TinyML for

SOBRIETY TEST

Ivan Ruby, Ivan Balsa and Salomão David

TinyML

1. Table of Contents

2. 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.

3. Project name

The project here proposed has the name *Sobriety test with TinyML*.

4. Team Members

The team is comprised by Ivan Gonçalves Ruby from the Concordia University, Ivan Balsa from the University Carlos III de Madrid and Salomão David from the International Centre for Theoretical Physics.

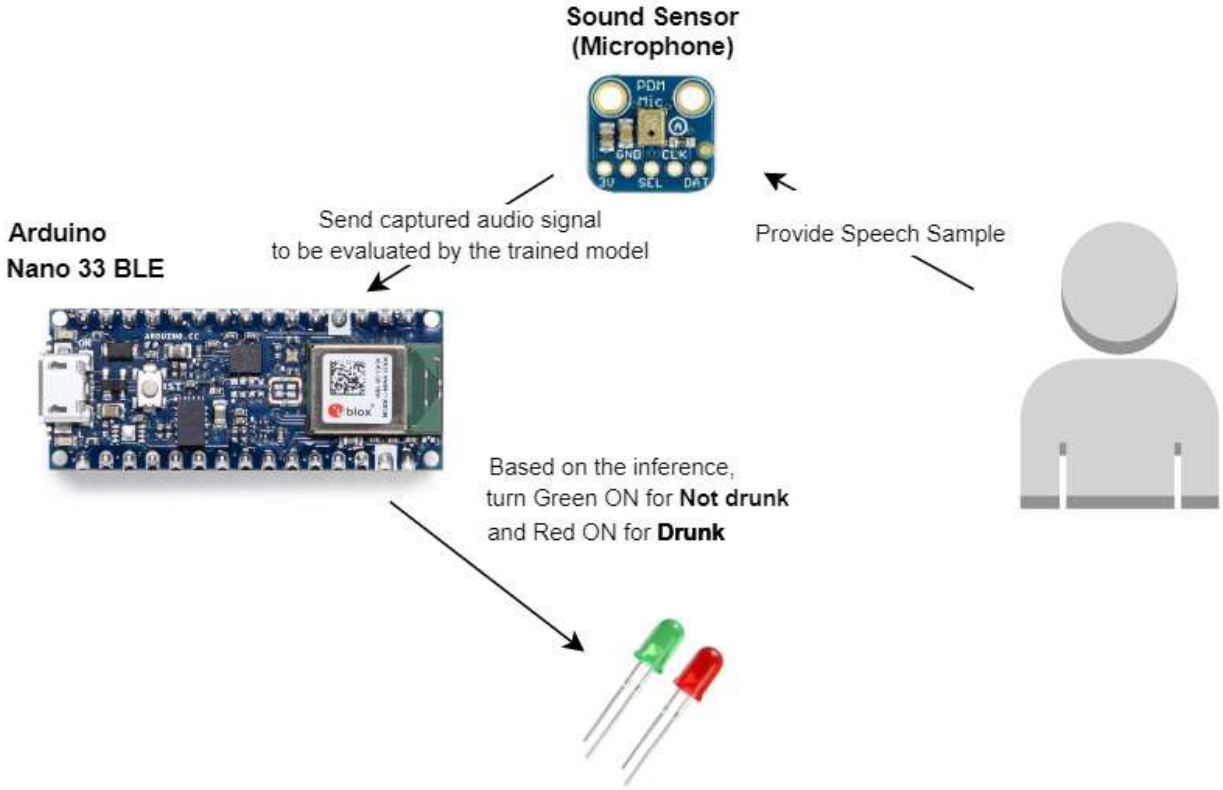
5. Project Objectives

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.

6. Longer project description

1. Block Diagram



2. Hardware to be utilized

- 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 will extract audio samples from multiple videos and audios available on the internet.

4. Preprocessing

Given the nature of the samples we will collect, we will have to analyze, label and process the audio files. The analysis and labelling process will consist of identifying whether the speaker is intoxicated or not as well as collecting metadata from the sample (gender, age, race). When processing the files, we will ensure all audio files have the same, preset length or duration to ensure consistency.

5. 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.

6. Optimizations

In Machine Learning, the purpose of optimization is to adjust hyperparameters so as to minimize the cost function. We will start by using the Stochastic Gradient Descent and Adam optimizers since these are the ones we have been more exposed during the class activities.

7. In system inference

To make our solution portable, we will deploy the trained model on the microcontroller. We will use the Edge Impulse platform to create a library of the model to be used in an Arduino platform. This will be done by converting the initial codebase into an Arduino Sketch that we will transfer to our Microcontroller. At this point, we do not have metrics on how much space our model may occupy as well as the time it will take to create an inference. We expect to have this data in the very near future.

8. Issues or roadblocks the team envision and potential solutions

Our initial concern was regarding data collection given the difficulty in finding existing audio samples of intoxicated English-speaking individuals. As described earlier in this document, our solution will be to source samples from public media and appropriately label them.

Given our reduced experience and exposure to a project of this nature, we also anticipate concerns with tools and processes, identifying the appropriate parameters and ensuring we will use a correct approach in the design of the model. For this point, we consider practice as the only remedy. Similarly, we will study other existing projects to understand what, how and why they made certain decisions that can be useful in our project.

9. The top unresolved question(s) the team have at this point

At this point, our main unresolved question is regarding the impact of the characteristics of the speakers we will have samples of. Since English is spoken all over the world, used as a *lingua-franca*, we might have speakers with different native languages, accents and other characteristics that may affect the accuracy of our model. Furthermore, we are unclear whether there should be a distinction between male and female speakers.

10. 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