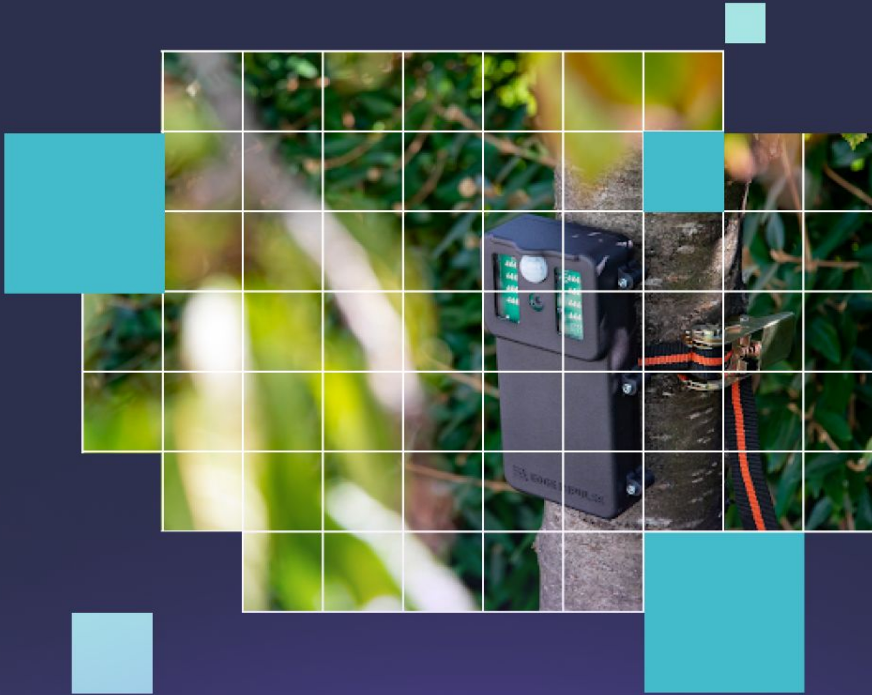
EDGE IMPULSE

The Future of Embedded ML

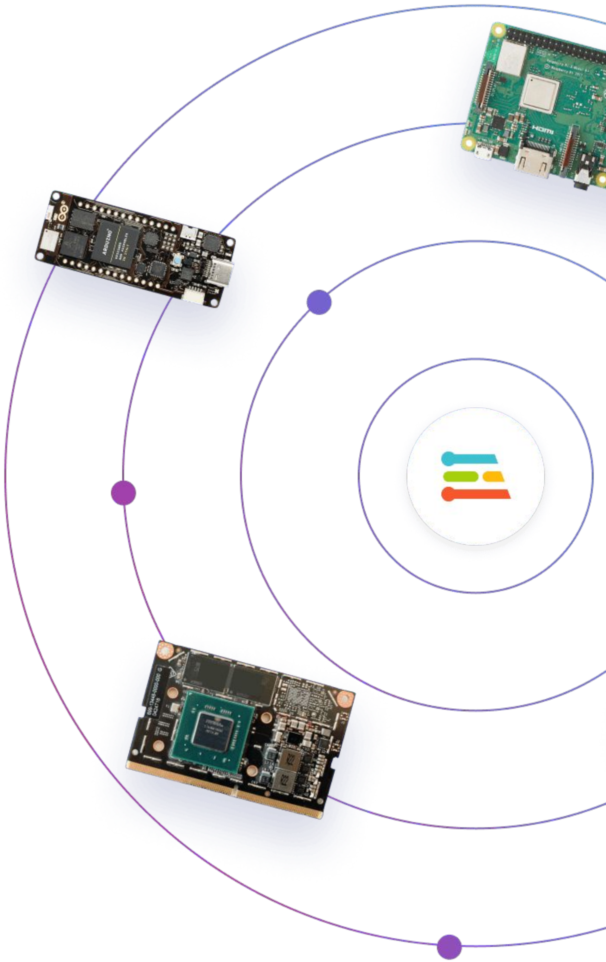
Alessandro Grande
Head of Product

ICTP, Trieste - July 3, 2023

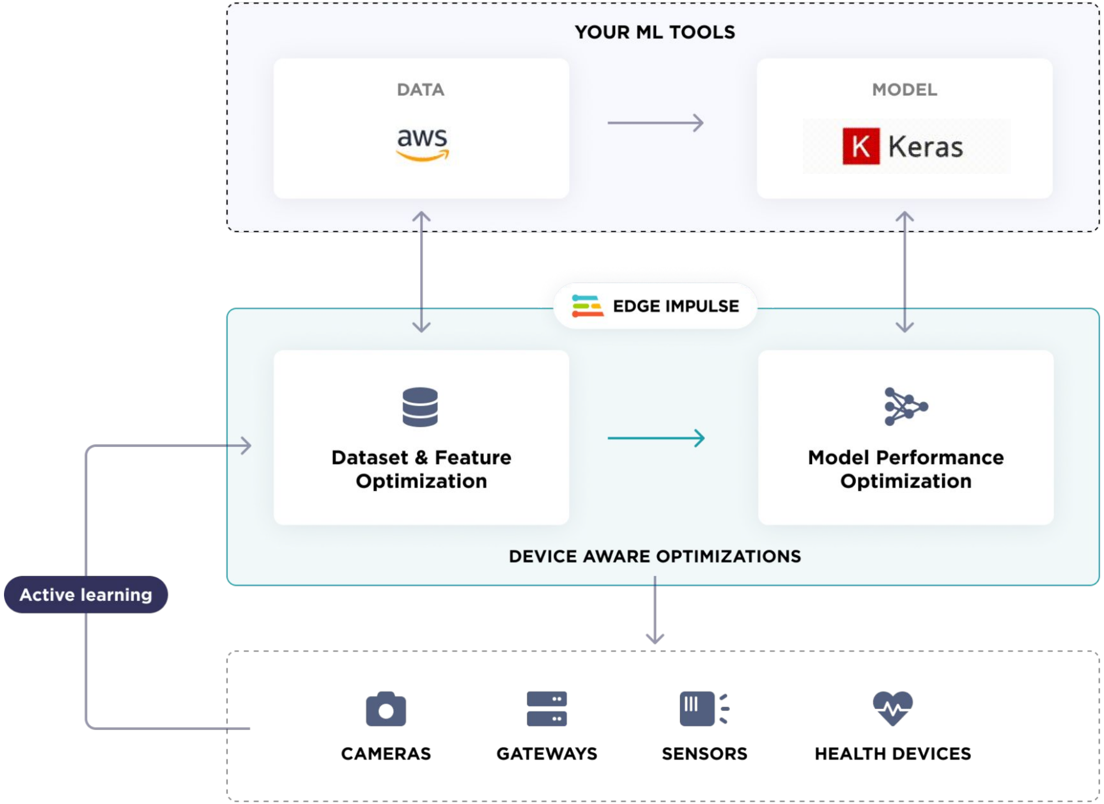


Agenda

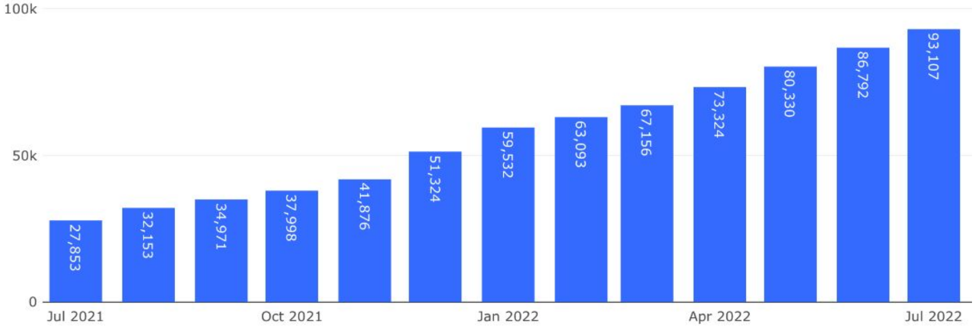
1. Intro to Edge Impulse
2. Customer challenges
3. What’s beneath the surface
4. Resources
5. Next steps



