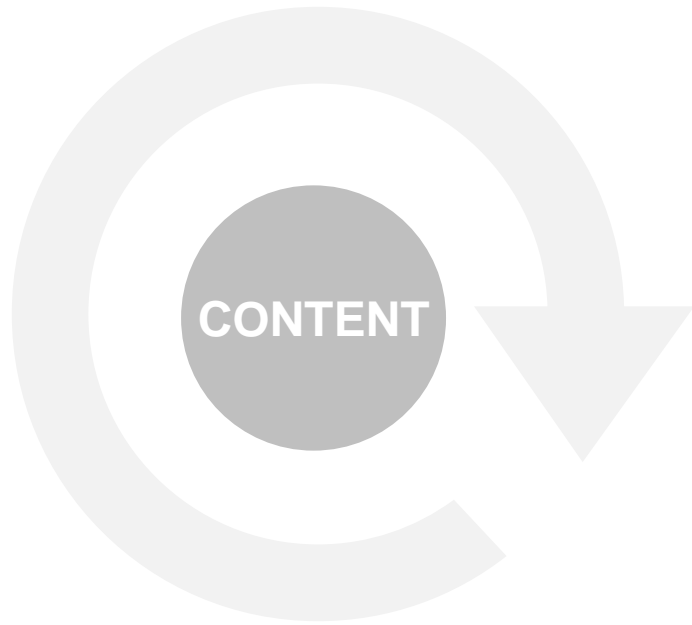# Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data

Chloë Brown, Jagmohan Chauhan, Andreas Grammenos, Jing Han, Apinan Hasthanasombat,

Dimitris Spathis, **Tong Xia,** *Pietro Cicuta, Cecilia Mascolo*

Computer Science and Technology Department

UNIVERSITY OF CAMBRIDGE

European Research Council
Supporting top researchers from anywhere in the world

CONTENT

1
Motivations

2
Date collection

3
Results

4
Discussion

01
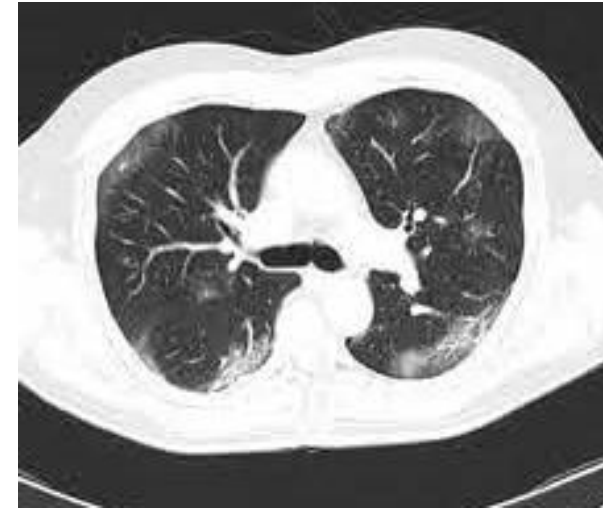
Motivation

What is the research question?

Why is it promising?

Covid-19 Diagnosis

☐ **Accessibility:** this test still requires an office visit and thus breach of social distancing and self-isolation[1].

CT Scan

RT-PCR

☐ **Accessibility of X-ray or CT Scan:** it needs a radiologist to perform the diagnosis, and still requires a visit to a well-equipped clinical facility [2].

[1] Kucirka, Lauren M., et al. "Variation in false-negative rate of reverse transcriptase polymerase chain reaction–based SARS-CoV-2 tests by time since exposure." *Annals of Internal Medicine* (2020).
[2] Imran, Ali, et al. "AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app." arXiv preprint arXiv:2004.01275 (2020).

Covid-19 Sounds

*A **non-invasive and ubiquitous testing protocol**, which would allow individual prescreening 'anywhere', 'anytime', in real-time, and available to 'anyone'.*

IS IT POSSIBLE TO DETECT COVID FROM HUMAN SOUNDS?

- ✓ Lung is the main organ involved and infected by virus, which leads to some changes in respiratory sounds[3].

- ✓ Audio-based methods are promising in detecting cough-related disease like pertussis[4], croup[5], and tuberculosis[6].

