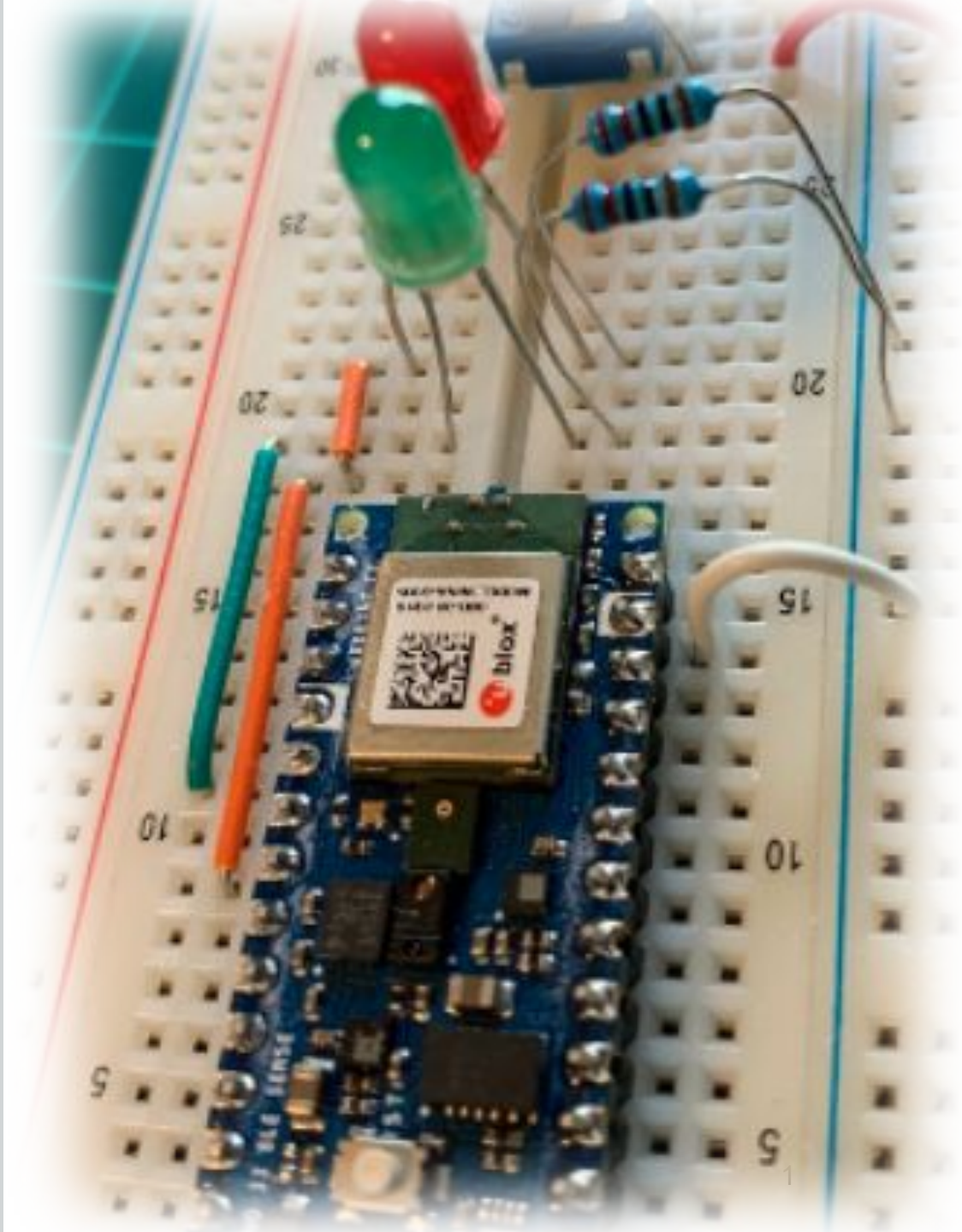
IESTI01 – TinyML

Embedded Machine Learning

12. Introduction to Edge Impulse – CNN with Cifar-10

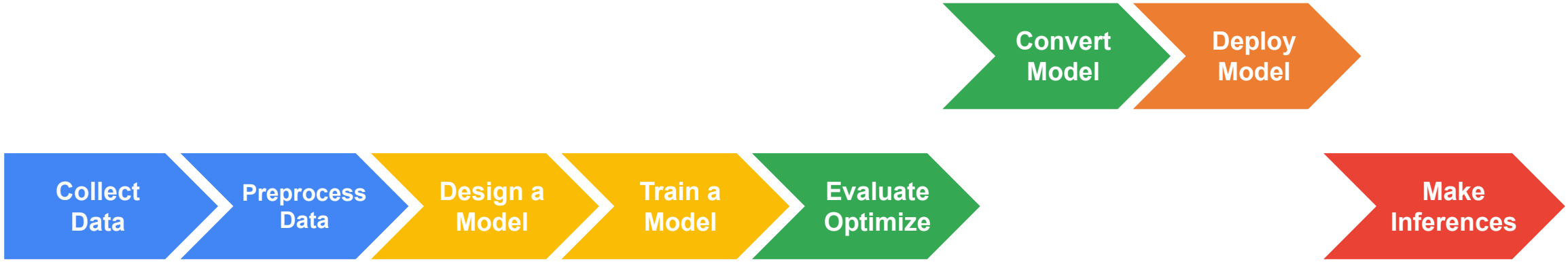


Prof. Marcelo Rovai
UNIFEI

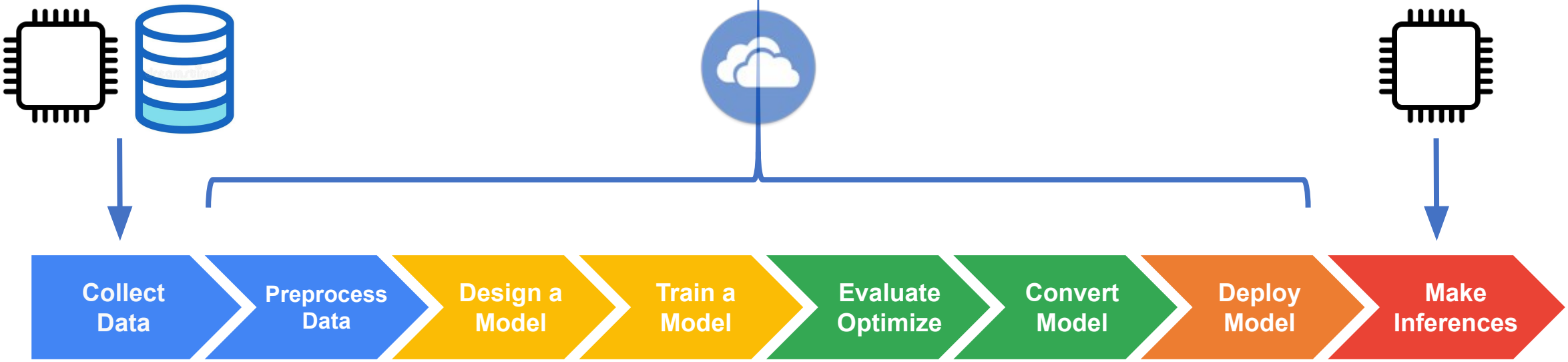
Embedded Machine Learning (TinyML) Workflow Review

Machine Learning Workflow

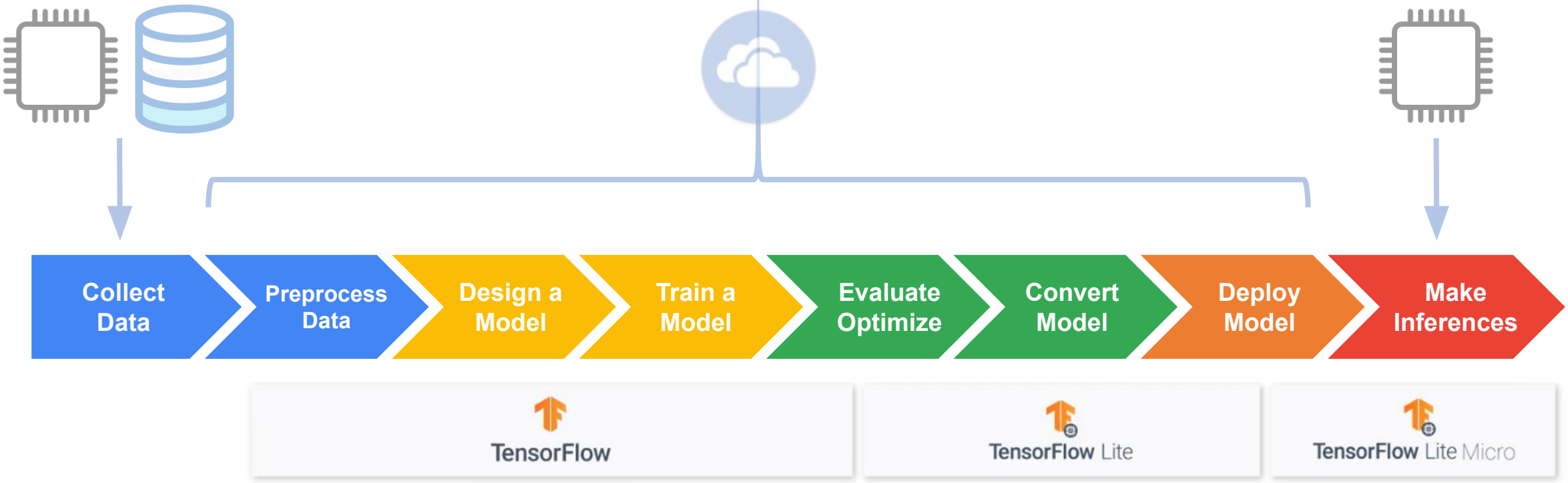
Tiny Machine Learning Workflow (“What”)



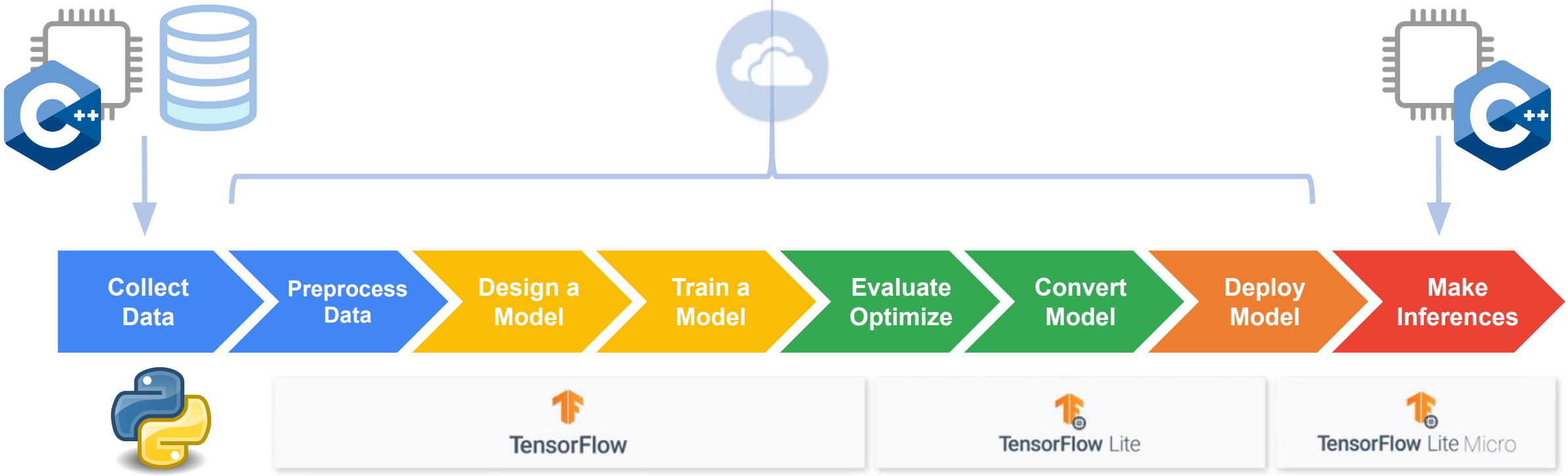
Tiny Machine Learning Workflow ("Where")



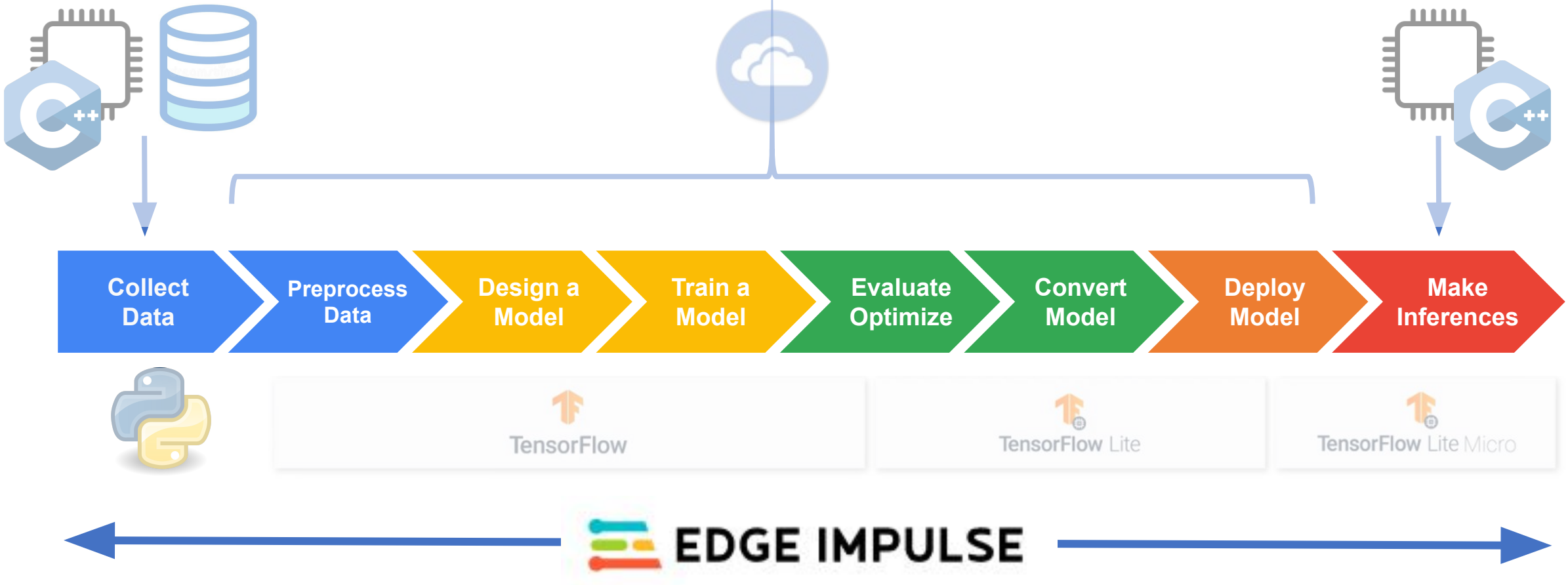
Machine Learning Workflow ("How")

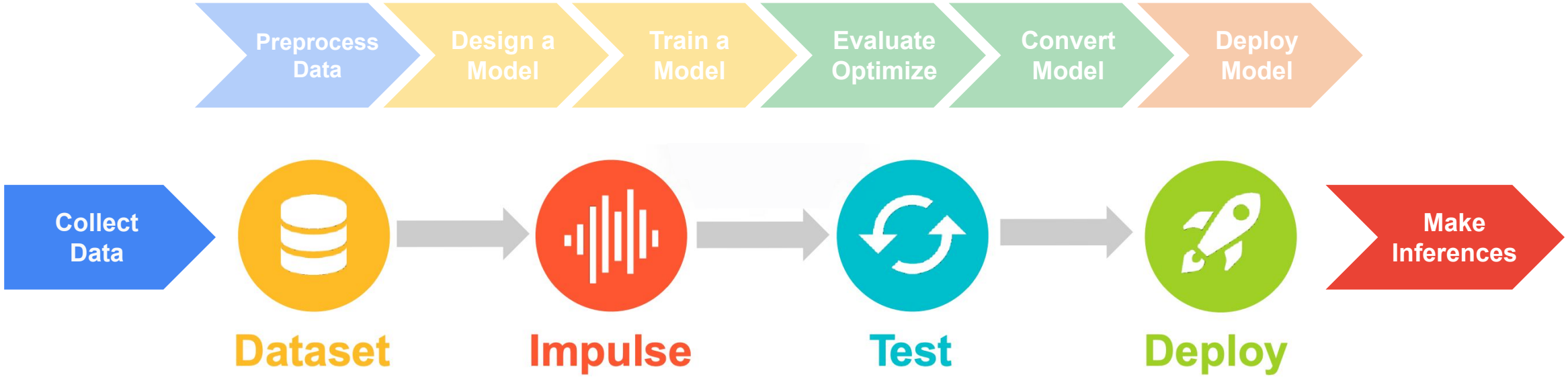


Machine Learning Workflow ("How")