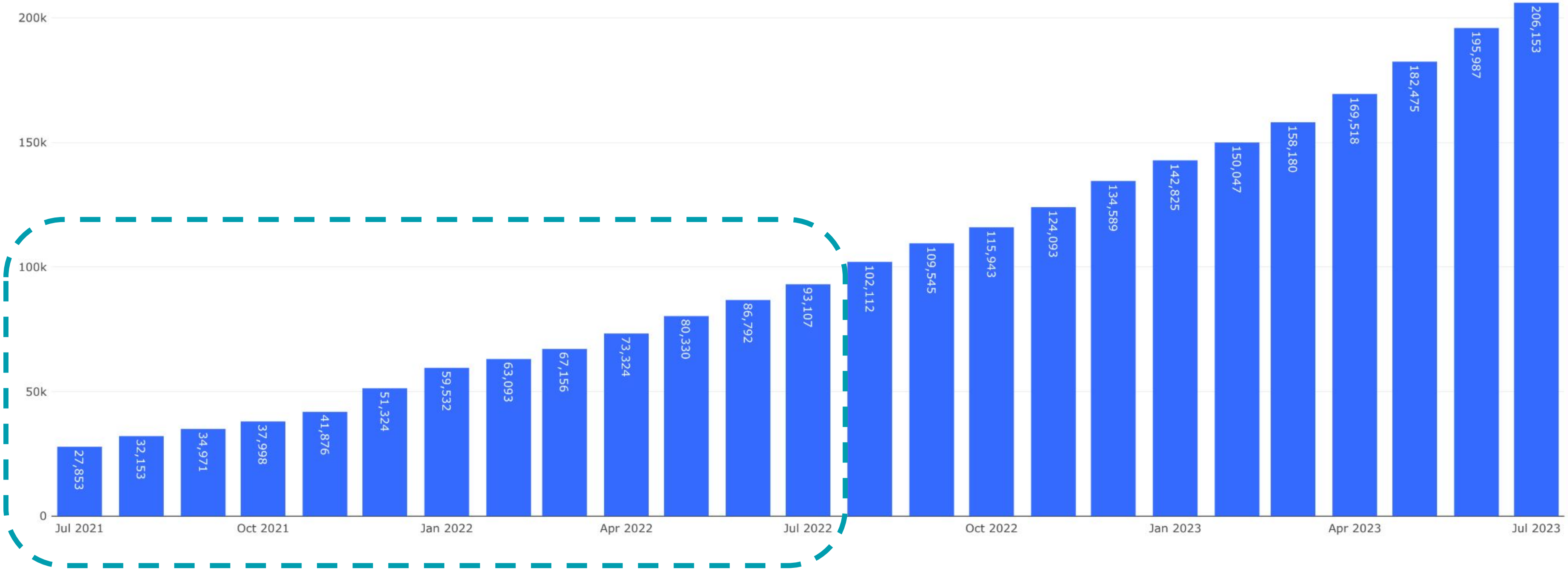
The edge AI platform



Number of Projects on Edge Impulse



Number of Projects on Edge Impulse



TinyML Use Cases

Health

Industrial

Wearables

Infrastructure

Buildings

ŌURA

KNOW LABS

NOWATCH

Hyfe AI

Brambles

NASA

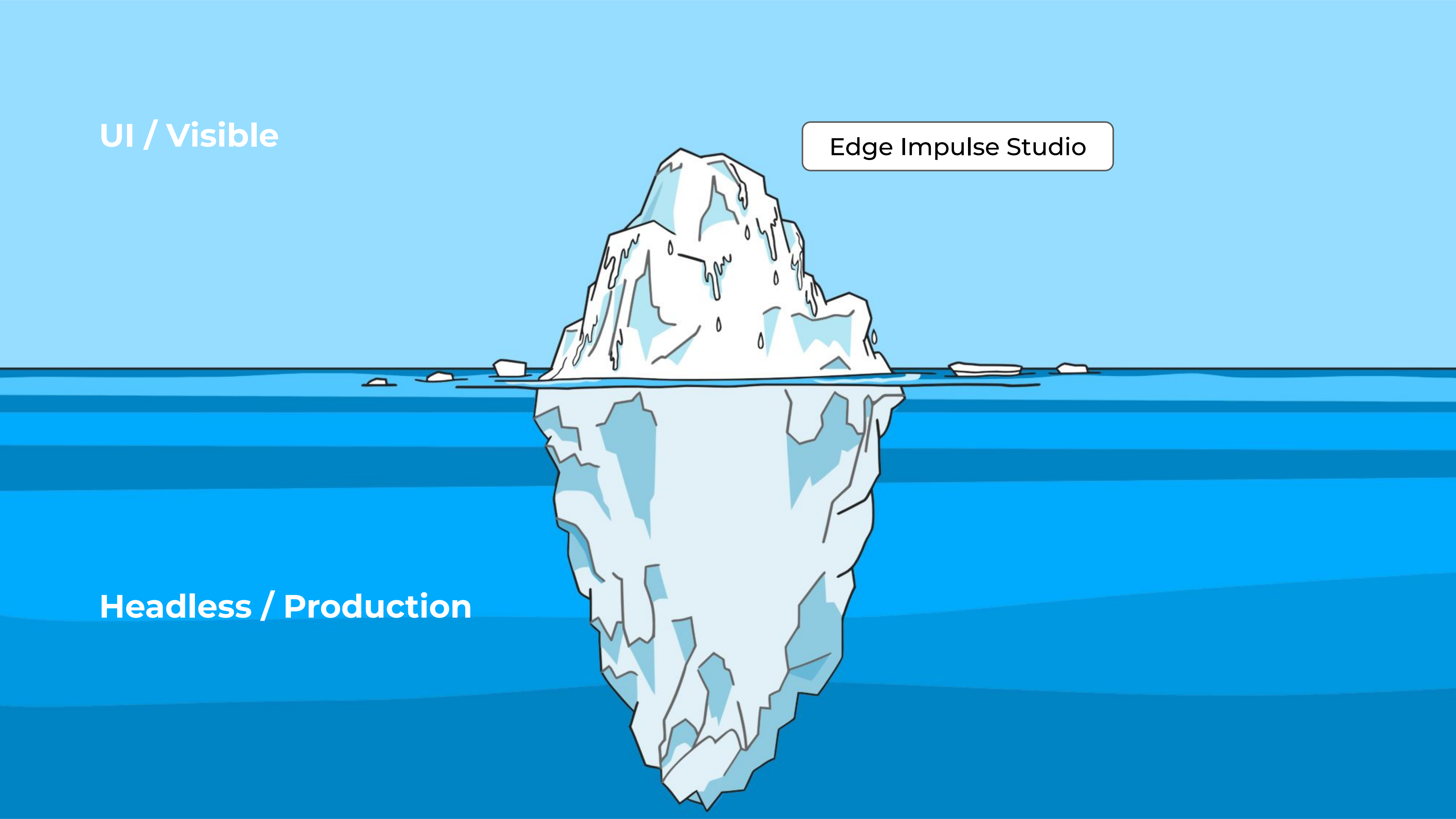
Lexmark™

ECOLAB®

IZOELEKTRO

Production Challenges

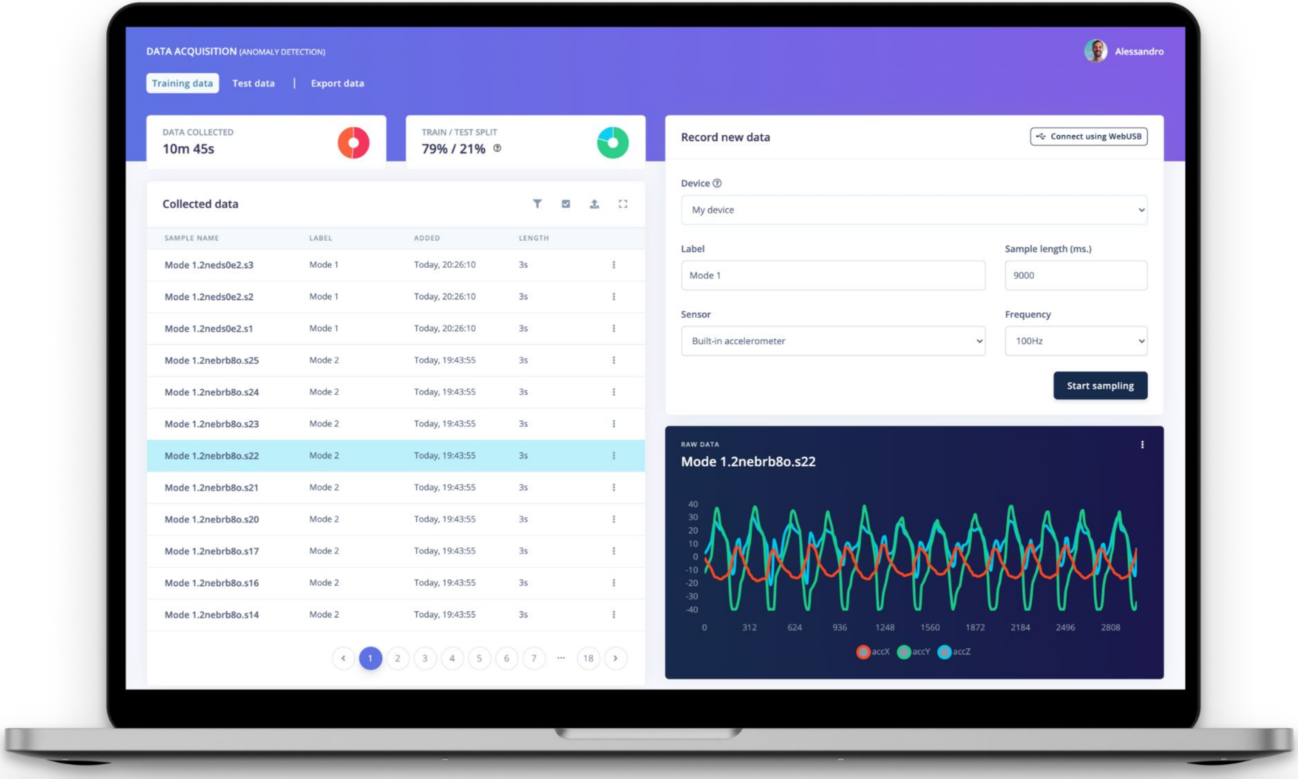
1. Data collection
2. Data quality analysis
3. Feature extraction and DSP
4. Deployment
5. Monitoring performance



UI / Visible

Edge Impulse Studio

Headless / Production



DATA ACQUISITION (ANOMALY DETECTION)

DATA COLLECTED
10m 45s

TRAIN / TEST SPLIT
79% / 21%

Collected data

SAMPLE NAME	LABEL	ADDED	LENGTH
Mode 1.2neds0e2.s3	Mode 1	Today, 20:26:10	3s
Mode 1.2neds0e2.s2	Mode 1	Today, 20:26:10	3s
Mode 1.2neds0e2.s1	Mode 1	Today, 20:26:10	3s
Mode 1.2nebrb8o.s25	Mode 2	Today, 19:43:55	3s
Mode 1.2nebrb8o.s24	Mode 2	Today, 19:43:55	3s
Mode 1.2nebrb8o.s23	Mode 2	Today, 19:43:55	3s
Mode 1.2nebrb8o.s22	Mode 2	Today, 19:43:55	3s
Mode 1.2nebrb8o.s21	Mode 2	Today, 19:43:55	3s
Mode 1.2nebrb8o.s20	Mode 2	Today, 19:43:55	3s
Mode 1.2nebrb8o.s17	Mode 2	Today, 19:43:55	3s
Mode 1.2nebrb8o.s16	Mode 2	Today, 19:43:55	3s
Mode 1.2nebrb8o.s14	Mode 2	Today, 19:43:55	3s

