



We put ML into real products

Louis Moreau, Senior DevRel Engineer
[@luisomoreau](https://twitter.com/luisomoreau)



A paradigm shift

Traditional programming

Data

+

Rules

Adapt the rules

```
Block 1
This example code is in the public domain.
http://www.arduino.cc/en/Tutorial/Blink

void setup() {
  // initialize digital pin BUILTIN as an output:
  pinMode(BUILTIN, OUTPUT);
}

// the loop function runs over and over again forever
void loop() {
  digitalWrite(BUILTIN, HIGH); // turn the LED on (HIGH is the voltage level)
  delay(1000);               // wait for a second
  digitalWrite(BUILTIN, LOW); // turn the LED off by making the voltage LOW
  delay(1000);               // wait for a second
}
```

Outcomes

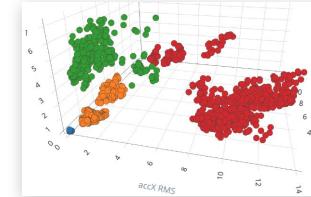
Machine learning

Data

+

Outcomes

Collect more data



Rules

Benefits of edge ML



Innovation

Add new differentiating features, become a market leader by standing out from your competition



Privacy

Data stays on the device, gets processed locally and drives remote alerts, notifications, and actions



Power

Stay operational for longer periods of time



Cost

Save on storage and compute costs by not sending raw data constantly to the cloud



Reliability

Be operational in low connectivity environments



Bandwidth & Latency

Process data real-time on the edge device, without having to wait for a response back from the cloud

Platform



Powering the largest edge ecosystem with MLOps

40,000+

Developers

90,000+

Projects

1,000+

Enterprises

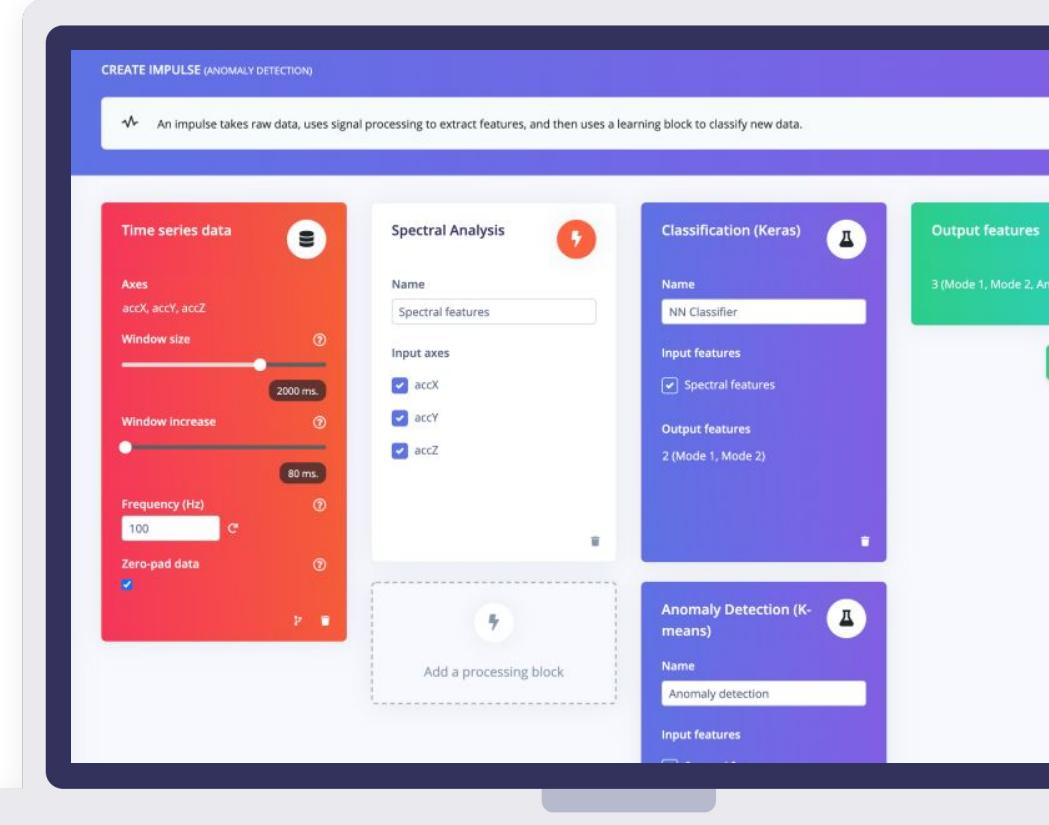
TRUSTED BY LEADING ENTERPRISES





The developer-first edge ML platform

- Royalty-free business model, therefore no impact on BOM cost
- Your IP, stays your IP
- Total explainability, no black boxes