[3] Hui Huang, et al. The respiratory sound features of COVID-19 patients fill gaps between clinical data and screening methods[J]. medRxiv, 2020.
[4] R. X. A. Pramono, et al, "A cough-based algorithm for automatic diagnosis of pertussis," PloS one, vol. 11, no. 9, 2016
[5] R. V. Sharan, et al "Automatic croup diagnosis using cough sound recognition," IEEE Transactions on Biomedical Engineering, vol. 66, no. 2, pp. 485–495, 2018.
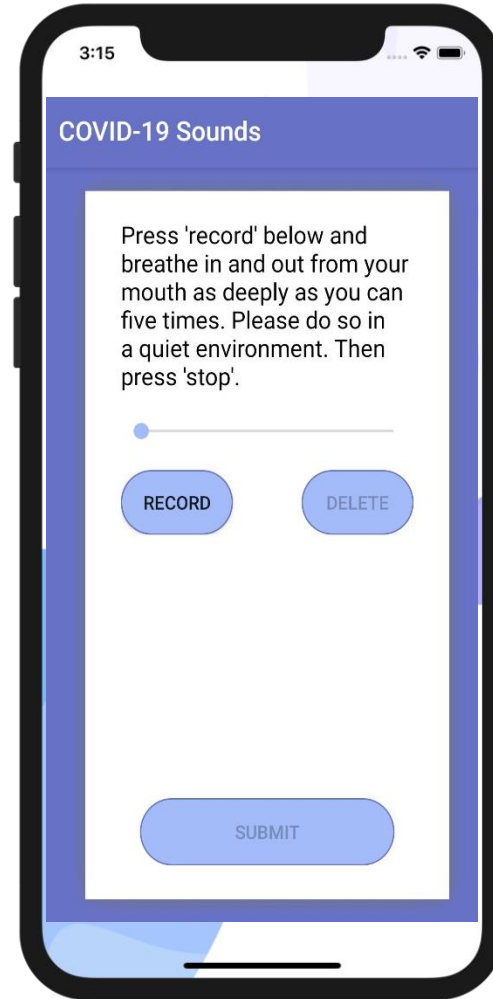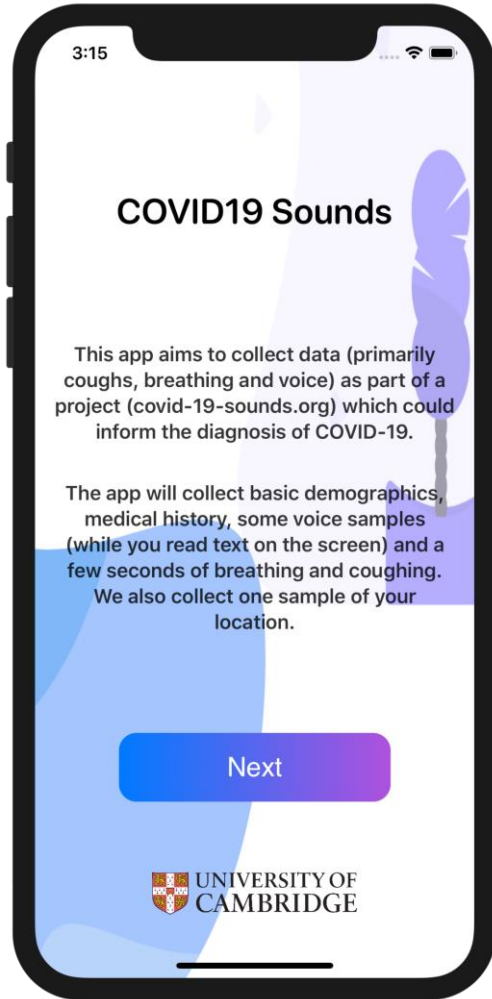[6] G. Botha, et al, "Detection of tuberculosis by automatic cough sound analysis," Physiological measurement, vol. 39,no. 4, p. 045005, 2018

What we do

IS IT EASY TO DIFFERENTIATE COVID FROM THE HEALTHY BY SOUNDS?

✓ *Complex latent sounds features!*

Harness the power of big data and machine learning to enable smartphone-based screening and diagnosis.

02

Data Collection

We develop an application

Date collection App

Sign-up survey

Demographics, Medical history, Smoking history，etc.

Daily survey

✓ **Symptoms**: fever, loss of taste, etc.
✓ **Testing** state: positive, negative, never tested, etc.
✓ **Sounds:** breathe，cough and some voice.

COVID-19 Sounds App - University of Cambridge (covid-19-sounds.org)

Dataset

30,000 audio recordings (and counting!)