‹ 1 2 3 4 5 6 7 ··· 18 ›

Record new data

Device: My device

Label: Mode 1

Sample length (ms.): 9000

Sensor: Built-in accelerometer

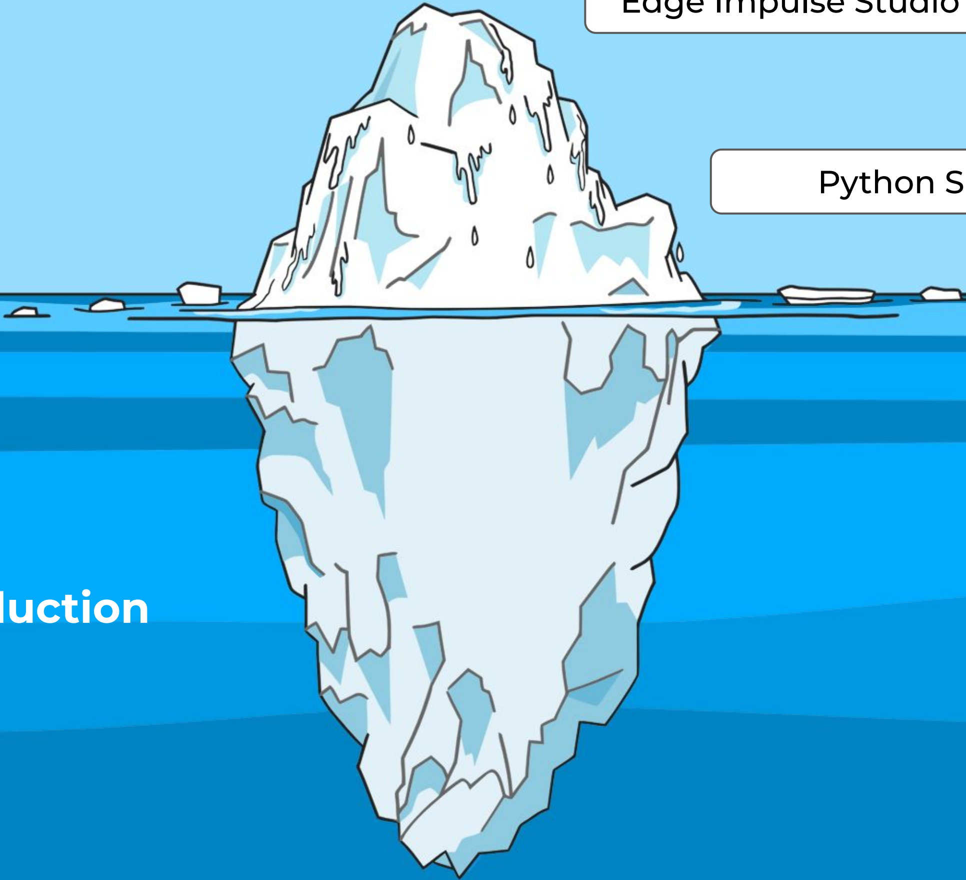
Frequency: 100Hz

RAW DATA

Mode 1.2nebrb8o.s22

studio.edgeimpulse.com/evaluate

UI / Visible



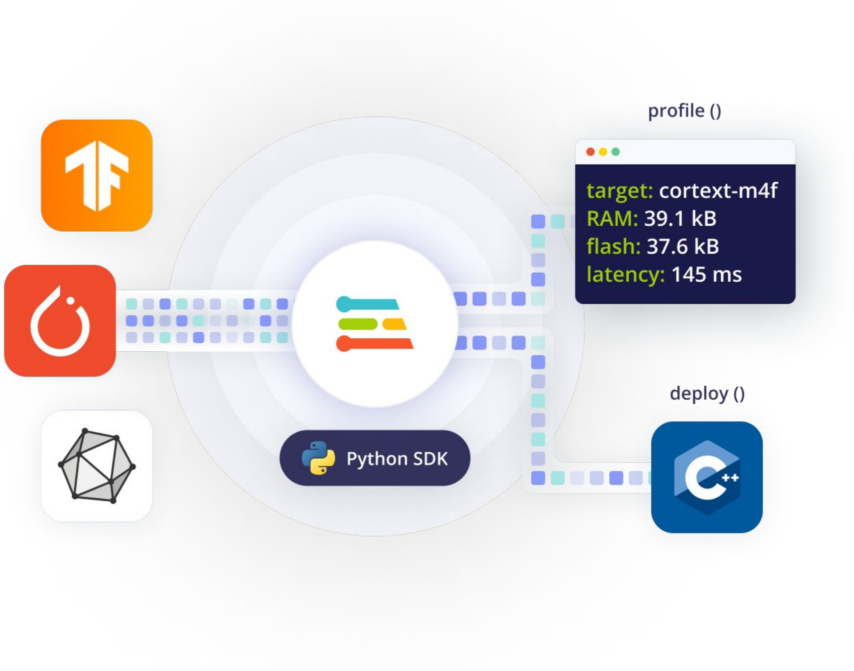
Edge Impulse Studio

Python SDK

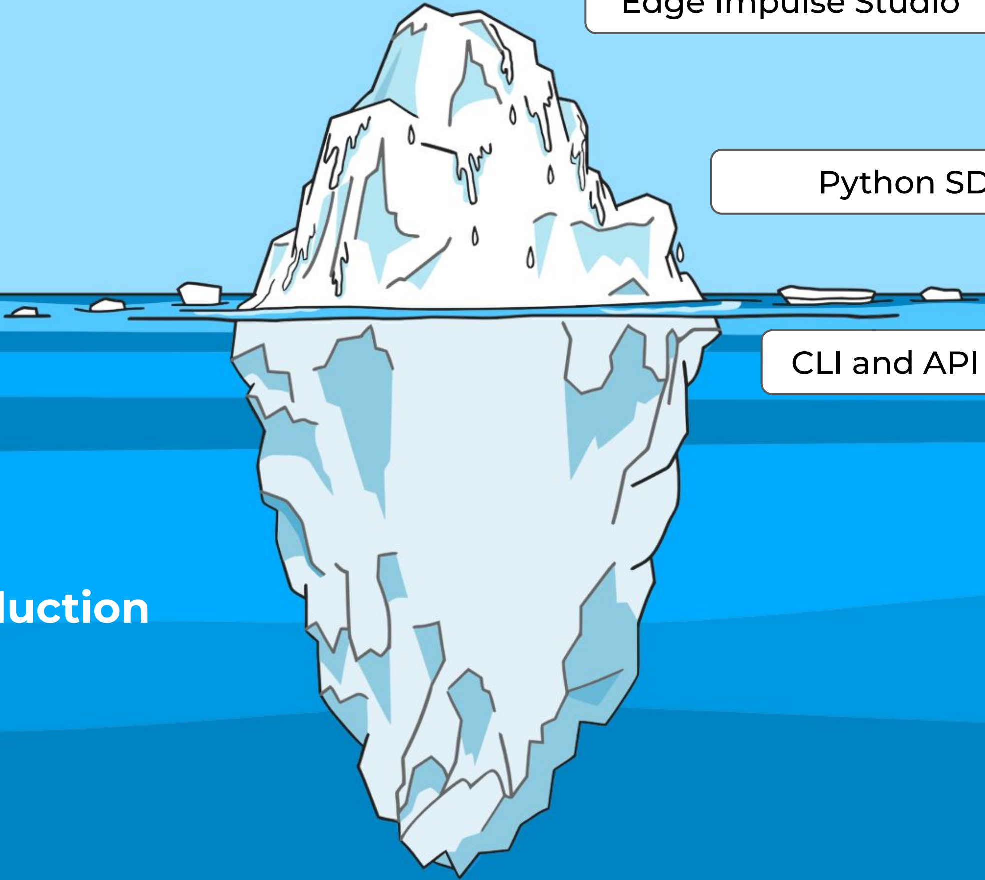
Headless / Production

BYOM & Python SDK

- Profile on-device performance of any trained model
- Analyze the impact of architectural decisions
- Generate optimized C++ libraries
- Deploy to any edge device



```
target: cortext-m4f
RAM: 39.1 kB
flash: 37.6 kB
latency: 145 ms
```



