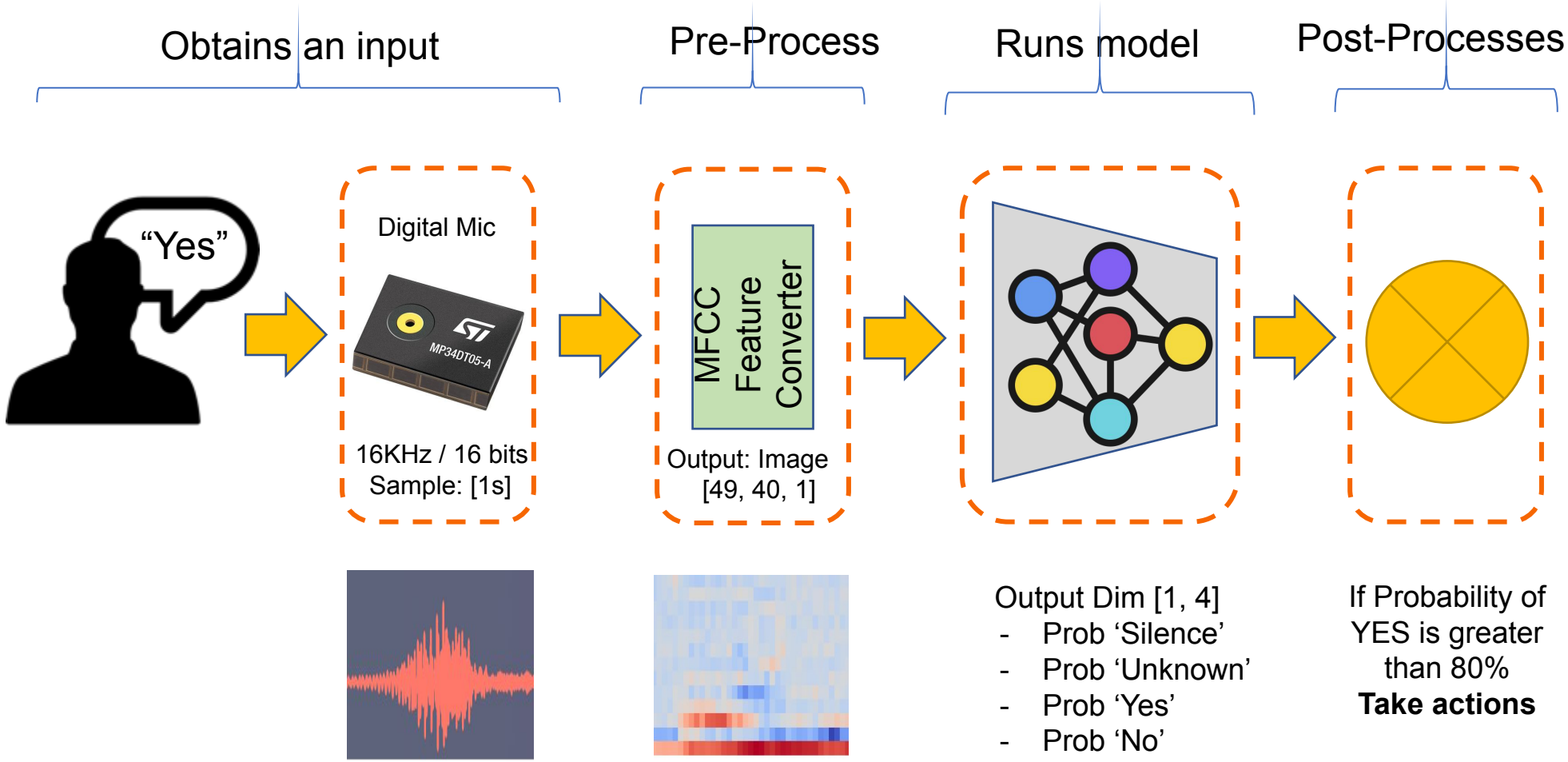
KeyWord Spotting (KWS) - Inference



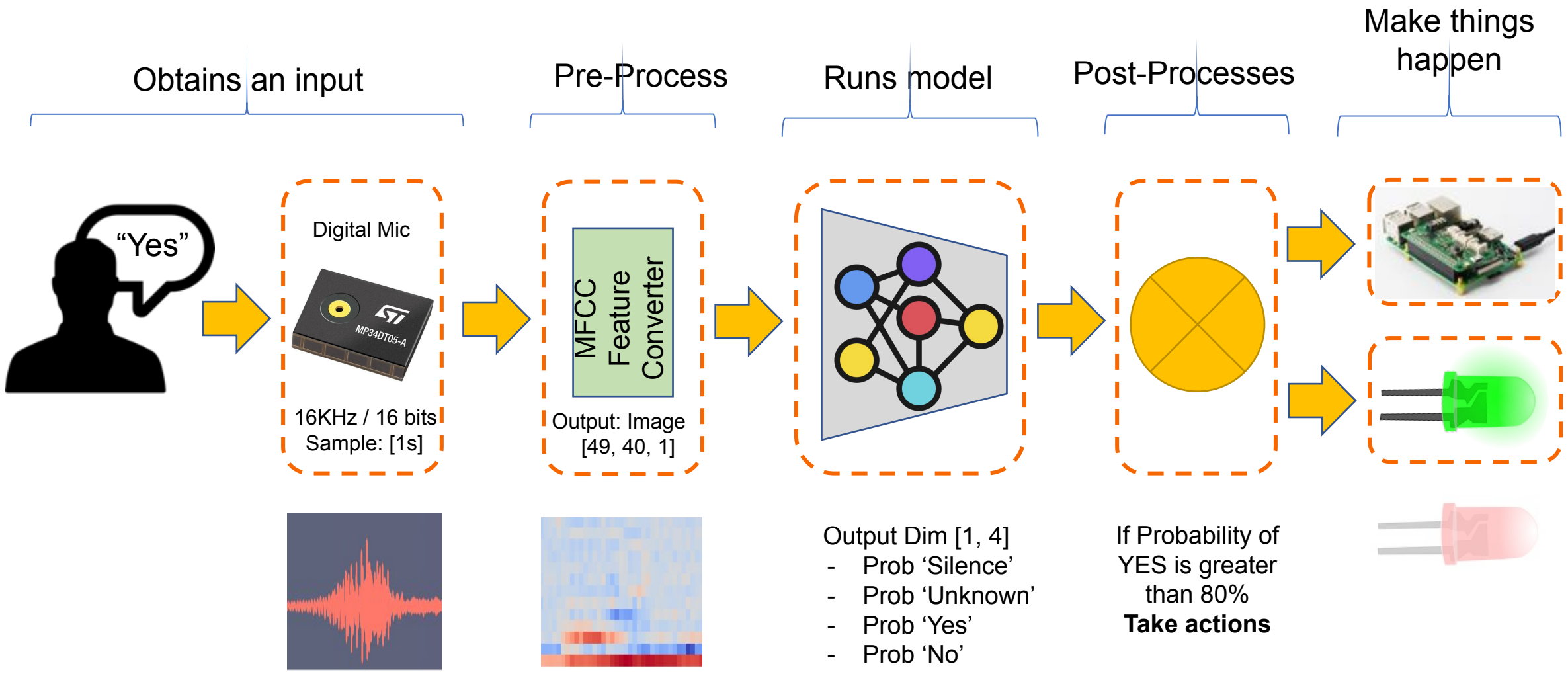
KeyWord Spotting (KWS) - Inference



KeyWord Spotting (KWS) - Inference



KeyWord Spotting (KWS) - Inference



KeyWord Spotting (KWS) - **Model**

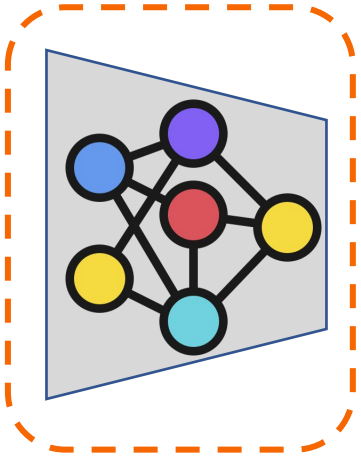
Obtains an input

Pre-Process

Runs model

postprocesses

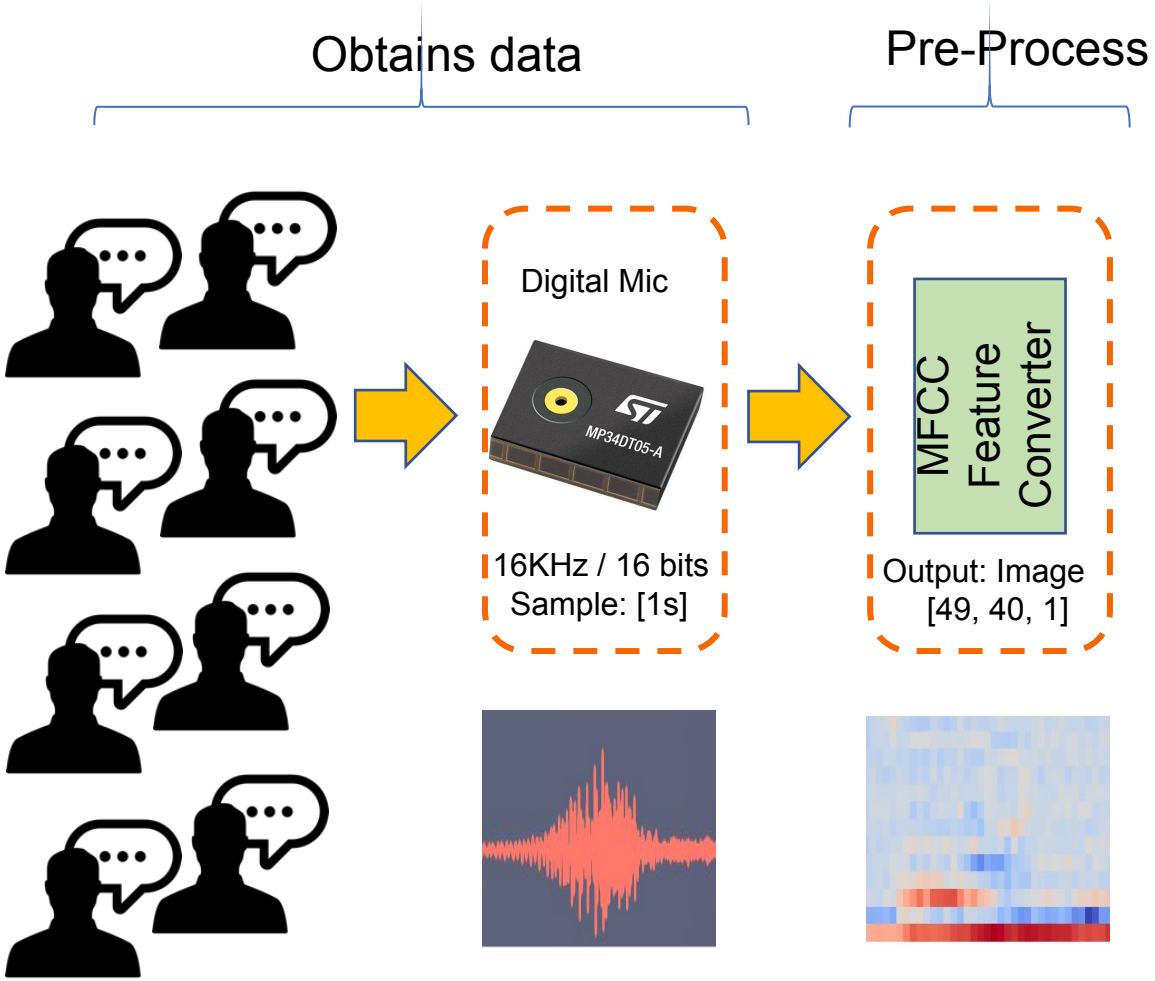
