

## Bias and privacy in AI's cough-based COVID-19 recognition

We read with interest the Comment by Coppock and colleagues,<sup>1</sup> in which the authors express their thoughtful opinion about several simultaneous works by independent research groups worldwide (eg, Massachusetts Institute of Technology, National Research Council of Canada, University of Cambridge, and Swiss Federal Institute of Technology Lausanne). One of these works was our own; a pioneering, multicentre, international study<sup>2</sup> with a clinically validated dataset of forced coughs alongside quantitative RT-PCR from participants who physically attended a test centre. Participant control was performed on site by health personnel at the partner health centres that contributed to this study.<sup>2</sup>

The Comment<sup>1</sup> focuses on human audio biometrics in general, albeit eight out of ten referenced works made use of coughs as their audio source. The use of cough sounds to detect respiratory system abnormalities had been investigated for at least 3 years before the COVID-19 pandemic.<sup>3</sup> Cough analysis is a particular form of audio biometrics, and other forms of audio biometrics cannot be discussed interchangeably.

Although person authentication via speech has been achieved to some degree; to date, recognising an individual with ease in a large database solely by the sound of their cough is inconclusive. The same applies to inferring emotional traits.

The prospect of re-identifying patients from cough sounds raises several concerns. Will participants whose data is used in developing these pre-screening systems be able to benefit from it in the future in an unbiased manner? Additionally, how can personal biometric data be made public, ensuring that it always remains non-identifiable? On the basis of the current health context, official calls

have been made not to trivialise the privacy and protection of patient data.<sup>4,5</sup> Subsequent research initiatives from public bodies must now help to enable inclusive research clusters and secure collaborative infrastructures in the domain of audio biometrics.

Our training, development (validation), and holdout (test) sets do not contain data from the same participant (ie, they are participant-independent) to avoid spurious discerning patterns that could compromise classification scores.<sup>2</sup> However, a general assumption is that a competent biometrics classifier should maximise the distinction between divergent patterns within participants, while minimising that of different participants within the same class. The inclusion of divergent observations (negative and positive) from the same participant, in the training set only, could help to satisfy this assumption; however, this effect requires further study.

To conclude, there is hope that rapid, point-of-need pre-screening for COVID-19 via forced cough sounds captured from smartphones or portable devices could be feasible in the short term. Online interventional studies are now necessary to explore the potential of this novel technology and to evaluate its real-life performance, health impact assessment, end-user utility, and acceptability.

We declare no competing interests.

Copyright © 2021 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

Humberto Perez-Espinosa,  
Eva Timonet-Andreu,  
\*Javier Andreu-Perez  
javier.andreu@essex.ac.uk

Faculty of Science and Health, Centre for Computational Intelligence, University of Essex, Colchester CO4 3SQ, UK (HP-E, JA-P); National Council of Science and Technology, CICESE-UT3, Research Transfer Unit, Ensenada, Mexico (HP-E); Costa del Sol Health Agency, Biomedical Research Institute of Malaga, Malaga, Spain (ET-A)

1 Coppock H, Jones L, Kiskin I, Schuller B. COVID-19 detection from audio: seven grains of salt. *Lancet Digit Health* 2021; **3**: e537–38.

2 Andreu-Perez J, Perez-Espinosa H, Timonet E, et al. A generic deep learning based cough analysis system from clinically validated samples for point-of-need COVID-19 test and severity levels. *IEEE Trans Serv Comput* 2021; published online Feb 23. <https://doi.org/10.1109/TSC.2021.3061402>.

3 Botha GHR, Theron G, Warren RM, et al. Detection of tuberculosis by automatic cough sound analysis. *Physiol Meas* 2018; **39**: 045005.

4 WHO. Joint statement on data protection and privacy in the COVID-19 response. Nov 19, 2020. <https://www.who.int/news/item/19-11-2020-joint-statement-on-data-protection-and-privacy-in-the-covid-19-response> (accessed Aug 2, 2021).

5 Pierucci A, Walter J-P. Joint Statement on the right to data protection in the context of the COVID-19 pandemic. May 14, 2020. <https://www.coe.int/en/web/kyiv/-/joint-statement-on-the-right-to-data-protection-in-the-context-of-the-covid-19-pandemic> (accessed Aug 2, 2021).