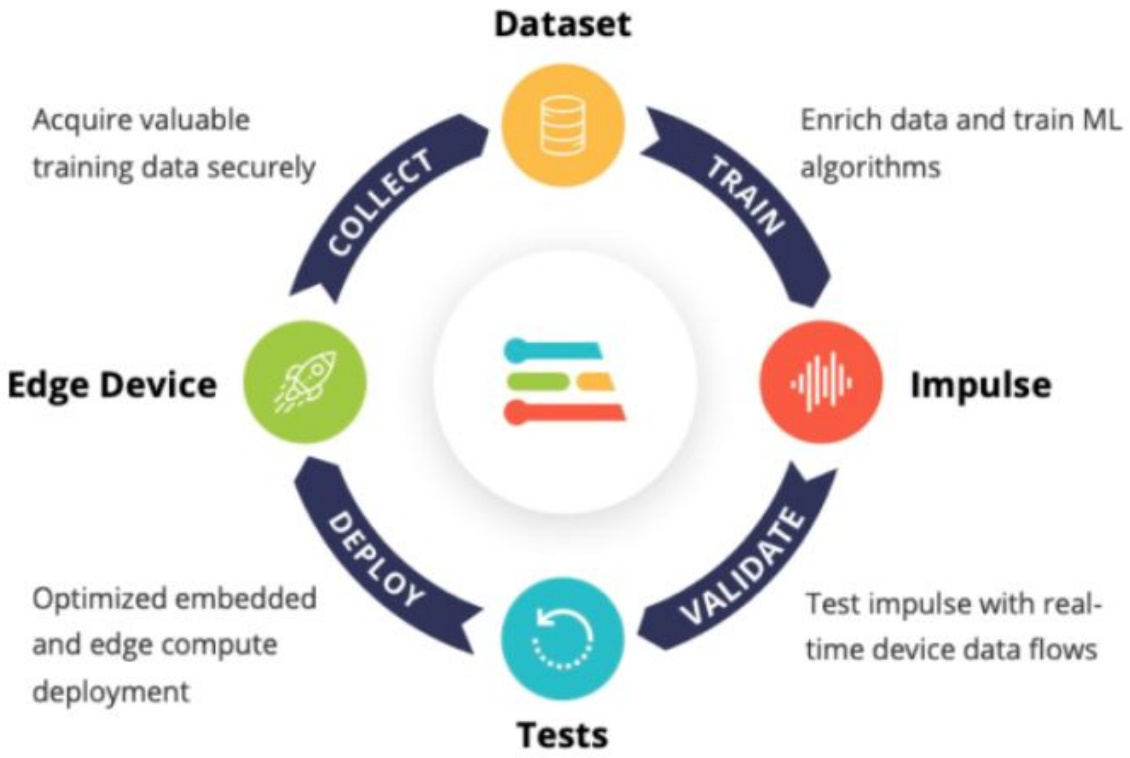


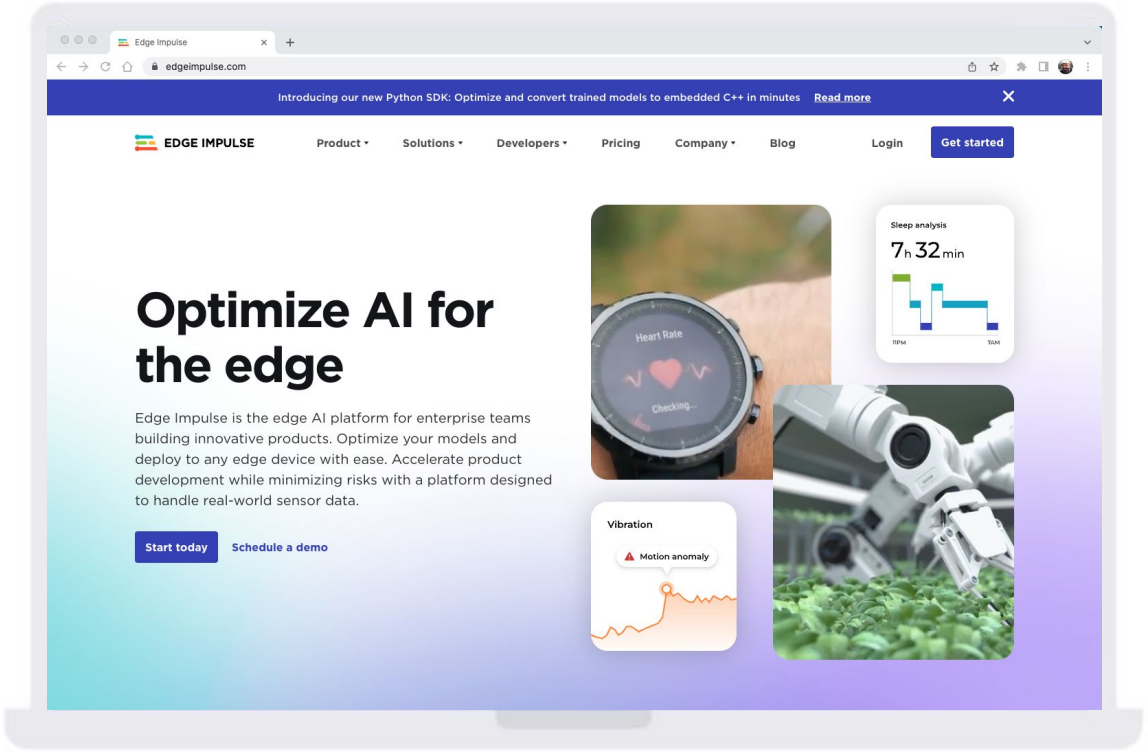
Machine Learning Workflow ("How")




Preprocess Data
Design a Model
Train a Model
Evaluate Optimize
Convert Model
Deploy Model
Collect Data
Dataset
Impulse
Test
Deploy
Make Inferences

EI Studio - Embedded ML platform ("AutoML")





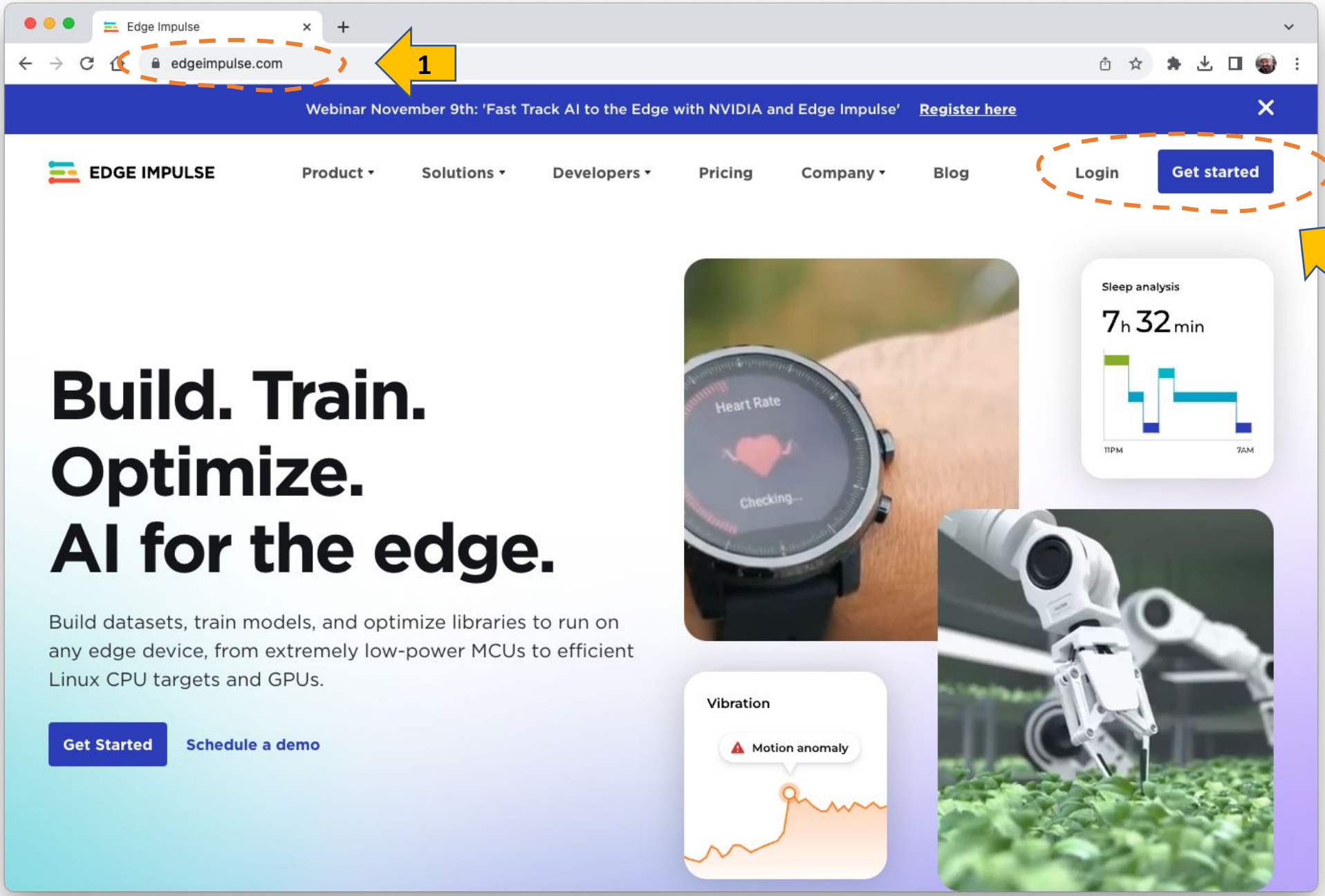
Learn more at http://edgeimpulse.com

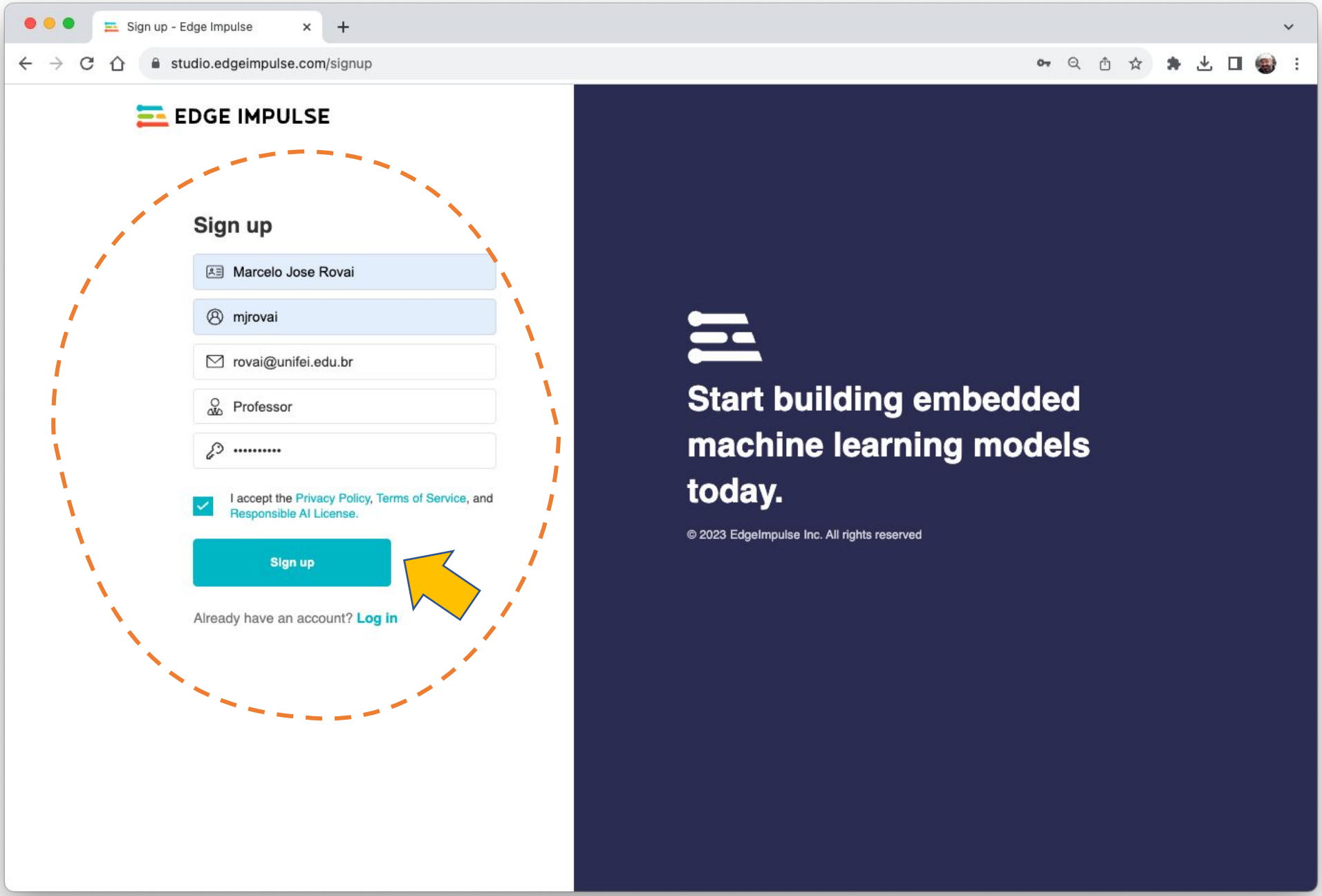
[2023 SciTinyML Workshop - Edge Impulse Overview by Shawn Hymel]