November 1, 2020 — As of today, the **COVID-19 Sounds study app** has more than 16,000 people contributing over 30,000 audio recordings and symptoms to the fight against COVID! This makes our study one of the biggest crowdsourced health sound projects in the world. Thanks to all of you who are contributing to cutting edge AI-driven COVID research!
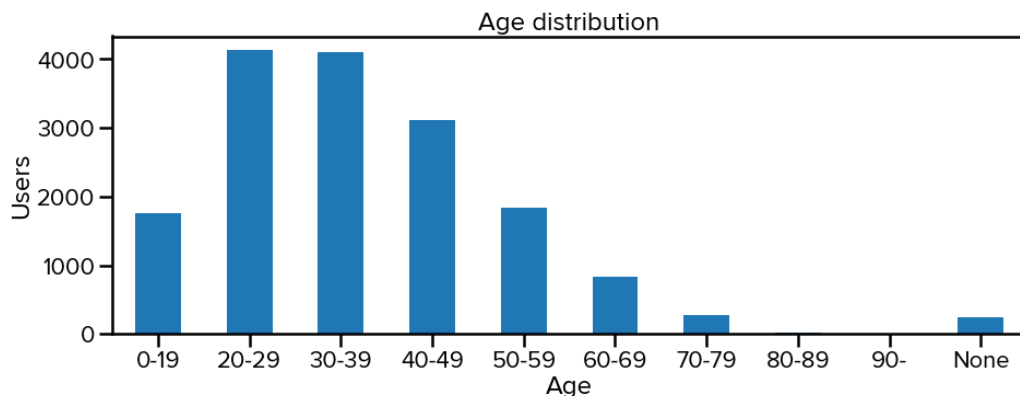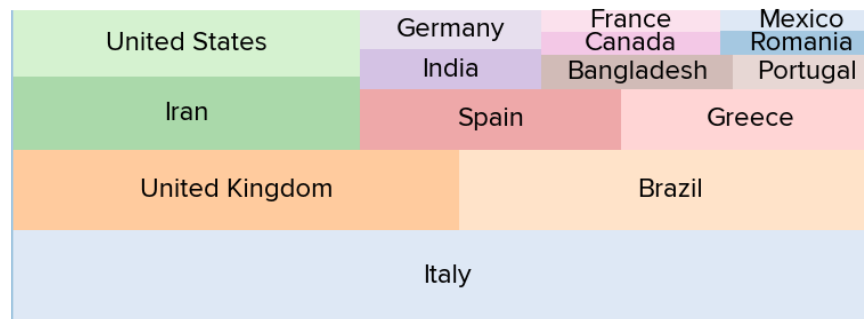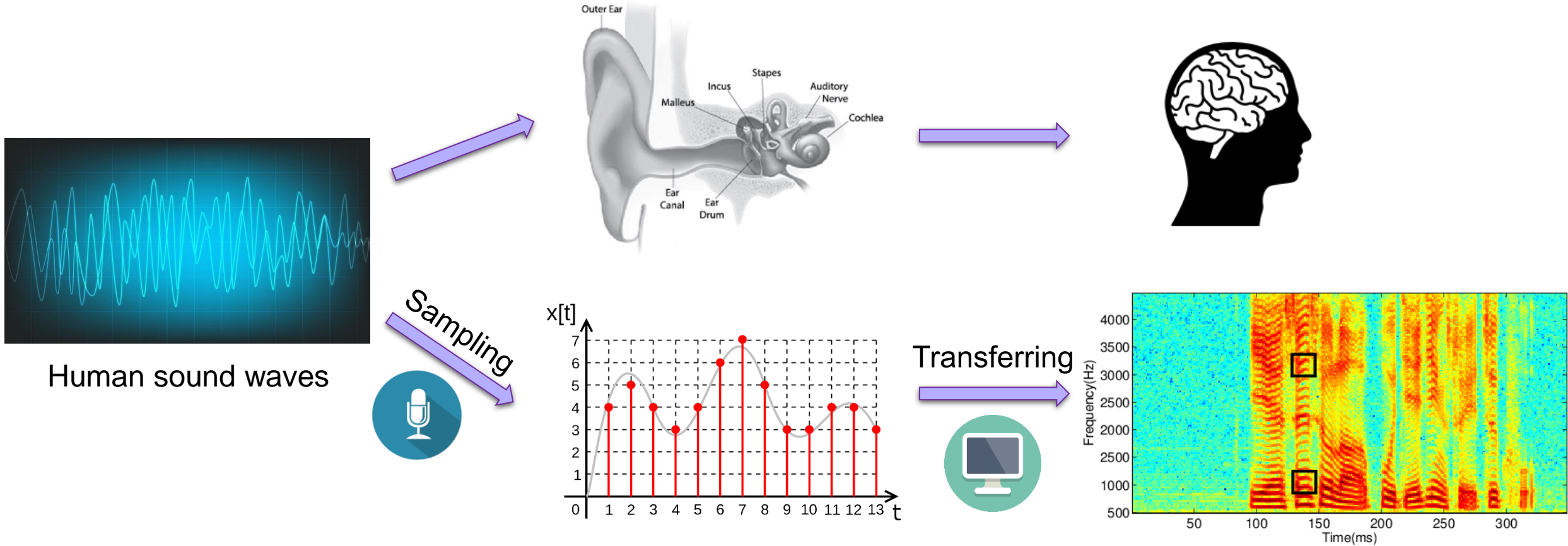

