A Study on Creating & Classifying Datasets of Toyota Diesel Engine Sounds from Human Voices with Edge Impulse and Arduino

Sajjad

1. Introduction

"Can machines talk?", was the first topic of our discussion with Mr. Jerker. Intrigued by the question, we decided to work with this theme in mind. Identifying the voices between a human and an engine is the first step towards finding that answer.

Despite all the efforts, the most challenging task was finding a workable dataset for car engine sounds. To overcome that, a *Zoom Pod Trak P4* recorder was used to record engine sound data from a Toyota Corolla 2008 1.4 Diesel. The human sound datasets were also manually recorded. Utilizing the datasets, Edge Impulse was used for training a model to recognize specific engine sound patterns through Arduino.

2. **Objectives**:

The main objective was to create an usable dataset for car engine sound data and use Edge computing to accurately and efficiently distinguish engine sounds from human voices with a focus on Toyota diesel engines. Identifying and analyzing engine sounds can be crucial for vehicle maintenance and early detection of potential issues. This project applies machine learning, IoT, and embedded systems to solve a practical problem in automotive engineering.

3. Methodology

3.1 Research and Planning:

The research involved extensive search for existing datasets of engine sounds. However, very few such datasets were available during the research period. Which influenced us in creating our own datasets from scratch. We used our own datasets for this study and carefully conducted the plannings.

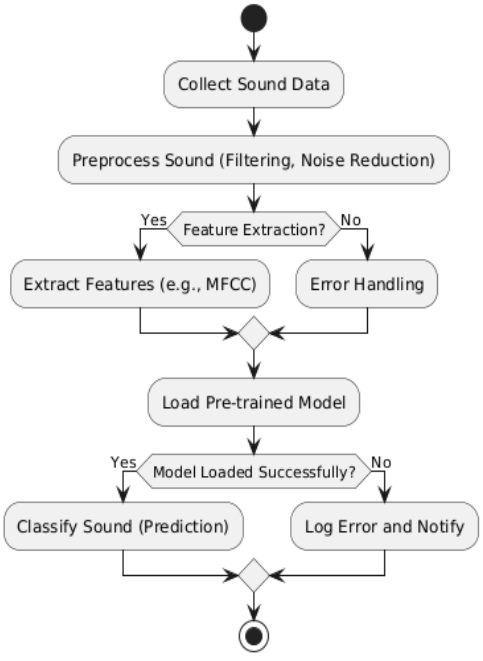


Figure: Data Collection Flow Chart

3.2 Data Collection:

Nearly 120 WAV files of Idle engine sound recording were created with an average duration of 4 seconds. The data recording was done by placing a *Zoom Pod Trak P4* recorder 25 cm above the engine. Four different human voices each speaking the phrase “Turku” and “Vaasa” were recorded using the microphone connected with the Arduino. The total amount of collected data was 981 WAV files and the duration was 34 minutes. The data was then processed for feature extraction using Edge Impulse.

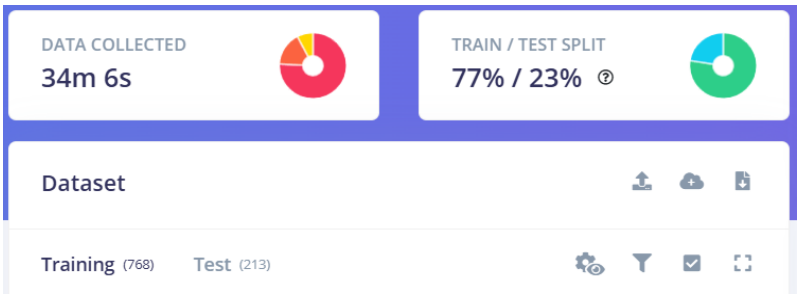


Figure: Collected Data

3.3 Model Training:

The model was trained using KNN classifier machine learning model in Edge Impulse to distinguish between the engine sounds and human voices. I focused on feature extraction techniques such as Mel-frequency cepstral coefficients (MFCC) to represent sound patterns. The Datasets were split in 80/20 percent for training and testing.

