Sounds of COVID-19: Exploring Realistic Performance of Audio-based Digital Testing

Tong Xia, *Jing Han, Dimitris Spathis, Erika Bondareva, Chloë Brown, Jagmohan Chauhan, Ting Dang, Andreas Grammenos, Apinan Hasthanasombat, Andres Floto, Pietro Cicuta, Cecilia Mascolo*

UNIVERSITY OF CAMBRIDGE

European Research Council
Supporting top researchers from anywhere in the world

William gates building, West Cambridge

Cecilia Mascolo

Full Professor of Mobile Systems in the Department of Computer Science and Technology, University of Cambridge

- Mobile health

Tong Xia

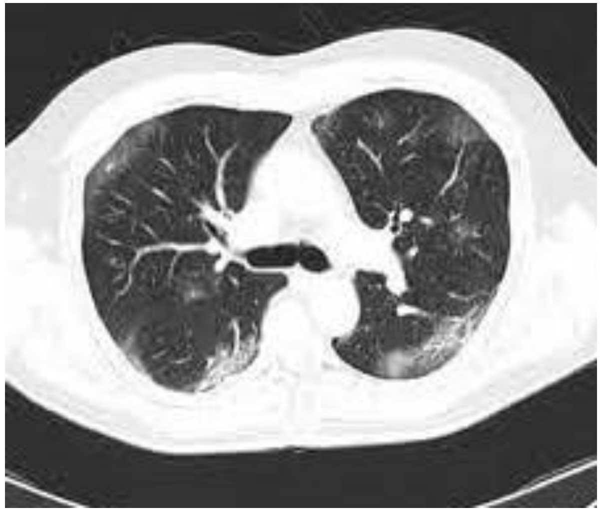
A third-year PhD student in computer science with interests in

- Trustworthy machine learning
- Deep learning for healthcare
- Respiratory health

Covid-19 Testing

- **Accessibility:** low efficient, expensive, labor-consuming

LFT, RT-PCR



CT Scan

- **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].

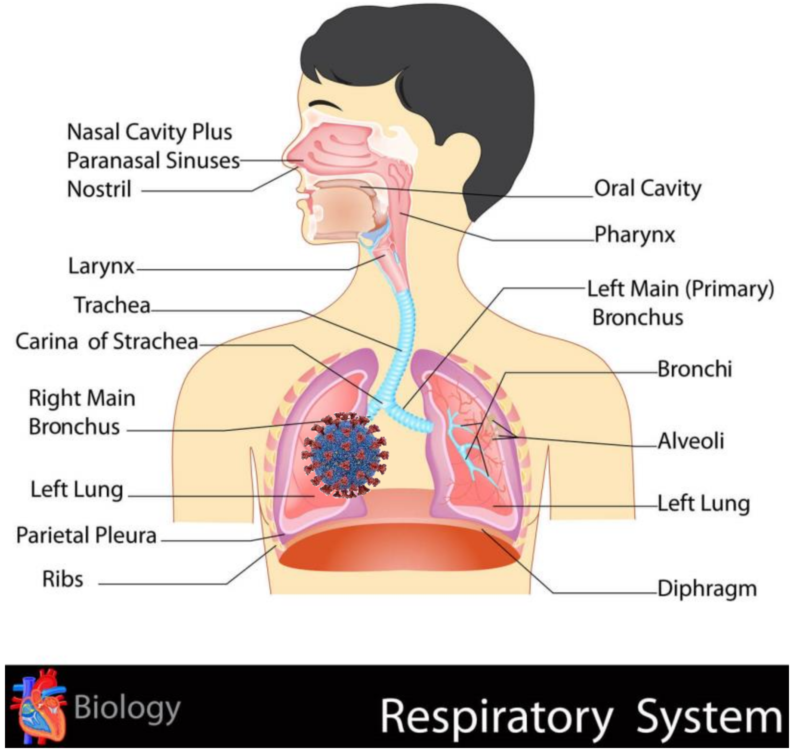
[1] Kucirka, Lauren M., et al. "Variation in false-negative rate of reverse transcriptase polymerase chain reaction–based SARS-CoV-2 tests by time since exposure." *Annals of Internal Medicine* (2020).
[2] Imran, Ali, et al. "AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples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

- ✓ 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

01

Data Collection

We develop an application

UNIVERSITY OF CAMBRIDGE

April 2020

COVID-19 Sounds App

Upload short recordings of cough and breathing and report symptoms to help researchers from the University of Cambridge detect if a person is suffering from COVID-19. Healthy and *non-healthy* participants welcome.

Cecilia Mascolo

Cecilia is Professor of Mobile Systems. She is an expert in mobile health and mobile data analysis.

Pietro Cicuta

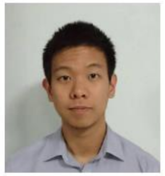
Pietro is Professor of Biological Physics at the Cavendish Laboratory, Cambridge.

Andres Floto

Andres is Professor of Respiratory Biology and Research Director of the Cambridge Centre for Lung Infection at Papworth Hospital.

Andreas Grammenos

Andreas is a PhD student in Computer Science with interests in machine learning, data science, and systems.



Apinan Hasthanasombat

Api is a PhD Student in Computer Science with interests in causal inference and its applications to systems data and design.

Dimitris Spathis

Dimitris is a PhD Student in Computer Science with interests in machine learning and health data science.

Chloe Brown

Chloe has a PhD in Computer Science and is a junior doctor.

Jagmohan Chauhan

Jagmohan is a postdoctoral researcher in Mobile Systems.

Jing Han

Jing is a postdoctoral researcher with interests in deep learning and digital health.

Tong Xia

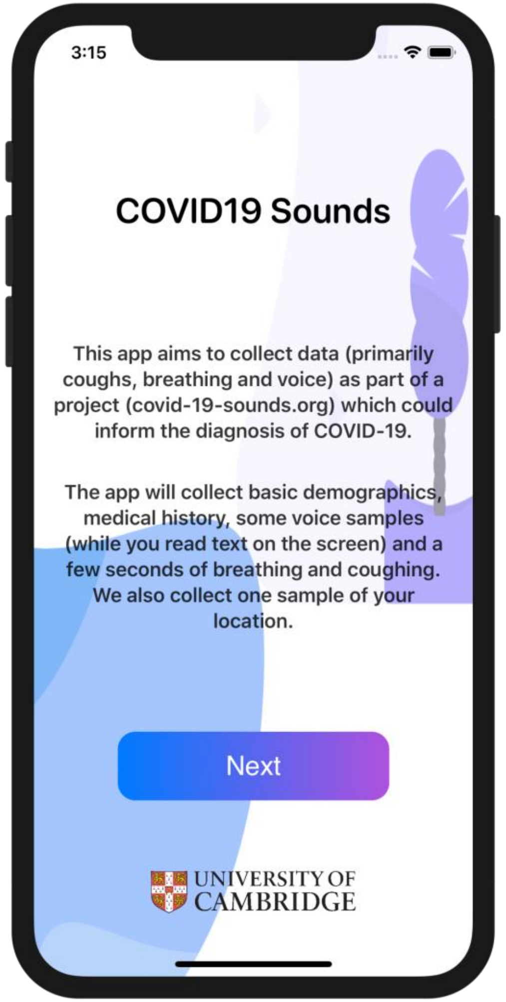
Tong is a PhD student Computer Science. She is interested in data mining and machine learning.