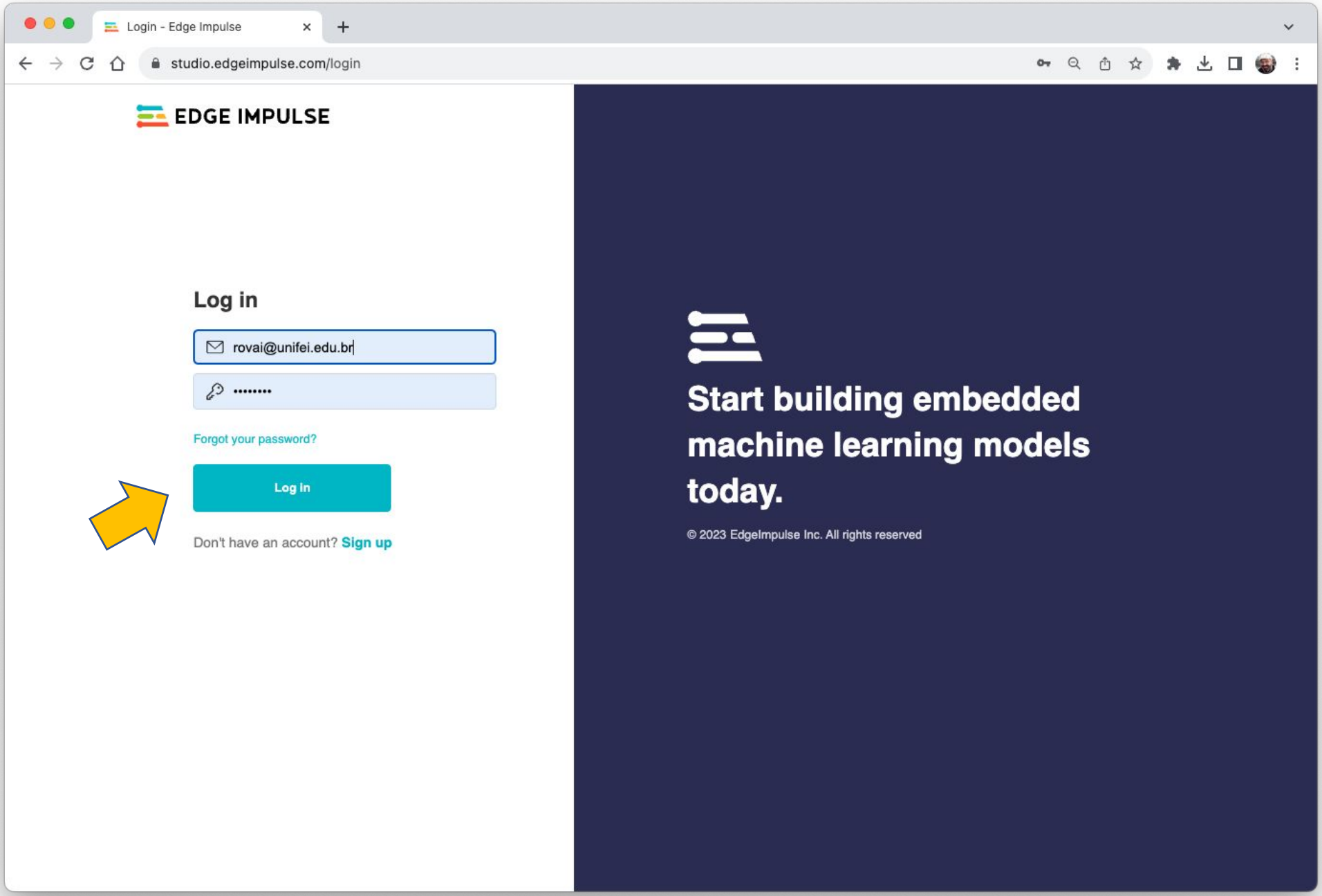
Cifar10
Edge Impulse Studio

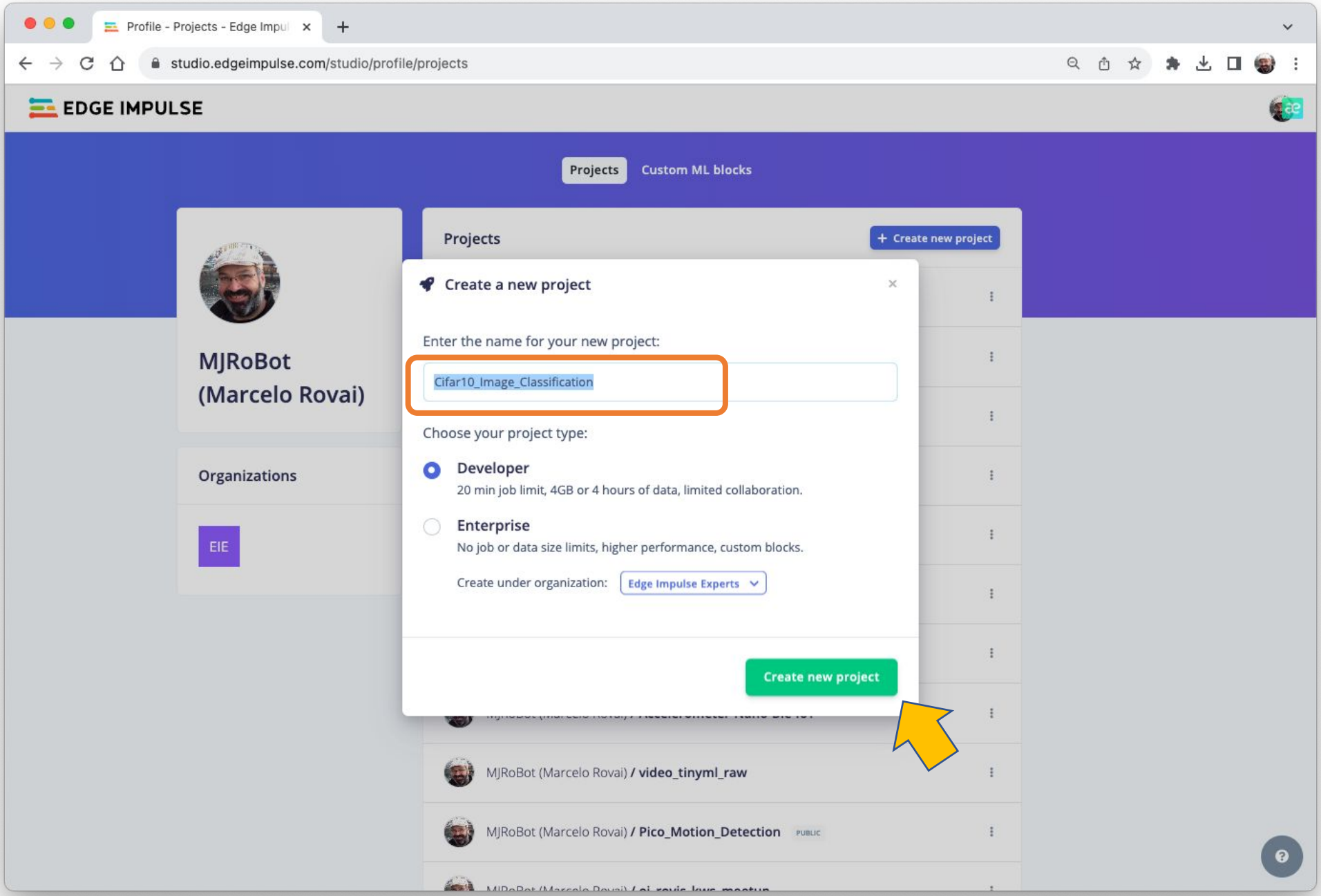
Dataset: https://github.com/YoongiKim/CIFAR-10-images

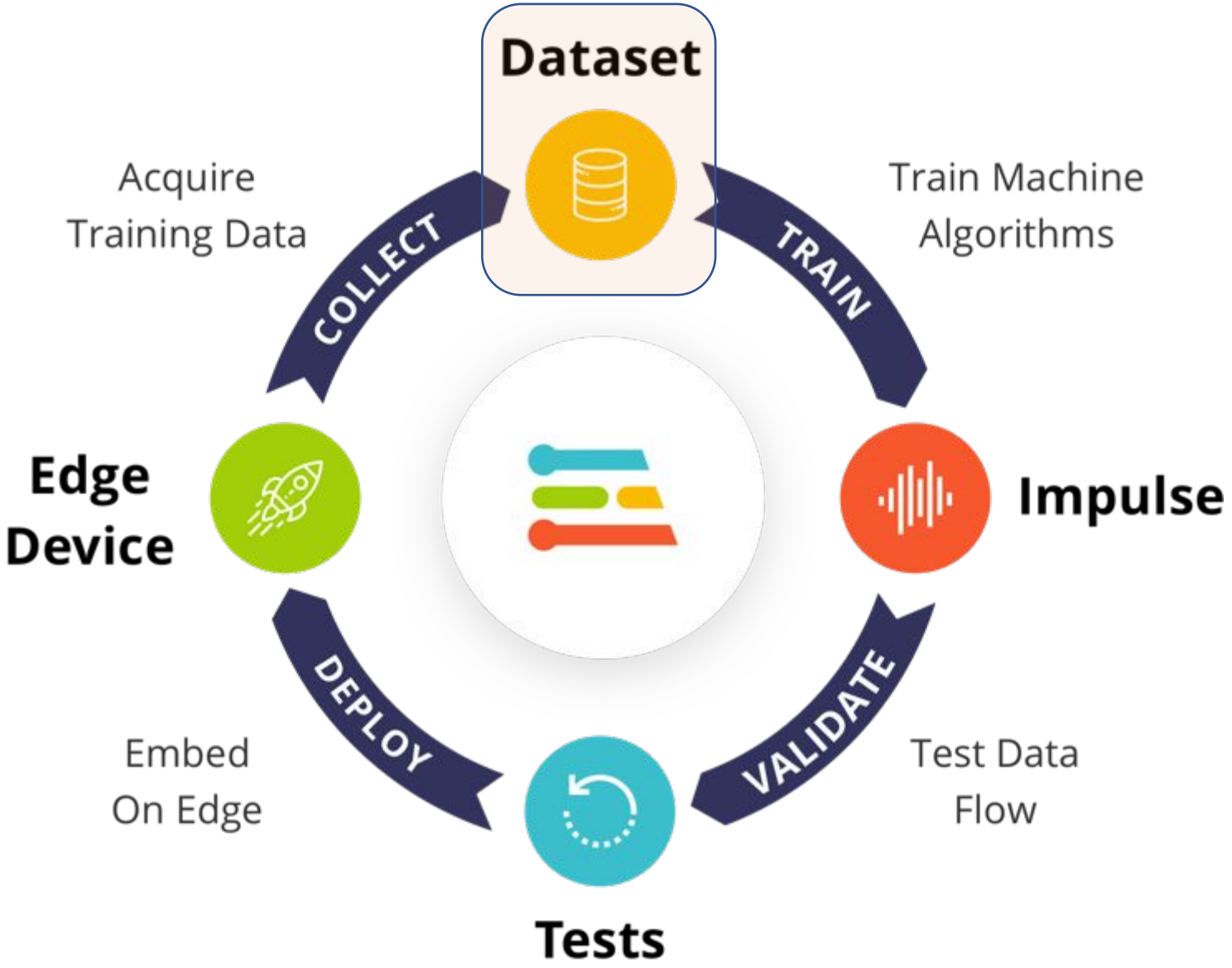
EI Studio Public Project: https://studio.edgeimpulse.com/public/51070/latest


Edge Impulse
edgeimpulse.com
1
Webinar November 9th: 'Fast Track AI to the Edge with NVIDIA and Edge Impulse' Register here
EDGE IMPULSE
Product
Solutions
Developers
Pricing
Company
Blog
Login
Get started
2
Build. Train.
Optimize.
AI for the edge.
Build datasets, train models, and optimize libraries to run on any edge device, from extremely low-power MCUs to efficient Linux CPU targets and GPUs.
Get Started
Schedule a demo
Heart Rate
Checking...
Sleep analysis
7h 32 min
11PM
7AM
Vibration
Motion anomaly


Sign up - Edge Impulse
studio.edgeimpulse.com/signup
EDGE IMPULSE
Sign up
Marcelo Jose Rovai
mjrovai
rovai@unifei.edu.br
Professor
I accept the Privacy Policy, Terms of Service, and Responsible AI License.
Sign up
Already have an account? Log in
Start building embedded machine learning models today.
© 2023 EdgeImpulse Inc. All rights reserved


Login - Edge Impulse
studio.edgeimpulse.com/login
EDGE IMPULSE
Log in
rovai@unifei.edu.br
Forgot your password?
Log In
Don't have an account? Sign up
Start building embedded machine learning models today.
© 2023 EdgeImpulse Inc. All rights reserved

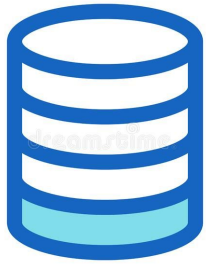

Profile - Projects - Edge Impul
studio.edgeimpulse.com/studio/profile/projects
EDGE IMPULSE
Projects
Custom ML blocks
MJRoBot
(Marcelo Rovai)
Organizations
EIE
Projects
+ Create new project
Create a new project
Enter the name for your new project:
Cifar10_Image_Classification
Choose your project type:
Developer
20 min job limit, 4GB or 4 hours of data, limited collaboration.
Enterprise
No job or data size limits, higher performance, custom blocks.
Create under organization:
Edge Impulse Experts
Create new project
MJRoBot (Marcelo Rovai) / video_tinyml_raw
MJRoBot (Marcelo Rovai) / Pico_Motion_Detection
PUBLIC

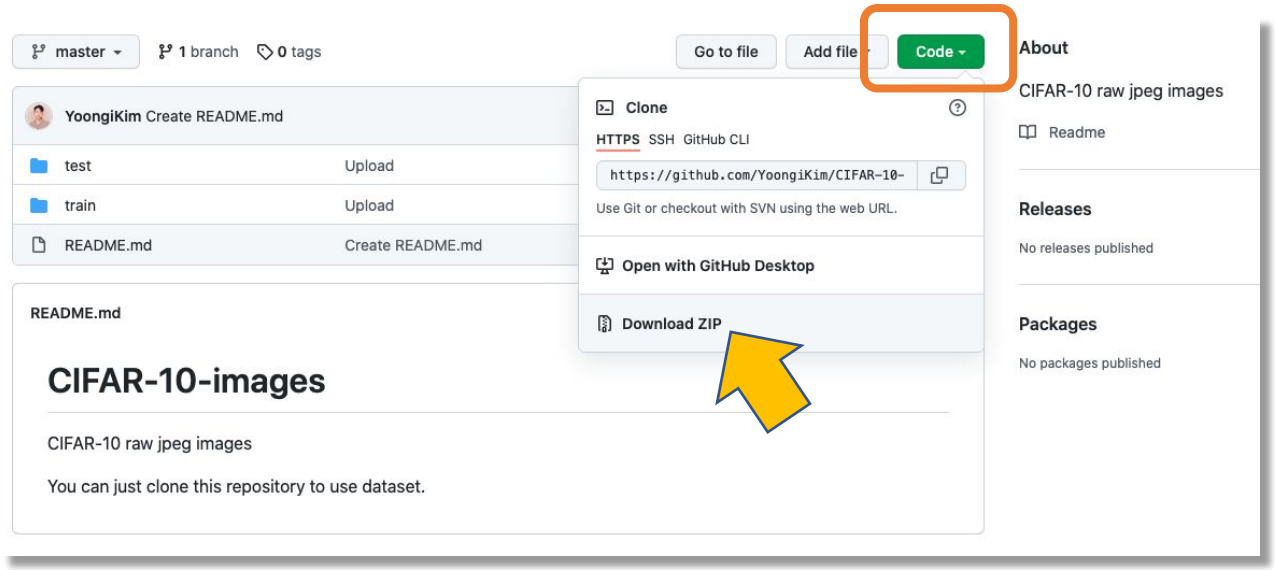


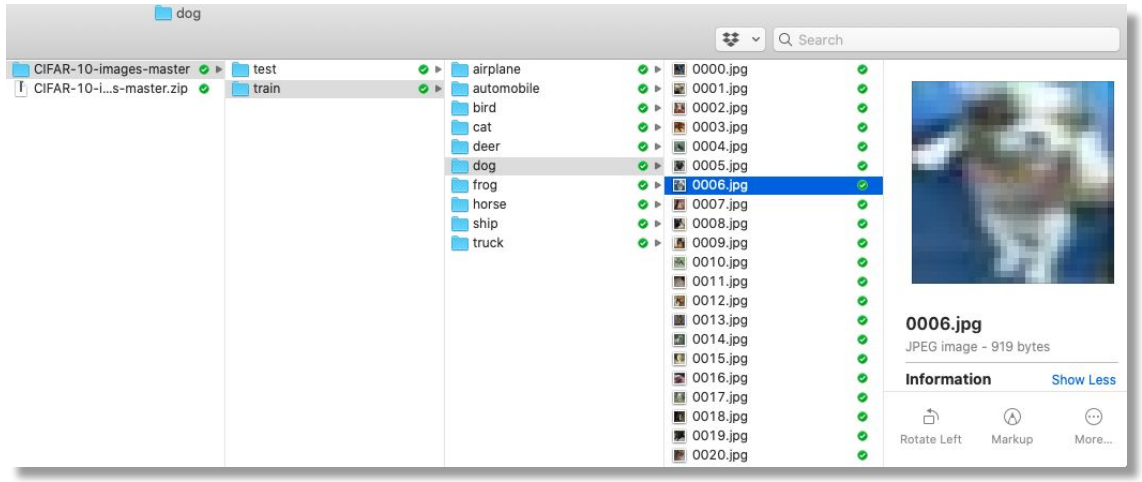
- Pre-Processing Data
- Design a Model
- Train a Model

