

ĐỒ ÁN CUỐI KỲ

HỆ THÔNG NHẬN DIỆN TIẾNG KHÓC CỦA  
TRẺ SƠ SINH SỬ DỤNG ESP32

MÔN HỌC: NHẬP MÔN LẬP TRÌNH ĐIỀU KHIỂN THIẾT  
BỊ THÔNG MINH - 20\_23  
NGƯỜI THỰC HIỆN: NGUYỄN HỮU THIỆN – 20120194

# NHỮNG TÀI NGUYÊN ĐƯỢC SỬ DỤNG

## 1. Phần cứng + •

- 1 Kit RF thu phát Wifi BLE ESP32 NodeMCU-32S CH340 Ai-Thinker
- 1 Cáp Micro USB
- 1 Sipeed MSM261S4030H0 High Sensitivity I2S Digital Microphone Module
- 1 Dây Cắm Test Board (5 sợi)

## 2. Phần mềm

- Arduino IDE 2.2.1
- EDGE IMPULSE

+

•

◦

# I. CÁCH THU THẬP DỮ LIỆU

- Dữ liệu được thu thập từ các nguồn sau:
  - + <https://freesound.org/>
  - + <https://www.kaggle.com/datasets/raiyanjahangir1939/infant-cry-sounds>
  - + [https://github.com/giulbia/baby\\_cry\\_detection/tree/master/data](https://github.com/giulbia/baby_cry_detection/tree/master/data)
- Là các đoạn ghi âm, audio ngắn (có độ dài khoảng 5 giây) và tất cả được chuyển thành định dạng tệp .wav
- Được gán nhãn trong 3 loại nhãn sau:
  1. cry (108 files)
  2. noise (108 files)
  3. unknow (108 files)

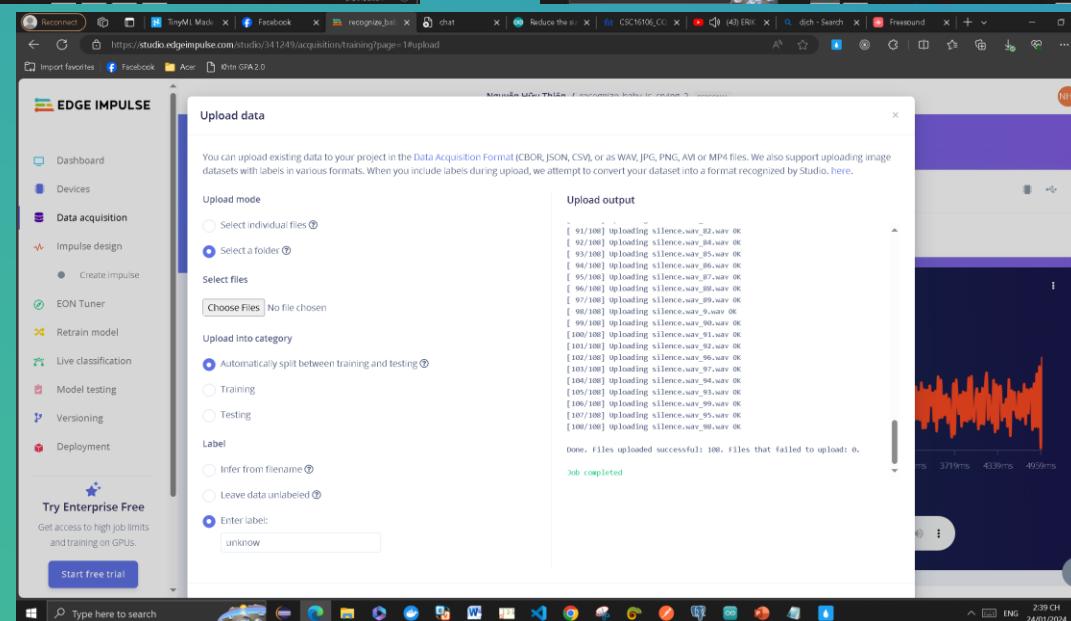
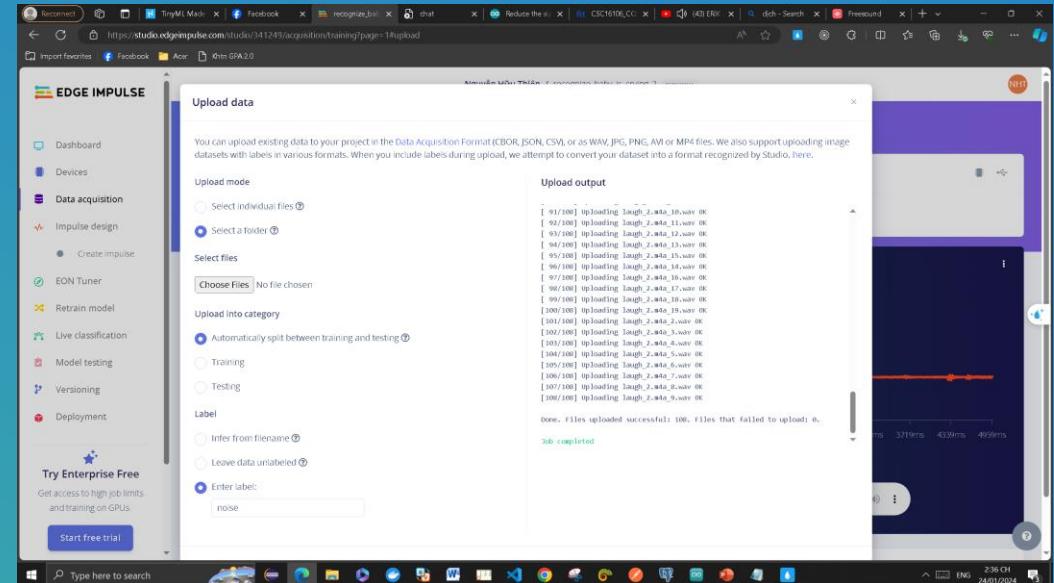
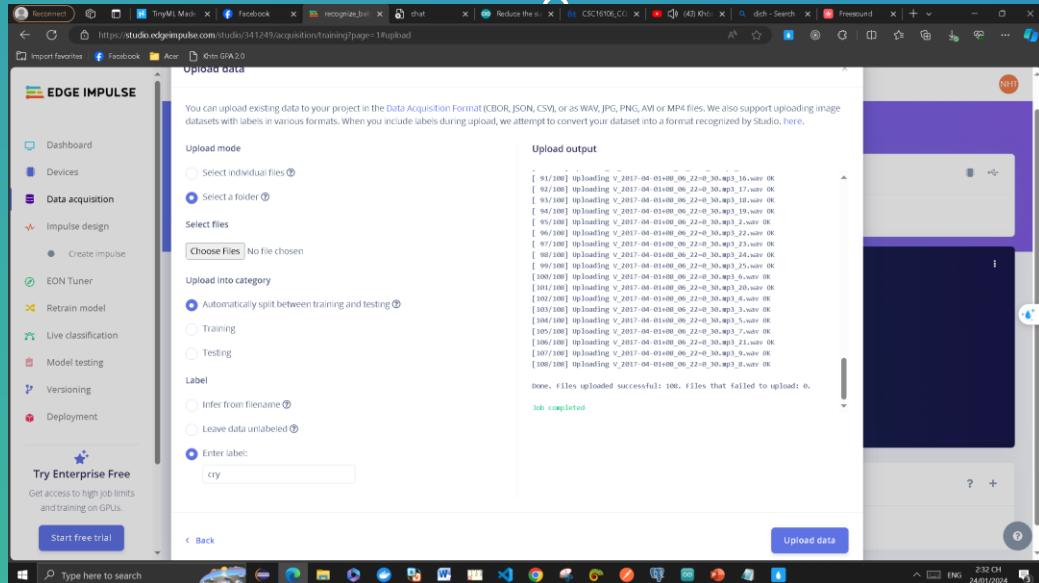
+

