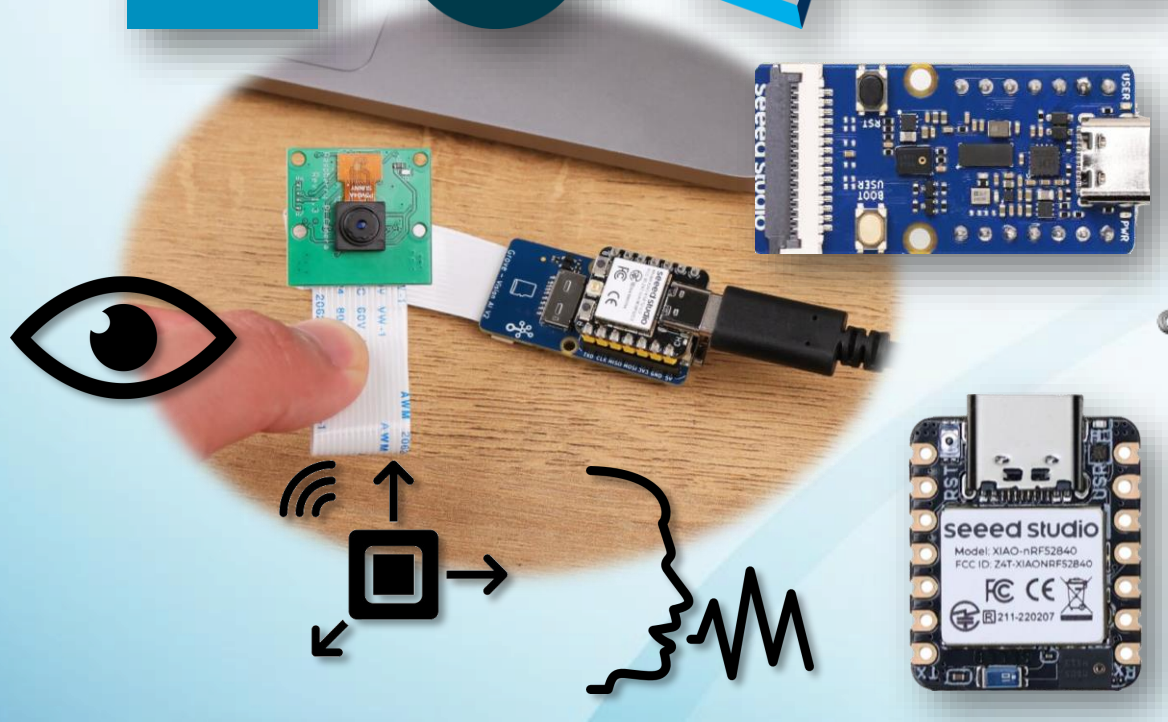
OmniXRI TinyML 小學堂 2025



Cortex-M Processor

Ethos-U MicroNPU

【第 6 講】
TinyML 開發流程

歐尼克斯實境互動工作室 (OmniXRI Studio)
許哲豪 (Jack Hsu)

簡報大綱

- 6.1. Edge Impulse簡介
- 6.2. 資料集建立
- 6.3. 模型選用與訓練
- 6.4. 模型優化及部署

本課程完全免費，請勿移作商業用途！
歡迎留言、訂閱、點讚、轉發，讓更多需要的朋友也能一起學習。

完整課程大綱： https://omnixri.blogspot.com/2025/03/omnixri-tinyml-2025-0.html
課程直播清單： https://www.youtube.com/@omnixri1784/streams

6.1. Edge Impulse 簡介

TinyML 開發流程（簡易版）

- 選用開發板
  - Core
  - GPIO
  - Sensor
  - Comm.
- 選開發平台
  - MCU
  - TinyML

- 資料收集
  - 手機
  - 電腦
  - 開發板
- 資料標註
  - 分割
  - 訓練/驗證/測試集

- DSP前處理
- 選用模型
- 模型訓練
- 調整參數
- 模型優化

（本地 / 雲端）

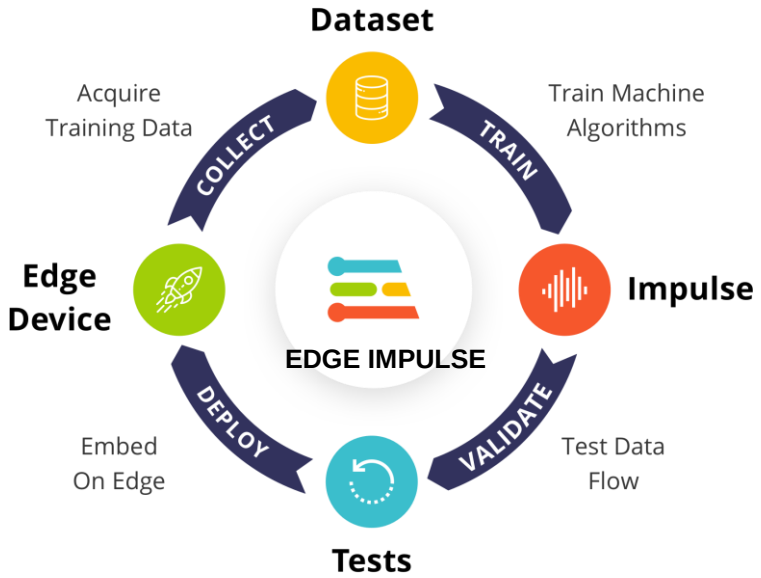
- 格式轉換
  - C/C++
  - Arduino
  - TFLite
  - Others
- 編譯執行
  - No Code
  - Low Code

Edge Impulse Studio

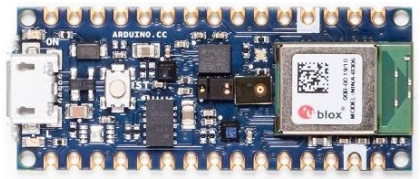
MCU等級AI供應商為tinyML主要成員

提供眾多API及SDK方便開發應用

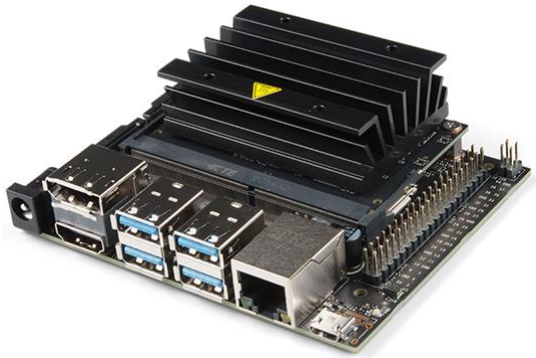
- 動作辨識、關鍵詞偵測、物件偵測等
- 支援多種小型AI及MCU / MPU開發
- 提供豐富線上開發工具



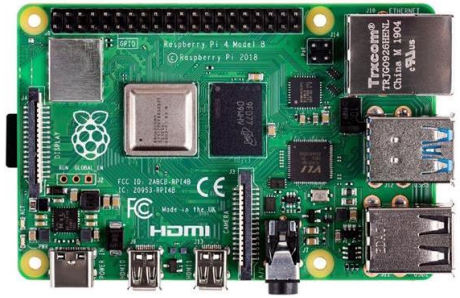
Seeed Grove Vision AI Module V2 kit



Arduino Nano 33 BLE Sense



Nvidia Jetson Nano



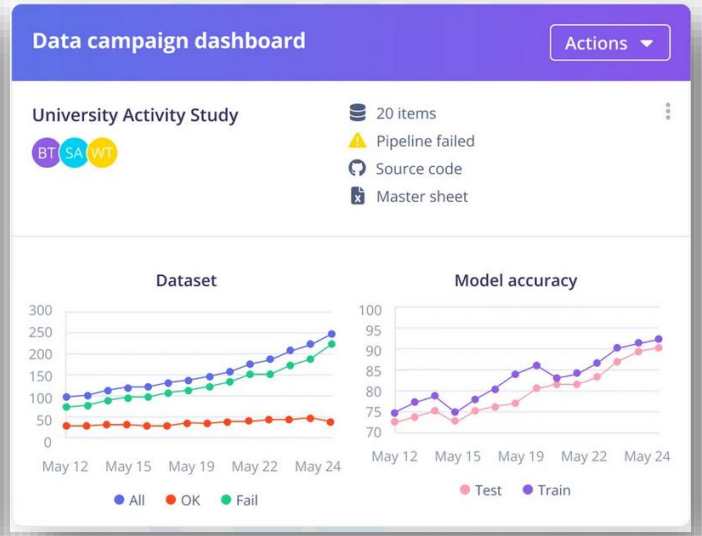
Raspberry Pi 4

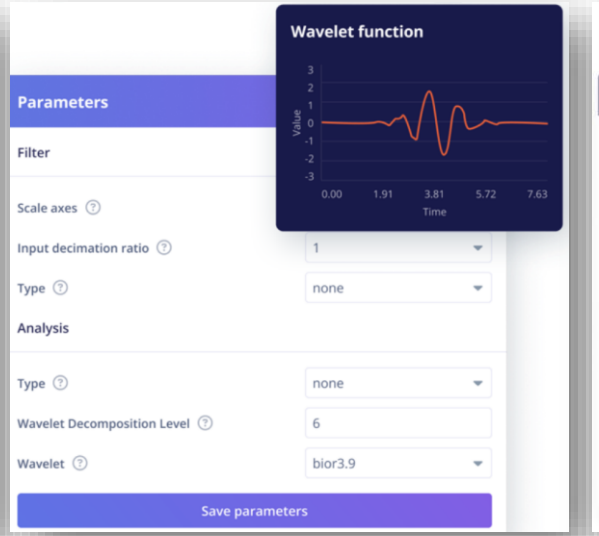
Seeed Xiao nRF52840 Sense

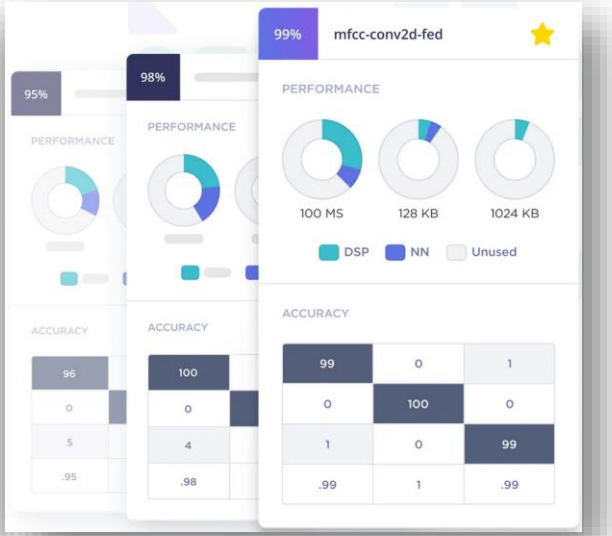
資料來源：https://docs.edgeimpulse.com/docs/

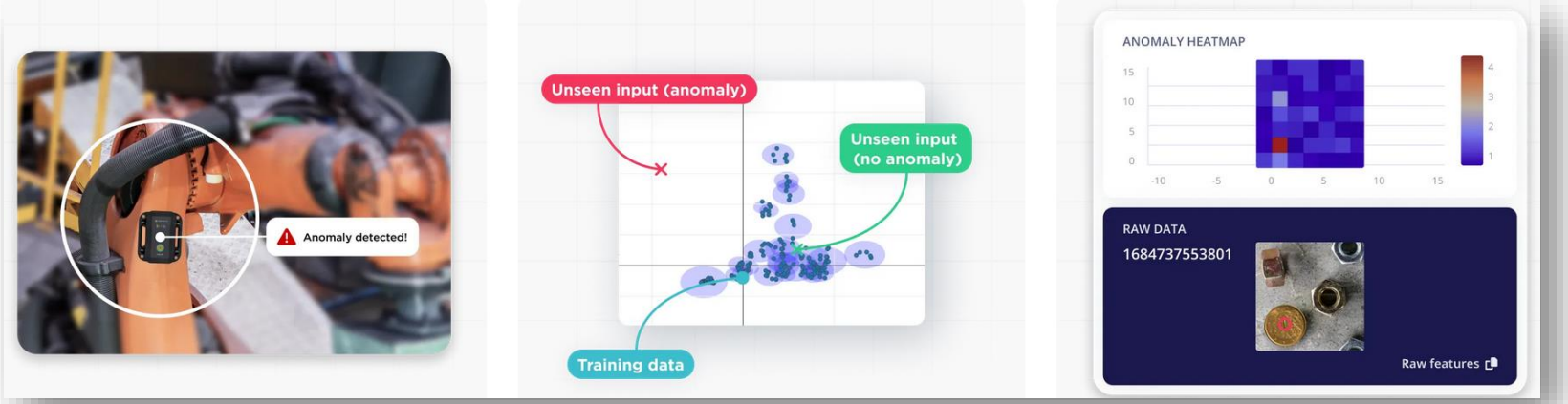
Edge Impulse - 主要功能

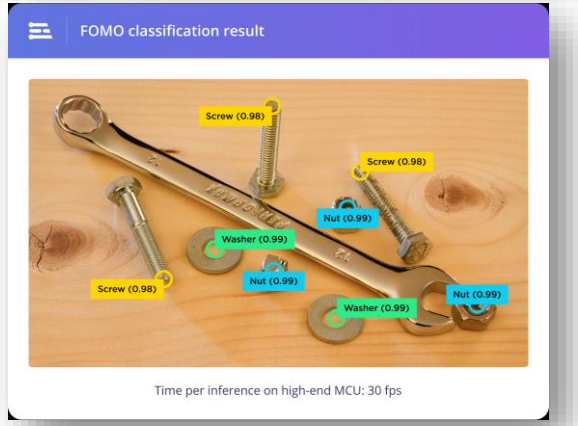
```python
# Quick integration for any Python script
import edgeimpulse as ei

# 1. Set API key from Edge Impulse project
ei.API_KEY = "ei_dae2..."

# 2. Profile model performance on target device
results = ei.model.profile(model="/path/to/model",
                           device="cortex-m4f-80mhz")

# 3. View profiling results
results.summary()

# 4. Optimize model and convert to C++ library
from ei.model.output_type import Classification
ei.model.deploy(model="/path/to/model",
                model_output_type=Classification(),
                output_directory="/library")
```



