TinyML for

SOBRIETY TEST

Ivan Ruby, Ivan Balsa and Salomão David

TinyML

Table of Contents

1. Introduction

Over time, excessive alcohol use can lead to the development of chronic diseases and other serious problems including death and disability. Machine learning (ML) - research driven interventions leveraging smart-breathalyzer have been conducted in recent years, aiming to help reduce death and disability. In this project we aim to develop a digital phenotype of long-term smart-breathalyzer behavior to predict individuals' alcohol concentration based on the voice pattern submitted to an Arduino.

To do so, we will make use of an existing corpus of voice clips collected from sober as well as intoxicated individuals. Given the limited availability of datasets with English speech samples of intoxicated individuals, we will collect and process audio samples from publicly available clips on the internet.

Although we acknowledge the concerns with accent variations, age and cultural differences in language, we are optimist about the accuracy of our model in classifying voice samples as drunk or not drunk. Our working hypothesis is that since we will only have two classes (drunk and not drunk) the variations in the spoken language will be less impacting. This is a hypothesis we will test during the implementation of the project and will document its results.

2. Project Objectives

With COVID-19 we have seen several changes to old practices, being some of them social distancing, wearing masks and glove a standard practice. The alcohol and drug testing needs to adjust to protect both the test subject and the tester. Since COVID-19 predominantly spread through droplets from coughing and sneezing, to avoid forceful exhales breathalyzers have been

recognized as potential agent for transmission. Hence, this project objective is to assess and evaluate the level of an individual sobriety or intoxication level as alternative to breathalyzers.

The project aims to *provide a quick and non-intrusive method to guarantee security and wellbeing in public and private spaces*. Using this self-operated system, a user can be prompted to read a test phrase out loud and, based on the characteristics of the voice captured, the intoxication level can be predicted.