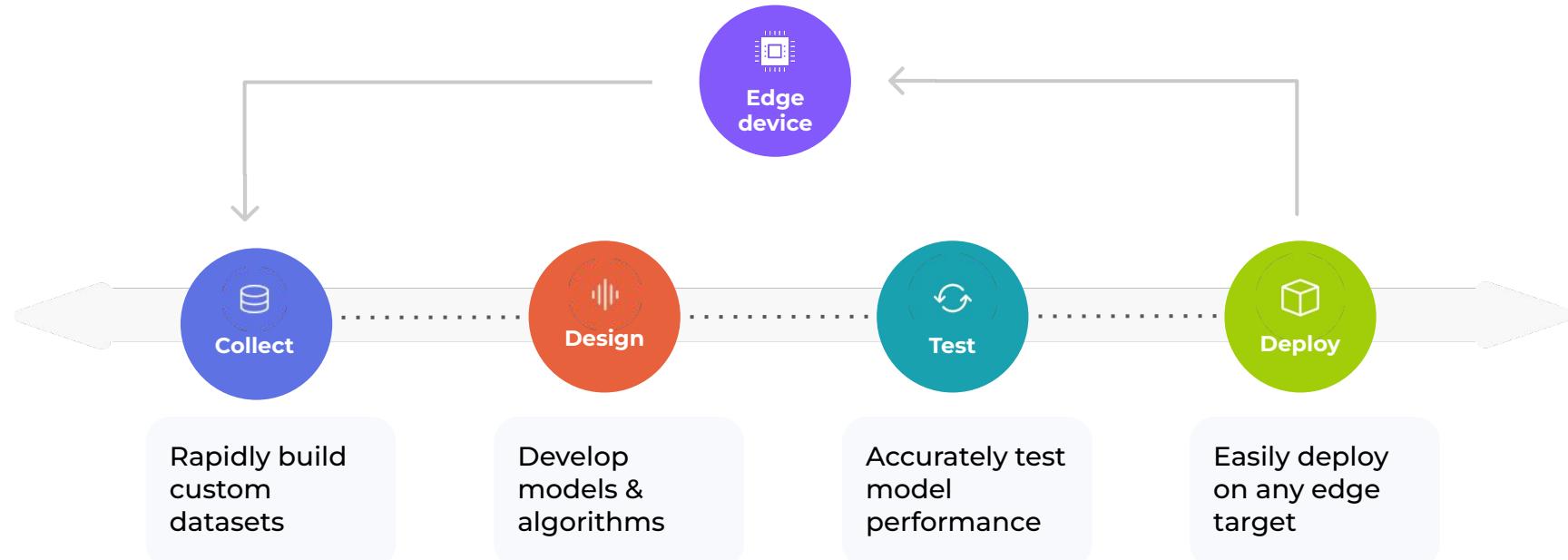


Any sensor, any data, any use case

	Ultra low power	Low-end MCU	High-end MCU	NPU	MPU	GPU
Memory	Anomaly detection 10kB	Sensor fusion classification 18kB	Audio classification 50kB	Image classification 256kB	Complex image or voice 1MB+	Video classification 1GB+
Sensor	✓	✓	✓	✓	✓	✓
Audio	✓	✓	✓	✓	✓	✓
Image			✓	✓	✓	✓
Video					✓	✓

Develop edge ML applications with Edge Impulse

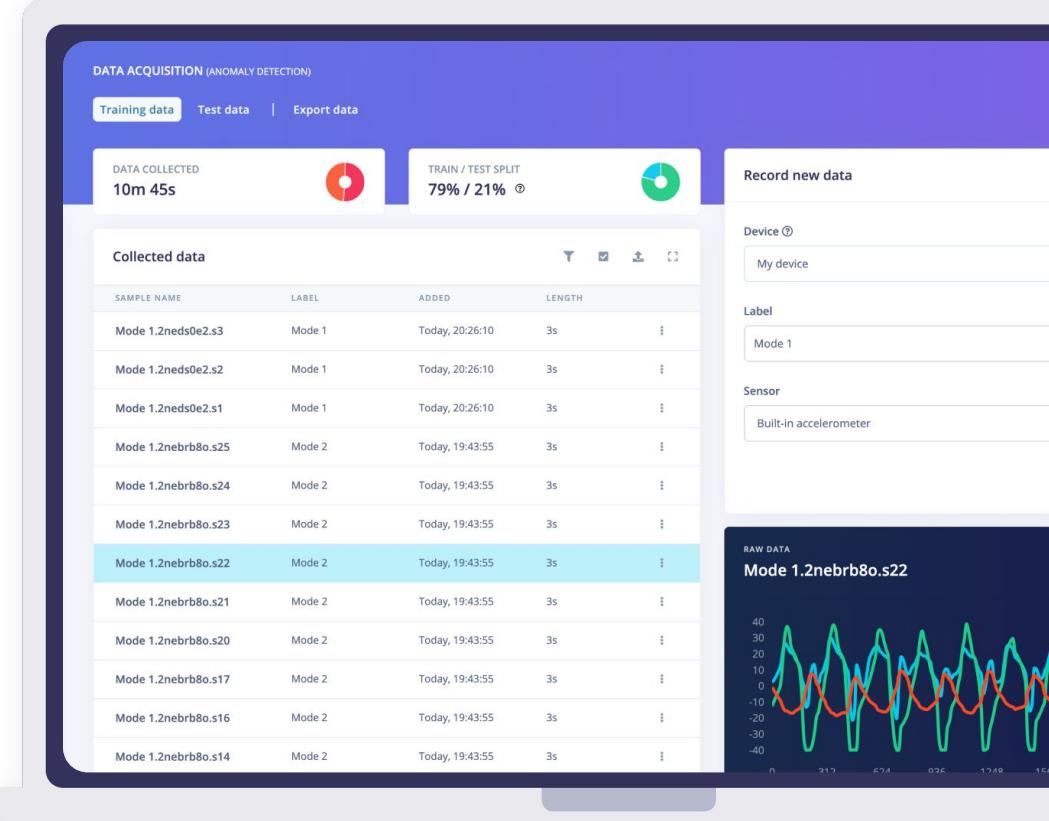
The infrastructure and integrations your data science and ML teams need



