



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Original breve

## Rendimiento diagnóstico de algoritmos de inteligencia artificial para detección de compromiso pulmonar por COVID-19 basados en radiografía portátil

Ricardo Luis Cobeñas\*, María de Vedia, Juan Florez, Daniela Jaramillo, Luciana Ferrari y Ricardo Re

Departamento de Diagnóstico por Imágenes, Centro de Educación Médica e Investigaciones Clínicas Norberto Quirno (CEMIC), Buenos Aires, Argentina

### INFORMACIÓN DEL ARTÍCULO

*Historia del artículo:*

Recibido el 11 de febrero de 2022

Aceptado el 27 de abril de 2022

On-line el xxx

*Palabras clave:*

Inteligencia artificial

COVID-19

Radiografía de tórax

Neumonía

Aprendizaje automático

Pulmón

### R E S U M E N

**Introducción y objetivo:** Evaluar el rendimiento diagnóstico de diferentes algoritmos de inteligencia artificial (IA) para la identificación de compromiso pulmonar por SARS-CoV-2 basados en radiografía (Rx) de tórax portátil.

**Material y método:** Estudio observacional prospectivo que incluyó pacientes ingresados por sospecha de infección por COVID-19 en un hospital universitario entre julio y noviembre de 2020. El patrón de referencia de compromiso pulmonar por SARS-CoV-2 comprendió una PCR positiva y síntomas respiratorios bajos.

**Resultados:** Se incluyeron 493 pacientes, 140 (28%) con PCR positiva y 32 (7%) con neumonía por SARS-CoV-2. El algoritmo AI-B tuvo el mejor rendimiento diagnóstico (áreas bajo la curva ROC AI-B 0,73 vs. AI-A 0,51 vs. AI-C 0,57). Utilizando un umbral de detección superior al 55%. AI-B presentó mayor precisión que el especialista (área bajo la curva de 0,68 [IC 95%: 0,64-0,72] vs. 0,54 [IC 95%: 0,49-0,59]).

**Conclusión:** Los algoritmos de IA basados en Rx portátiles permiten una precisión diagnóstica comparable a la humana para la detección de compromiso pulmonar por SARS-CoV-2.

© 2022 Elsevier España, S.L.U. Todos los derechos reservados.

## Diagnostic performance of artificial intelligence algorithms for detection of pulmonary involvement by COVID-19 based on portable radiography

### A B S T R A C T

**Introduction and objectives:** To evaluate the diagnostic performance of different artificial intelligence (AI) algorithms for the identification of pulmonary involvement by SARS-CoV-2 based on portable chest radiography (RX).

**Material and methods:** Prospective observational study that included patients admitted for suspected COVID-19 infection in a university hospital between July and November 2020. The reference standard of pulmonary involvement by SARS-CoV-2 comprised a positive PCR test and low-tract respiratory symptoms.

**Results:** 493 patients were included, 140 (28%) with positive PCR and 32 (7%) with SARS-CoV-2 pneumonia. The AI-B algorithm had the best diagnostic performance (areas under the ROC curve AI-B 0.73, vs. AI-A 0.51, vs. AI-C 0.57). Using a detection threshold greater than 55%, AI-B had greater diagnostic performance than the specialist [(area under the curve of 0.68 (95% CI 0.64-0.72), vs. 0.54 (95% CI 0.49-0.59)].

**Conclusion:** AI algorithms based on portable RX enabled a diagnostic performance comparable to human assessment for the detection of SARS-CoV-2 lung involvement.

© 2022 Elsevier España, S.L.U. All rights reserved.

\* Autor para correspondencia.

Correo electrónico: ricardocobenas@gmail.com (R.L. Cobeñas).

## Introducción

La utilización de la radiografía (Rx) portatil durante la pandemia COVID-19 significó un recurso indispensable en este contexto, reduciendo los riesgos de contaminación asociados al traslado. Es por este motivo que se estableció la realización de estudios portátiles tanto a pacientes internados en sala de internación como a consultas externas ambulatorias.