Make things happen



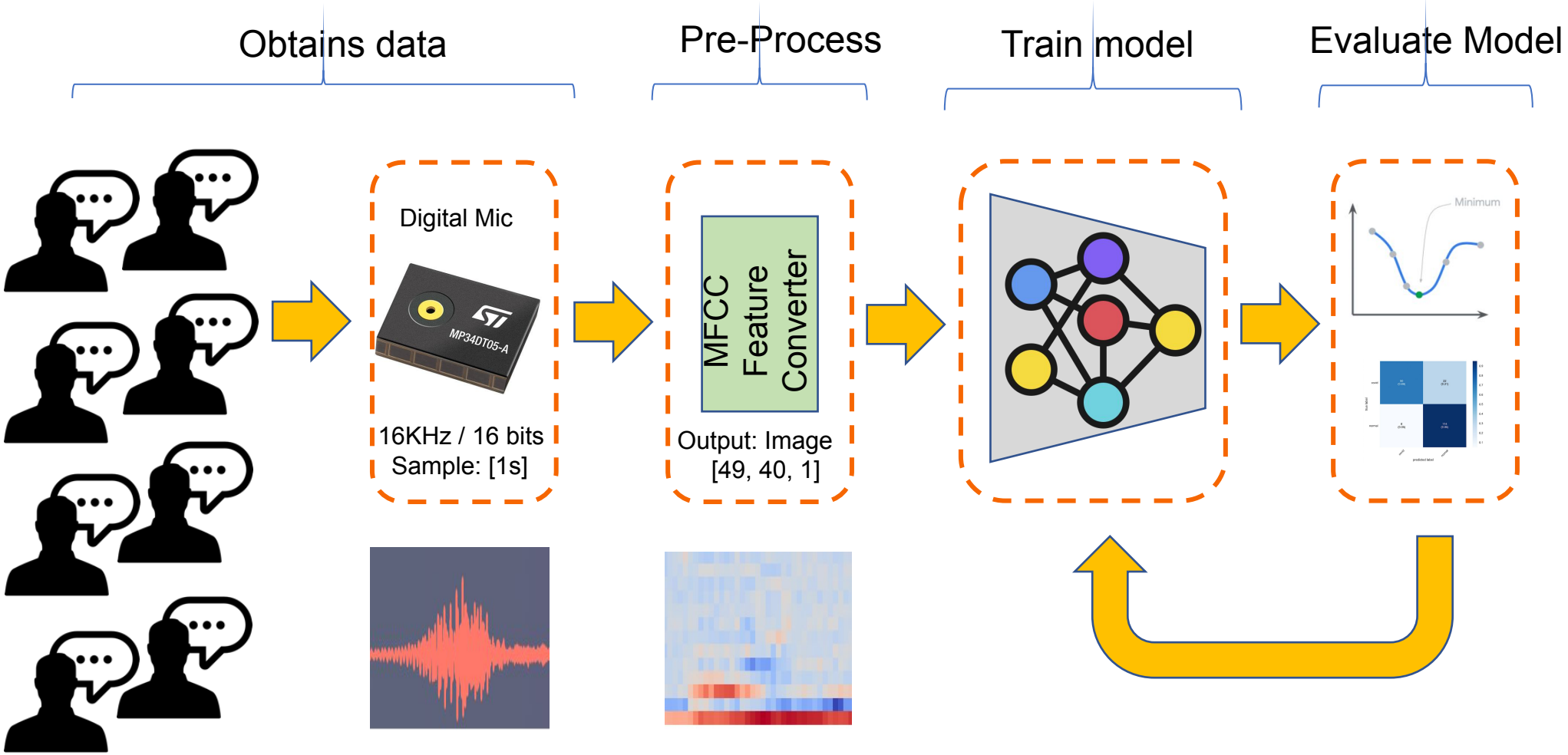
Output Dim [1, 4]
- Prob ‘Silence’
- Prob ‘Unknown’
- Prob ‘Yes’
- Prob ‘No’

If Probability of YES is greater than 80%
Take actions

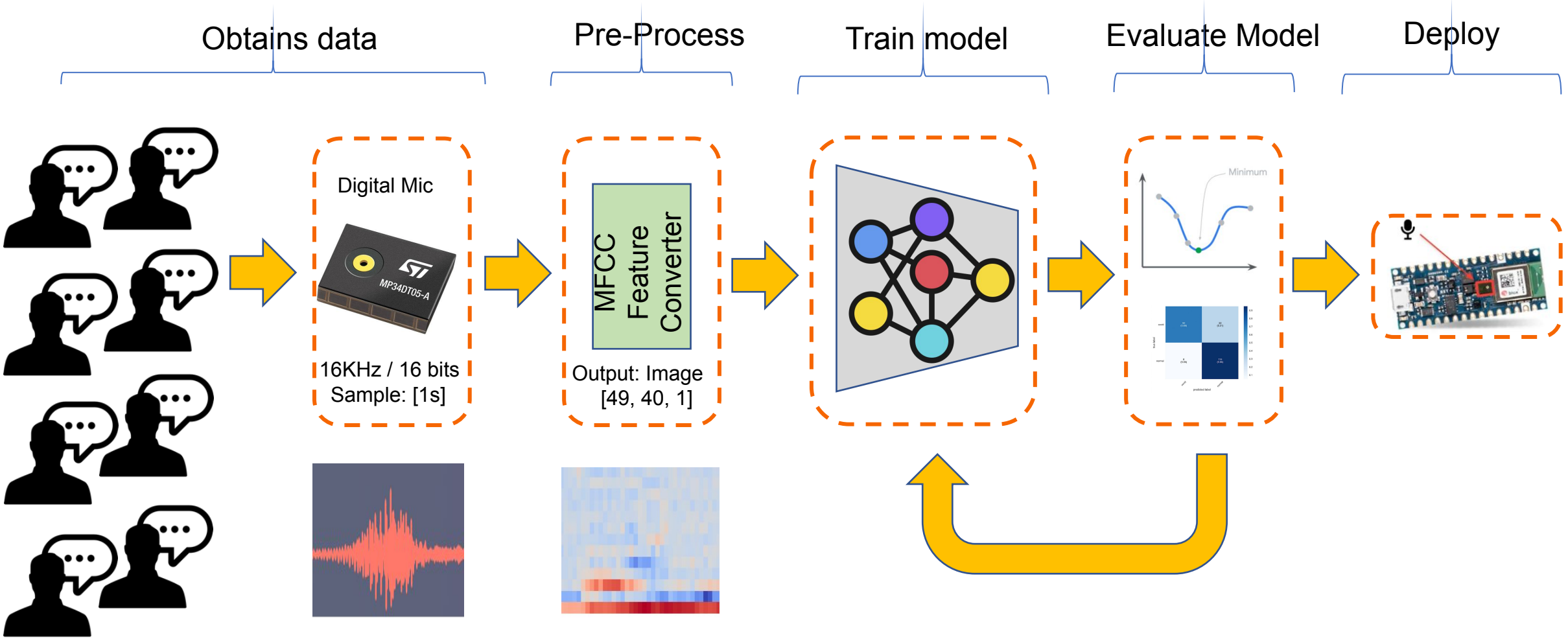
KeyWord Spotting (KWS) – Create Model (Training)



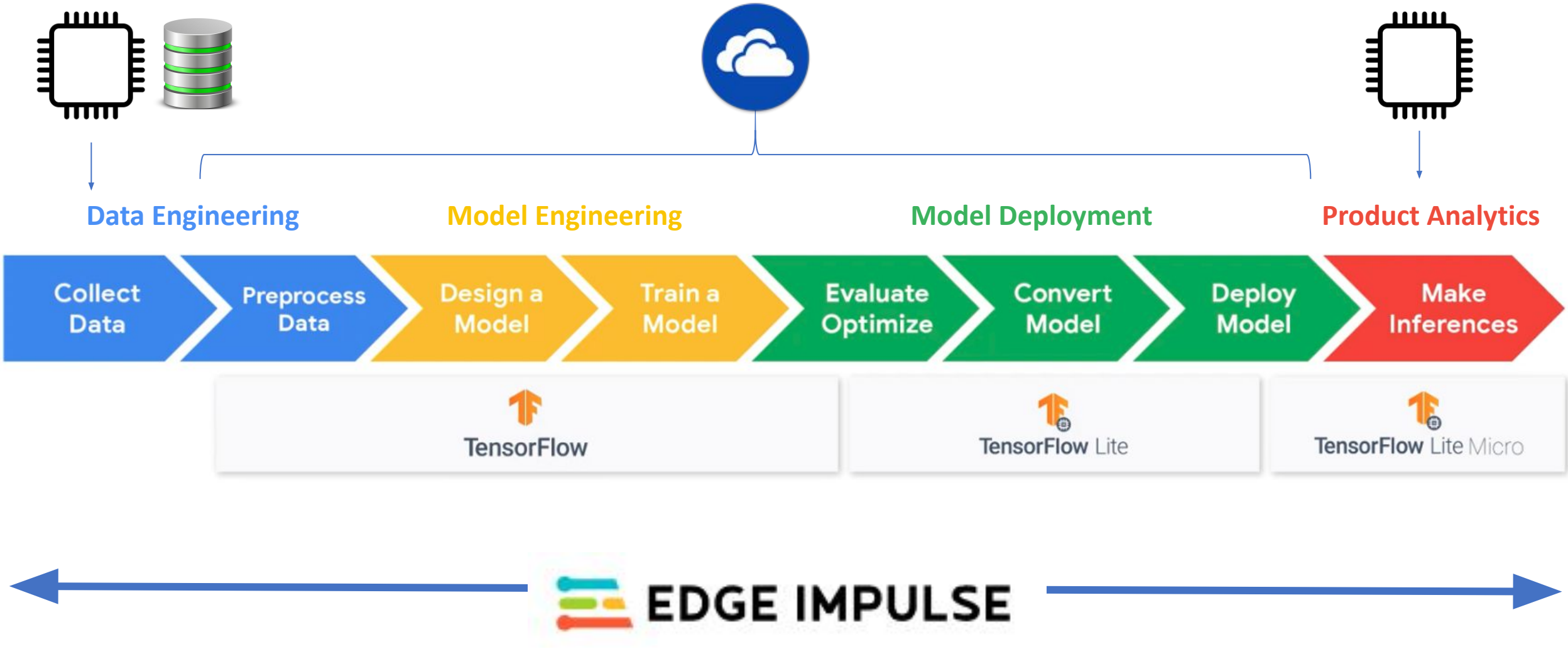
KeyWord Spotting (KWS) – Create Model (Training)