Collect

Build valuable datasets at scale

- Tools to collect data directly from devices



Collect

Build valuable datasets at scale

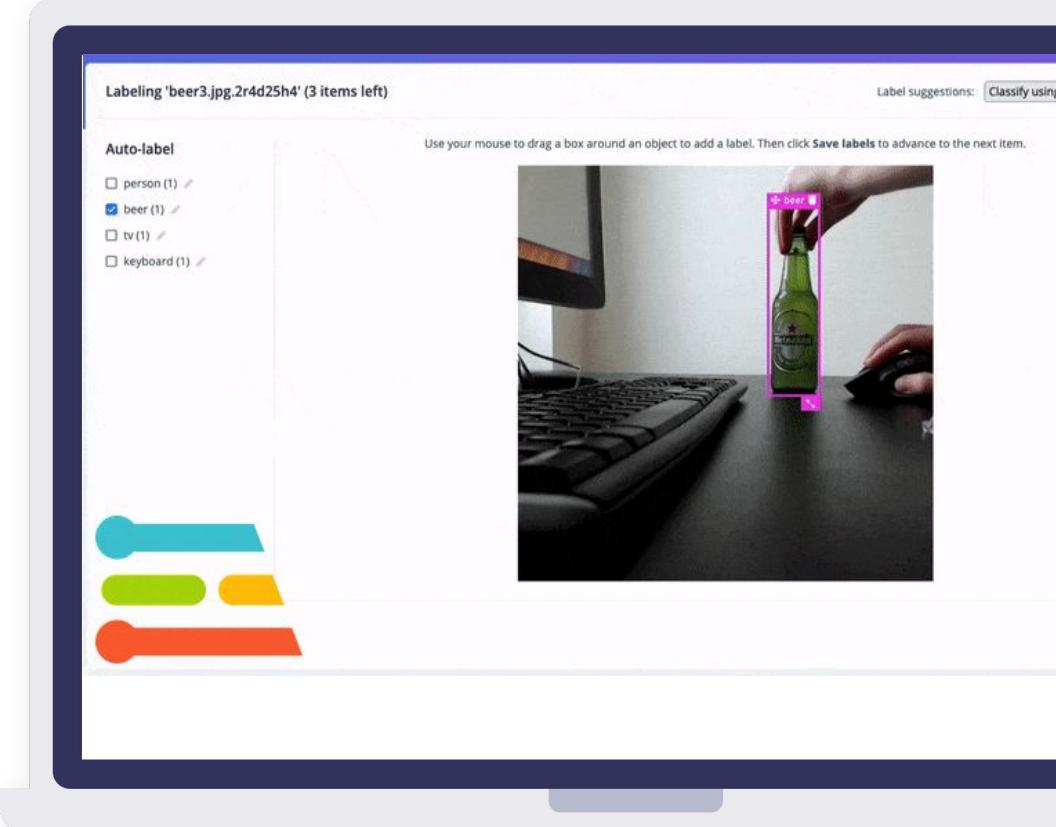
- Tools to collect data directly from devices
- Assisted-labeling tools