○

●

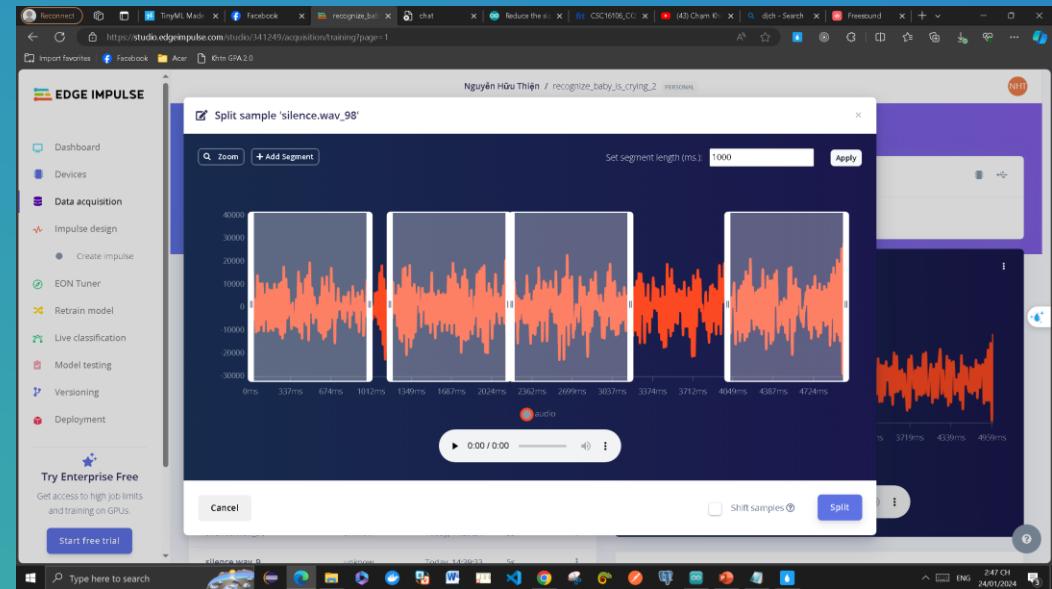
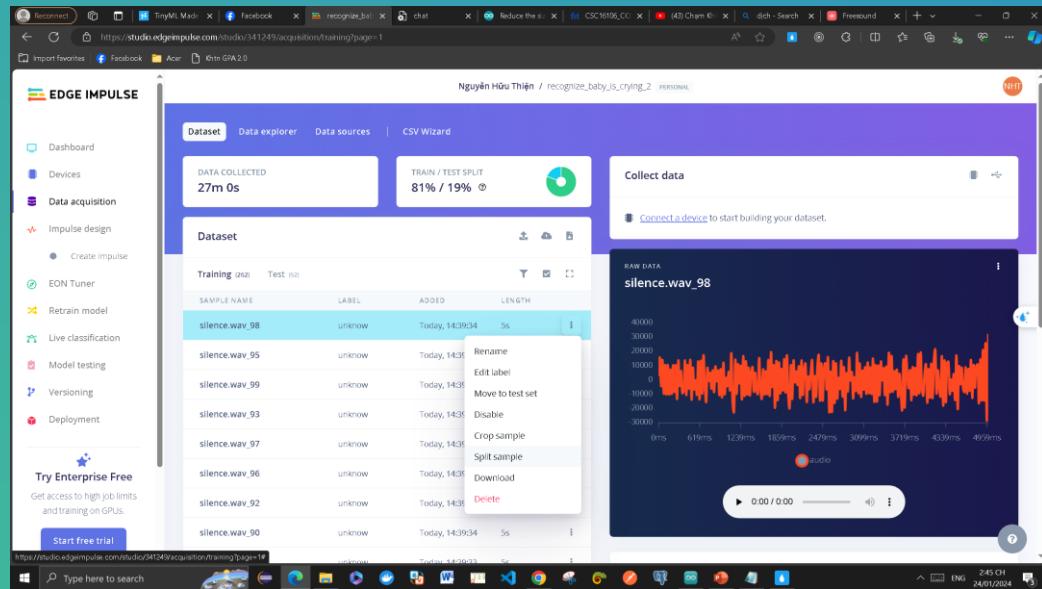
# II. KHẢO SÁT DỮ LIỆU

## B1: Upload dataset<sup>†</sup> vào EDGE IMPULSE



## II. KHẢO SÁT DỮ LIỆU (CONT)

B2: Data đã upload được chia thành từng mẫu nhỏ hơn (có độ dài 1 giây), đồng thời ta loại bỏ các phần âm thanh gây nhiễu hoặc không có tác dụng cho việc huấn luyện mô hình



## II. KHẢO SÁT DỮ LIỆU (CONT)

Kết quả thu được sau bước làm sạch dữ liệu là:

The screenshot shows the Edge Impulse Studio interface. On the left, there's a sidebar with various options like Dashboard, Devices, Data acquisition, and Model testing. The main area displays a dataset summary: "DATA COLLECTED 10m 57s" and "TRAIN / TEST SPLIT 82% / 18%". Below this is a table titled "Dataset" with columns for SAMPLE NAME, LABEL, ADDED, and LENGTH. The table lists 15 entries, all labeled "cry" and added today at 10:54:05, each lasting 1 second. At the bottom of the table is a navigation bar with page numbers from 1 to 45. To the right of the table, there's a "Collect data" section with a message: "Connect a device to start building your dataset." Below that is a "RAW DATA" section with the instruction "Click on a sample to load...". A "Start free trial" button is visible at the bottom left of the sidebar.

### III. THIẾT KẾ MÔ HÌNH MÁY HỌC

+  
•

B3: Sử dụng thuật toán học KWS (Keyword Spotting)

hay còn gọi là thuật toán nhận diện từ khóa. Có thể ta đang nghĩ đến vấn đề phân loại âm thanh thay vì KWS. Ban đầu, mình cũng nghĩ như vậy, nhưng dường như KWS mang lại độ chính xác cao hơn. Có thể là vì tiếng khóc của trẻ sơ sinh có nhịp tương tự như ngôn ngữ tự nhiên của chúng ta.

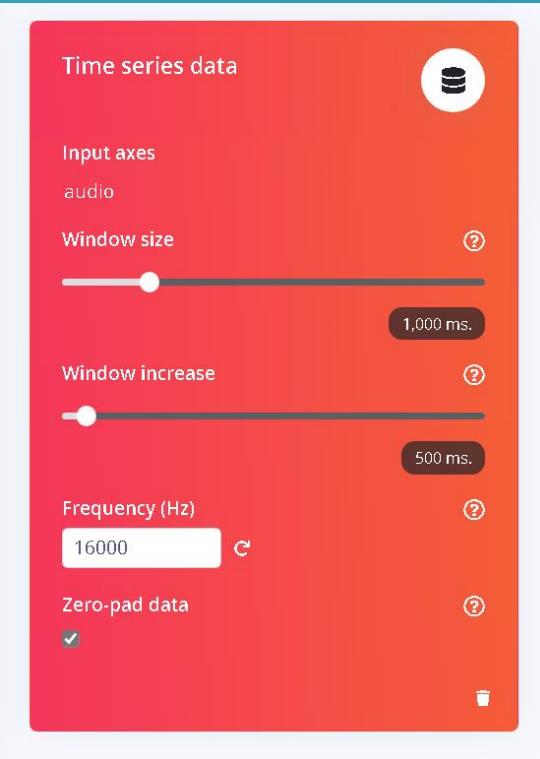
+  
•  
○

### III. THIẾT KẾ MÔ HÌNH MÁY HỌC (CONT)

B4: Create impulse<sup>+</sup> .

1. Tại Time series data:

- Window size = 1000ms (Vì tất cả các mẫu dữ liệu sau khi đã xử lý đều có độ dài 1 giây)
- Window increase = 500ms (Tăng cường dữ liệu mỗi 0.5 giây)
- Frequency (Hz) = 16000 (EDGE IMPULSE yêu cầu bắt buộc 16000 đối với thuật toán KWS)