UI / Visible

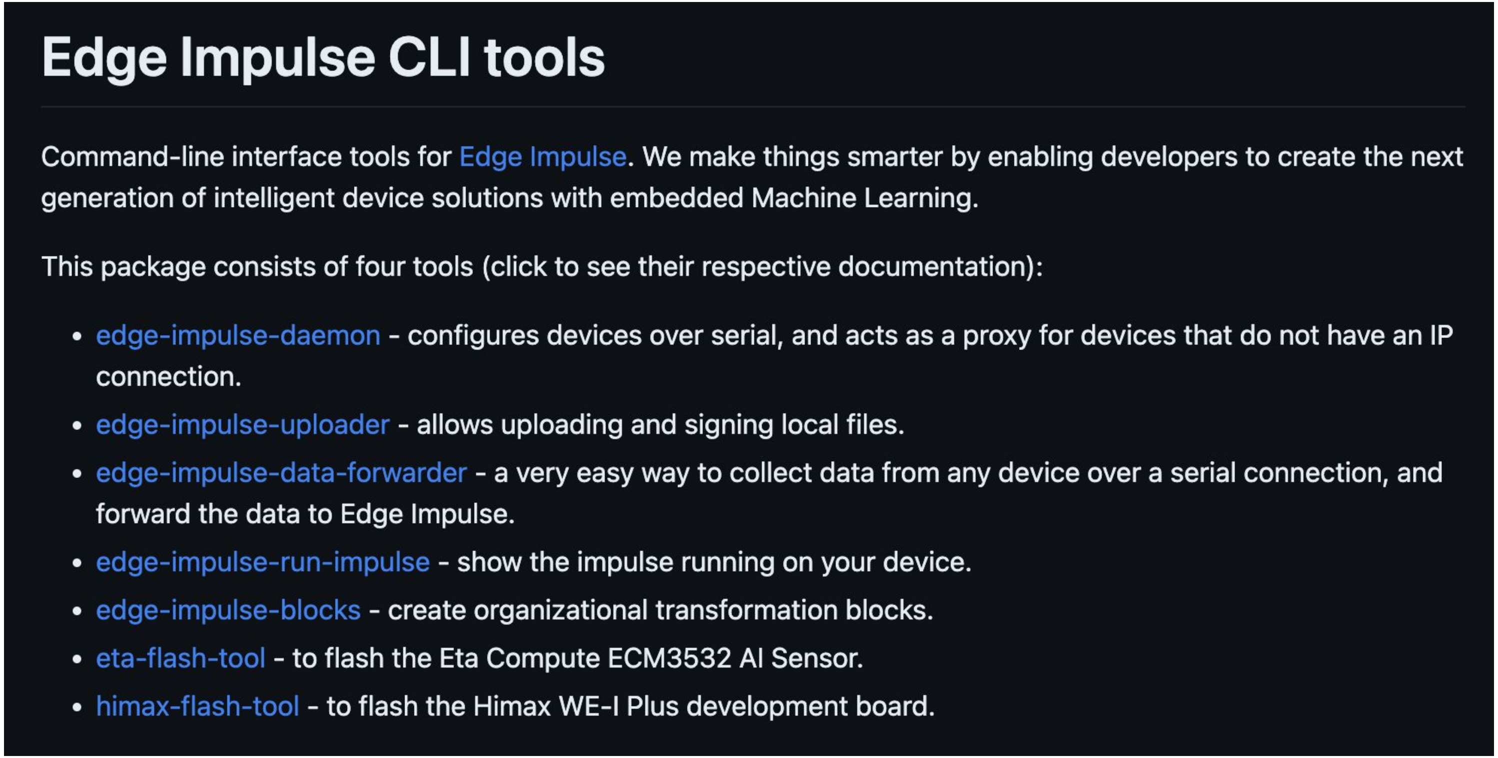
Edge Impulse Studio

Python SDK

CLI and API bindings

Headless / Production

CLI



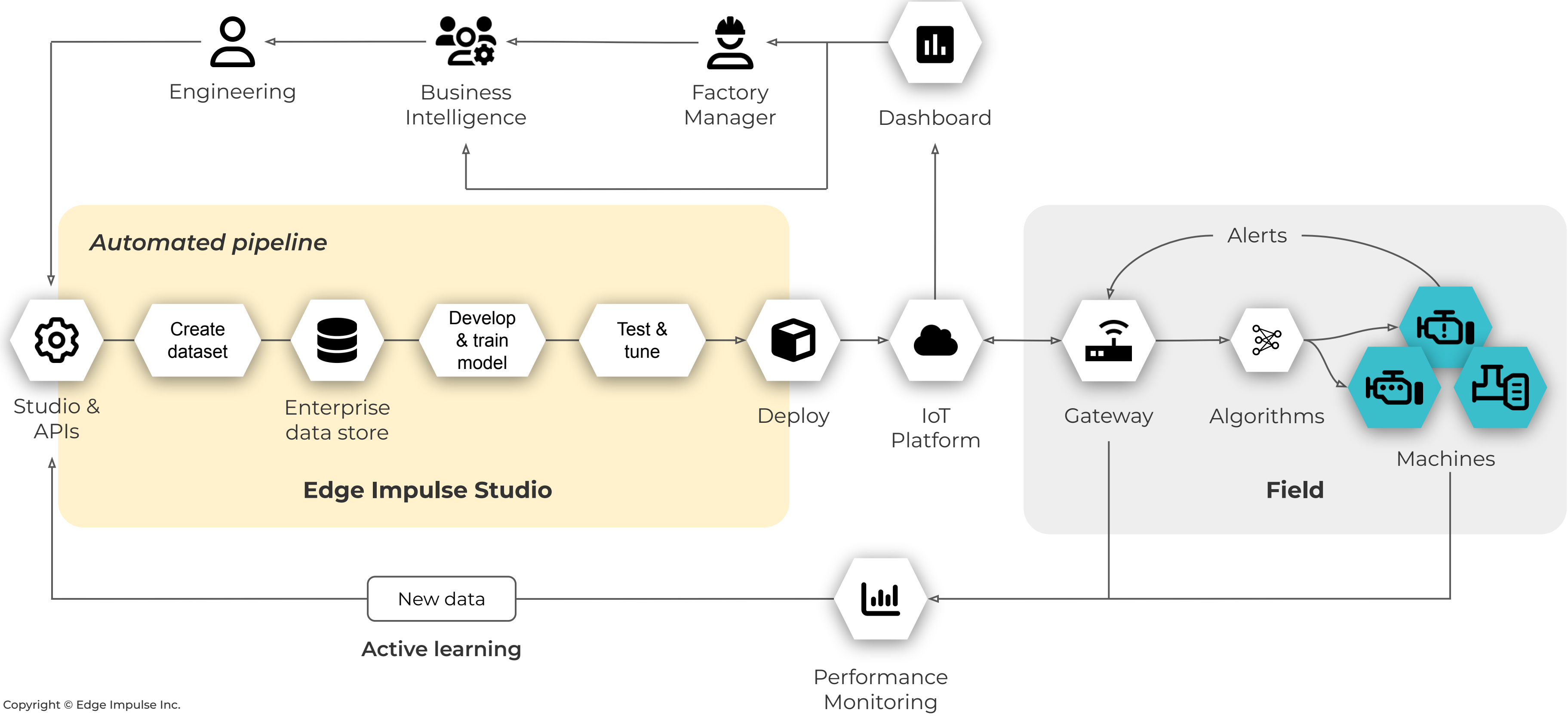
Edge Impulse CLI tools

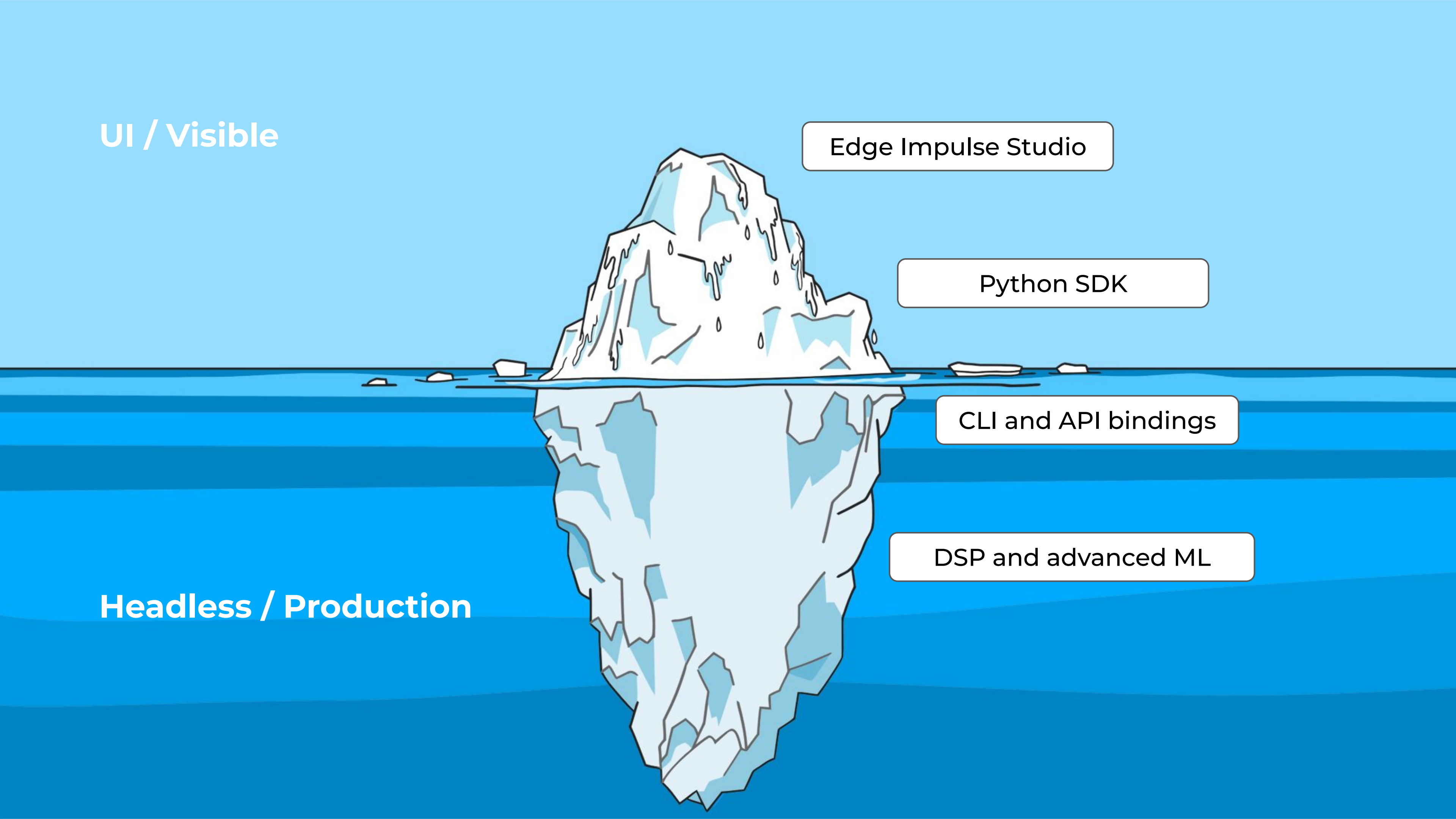
Command-line interface tools for Edge Impulse. We make things smarter by enabling developers to create the next generation of intelligent device solutions with embedded Machine Learning.

This package consists of four tools (click to see their respective documentation):

- edge-impulse-daemon - configures devices over serial, and acts as a proxy for devices that do not have an IP connection.
- edge-impulse-uploader - allows uploading and signing local files.
- edge-impulse-data-forwarder - a very easy way to collect data from any device over a serial connection, and forward the data to Edge Impulse.
- edge-impulse-run-impulse - show the impulse running on your device.
- edge-impulse-blocks - create organizational transformation blocks.
- eta-flash-tool - to flash the Eta Compute ECM3532 AI Sensor.
- himax-flash-tool - to flash the Himax WE-I Plus development board.

Embedded ML in the Real World





UI / Visible

Edge Impulse Studio

Python SDK

CLI and API bindings

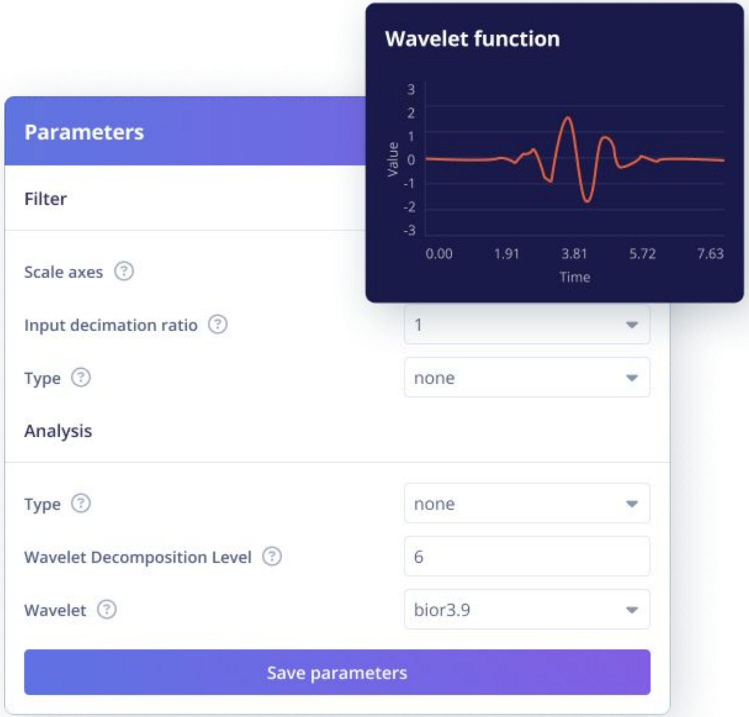
DSP and advanced ML

Headless / Production

Interactive Feature Engineering

Real-time visualization of DSP

- Immediate feedback loop enabling tactile exploration by domain expert
- Service-based architecture for real time DSP on individual samples (separate from job-based system for batched data)



Parameters

Filter

Scale axes

Input decimation ratio: 1

Type: none

Analysis

Type: none

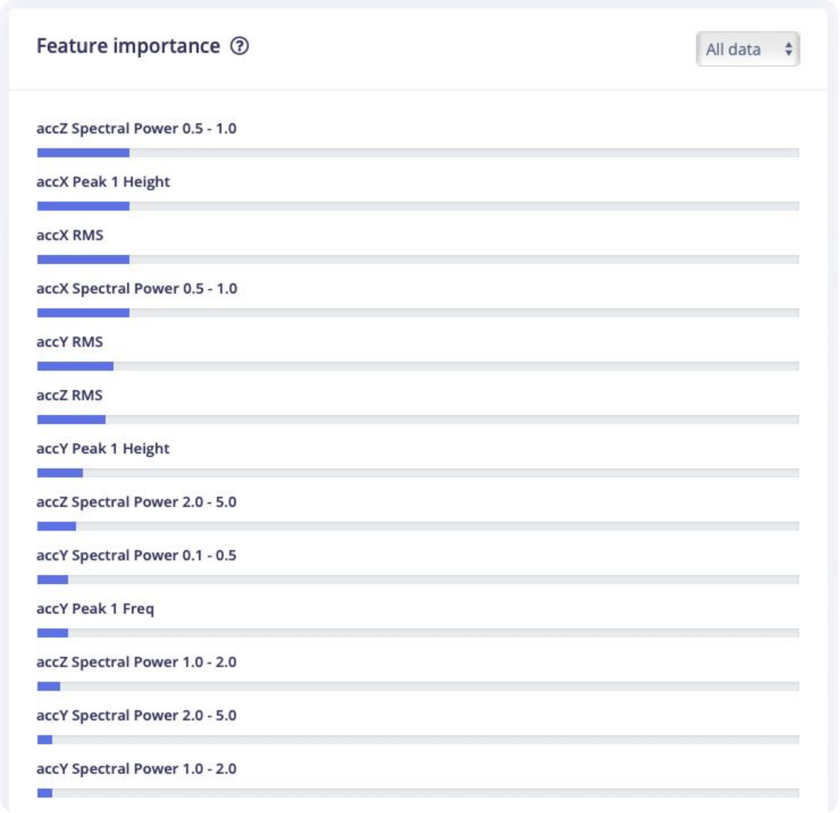
Wavelet Decomposition Level: 6

Wavelet: bior3.9

Feature Importance

Don’t use everything

- Uses recursive feature elimination with cross-validation (RFECV)
- Only computed for relatively low-dimensionality data



Feature importance

accZ Spectral Power 0.5 - 1.0

accX Peak 1 Height

accX RMS

accX Spectral Power 0.5 - 1.0

accY RMS

accZ RMS

accY Peak 1 Height

accZ Spectral Power 2.0 - 5.0

accY Spectral Power 0.1 - 0.5

accY Peak 1 Freq

accZ Spectral Power 1.0 - 2.0

accY Spectral Power 2.0 - 5.0

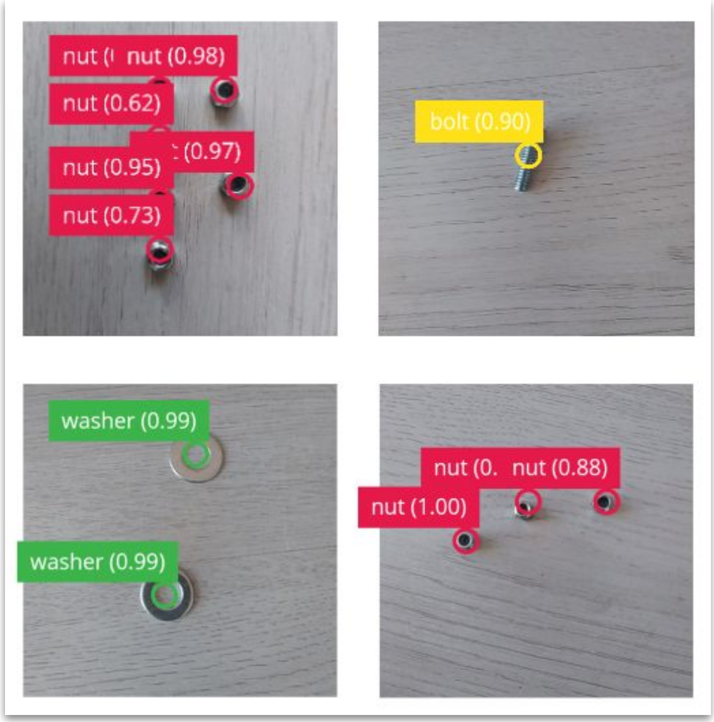
accY Spectral Power 1.0 - 2.0

MODEL HIGHLIGHT

FOMO: Faster Objects, More Objects

- **20x average performance improvement**
- Object detection on MCUs
- Ultra fast on embedded Linux
- Better at detecting smaller and more numerous objects
- Capable of segmentation and counting objects

	Cortex-M4	Cortex-M7	Cortex-A	Nvidia
FOMO	2 fps	15-30 fps	60+ fps	150+ fps
SSD	NA	NA	3 fps	20 fps

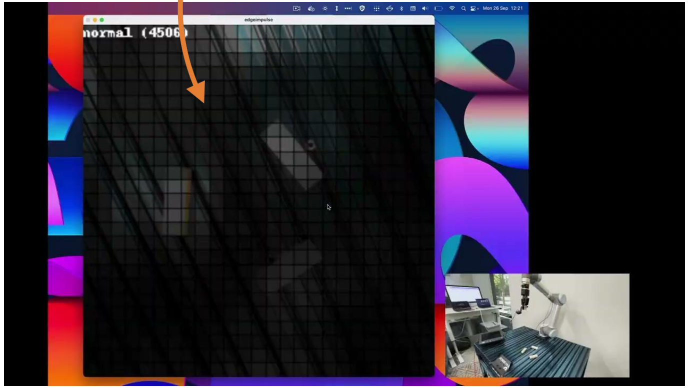


MODEL HIGHLIGHT

FOMO: Faster Objects, More Objects

- Remove classification head, replace with GMMs
- Only requires training on normal data
- Each cell tells you the chance that it's an anomaly
- Same performance as FOMO:
  Up to 30fps. on Cortex-M7, <200K RAM