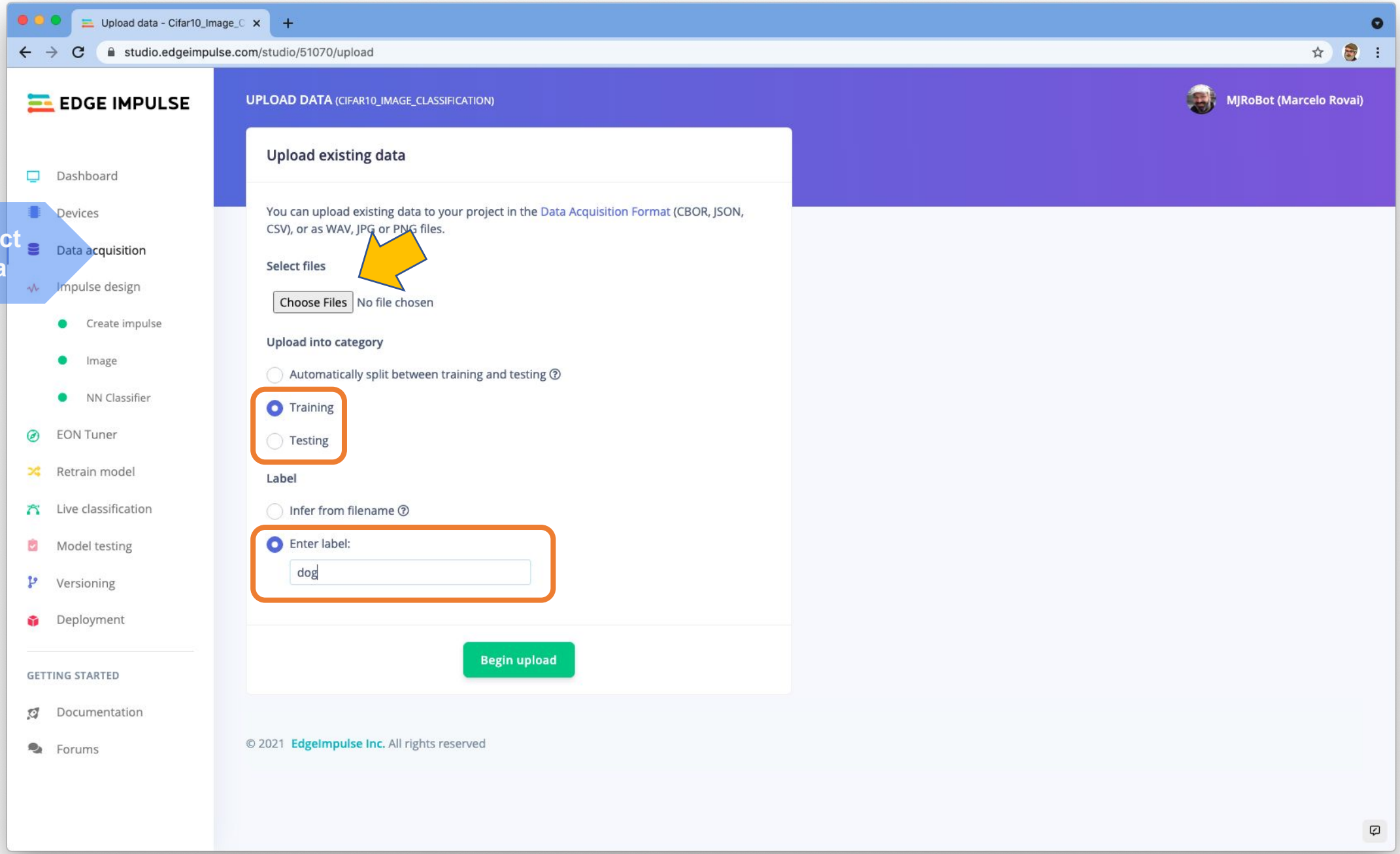
Download Dataset

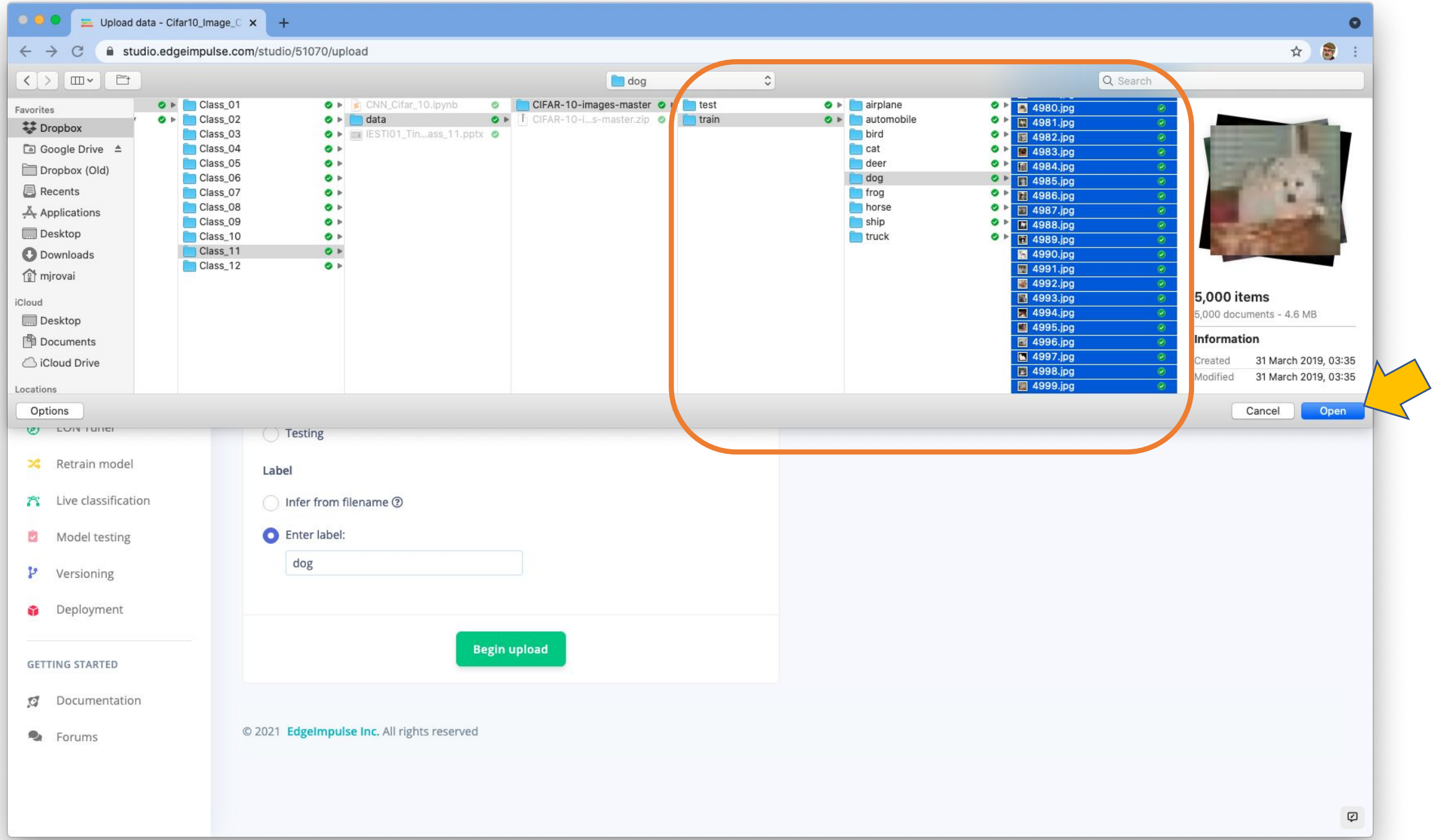
https://github.com/YoongiKim/CIFAR-10-images

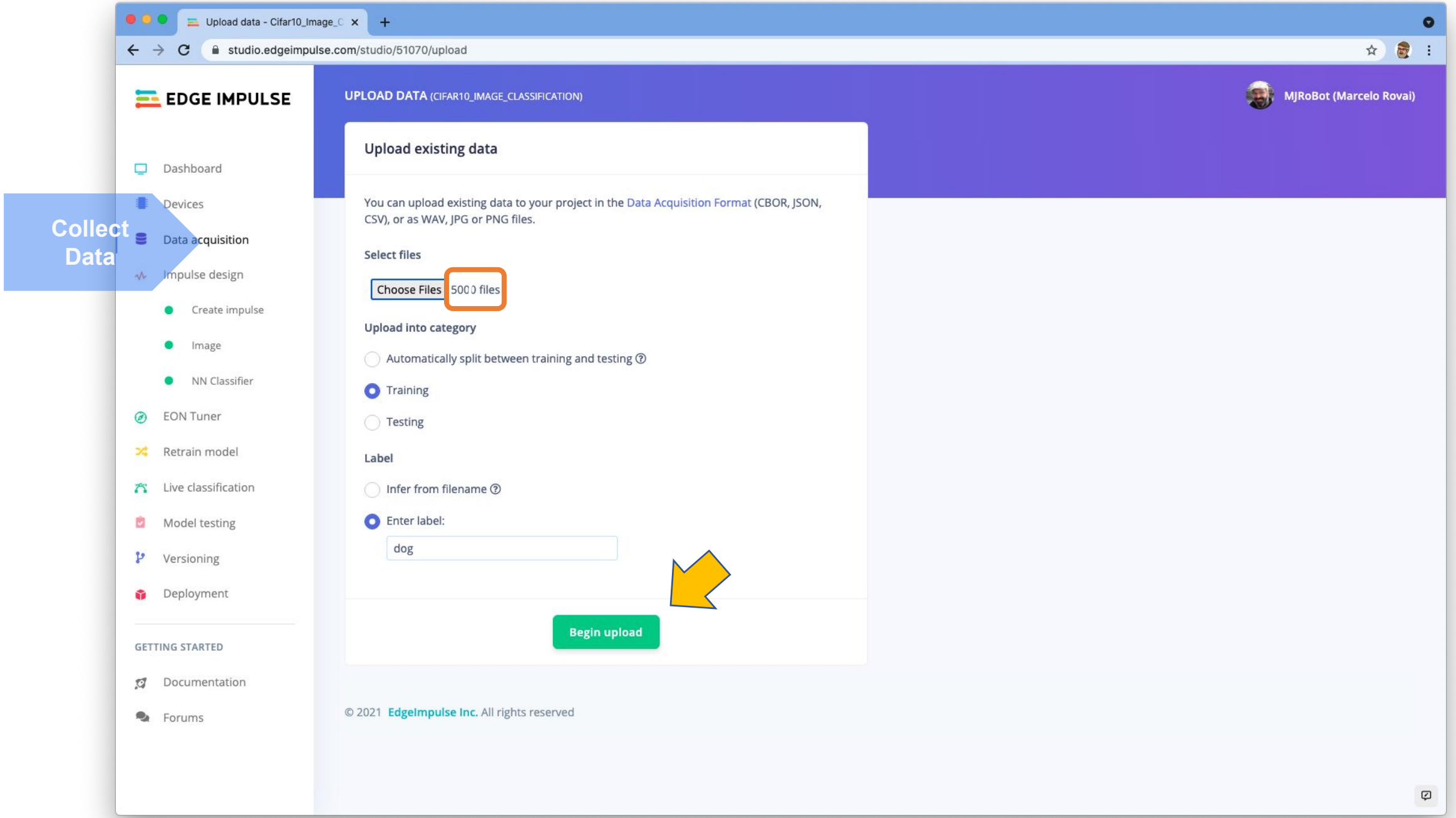




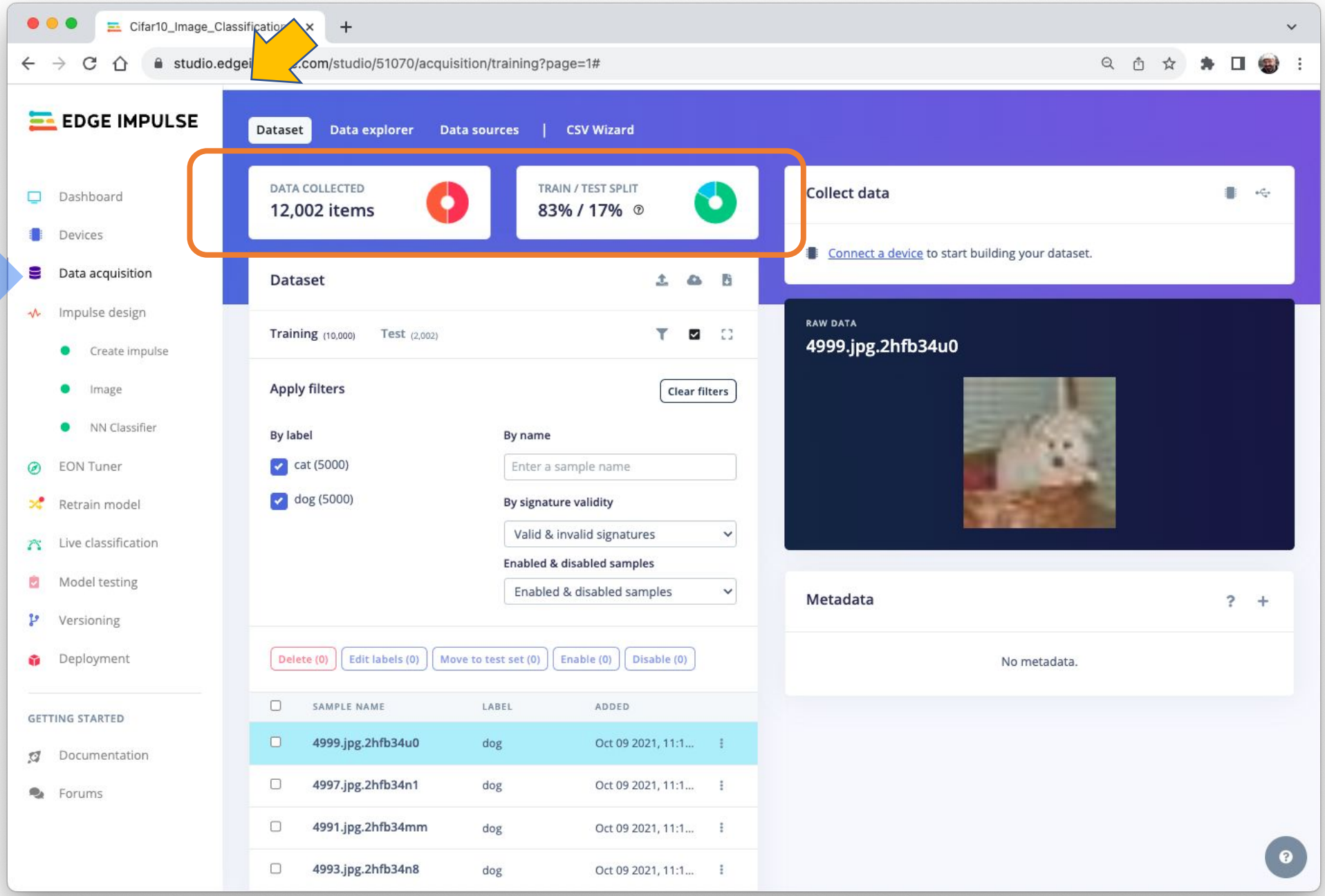


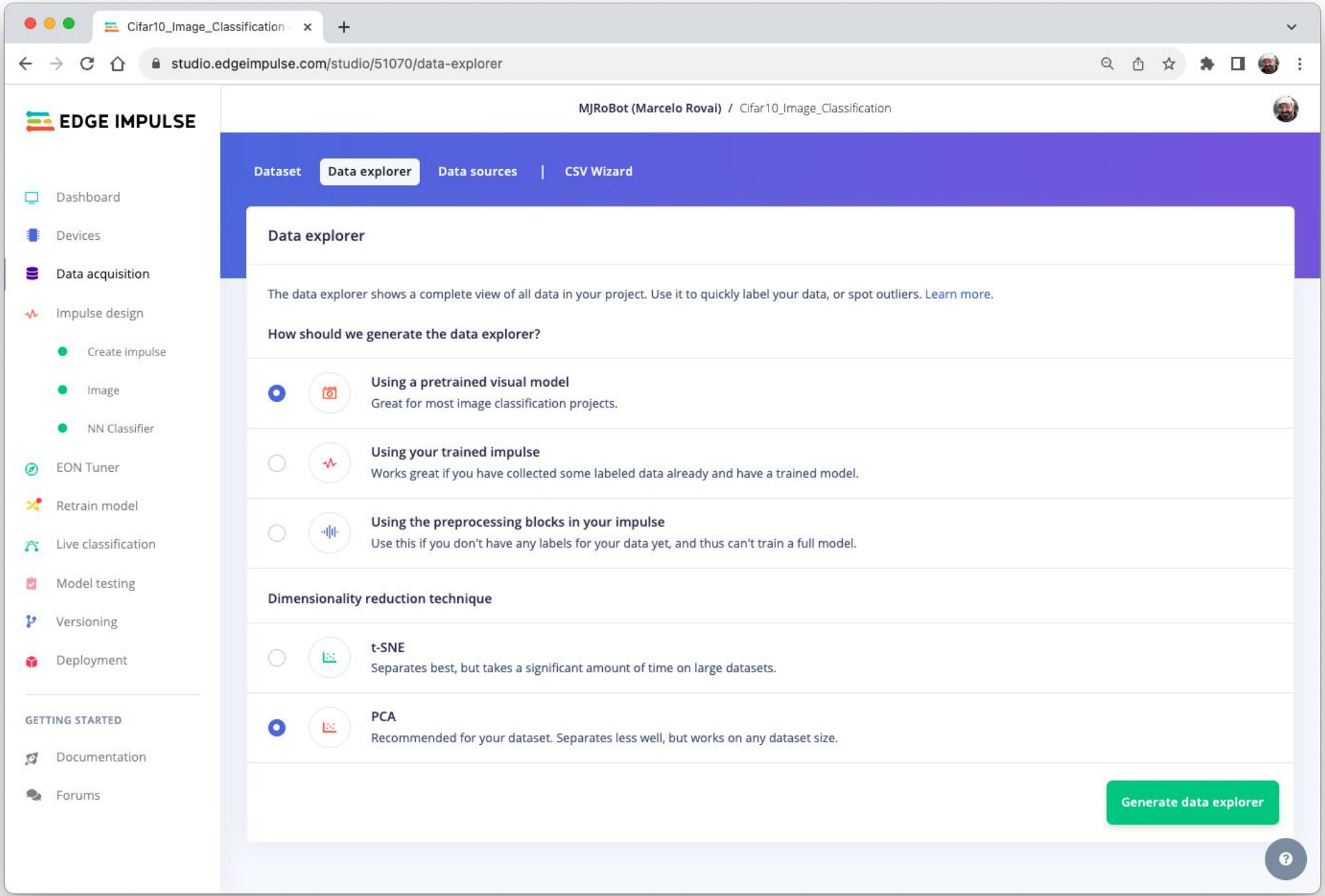

Upload data - Cifar10_Image_C
studio.edgeimpulse.com/studio/51070/upload
EDGE IMPULSE
UPLOAD DATA (CIFAR10_IMAGE_CLASSIFICATION)
MJRoBot (Marcelo Rovai)
Collect Data
Dashboard
Devices
Data acquisition
Impulse design
Create impulse
Image
NN Classifier
EON Tuner
Retrain model
Live classification
Model testing
Versioning
Deployment
GETTING STARTED
Documentation
Forums
Upload existing data
You can upload existing data to your project in the Data Acquisition Format (CBOR, JSON, CSV), or as WAV, JPG or PNG files.
Select files
Choose Files
No file chosen
Upload into category
Automatically split between training and testing
Training
Testing
Label
Infer from filename
Enter label:
dog
Begin upload
© 2021 EdgeImpulse Inc. All rights reserved