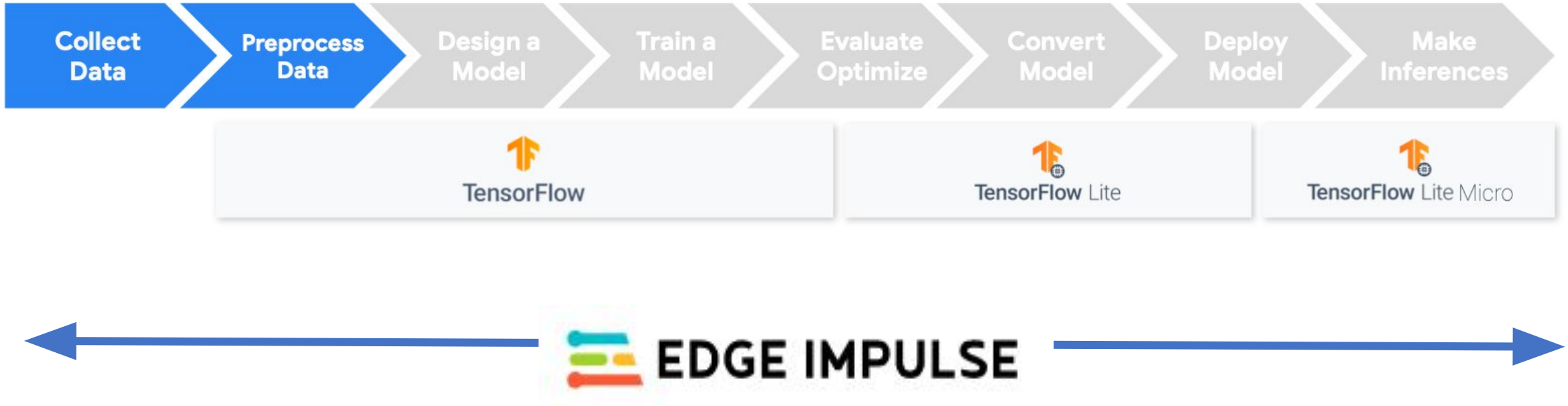


KeyWord Spotting (KWS) – Create Model (Training)

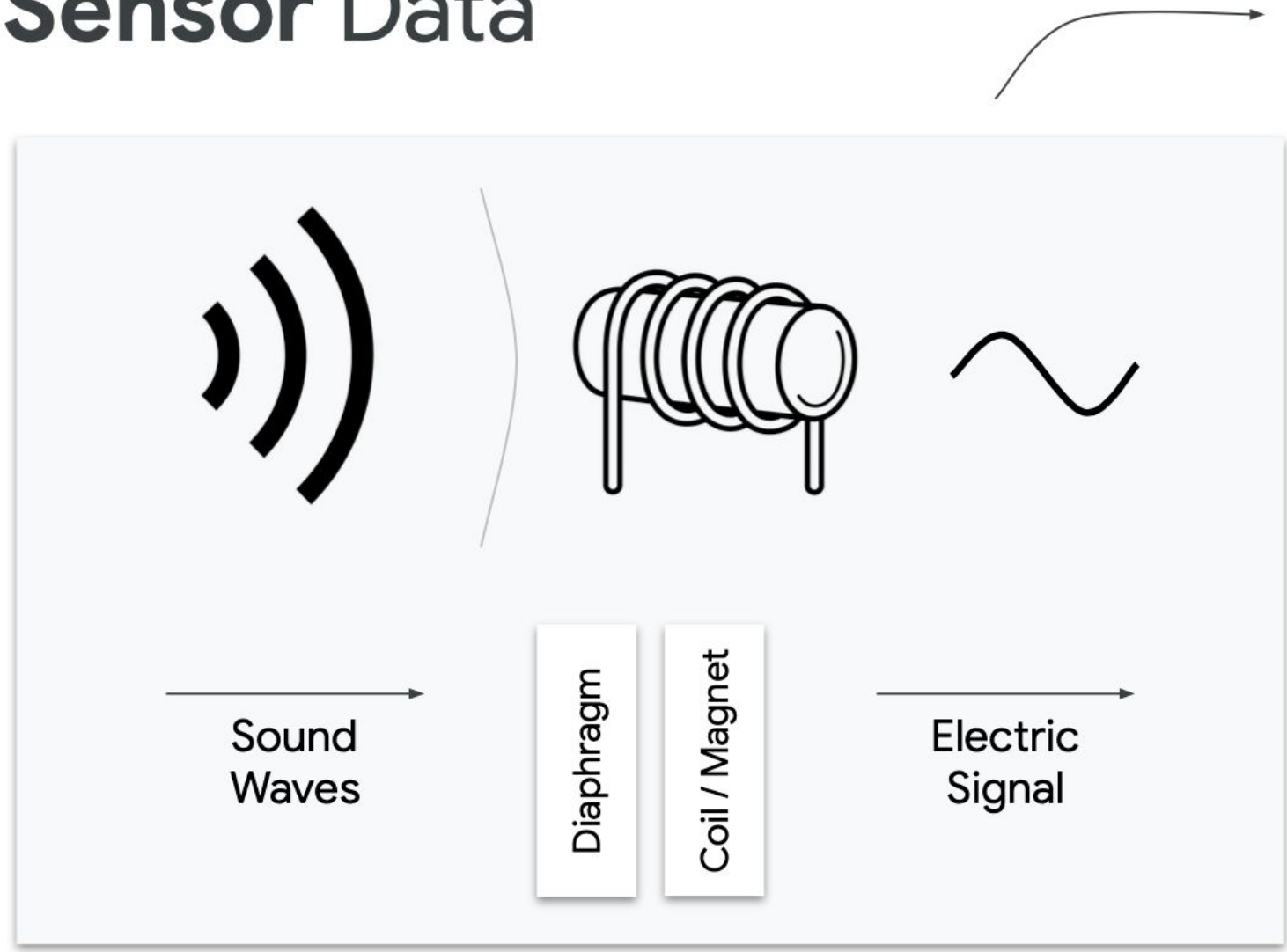


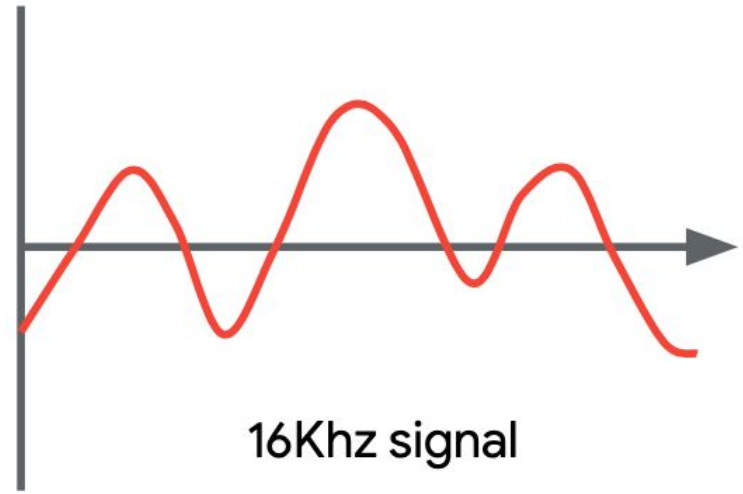
KWS Data Collection & Pre-Processing

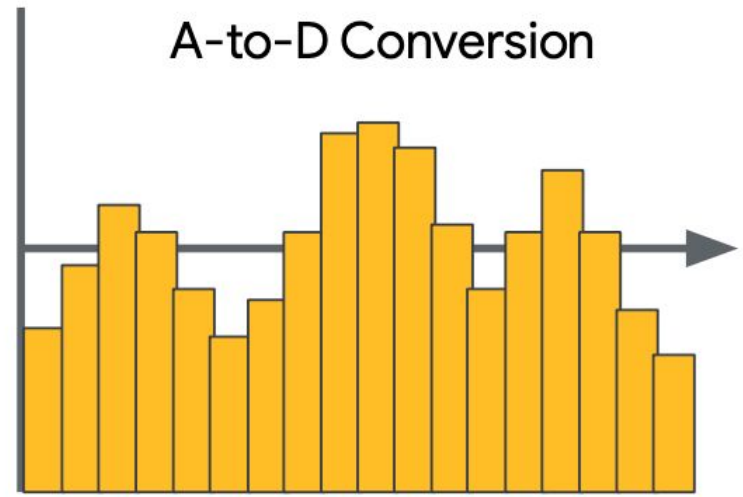

Data Engineering
Model Engineering
Model Deployment
Product Analytics
Collect Data
Preprocess Data
Design a Model
Train a Model
Evaluate Optimize
Convert Model
Deploy Model
Make Inferences
TensorFlow
TensorFlow Lite
TensorFlow Lite Micro
EDGE IMPULSE


Collect Data
Preprocess Data
Design a Model
Train a Model
Evaluate Optimize
Convert Model
Deploy Model
Make Inferences
TensorFlow
TensorFlow Lite
TensorFlow Lite Micro
EDGE IMPULSE