3. Implementation

1. Block Diagram

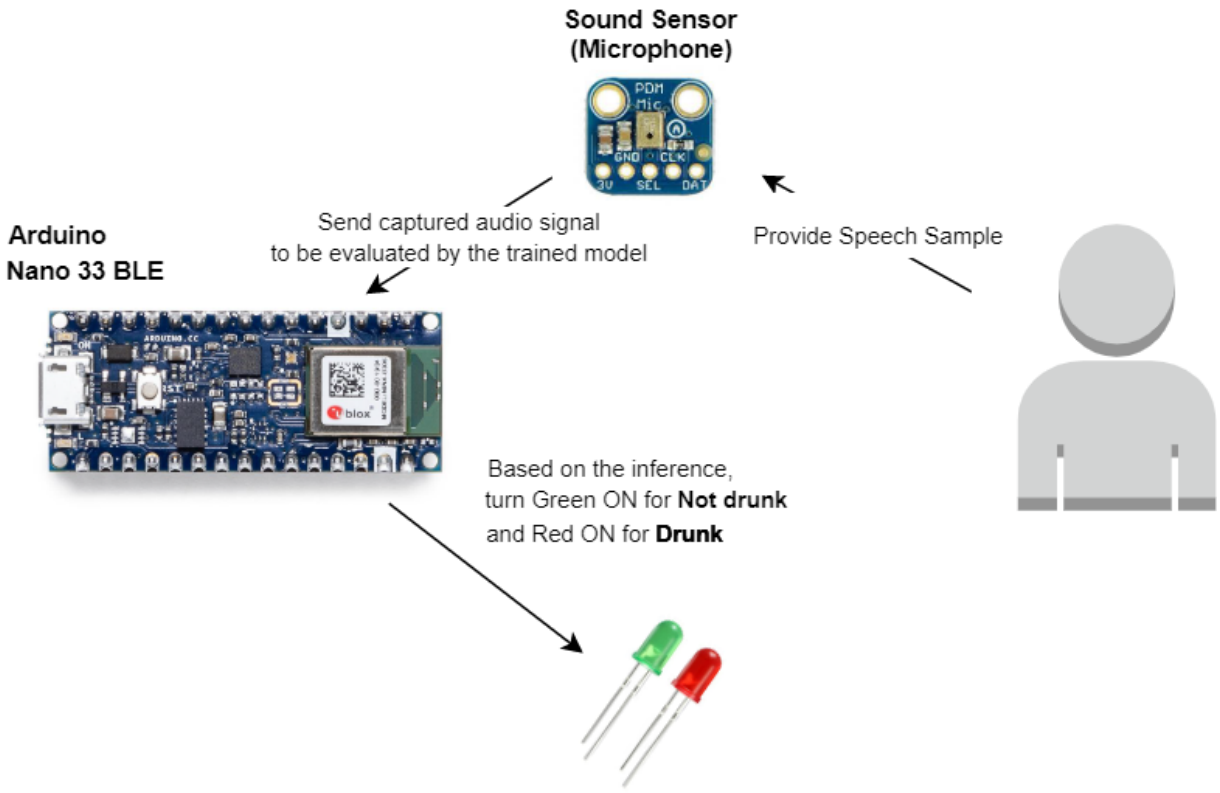


Figure 1. Circuit diagram

2. Hardware

- Arduino Nano 33 BLE Sense
- PDM MEMS Microphone
- 2x LEDs (Green and Red)
- Miscellaneous: Breadboard, resistors, connecting cables

3. Data collection

In our initial exploration of available data, we identified a dataset (Alcohol Language Corpus) containing samples of 162 speakers, either sober or with a self-chosen level of intoxication [1]. Although initially promising, this collection of audio samples costs EUR1200 and only contains samples of German speakers. For this project, we would like to train and test our model with English samples and we have limited financial resources. Therefore, we had to discard this alternative and continue our search. Furthermore, we could not identify a comparable dataset in English but found publications and projects with a similar goal where authors described how they circumvented the dataset availability limitation. In [2], the authors decided to self-source samples of intoxicated individuals in two ways. First, they captured speech samples from individuals after consuming a set quantity of alcohol. Next, they extracted audio samples from public videos and audios with speakers deemed to be intoxicated with alcohol.

For our project, we did not find it feasible to collect speech samples from individuals. This is so, given the constraints with time and finding individuals willing to provide these samples. Therefore, we embarked on a journey to collect audio samples from multiple media publicly available on the internet.

We searched the popular video platform Youtube (https://www.youtube.com) for videos labelled with combinations of the following keywords: drunk people speaking interview. Doing

so, we were able to create a list of candidate videos from which we filtered only those that fit our inclusion criteria:

- Adult speakers
- Minimal cross-talking (people talking over each other)
- No expletives (curse words)
- Only one sample per individual (avoid multiple samples from the same individual)
- 10 clear seconds of same individual speaking

From this process, we were able to collect 23 videos from which a 10-second clip of a single speaker was extracted. This way, we were able to collect **23 audio samples of 10 seconds each**.