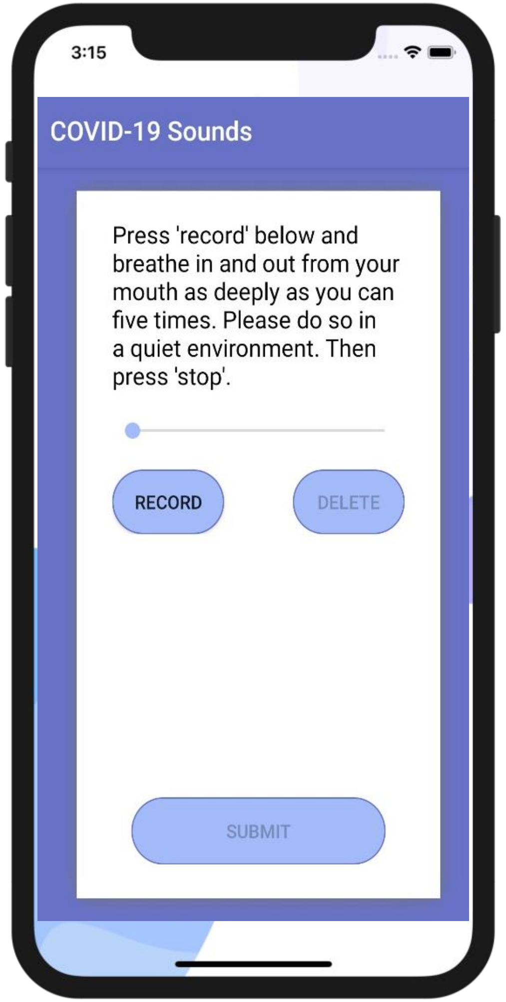
Ting Dang

Ting is a post-doctoral researcher with interests in speech-processing, affective computing, and audio based health diagnosis.

https://covid-19-sounds.org/en/index.html

Date collection App





Sign-up survey

Demographics, Medical history, Smoking history， etc.

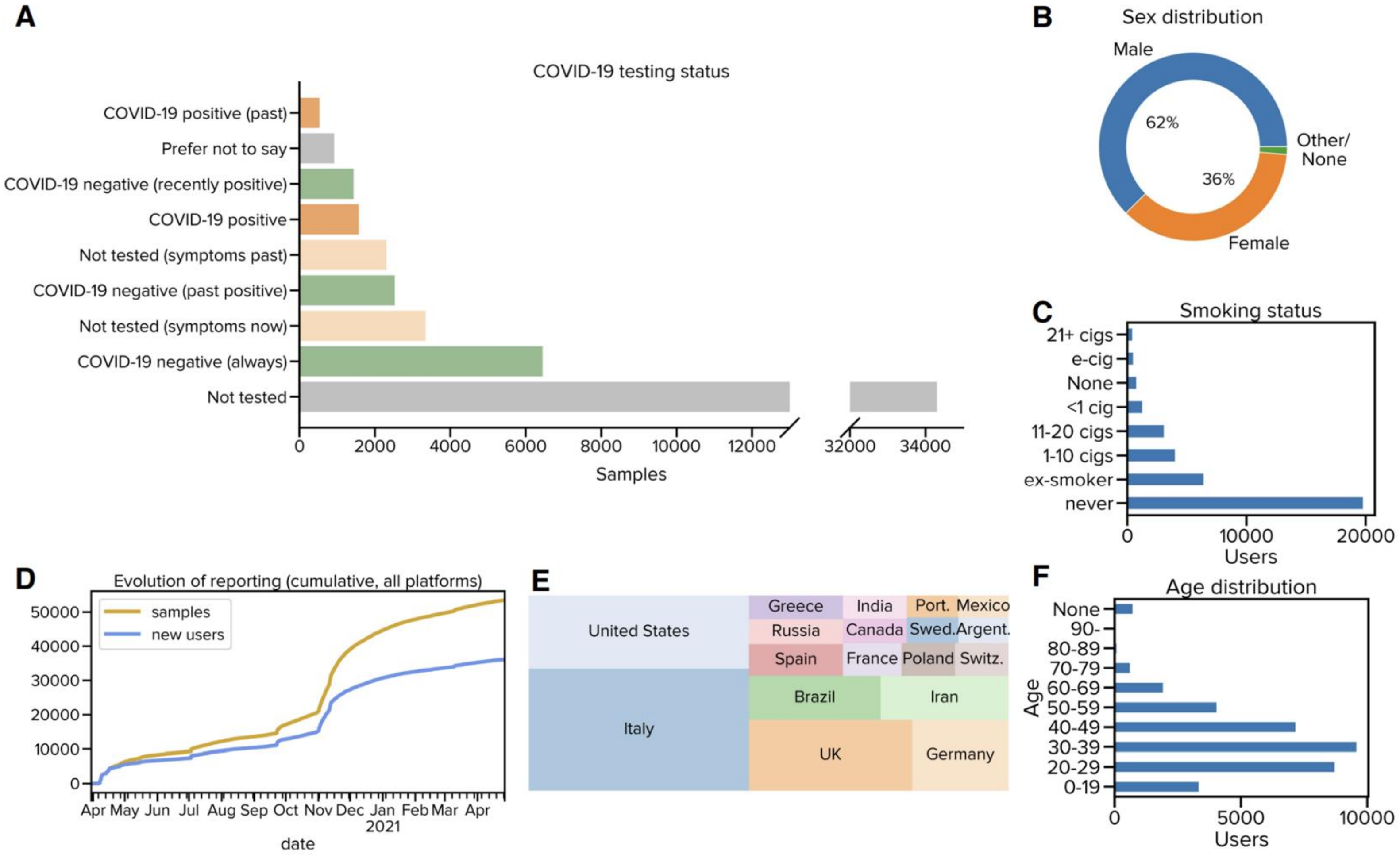
Daily survey

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

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

Dataset

- 53,449 audio samples
- over 552 hours in total
- from 36,116 users
- 2,106 positive samples



Tong et al. COVID-19 Sounds: A Large-Scale Audio Dataset for Digital Respiratory Screening. NeurIPS Dataset Track 2021