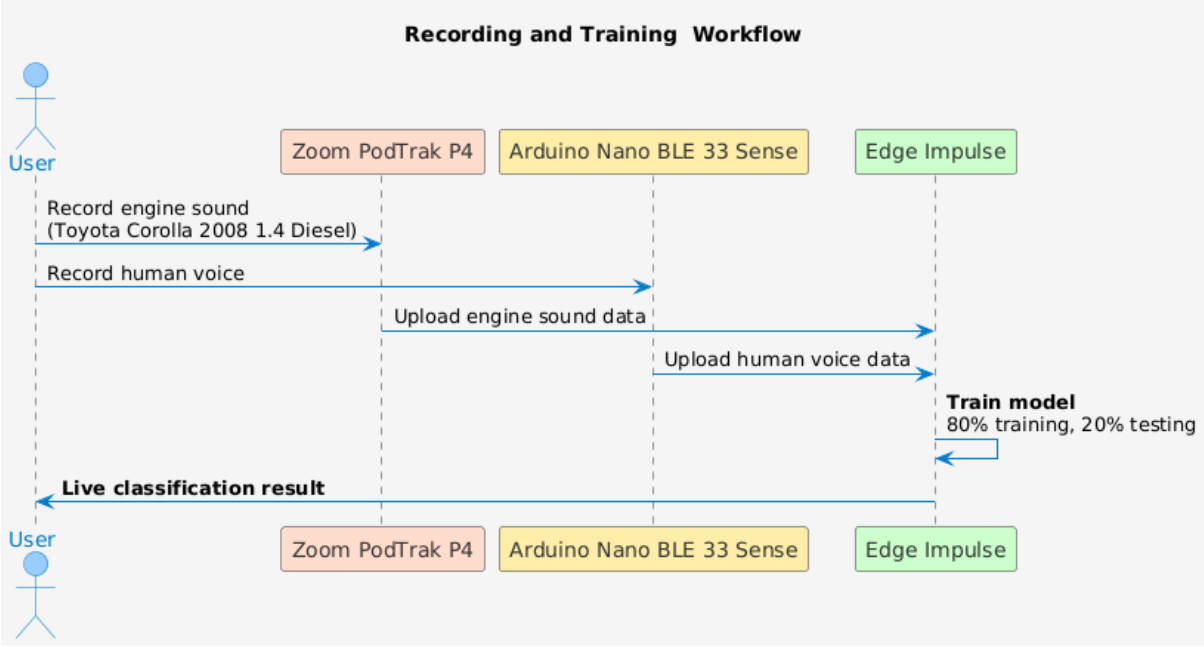
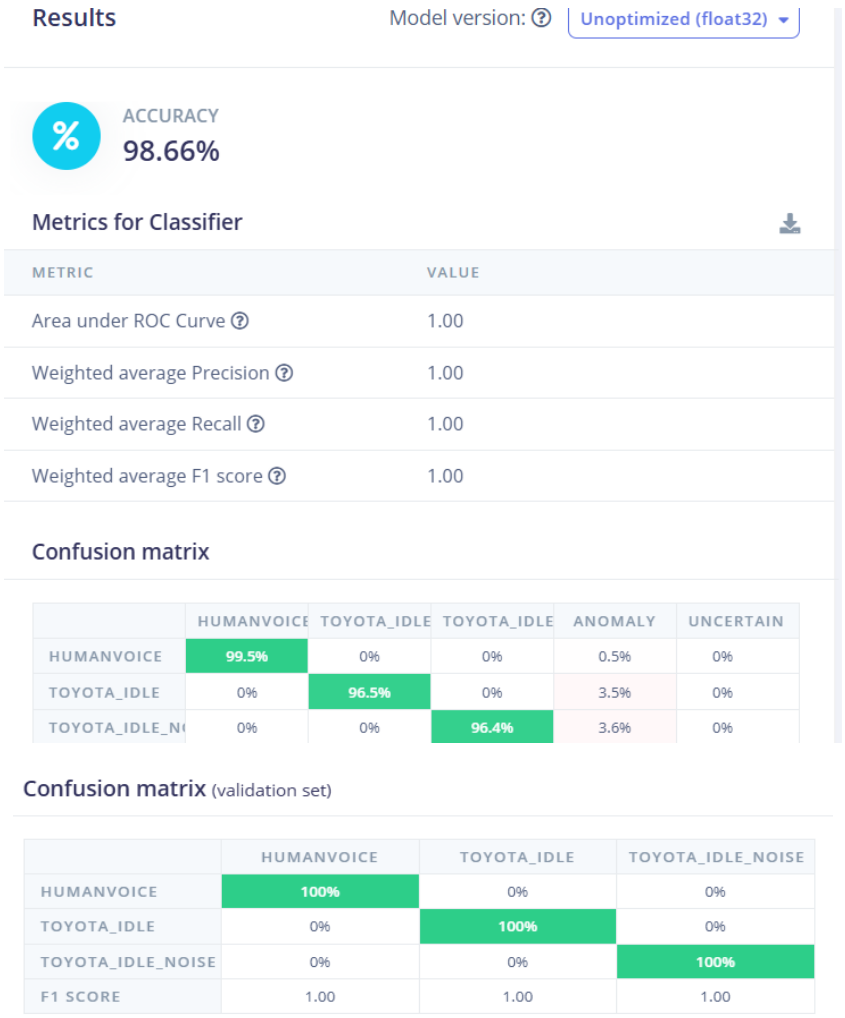