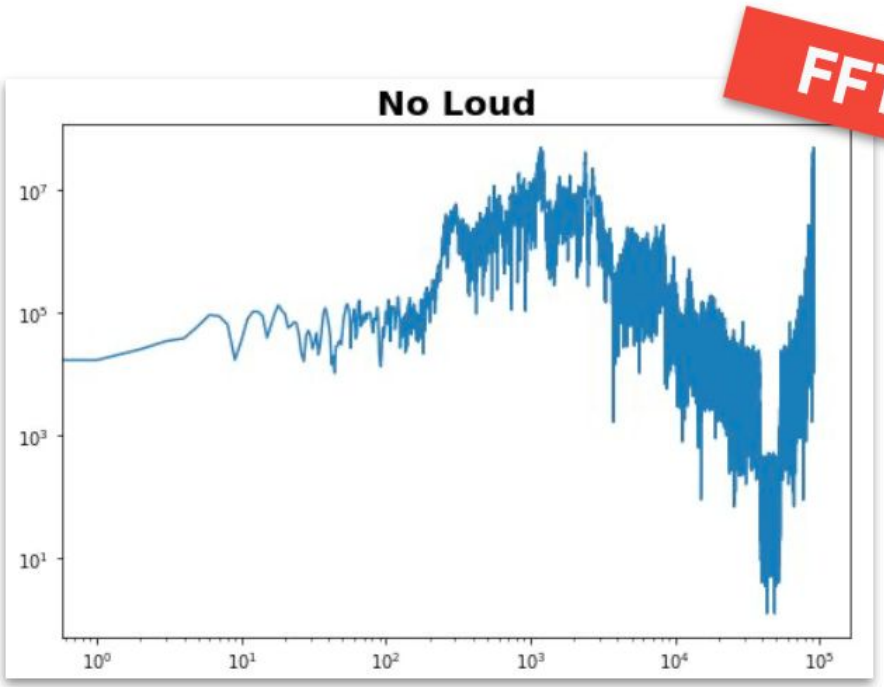
Sensor Data

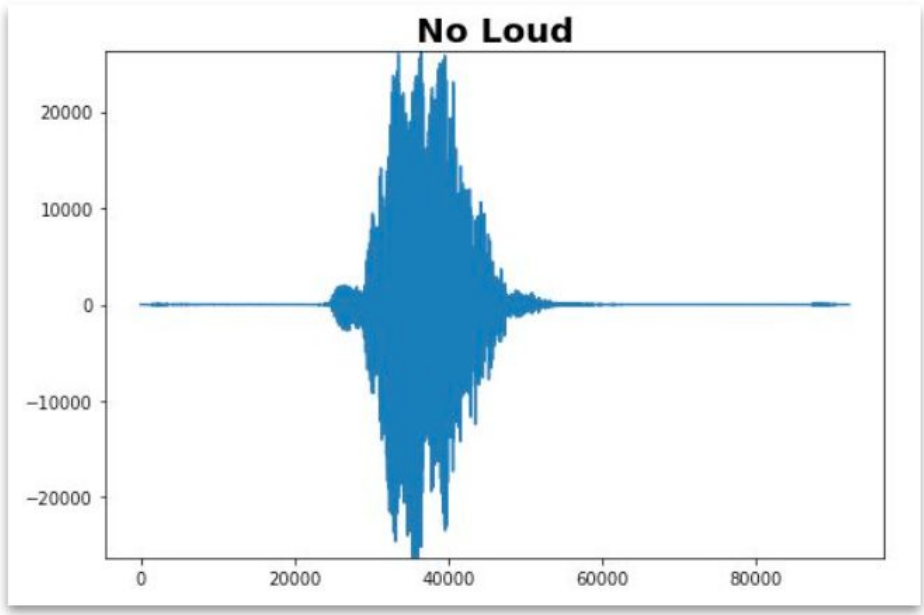

Sound
Waves
Diaphragm
Coil / Magnet
Electric
Signal


16Khz signal


A-to-D Conversion

Signal **Components**?

