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

Harry Coppock, Lyn Jones, Ivan Kiskin, Björn Schuller

Angaben zur Veröffentlichung / Publication details:

Coppock, Harry, Lyn Jones, Ivan Kiskin, and Björn Schuller. 2021. "Bias and privacy in AI's cough-based COVID-19 recognition - authors' reply." *The Lancet Digital Health* 3 (12): e761. https://doi.org/10.1016/s2589-7500(21)00233-8.

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

Authors' reply

We thank Humberto Perez-Espinosa and colleagues for their constructive points regarding our Comment,[1] which raised concerns over the work on COVID-19 detection from bioacoustic recordings. We take this opportunity to note that the study by Perez-Espinosa and colleagues[2] represented one of the superior COVID-19 audio datasets that were collected. Although the study was not completely free from the "seven grains of salt" detailed in our Comment,[1] it was large scale, validated by quantitative RT-PCR, and the participants were blinded. We also applaud the recording of cycle threshold, which allowed for the comparison between model performance and viral load. We agree with the authors that participants of studies used to develop deep learning algorithms will not be able to benefit from the screening tool in a completely unbiased manner. Effort should be made to reduce this effect through careful training procedures and further scientific breakthroughs in explainable artificial intelligence and debiasing systems. In answer to the authors' concern for the privacy of participants in publicly available datasets, we admit that privacy law is not our area of expertise and we would look towards an expert in the field to comment on this. However, we note that there are a multitude of publicly available datasets containing sensitive biometric data—eg, the COVID-19 auditory respiratory dataset, COVID-19 Sounds.[3] Nevertheless, if publication of a dataset is not possible, effort should be made to evaluate each study's model on other datasets, and to invite other research groups to evaluate their trained model on that dataset in a manner that keeps the data private and secure.

We note that Perez-Espinosa and colleagues state that their "training, development (validation), and holdout (test) sets do not contain data from the same participant". This statement is a vital piece of information to include when writing up studies. The authors make an important point regarding the variability between participants; however, positive and negative cases from the same participant should still exist purely within one set and not cross train or test boundaries. We contend that absence of current published evidence does not eliminate the possibility that identity could be determined from cough audio; therefore, it cannot be considered sufficient justification for including the same individuals in both training and test sets. Given the advances of deep learning in pattern recognition, combined with audio recordings containing data besides bioacoustic information, we argue with confidence that cough recordings allow for algorithms to infer user identity to a high level. Additionally, we know from first-hand experience that when disjoint user sets are not ensured, classification performance substantially increases. This finding was shown in the study by Han and colleagues,[4] in which bias was systematically added back into the dataset. When disjoint user sets were not ensured, sensitivity scores for COVID-19 increased from 0·65 (95% CI 0·58–0·72) to 0·84 (0·75–0·92) for test users who were also present in the training set. The complexity of deep neural networks allows for memorisation of data to a degree, which results in inflated scores when not training and testing on disjoint sets.[5] Thus, creating a disjoint test set is a fundamental prerequisite for reporting representative performance figures.

Furthermore, although we agree with the authors that cough analysis is a distinct form of audio biometrics, it is within the same field of human respiratory sounds, meaning it is also subject to the issues detailed in our Comment.[1]

We declare no competing interests.

*Harry Coppock, Lyn Jones, Ivan Kiskin, Björn Schuller
harry.coppock@imperial.ac.uk

Department of Computer Science, Imperial College London, London SW7 2RH, UK, (HC, BS); Radiology Department, North Bristol NHS Trust, Bristol, UK (LJ); Department of Engineering Science, University of Oxford, Oxford, UK (IK)

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. Xia T, Spathis D, Brown C, et al. COVID-19 sounds: a large-scale audio dataset for digital COVID-19 detection. Aug 20, 2021. https://openreview.net/forum?id=9KArJb4r5ZQ (accessed Nov 2, 2021).
4. Han J, Xia T, Spathis D, et al. Sounds of COVID-19: exploring realistic performance of audio-based digital testing. June 29, 2021. https://arxiv.org/abs/2106.15523 (accessed Nov 2, 2021).
5. Carlini N, Liu C, Erlingsson Ú, Kos J, Song D. The secret sharer: evaluating and testing unintended memorization in neural networks. https://www.usenix.org/conference/usenixsecurity19/presentation/carlini (accessed Nov 2, 2021).