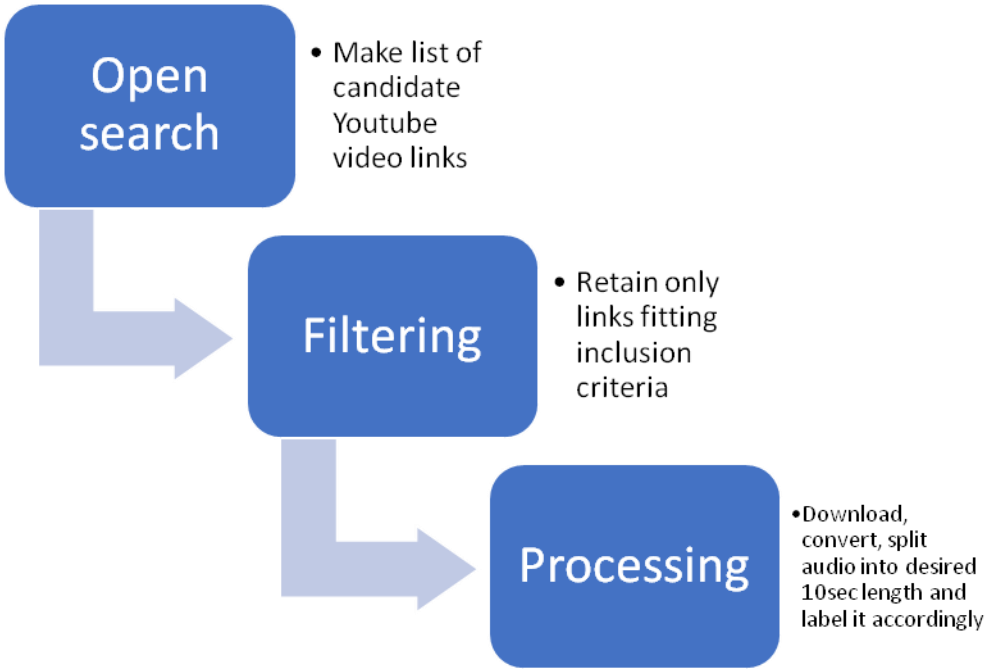


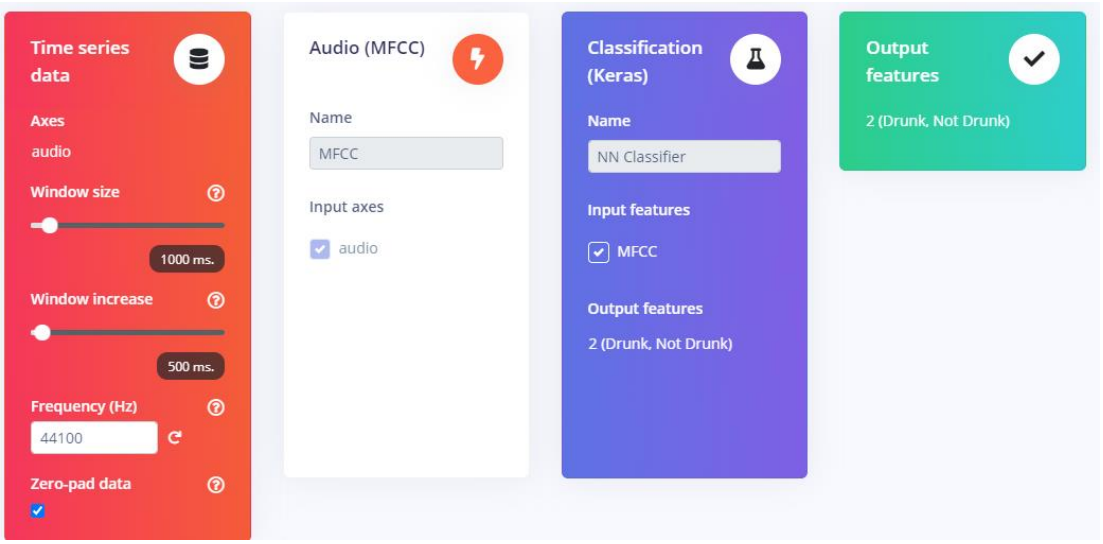
Figure 2. Summary of data collection process

To complement our dataset with samples of non-intoxicated speakers, we combined the earlier mentioned approach of collection and extraction of audio samples from Youtube videos (with regular conversations) with an existing public speech dataset LibreSpeech (https://www.openslr.org/12). According to the authors' website, the dataset consists of a "corpus of approximately 1000 hours of 16kHz read English speech". This way, we could have a

combination of slow-paced speech (LibreSpeech dataset) as well as more naturally paced speech.

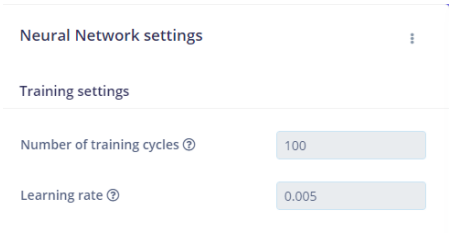
4. Model design

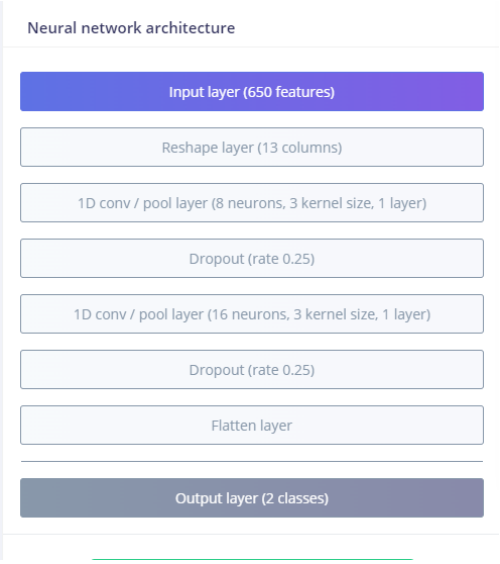
The goal for our model is for it to be able to assess whether a given sample can be classified as an alcohol-intoxication state: drunk or not drunk. Therefore, our design will consist of a Classification task, where the expected output is a label identifying a group associated with the training data.