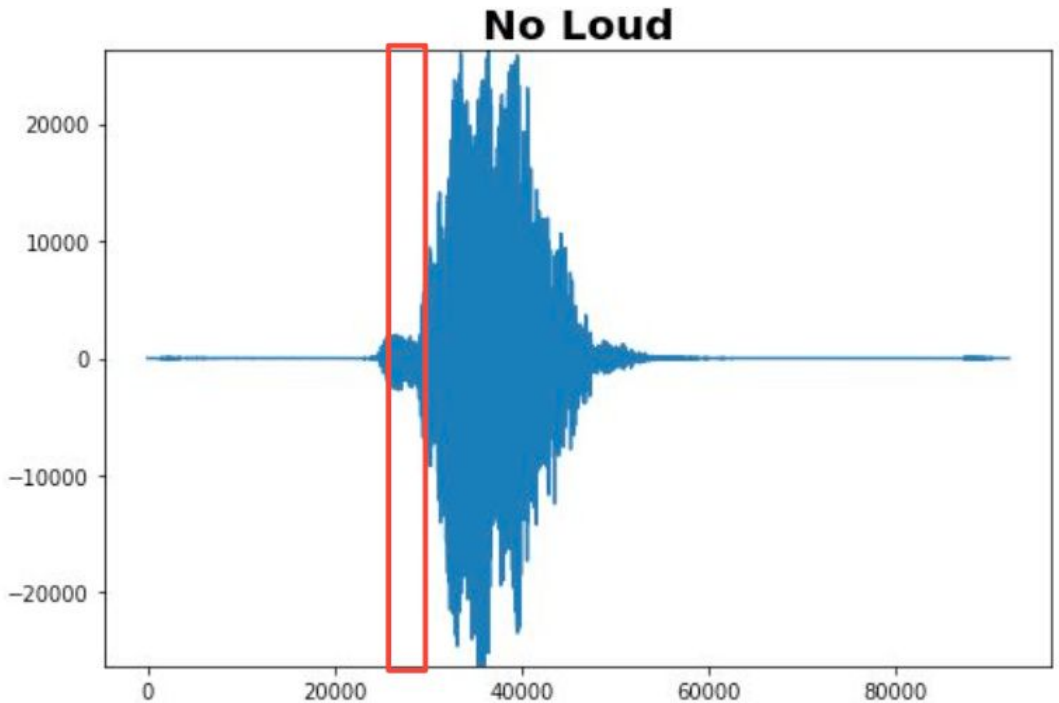




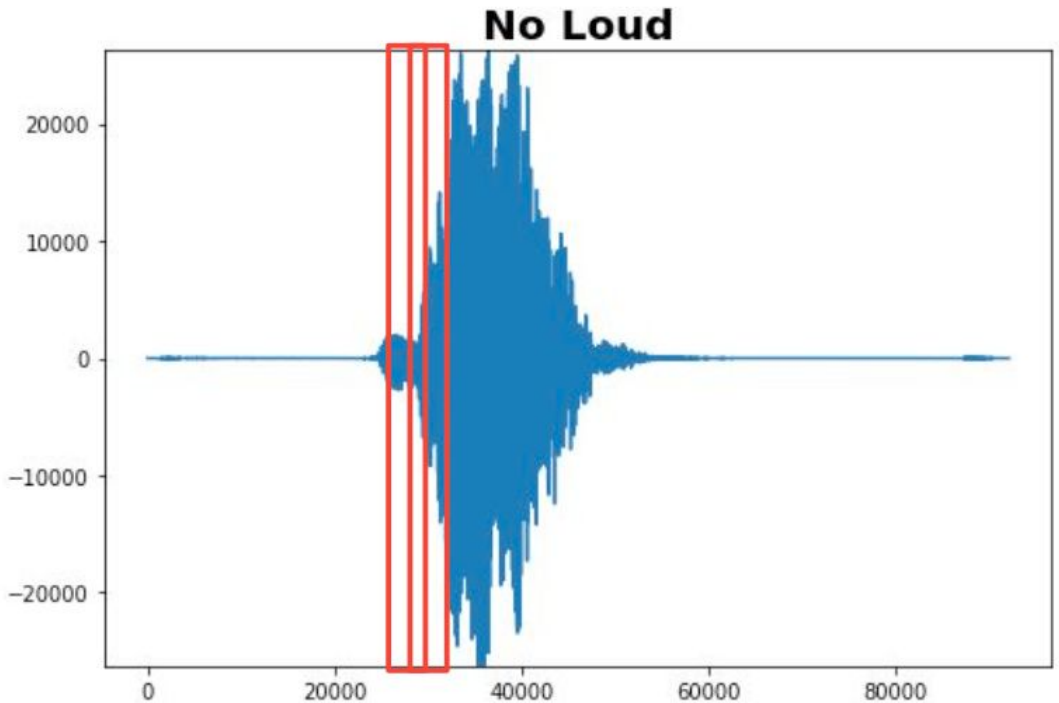
Frequency

Time

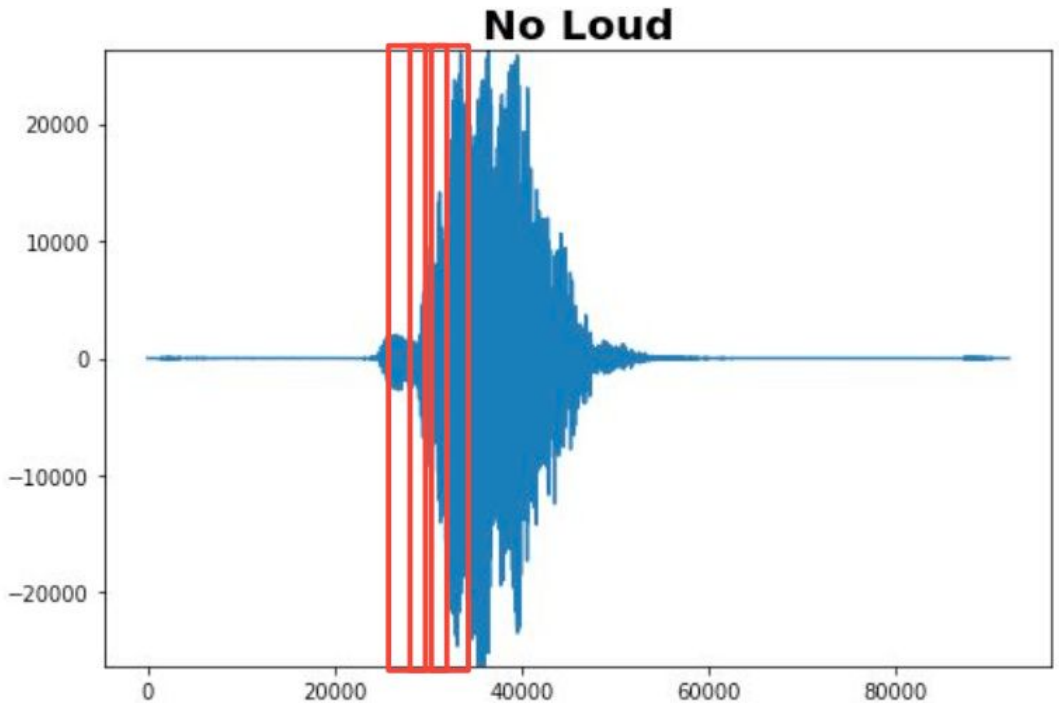
Data Preprocessing



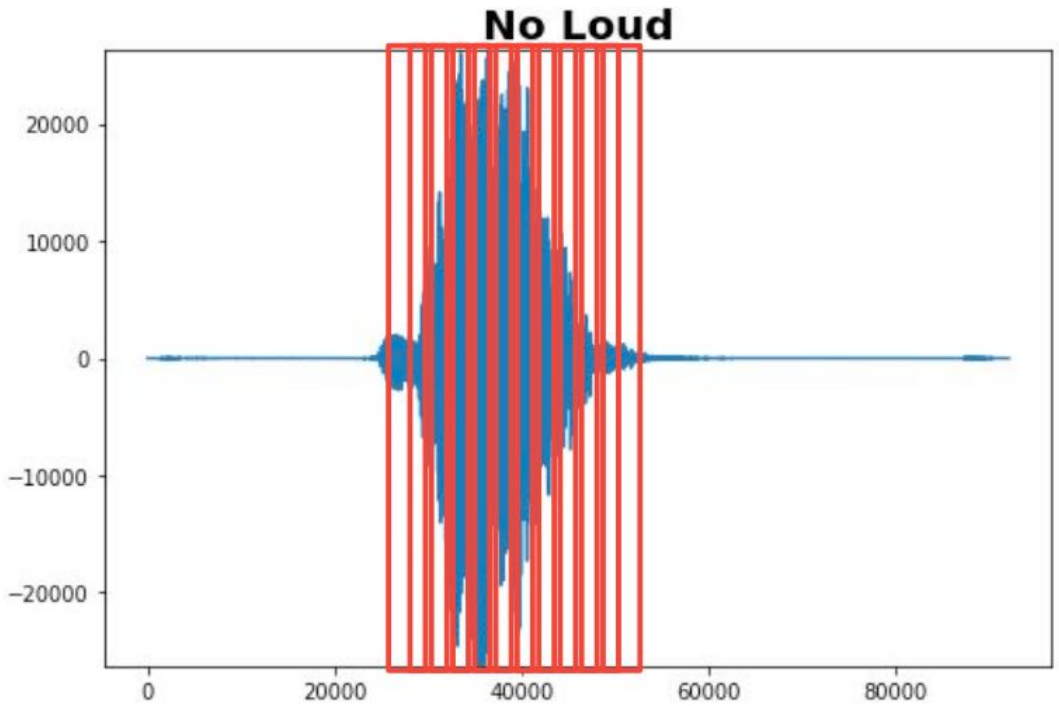
Data Preprocessing

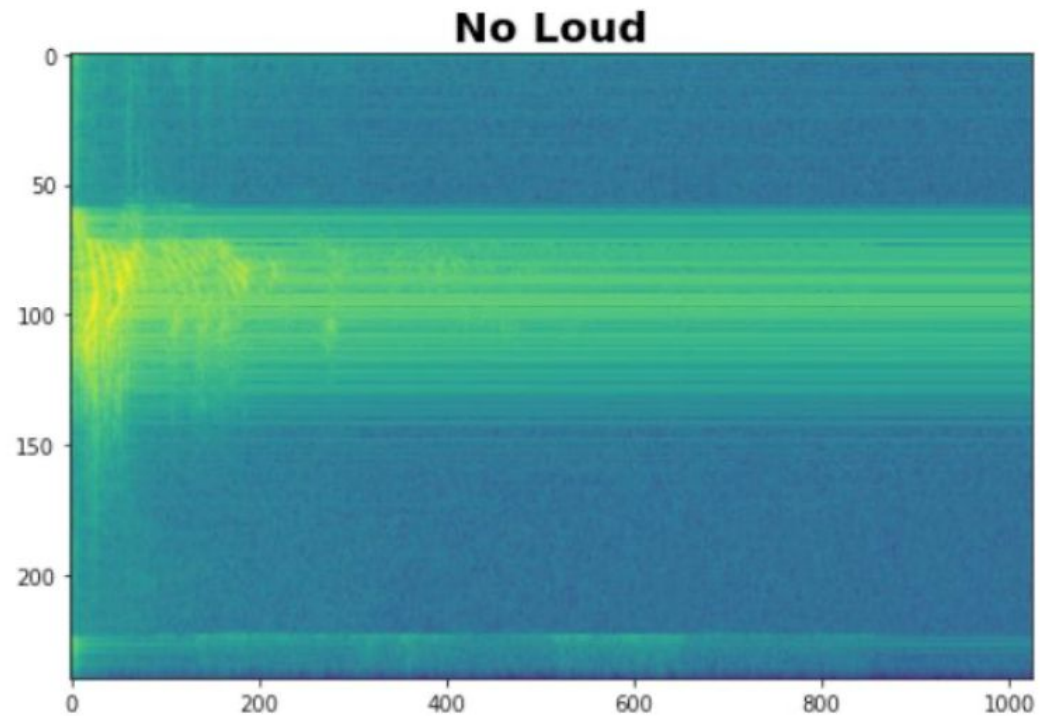