Each cell is an anomaly detector

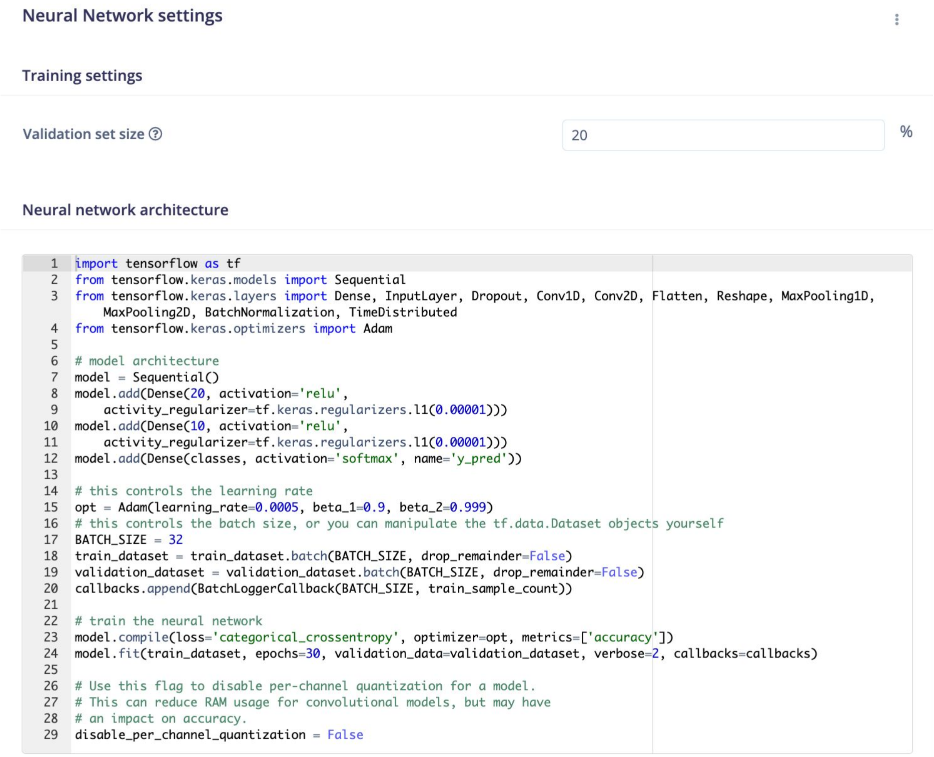


normal (4506)

FEATURE HIGHLIGHT

Keras Expert Mode

- For advanced users
- Use Keras standard API
- Customize NN architecture and take full control over training procedure



Neural Network settings

Training settings

Validation set size 20 %

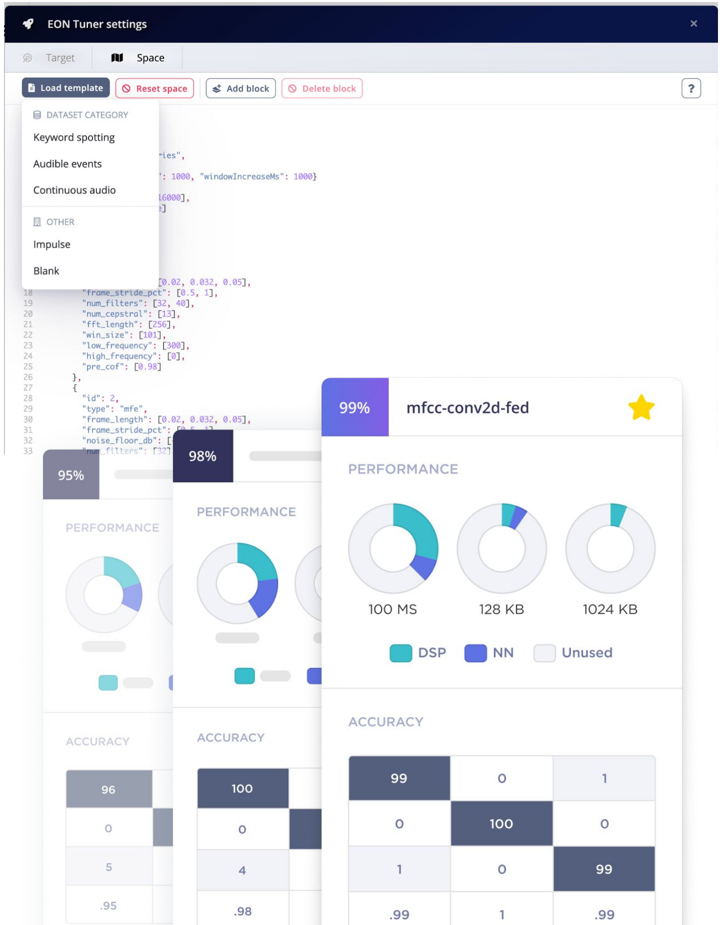
Neural network architecture

```
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, InputLayer, Dropout, Conv1D, Conv2D, Flatten, Reshape, MaxPooling1D,
    MaxPooling2D, BatchNormalization, TimeDistributed
from tensorflow.keras.optimizers import Adam

# model architecture
model = Sequential()
model.add(Dense(20, activation='relu',
    activity_regularizer=tf.keras.regularizers.l1(0.00001)))
model.add(Dense(10, activation='relu',
    activity_regularizer=tf.keras.regularizers.l1(0.00001)))
model.add(Dense(classes, activation='softmax', name='y_pred'))

# this controls the learning rate
opt = Adam(learning_rate=0.0005, beta_1=0.9, beta_2=0.999)
# this controls the batch size, or you can manipulate the tf.data.Dataset objects yourself
BATCH_SIZE = 32
train_dataset = train_dataset.batch(BATCH_SIZE, drop_remainder=False)
validation_dataset = validation_dataset.batch(BATCH_SIZE, drop_remainder=False)
callbacks.append(BatchLoggerCallback(BATCH_SIZE, train_sample_count))

# train the neural network
model.compile(loss='categorical_crossentropy', optimizer=opt, metrics=['accuracy'])
model.fit(train_dataset, epochs=30, validation_data=validation_dataset, verbose=2, callbacks=callbacks)

# Use this flag to disable per-channel quantization for a model.
# This can reduce RAM usage for convolutional models, but may have
# an impact on accuracy.
disable_per_channel_quantization = False
```

EON Tuner

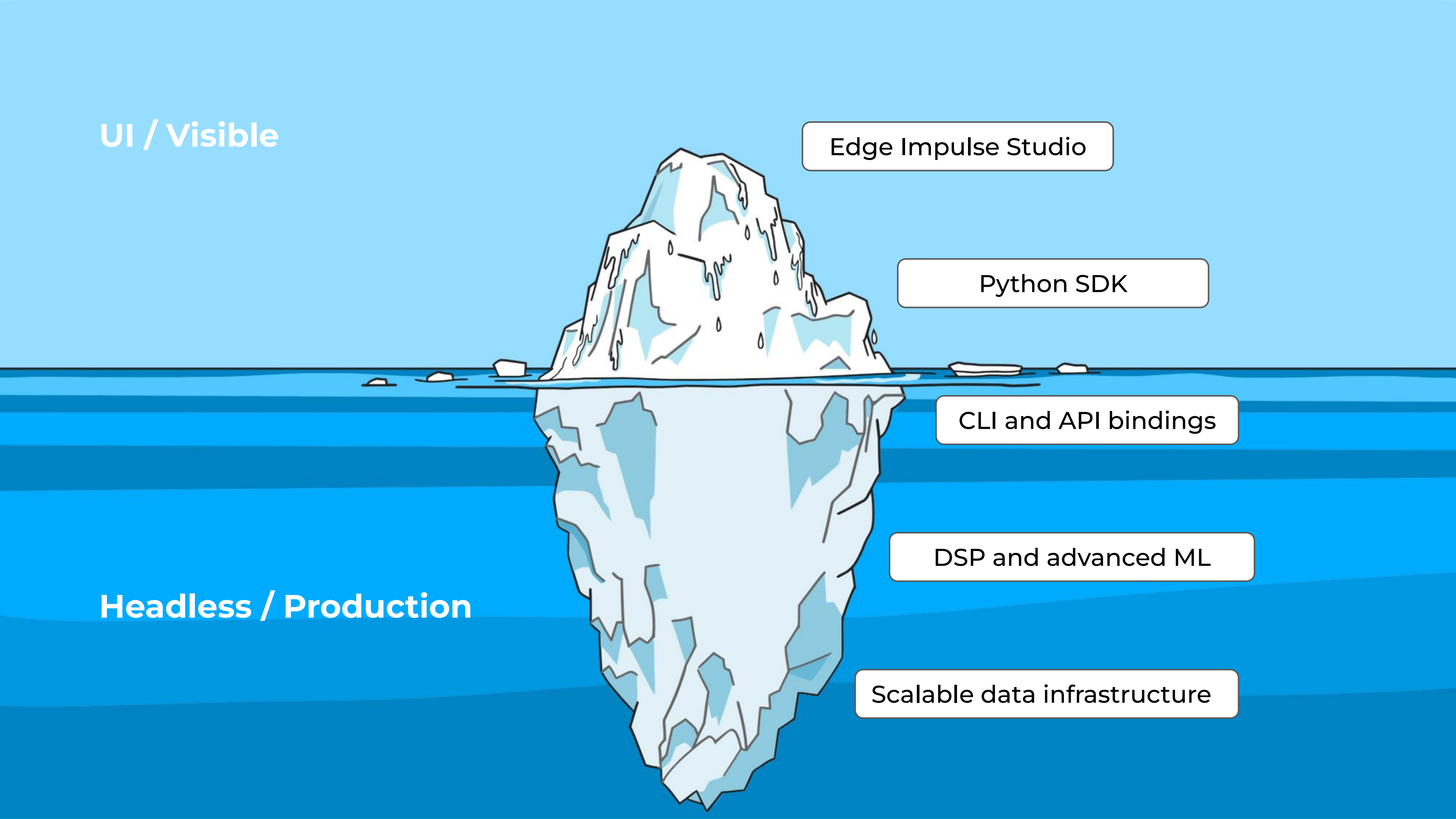
Establish a baseline quickly

- Search space based on prior knowledge of data modalities
- Reusable workers to minimize startup cost
- Customize search space



```
ies",
': 1000, "windowIncreaseMs": 1000}
16000],
e]
[0.02, 0.032, 0.05],
        "frame_stride_pct": [0.5, 1],
        "num_filters": [32, 40],
        "num_cepstral": [13],
        "fft_length": [256],
        "win_size": [101],
        "low_frequency": [300],
        "high_frequency": [0],
        "pre_cof": [0.98]
      },
      {
        "id": 2,
        "type": "mfe",
        "frame_length": [0.02, 0.032, 0.05],
        "frame_stride_pct": [0.5, 1]
        "noise_floor_db": [
        "num_filters": [32]
```