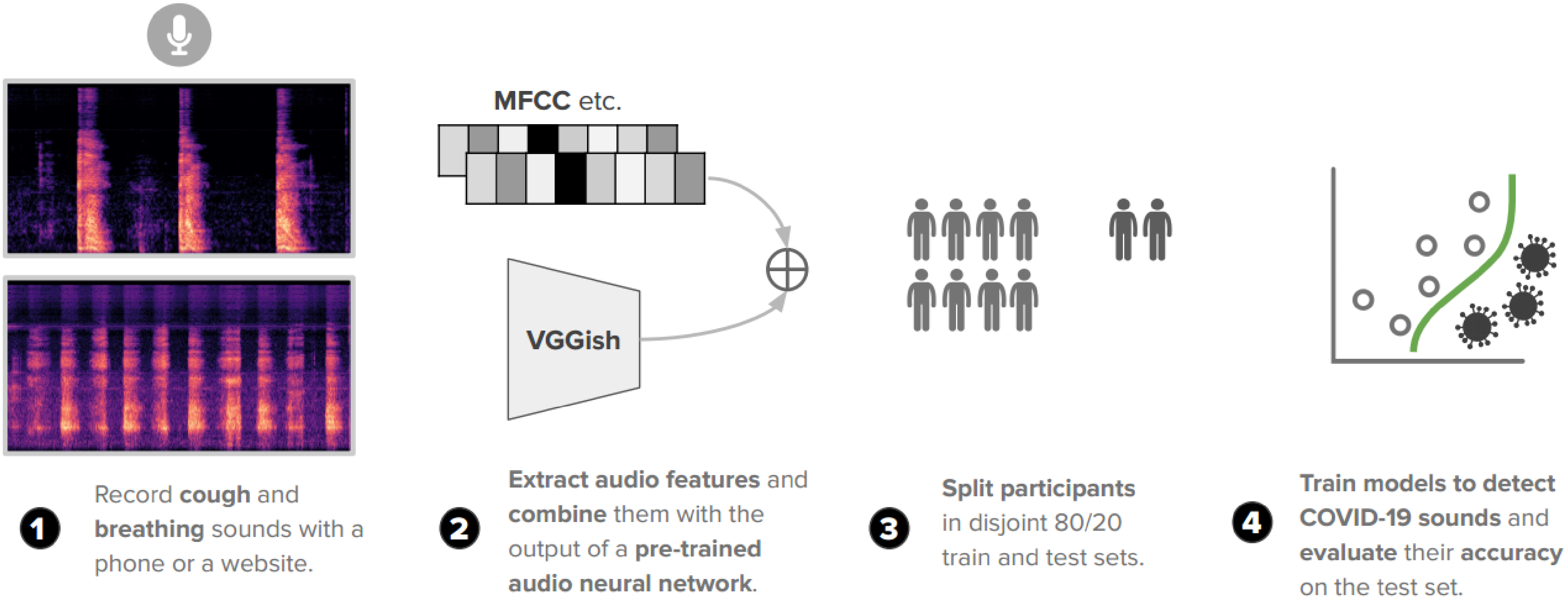
02

Methodology and Results

What method do we use

How about the performance

Framework

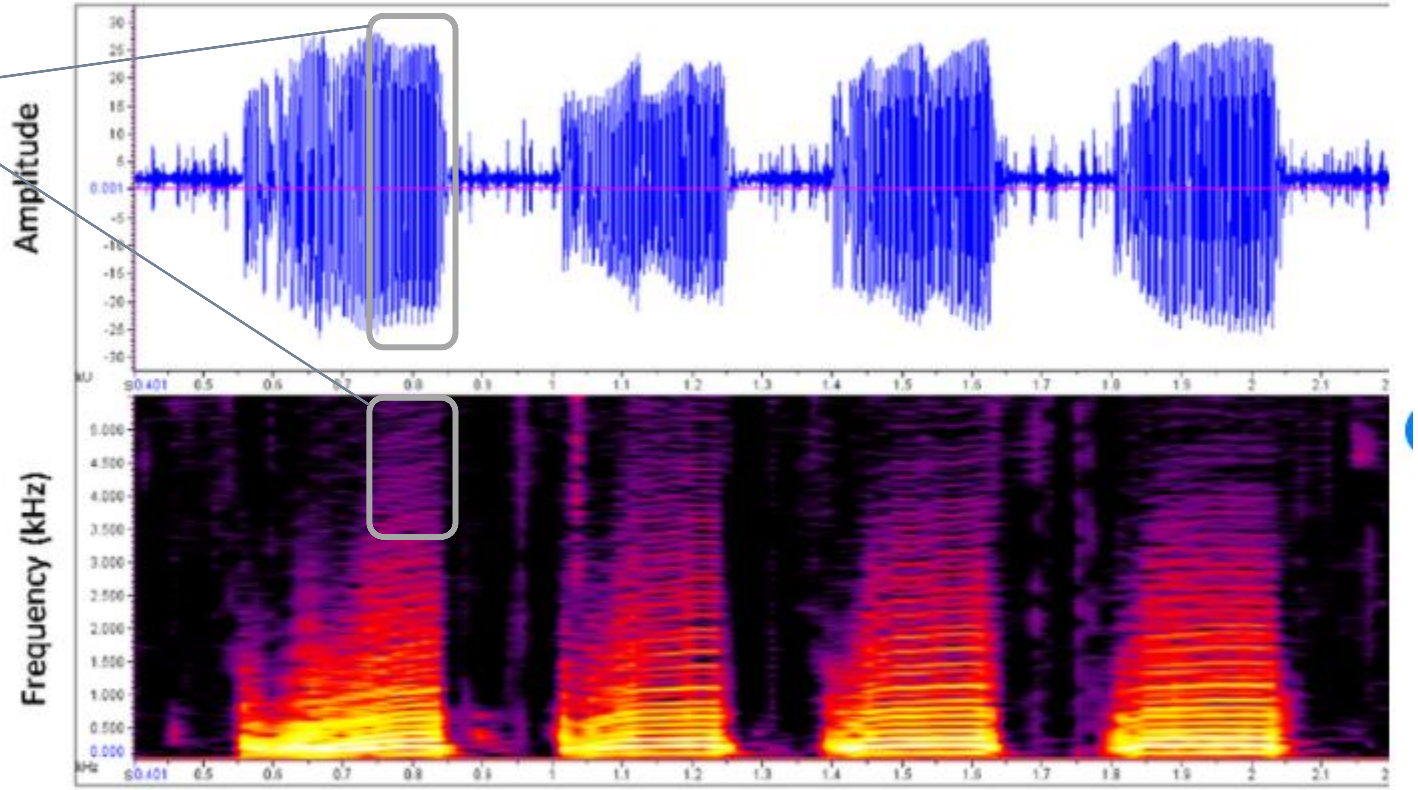


- Distinguish COVID-19 positive from negative
  - AUC-ROC
  - Sensitivity(TPR)
  - Specificity(1-FPR)

		Has the condition	Does not have the condition
Result from screening test	Positive	a True positive	b False positive
	Negative	c False negative	d True negative

Spectrogram

High cross-zero rate corresponding to energy distributed in high frequency band



Short-time Fourier Transform (**STFT**), is used to determine the sinusoidal frequency and phase content of local sections of a signal as it changes over time.

A **spectrogram** is a visual representation of the spectrum of frequencies of a signal as it varies with time.