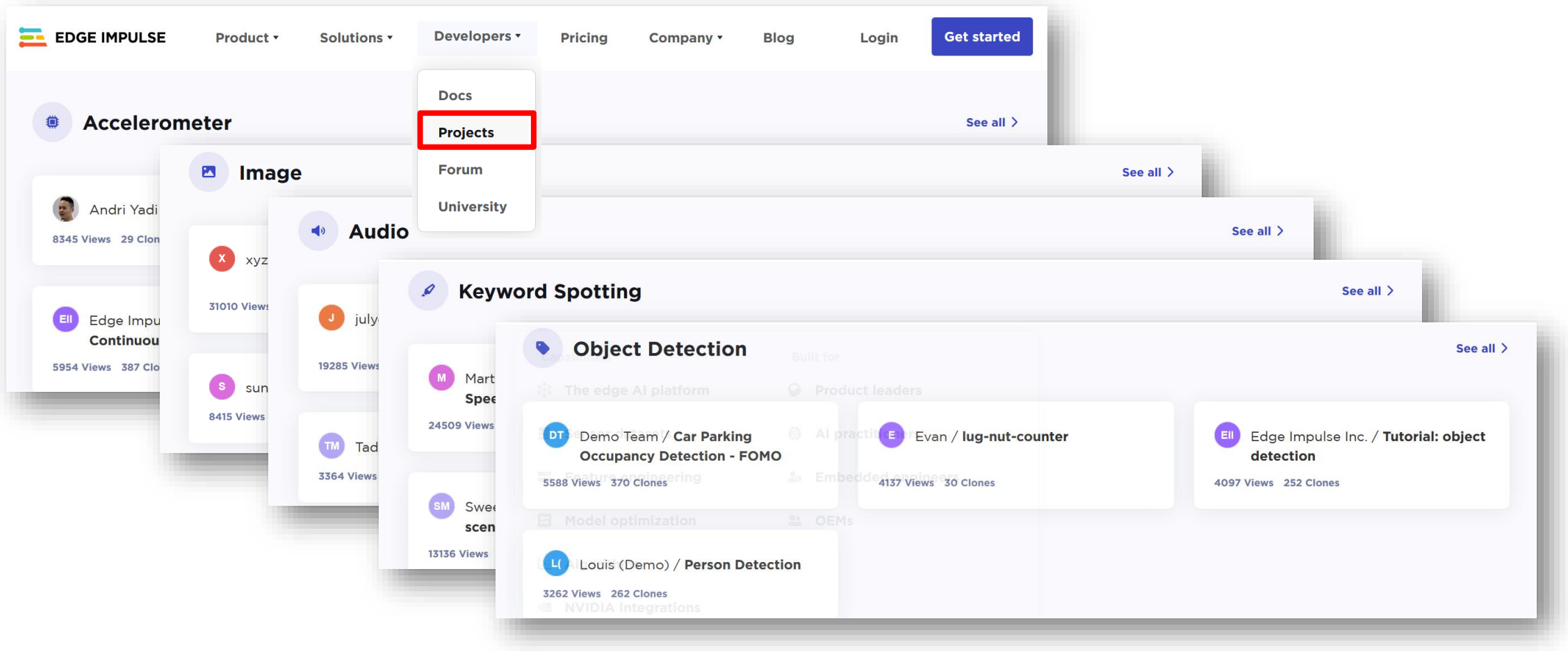






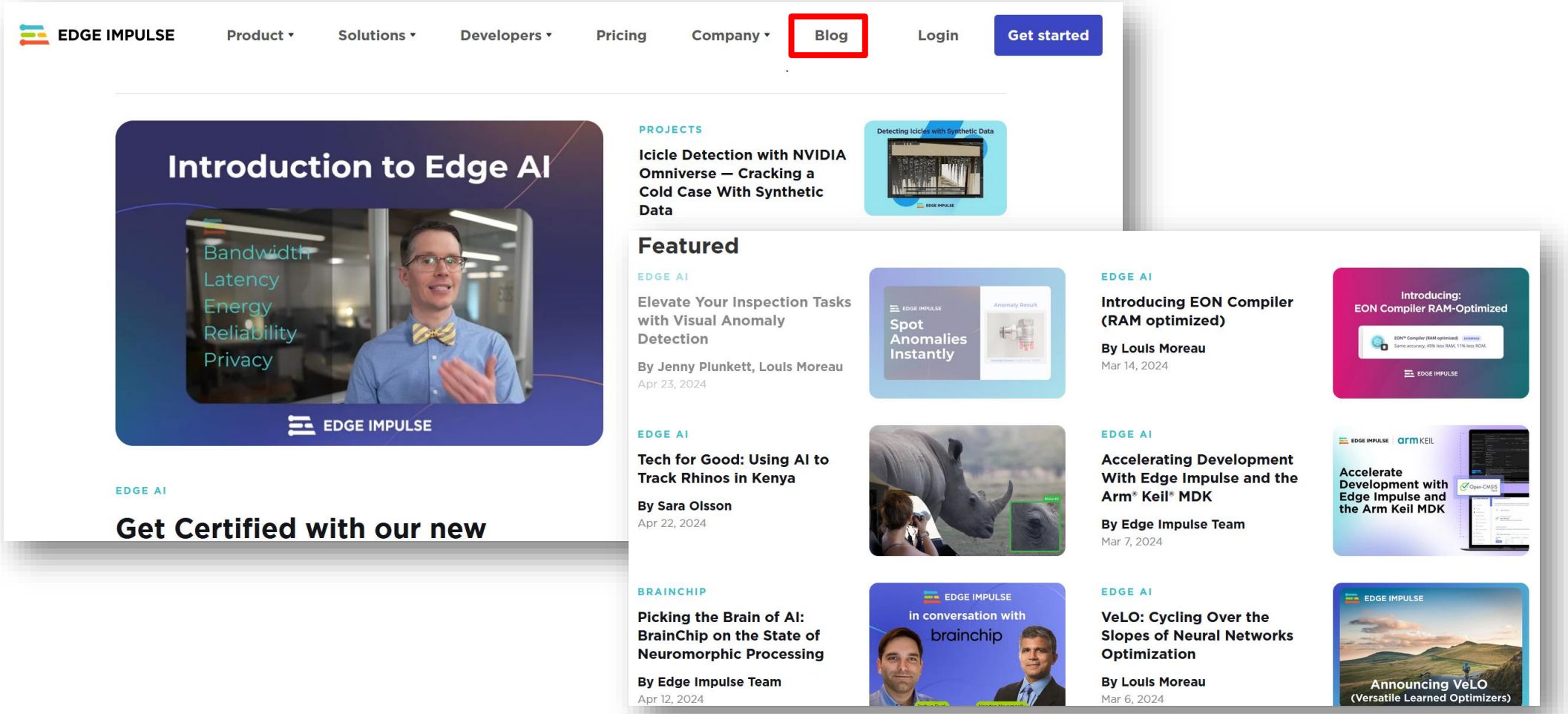


Edge Impulse – 公用專案



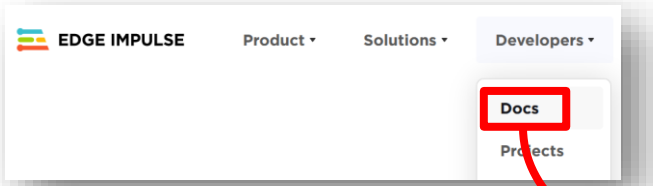
https://edgeimpulse.com/projects/overview

Edge Impulse – 部落格文章



https://edgeimpulse.com/blog/

Edge Impulse – 說明文件



專家案例

https://docs.edgeimpulse.com/docs

EDGE IMPULSE

Getting Started
- For beginners
- For ML practitioners
- For embedded engineers

Frequently asked questions

TUTORIALS
- End-to-end tutorials
- Advanced inferencing
- API examples
- ML & data engineering
- Lifecycle Management
- Expert network projects

EDGE IMPULSE STUDIO
- Organization hub
- Project dashboard
- Devices
- Data acquisition
- Impulse design

Bring your own model (BYOM)
- Processing blocks
- Learning blocks
- EON Tuner
- Retrain model
- Live classification
- Model testing
- Performance calibration
- Deployment

TOOLS
- API and SDK references
- Edge Impulse CLI
- Edge Impulse for Linux
- Edge Impulse Python SDK

RUN INFERENCE
- C++ library
- Arm Keil MDK CMSIS-PACK
- Arduino library
- DRP-AI library
- IAR Library
- Particle Library

OpenMV
- Cube.MX CMSIS-PACK
- Linux EIM Executable
- WebAssembly
- Docker container
- Edge Impulse firmwares
- Hardware specific tutorials

EDGE AI HARDWARE
- Overview
- Production-ready
- MCU
- MCU + AI Accelerators
- CPU
- CPU + AI Accelerators
- GPU
- Mobile Phone
- Porting Guide

INTEGRATIONS
- Arduino Machine Learning Tools
- NVIDIA Omniverse
- Embedded IDEs - Open-CMSIS

Scailable
- Weights & Biases

PRE-BUILT DATASETS
- Continuous gestures
- Running faucet
- Keyword spotting
- TensorFlow Lite reference models

TIPS & TRICKS
- Increasing model performance
- Data augmentation
- Inference performance metrics
- Optimize compute time
- Adding parameters to custom blocks

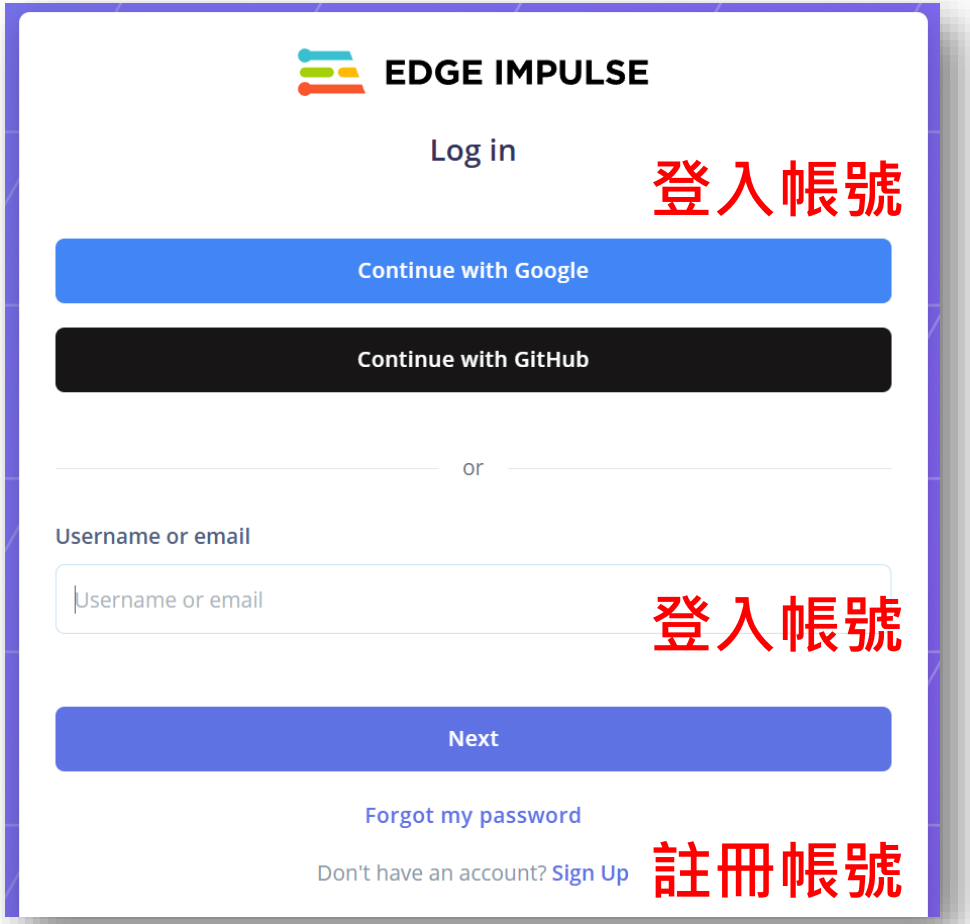
CONCEPTS
- Glossary
- ML Concepts
- Edge AI
- What is embedded ML, anyway?
- What is edge machine learning (edge ML)?

Edge Impulse – Expert network projects

- Featured Machine Learning Projects
- Prototype and Concept Projects
- Image Projects
- Audio Projects
- Predictive Maintenance & Fault Classification
- Accelerometer & Activity Projects
- Air Quality & Environmental Projects
- Novel Sensor Projects
- Software Integration Demos

https://docs.edgeimpulse.com/experts

Edge Impulse - 註冊/登入帳號



免費帳號使用限制

- **2組**私人專案
- 公用專案不限
- 最多**3個**成員(限公用專案)
- 每個專案 CPU 計算用記憶體 4GB 或 **4小時**資料
- 每個專案計算時間**20分鐘**
- EON Tuner, Compiler 有限制

https://studio.edgeimpulse.com/login

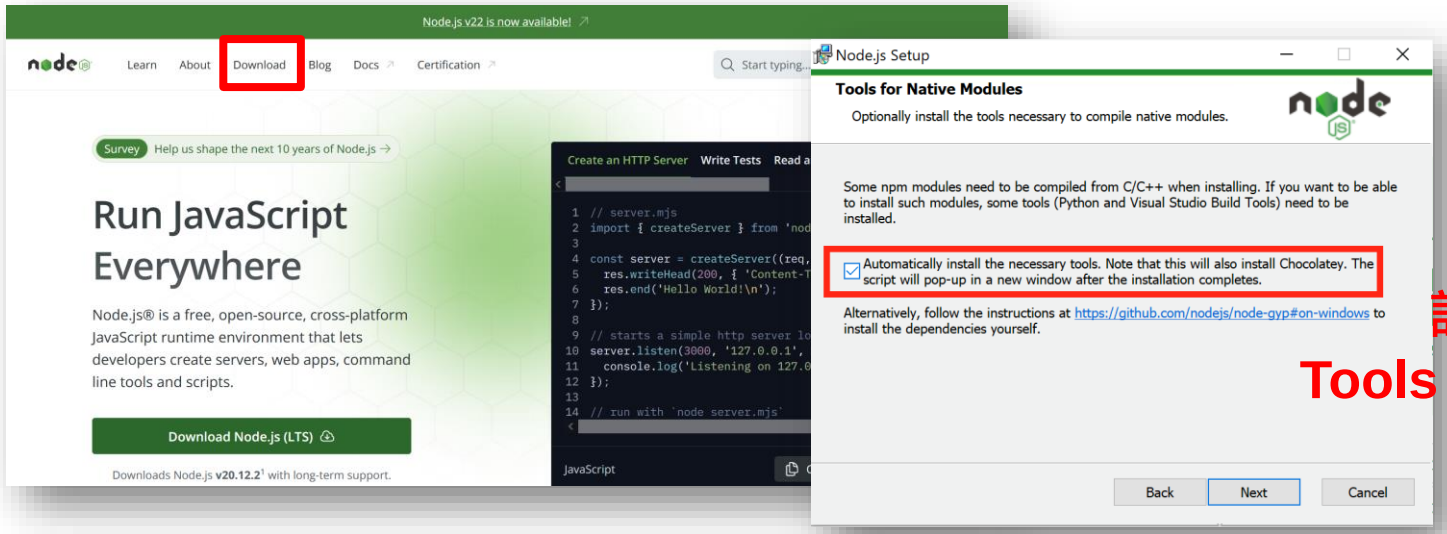
https://edgeimpulse.com/pricing

下載及安裝必要工具

1. Python 3.x

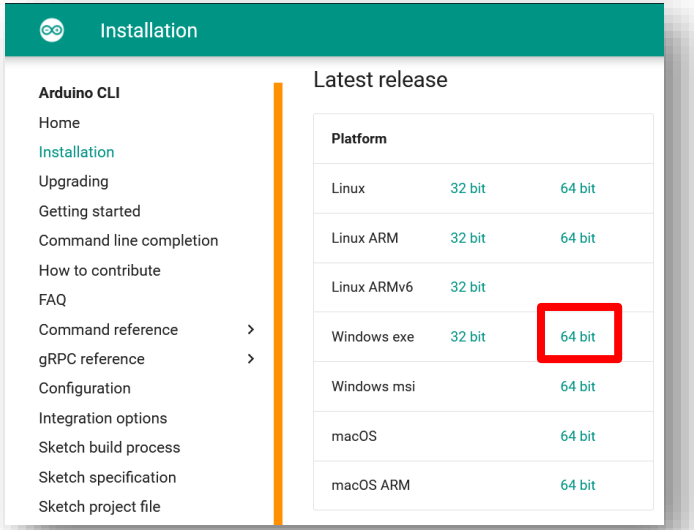
2. Node.js

(若後續 Edge Impulse 無法安裝 可換回舊版，例： v18.20.2 LTS)



3. Arduino CLI

(選配，依不同開發板需求)



