

## Detección del COVID-19 por audio: siete granos de sal

*Harry Coppock, Lyn Jones, Ivan Kiskin, Björn Schuller*

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

The Lancet, [doi.org/10.1016/S2589-7500\(21\)00141-2](https://doi.org/10.1016/S2589-7500(21)00141-2), 21 de julio, 2021.

---

Los testeos masivos digitales para COVID-19, a través de una aplicación de un teléfono móvil, podrían ser posibles a través de la máquina de aprendizaje y su capacidad para identificar patrones en los datos. El COVID-19 parece conferir características únicas en el audio producido por las personas infectadas,<sup>1</sup> y el aprendizaje por las máquinas para la detección del COVID-19 a partir de la respiración, la tos y las grabaciones de audio de voz, ha dado resultados prometedores.<sup>2-4</sup>

En esta crítica, presentamos 7 temas principales con esta investigación y argumentamos que se necesita más investigación antes de sacar conclusiones sobre la detectabilidad del COVID-19 a partir de un audio. Muchos de estos problemas se relacionan con una sola pregunta: ¿son las representaciones de audio aprendidas, que se correlacionan con COVID-19 en las diversas recopilaciones de conjuntos de datos, verdaderos biomarcadores de audio causados por COVID-19?

Una preocupación es que los algoritmos de aprendizaje automático pueden simplemente distinguir entre individuos sanos e individuos que generalmente no se encuentran bien, en lugar de estar detectando el propio COVID-19. Distinguir entre individuos sanos y enfermos es mucho más fácil, pero menos útil. Aunque algunos investigadores han estudiado este problema, por ejemplo, mediante la construcción de tareas en las que los modelos entrenados clasifican entre individuos con otras enfermedades respiratorias y personas con COVID-19,<sup>2,5</sup> ningún estudio ha podido concluir que el propio COVID-19 se está identificando realmente.

La corrupción ambiental es otra preocupación. Cuando un evento de audio, como una tos, se graba, la información sobre el entorno acústico circundante también es incluida en el archivo de audio. El ruido de fondo introduce un potencial de sesgo en el conjunto de datos de audio. Pueden existir sesgos de adiciones de audio inadvertidas en muchas formas; un ejemplo en este contexto podría ser que las personas positivas para COVID-19 tienen más probabilidades de ser en interiores, tal vez en un entorno médico en el momento de la grabación, que aquellos sin COVID-19, mientras que las personas negativas para COVID-19 tienen más probabilidades de estar al aire libre, o tal vez en un lugar de trabajo, que los con COVID-19. Cada escenario tiene su propio modo de audio ambiental identificable. Cualquier asociación puede ser identificada por el modelo de aprendizaje automático, ya que intenta hacer una predicción de COVID-19. Este sesgo puede ser sutil y, una vez presente en el conjunto de datos, es muy difícil de detectar y eliminar. Esta sutileza podría explicar por qué no se ha abordado como una limitación en cualquiera de los estudios hasta ahora. La investigación futura debe intentar controlar el entorno de audio en el momento de la grabación para abordar este problema.

La conciencia de los participantes, o falta de cegamiento, es un problema. En la mayoría de los conjuntos de datos recopilados, la mayoría los participantes conocían su estado de COVID-19 en el momento de la grabación,<sup>2,3,5-9</sup> lo que es problemático, porque la información que genera el saber el estado de saber que tienen COVID-19, puede filtrarse en los modelos de aprendizaje de la máquina, través de la emoción en el participante, en la voz y comportamiento. Al igual que con la corrupción ambiental, la conciencia de los participantes permite modelos de aprendizaje automático para evitar la difícil tarea de identificar las características del verdadero COVID-19, a favor de una ruta más fácil, para lograr una puntuación de clasificación alta; en este caso, asociando señales emocionales con el estado de COVID-19.

Otro problema es la validez de los conjuntos de datos. Los métodos de aprendizaje automático supervisado se basan en un terreno preciso de etiquetado de la verdad o realidad de las instancias. La herramienta resultante de la máquina de aprendizaje solo puede ser tan precisa como estas etiquetas. Sin embargo, la validez de las etiquetas utilizadas en los conjuntos de datos de audio recopilados en el COVID-19 hasta ahora son cuestionables, porque la mayoría de los conjuntos de datos permiten a los participantes autoinformar su situación de COVID-19 y fallan en registrar el tipo de prueba que los participantes usaron;<sup>2,8,9</sup> como una PCR o la prueba de flujo lateral, menos precisa. A pesar de que algunos estudios exigen una prueba de PCR validada,<sup>6-8,10</sup> los conjuntos de datos son pequeños y ninguno se ha puesto a disposición del público en el momento de escribir y publicar. Para resaltar la gravedad de este problema, observamos que algunos conjuntos de datos han aceptado la autoevaluación como medio de etiquetar el conjunto de datos,<sup>3</sup> y otros ni siquiera han logrado detallar cómo se determinó el estado de COVID-19.<sup>5</sup> Además, encontramos pocos trabajos que investiguen la sensibilidad del algoritmo, relacionada al umbral de ciclo de la PCR (Ct) y, por tanto, carga viral.