Dee Learning Model

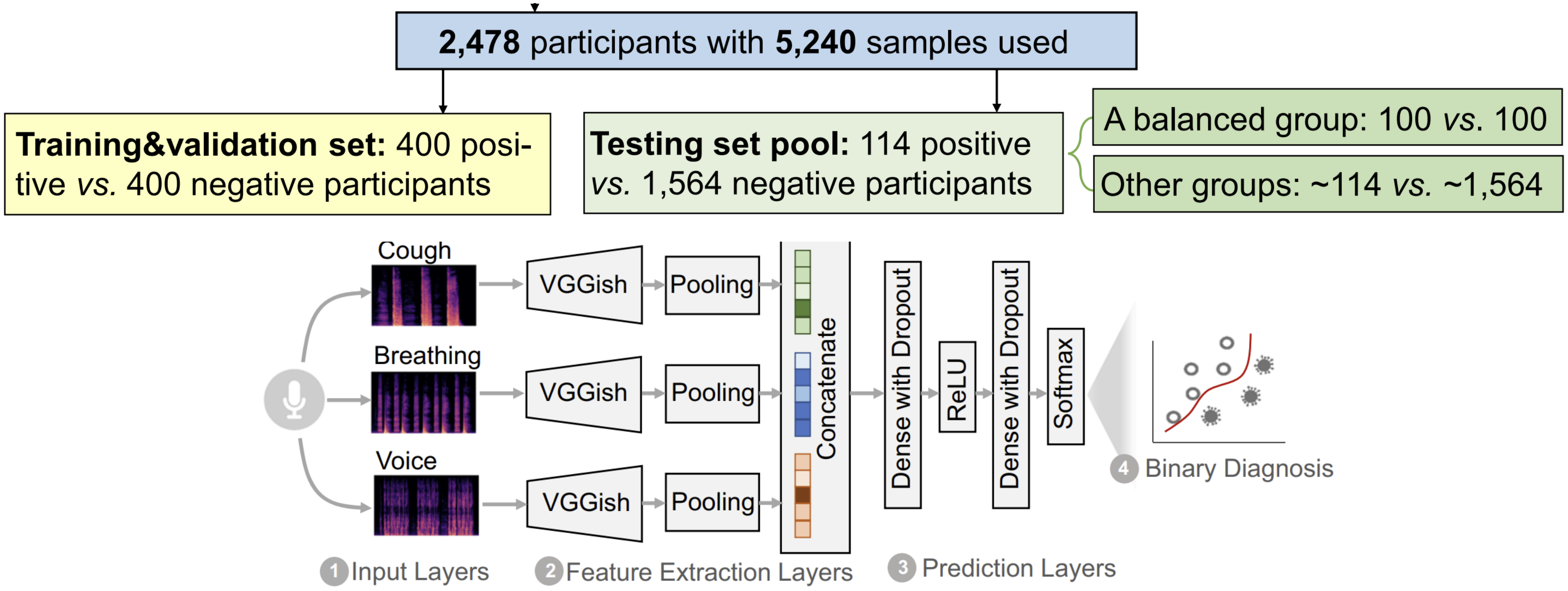


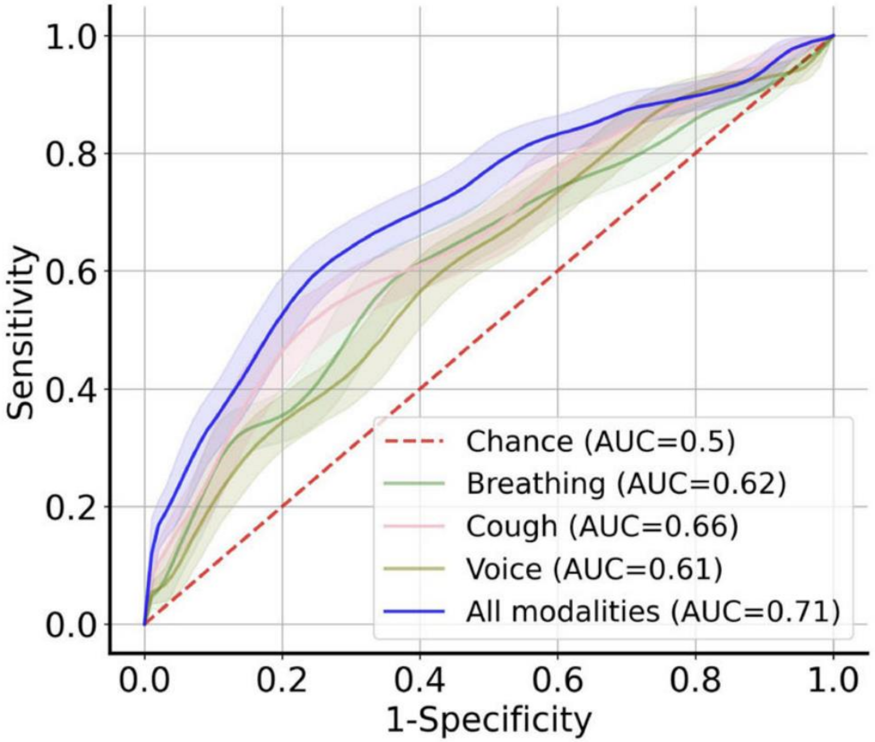
Fig. 5 Overview architecture of the deep learning model. A convolutional neural network using cough, breathing, and voice sounds as input, to predict COVID-19 as a binary outcome. *VGGish* is a neural network pre-trained on the Audioset dataset, *Pooling* is an aggregation operator, *Dense* is a fully connected neural network layer, *Dropout* is a randomised operation that reduces overfitting, *ReLU* is a rectified linear unit activation, *Softmax* is the logistic function.

Jing et al. Sounds of COVID-19: exploring realistic performance of audio-based digital testing. npj Digital Medicine 2022.

Results

*Sensitivity of LFT from individual studies ranged from 37.7% (95% CI 30.6-45.5) to 99.2% (95% CI 95.5-99.9)

a.



b.

	ROC-AUC	Sensitivity	Specificity
Breathing only	0.62(0.56-0.68)	0.64(0.56-0.71)	0.56(0.48-0.63)
Cough only	0.66(0.60-0.71)	0.59(0.51-0.66)	0.66(0.58-0.73)
Voice only	0.61(0.55-0.67)	0.57(0.49-0.64)	0.60(0.52-0.67)
All modalities	**0.71(0.65-0.77)**	**0.65(0.58-0.72)**	**0.69(0.62-0.76)**

c.

Subgroup	#Pos./Neg.	ROC-AUC	Sensitivity	Specificity
Gender				
Male	58/52	0.71(0.63-0.78)	0.59(0.49-0.68)	0.74(0.63-0.83)
Female	42/46	0.73(0.65-0.80)	0.71(0.61-0.81)	0.65(0.55-0.75)
Age				
16-39	55/54	0.65(0.56-0.73)	0.57(0.46-0.68)	0.65(0.55-0.75)
40-59	36/34	0.76(0.67-0.85)	0.72(0.61-0.82)	0.68(0.55-0.81)
60-	4 /6	0.91(0.77-1.0)	0.88(0.60-1.0)	0.88(0.69-1.0)
Symptom				
Asymptomatic	18/73	0.75(0.60-0.88)	0.50(0.25-0.76)	0.85(0.77-0.92)
Symptomatic	144/89	0.66(0.59-0.73)	0.67(0.59-0.74)	0.56(0.45-0.66)

Fig. 2 Model performance. a Receiver-operating characteristic curve for the binary classification task of diagnosing COVID-19. **b** ROC-AUC, sensitivity and specificity with 95% confidence intervals in brackets for the combination of all modalities or each single modality separately. **c** Subgroup performance comparison under three modalities. For gender and age group, # denotes the number of unique positive/negative participants. Note that some participants provided multiple samples, which could be either asymptomatic or symptomatic.