Age distribution

Sex distribution

Smoking status

COVID-19 testing status

9
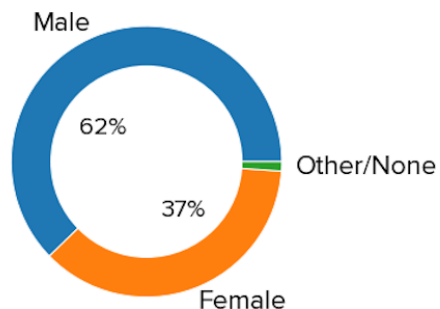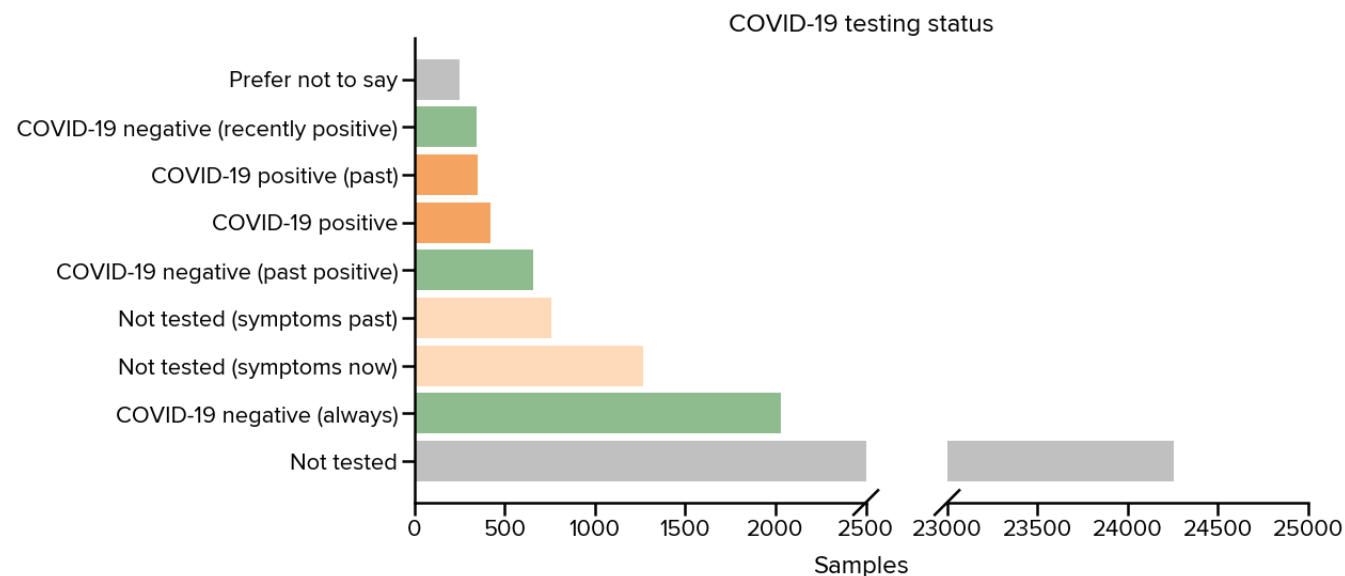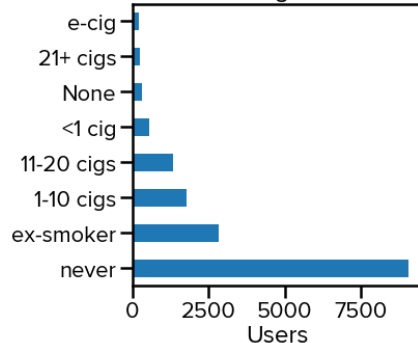
03

Results

What method do we use

How about the performance

How computer understand sounds?