Collect

Build valuable datasets at scale

- Tools to collect data directly from devices
- Assisted-labeling tools



Collect

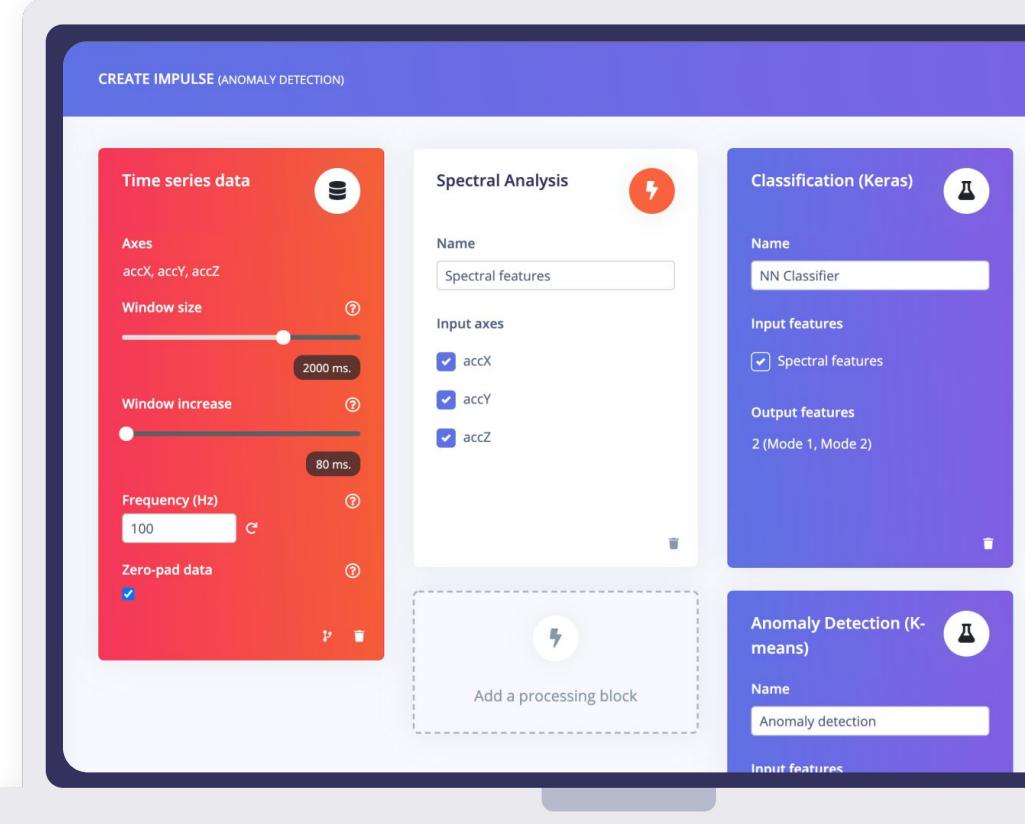
Build valuable datasets at scale

- The infrastructure data science teams need
- Assisted-labeling tools
- Integrations with most widely used cloud data buckets

The screenshot shows the Edge Impulse web application. On the left is a sidebar with icons for Dashboard, Devices, Data sources (selected), Data acquisition, Impulse design, Create impulse, EON Tuner, Retrain model, Live classification, Model testing, Versioning, Deployment, Documentation, and Forums. The main area has a blue header 'DATA SOURCES' and a sub-header 'Active data'. A modal window is open, titled 'Automatically fetch new data'. It asks 'Where does your data live?' with options: 'Amazon S3' (selected), 'Google Cloud Storage', 'Somewhere else (S3 compatible)', 'Upload portal in your Edge Impulse organization', and 'Don't import data'. Below each option is a brief description. A green button at the bottom right of the modal says 'Next, provide credentials'.

Advanced algorithm and ML expertise

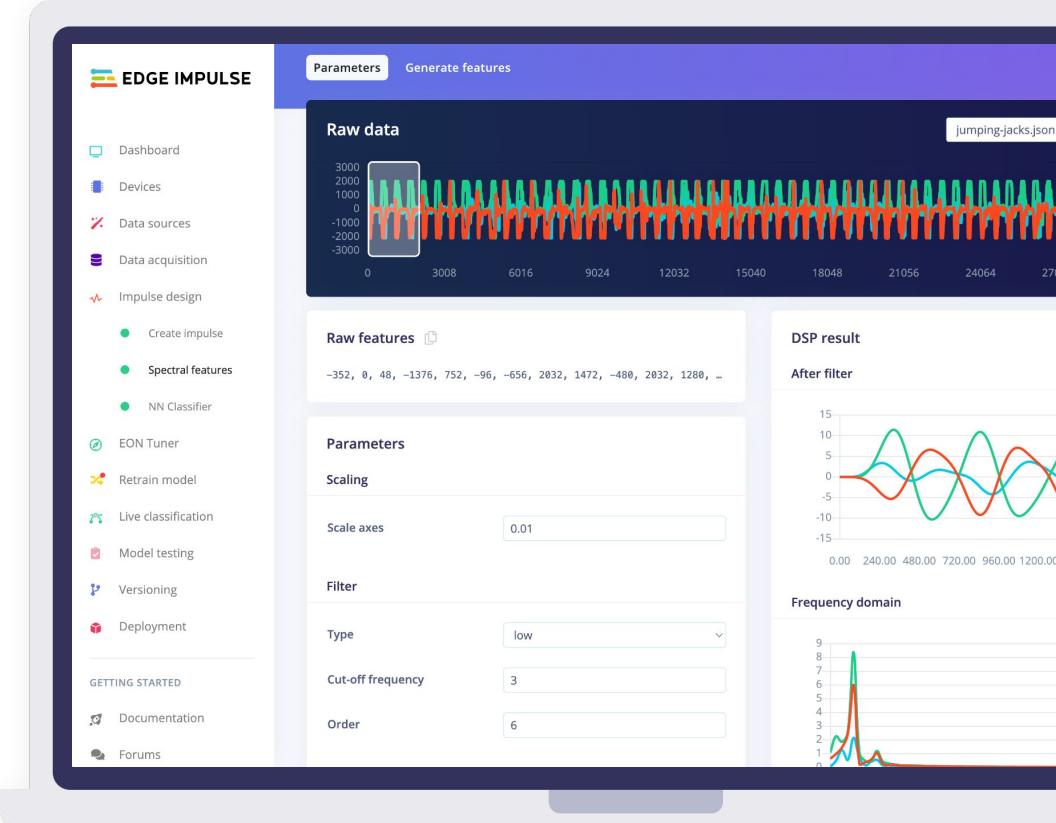
- Advanced algorithm and DSP expertise
- No black boxes
- Explainable AutoML
- Knowledge sharing and collaboration between teams