Data Preprocessing

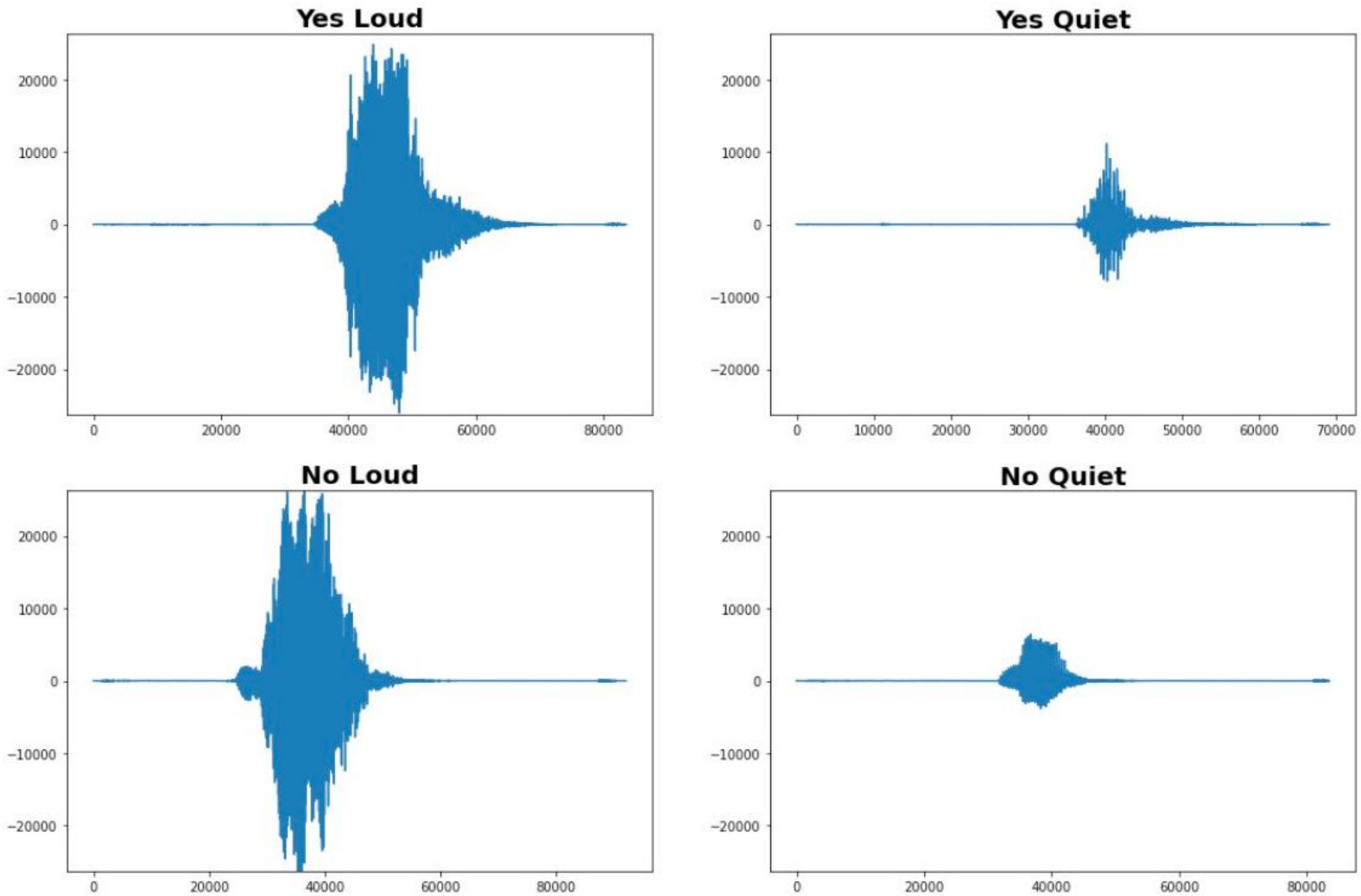


Data Preprocessing: **Spectrograms**

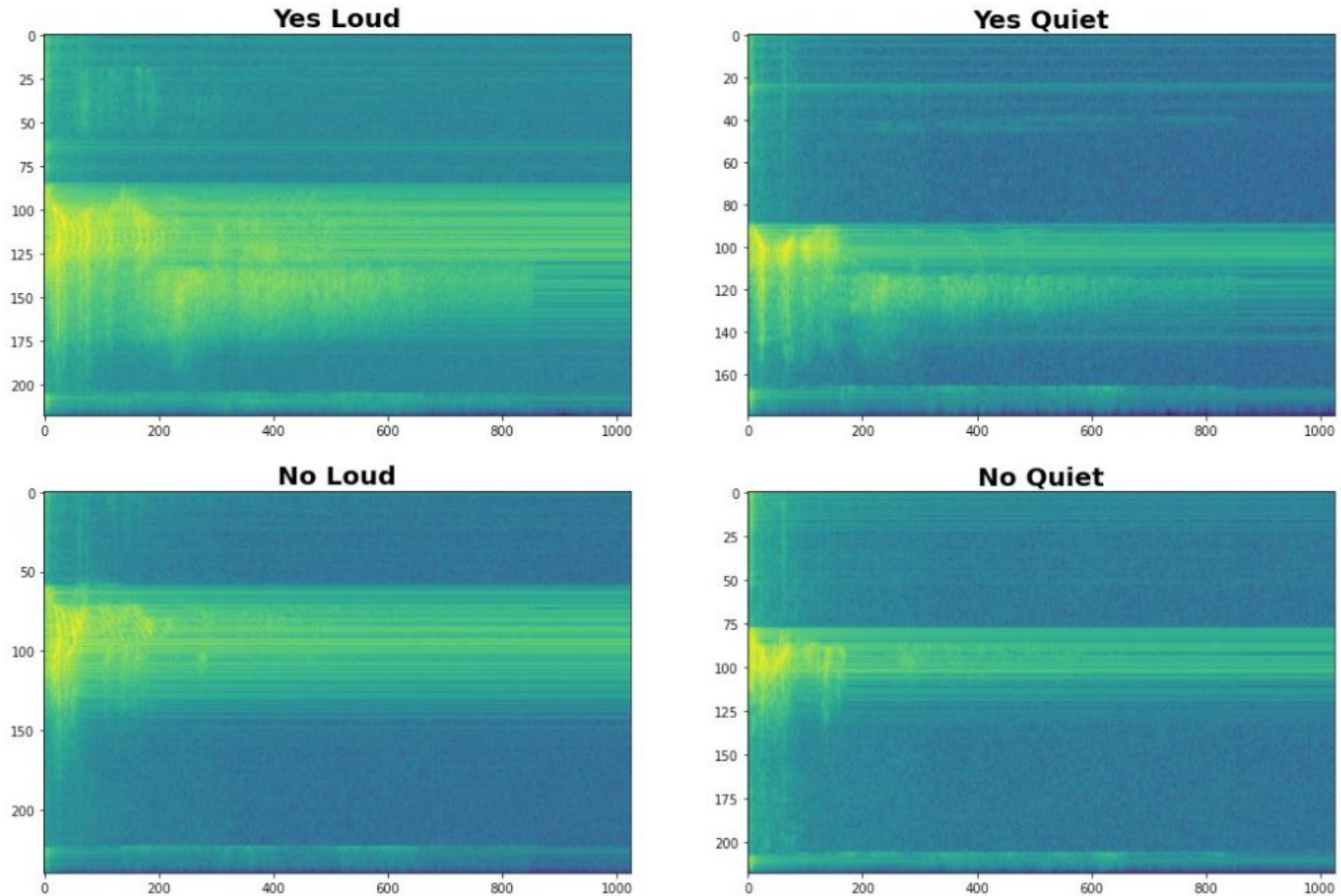




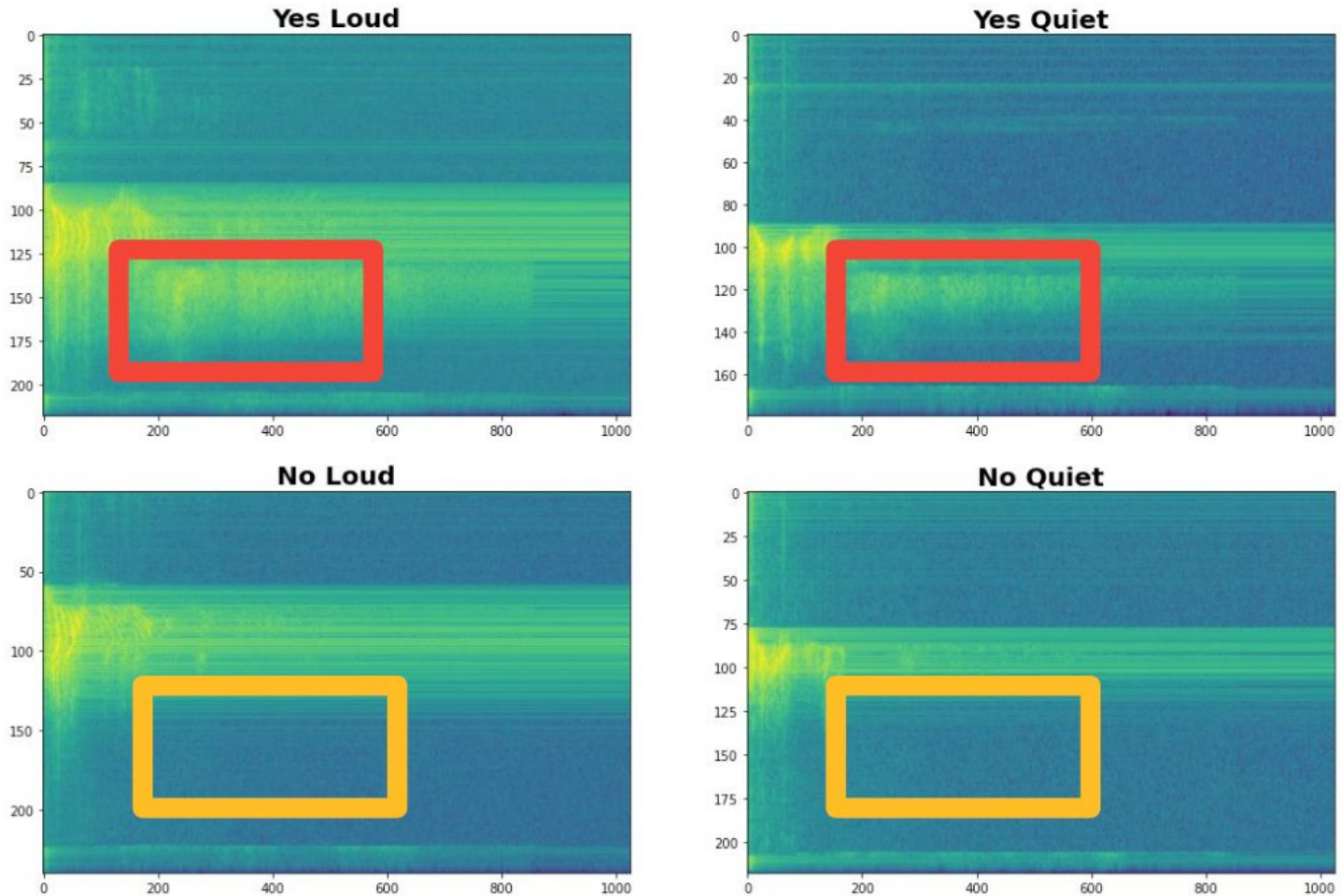
Data Preprocessing: **Spectrograms**

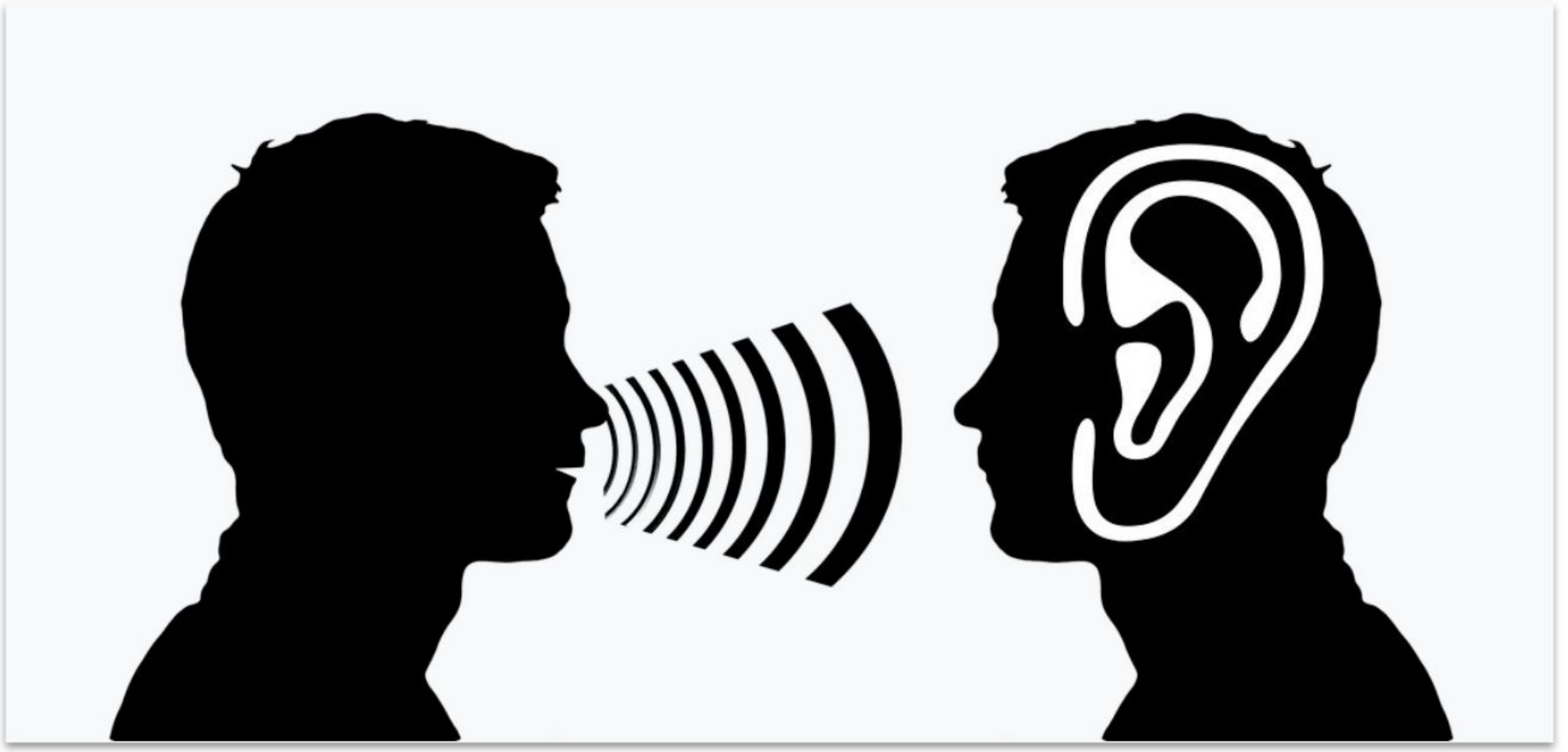