Human sound waves

Sampling

Transferring
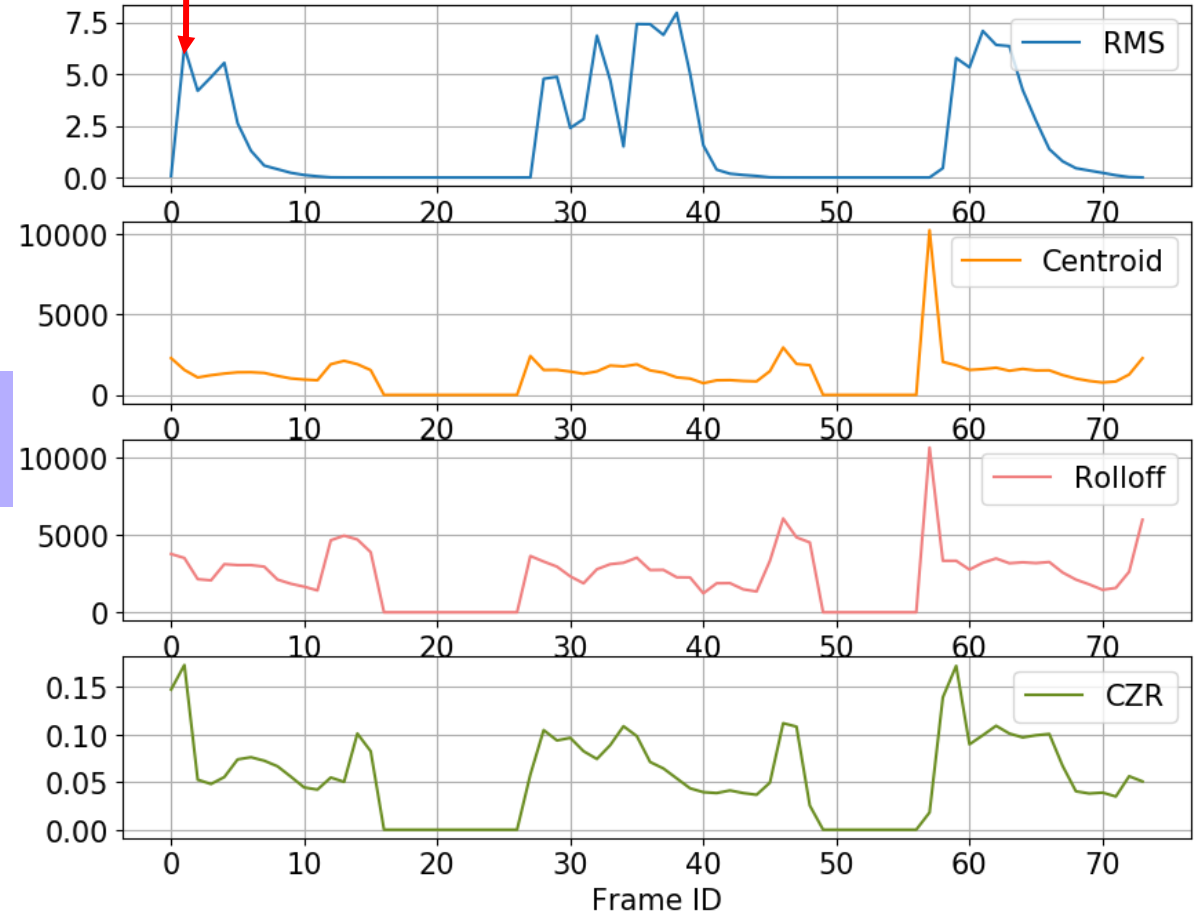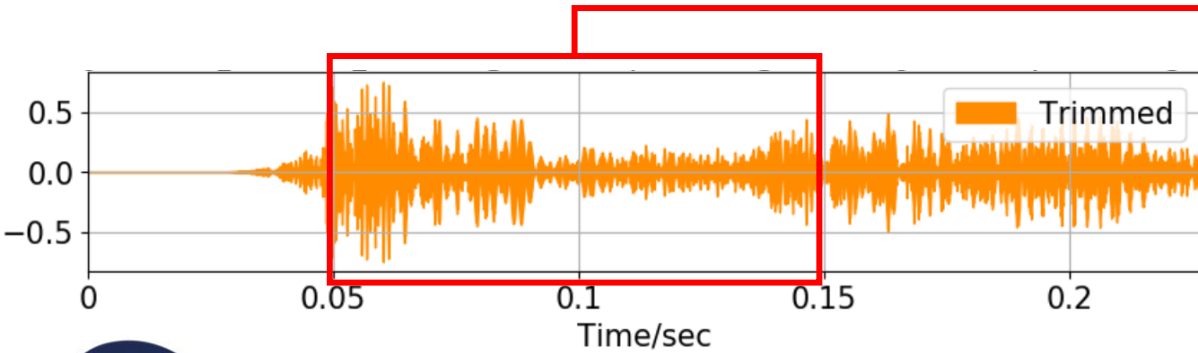
www.nidcd.nih.gov/health/how-do-we-hear
www.analog.com/en/design-center/landing-pages/001/beginners-guide-to-dsp.html
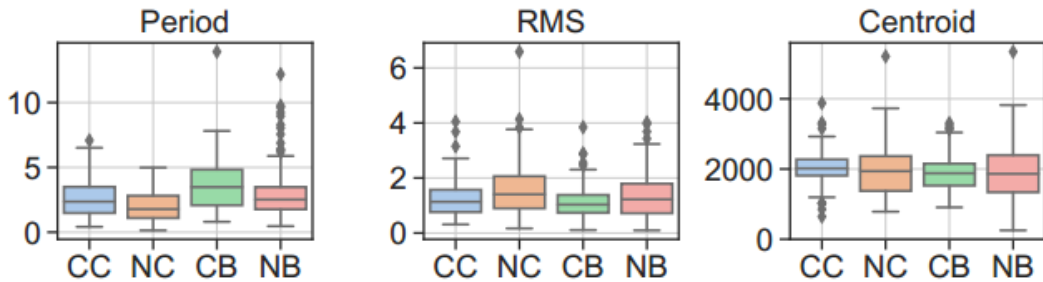
What features we use?