99	0	1
0	100	0
1	0	99
.99	1	.99



UI / Visible

Edge Impulse Studio

Python SDK

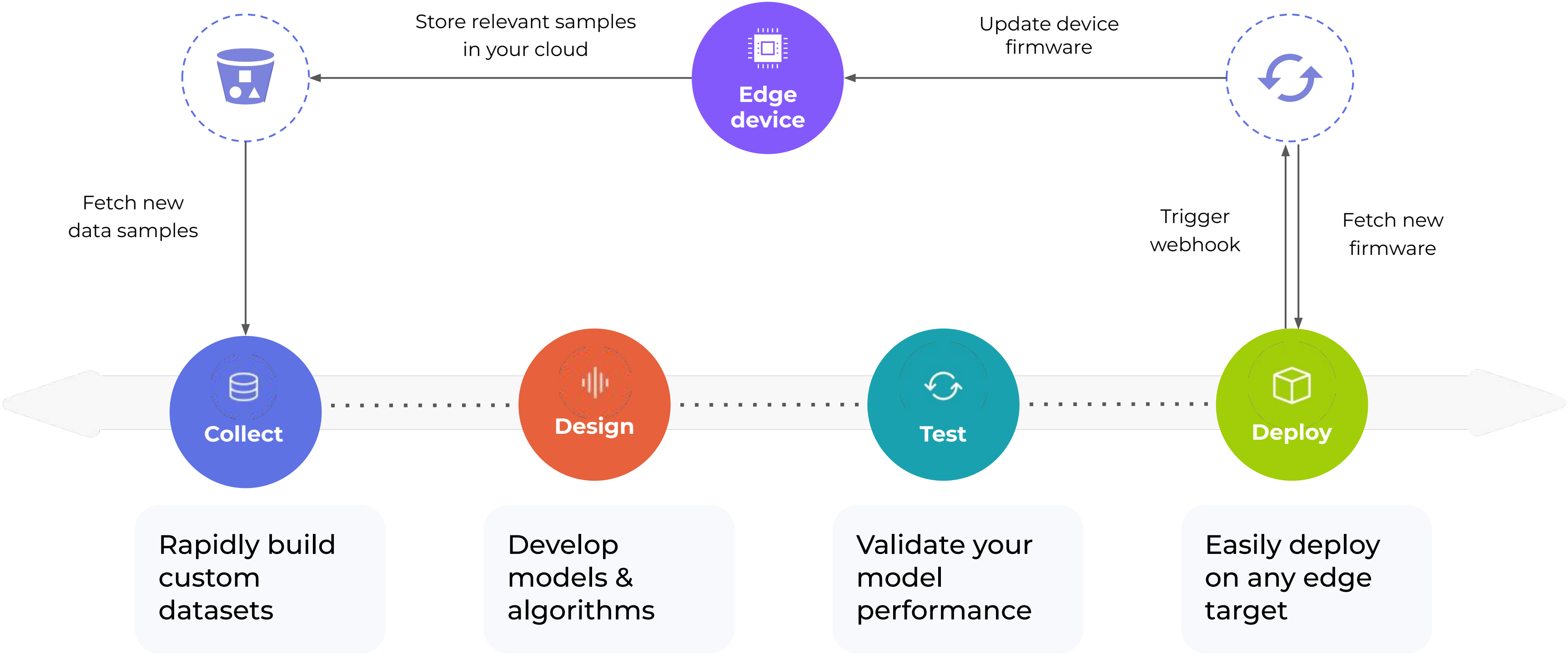
CLI and API bindings

DSP and advanced ML

Headless / Production

Scalable data infrastructure

Data-Centric ML



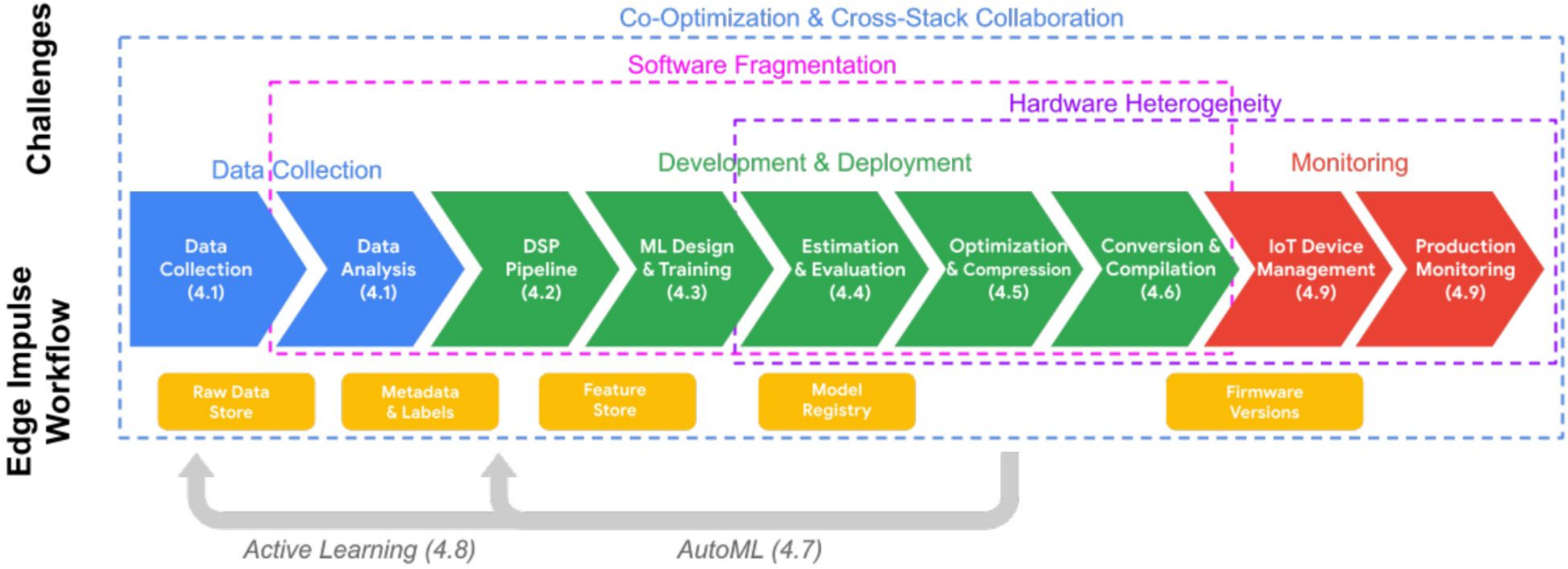
Rapidly build custom datasets

Develop models & algorithms

Validate your model performance

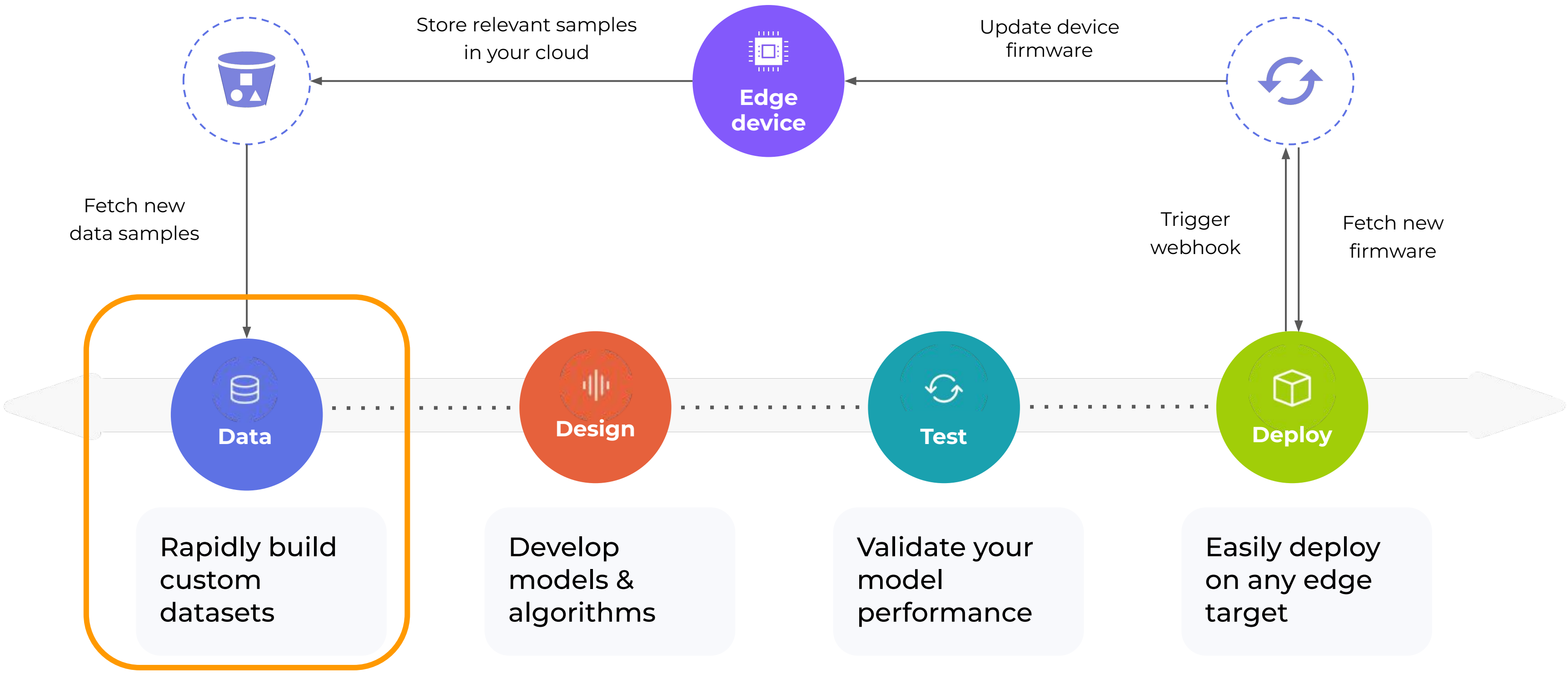
Easily deploy on any edge target

Active Learning with Edge Impulse



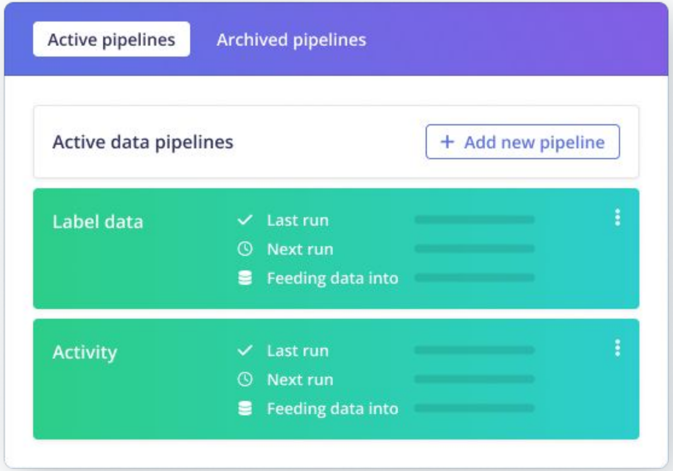
Source: Shawn Hymel, et al. Edge Impulse: An MLOps Platform for Tiny Machine Learning, November 2022. arXiv:2212.03332

Active Learning with Edge Impulse

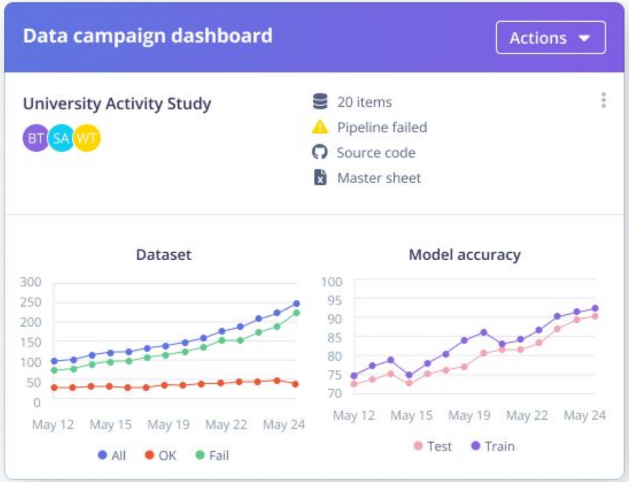

Store relevant samples in your cloud
Edge device
Update device firmware
Fetch new data samples
Trigger webhook
Fetch new firmware
Data
Design
Test
Deploy
Rapidly build custom datasets
Develop models & algorithms
Validate your model performance
Easily deploy on any edge target

Working with Data

- Data pipelines and transformation - enabling data preparation at scale
- Data campaign dashboards - optimize performance and share learnings



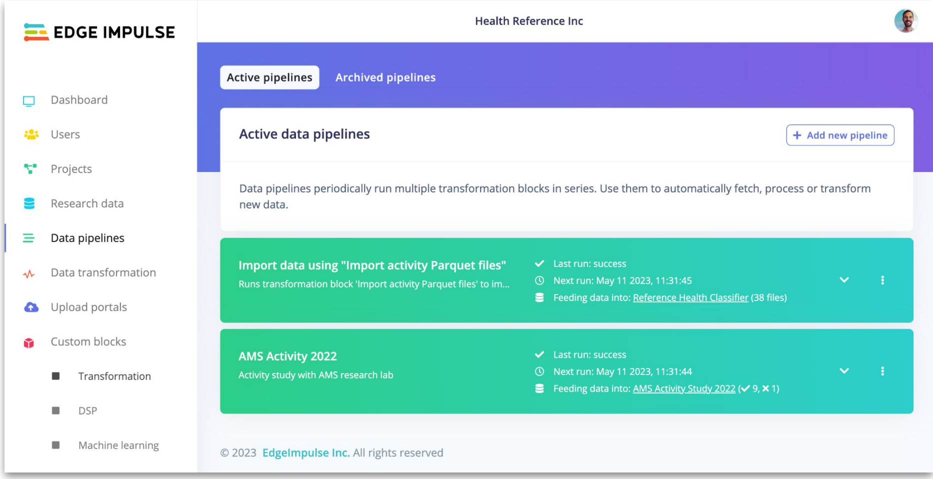
Data pipelines and transformations



Data dashboards

Data preparation

- Fetch data
- **Basic checks**: Are all files present? Do all files start / end around the same time? All expected labels for the study present?
- **Advanced checks**: Correlation between different devices (e.g. HR from PPG, and HR from Polar)?
- Runs automatically at a set interval (or on-demand, or triggered from code)
- Sends email on new data



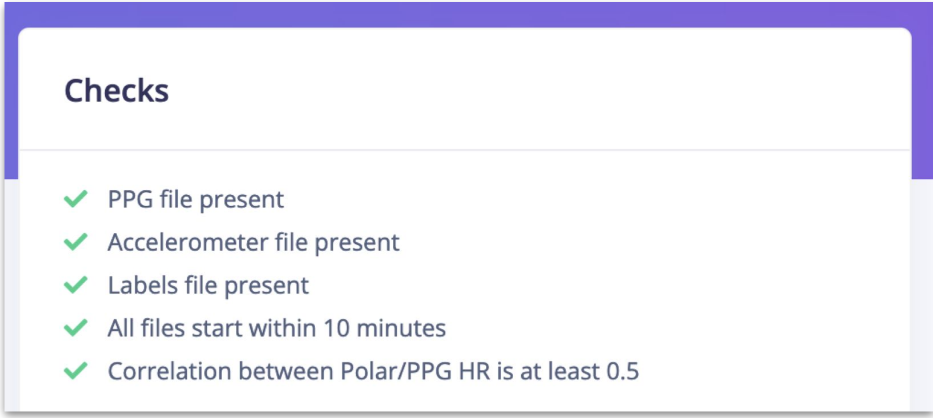
Active data pipelines

Data pipelines periodically run multiple transformation blocks in series. Use them to automatically fetch, process or transform new data.