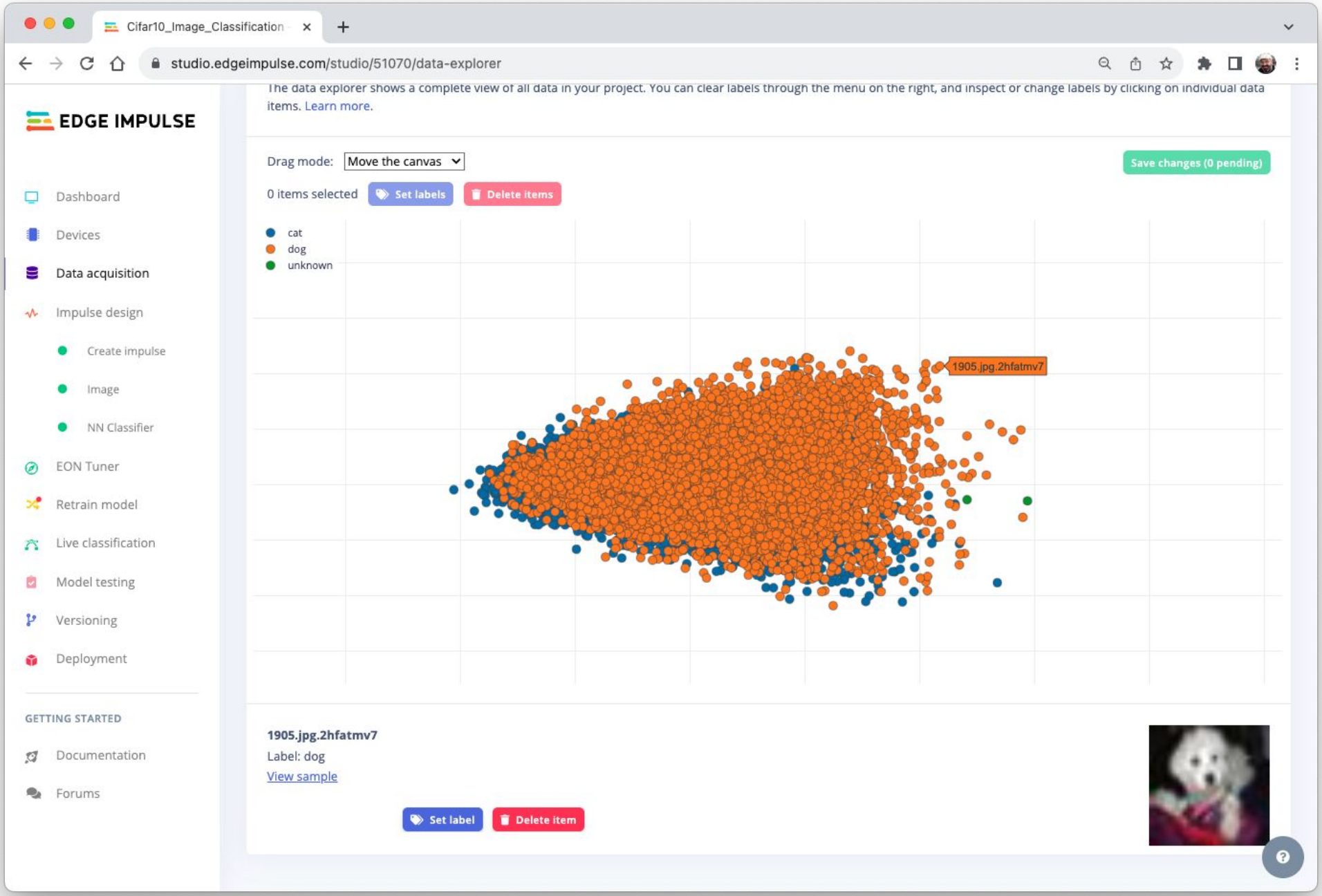

Upload data - Cifar10_Image_C
studio.edgeimpulse.com/studio/51070/upload
dog
Search
Favorites
Dropbox
Google Drive
Dropbox (Old)
Recents
Applications
Desktop
Downloads
mjrovai
iCloud
Desktop
Documents
iCloud Drive
Locations
Class_01
Class_02
Class_03
Class_04
Class_05
Class_06
Class_07
Class_08
Class_09
Class_10
Class_11
Class_12
CNN_Cifar_10.ipynb
data
IESTI01_Tin...ass_11.pptx
CIFAR-10-images-master
CIFAR-10-i...s-master.zip
test
train
airplane
automobile
bird
cat
deer
dog
frog
horse
ship
truck
4980.jpg
4981.jpg
4982.jpg
4983.jpg
4984.jpg
4985.jpg
4986.jpg
4987.jpg
4988.jpg
4989.jpg
4990.jpg
4991.jpg
4992.jpg
4993.jpg
4994.jpg
4995.jpg
4996.jpg
4997.jpg
4998.jpg
4999.jpg
5,000 items
5,000 documents - 4.6 MB
Information
Created 31 March 2019, 03:35
Modified 31 March 2019, 03:35
Options
Cancel
Open
Retrain model
Live classification
Model testing
Versioning
Deployment
GETTING STARTED
Documentation
Forums
Testing
Label
Infer from filename
Enter label:
dog
Begin upload
© 2021 EdgeImpulse Inc. All rights reserved

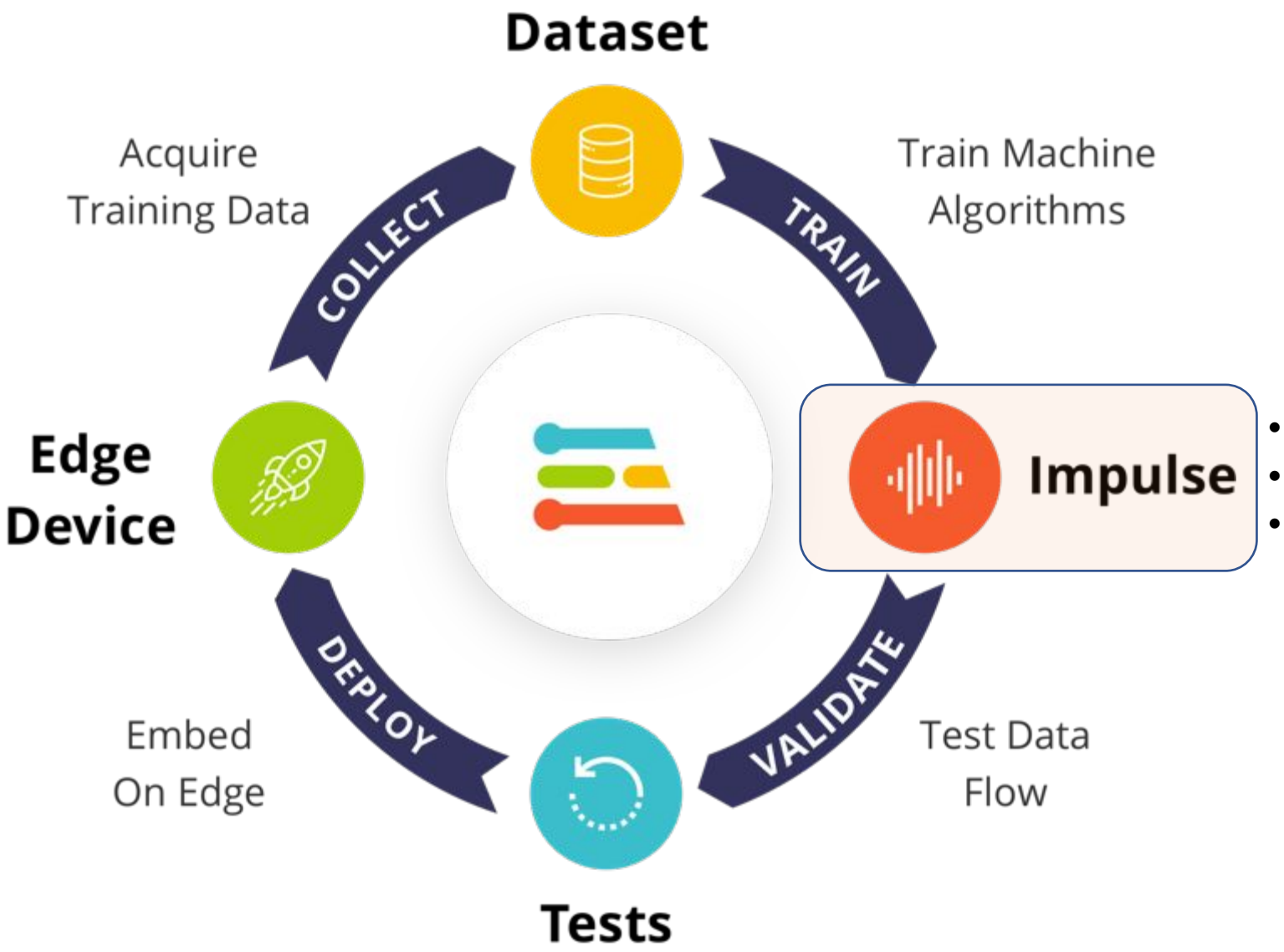

Upload data - Cifar10_Image_C
studio.edgeimpulse.com/studio/51070/upload
EDGE IMPULSE
UPLOAD DATA (CIFAR10_IMAGE_CLASSIFICATION)
MJRoBot (Marcelo Rovai)
Collect Data
Dashboard
Devices
Data acquisition
Impulse design
Create impulse
Image
NN Classifier
EON Tuner
Retrain model
Live classification
Model testing
Versioning
Deployment
GETTING STARTED
Documentation
Forums
Upload existing data
You can upload existing data to your project in the Data Acquisition Format (CBOR, JSON, CSV), or as WAV, JPG or PNG files.
Select files
Choose Files
5000 files
Upload into category
Automatically split between training and testing
Training
Testing
Label
Infer from filename
Enter label:
dog
Begin upload
© 2021 EdgeImpulse Inc. All rights reserved

Collect Data

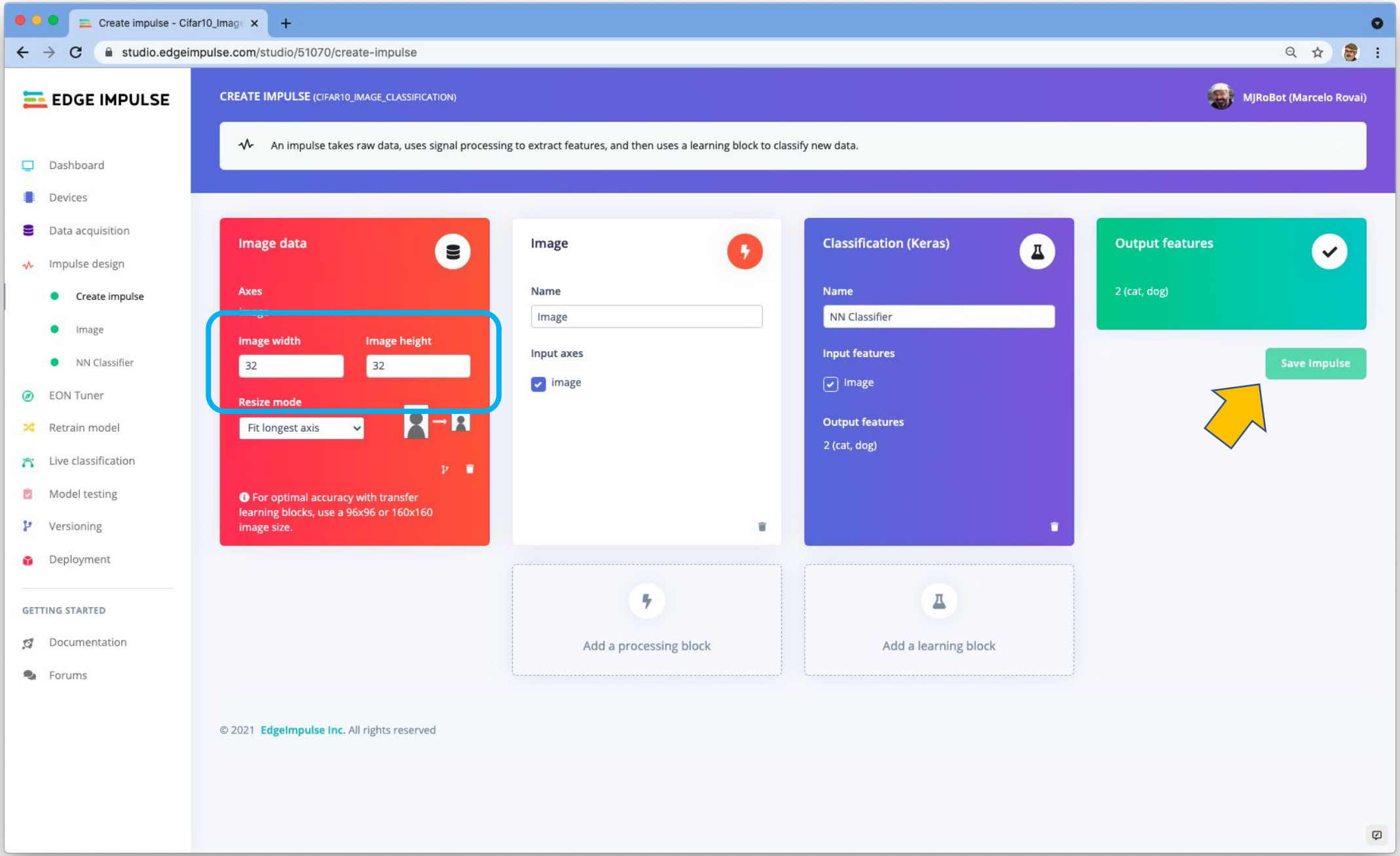


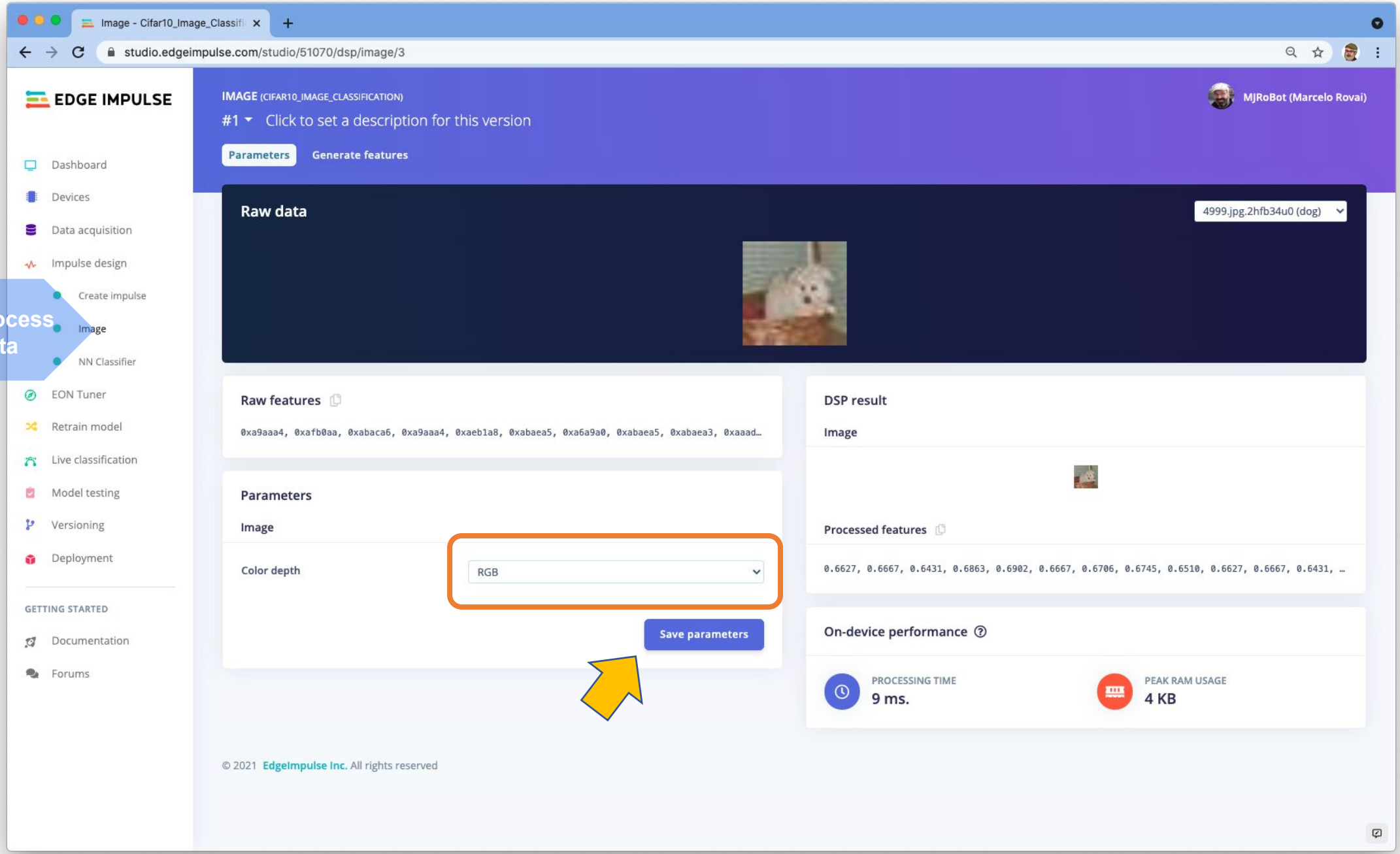

Cifar10_Image_Classification
studio.edgeimpulse.com/studio/51070/data-explorer
EDGE IMPULSE
MJRoBot (Marcelo Rovai) / Cifar10_Image_Classification
Dashboard
Devices
Data acquisition
Impulse design
Create impulse
Image
NN Classifier
EON Tuner
Retrain model
Live classification
Model testing
Versioning
Deployment
GETTING STARTED
Documentation
Forums
Dataset
Data explorer
Data sources
CSV Wizard
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 visual model
Great for most image classification projects.
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
Separates best, but takes a significant amount of time on large datasets.
PCA
Recommended for your dataset. Separates less well, but works on any dataset size.
Generate data explorer