Indicators showing the features of a cough w.r.t. energy distribution, central frequency, etc.
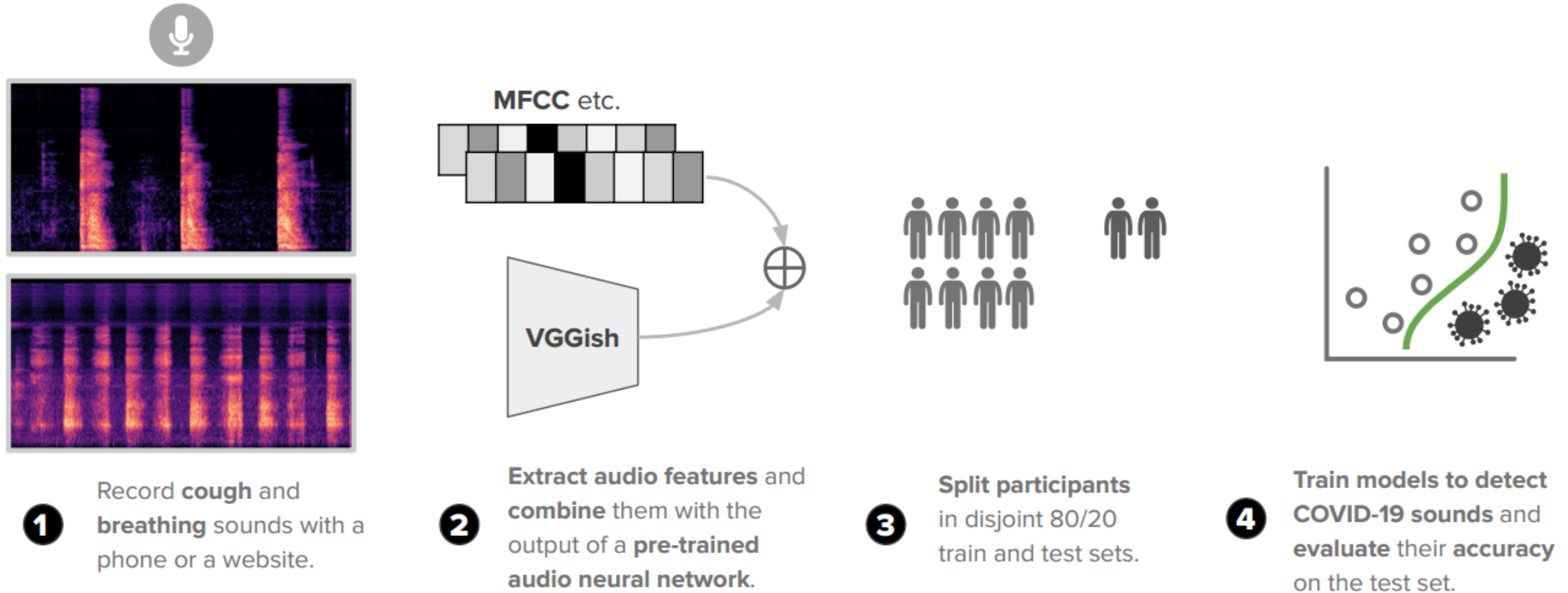
CC&CB : COIVD   NC&NB: Non-COVID

We can see difference when comparing COVID and Non-COVID Cough&Breathe in our data!

Model Framework

1 Record **cough** and **breathing** sounds with a phone or a website.

2 **Extract audio features** and **combine** them with the output of a **pre-trained audio neural network**.

3 **Split participants** in disjoint 80/20 train and test sets.

4 **Train models to detect COVID-19 sounds** and **evaluate** their **accuracy** on the test set.