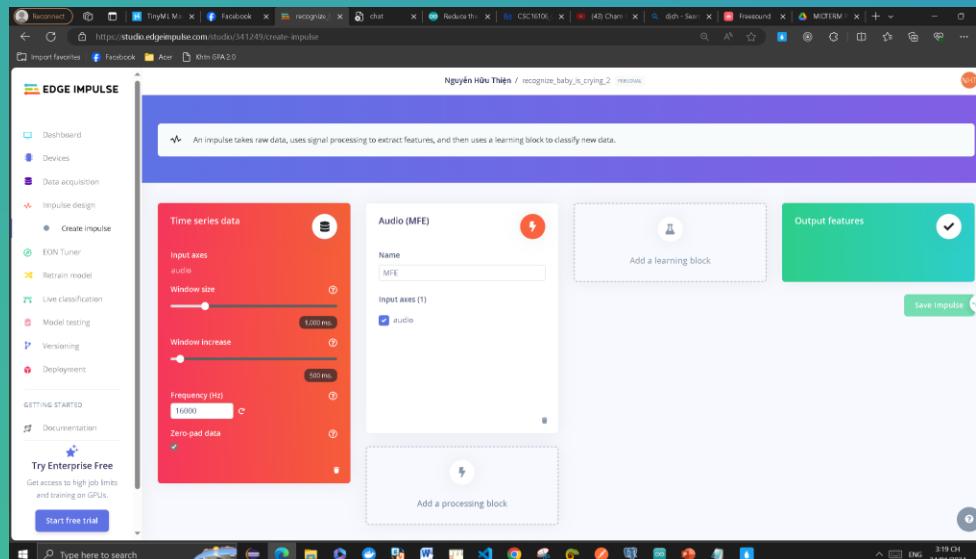
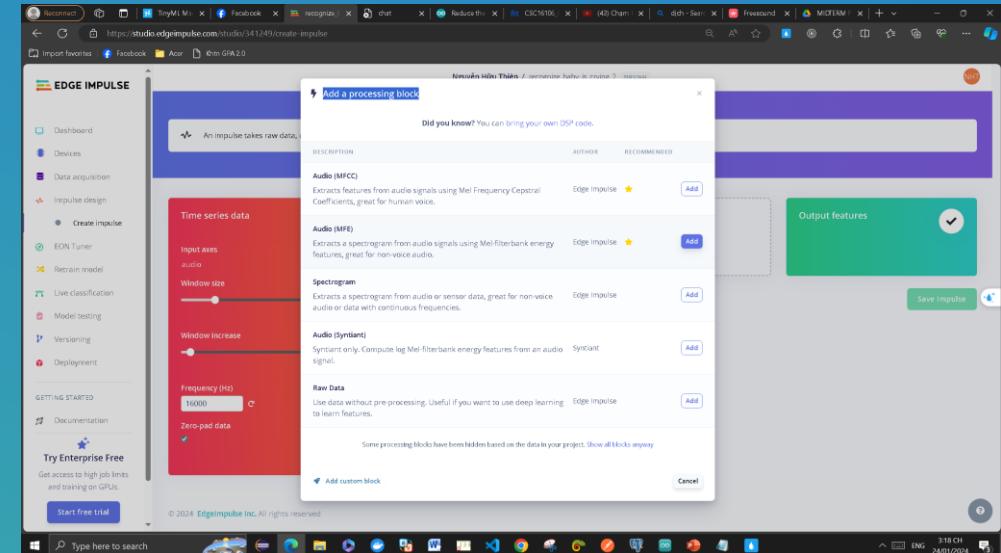
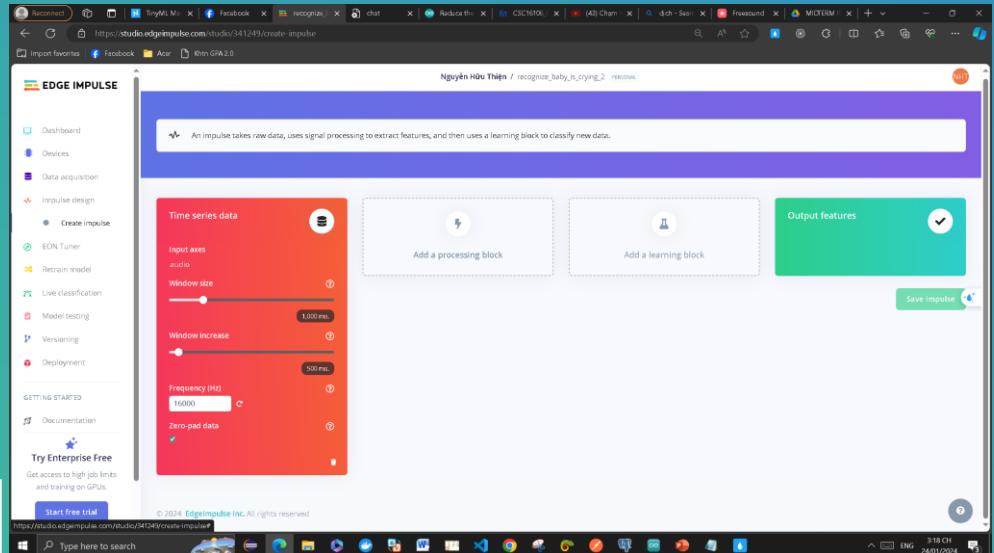
+

o

.

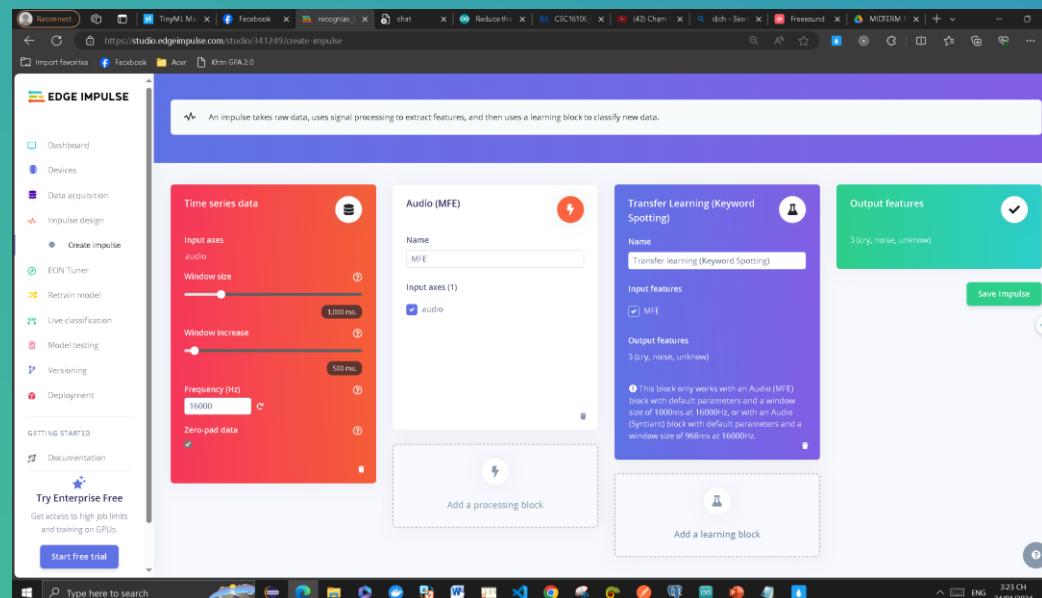
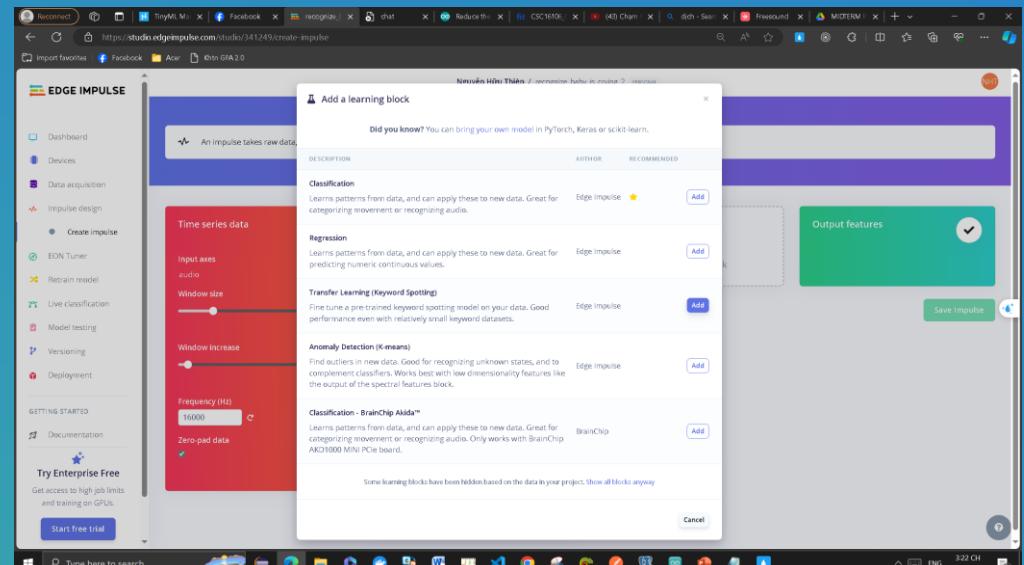
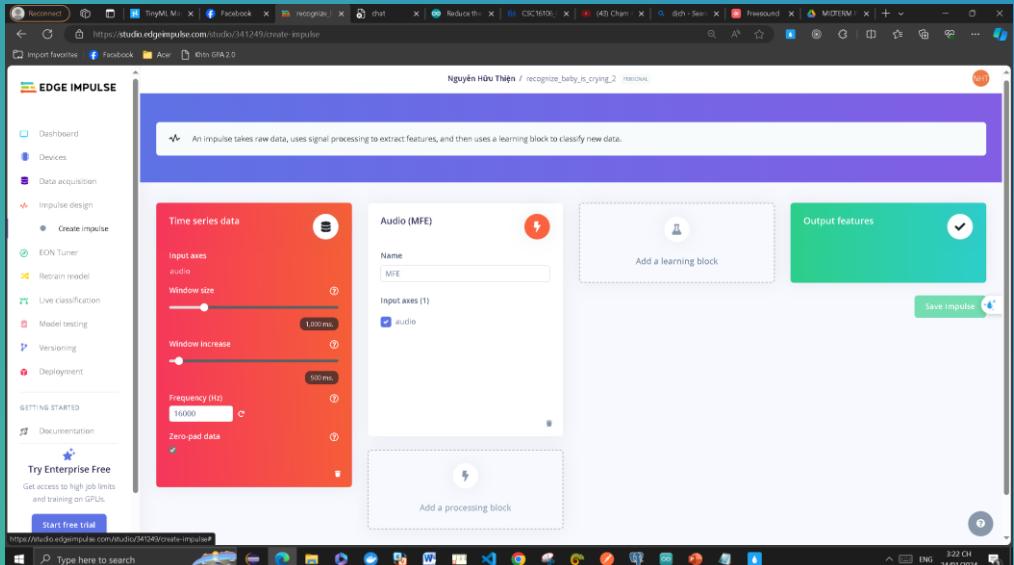
# III. THIẾT KẾ MÔ HÌNH MÁY HỌC (CONT)