安裝時要注意
記得一定要勾選
Tools for Native Modules

以 node -v 及 npm -v
檢查安裝好的版本

安裝 Edge Impulse 工作環境

- 安裝Edge Impulse CLI

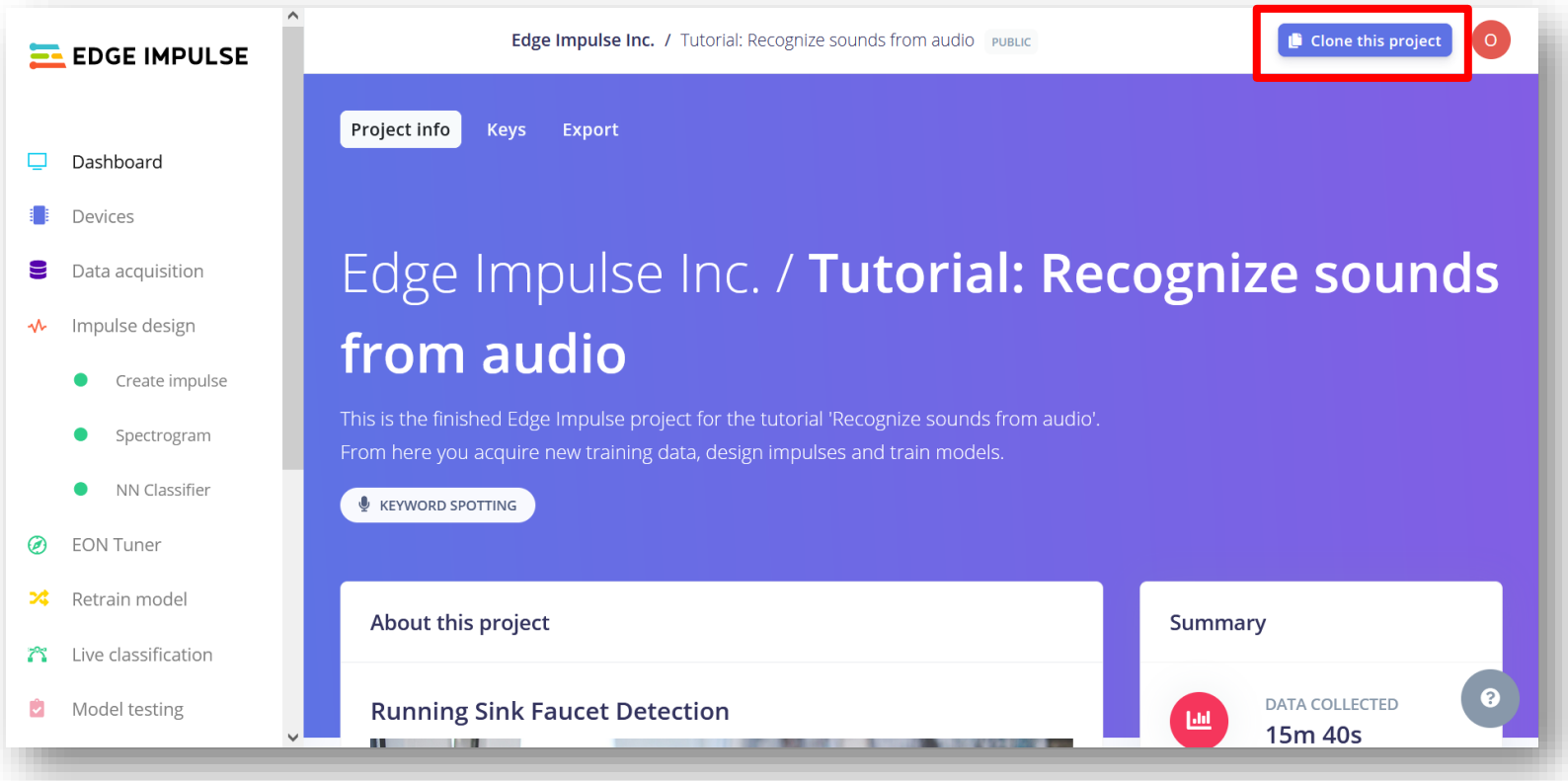
  npm install -g edge-impulse-cli –force

- 安裝完成後會得到下列工具
  - edge-impulse-daemon 透過 COM 埠和雲端平台連接
  - edge-impulse-uploader 允許上傳本機檔案
  - edge-impulse-data-forwarder 透過 COM 埠上傳開發板感測器數值
  - edge-impulse-run-impulse 顯示裝置上運行的模型
  - edge-impulse-blocks 建立自定義區塊
  - himax-flash-tool 燒錄 Haimax WE-I Plus

https://docs.edgeimpulse.com/docs/tools/edge-impulse-cli/cli-installation

啟動公用專案範例 – 水流聲辨識

Edge Impulse Inc. / Tutorial: Recognize sounds from audio



複製專案到
個人帳號

水流聲辨識

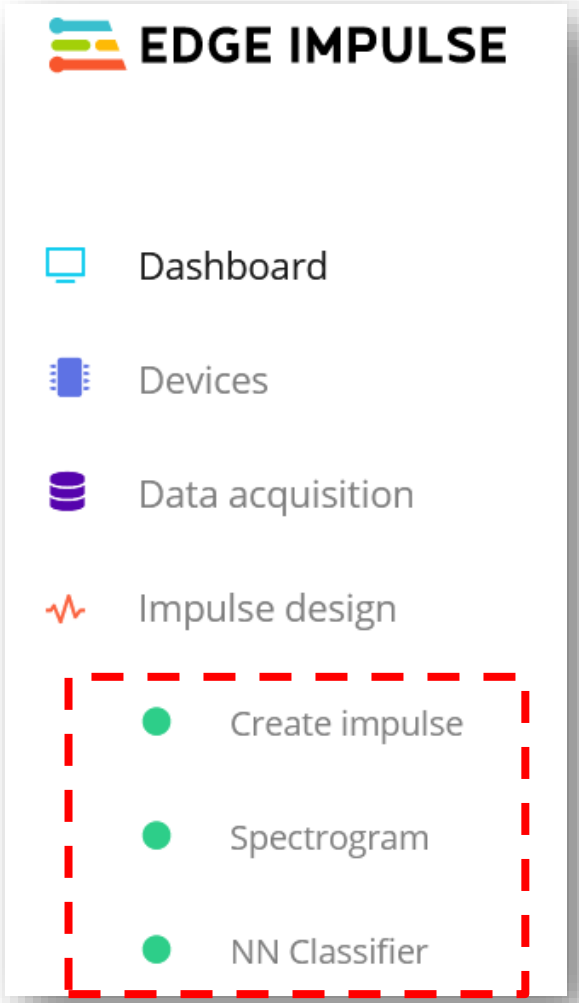
https://studio.edgeimpulse.com/public/14301/latest

專案工作流程

水流聲辨識

1. 準備開發板
2. 收集資料
3. 建立資料集
4. 設計及選用模型
5. 配置MFE模塊
6. 配置神經網路
7. 分類新資料
8. 模型測試
9. 模型故障排除
10. 部署到指定裝置

Edge Impulse Studio – 專案功能表



專案儀錶板

連線裝置

資料擷取

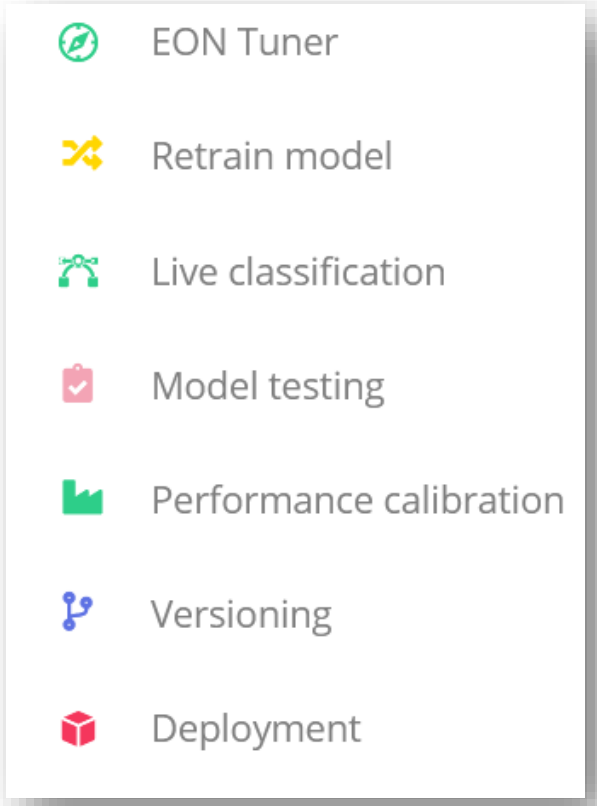
流程設計

建立流程

資料頻譜

神經網路分類器

不同類型專案項目不同



EON調整器

重新訓練模型

立即分類

模型測試

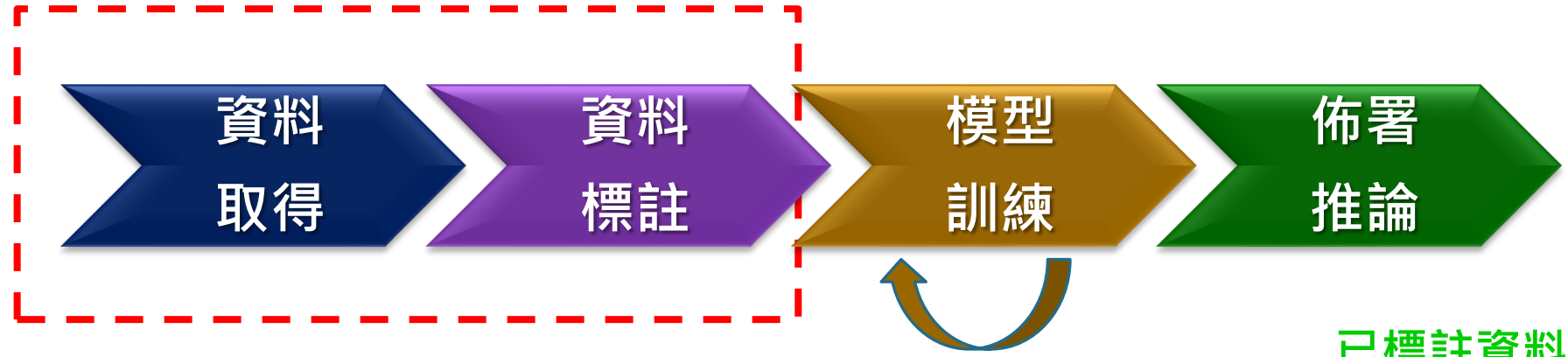
效能效正

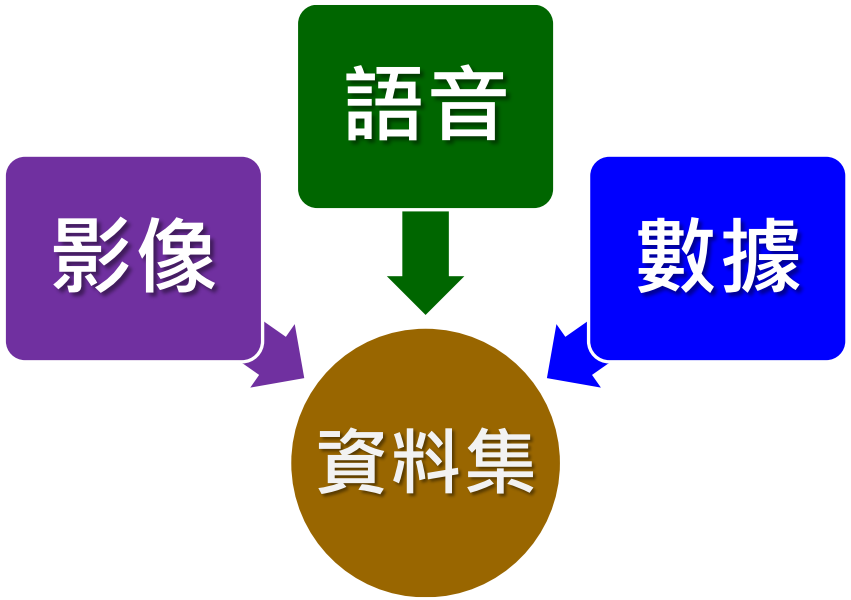
版本管理

程式部署

6.1. 資料集建立

常見資料類型





常見資料結構

格式化資料

- 一維時序資料（感測器）

  光電、溫度、濕度、壓力、

  聲音、速度、運動...
- 二維 / 三維影像資料

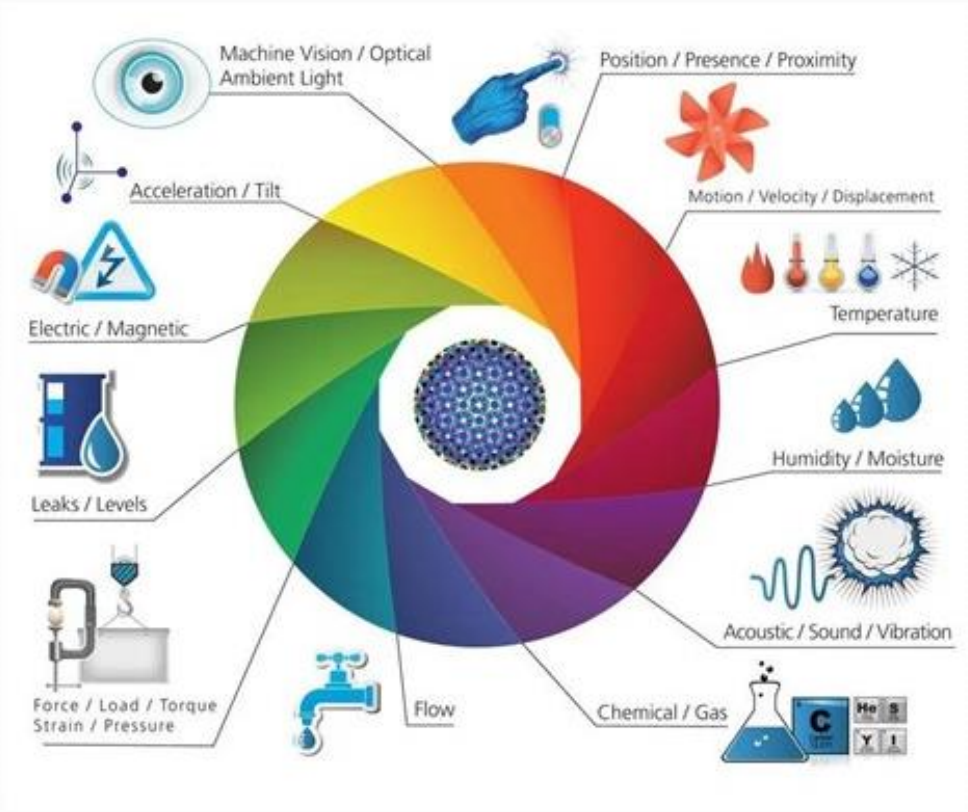
  可見光、紅外光、深度影像
- 影像序列資料（影片）

  原始、壓縮、串流
- 多維表單數據（資料庫）

非格式化資料

- 自然語言、文章
- 影像、影片內容
- 非定時定量收集的資料
- 偶發性資料

常見感測器



影像來源：https://www.edntaiwan.com/20190201ta31-designers-guide-to-iiot-sensor-systems/

- A. 電阻、電容、電感式
- B. 光電式
  - 發射接收式、紅外熱幅射式
- C. 壓電式
- D. 電聲式
  - 超音波、麥克風
- E. 微機電式
  - 運動類、環境類
- F. 電磁波式
- G. 影像式

資料來源：https://omnixri.blogspot.com/2022/11/20221108.html

資料取得方式

資料集來源

- 公開資料集
- 私有資料庫
- 網路收集
- 自行拍攝取像
  - 專業相機（手機、相機、攝影機）
  - 網路攝影機（USB）
  - 開發板專用相機 （CSI）
    - 可見光
    - 紅外線
- 感測器資料
  - 網路連接、MQTT