Import data using "Import activity Parquet files"
Runs transformation block 'Import activity Parquet files' to im...
✓ Last run: success
Next run: May 11 2023, 11:31:45
Feeding data into: Reference Health Classifier (38 files)

AMS Activity 2022
Activity study with AMS research lab
✓ Last run: success
Next run: May 11 2023, 11:31:44
Feeding data into: AMS Activity Study 2022 (✓ 9, ✗ 1)

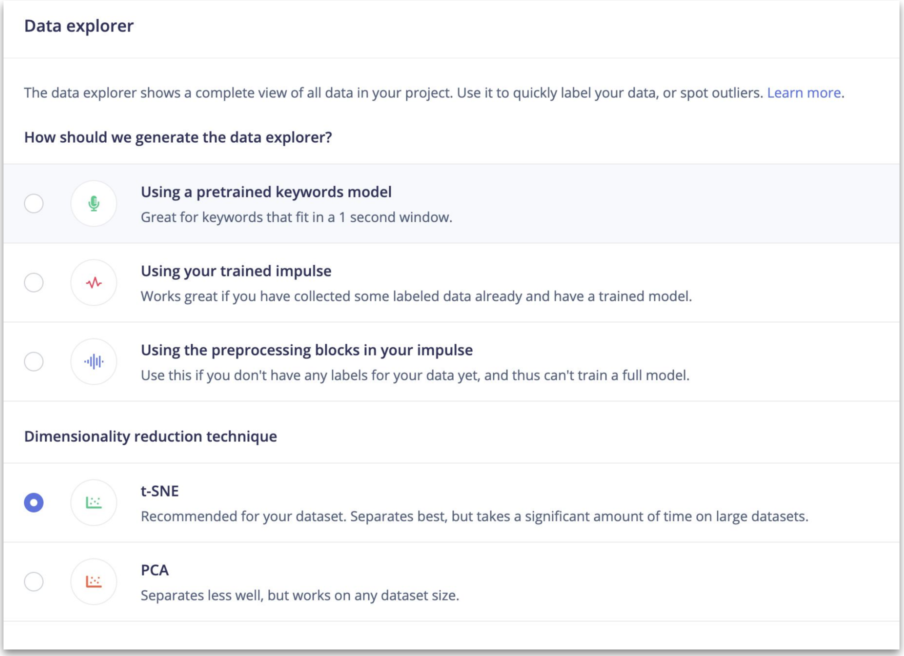
© 2023 EdgeImpulse Inc. All rights reserved



Checks

- ✓ PPG file present
- ✓ Accelerometer file present
- ✓ Labels file present
- ✓ All files start within 10 minutes
- ✓ Correlation between Polar/PPG HR is at least 0.5

Visualize data and uncover critical insights



Data explorer

The data explorer shows a complete view of all data in your project. Use it to quickly label your data, or spot outliers. Learn more.

How should we generate the data explorer?

○ **Using a pretrained keywords model**
Great for keywords that fit in a 1 second window.

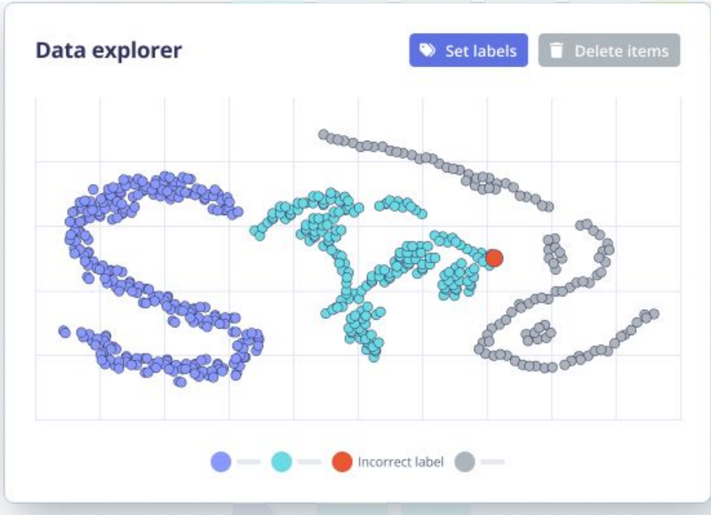
○ **Using your trained impulse**
Works great if you have collected some labeled data already and have a trained model.

○ **Using the preprocessing blocks in your impulse**
Use this if you don't have any labels for your data yet, and thus can't train a full model.

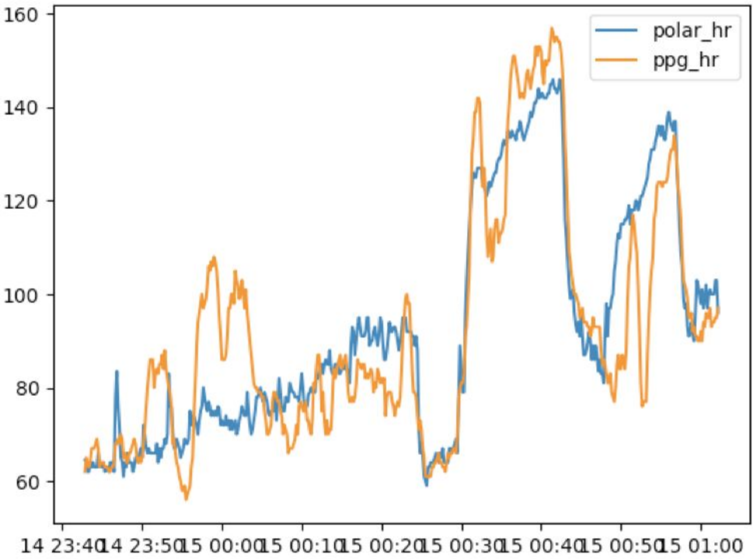
Dimensionality reduction technique

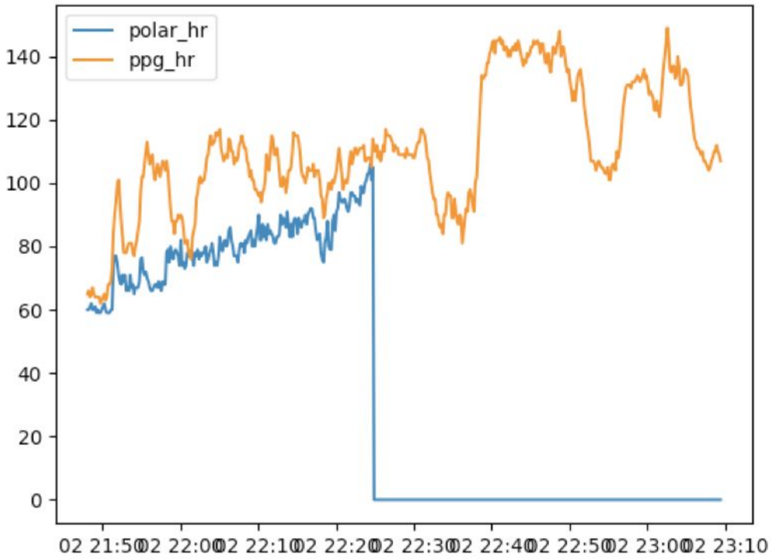
● **t-SNE**
Recommended for your dataset. Separates best, but takes a significant amount of time on large datasets.

○ **PCA**
Separates less well, but works on any dataset size.



Validating data: correlation





Fixes many issues: data uploaded for wrong participant, device failure and can be used to collect clock drift. Here implemented for PPG (derive HR) + Polar H10.

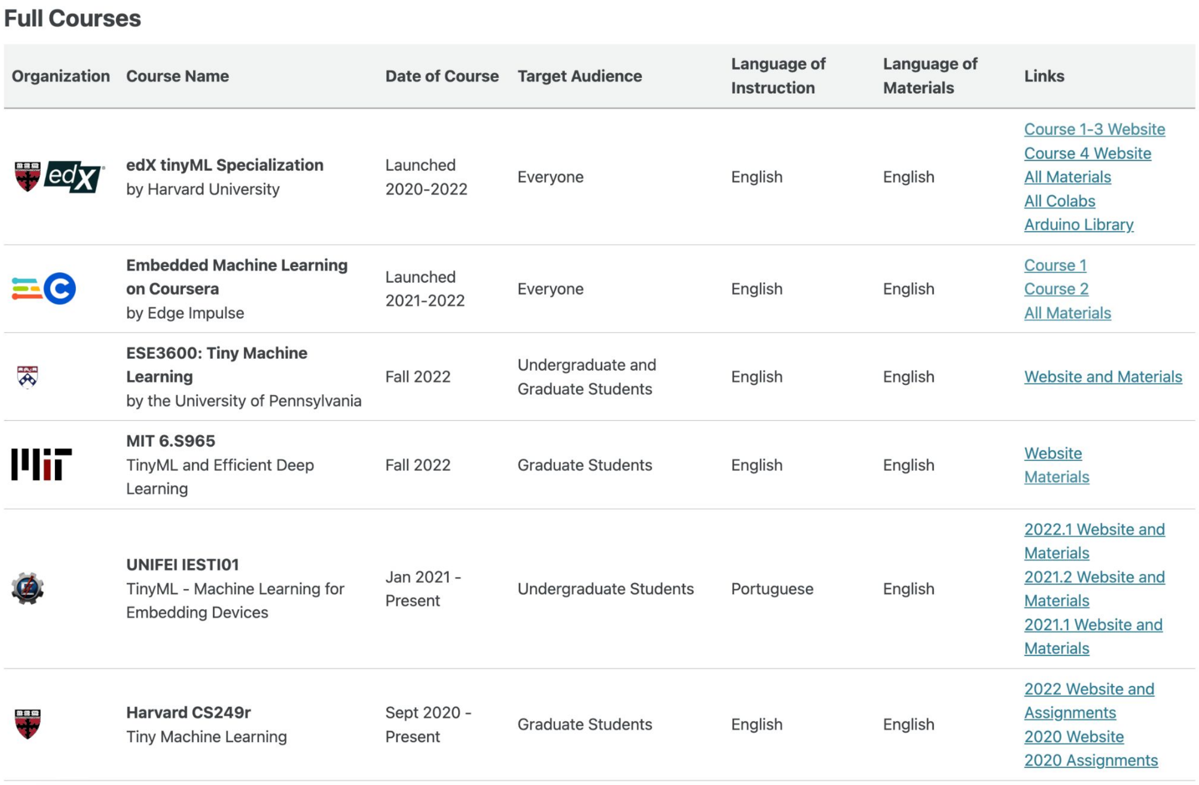
Resources

(Courses)

tinyml.seas.harvard.edu

Curriculum and Content

tinyml.seas.harvard.edu



Full Courses

Organization	Course Name	Date of Course	Target Audience	Language of Instruction	Language of Materials	Links
	edX tinyML Specialization by Harvard University	Launched 2020-2022	Everyone	English	English	Course 1-3 Website Course 4 Website All Materials All Colabs Arduino Library
	Embedded Machine Learning on Coursera by Edge Impulse	Launched 2021-2022	Everyone	English	English	Course 1 Course 2 All Materials
	ESE3600: Tiny Machine Learning by the University of Pennsylvania	Fall 2022	Undergraduate and Graduate Students	English	English	Website and Materials
	MIT 6.S965 TinyML and Efficient Deep Learning	Fall 2022	Graduate Students	English	English	Website Materials
	UNIFEI IESTI01 TinyML - Machine Learning for Embedding Devices	Jan 2021 - Present	Undergraduate Students	Portuguese	English	2022.1 Website and Materials 2021.2 Website and Materials 2021.1 Website and Materials
	Harvard CS249r Tiny Machine Learning	Sept 2020 - Present	Graduate Students	English	English	2022 Website and Assignments 2020 Website 2020 Assignments