## 2. Add a processing block (Audio (MFE))



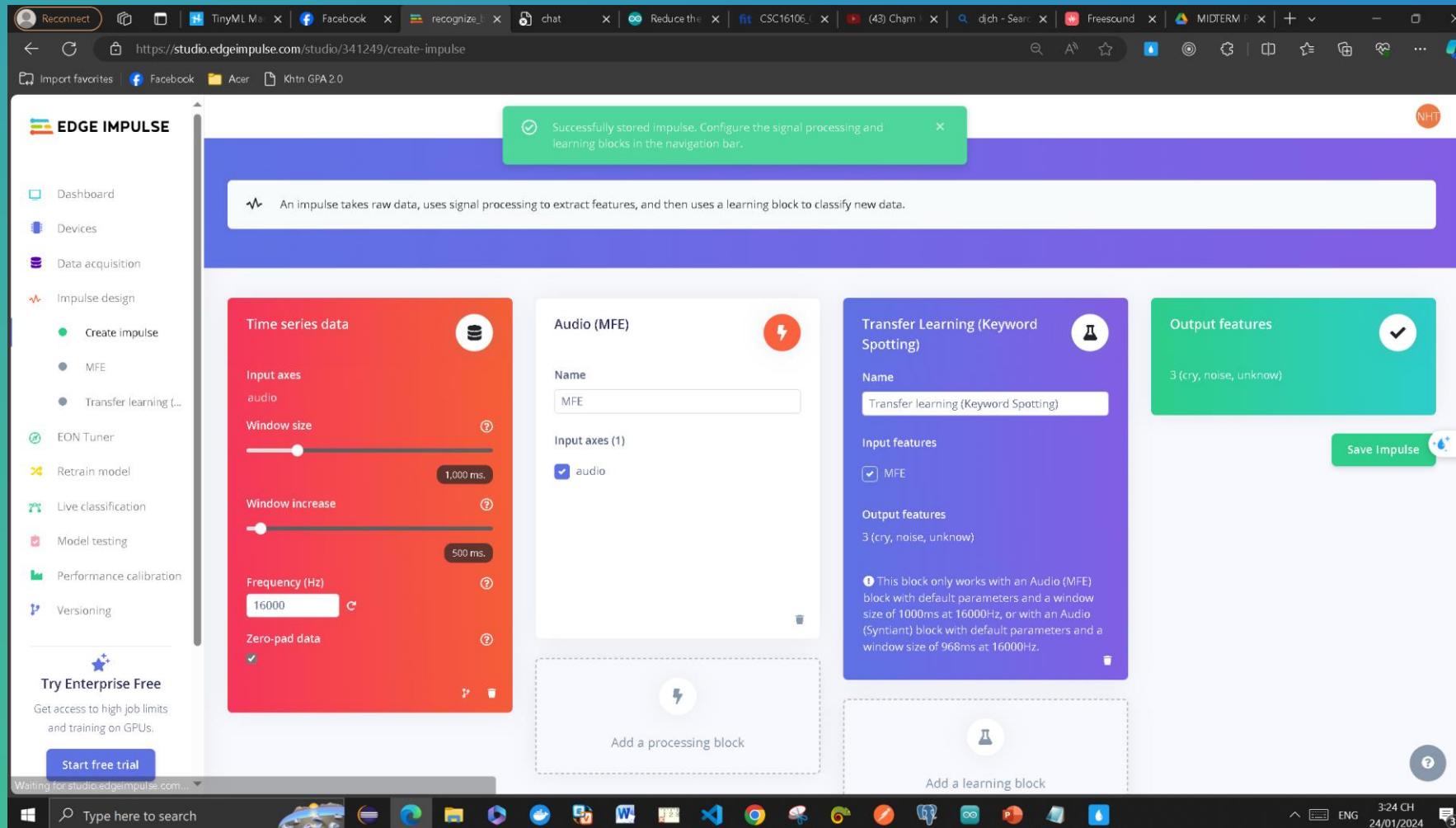
# III. THIẾT KẾ MÔ HÌNH MÁY HỌC (CONT)

## 3. Add a learning block (Transfer Learning (Keyword Spotting))



# III. THIẾT KẾ MÔ HÌNH MÁY HỌC (CONT)

## 4. Save Impulse



### III. THIẾT KẾ MÔ HÌNH MÁY HỌC (CONT)

B5: Chuyển sang mục MFE của Impulse design, tại tab Parameters

- Chọn Autotune parameters
- Nhấn Save parameters

The screenshot shows the Edge Impulse studio interface. On the left, a sidebar lists various project sections: Dashboard, Devices, Data acquisition, Impulse design, Create impulse, MFE (selected), Transfer learning (K...), EON Tuner, Retrain model, Live classification, Model testing, Performance calibration, Versioning, Deployment, and Getting Started. A 'Try Enterprise Free' section with a 'Start free trial' button is also present.

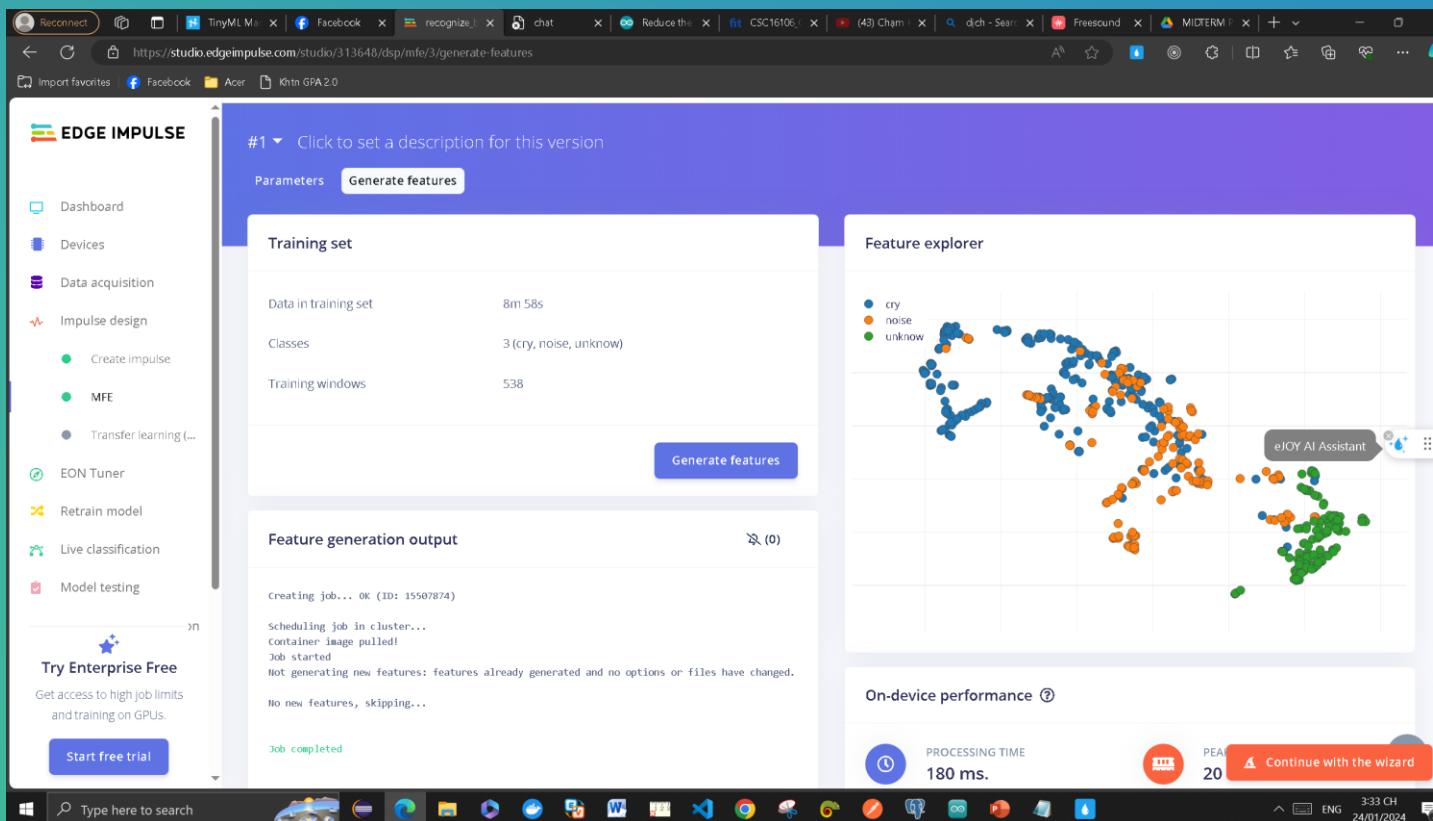
The main workspace is titled '#1 Click to set a description for this version'. It features a 'Parameters' tab (selected) and a 'Generate features' tab. The 'Raw data' section displays a waveform for a file named '3-152007-A.s3 (cry)' with a duration of 0:00 / 0:01. Below it, the 'Raw features' section shows a list of numerical values: 1387.9588, 2449.8199, 2696.1122, 1735.1409, 818.4865, 1025.5708, 198.0736, -759.6915, -622.8782, -785.7448, -1740.7986, -1557.36. To the right, a 'Label' field is set to 'cry'. The 'DSP result' section displays two plots: 'Mel Energies (DSP Output)' and 'FFT Bin Weighting'. The 'Mel Energies' plot shows energy distribution over time (0.0 to 1.0 seconds) and frequency (45 to 7482 Hz). The 'FFT Bin Weighting' plot shows the relationship between the FFT Bin Index (0 to 120) and the Output Row Index (0 to 30). At the bottom, there is a 'Save parameters' button and a 'Continue with the wizard' button.