Cifar10_Image_Classification
studio.edgeimpulse.com/studio/51070/data-explorer
EDGE IMPULSE
Dashboard
Devices
Data acquisition
Impulse design
Create impulse
Image
NN Classifier
EON Tuner
Retrain model
Live classification
Model testing
Versioning
Deployment
GETTING STARTED
Documentation
Forums
The data explorer shows a complete view of all data in your project. You can clear labels through the menu on the right, and inspect or change labels by clicking on individual data items. Learn more.
Drag mode: Move the canvas
Save changes (0 pending)
0 items selected
Set labels
Delete items
cat
dog
unknown
1905.jpg.2hfatmv7
1905.jpg.2hfatmv7
Label: dog
View sample
Set label
Delete item



- Pre-Processing Data
- Design a Model
- Train a Model


Create impulse - Cifar10_Imag
studio.edgeimpulse.com/studio/51070/create-impulse
EDGE IMPULSE
CREATE IMPULSE (CIFAR10_IMAGE_CLASSIFICATION)
MJRoBot (Marcelo Rovai)
An impulse takes raw data, uses signal processing to extract features, and then uses a learning block to classify new data.
Dashboard
Devices
Data acquisition
Impulse design
Create impulse
Image
NN Classifier
EON Tuner
Retrain model
Live classification
Model testing
Versioning
Deployment
GETTING STARTED
Documentation
Forums
Image data
Axes
Image width
32
Image height
32
Resize mode
Fit longest axis
For optimal accuracy with transfer learning blocks, use a 96x96 or 160x160 image size.
Image
Name
Image
Input axes
image
Classification (Keras)
Name
NN Classifier
Input features
Image
Output features
2 (cat, dog)
Output features
2 (cat, dog)
Save Impulse
Add a processing block
Add a learning block
© 2021 EdgeImpulse Inc. All rights reserved



Preprocess Data

Preprocess Data

Image - Cifar10_Image_Classifi

studio.edgeimpulse.com/studio/51070/dsp/image/3/generate-features

EDGE IMPULSE

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

GETTING STARTED

- Documentation
- Forums

IMAGE (CIFAR10_IMAGE_CLASSIFICATION)

#1 Click to set a description for this version

MJRoBot (Marcelo Rovai)

Parameters | Generate features

Training set

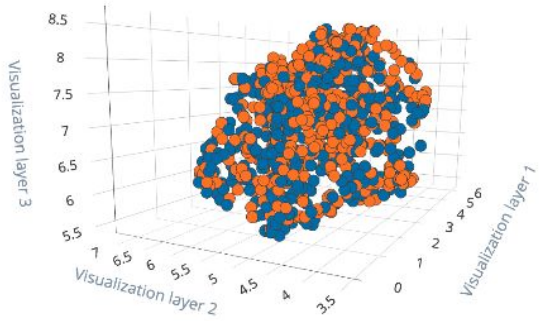
Data in training set	10,000 items
Classes	2 (cat, dog)

Generate features

Feature explorer (10,000 samples)

X Axis: Visualization layer 1 | Y Axis: Visualization layer 2 | Z Axis: Visualization layer 3

cat
dog



On-device performance

PROCESSING TIME
9 ms.

PEAK RAM USAGE
4 KB

colab.research.google.com

CNN_Cifar_10.ipynb

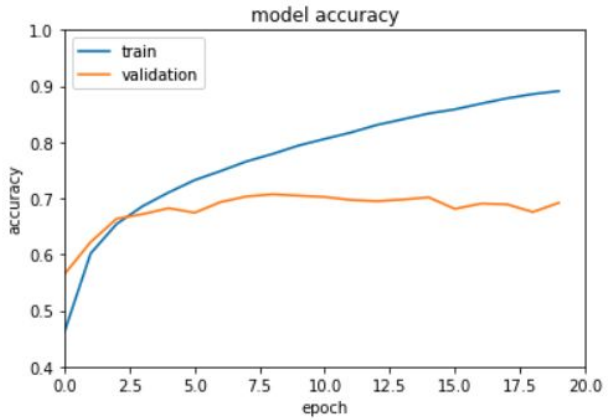
File Edit View Insert Runtime Tools Help

Share

+ Code + Text Copy to Drive

Connect GPU

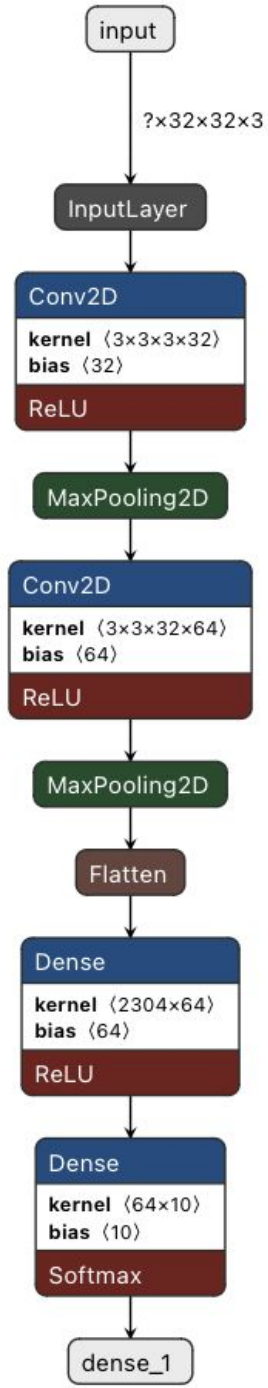
```
10 plt.show()
```

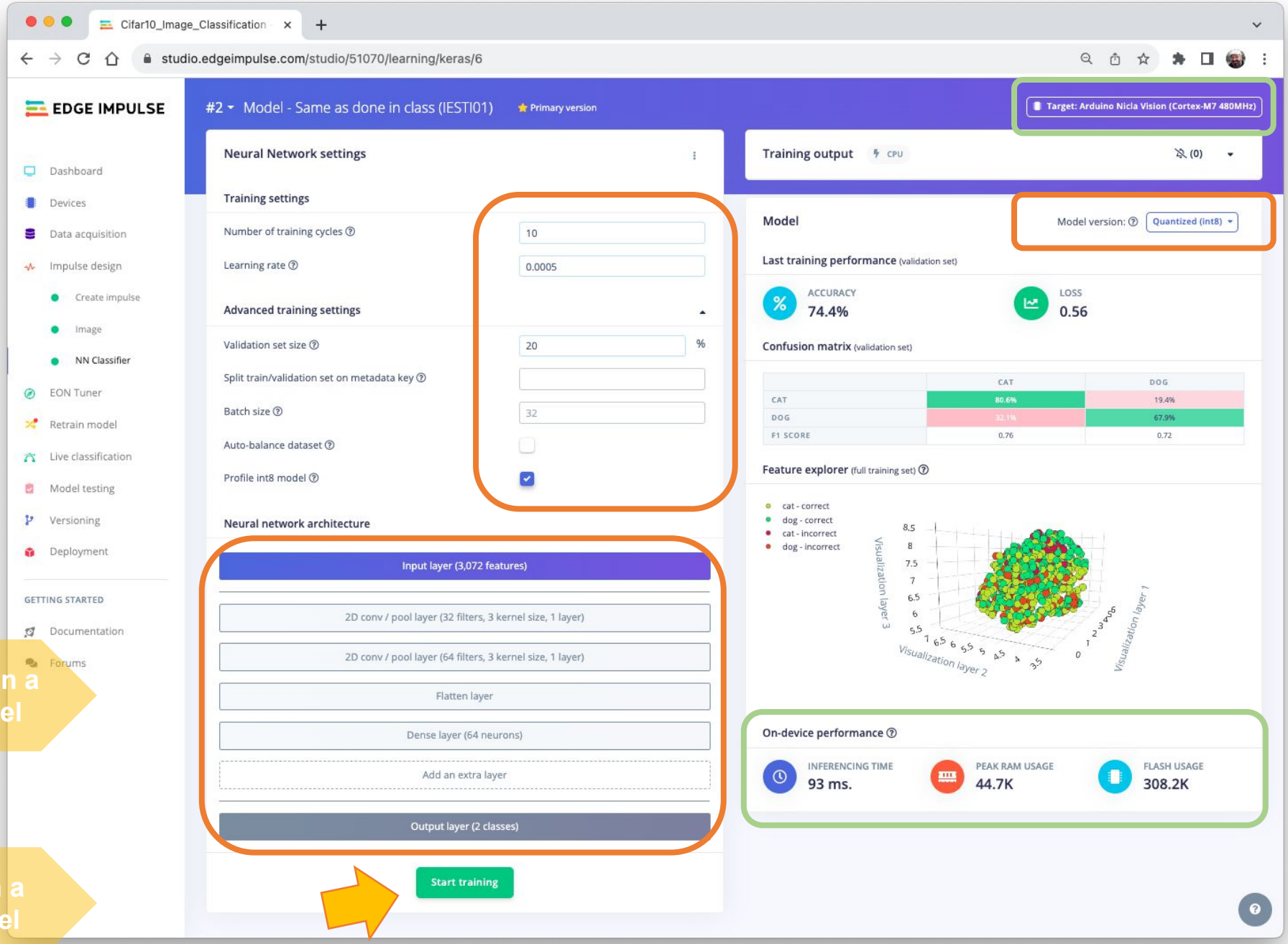


Evaluate Model

```
[ ]  1 print("Train data: ", model.evaluate(train_images, train_labels)[1])
     2 print("Eval data: ", model.evaluate(val_images, val_labels)[1])
     3 print("Test data: ", model.evaluate(test_images, test_labels)[1])
```

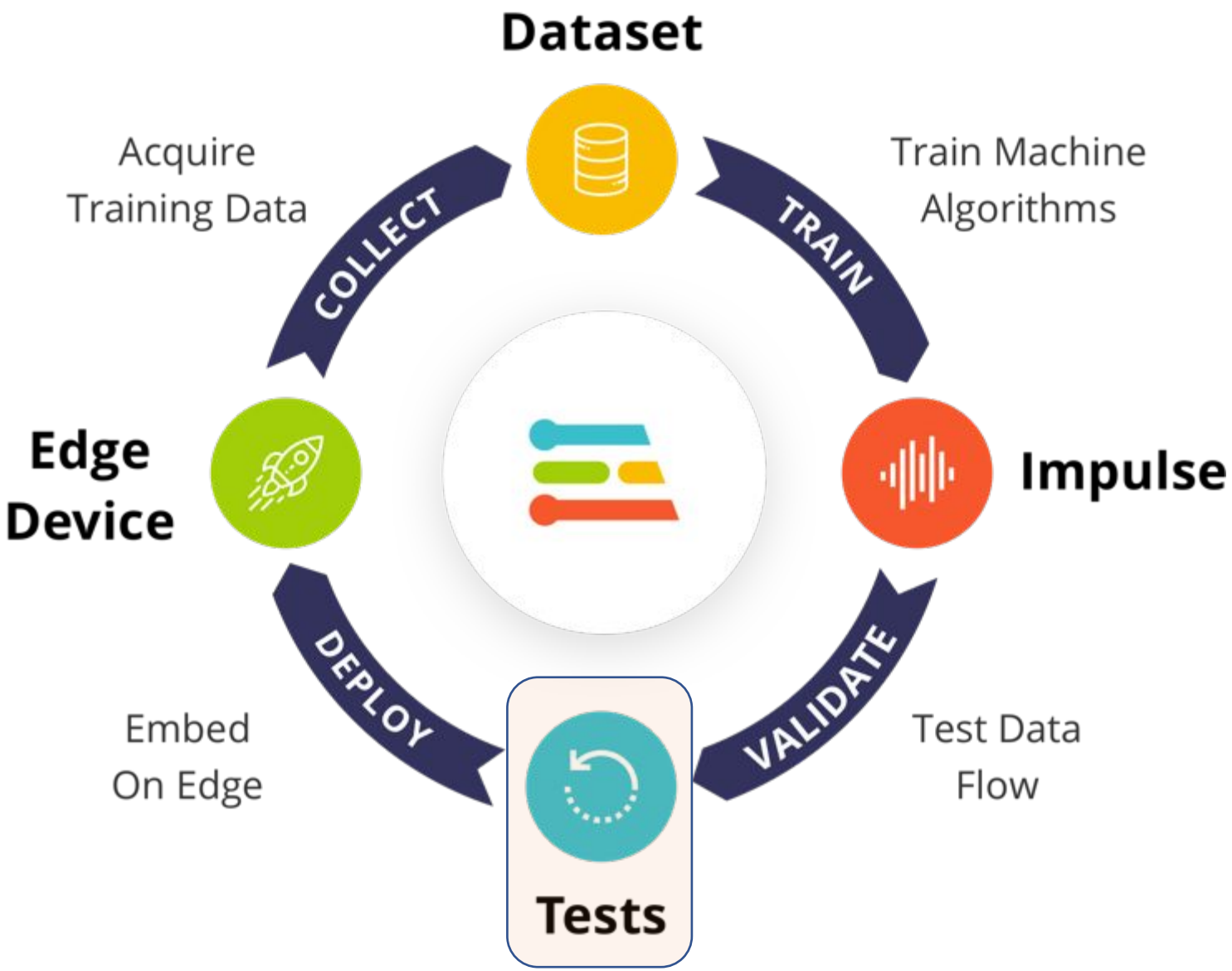
```
1407/1407 [==============================] - 6s 4ms/step - loss: 0.2313 - accuracy: 0.9199
Train data:  0.9198889136314392
157/157 [==============================] - 1s 4ms/step - loss: 1.2736 - accuracy: 0.6920
Eval data:  0.6919999718666077
313/313 [==============================] - 1s 4ms/step - loss: 1.3474 - accuracy: 0.6807
Test data:  0.6807000041007996
```





Design a Model

Train a Model



- Pre-Processing Data
- Design a Model
- Train a Model

Model testing - Cifar10_Image | Upload data - Cifar10 Classific

studio.edgeimpulse.com/studio/51070/validation

EDGE IMPULSE

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

GETTING STARTED

- Documentation
- Forums

Evaluate
Optimize

Test data

Classify all

Set the 'expected outcome' for each sample to the desired outcome to au... ...core the impulse.