**A systematic review of the sensitivity and specificity of lateral flow devices in the detection of SARS-CoV-2. BMC Infect Dis.*

Results of confounding factors

- How to use the **data**?
  - Random splitting?
  - User-independent?
  - Demographics?

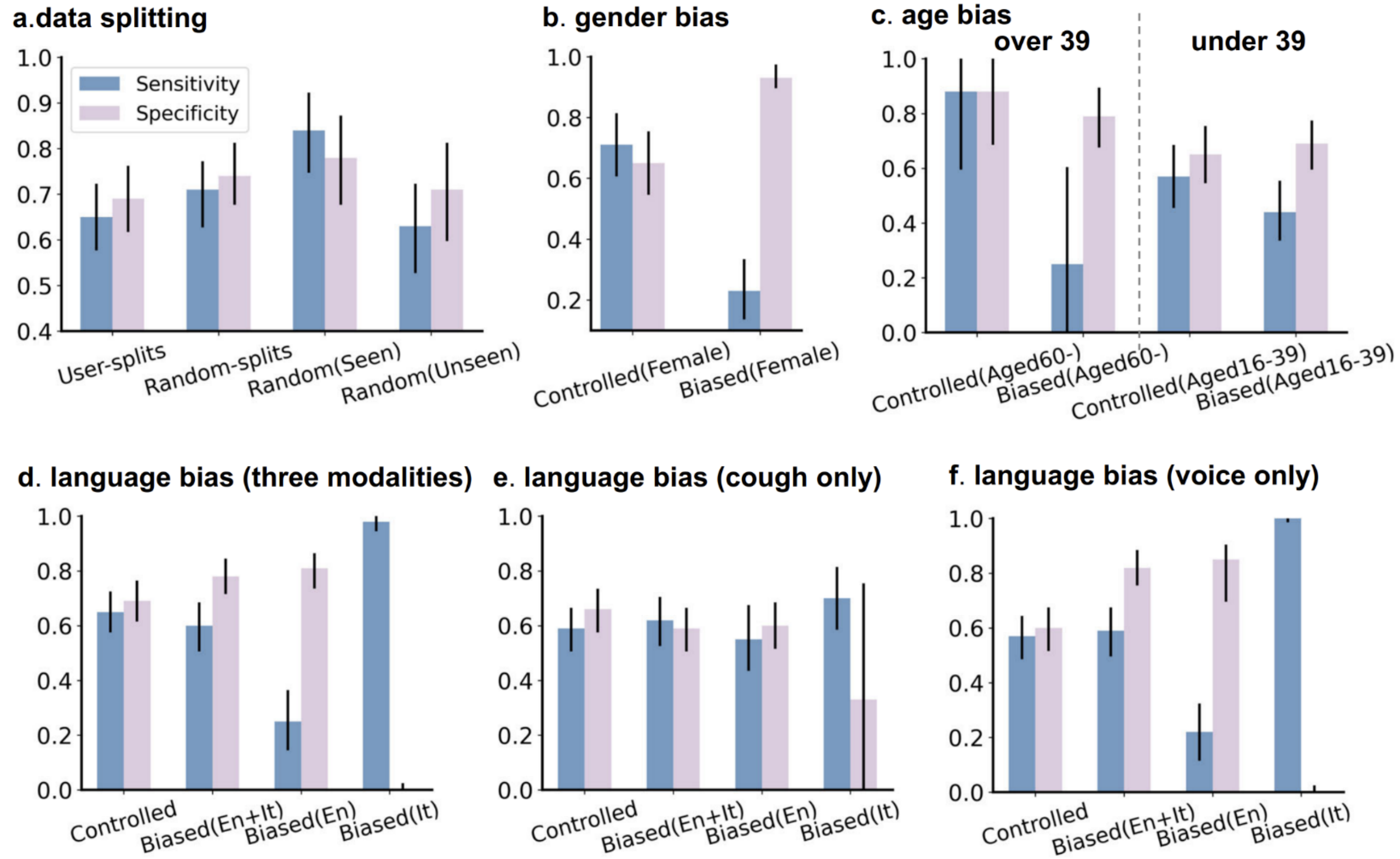


Fig. 4 Performance comparison. Sensitivity (blue) and specificity (pink) are presented with sold liner showing the 95% CIs. If not particularly mentioned, the results are based on the combination of three sound types. **a** User-independent splits vs. sample-level random splits: (Seen) denotes the performance on samples whose other samples were used for training, otherwise the performance is notated by (Unseen). **b** Controlled demographics vs gender bias: (Female) denotes the female subgroup. c, Controlled demographics vs two types of gender biases: all negative participants in training set aged over 39 or under 39. (Aged 60-) and (Aged 16–39) denote the elder and the younger subgroup. **d**–**f** Model for English-speakers vs model for biased English- and Italian-speakers: (En) and (It) denote two subgroups from the testing set.

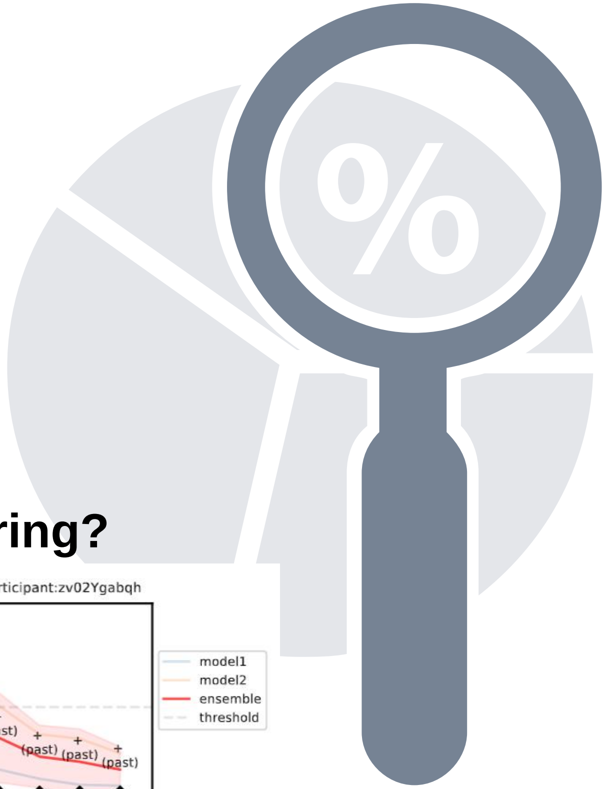
03

Discussion

More Work

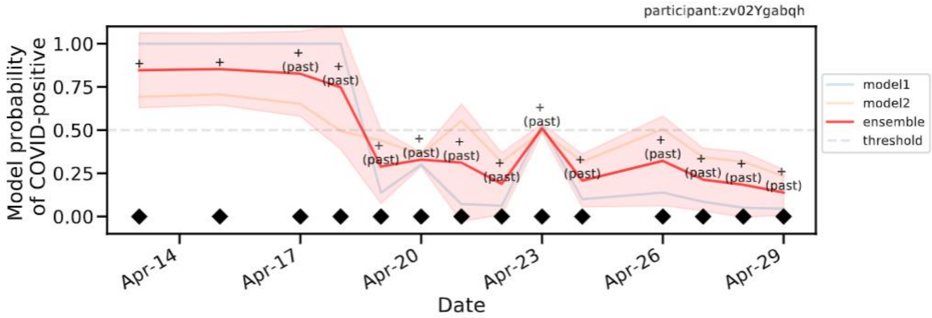
We are working on

Machine v.s Doctor?