### III. THIẾT KẾ MÔ HÌNH MÁY HỌC (CONT)

Chuyển sang tab Generate features

1. Chọn Generate features

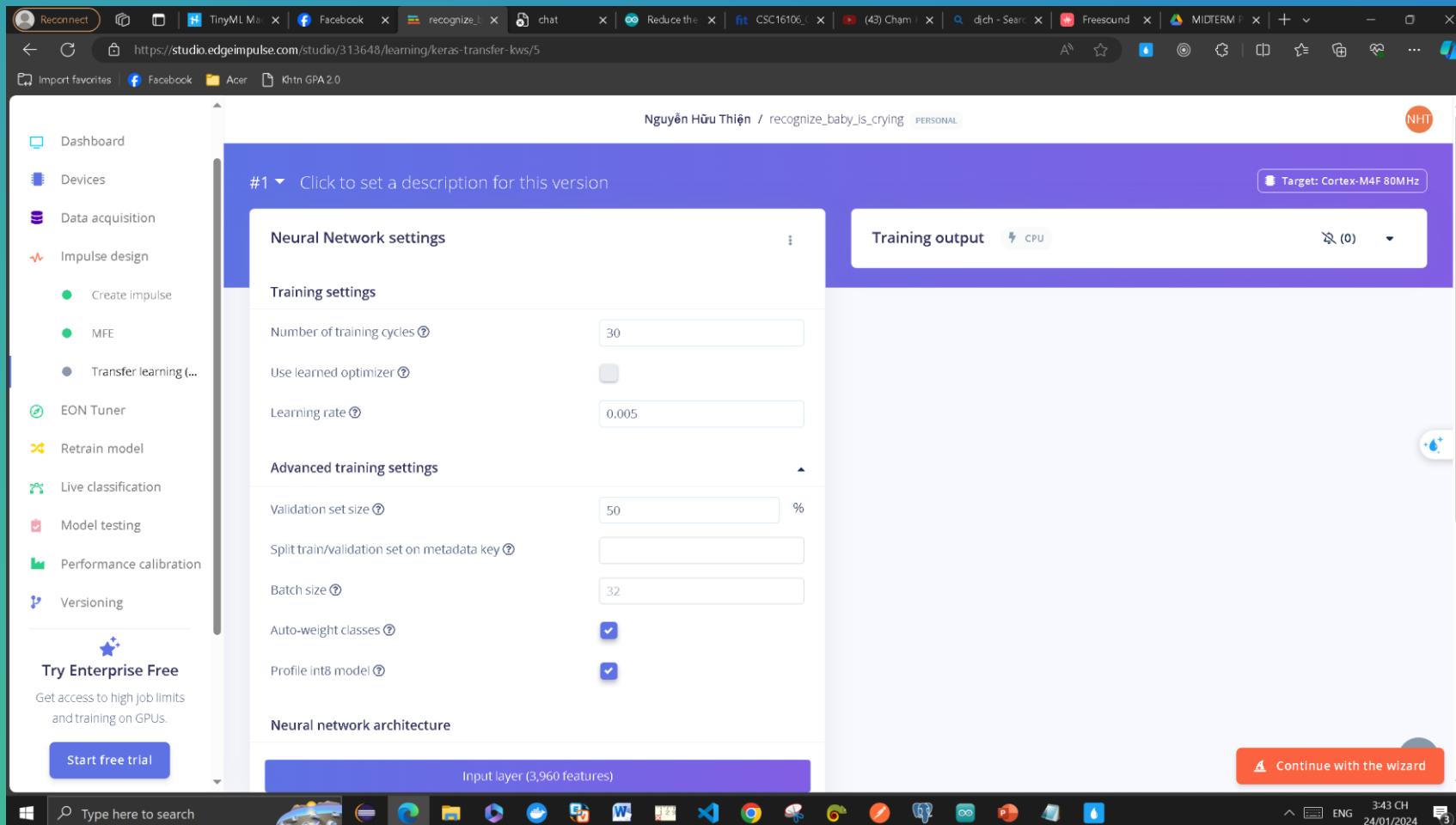
2. Quan sát Feature explorer ta thấy chỉ có các dữ liệu có nhãn unknown được phân thành 1 cụm rõ ràng, trong khi các dữ liệu có nhãn là cry và noise bị hòa trộn vào nhau => Điều này làm những dự đoán trong tương lai về việc phân biệt tiếng khóc hay tiếng ồn có thể gây nhầm lẫn



### III. THIẾT KẾ MÔ HÌNH MÁY HỌC (CONT)

B6: Chuyển sang mục Transfer learning (...) của Impulse design

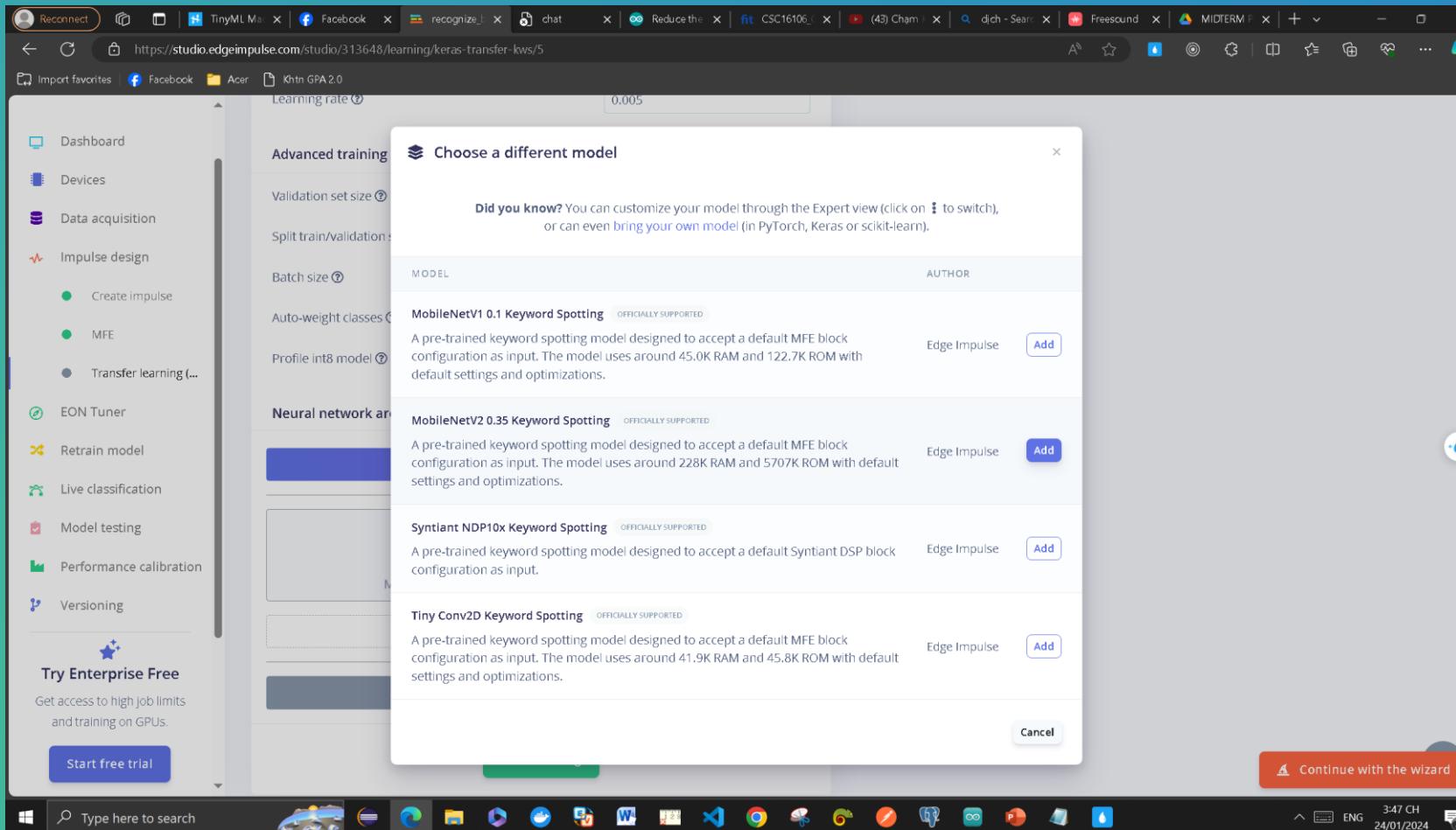
1. Cấu hình các thông số như hình:



### III. THIẾT KẾ MÔ HÌNH MÁY HỌC (CONT)

2. Tại Neural network architecture sử dụng model (MobileNetV2 0.35 Keyword Spotting vì đây là một model mới nên sẽ có hiệu suất huấn luyện tốt hơn các model khác)

3. Chọn Start training để bắt đầu huấn luyện mô hình



# IV. ĐÁNH GIÁ MÔ HÌNH

## 1. Kết quả huấn luyện mô hình

- Có kết quả khá tốt với accuracy khoảng 95.5% và loss chỉ có 0.1

The screenshot shows the Edge Impulse Studio interface with the following details:

- Left Sidebar:** Contains navigation links such as Dashboard, Devices, Data acquisition, Impulse design, Create impulse, MFE, Transfer learning, EON Tuner, Retrain model, Live classification, Model testing, Performance calibration, Versioning, and Try Enterprise Free.
- Main Area - Neural network architecture:** Shows the "Profile int8 model" section with a checked checkbox. It displays the "Input layer (3,960 features)" and "Output layer (3 classes)". Below these are sections for "MobileNetV2 0.35 (final layer: 128 neurons, 0.1 dropout)" and "Choose a different model". A "Start training" button is located at the bottom right of this section.
- Right Area - Last training performance:** Displays validation set performance metrics:
  - ACCURACY: 95.5%
  - LOSS: 0.10
- Confusion matrix (validation set):**

	CRY	NOISE	UNKNOW
CRY	96.9%	3.1%	0%
NOISE	5.6%	90.1%	4.2%
UNKNOW	1.5%	0%	98.5%
F1 SCORE	0.97	0.92	0.97
- Data explorer (full training set):** A scatter plot showing data points categorized by class and correctness. The legend includes:
  - cry - correct (light green)
  - noise - correct (medium green)
  - unknow - correct (dark green)
  - cry - incorrect (red)
  - noise - incorrect (dark red)
  - unknow - incorrect (brown)
- On-device performance:** Shows performance metrics:
  - INFERENCE TIME: 920 ms.
  - PEAK RAM USAGE: 216.3K
  - FLASH USAGE: (not explicitly shown in the screenshot)
- Bottom Right:** A red button labeled "Continue with the wizard".

# IV. ĐÁNH GIÁ MÔ HÌNH (CONT)

## 2. Kết quả khi chạy<sup>+</sup> trên tập test

- Chuyển sang mục Model testing, sau đó chọn Classify all.
- Kết quả: mô hình dự đoán chính xác đến 93.28%

=> Hứa hẹn sẽ có khả năng nhận diện được tiếng khóc của em bé khá tốt khi mô hình được triển khai trên thiết bị nhúng

The screenshot shows the Edge Impulse Studio interface on a Windows desktop. The main window displays the 'Model testing output' tab, which shows the process of classifying data for transfer learning (keyword spotting) and retraining a float32 model. The 'Model testing results' section indicates an accuracy of 93.28% with a confusion matrix:

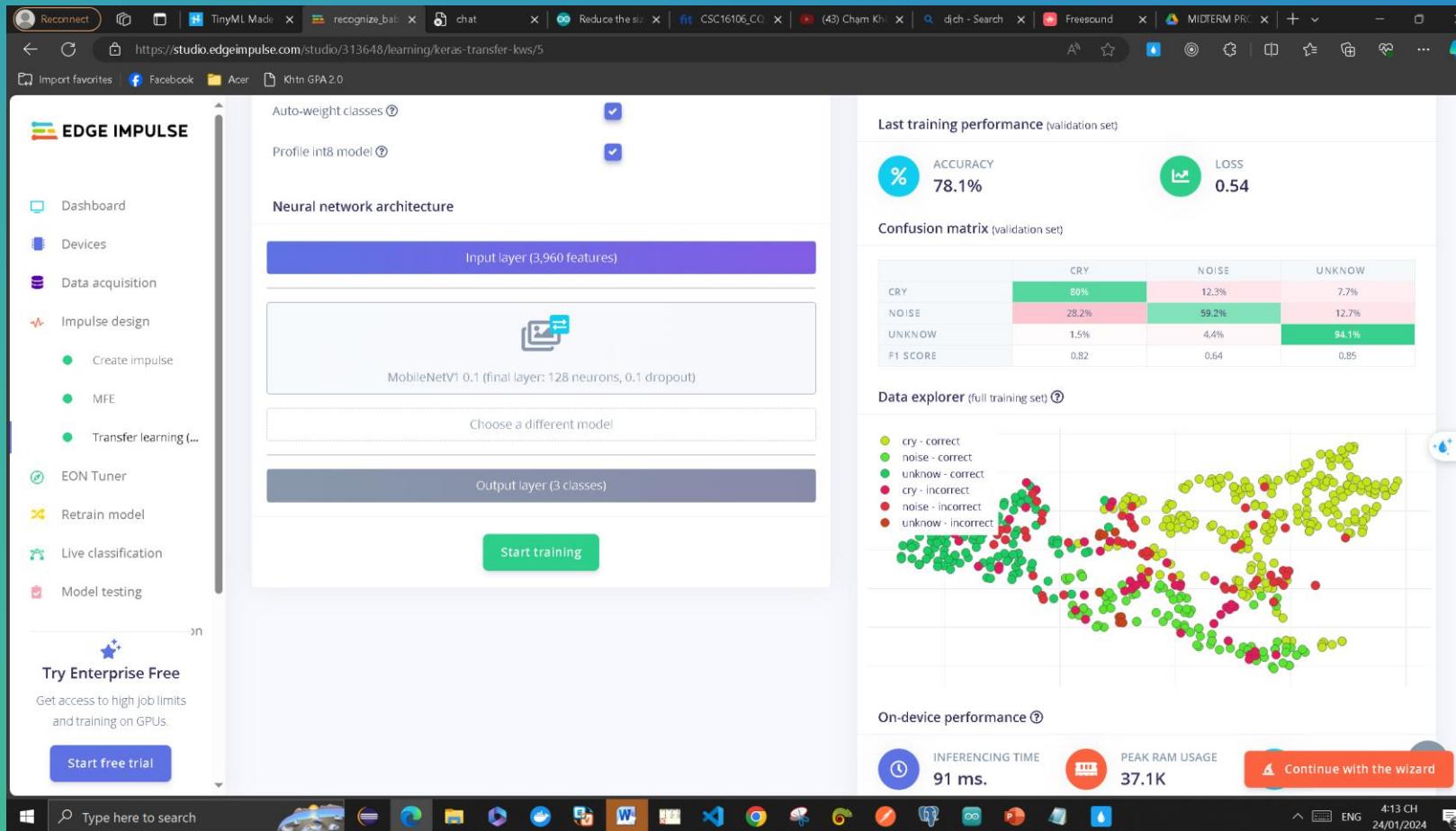
	CRY	NOISE	UNKNOWN	UNCERTAIN
CRY	90.9%	3.6%	1.8%	3.6%
NOISE	6.5%	90.3%	3.2%	0%
UNKNOWN	0%	0%	100%	0%
F1 SCORE	0.93	0.92	0.97	

The left sidebar shows the navigation menu with 'Model testing' selected. The bottom taskbar shows various open applications including a browser tab for 'TinyML Made' and system status icons.