取像要領

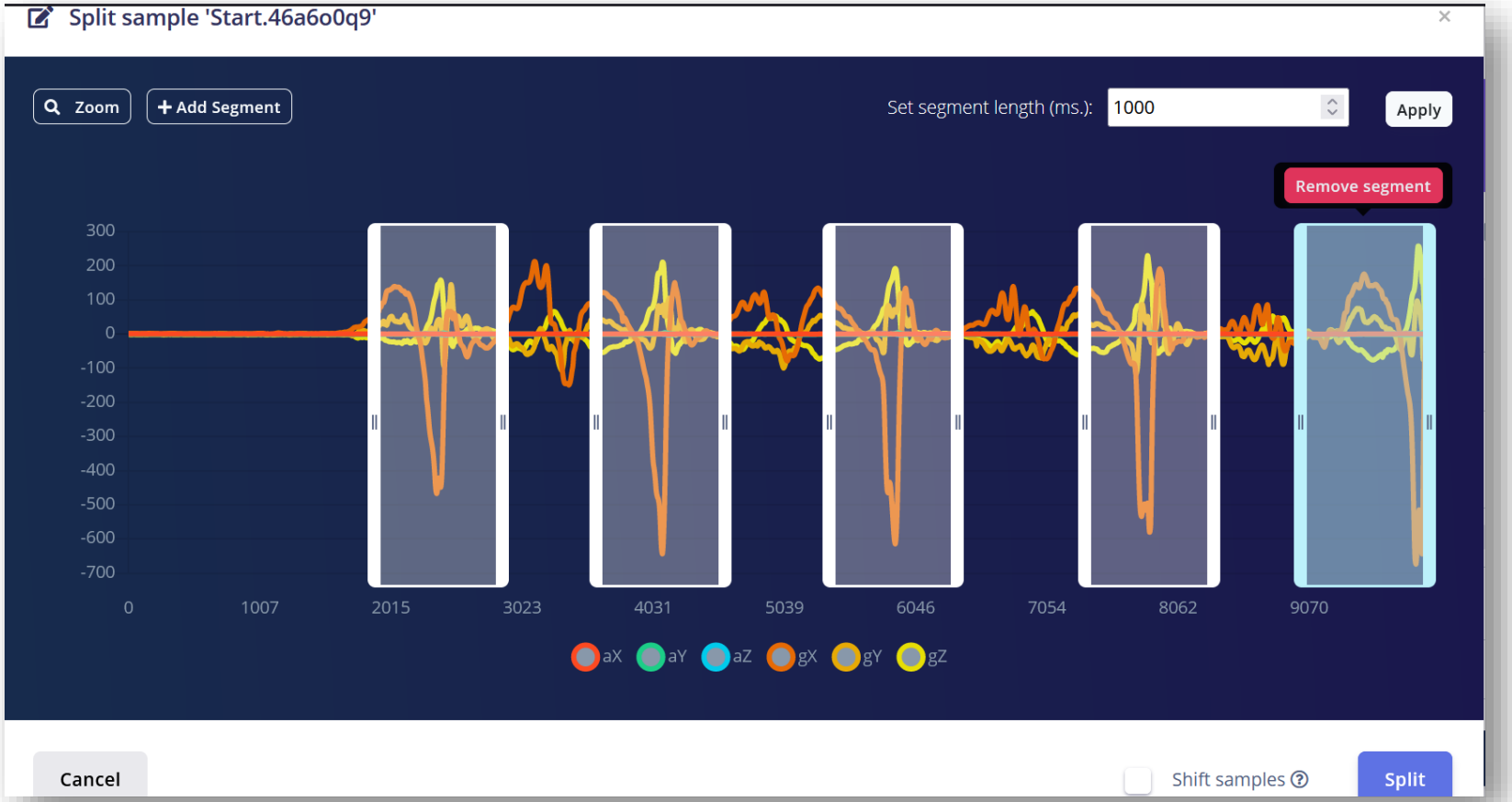
- 取像品質近似
  - 解析度、模糊度...
- 多視點
  - 遠近、角度、 物件正反面...
- 多光源差異
  - 環境亮度、陰影、光圈、快門...
- 多背景、品種