Is the model robust in the wild?

How to realize continual monitoring?



How to handle the data sensitivity?

Machine v.s Doctor

A comparative study of listening performance for COVID-19 between clinicians and machine learning

Jing Han[1*] Marco Montagna[2] Andreas Grammenos[1] Tong Xia[1]
Erika Bondareva[1] Chloë Siegele-Brown[3] Jagmohan Chauhan[3] Ting Dang[1]
Dimitris Spathis[1] Andres Floto[1] Pietro Cicuta[1] Cecilia Mascolo[1]

[1]University of Cambridge, UK
[2]Vita-Salute San Raffaele University, Italy
[3]University of Southampton, UK
[*] Corresponding author: jh2298@cam.ac.uk

- 24 audio samples (12 are tested positive)
- 36 respiratory clinicians
- The machine learning model outperformed the clinicians

Under-submission.

Machine v.s Doctor

Table 1: Performance comparison on 36 doctors and our deep learning-based model in terms of sensitivity (se), specificity (sp), and accuracy (acc), with 95% confidence intervals (CI) reported in brackets. Best performance of one clinician and our model in terms of accuracy is highlighted.

U_{id}	acc	se	sp	U_{id}	acc	se	sp
ID-1	.29(.12-.46)	.42(.14-.70)	.17(.00-.40)	ID-2	.54(.33-.75)	.50(.21-.80)	.58(.30-.86)
ID-3	.25(.08-.42)	.33(.08-.62)	.17(.00-.40)	ID-4	.33(.17-.54)	.25(.00-.50)	.42(.13-.70)
ID-5	.46(.29-.67)	.50(.22-.79)	.42(.14-.73)	ID-6	.42(.21-.62)	.33(.08-.64)	.50(.22-.78)
ID-7	.46(.29-.67)	.50(.22-.79)	.42(.14-.73)	ID-8	.62(.42-.79)	.75(.50-1.0)	.50(.21-.75)
ID-9	.25(.08-.42)	.33(.08-.62)	.17(.00-.40)	ID-10	.62(.42-.79)	.75(.50-1.0)	.50(.21-.75)
ID-11	.33(.17-.54)	.25(.00-.50)	.42(.13-.70)	ID-12	.42(.21-.62)	.33(.08-.64)	.50(.22-.78)
ID-13	.54(.33-.75)	.50(.21-.80)	.58(.30-.86)	ID-14	.42(.21-.58)	.25(.00-.50)	.58(.31-.85)
ID-15	.54(.33-.75)	.83(.60-1.0)	.25(.00-.50)	ID-16	.50(.33-.71)	.33(.08-.62)	.67(.38-.92)
ID-17	.50(.33-.71)	.33(.08-.62)	.67(.38-.92)	ID-18	.62(.42-.83)	.58(.30-.88)	.67(.38-.92)
ID-19	.58(.38-.79)	.50(.21-.77)	.67(.40-.92)	ID-20	.62(.42-.83)	.58(.30-.88)	.67(.38-.92)
ID-21	.58(.38-.79)	.50(.21-.77)	.67(.40-.92)	ID-22	.50(.29-.71)	.75(.50-1.0)	.25(.00-.53)
ID-23	.54(.33-.75)	.42(.14-.71)	.67(.38-.91)	ID-24	.50(.29-.71)	.50(.18-.78)	.50(.20-.80)
ID-25	.62(.42-.83)	.50(.20-.78)	.75(.50-1.0)	ID-26	.54(.38-.75)	.33(.08-.62)	.75(.50-1.0)
ID-27	.50(.29-.71)	.50(.18-.78)	.50(.20-.80)	ID-28	.46(.29-.67)	.42(.12-.70)	.50(.22-.75)
ID-29	.29(.12-.46)	.42(.14-.70)	.17(.00-.40)	ID-30	.62(.42-.79)	.75(.50-1.0)	.50(.21-.75)
ID-31	.58(.38-.79)	.42(.12-.71)	.75(.50-1.0)	ID-32	.58(.38-.79)	.58(.29-.86)	.58(.31-.85)
ID-33	.50(.29-.71)	.33(.08-.62)	.67(.38-.92)	ID-34	.50(.29-.71)	.33(.08-.62)	.67(.38-.91)
ID-35	.62(.42-.83)	.58(.27-.85)	.67(.42-.92)	ID-36	**.71(.50-.88)**	**.67(.36-.92)**	**.75(.50-1.0)**
our Model	**.79(.62-.92)**	**.75(.46-1.0)**	**.83(.58-1.0)**				

Our method outperforms clinicians.