Design

Advanced algorithm and ML expertise

- Advanced algorithm and DSP expertise
- No black boxes
- Explainable AutoML
- Knowledge sharing and collaboration between teams



Design

Advanced algorithm and ML expertise

- Advanced algorithm and DSP expertise
- No black boxes
- Explainable AutoML
- Knowledge sharing and collaboration between teams

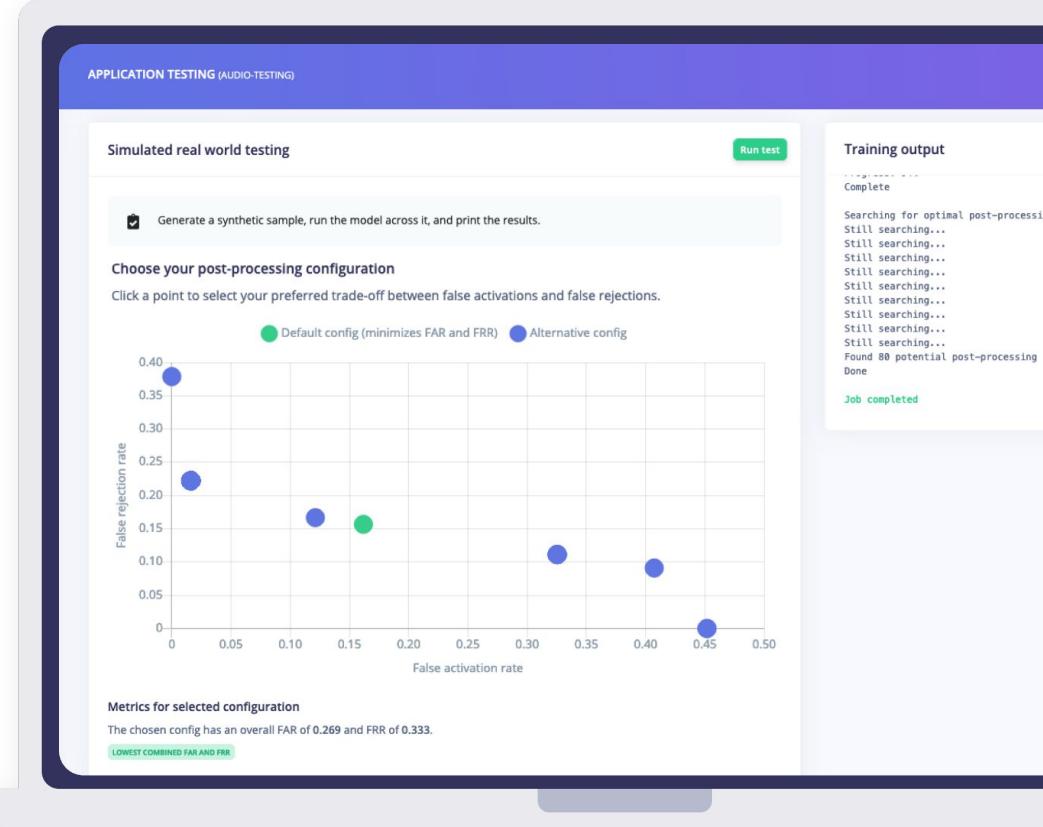
The screenshot displays the Edge Impulse software interface, which is a platform for developing machine learning models for edge devices. The interface is divided into several sections:

- Left sidebar:** A navigation menu with icons and text links for various features: Dashboard, Devices, Data sources, Data acquisition, Impulse design (with sub-options: Create impulse, Spectral features, NN Classifier), EON Tuner, Retrain model, Live classification, Model testing, Versioning, Deployment, Documentation, and Forums.
- Central area:**
 - Neural Network settings:** Includes "Training settings" (Number of training cycles: 40, Learning rate: 0.0005, Validation set size: 20%, Auto-balance dataset: checked) and "Neural network architecture".
 - The architecture shows an **Input layer (33 features)**, followed by a sequence of layers: Dense layer (16 neurons), Dense layer (8 neurons), Dropout (rate 0.25), Dense layer (8 neurons), and an **Output layer (4 classes)**.
 - An option to "Add an extra layer" is available.
 - Training output:** Shows "Model version: C" and "Last training performance (validation set)". It includes an accuracy metric of **99.6%** and a "Confusion matrix (validation set)" table.
 - Feature explorer:** A 3D scatter plot showing feature relationships between accX RMS, accY RMS, and accZ RMS. The legend indicates four classes: jumping-jacks - correct (yellow), no-movement - correct (green), rest - correct (blue), and squats - correct (red). A legend also shows red dots for rest - incorrect.