**** 建立足夠多樣性的資料集 ****

資料清洗

- 去冗、填空、修補、抑雜訊

感測器資料標註



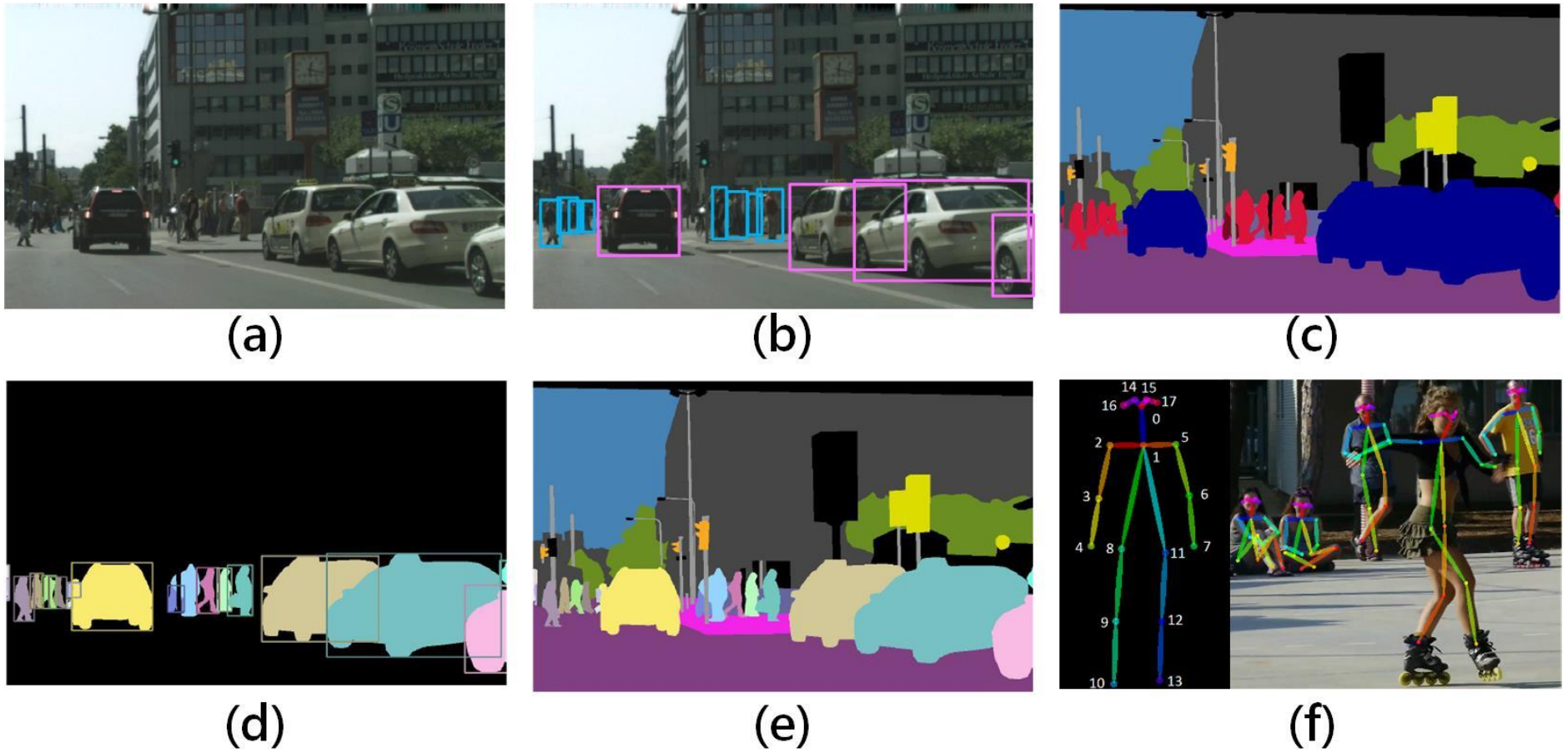
一維？多維？

連續？片段？

手動/自動分割？

特徵時長？

影像資料標註

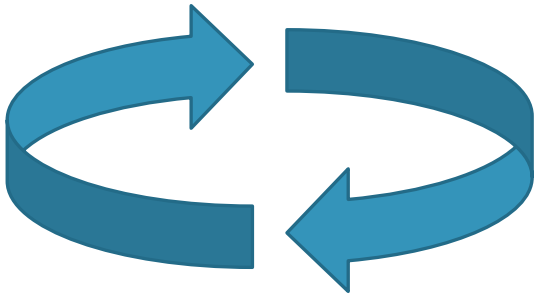


(a)原始影像/影像分類，(b)物件偵測，(c)語義分割，(d)實例分割，(e)全景分割，(f)人體姿態（關鍵點）。

訓練、驗證及測試資料集

已標註之資料可依一定比例分配到訓練集、驗證集及測試集中。**絕對不可把測試集加入訓練集或驗證集中。**

分配比例**沒有一定標準**，小規模資料集建議**訓練:驗證:測試**為6:2:2。大規模（數百萬筆）資料集，驗證、測試各留數萬筆即可。

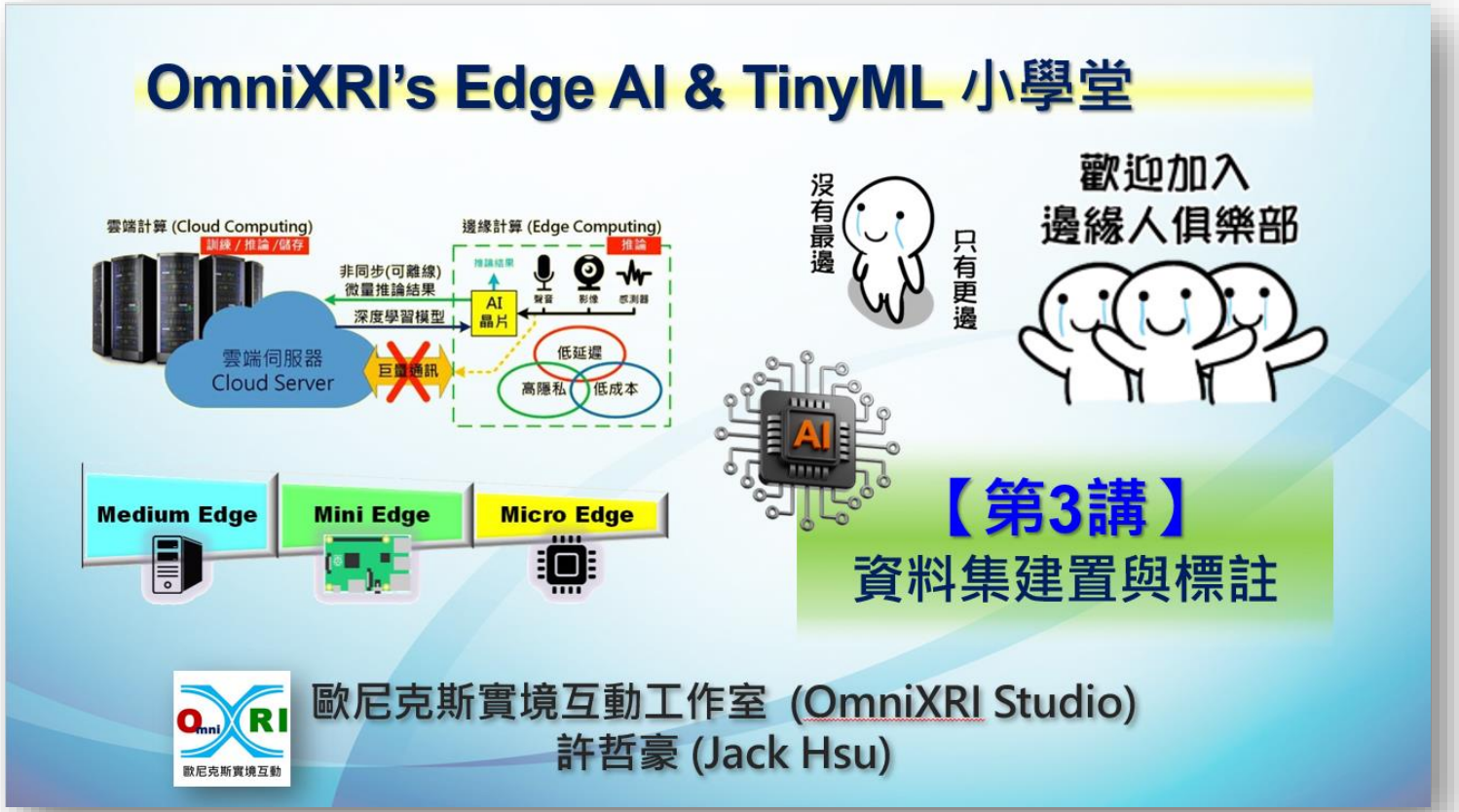


訓練集：就像在學校學習，了解問題，找出學習重點。

驗證集：就像模擬考，測試學習成果，找出修正方向。

測試集：就像大考，出題和模擬考接近程度會影響最後成績。

更多資料集建置與標註



3.1. 資料集建置

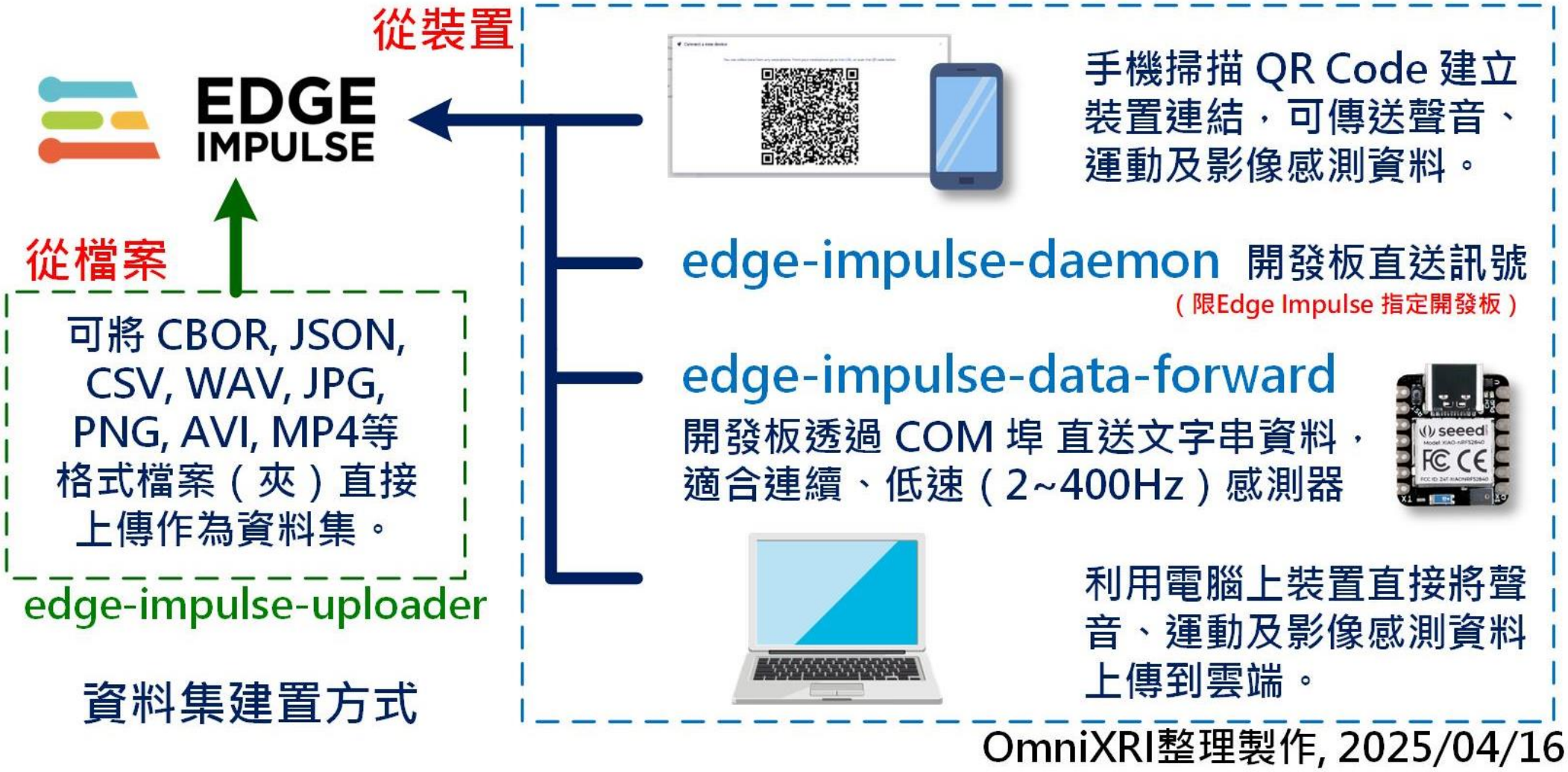
3.2. 公開資料集

3.3. 資料集標註

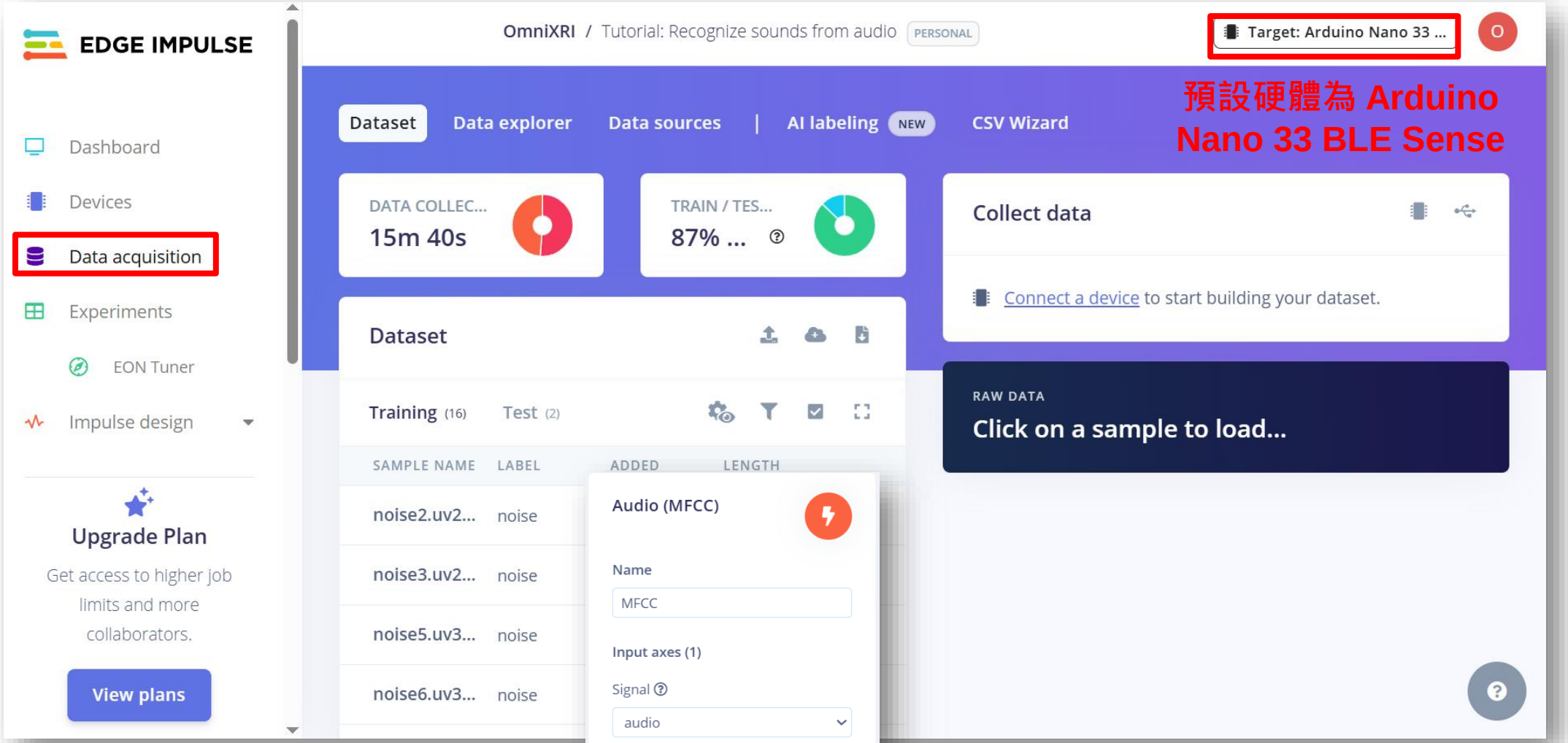
3.4. 資料集迷思

OmniXRI's Edge AI & TinyML 小學堂 (2024)【第3講】資料集建置與標註
https://youtu.be/d655nS-0XmM

Edge Impulse 資料集建置方式

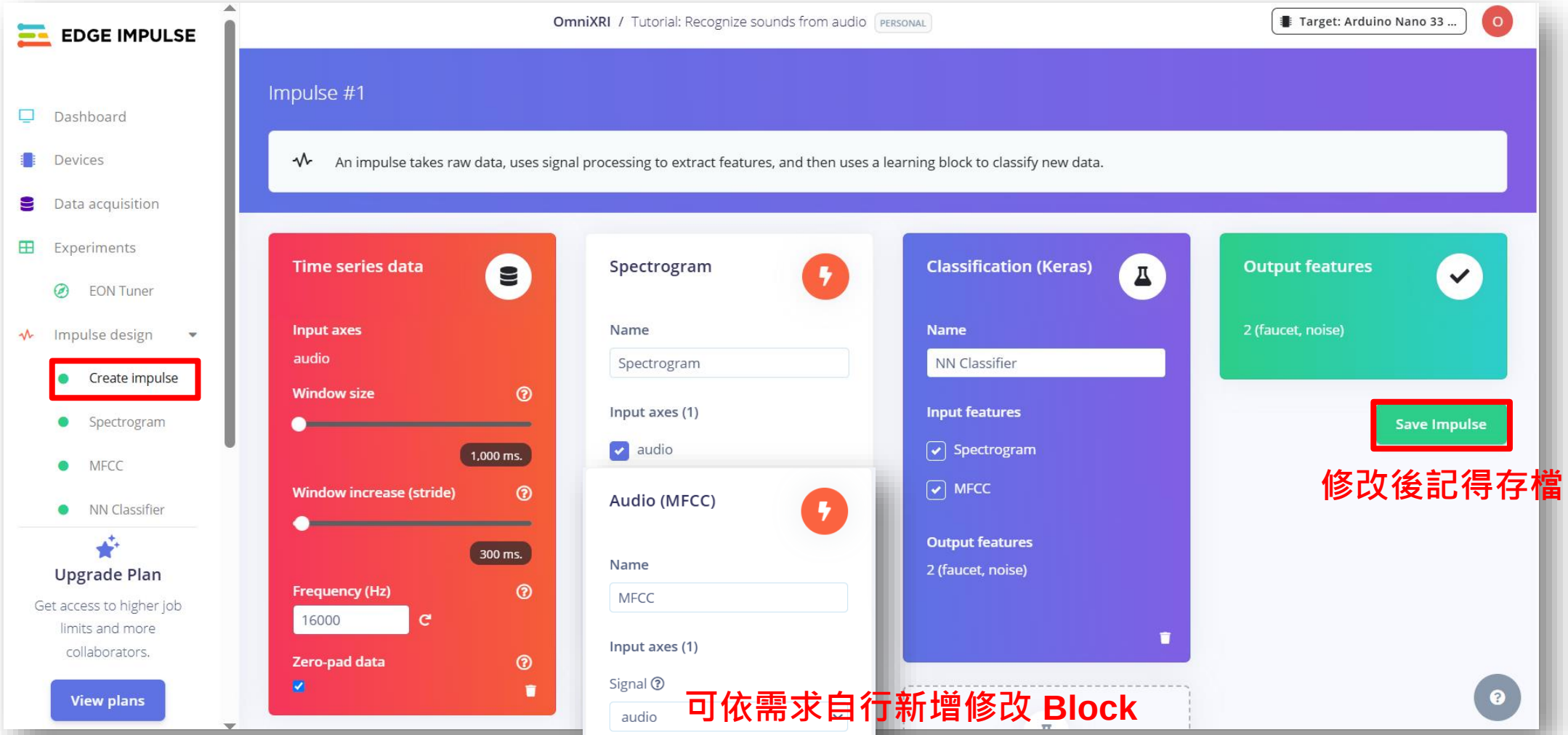


水流聲辨識範例 – 資料收集

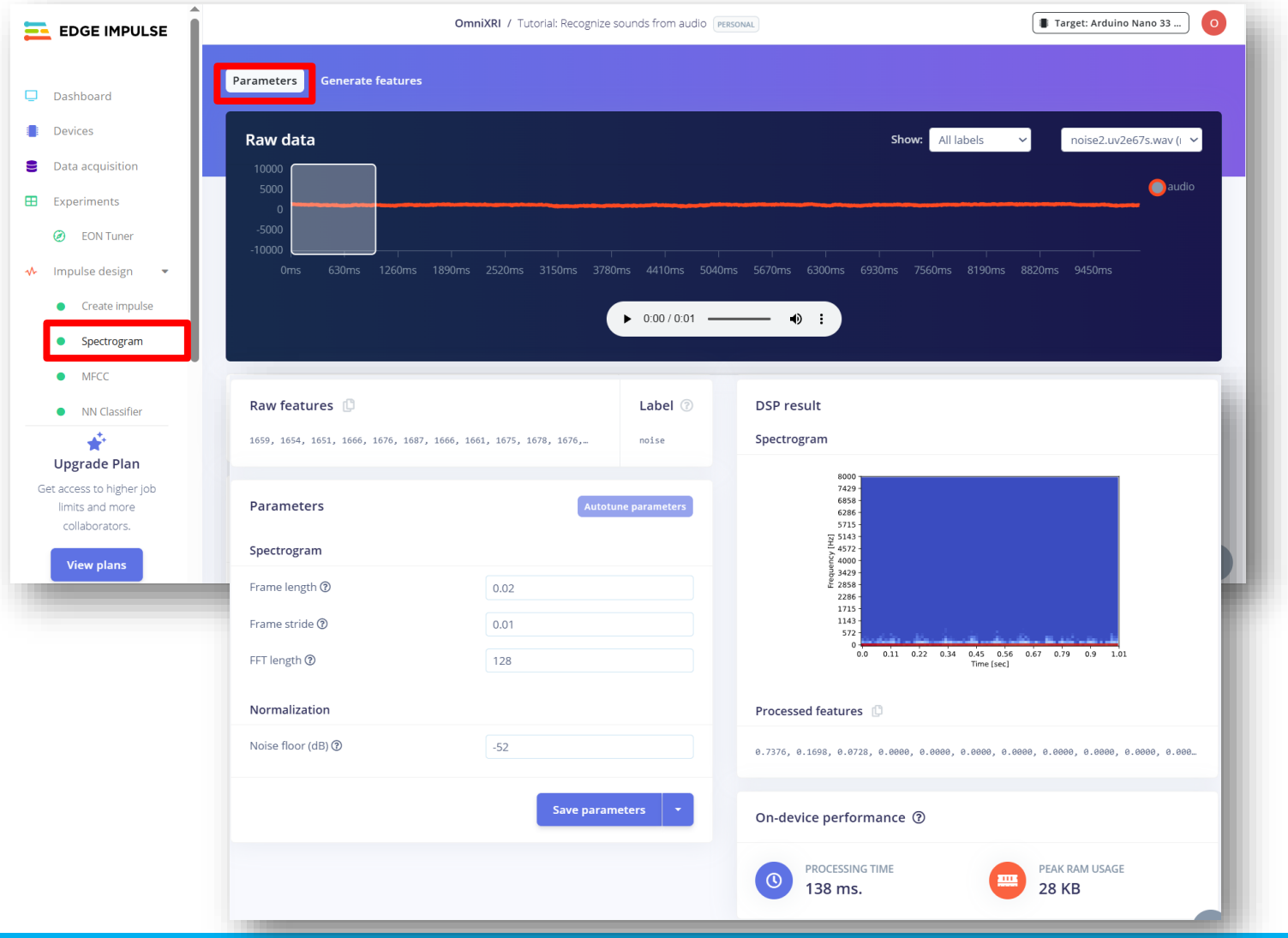


6.2. 模型選用與訓練

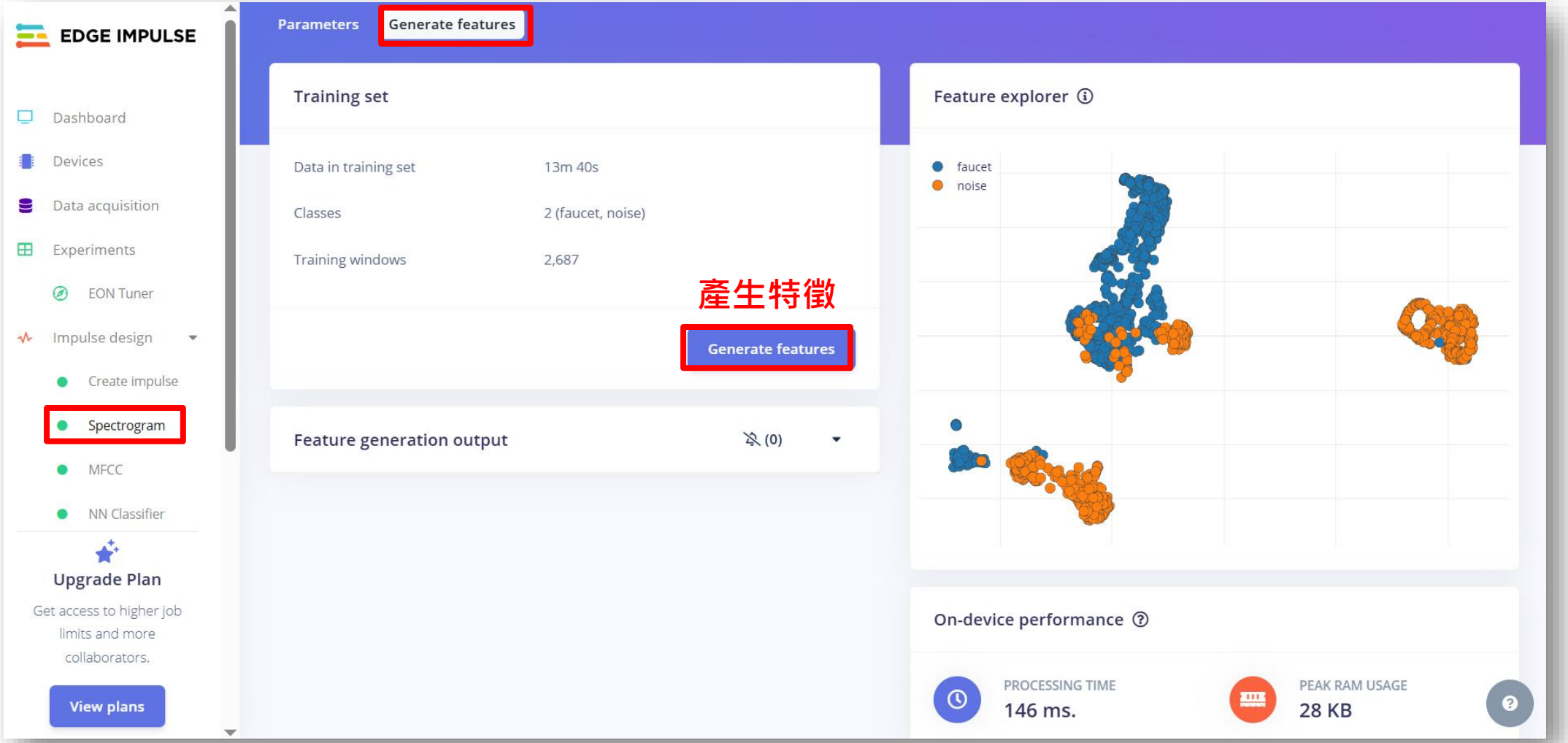
選擇模型及設定必要參數



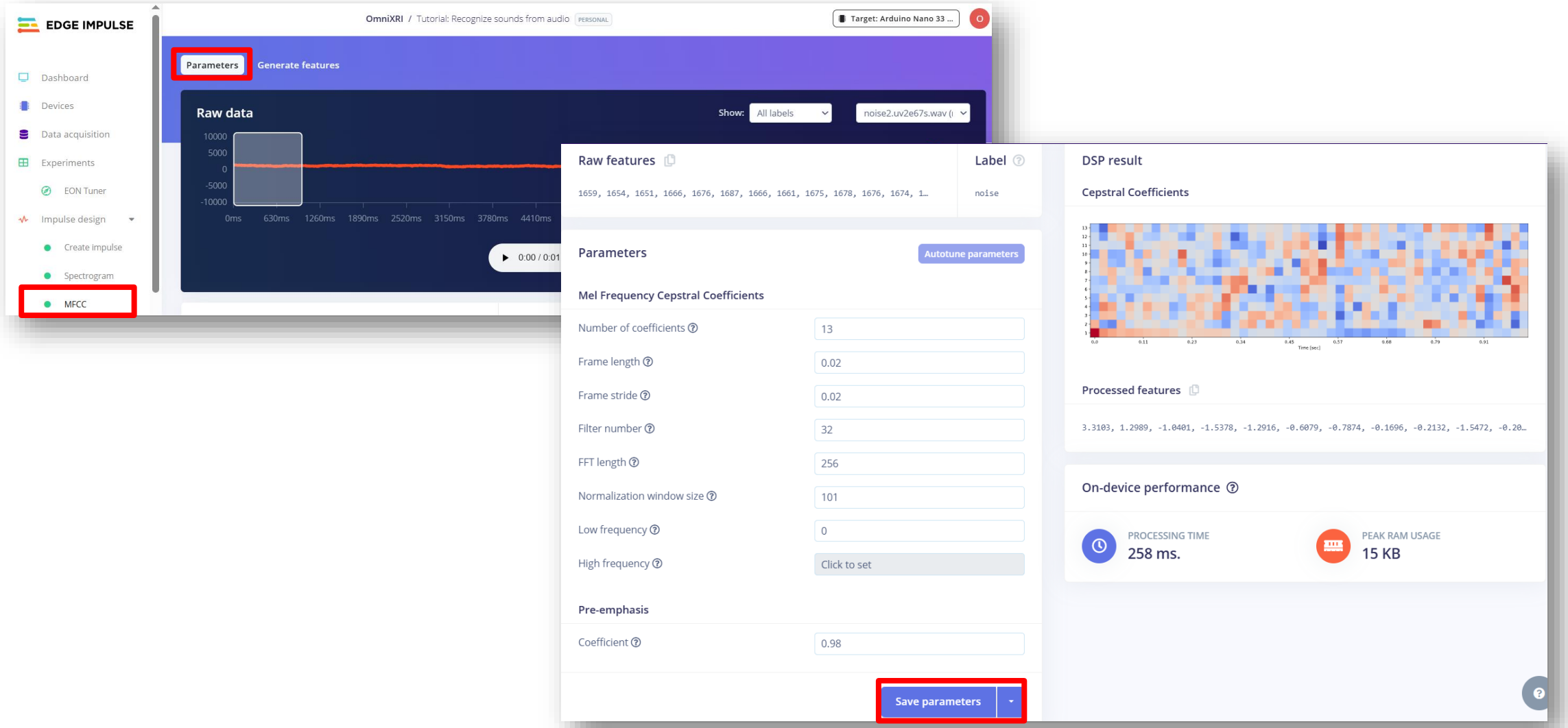
提取資料特徵 - Spectrogram



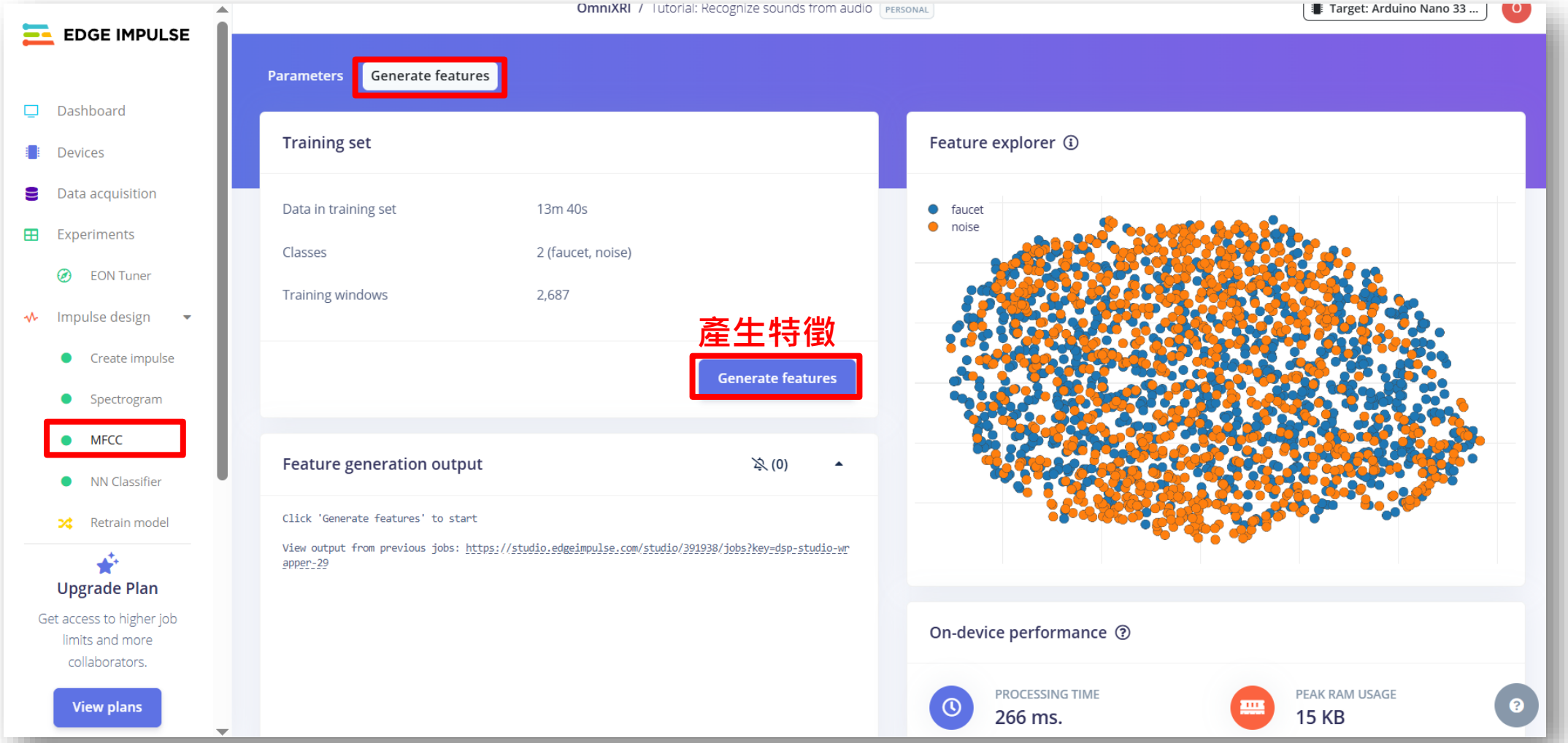
產生特徵結果 - Spectrogram



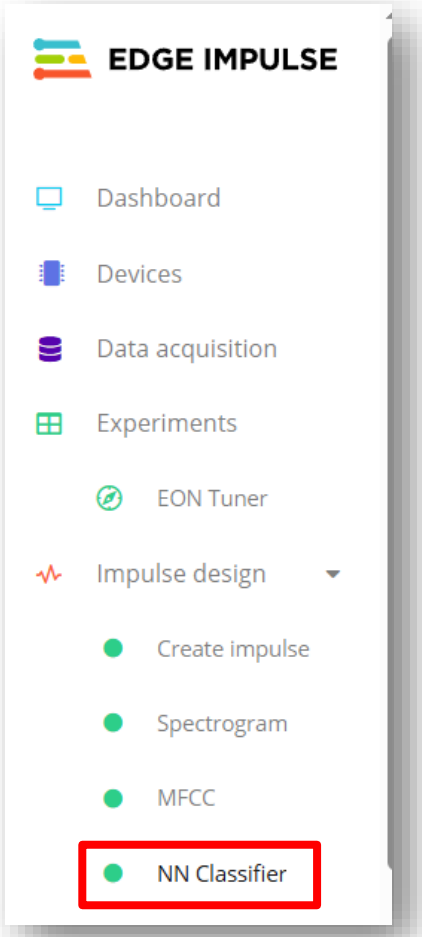
提取資料特徵 - MFCC


EDGE IMPULSE
OmniXRI / Tutorial: Recognize sounds from audio
PERSONAL
Target: Arduino Nano 33 ...
Dashboard
Devices
Data acquisition
Experiments
EON Tuner
Impulse design
Create impulse
Spectrogram
MFCC
Parameters
Generate features
Raw data
Show:
All labels
noise2.uv2e67s.wav
10000
5000
0
-5000
-10000
0ms
630ms
1260ms
1890ms
2520ms
3150ms
3780ms
4410ms
0:00 / 0:01
Raw features
1659, 1654, 1651, 1666, 1676, 1687, 1666, 1661, 1675, 1678, 1676, 1674, 1…
Label
noise
Parameters
Autotune parameters
Mel Frequency Cepstral Coefficients
Number of coefficients
13