Figure: Workflow Diagram

4. **Results**:
   The model was able to identify Toyota diesel engine sounds with a certain level of accuracy. The human voices were classified separately from the engine sounds. The system performed well in controlled environments, but its accuracy was slightly affected by background noise.



Results

Model version: Unoptimized (float32)

ACCURACY
98.66%

Metrics for Classifier

METRIC	VALUE
Area under ROC Curve	1.00
Weighted average Precision	1.00
Weighted average Recall	1.00
Weighted average F1 score	1.00

Confusion matrix

	HUMANVOICE	TOYOTA_IDLE	TOYOTA_IDLE	ANOMALY	UNCERTAIN
HUMANVOICE	99.5%	0%	0%	0.5%	0%
TOYOTA_IDLE	0%	96.5%	0%	3.5%	0%
TOYOTA_IDLE_N	0%	0%	96.4%	3.6%	0%

Confusion matrix (validation set)

	HUMANVOICE	TOYOTA_IDLE	TOYOTA_IDLE_NOISE
HUMANVOICE	100%	0%	0%
TOYOTA_IDLE	0%	100%	0%
TOYOTA_IDLE_NOISE	0%	0%	100%
F1 SCORE	1.00	1.00	1.00

Figure: Results

5. **Implemented Frameworks/Tools**:

   5.1 Edge Impulse: Edge Impulse is a leading platform for developing edge AI solutions. It provides an end-to-end framework for collecting data, training machine learning models, and deploying them to edge devices.

   5.2 Arduino: The Arduino Nano 33 BLE Sense is a compact development board equipped with a powerful nRF52840 microcontroller and a variety of sensors, including those for motion, temperature, humidity, barometric pressure, and light. It supports Bluetooth 5.0 Low Energy, making it ideal for IoT and wearable applications.

6 Evaluation:

The results were promising, with the model achieving good classification accuracy in most cases. However, background noise and variations in human voices slightly reduced its accuracy, suggesting the need for further data preprocessing or noise filtering techniques. This study helped me in learning how to work with Edge Impulse and Arduino for real-time applications. To improve further, I plan to focus on noise reduction techniques and more advanced machine learning models to improve classification accuracy in real-world scenarios. A future work on this could involve studying if the models could be used to identify different states of different engine models.

This internship has given me a practical understanding of how machine learning and sensor networks can be used to solve problems in real-time systems, which will be valuable for my career in technology and engineering.

7 Conclusions:

The project successfully demonstrated the feasibility of identifying Toyota diesel engine sounds using edge computing model on an Arduino, and distinguishing it from human voices. While the system performed well in controlled environments, there is potential for improvement in real-world applications where background noise is more prevalent. Some recommendations could include-

- Implementing additional noise filtering techniques to improve performance in noisy environments.
- Exploring other machine learning models to enhance classification accuracy.
- Extending the system to monitor other vehicle sounds for diagnostic purposes.
- Testing the system in real-world driving conditions to evaluate its robustness.

8 References

- Edge Impulse Documentation (https://www.edgeimpulse.com/docs).
- Arduino Documentation (https://www.arduino.cc/).