Test

Go to market faster, with confidence

- Hardware-aware development
- Full visibility across the whole ML pipeline
- Test your development against 24hrs of real world data
- Tune the post-processing algorithm to perform optimally



Deploy

Deploy to any edge device with ease

- The largest silicon ecosystem
- Award-winning compiler
- Get access to full source code
- Full firmware integration for a number of devices
- Digital twin for performance profiling and analysis
- Enable brownfield and future greenfield

The screenshot shows a user interface for deploying machine learning models. At the top, it says "DEPLOYMENT (TUTORIAL: CONTINUOUS MOTION RECOGNITION)". Below that, a section titled "Deploy your impulse" explains that you can deploy to any device without an internet connection, minimizing latency and power consumption, with a link to "Read more". A "Create library" section describes turning the impulse into optimized source code for any device, with icons for C++ library, Arduino library, and Cube.MX CMSIS-PACK. Another section, "Build firmware", shows three development boards: a Raspberry Pi, a BeagleBoard, and a STM32Cube board, with the STM32Cube board labeled "END OF LIFE".

DEPLOYMENT (TUTORIAL: CONTINUOUS MOTION RECOGNITION)

Deploy your impulse

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](#).

Create library

Turn your impulse into optimized source code that you can run on any device.

C++ library

Arduino library

Cube.MX CMSIS-PACK

WebAssembly

NVIDIA.

TensorRT library

Build firmware

Or get a ready-to-go binary for your development board that includes your impulse.