Frame length
0.02
Frame stride
0.02
Filter number
32
FFT length
256
Normalization window size
101
Low frequency
0
High frequency
Click to set
Pre-emphasis
Coefficient
0.98
Save parameters
DSP result
Cepstral Coefficients
Time [sec]
Processed features
3.3103, 1.2989, -1.0401, -1.5378, -1.2916, -0.6079, -0.7874, -0.1696, -0.2132, -1.5472, -0.20…
On-device performance
PROCESSING TIME
258 ms.
PEAK RAM USAGE
15 KB

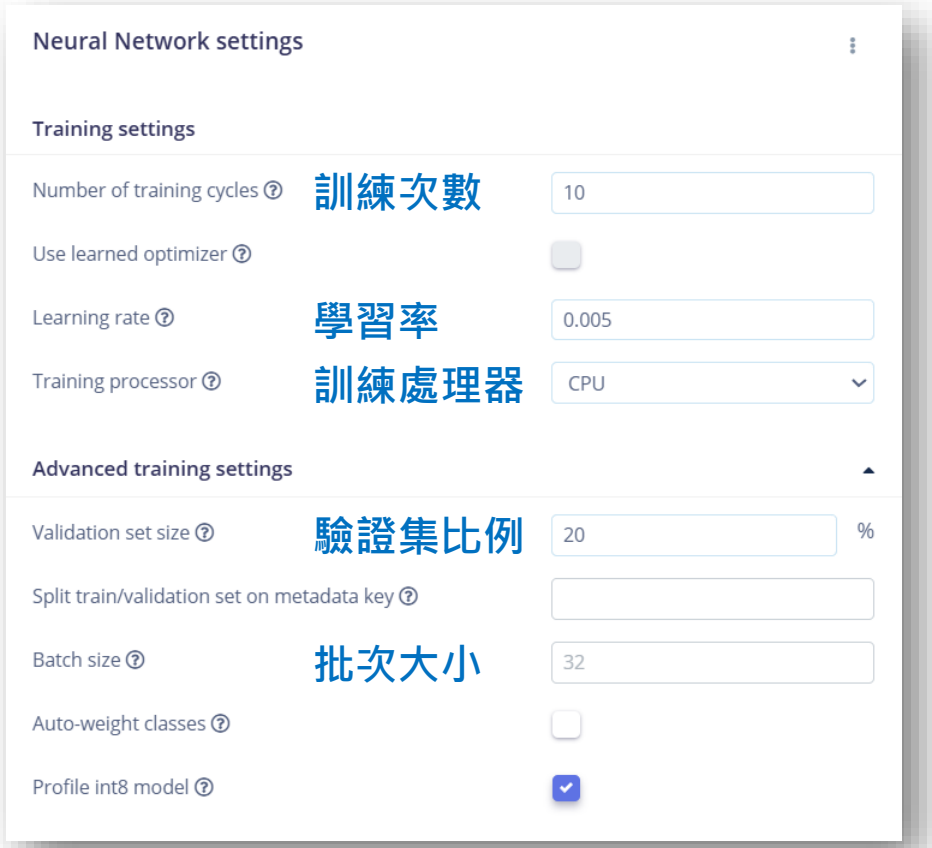
產生特徵結果 - MFCC


EDGE IMPULSE
OmniXRI / Tutorial: Recognize sounds from audio
PERSONAL
Target: Arduino Nano 33 ...
Parameters
Generate features
Dashboard
Devices
Data acquisition
Experiments
EON Tuner
Impulse design
Create impulse
Spectrogram
MFCC
NN Classifier
Retrain model
Upgrade Plan
Get access to higher job limits and more collaborators.
View plans
Training set
Data in training set
13m 40s
Classes
2 (faucet, noise)
Training windows
2,687
產生特徵
Generate features
Feature generation output
(0)
Click 'Generate features' to start
View output from previous jobs: https://studio.edgeimpulse.com/studio/391938/jobs?key=dsp-studio-wrapper-29
Feature explorer
faucet
noise
On-device performance
PROCESSING TIME
266 ms.
PEAK RAM USAGE
15 KB

設定分類訓練相關參數



神經網路參數設定



神經網路結構設定

模型訓練結果

Training output (0)

Model

Model version: Quantized (int8)

Last training performance (validation set)

ACCURACY	LOSS
99.4%	0.02

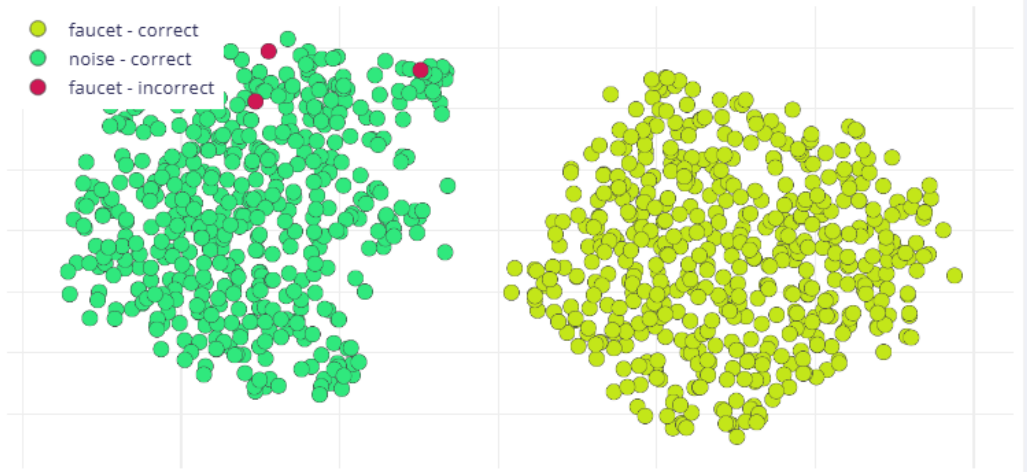
Confusion matrix (validation set)

	FAUCET	NOISE
FAUCET	99.3%	0.7%
NOISE	0.4%	99.6%
F1 SCORE	0.99	0.99

Metrics (validation set)

METRIC	VALUE
Area under ROC Curve	0.99
Weighted average Precision	0.99
Weighted average Recall	0.99
Weighted average F1 score	0.99