# V. KINH NGHIỆM HUẤN LUYỆN MÔ HÌNH

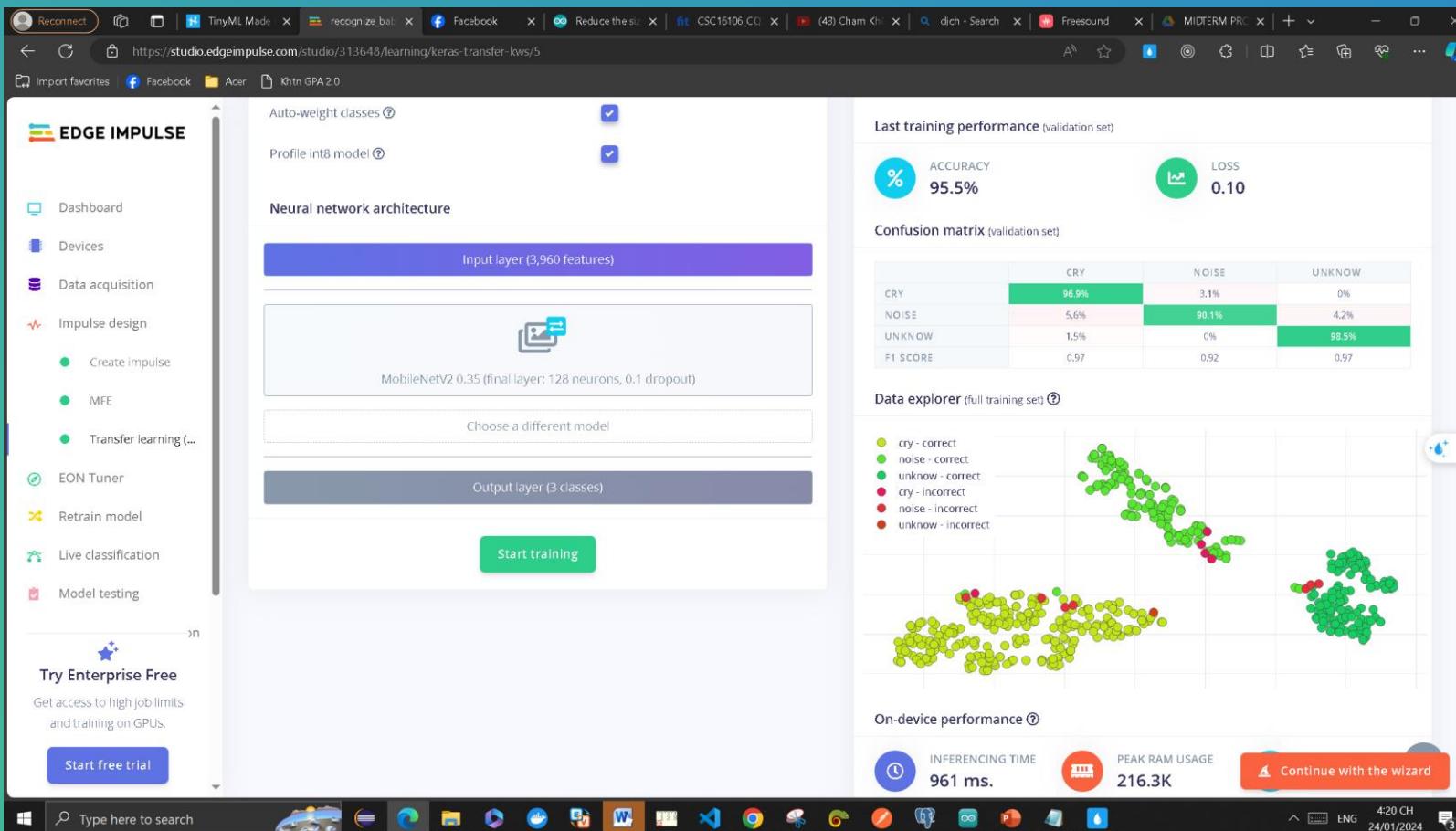
1. Huấn luyện mô hình với Neural network architecture (MobileNetV1 0.1 Keyword Spotting)

=> Kết quả là:



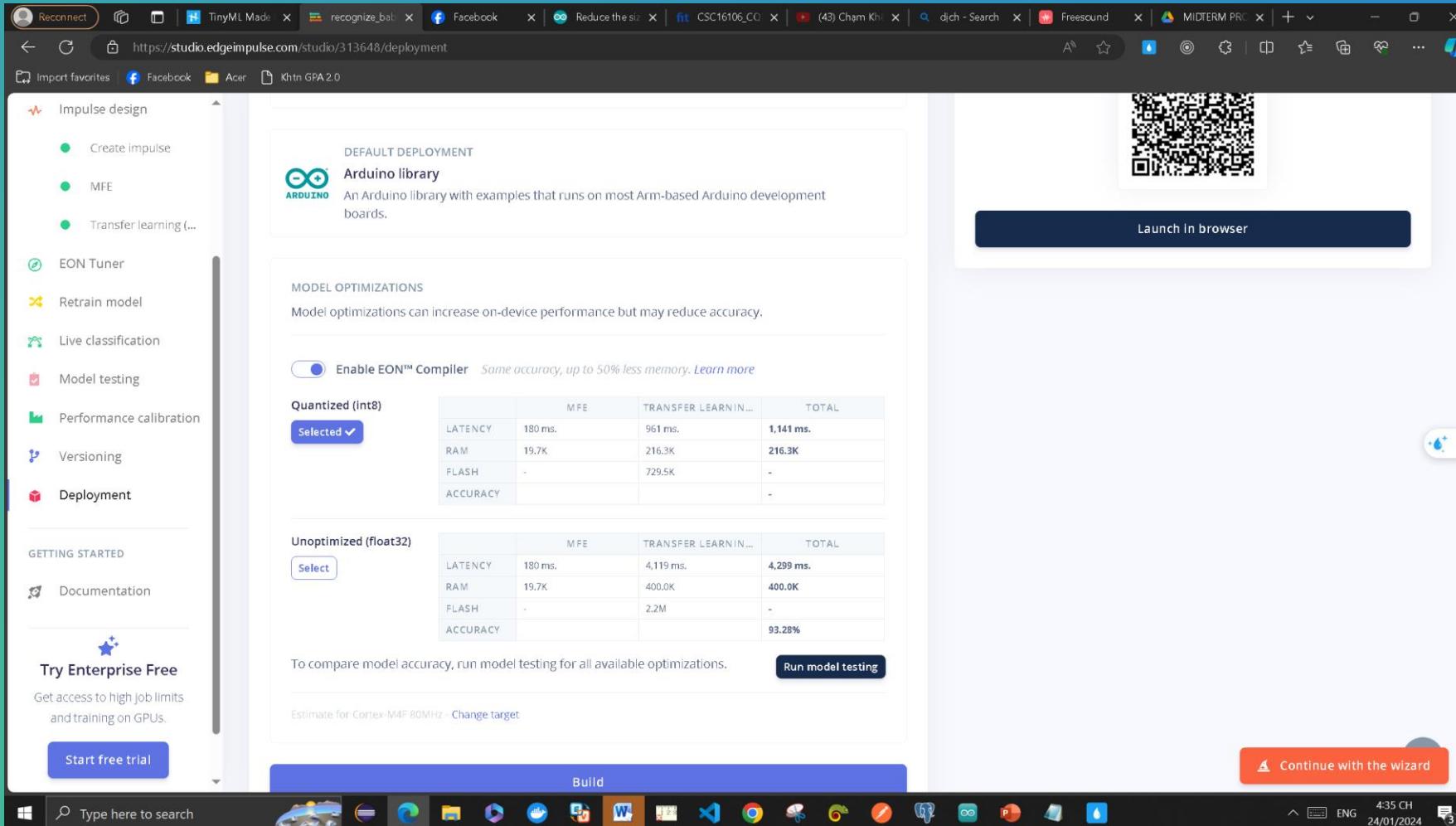
# V. KINH NGHIỆM HUẤN LUYỆN MÔ HÌNH (CONT)

2. Huấn luyện mô hình với Neural network architecture (MobileNetV2 0.35 Keyword Spotting)  
=> Kết quả là: Accuracy 95.5% cao hơn rất nhiều (78.1% với MobileNetV1 0.1 Keyword Spotting)