SAMPLE NAME	EXPECTED OUTCOME	LENGTH	ACCURACY	RESULT
testing.2hfe5uat	testing	-		1 cat
0999.jpg.2hfb7a...	cat	-	100%	1 cat
0997.jpg.2hfb7a...	cat	-	100%	1 cat
0998.jpg.2hfb7a...	cat	-	100%	1 cat
0996.jpg.2hfb7a...	cat	-	0%	1 uncertain
0993.jpg.2hfb7a...	cat	-	100%	1 cat
0995.jpg.2hfb7a...	cat	-	0%	1 dog
0991.jpg.2hfb7a...	cat	-	100%	1 cat
0994.jpg.2hfb7a...	cat	-	100%	1 cat
0992.jpg.2hfb7a...	cat	-	100%	1 cat
0990.jpg.2hfb79...	cat	-	100%	1 cat
0988.jpg.2hfb79...	cat	-	100%	1 cat
0985.jpg.2hfb79...	cat	-	0%	1 uncertain

Model testing output

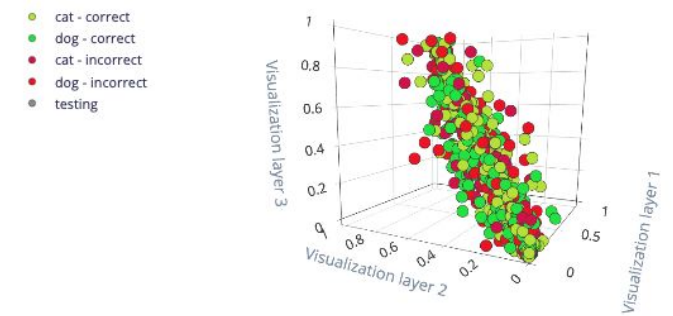
```
Classifying data for NN Classifier...
Copying features from processing blocks...
Copying features from DSP block...
Copying features from DSP block OK
Copying features from processing blocks OK

Classifying data for float32 model...
Scheduling job in cluster...
Job started
Classifying data for NN Classifier OK


Job completed
```