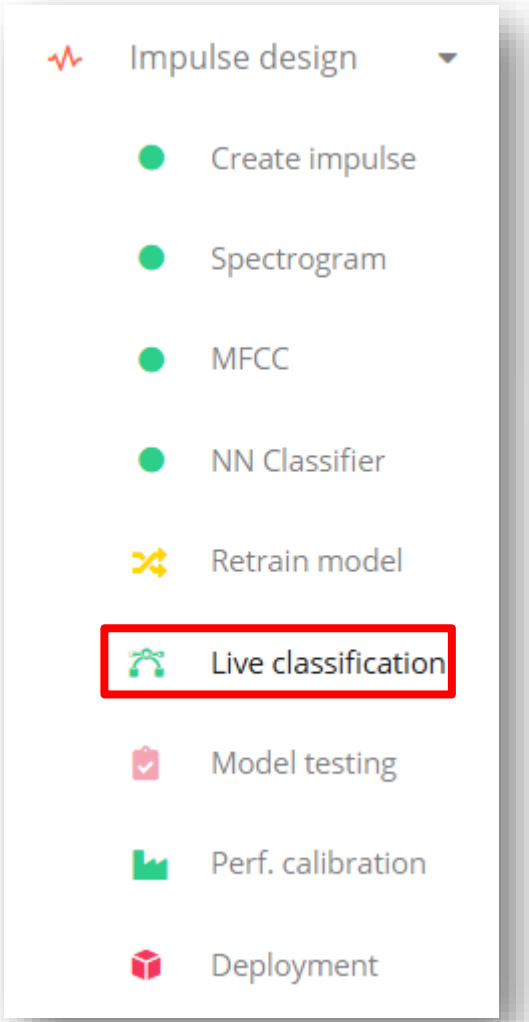
Data explorer (full training set)

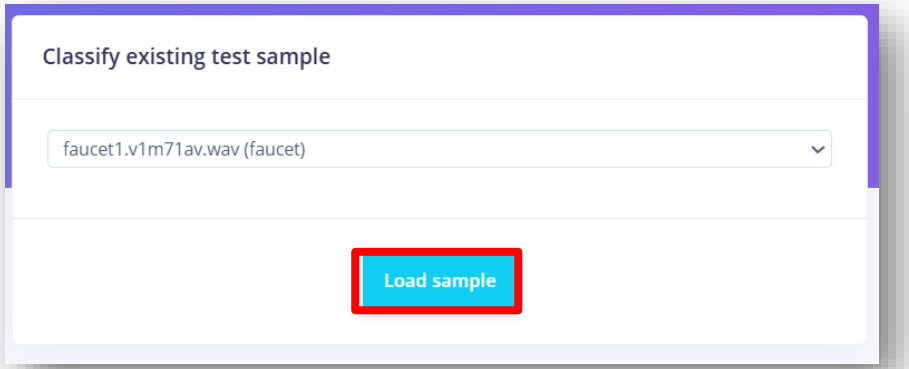


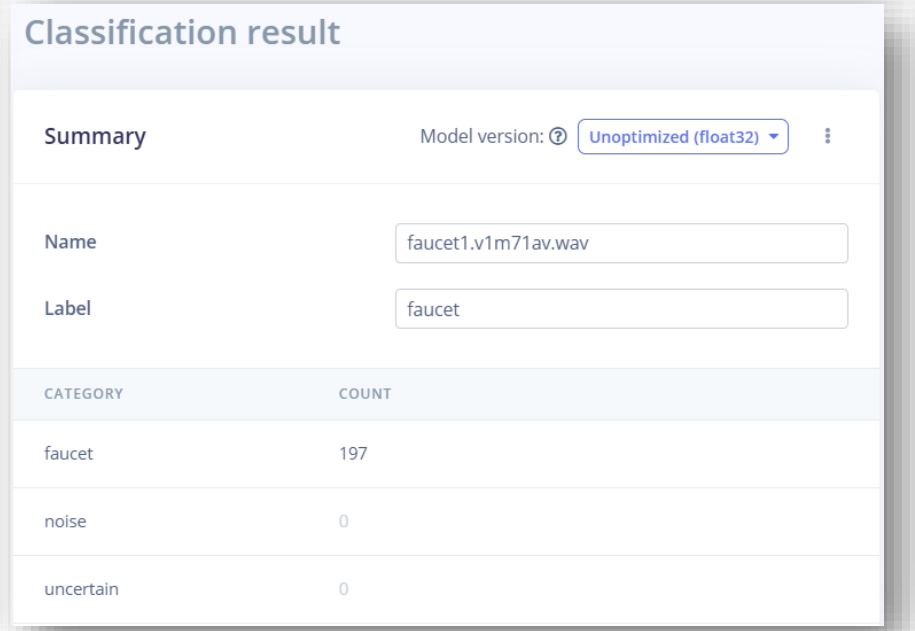
On-device performance Engine: EON™ Compiler (RAM optimized)

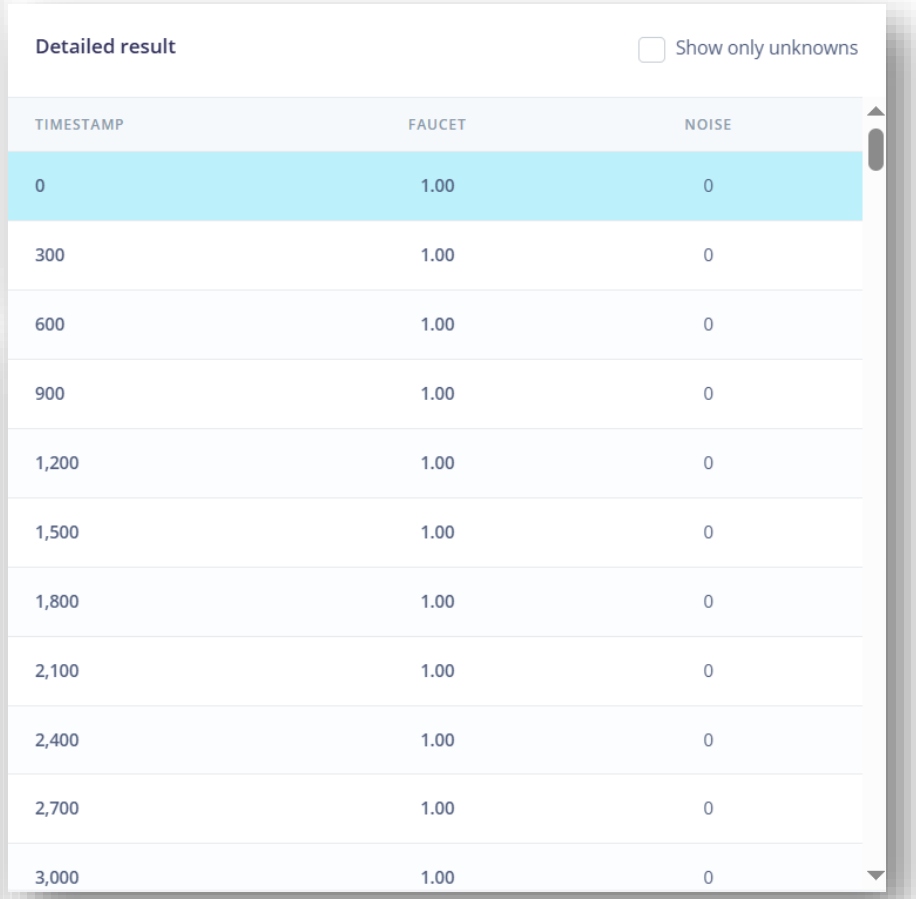
INFERENCING TIME	PEAK RAM USAGE	FLASH USAGE
16 ms.	22.0K	42.9K

線上測試（從內建測試集）


Impulse design
Create impulse
Spectrogram
MFCC
NN Classifier
Retrain model
Live classification
Model testing
Perf. calibration
Deployment

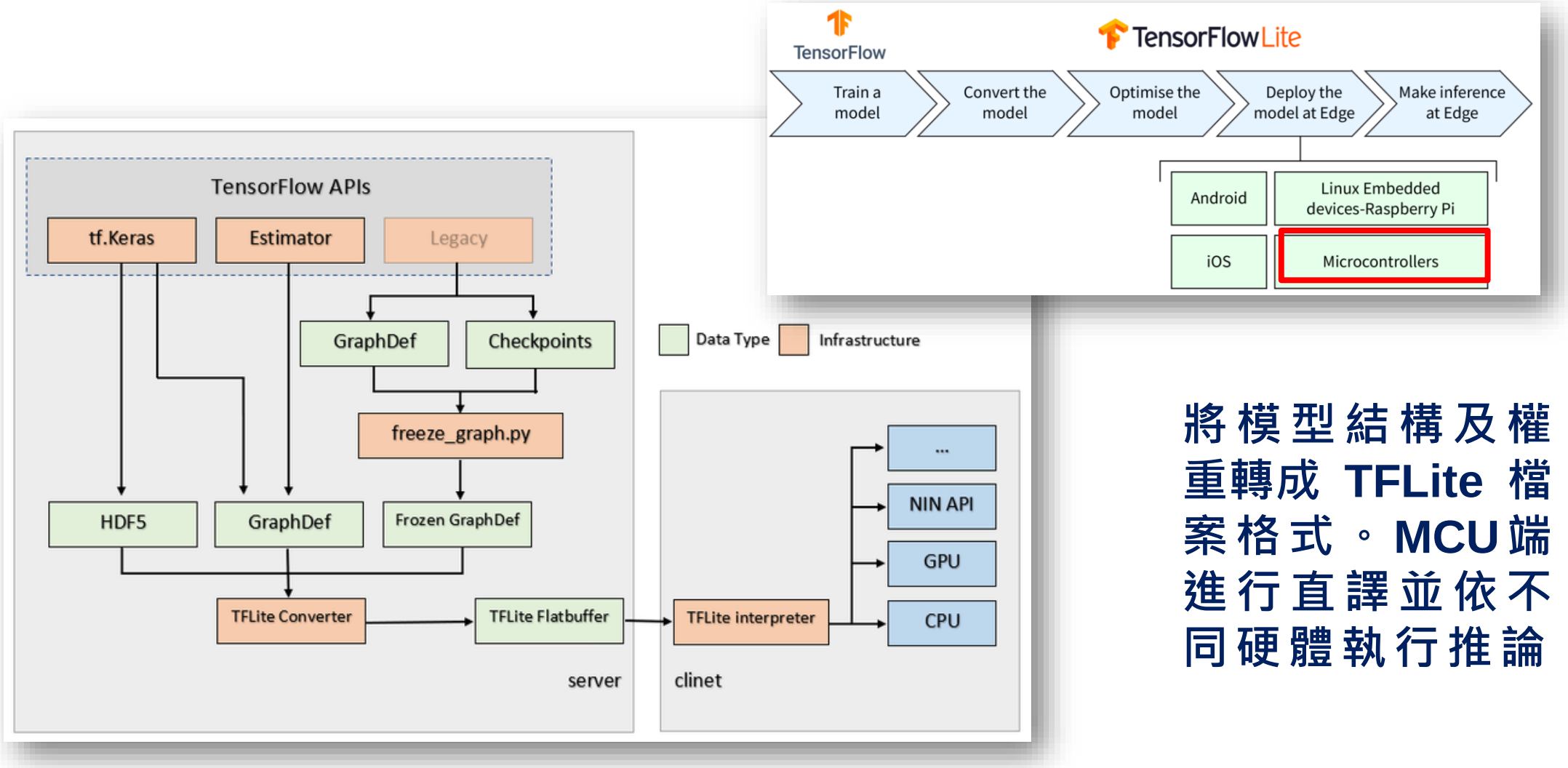

Classify existing test sample
faucet1.v1m71av.wav (faucet)
Load sample


Classification result
Summary
Model version:
Unoptimized (float32)
Name
faucet1.v1m71av.wav
Label
faucet
CATEGORY
COUNT
faucet
197
noise
0
uncertain
0


Detailed result
Show only unknowns
TIMESTAMP
FAUCET
NOISE
0
1.00
0
300
1.00
0
600
1.00
0
900
1.00
0
1,200
1.00
0
1,500
1.00
0
1,800
1.00
0
2,100
1.00
0
2,400
1.00
0
2,700
1.00
0
3,000
1.00
0

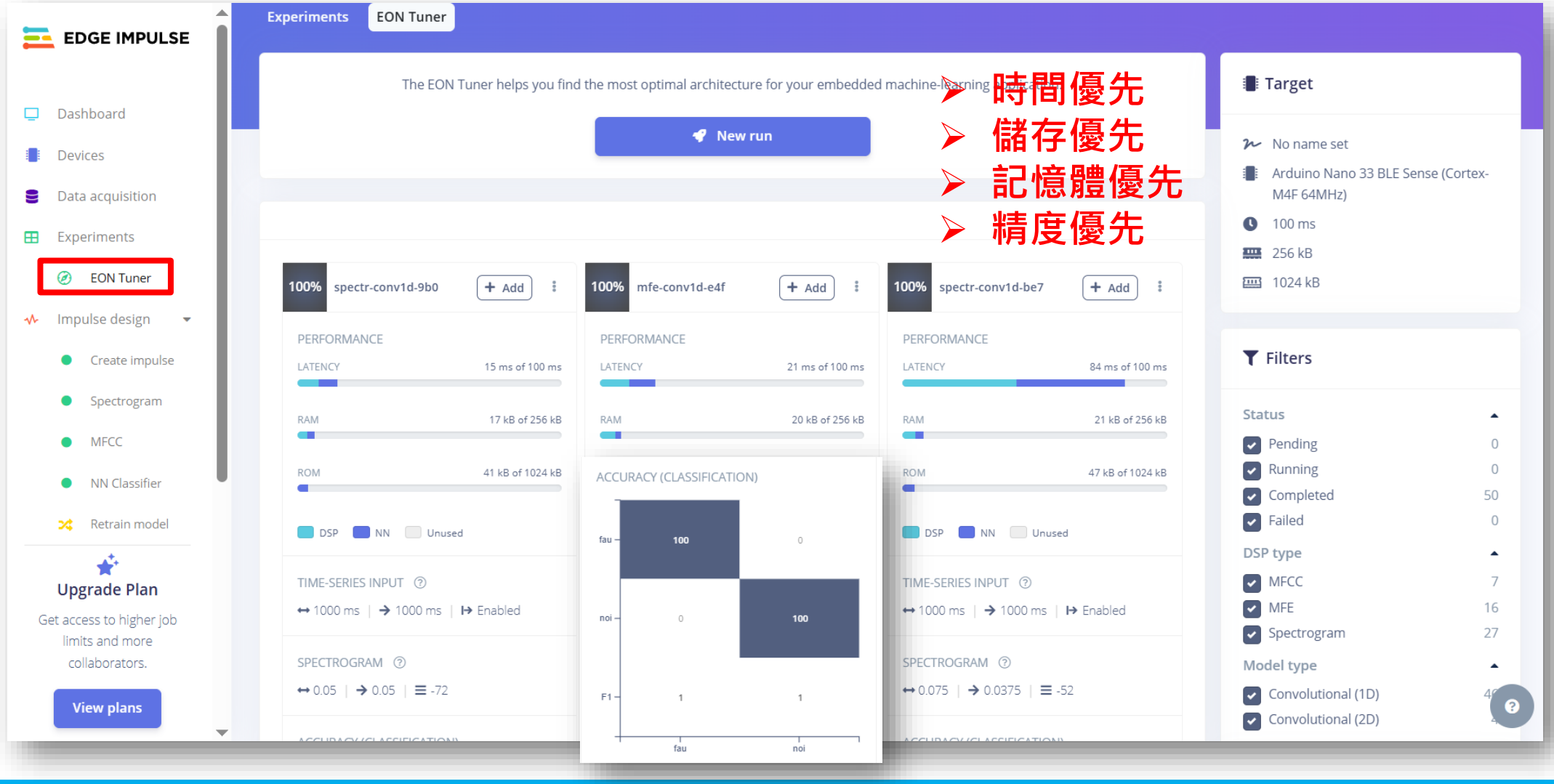
6.3. 模型優化及部署

TensorFlow Lite 轉換程序



將模型結構及權重轉成 TFLite 檔案格式。MCU端進行直譯並依不同硬體執行推論。

目標優化 – EON Tuner (AutoML)