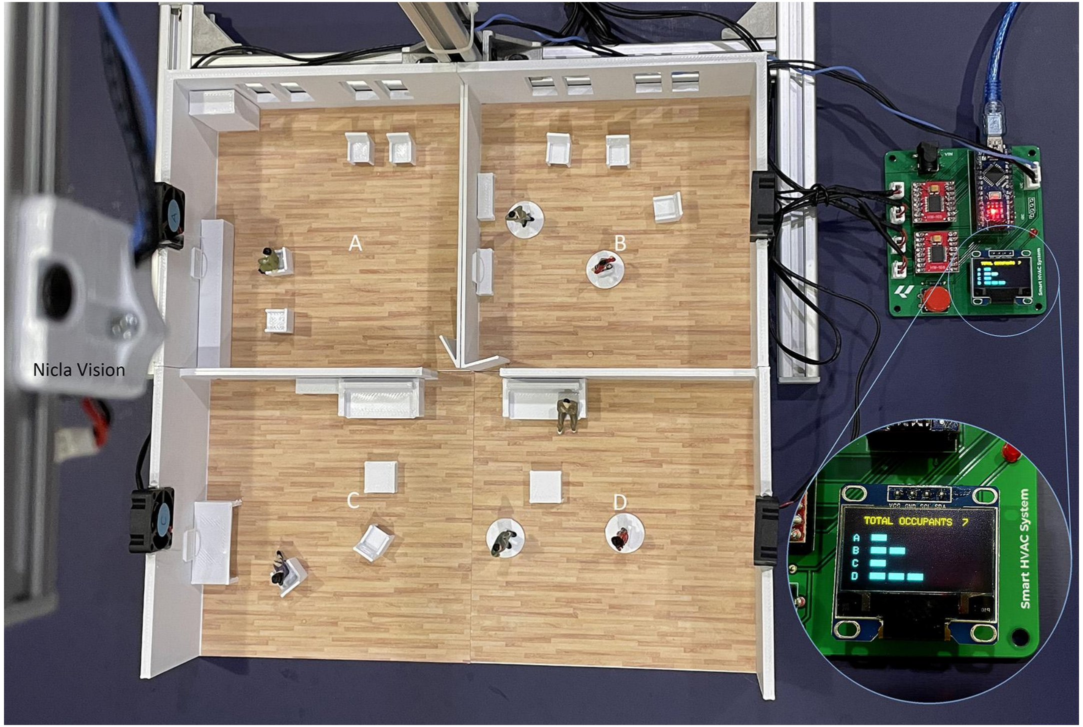
Resources

(Projects)

www.edgeimpulse.com/projects
docs.edgeimpulse.com/experts

Project: Smart HVAC

- **Creator:** Jallson Suryo
- **Description:** Set heating/cooling based on number of people in each room
- **Hardware:** Arduino Nicla Vision
- **Model:** FOMO



docs.edgeimpulse.com/experts/featured-machine-learning-projects/arduino-nicla-vision-smart-hvac

Project: Artificial Nose

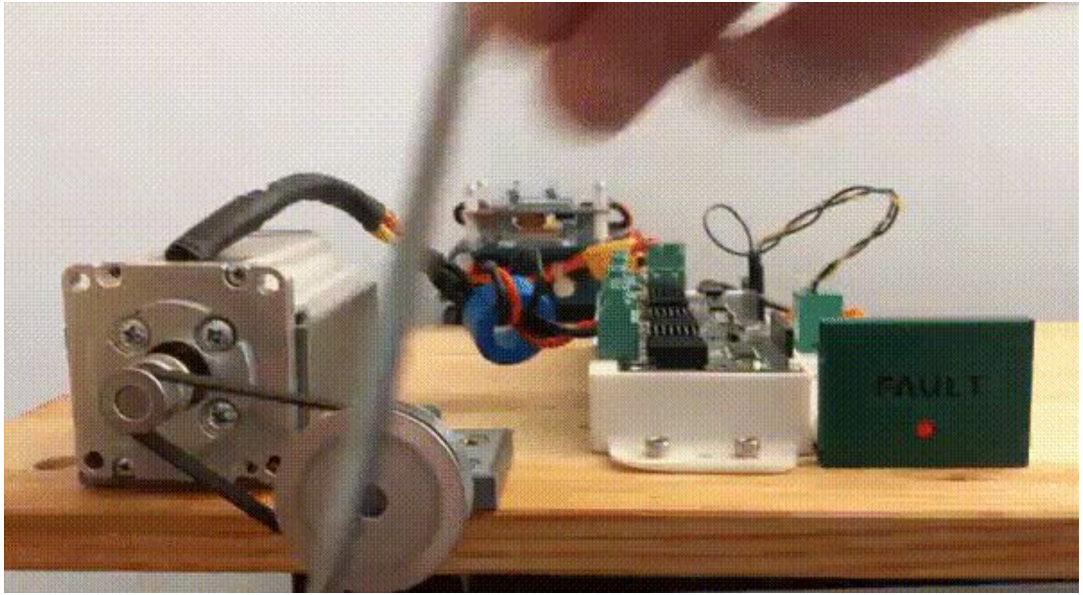
- **Creator:**
  Benjamin Cabé
- **Description:**
  Classify different odors based on gas data
- **Hardware:**
  Seeed Studio Wio Terminal
- **Model:**
  DNN

TinyML-powered
artificial nose
👃
kartben/artificial-nose

github.com/kartben/artificial-nose

Project: Motor Anomaly Detection

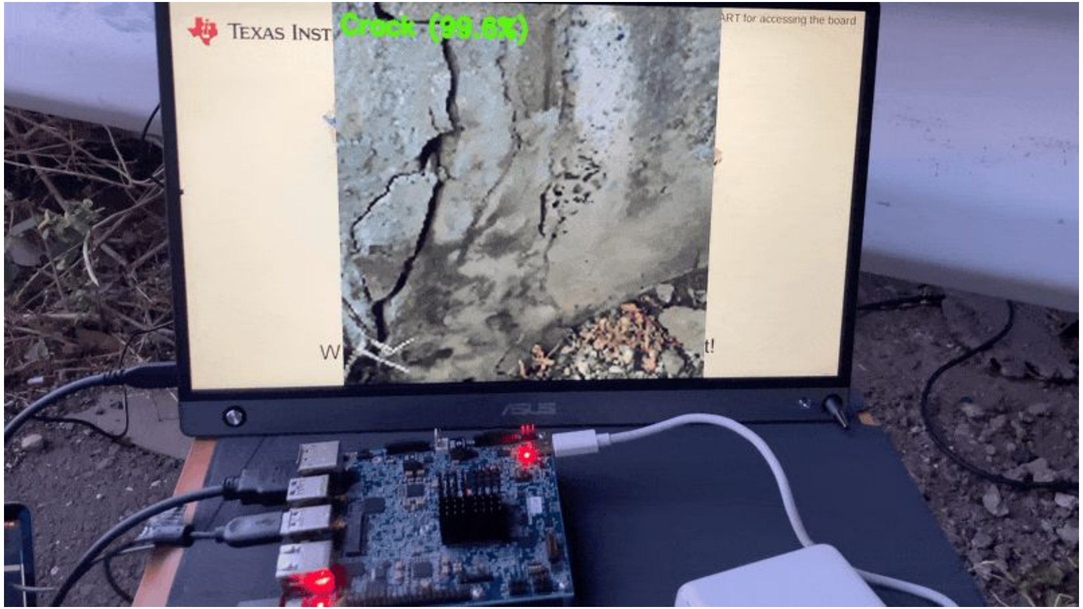
- **Creator**:
  Avi Brown
- **Description**:
  Identify anomalies based on motor current and voltage
- **Hardware**:
  Raspberry Pi Pico
- **Model**:
  K-means clustering



docs.edgeimpulse.com/experts/prototype-and-concept-projects/brushless-dc-motor-anomaly-detection

Project: Concrete Surface Crack Detection

- **Creator:**
  Naveen Kumar
- **Description:**
  Identify surface cracks in concrete structures
- **Hardware:**
  TI TDA4VM
- **Model:**
  MobileNetV2 with CAM



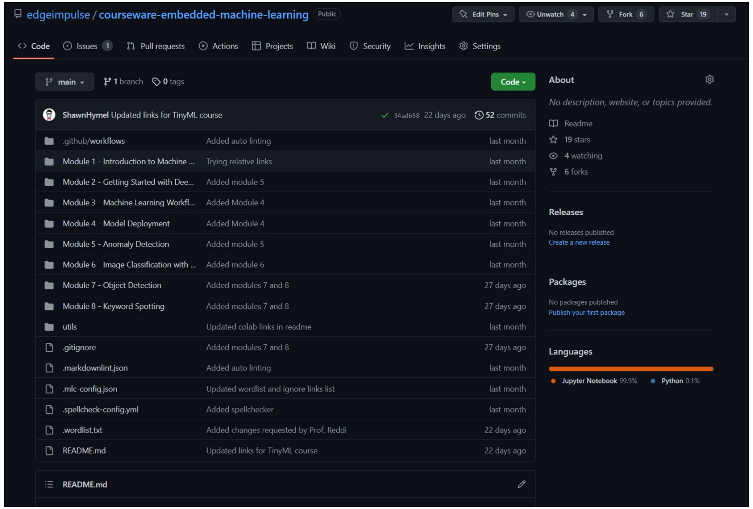
docs.edgeimpulse.com/experts/prototype-and-concept-projects/surface-crack-detection-ti-tda4vm

University Program

edgeimpulse.com/university

1. Free hardware kits
2. Content to build curriculum
3. Access to expert network
4. Discount to enterprise edition

Deadline July 16


edgeimpulse / courseware-embedded-machine-learning Public

ShawnHymel Updated links for TinyML course ✓ 34adb58 22 days ago 52 commits

.github/workflows	Added auto linting	last month
Module 1 - Introduction to Machine ...	Trying relative links	last month
Module 2 - Getting Started with Dee...	Added module 5	last month
Module 3 - Machine Learning Workfl...	Added Module 4	last month
Module 4 - Model Deployment	Added Module 4	last month
Module 5 - Anomaly Detection	Added module 5	last month
Module 6 - Image Classification with ...	Added module 6	last month
Module 7 - Object Detection	Added modules 7 and 8	27 days ago
Module 8 - Keyword Spotting	Added modules 7 and 8	27 days ago
utils	Updated colab links in readme	last month
.gitignore	Added modules 7 and 8	27 days ago
.markdownlint.json	Added auto linting	last month
.mlc-config.json	Updated wordlist and ignore links list	last month
.spellcheck-config.yml	Added spellchecker	last month
.wordlist.txt	Added changes requested by Prof. Reddi	22 days ago
README.md	Updated links for TinyML course	22 days ago

README.md

About

No description, website, or topics provided.

Readme
19 stars
4 watching
6 forks

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Languages

● Jupyter Notebook 99.9% ● Python 0.1%

Let’s simplify embedded ML for the next generation of engineers together

Thanks!

EDGE IMPULSE

hello@edgeimpulse.com

3031 Tisch Way
110 Plaza West
San Jose, CA 95128
USA