Raspberry Pi

BeagleBoard

STM32Cube END OF LIFE



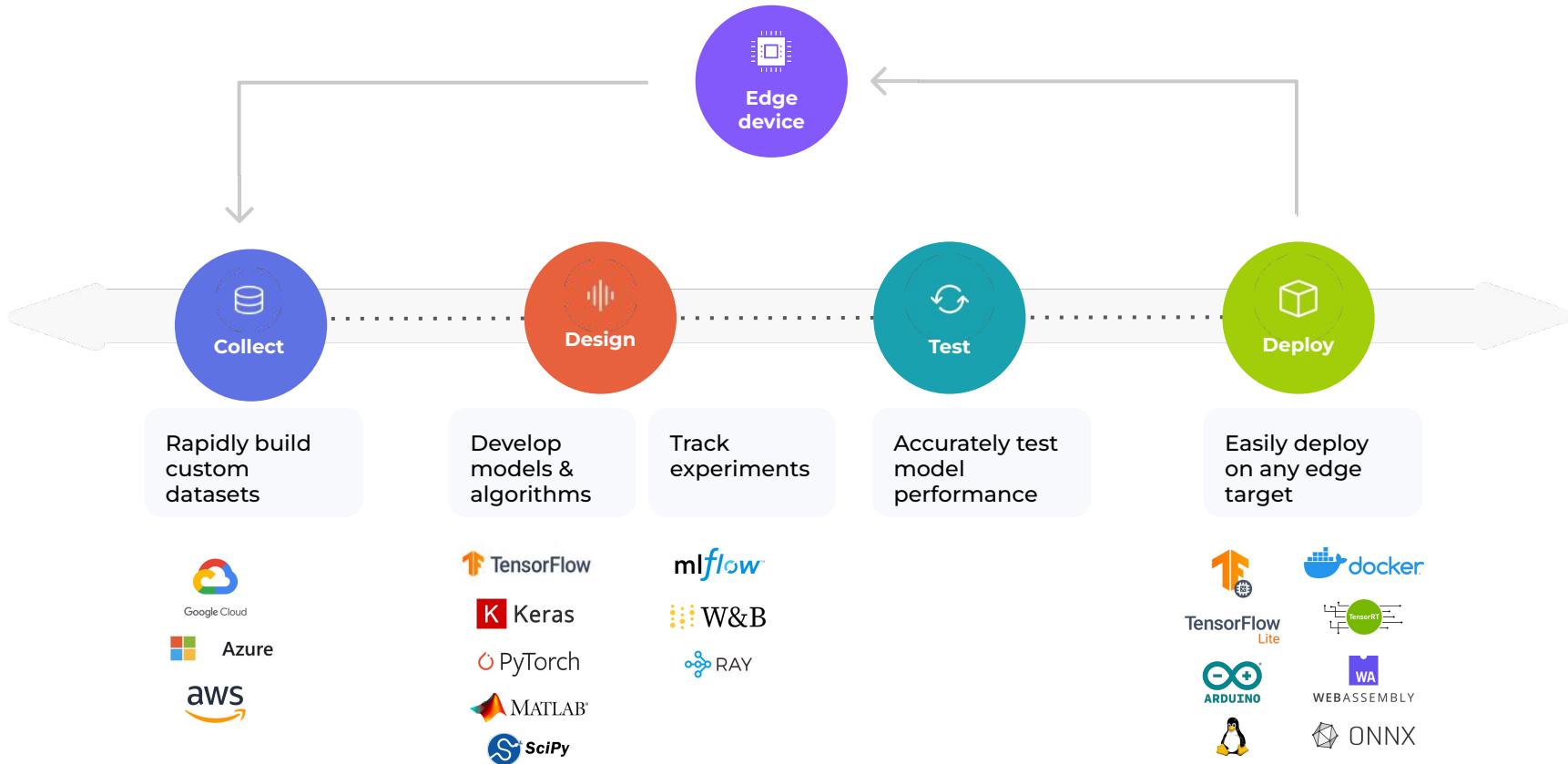
Edge device

Comprehensive hardware support

Benefit from the leading edge ML silicon ecosystem

Integrations

arm NVIDIA intel RISC-V

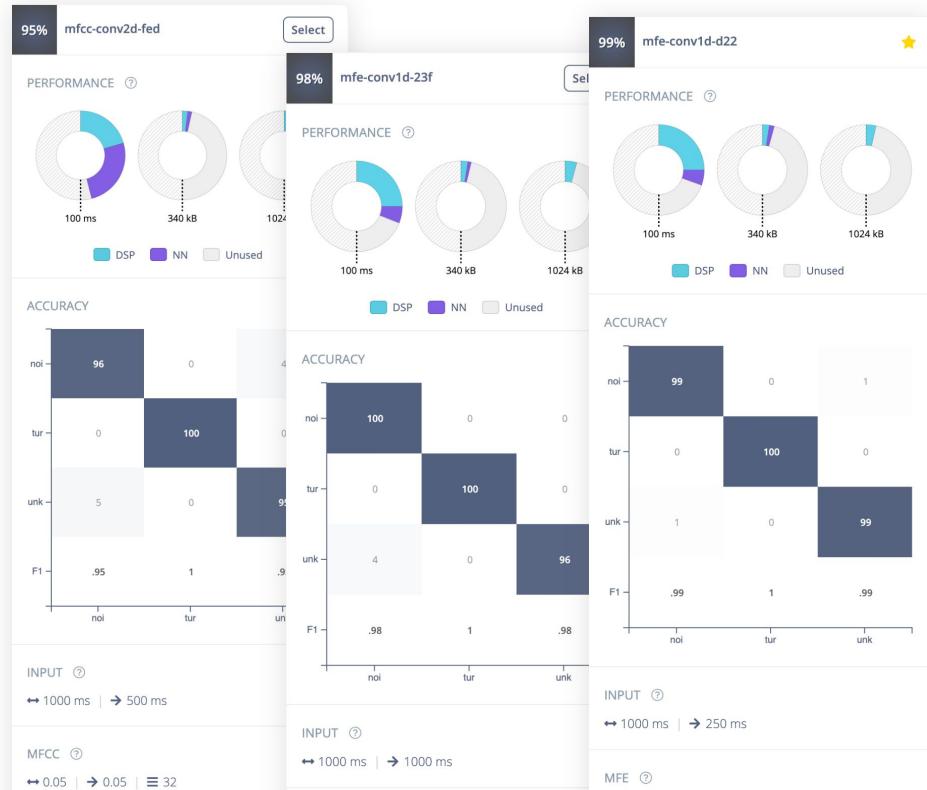


Features



Optimal ML solutions with EON Tuner

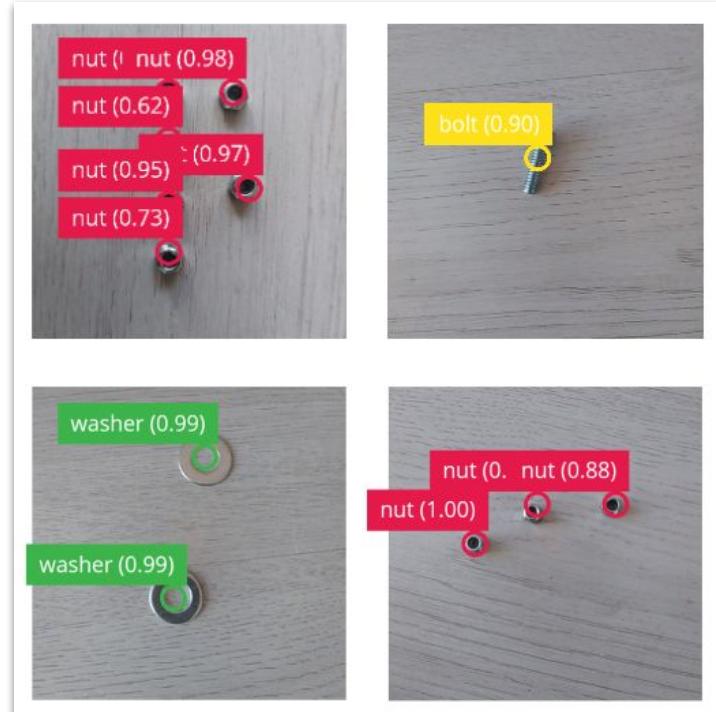
- Make the most of limited and heterogeneous compute
- Find the perfect balance between feature extraction and model architecture
- Recommendations based on real performance metrics
- Built for constrained use cases to high end HW and complex use cases (e.g.: high-end CV)
- Suggestions on optimal hardware target for use case