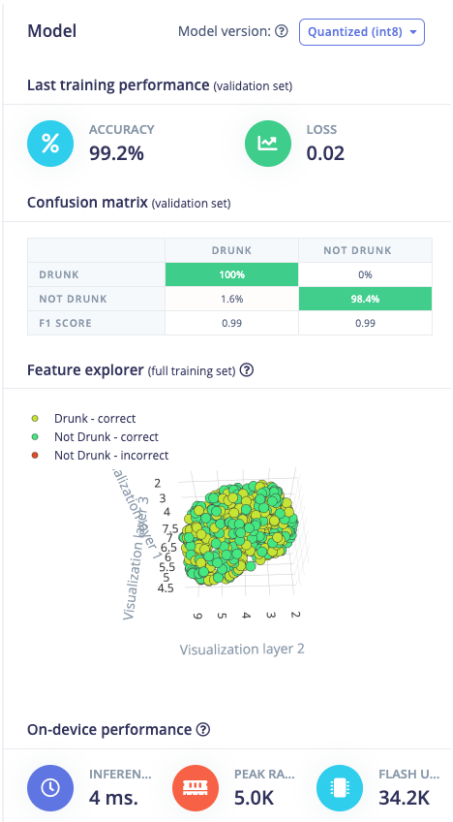
We decided to use this disposition to our project. The layer of Audio (MFCC) works so good with human voices that is what we want. Then we extract the features:

Then we selected our Neural Network settings to start training the program.



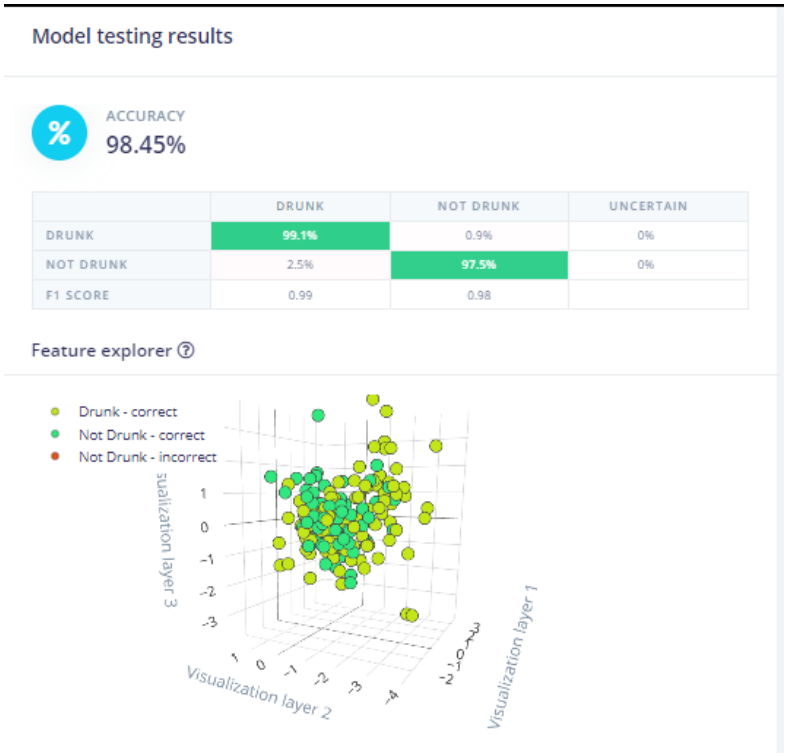


After this we will make the first train for our project:



We can see an accuracy of 99.2% and loss of 0.02. The prediction for drunk audios has a 100% of accuracy, and the accuracy for not drunk audios is also good (98,4%).

Finally, we just need to do the test of our program:



We obtained a 98,45% of accuracy on the model testing results what are a so good result.

5. Conclusion

To sum up, our goal was to create a model able to differentiate voices of drunk and non-drunk speakers. Checking the confusion matrix, we can see a good value for the accuracy (98.45%) what tells us that the result was satisfactory. However, we would need more work to make this model 100% functional. As we said before there are many different accents all around the world. An example is the group that worked on the project, where our accents of English have considerable variations.

To improve the model, a larger dataset would be needed. One that not only has quantity of samples but also variety of accents.

Nonetheless, we met the target and we consider the final result appropriate for the amount of data that we were able to collect.

6. References

[1] Schiel, F., Heinrich, Chr., Barfüßer, S., Gilg, Th. (2008). ALC - Alcohol Language Corpus. In: Proc. of LREC 2008, Marrakesch, Marokko.

[2] Miller, Joshua, Jillian Donahue, and Benjamin Schmitz. "Speech emotion and drunkenness detection using a convolutional neural network." Retrieved from http://www2.ece.rochester.edu/~zduan/teaching/ece477/projects/2018/JoshuaMiller_JillianDonahue_BenjaminSchmitz_ReportFinal.pdf