Data Preprocessing: **Spectrograms**



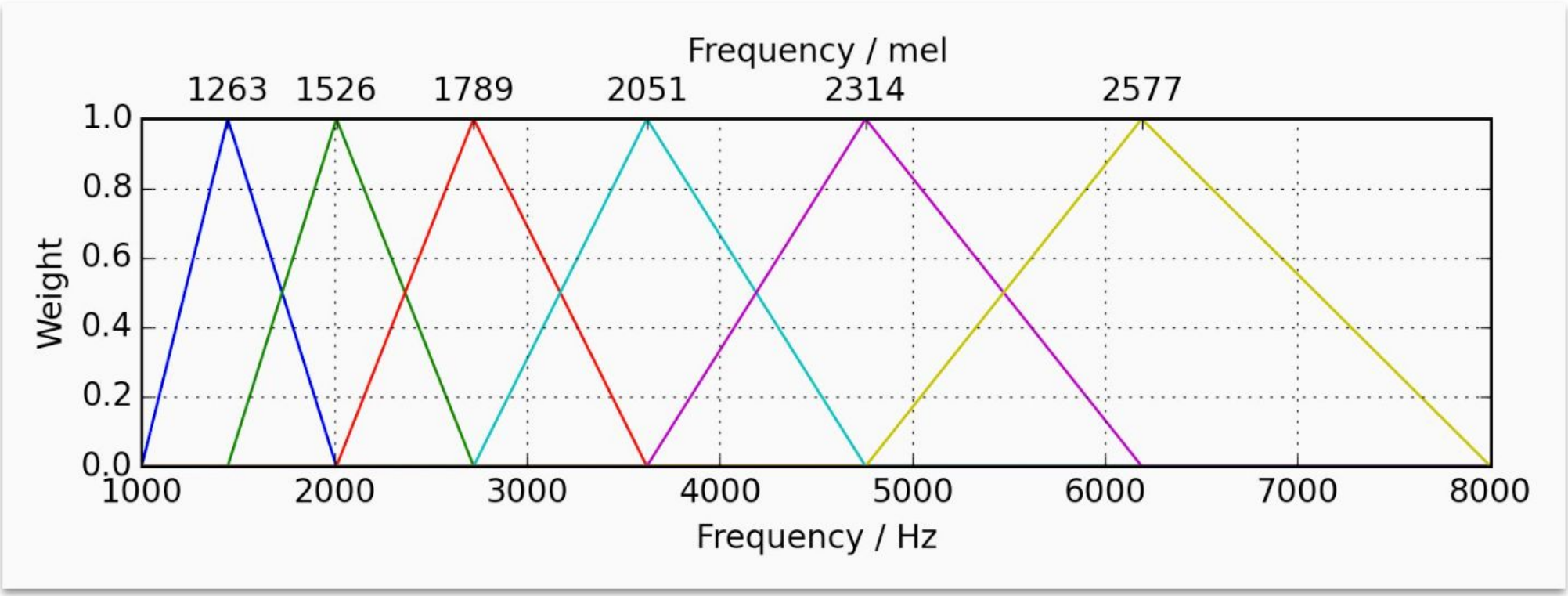
Data Preprocessing: **Spectrograms**



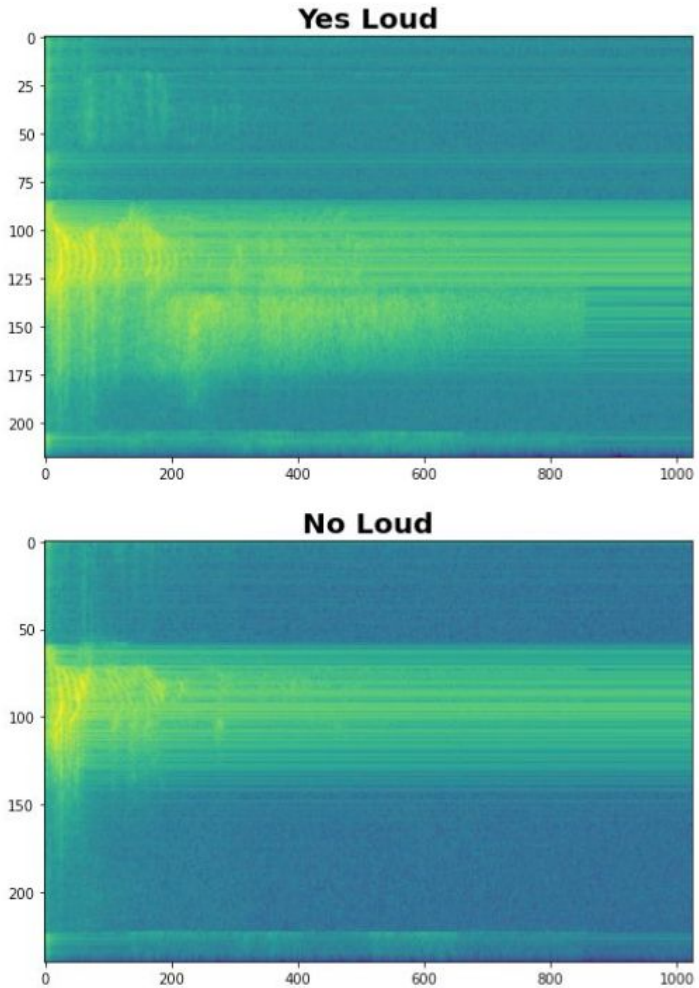


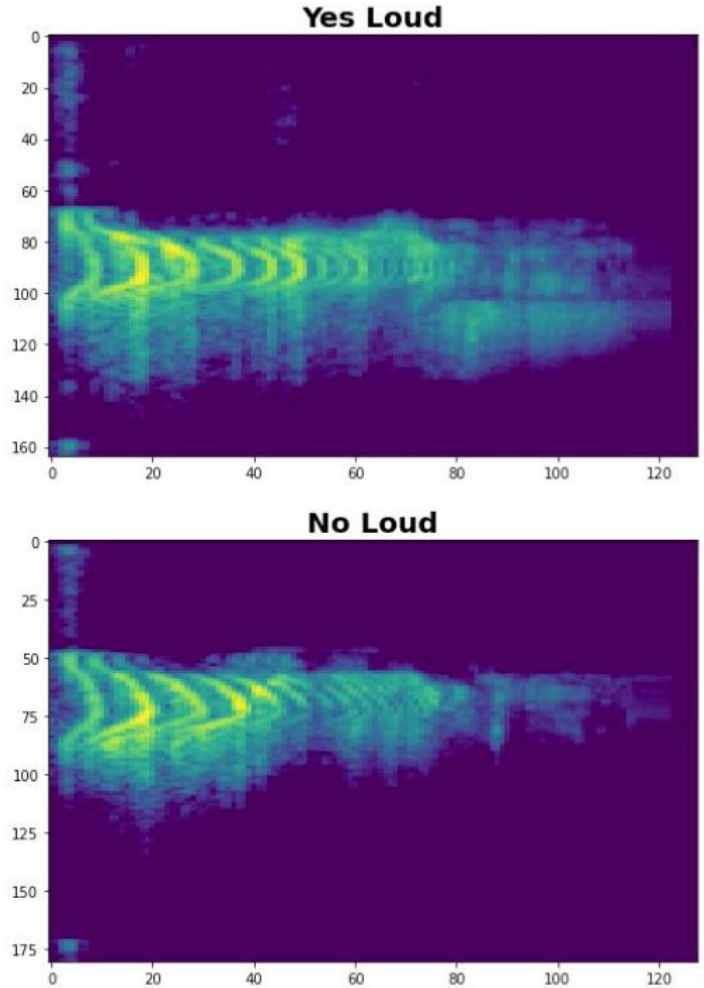
The lower band frequencies is much more crisper to us