FOMO: Faster objects, more objects

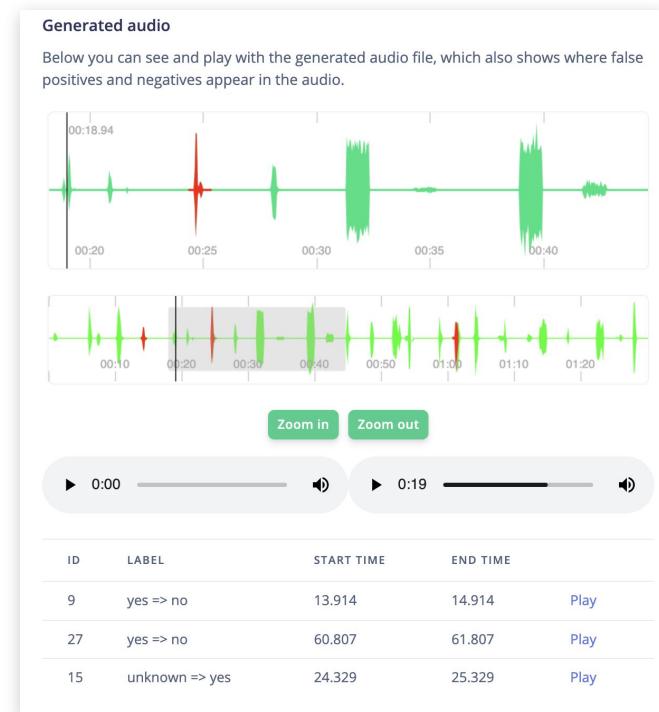
- **20x average performance improvement**
- Object detection on MCUs
- Ultra fast on embedded Linux
- Better at detecting smaller and more numerous objects
- Capable of segmentation and counting objects

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



Calibrate your application at scale

- Test on realistic samples: 24 hours of real world audio
- Understand the impact of post-processing while accounting for device constraints and latency
- Choose your ideal balance between false activations and false rejections
- Leverage genetic algorithms to design optimal post-processing configuration



Customers



Advantech increases manufacturing productivity

Visual inspection system to flag delays on the production line in real-time

Vision

- A reported 15% overall increase in production line efficiency
- Faster detection of idle time raises assembly-line productivity
- Managers free up time to focus on production planning and operations



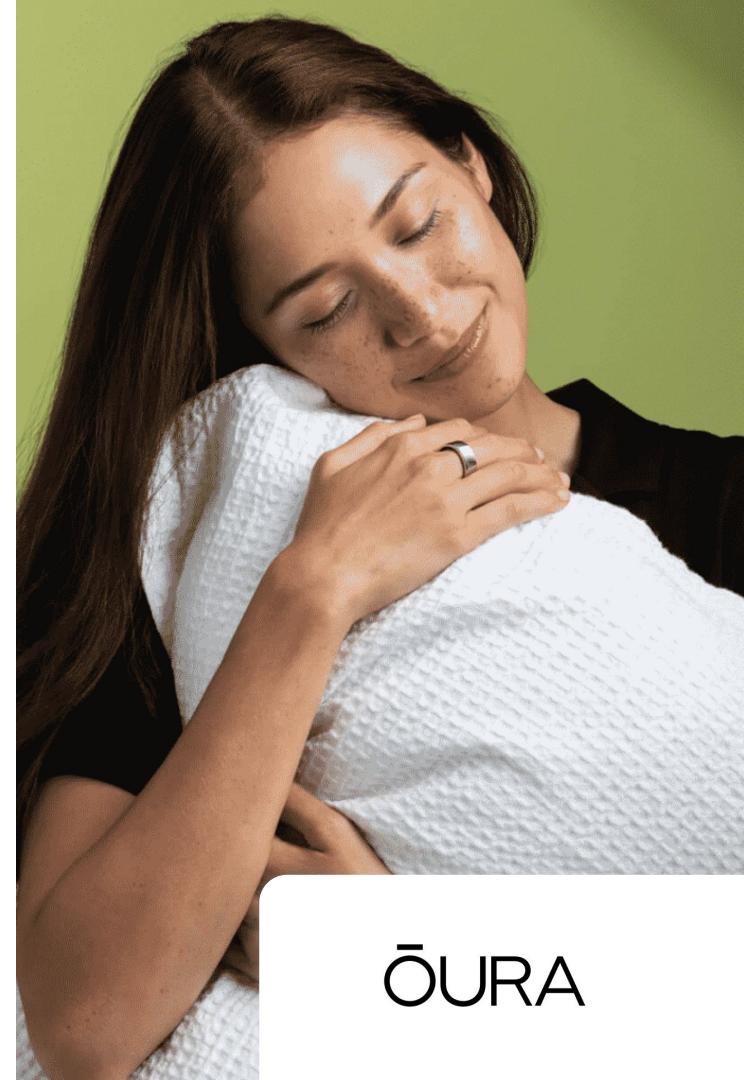
ADVANTECH

Oura goes deeper on deep sleep

Through the use of Edge Impulse's advanced data infrastructure, Oura rapidly improved their algorithm

Heart Motion Temperature

- Unprecedented sleep-scoring accuracy. A 17% point increase in scoring accuracy
- Improved correlation accuracy of 79%
- Data-driven development process enabled Oura's data science team to scale



Nordic and Izoelektro predict power line failure

Smart power grid monitoring that improves the operation, stability, and reliability of electricity distribution.”

Current

Motion

- Automated monitoring of poles and lines
- NB-IoT with 20 year battery life made possible by ML
- Avoid disastrous wildfires and reduce maintenance costs



 NORDIC®
SEMICONDUCTOR

 IZOELEKTRO

Where's my pallet?

Reduce power use on battery operated devices in pallets indoors and outdoors.

Implementation

- Gather vibration / accelerometer data from existing tracking devices in pallets (wood and plastic)
- Label, train model to identify if in a forklift, truck, idle, indoor / outdoor.
- Test in pallets in select service centers