Machine v.s Doctor

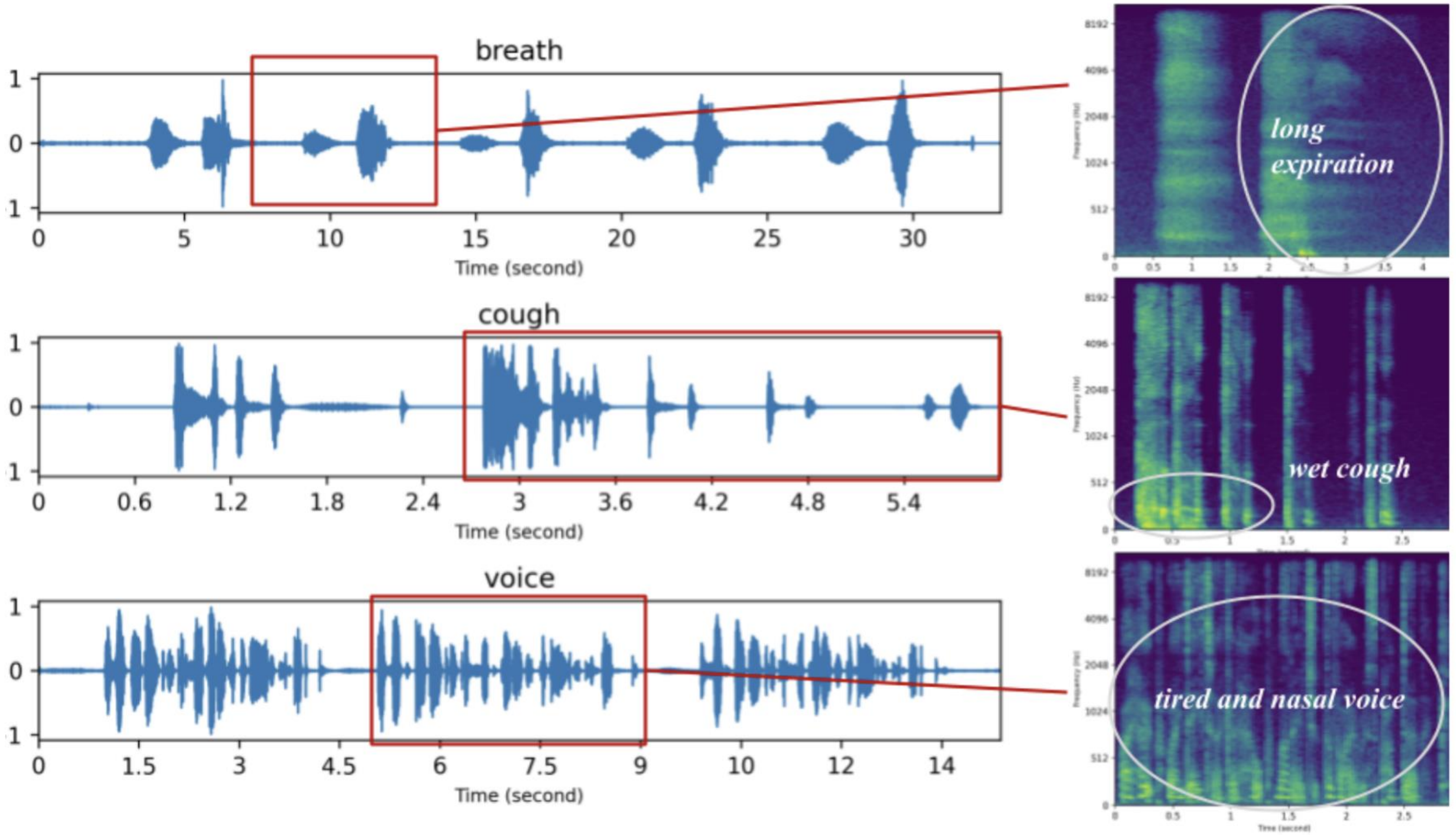


Figure 4: COVID-19 positive breath, cough, and voice recordings from two samples (08 and 18) and the reported cues from one of the clinicians.

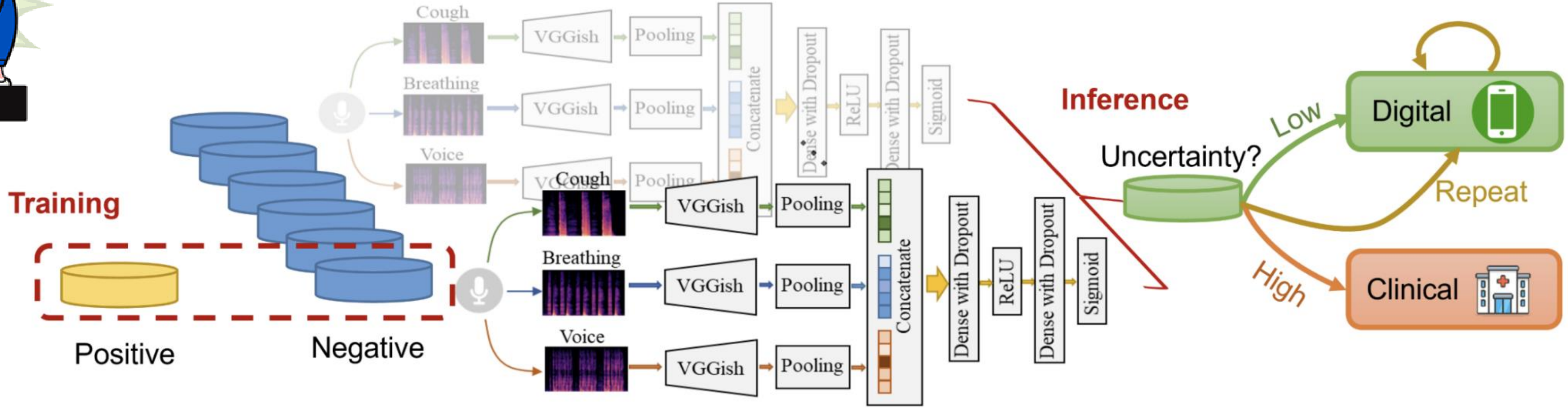
Uncertainty Estimation

Uncertainty-Aware COVID-19 Detection from Imbalanced Sound Data

Tong Xia, Jing Han, Lorena Qendro, Ting Dang, Cecilia Mascolo

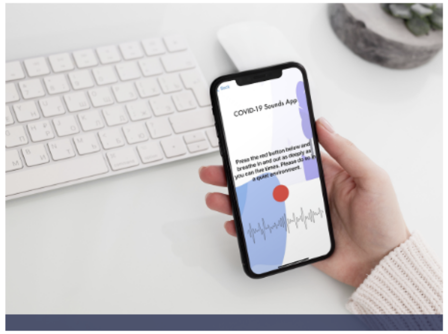
Department of Computer Science and Technology, University of Cambridge, UK

tx229@cam.ac.uk



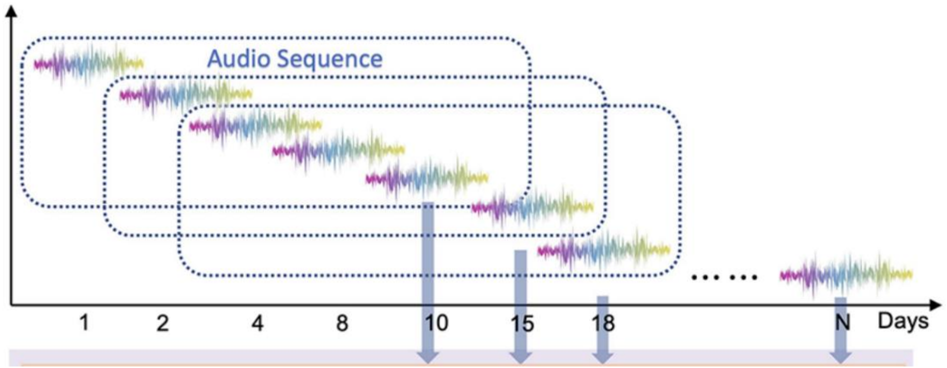
Tong et al. Uncertainty-Aware COVID-19 Detection from Imbalanced Sound Data. INTERSPPECH 2021.

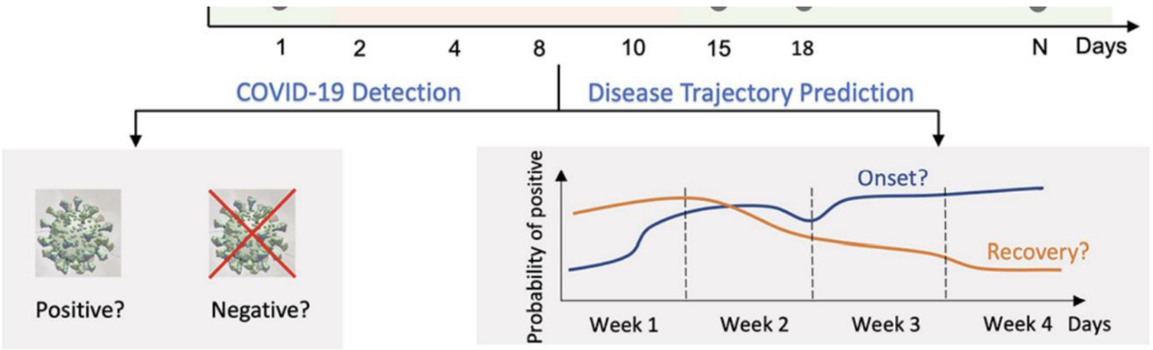
Progression Prediction



Exploring Longitudinal Cough, Breath, and Voice Data for COVID-19 Progression Prediction via Sequential Deep Learning: Model Development and Validation

Ting Dang [1]; Jing Han [1]; Tong Xia [1]; Dimitris Spathis [1]; Erika Bondareva [1]; Chloë Siegele-Brown [1]; Jagmohan Chauhan [1, 2]; Andreas Grammenos [1]; Apinan Hasthanasombat [1]; R Andres Floto [1]; Pietro Cicuta [1]; Cecilia Mascolo [1]





Model Training without Aggregating Data

Sensitive !