Mel Filterbanks

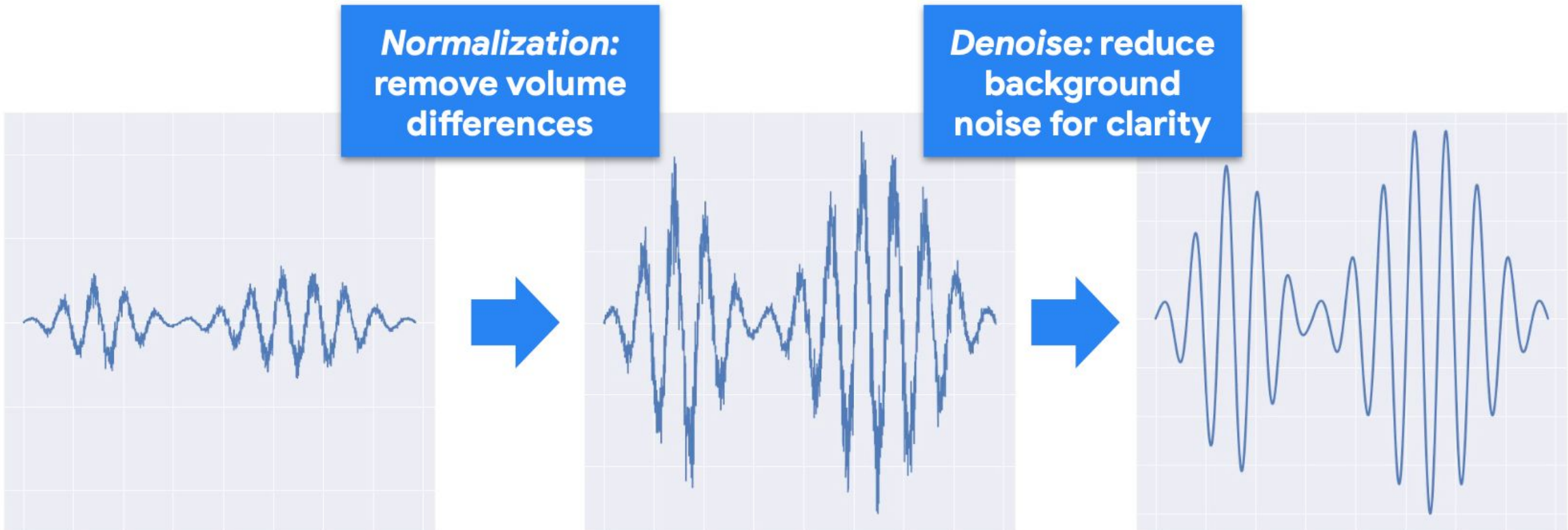


Spectrograms v. MFCCs





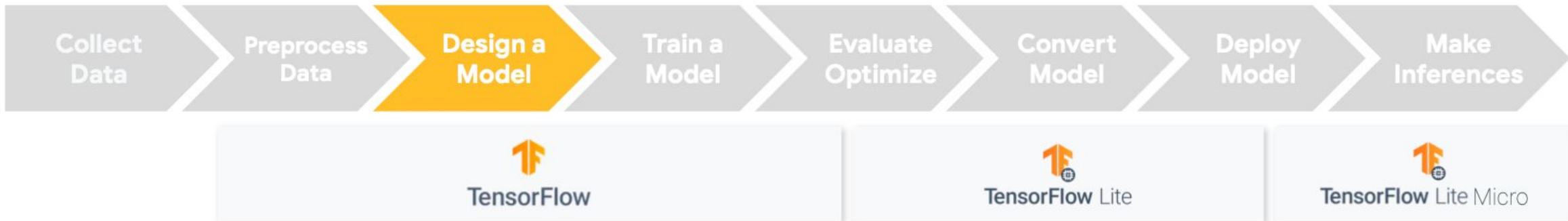
Additional **Feature Engineering**



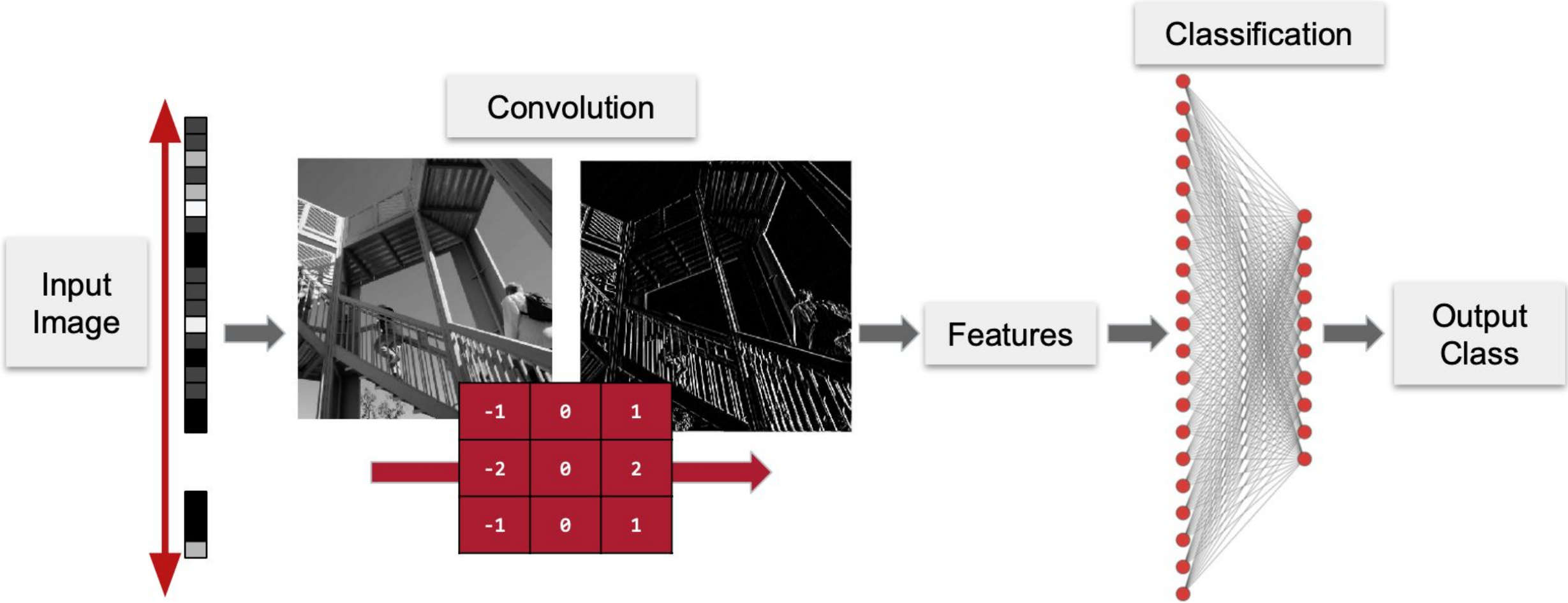
Spectrograms and MFCCs

SpectrogramsMFCCs.ipynb

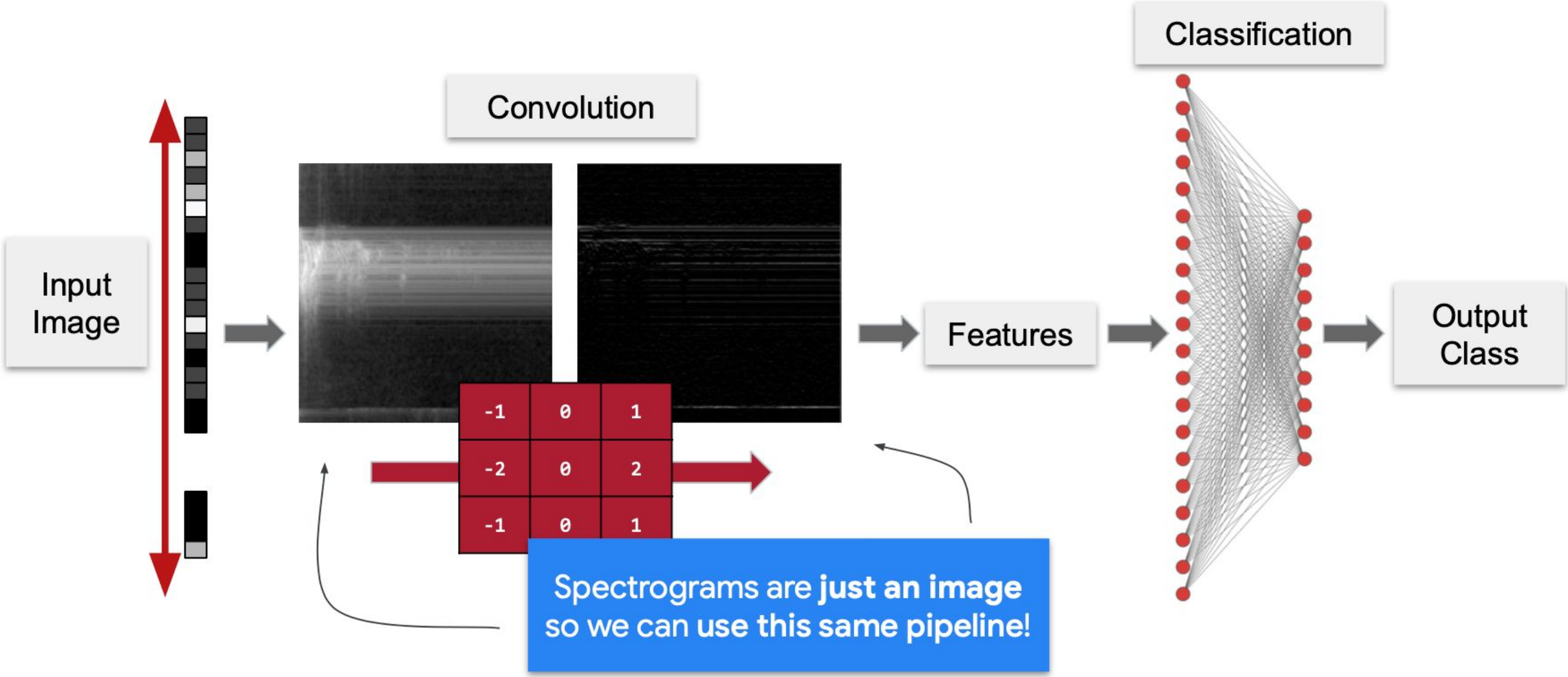
A Keyword Spotting Model


Collect Data
Preprocess Data
Design a Model
Train a Model
Evaluate Optimize
Convert Model
Deploy Model
Make Inferences
TensorFlow
TensorFlow Lite
TensorFlow Lite Micro

A model for **Keyword Spotting**

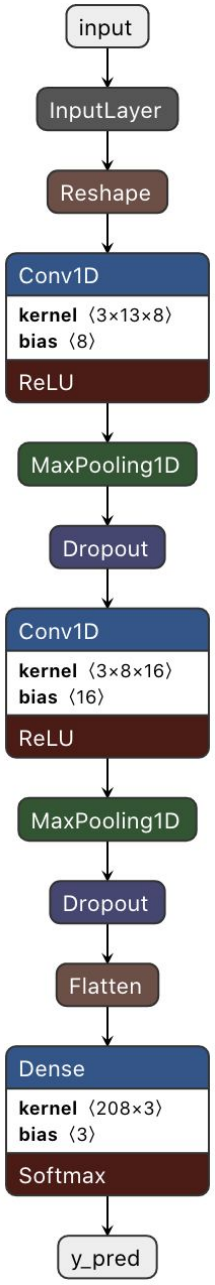


A model for **Keyword Spotting**



```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
reshape (Reshape)            (None, 50, 13)            0
_________________________________________________________________
conv1d (Conv1D)              (None, 50, 8)             320
_________________________________________________________________
max_pooling1d (MaxPooling1D) (None, 25, 8)             0
_________________________________________________________________
dropout (Dropout)            (None, 25, 8)             0
_________________________________________________________________
conv1d_1 (Conv1D)            (None, 25, 16)            400
_________________________________________________________________
max_pooling1d_1 (MaxPooling1 (None, 13, 16)            0
_________________________________________________________________
dropout_1 (Dropout)          (None, 13, 16)            0
_________________________________________________________________
flatten (Flatten)            (None, 208)               0
_________________________________________________________________
y_pred (Dense)               (None, 3)                 627
=================================================================
Total params: 1,347
Trainable params: 1,347
Non-trainable params: 0
_________________________________________________________________
```

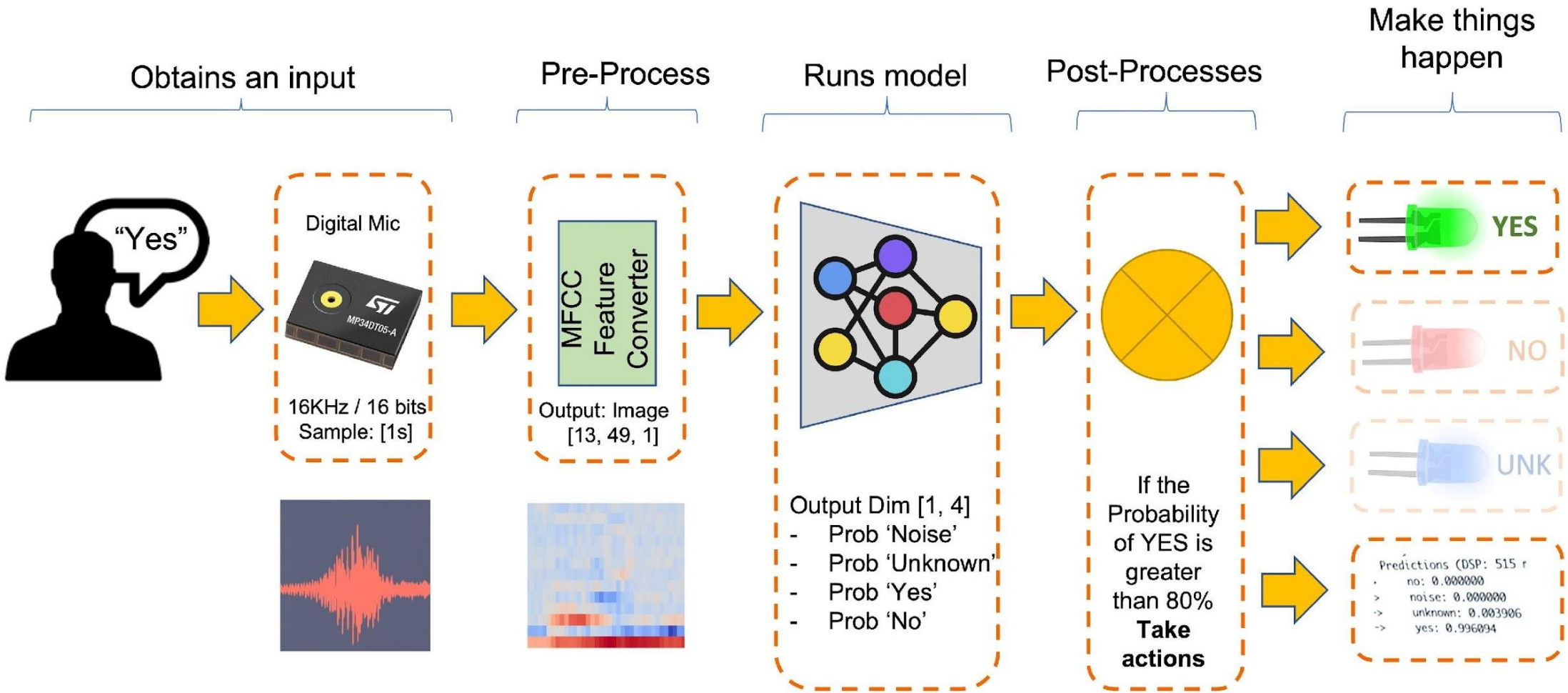
Model size: 200KB



KWS
Keyword Spotting Project

- First keyword: Yes
- Second keyword: No
- Noise Background Noise
- Unknow: (a mix of different words than Yes and No)

https://studio.edgeimpulse.com/public/292418/latest


Obtains an input
Pre-Process
Runs model
Post-Processes
Make things happen
"Yes"
Digital Mic
MP34DT05-A
16KHz / 16 bits
Sample: [1s]
MFCC
Feature
Converter
Output: Image
[13, 49, 1]
Output Dim [1, 4]
- Prob 'Noise'
- Prob 'Unknown'
- Prob 'Yes'
- Prob 'No'
If the Probability of YES is greater than 80% Take actions
YES
NO
UNK
Predictions (DSP: 515 r
no: 0.000000
noise: 0.000000
unknown: 0.003906
yes: 0.996094

Thanks


WALC 2023
Guayaquil, Ecuador
13 al 17 de Noviembre
26°
Taller sobre
Tecnologías de Redes Internet
para América Latina y el Caribe
Fundación
EsLaRed
Organización General