Preliminary Results

- ☐ 800+ positive and 1900+ negative users
- ☐ More than 4000 samples in total
- ☐ Well-controlled group in demographics
- ☐ Change the ratio of C/N closed to real-life

	AUC	Sensitivity	Specificity
Balanced group	0.77	0.75	0.67
Prevalence = 10%	0.74±0.12	0.73± 0.17	0.62±0.09
Prevalence = 5%	0.72±0.18	0.72±0.20	0.60±0.10
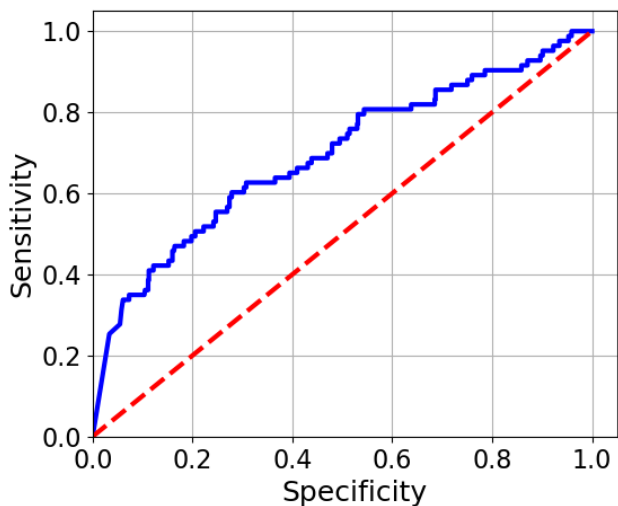
✓ Gain insights about how COVID effect human sounds
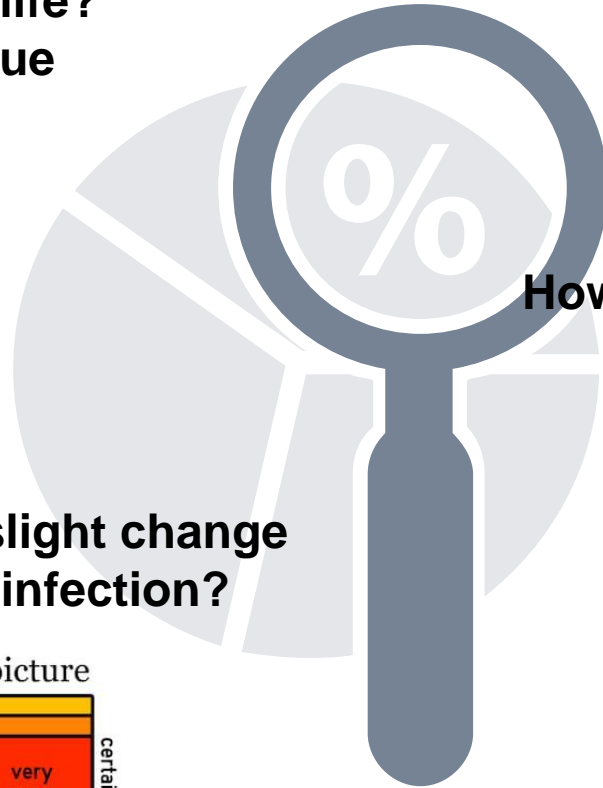✓ As a promising tool for ubiquitous pre-screening
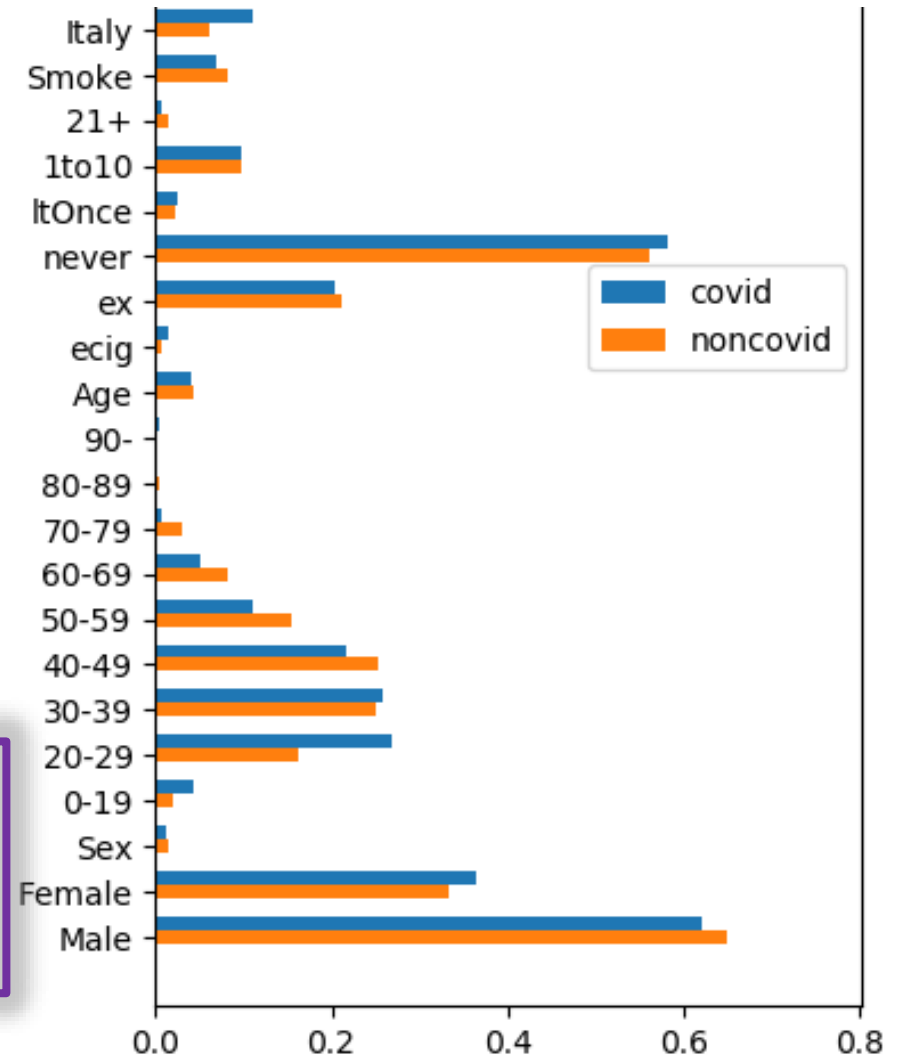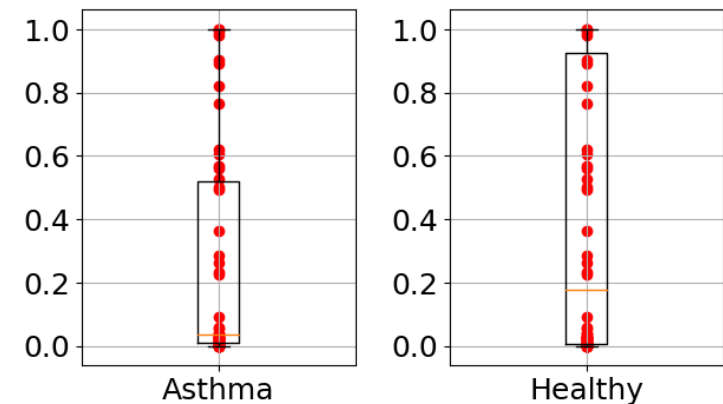
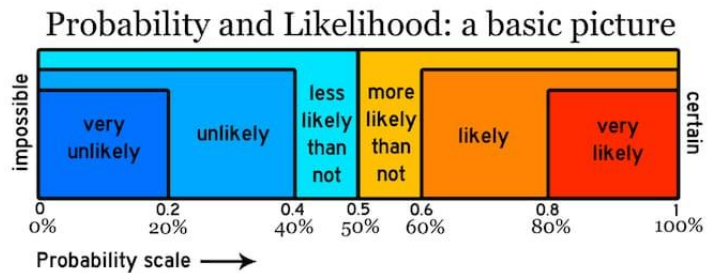# 04