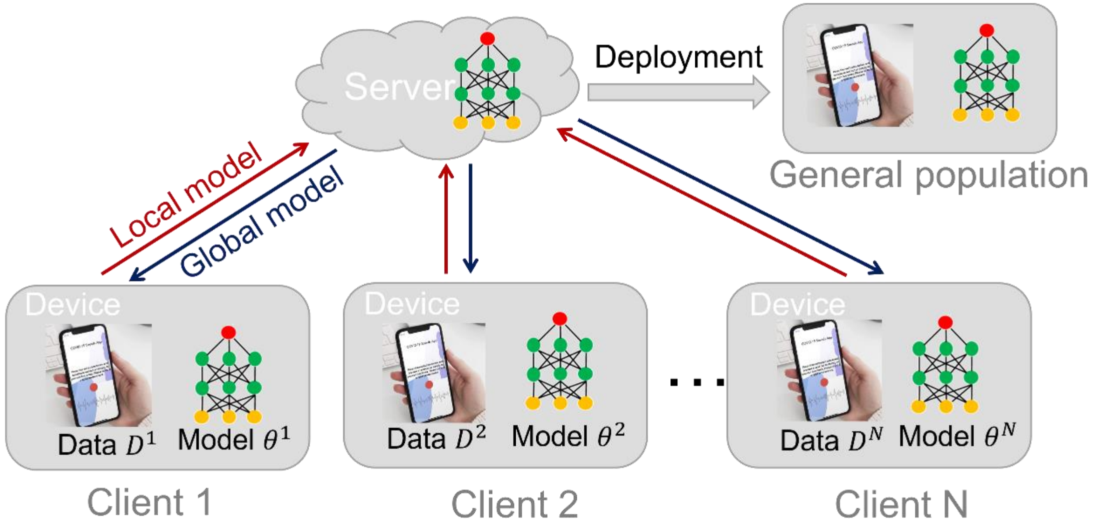
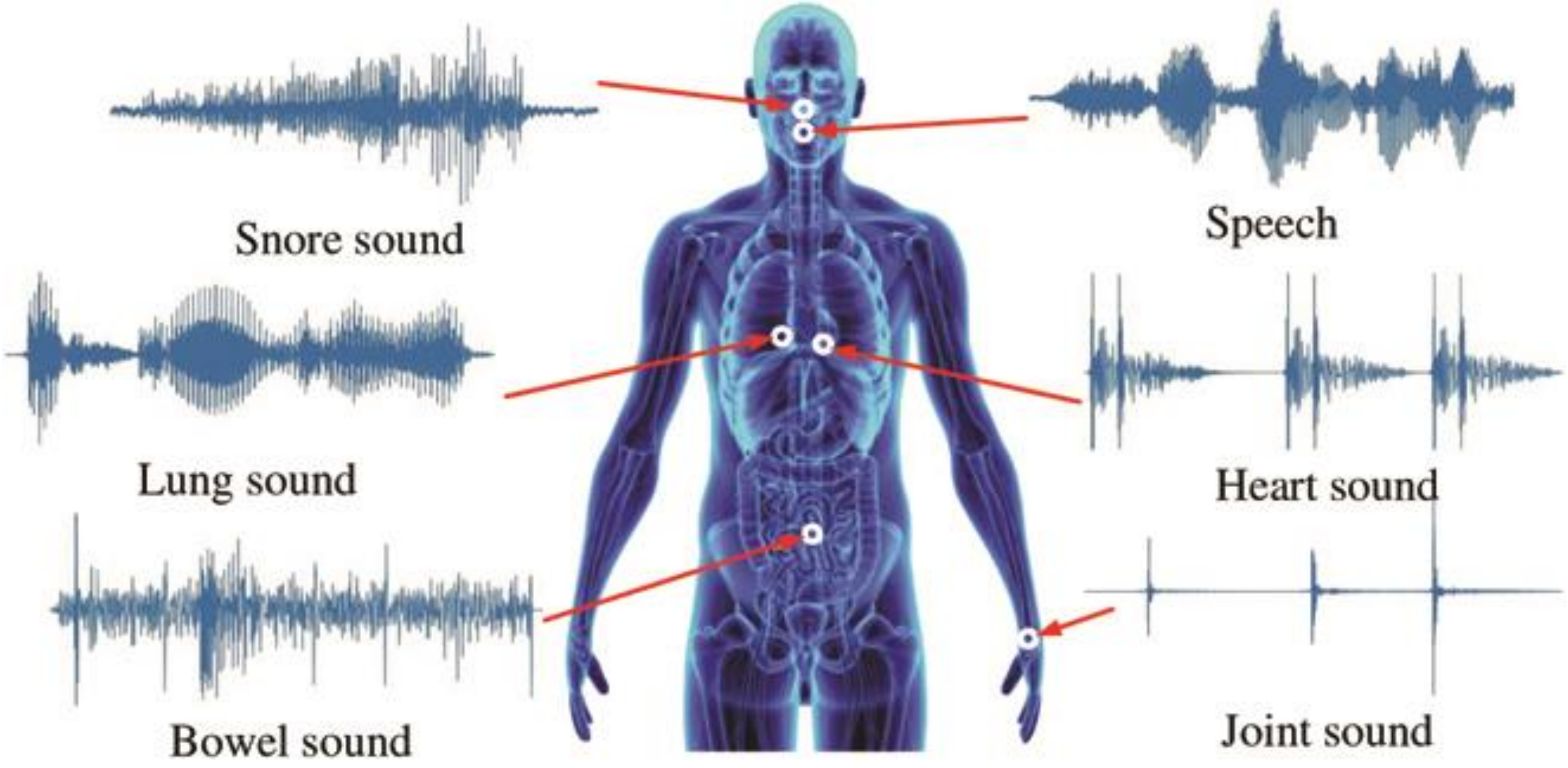


Fig. 1. Cross-device FL for mobile health, where models are trained on edge devices from private health sensing data, and the global model is aggregated from the clients' models

An overview of the body sound analysis.



These body sounds can be collected via the prevalent microphone equipment and/or wearable devices.
The sounds generated by our human body can reveal the health status in both physical and mental terms.

Qian K, et al. The Voice of the Body: Why AI Should Listen to It and an Archive.

Thank you!

Tong Xia

https://xtxiatong.github.io/

tx229@cam.ac.uk