La disponibilidad de una base de códigos, y de un conjunto de datos es otro motivo de preocupación. Una parte importante del proceso de la ciencia es la revisión por pares, y la reproducción de los resultados. Sin embargo, en un trabajo publicado sobre la detección de COVID-19 a partir del audio, es raro que los equipos hagan disponible públicamente su base de códigos o conjunto de datos, y esto ha impedido a otros grupos intenten replicar sus resultados. Solo los autores de cuatro<sup>2,4,8,9</sup> estudios han publicado su códigos o conjuntos de datos de forma parcial o total. ¿Pueden se pueden tomar en serio medidas de clasificación sobre un tema tan delicado, si no pueden ser controladas por otros equipos de investigación?

Las comorbilidades, los factores geográficos, étnicos y socioeconómicos son una posible preocupación en el contexto de la utilización del aprendizaje automático para detectar COVID-19. La influencia de estos factores en la propagación de COVID-19 es compleja. La prevalencia de la enfermedad ha sido desigual en las distintas regiones y demografías sociales, y los patrones en los conjuntos de datos recopilados, se acentúan por una irregular distribución de las probabilidades entre los grupos participantes en estos estudios. El sesgo se introduce a través de la resultante de las oportunidades de que que los modelos de aprendizaje automático infieran en el estado de COVID-19, a través de la asociación con las características de los datos demográficos.

Por último, existe un problema común sobre el nivel de control que se tiene sobre la población participante al desarrollar los modelos de aprendizaje automático. Los modelos de inteligencia artificial son competentes en la identificación de los hablantes. Por tanto, es imperativo que la formación, desarrollo, y los conjuntos de prueba sean de poblaciones independientes. Cuando las poblaciones participantes no son controladas, se informan puntuaciones de clasificación

infladas porque el modelo puede reconocer fácilmente la reaparición de los participantes y clasificar su estado de COVID-19 sobre la base de los casos en la fase de formación. Sin embargo, en varios conjuntos de datos no registran la identidad de los participantes,<sup>10,9</sup> resultando en una incapacidad para evitar este problema.

A pesar de estas preocupaciones, la evidencia sugiere que el COVID-19 produce características identificables en el audio de los infectados al hablar, al toser y al respirar, y se está avanzando a través de la recopilación de varios conjuntos de datos clínicamente validados, destinados a abordar estos 7 números. Estos conjuntos de datos nos acercarán a comprender si la aspiración de una prueba masiva, digital y gratuita para el COVID-19 podría convertirse en una realidad.

## Referencias

---

1 Quatieri T, Talkar T, Palmer J. A framework for biomarkers of COVID-19 based on coordination of speech-production subsystems. *IEEE Open Journal of Engineering in Medicine and Biology* 2020; **1**: 203–06.

2 Brown C, Chauhan J, Grammenos A, et al. Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data. *Proc Data Min Knowl Discov* 2020; 3474–84.

3 Laguarda J, Hueto F, Subirana B. COVID-19 artificial intelligence diagnosis using only cough recordings. *IEEE Open Journal of Engineering in Medicine and Biology* 2020; **1**: 275–81.

4 Coppock H, Gaskell A, Tzirakis P, Baird A, Jones L, Schuller B. End-to-end convolutional neural network enables COVID-19 detection from breath and cough audio: a pilot study. *BMJ Innov* 2021; **7**: 356–62.

5 Imran A, Posokhova I, Qureshi HN, et al. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Inform Med Unlocked* 2020; **20**: 100378.

6 Pinkas G, Karny Y, Malachi A, Barkai G, Bachar G, Aharonson V. SARS-CoV-2 detection from voice. *IEEE Open Journal of Engineering in Medicine and Biology* 2020; **1**: 268–74.

7 Pizzo DT, Esteban S, Scetta M. IATos: AI-powered pre-screening tool for COVID-19 from cough audio samples. *arXiv* 2021; published online April 27. <https://arxiv.org/abs/2104.13247> (preprint).

8 Bagad P, Dalmia A, Doshi J, et al. Cough against COVID: evidence of COVID-19 signature in cough sounds. *arXiv* 2020; published online Sept 23. <https://arxiv.org/abs/2009.08790> (preprint).

9 Orlandic L, Teijeiro T, Atienza D. The COUGHVID crowdsourcing dataset: a corpus for the study of large-scale cough analysis algorithms. *arXiv* 2020; published online Sept 24. <https://doi.org/10.1038/s41597-021-00937-4> (preprint).

10 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>.

**Traducción:** *Ramiro Heredia (ramiroherediamd@gmail.com)*