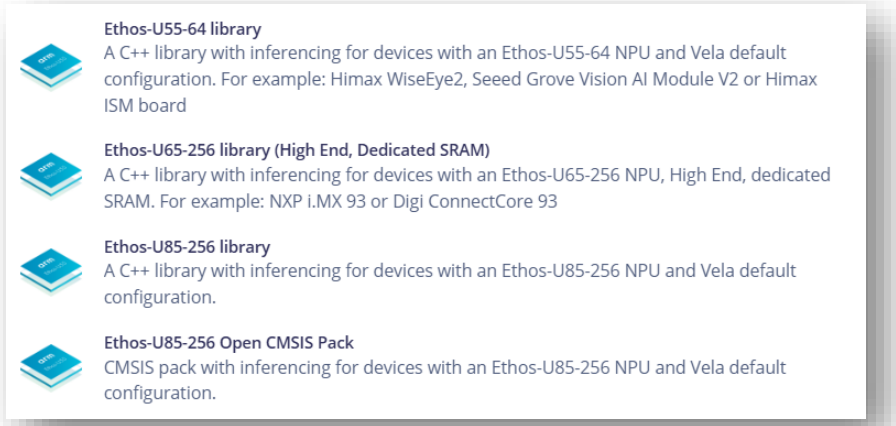


模型優化


MODEL OPTIMIZATIONS
Model optimizations can increase on-device performance but may reduce accuracy.
TensorFlow Lite
TensorFlow Lite
Quantized (int8)
Selected
整數8位元
SPECTROGRAM MFCC NN CLASSIFIER TOTAL
LATENCY 146 ms. 266 ms. 16 ms. 428 ms.
RAM 27.8K 15.4K 21.5K 27.8K
FLASH - - 53.5K -
ACCURACY -
Unoptimized (float32)
Select
浮點32位元
SPECTROGRAM MFCC NN CLASSIFIER TOTAL
LATENCY 146 ms. 266 ms. 467 ms. 879 ms.
RAM 27.8K 15.4K 71.0K 71.0K
FLASH - - 60.7K -
ACCURACY -
To compare model accuracy, run model testing for all available optimizations.
Run model testing
Estimate for Arduino Nano 33 BLE Sense (Cortex-M4F 64MHz) - Change target
開始建構
Build


MODEL OPTIMIZATIONS
Model optimizations can increase on-device performance but may reduce accuracy.
EON™ Compiler
EON Compiler
Same accuracy, 21% less RAM, 44% less ROM.
Quantized (int8)
Selected
SPECTROGRAM MFCC NN CLASSIFIER TOTAL
LATENCY 146 ms. 266 ms. 16 ms. 428 ms.
RAM 27.8K 15.4K 17.0K 27.8K
FLASH - - 33.1K -
ACCURACY -
Unoptimized (float32)
Select
SPECTROGRAM MFCC NN CLASSIFIER TOTAL
LATENCY 146 ms. 266 ms. 467 ms. 879 ms.
RAM 27.8K 15.4K 57.3K 57.3K
FLASH - - 33.8K -
ACCURACY -
To compare model accuracy, run model testing for all available optimizations.
Run model testing
Estimate for Arduino Nano 33 BLE Sense (Cortex-M4F 64MHz) - Change target
開始建構
Build

可部署方式



1. **客製化函式庫**

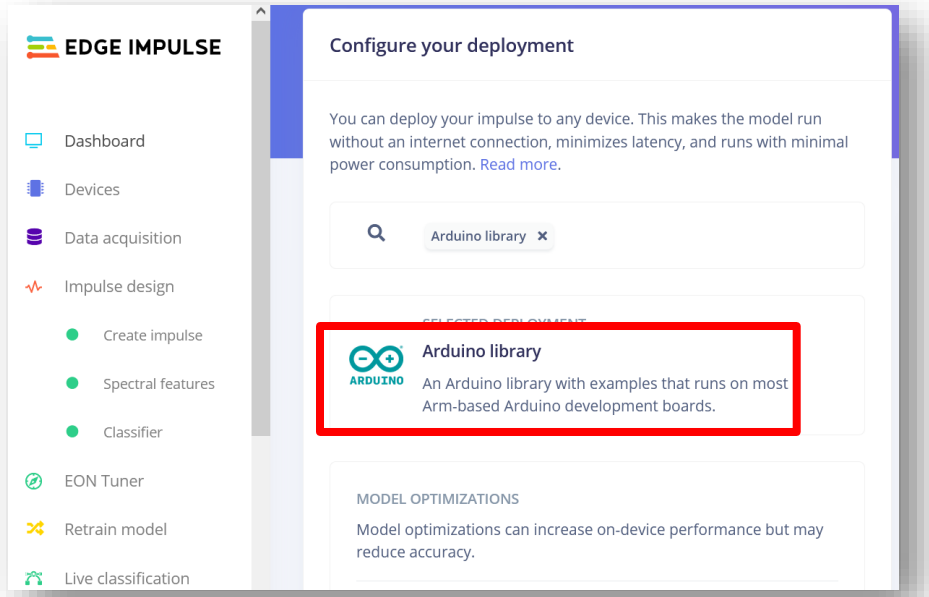
   C++, **Arduino**, WebAssembly, Cube.MX CMSIS-Pack, DAR-AI, OpenMV, **Ethos-U (NPU)**, Simplicity Studio, TensorRT, TIDL, Tensai Flow

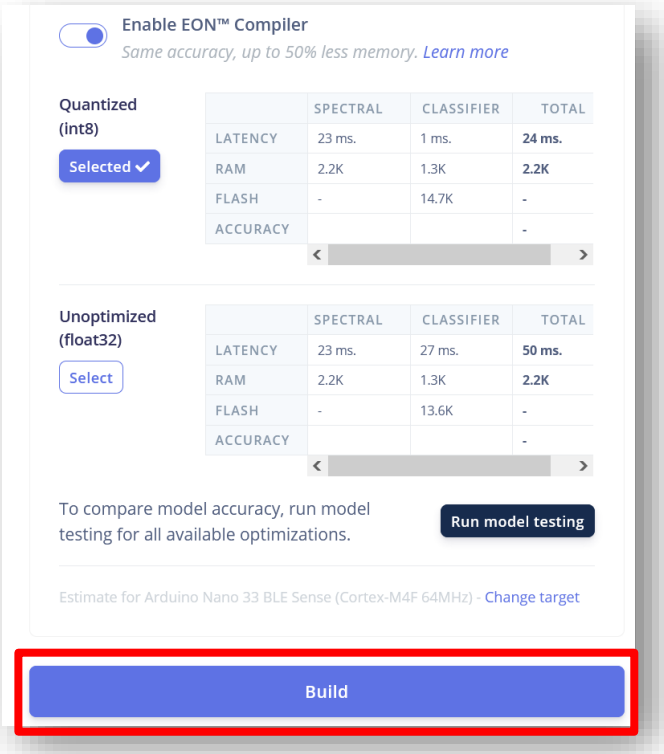
2. 預編譯韌體

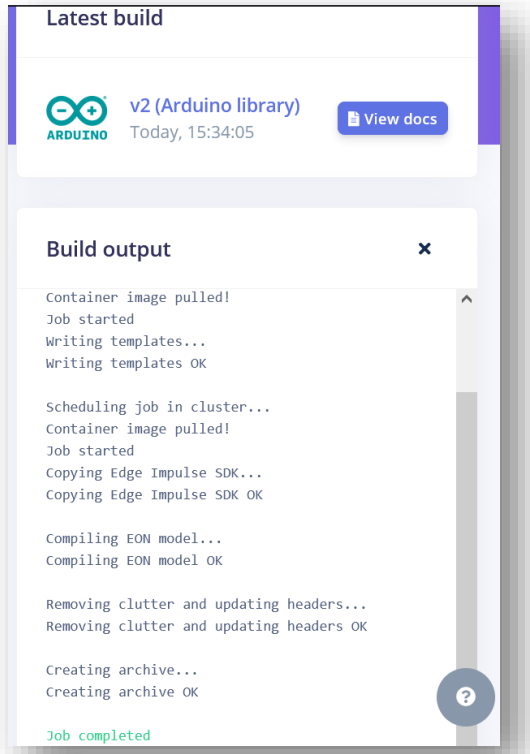
   搭配 edge-impulse-run-impulse 使用

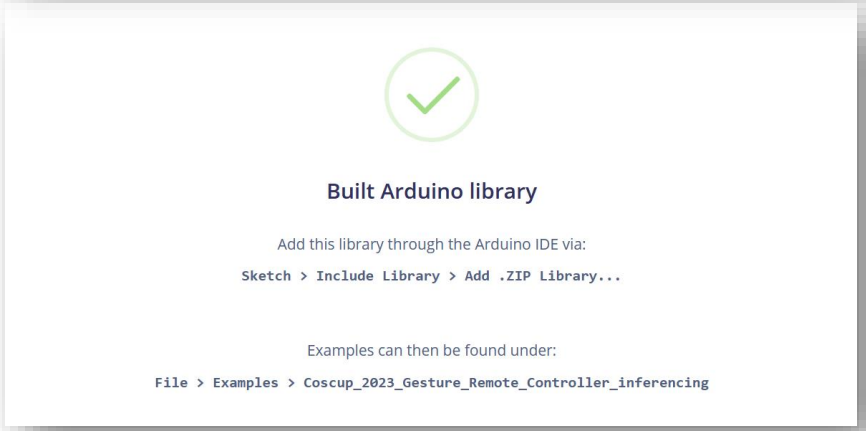
3. Linux 型式
4. Docker 型式
5. 在手機或電腦執行
6. 客製化部署方塊（企業版）

以 Arduino 型式輸出為例





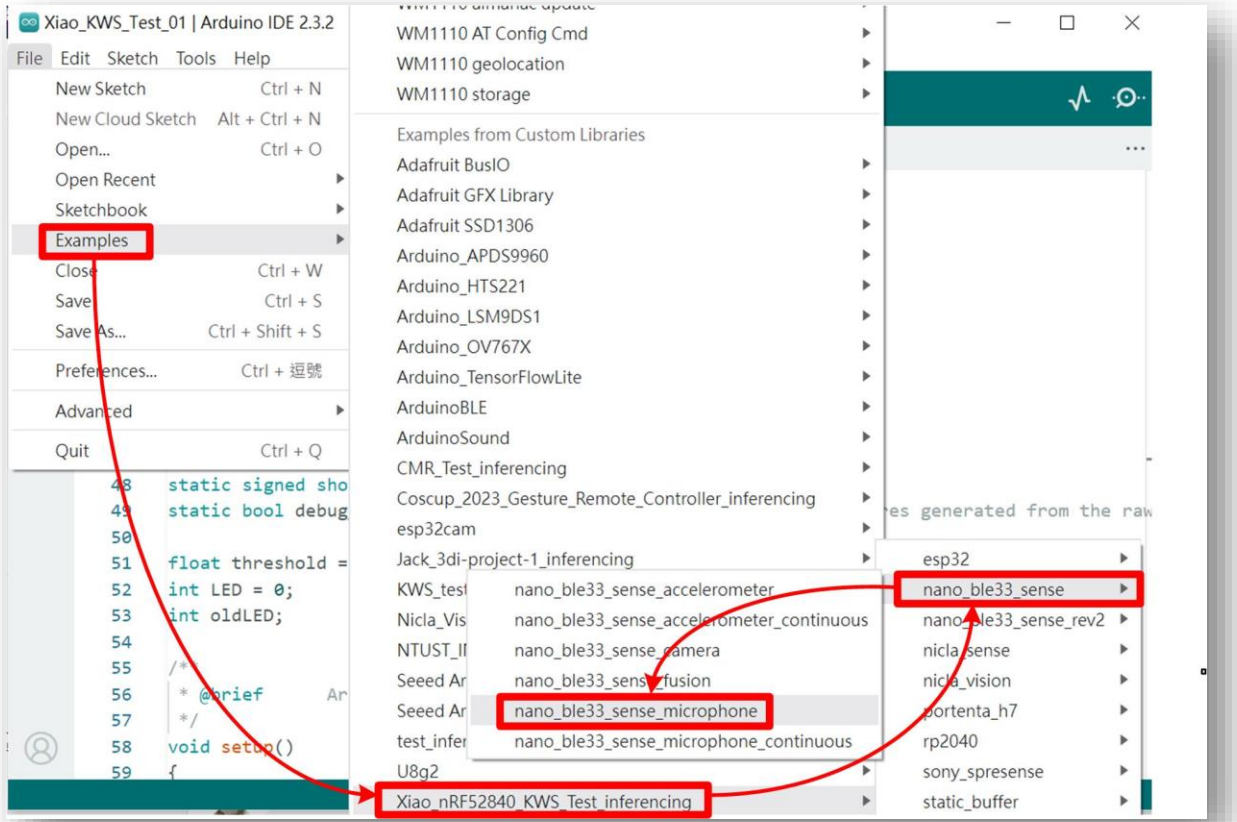




以 Arduino 函式庫(*.zip)方式輸出
預設檔名為「ei-專案名稱-arduino-版本序號.zip」

導入 Arduino 函式庫並進行推論測試

新增函式庫 Sketch > Include Library > Add .ZIP Library …
新增範例 File > Examples > ei-專案名稱_inferencing > nano_ble33_sense > nano_ble33_sense_microphone （原始範例輸出結果從 COM 輸出文字串）



新增程式碼 (Low Code)
可調整觸發門檻
可指定辨完成亮滅LED

```
51  float threshold = 0.50;
52  int LED = 0;
53  int oldLED;
```

```
117         //lets light up some LEDS
118         if (result.classification[ix].value > threshold) {
119           //now let's see what label were in
120           switch (ix) {
121             case 0: LED = 11; break; // RED LED
122             case 1: LED = 12; break; // BLUE LED
123             case 2: LED = 13; break; // GREEN LED
124             default: LED = 0;
125           }
126           //in Sense, LOW will light up the LED
127           if (LED != 0) {
128             digitalWrite (oldLED, HIGH); //if we enter a wo
129             digitalWrite (LED, LOW);
130             oldLED = LED;
131           }
132           else //turn off LED
133             digitalWrite (oldLED, HIGH);
134         }
```

參考文獻

- 許哲豪，臺灣科技大學資訊工程系「人工智慧與邊緣運算實務」（2021~2023）

https://omnixri.blogspot.com/p/ntust-edge-ai.html

- 許哲豪，OmniXRI's Edge AI & TinyML 小學堂 Youtube 直播課程總結

https://omnixri.blogspot.com/2024/06/omnixris-edge-ai-tinyml-youtube.html

- 許哲豪，歐尼克斯實境互動工作室系列發文─TinyML(MCU AI)系列

https://hackmd.io/1PK1URhIQ7GutcWgpgsWbg#TinyMLMCU-AI%E7%B3%BB%E5%88%97

- Edge Impulse Studio Document

https://docs.edgeimpulse.com/docs

- 許哲豪，【課程簡報】20240509_慈濟醫資_穿戴式人工智慧工作坊_利用TinyML技術快速搭建微型智慧應用

https://omnixri.blogspot.com/2024/05/20240509tinyml12.html

歐尼克斯實境互動工作室
(OmniXRI Studio)
許哲豪 (Jack Hsu)

Facebook : Jack Omnixri
FB社團： Edge AI Taiwan邊緣智能交流區
電子信箱：omnixri@gmail.com
部落格：https://omnixri.blogspot.tw
開　源：https://github.com/OmniXRI

YOUTUBE 直播： https://www.youtube.com/@omnixri1784/streams