Discussion

Any other question

We are working on

How robust is our model in real-life?
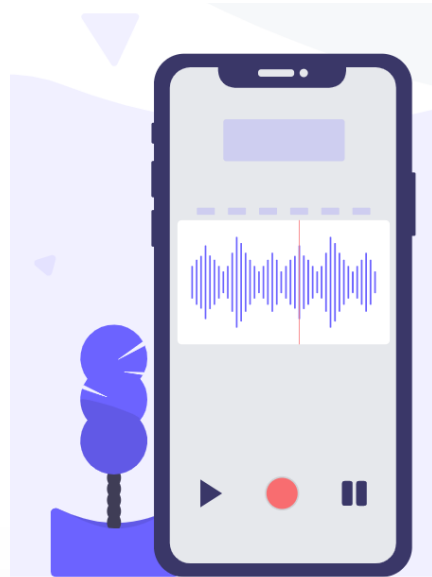**-- How to handle data quality issue**

**How to demonstrate that we detect COVID
instead of other respiratory diseases?**

**Can the model understand the slight change
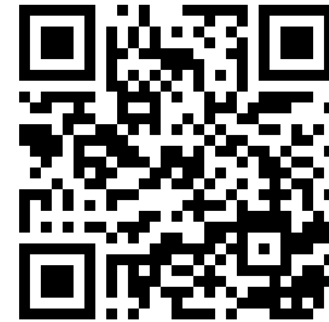in one's sound before and after infection?**

Probability and Likelihood: a basic picture

impossible | very unlikely | unlikely | less likely than not | more likely than not | likely | very likely | certain

0% 0.2 20% 0.4 40% 0.5 50% 0.6 60% 0.8 80% 1 100%

Probability scale ⟶

The probability of detecting non-COVID users into COVID

Thank You for Listening! Donate your sounds and help save lives

Tong Xia

www.cl.cam.ac.uk/~tx229/

www.covid-19-sounds.org/en/