# VI. TRIỂN KHAI MÔ HÌNH LÊN THIẾT BỊ NHÚNG

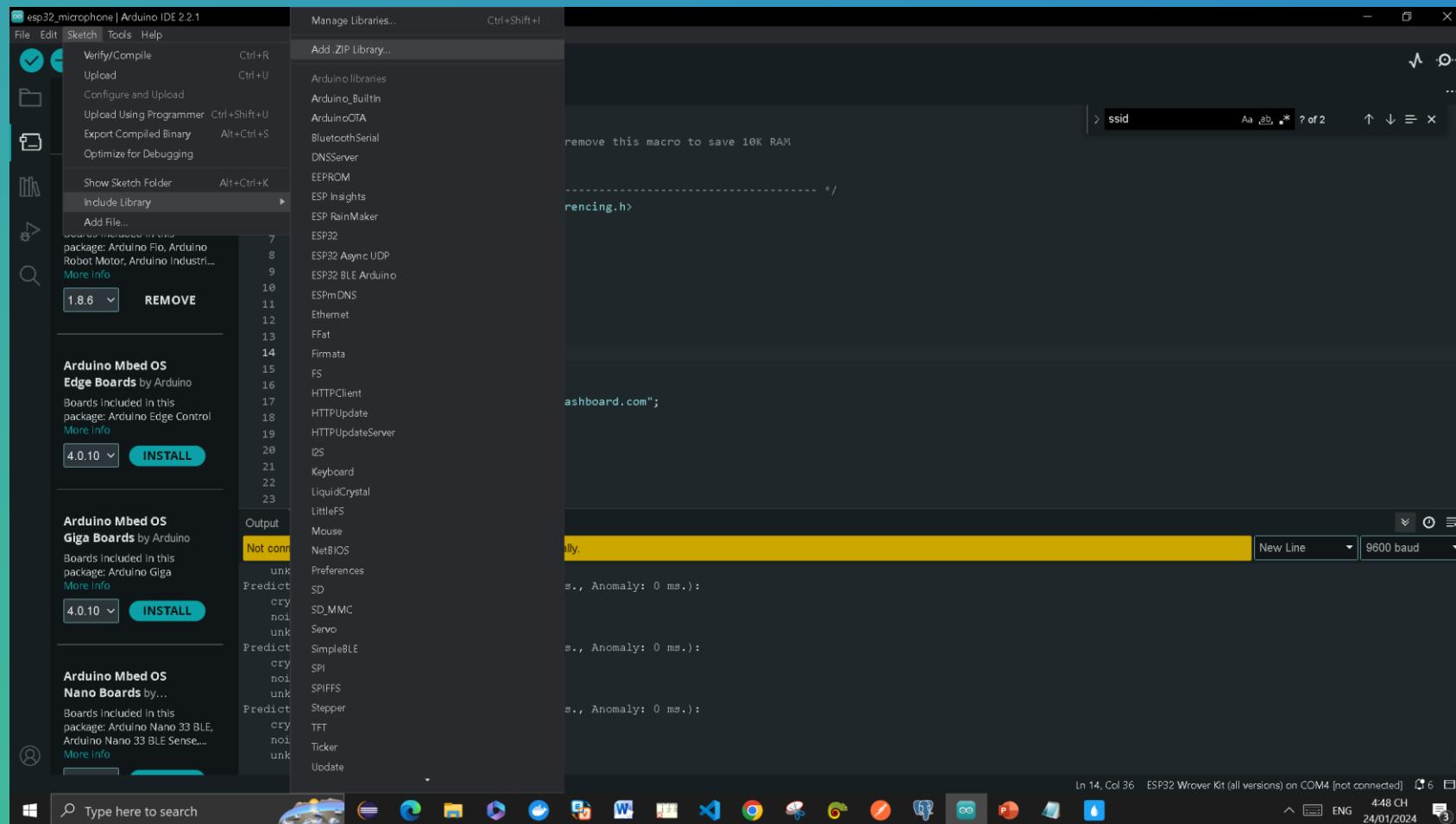
1. Chuyển sang mục Deployment
2. Search deployment options chọn Arduino library
3. Chọn Build để tải model xuống



# VI. TRIỂN KHAI MÔ HÌNH LÊN THIẾT BỊ NHÚNG (CONT)

## 4. Tại Arduino IDE<sup>+</sup>.

- Sketch -> Include<sup>O</sup>Library -> Add .ZIP Library...
- Chọn model vừa được deploy và tải về (.zip) ở bước trên



# VI. TRIỂN KHAI MÔ HÌNH LÊN THIẾT BỊ NHÚNG (CONT)

## 5. Sử dụng model trong Arduino IDE

- Cú pháp tên model: <Tên project của bạn trên EDGE IMPULSE>\_inferencing.h

-  Nguyen Huu Thien / recognize\_baby\_is\_crying PERSONAL NHT

- Ở đây tên project của mình là recognize\_baby\_is\_crying => import recognize\_baby\_is\_crying\_inferencing.h

```
4  /* Includes ----- */  
5  #include <recognize_baby_is_crying_inferencing.h>  
6  #include <WiFi.h>
```



# VI. TRIỂN KHAI MÔ HÌNH LÊN THIẾT BỊ NHÚNG (CONT)

## 6. Quan sát kết quả dự đoán trên Arduino IDE

The screenshot shows the Arduino IDE interface. On the left, the Boards Manager is open, displaying four sections:

- Arduino AVR Boards**: Version 1.8.6 installed.
- Arduino Mbed OS Edge Boards**: Version 4.0.10, with an **INSTALL** button.
- Arduino Mbed OS Giga Boards**: Version 4.0.10, with an **INSTALL** button.
- Arduino Mbed OS Nano Boards**: Version 4.0.0, with an **INSTALL** button.

The main workspace shows the code for `esp32_microphone.ino`:

```
106     ei_printf(": \n");
107     for (size_t ix = 0; ix < EI_CLASSIFIER_LABEL_COUNT; ix++) {
108         ei_printf(" %s: ", result.classification[ix].label);
109         ei_printf_float(result.classification[ix].value);
110         ei_printf("\n");
111     }
112     // if(result.classification[0].value > 0.6)
113     // {
114     //     client.publish("20120194","");
115     // }
116     #if EI_CLASSIFIER_HAS_ANOMALY == 1
117     ei_printf(" anomaly score: ");
118     ei_printf_float(result.anomaly);
119     ei_printf("\n");
120 #endif
121
122
123     static void audio_inference_callback(uint32_t n_bytes)
124     {
125         for(int i = 0; i < n_bytes>>1; i++) {
126             inference.buffer[inference.buf_count++] = sampleBuffer[i];
127
128             if(inference.buf_count >= inference.n_samples) {
129                 inference.buf_count = 0;
130                 inference.buf_ready = 1;
131             }
132         }
133     }
134 }
```

The **Serial Monitor** tab is active, showing the output:

```
Not connected. Select a board and a port to connect automatically.
unknown: 0.000000
Predictions (DSP: 669 ms., Classification: 13 ms., Anomaly: 0 ms.):
cry: 0.152344
noise: 0.363281
unknown: 0.484375
Predictions (DSP: 669 ms., Classification: 13 ms., Anomaly: 0 ms.):
cry: 0.640625
noise: 0.359375
unknown: 0.000000
Predictions (DSP: 669 ms., Classification: 13 ms., Anomaly: 0 ms.):
cry: 0.570312
noise: 0.429687
unknown: 0.000000
```

The status bar at the bottom indicates: Ln 118, Col 27 ESP32 Wrover Kit (all versions) on COM4 [not connected] 4:58 CH ENG 24/01/2024

## VII. TÀI LIỆU THAM KHẢO

- +
- 
- 

Đồ án được lấy ý tưởng từ bài viết:

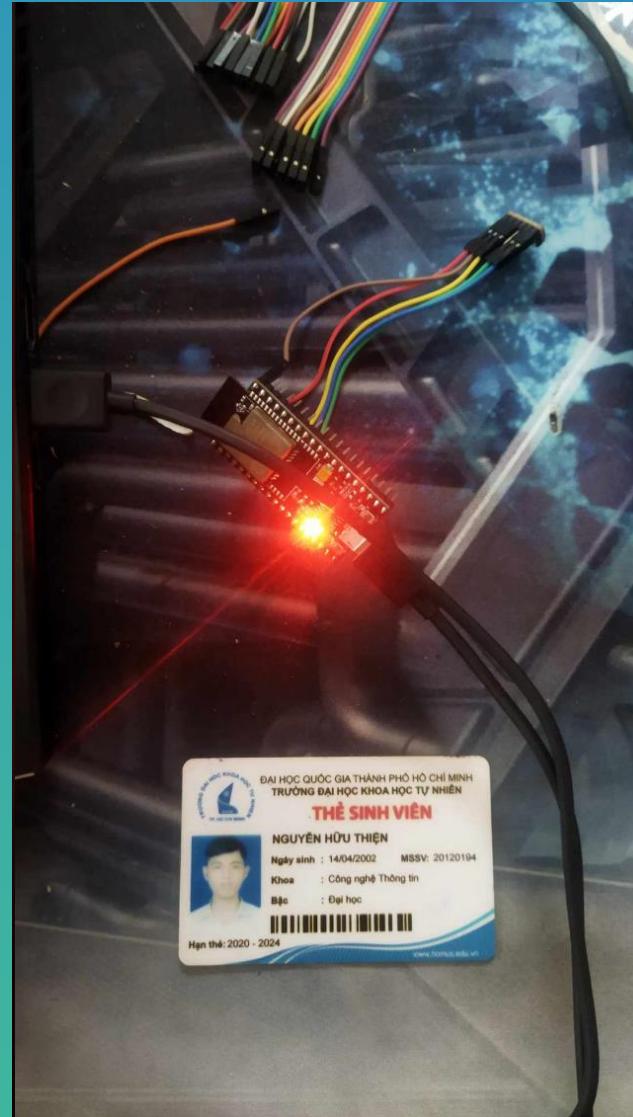
<https://www.hackster.io/nafihahmd/e-nanny-esp32-based-smart-baby-sleep-monitoring-system-9e68fb>

- +
- 
-

# VIII. DEMO SẢN PHẨM

Setup thiết bị

+  
●  
○



+  
●  
○