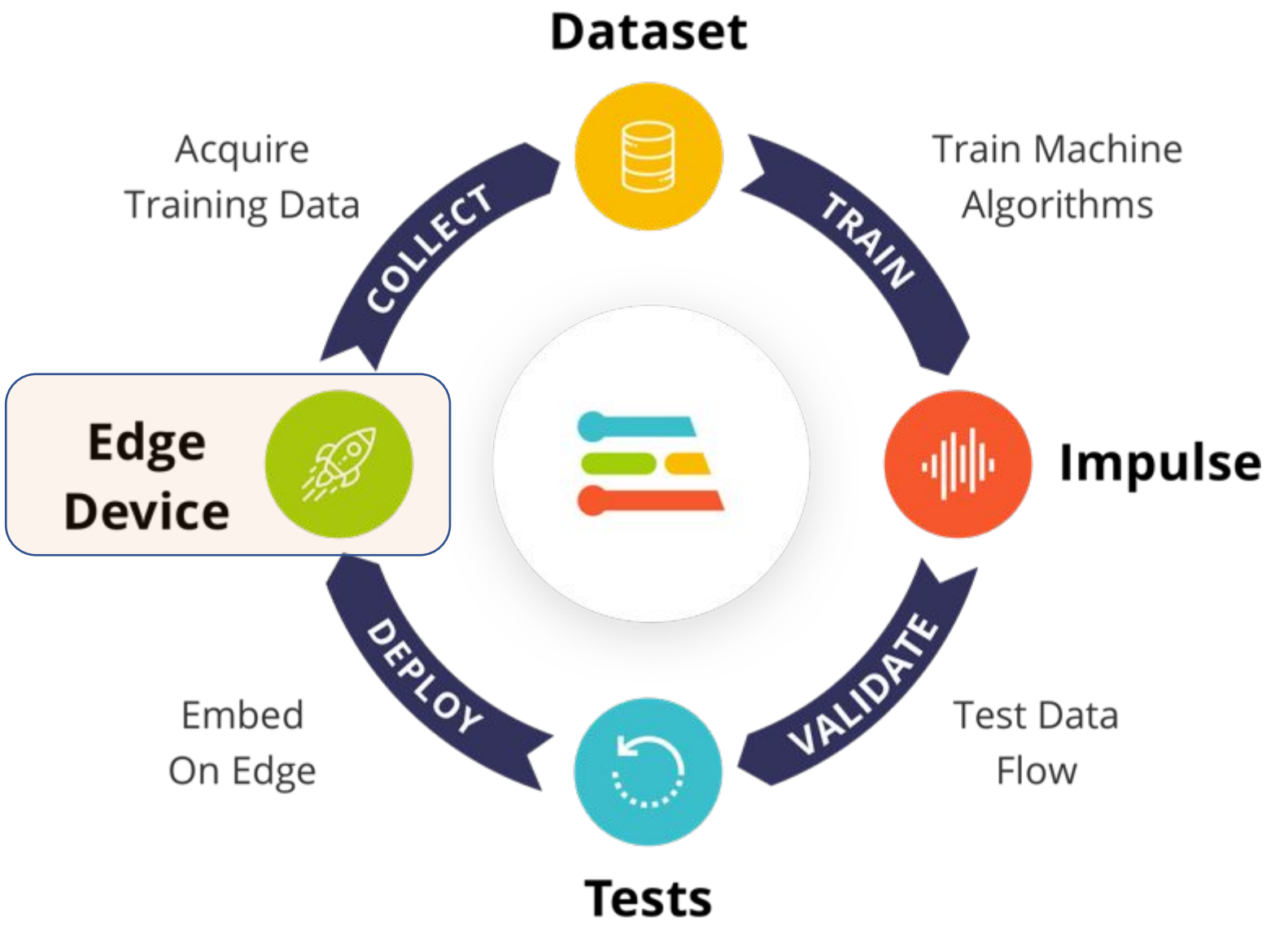
Model testing results

ACCURACY
64.95%

	CAT	DOG	UNCERTAIN
CAT	67.7%	16%	16.3%
DOG	21.2%	62.2%	16.6%
F1 SCORE	0.72	0.70	

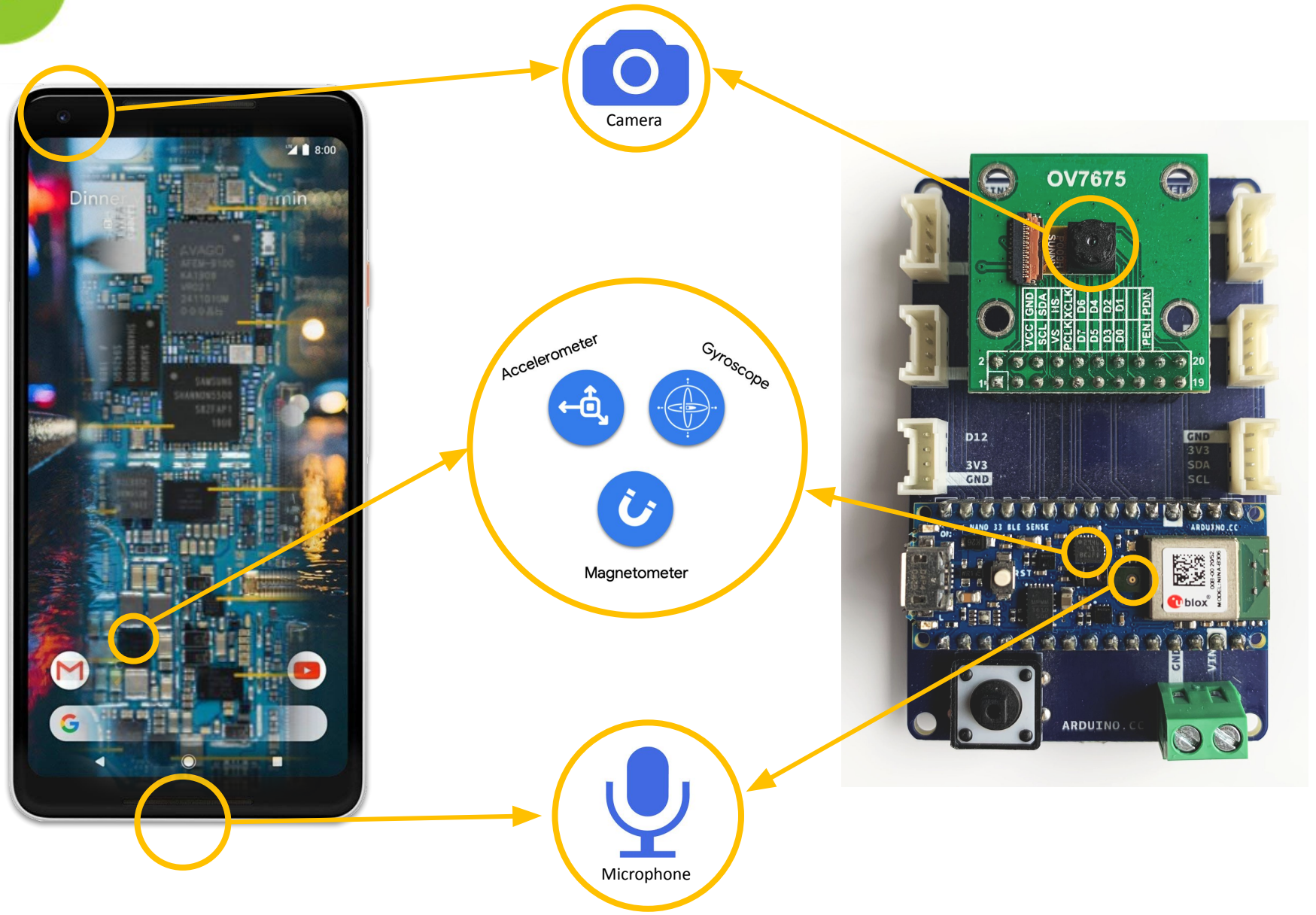
Feature explorer



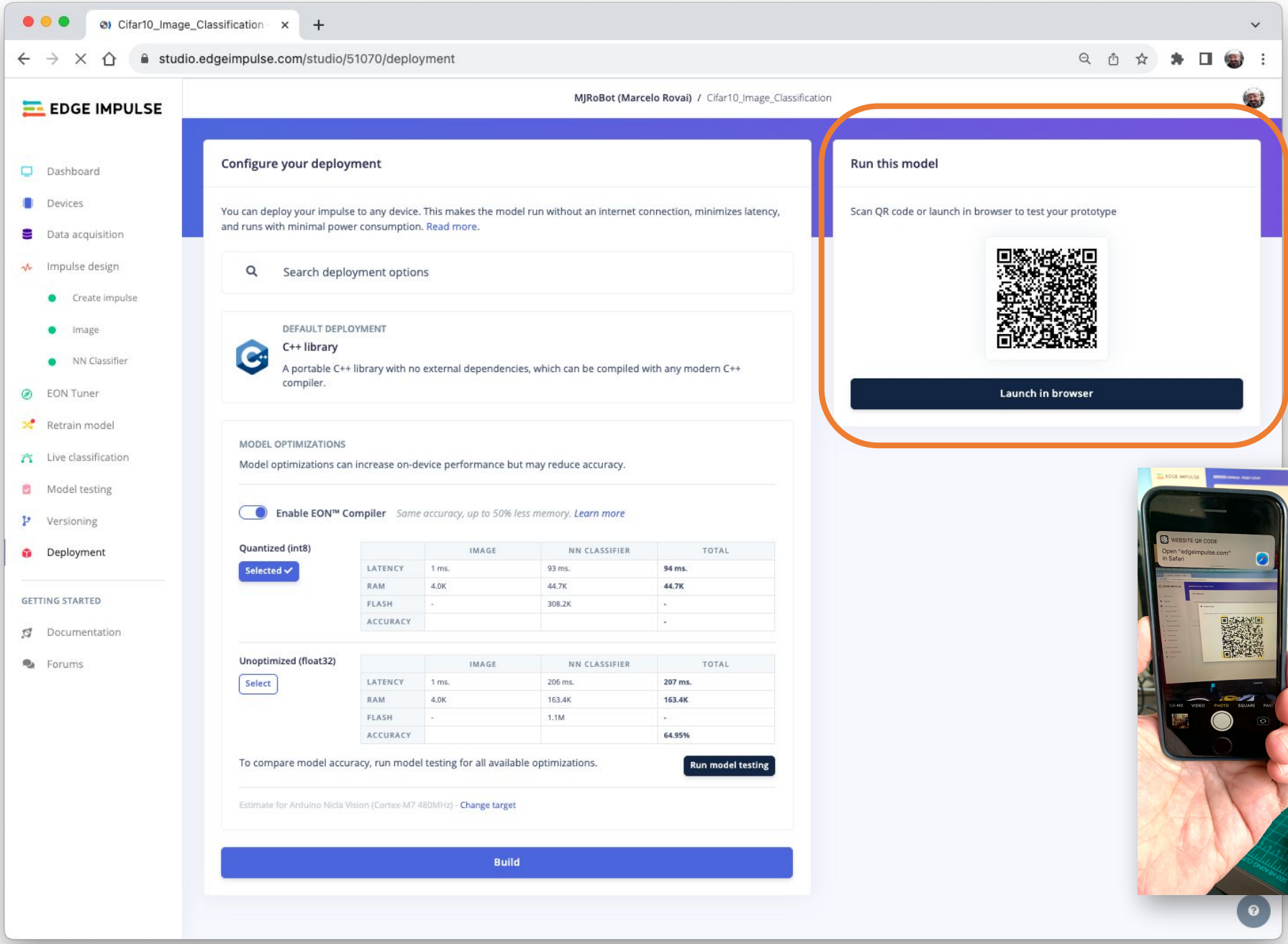


- Pre-Processing Data
- Design a Model
- Train a Model

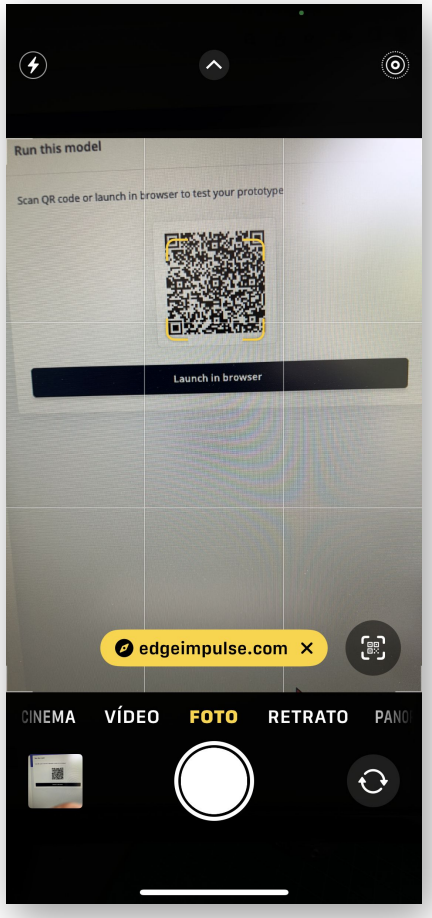
Edge Device & Sensors

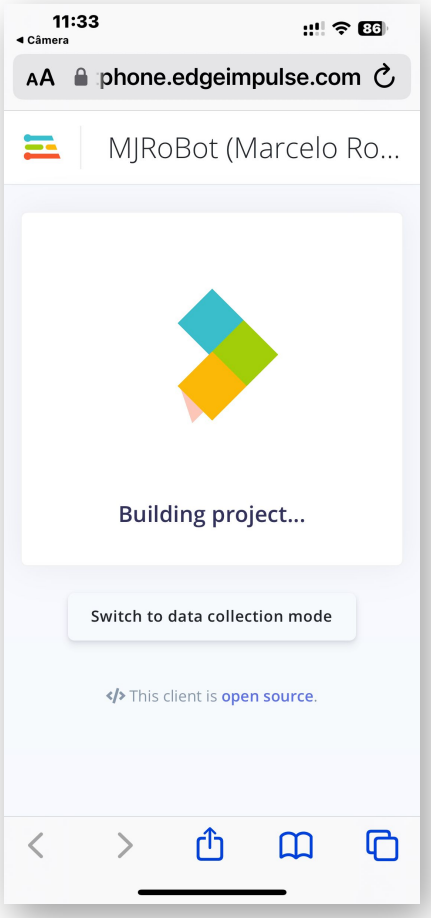


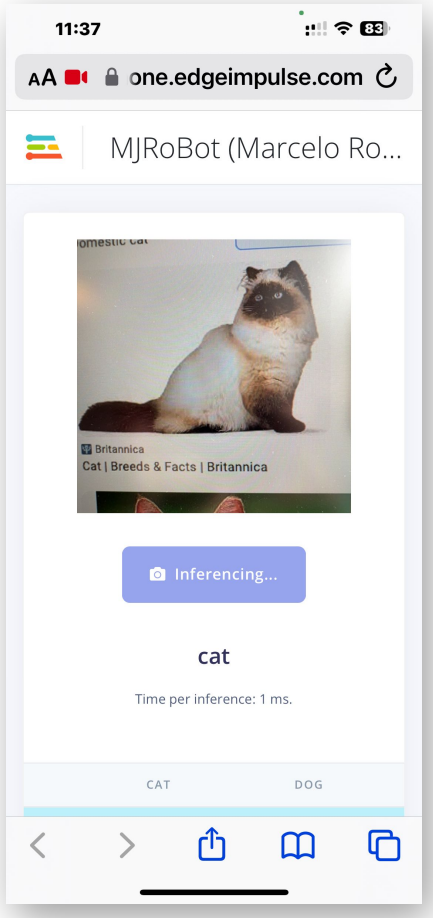
Deploy Model

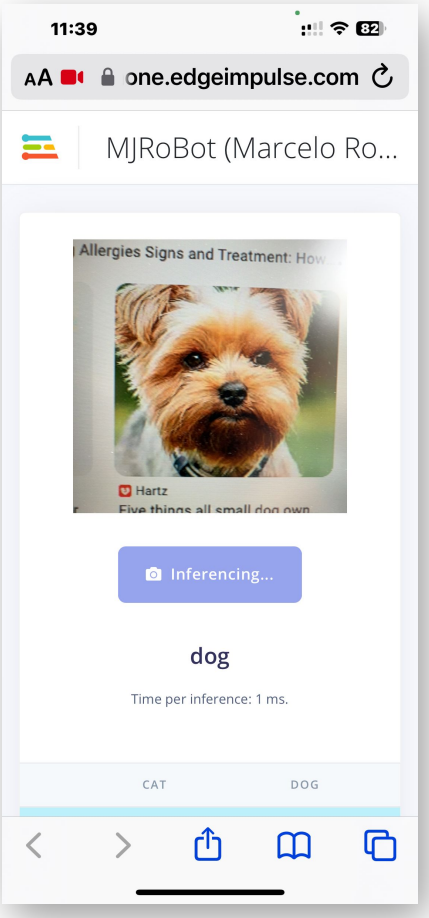


Make Inferences





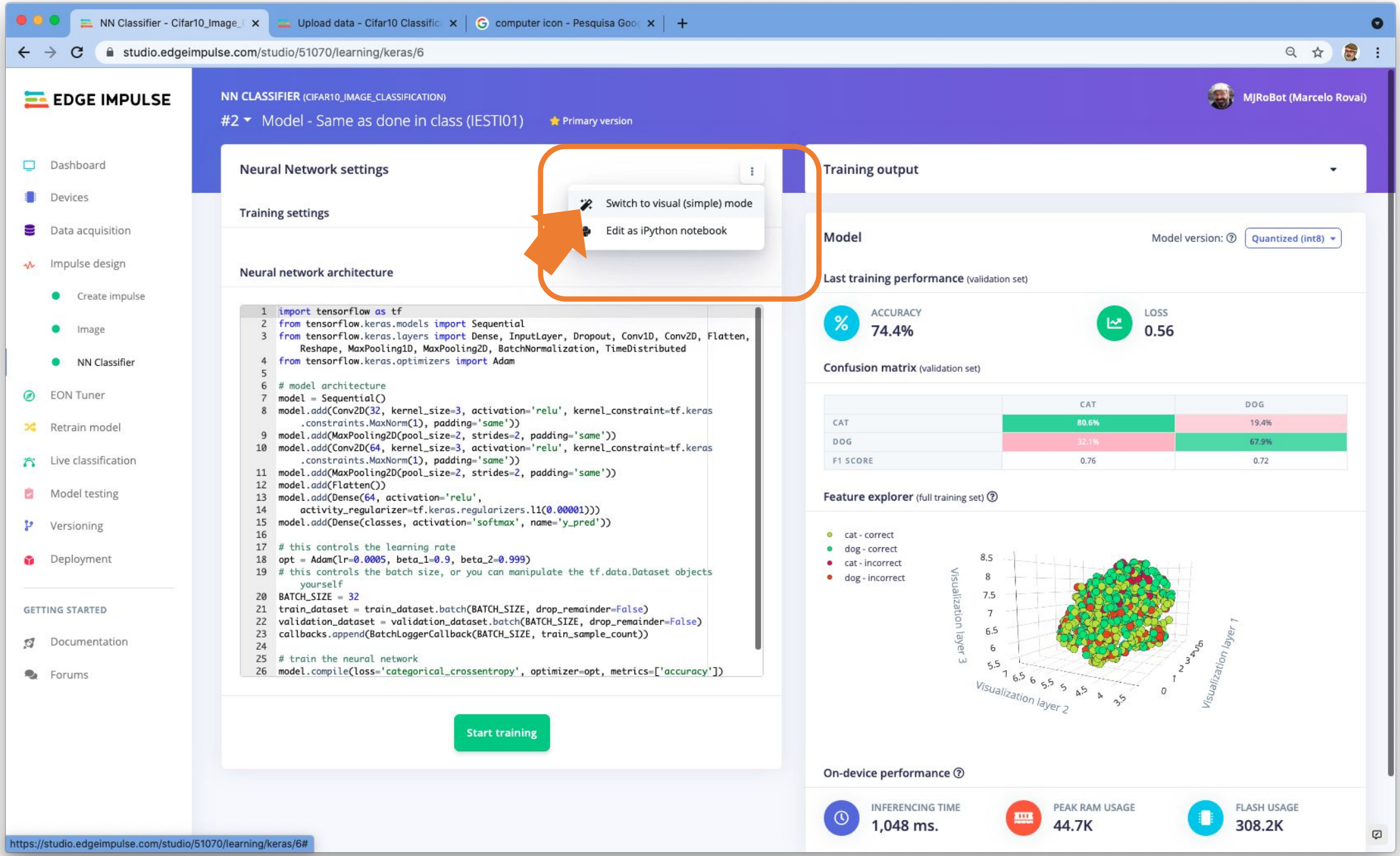




EI Studio “under the hood”

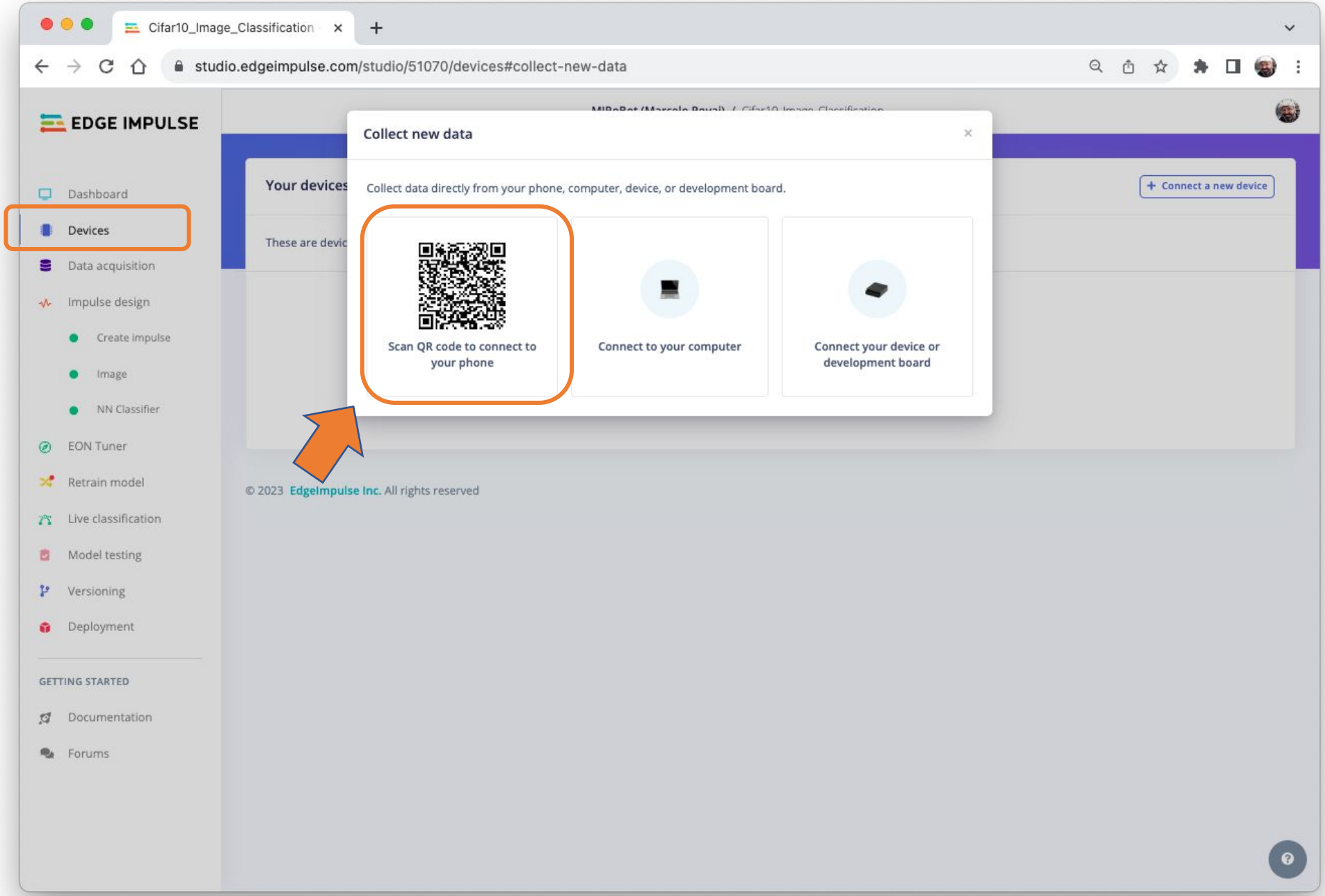
Code Time!

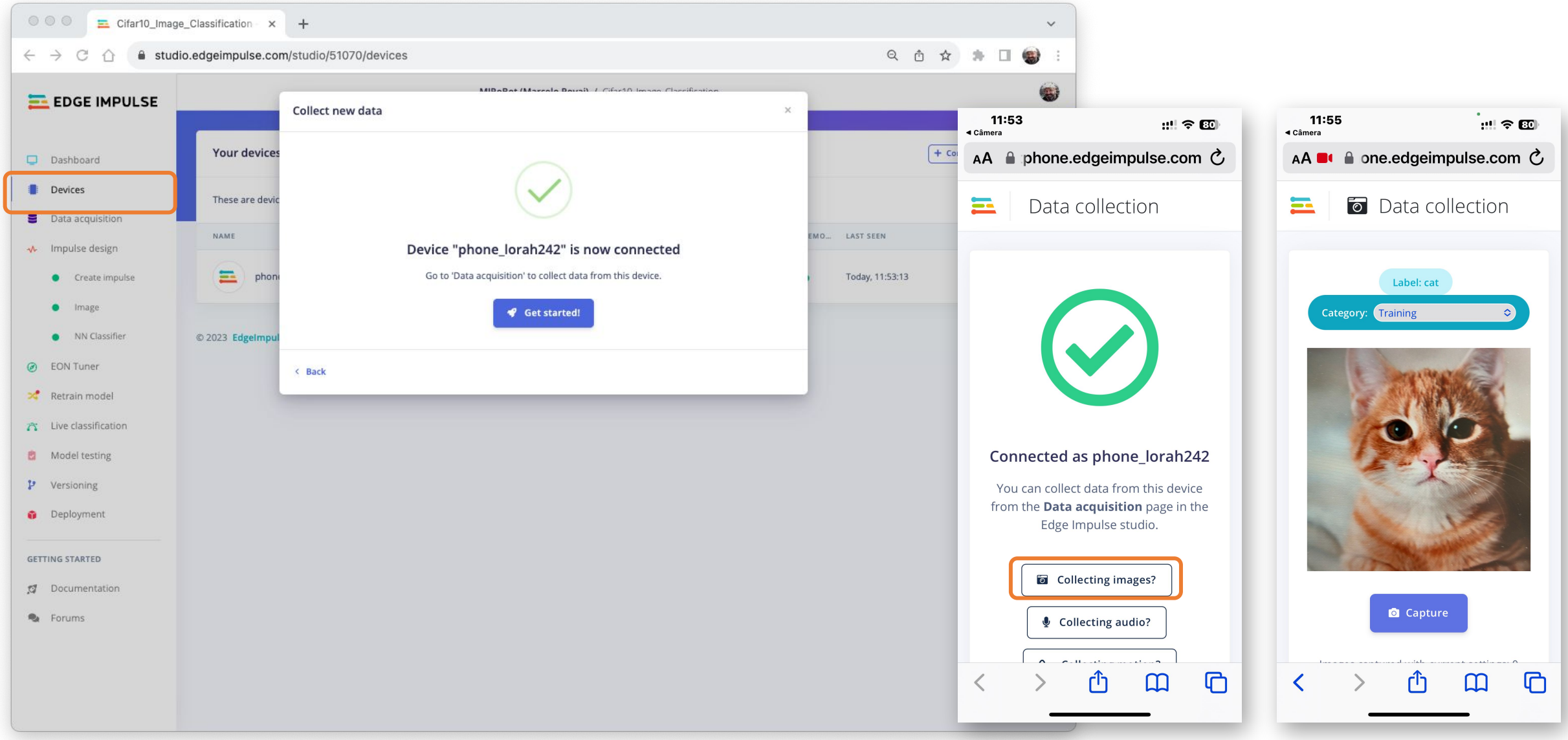
ei-cifar10_image_classification-nn-classifier-explained.ipynb


NN Classifier - Cifar10_Image_
Upload data - Cifar10 Classific
computer icon - Pesquisa Goo
studio.edgeimpulse.com/studio/51070/learning/keras/6
EDGE IMPULSE
NN CLASSIFIER (CIFAR10_IMAGE_CLASSIFICATION)
#2 Model - Same as done in class (IESTI01)
Primary version
MJRoBot (Marcelo Rovai)
Dashboard
Devices
Data acquisition
Impulse design
Create impulse
Image
NN Classifier
EON Tuner
Retrain model
Live classification
Model testing
Versioning
Deployment
GETTING STARTED
Documentation
Forums
Neural Network settings
Training settings
Switch to visual (simple) mode
Edit as iPython notebook
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(Conv2D(32, kernel_size=3, activation='relu', kernel_constraint=tf.keras.constraints.MaxNorm(1), padding='same'))
model.add(MaxPooling2D(pool_size=2, strides=2, padding='same'))
model.add(Conv2D(64, kernel_size=3, activation='relu', kernel_constraint=tf.keras.constraints.MaxNorm(1), padding='same'))
model.add(MaxPooling2D(pool_size=2, strides=2, padding='same'))
model.add(Flatten())
model.add(Dense(64, 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(lr=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'])
Start training
Training output
Model
Model version:
Quantized (int8)
Last training performance (validation set)
ACCURACY
74.4%
LOSS
0.56
Confusion matrix (validation set)
CAT
DOG
F1 SCORE
80.6%
19.4%
32.1%
67.9%
0.76
0.72
Feature explorer (full training set)
cat - correct
dog - correct
cat - incorrect
dog - incorrect
Visualization layer 1
Visualization layer 2
Visualization layer 3
On-device performance
INFERENCING TIME
1,048 ms.
PEAK RAM USAGE
44.7K
FLASH USAGE
308.2K
https://studio.edgeimpulse.com/studio/51070/learning/keras/6#

Connect device for data capture

Mobile Phone


Cifar10_Image_Classification
studio.edgeimpulse.com/studio/51070/devices#collect-new-data
EDGE IMPULSE
Dashboard
Devices
Data acquisition
Impulse design
Create impulse
Image
NN Classifier
EON Tuner
Retrain model
Live classification
Model testing
Versioning
Deployment
GETTING STARTED
Documentation
Forums
Your devices
These are devic
+ Connect a new device
Collect new data
Collect data directly from your phone, computer, device, or development board.
Scan QR code to connect to your phone
Connect to your computer
Connect your device or development board
© 2023 EdgeImpulse Inc. All rights reserved


Cifar10_Image_Classification
studio.edgeimpulse.com/studio/51070/devices
EDGE IMPULSE
Dashboard
Devices
Data acquisition
Impulse design
Create impulse
Image
NN Classifier
EON Tuner
Retrain model
Live classification
Model testing
Versioning
Deployment
GETTING STARTED
Documentation
Forums
Your devices
NAME
LAST SEEN
Today, 11:53:13
Collect new data
Device "phone_lorah242" is now connected
Go to 'Data acquisition' to collect data from this device.
Get started!
Back
11:53
phone.edgeimpulse.com
Data collection
Connected as phone_lorah242
You can collect data from this device from the Data acquisition page in the Edge Impulse studio.
Collecting images?
Collecting audio?
11:55
one.edgeimpulse.com
Data collection
Label: cat
Category: Training
Capture

Reading Material

Main references

- Harvard School of Engineering and Applied Sciences - CS249r: Tiny Machine Learning
- Professional Certificate in Tiny Machine Learning (TinyML) – edX/Harvard
- Introduction to Embedded Machine Learning - Coursera/Edge Impulse
- Computer Vision with Embedded Machine Learning - Coursera/Edge Impulse
- Fundamentals textbook: “Deep Learning with Python” by François Chollet
- Applications & Deploy textbook: “TinyML” by Pete Warden, Daniel Situnayake
- Deploy textbook “TinyML Cookbook” by Gian Marco Iodice

I want to thank Shawn Hymel and Edge Impulse, Pete Warden and Laurence Moroney from Google, Professor Vijay Janapa Reddi and Brian Plancher from Harvard, and the rest of the TinyMLedu team for preparing the excellent material on TinyML that is the basis of this course at UNIFEI.

The IESTI01 course is part of the **TinyML4D**, an initiative to make TinyML education available to everyone globally.

Thanks

UNIFEI