La tecnología informática aportó positivamente en esta etapa inicial de la pandemia, con la aparición de aplicaciones de seguimiento de pacientes, rastreo de contactos, escáner térmicos y cámaras de atención a distancia<sup>1</sup>. En este contexto, surgieron múltiples plataformas de inteligencia artificial (IA) con el objetivo de facilitar la detección de hallazgos radiológicos relacionados a la infección por COVID-19<sup>2</sup>.

En la neumonía por SARS-CoV-2, los algoritmos de IA detectan opacidades parcheadas bilaterales, que pueden variar en localización e intensificarse con el transcurso del tiempo. Son hallazgos similares a los de las neumonías virales, por lo que el análisis reviste un desafío tanto para el radiólogo como para el algoritmo en cuestión<sup>3</sup>.

En las economías en vías de desarrollo donde la gran mayoría de los especialistas en diagnóstico por imágenes se encuentran en los grandes centros urbanos, y donde además la accesibilidad a pruebas de PCR es limitada o presenta demoras significativas en la entrega de los resultados, es importante poder definir si los algoritmos de IA son una herramienta confiable, para servir de apoyo diagnóstico a los médicos de guardia y los centros periféricos de salud donde no se cuenta con disponibilidad de revisión por especialistas.

Por lo tanto, el objetivo de este trabajo fue evaluar el potencial que tienen diferentes algoritmos de IA para detectar compromiso pulmonar por COVID-19 en Rx de tórax portátiles frente.

## Material y método

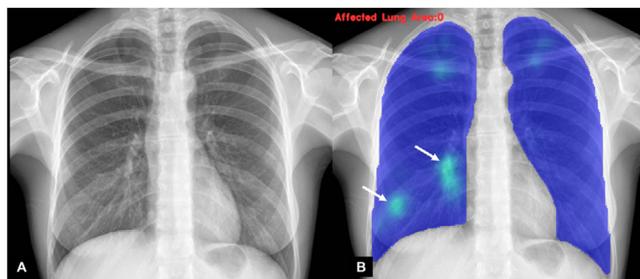
Estudio observacional prospectivo en pacientes consecutivos ingresados en el departamento de emergencias u hospitalizados por sospecha de infección por COVID-19 en un hospital universitario. La presencia de síntomas como fiebre, tos, disnea, anosmia y/o agnosia constituyeron en esta etapa de la pandemia los criterios de hisopado y realización de Rx de tórax.

Las imágenes de Rx de tórax fueron extraídas del sistema de comunicación y archivero de imágenes (PACS) en formato DICOM, mientras que los datos clínicos y de laboratorio fueron obtenidos a partir de la historia clínica electrónica.

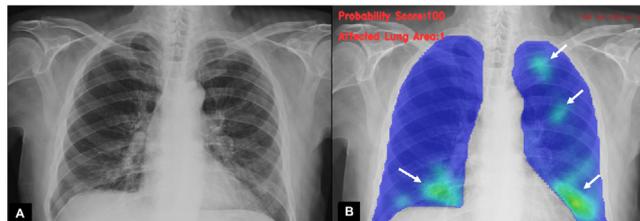
El patrón de referencia de compromiso pulmonar por COVID-19 fue definido como la presencia combinada de una prueba de PCR positiva y síntomas de infección pulmonar. De forma independiente y sin conocimiento de la sintomatología o antecedentes, se analizaron las Rx de tórax fuera de línea a fin de determinar cuáles de ellas mostraban hallazgos típicos de una neumonía por COVID-19.

Las Rx fueron analizadas independientemente por un médico especialista en diagnóstico por imágenes, así como también por 3 plataformas de IA con algoritmos de entrenamiento diferentes y acceso abierto específicamente diseñados para evaluar la detección de compromiso pulmonar por COVID-19 en dichas Rx de tórax. Las plataformas utilizadas fueron Pneuma Deep Health COVID (<http://pneuma.deephealth.thingtrack.com/>), CAD-4 (<https://www.delft.care/how-to-access/>) y ENTELAI (<https://covid.entelai.com/>).

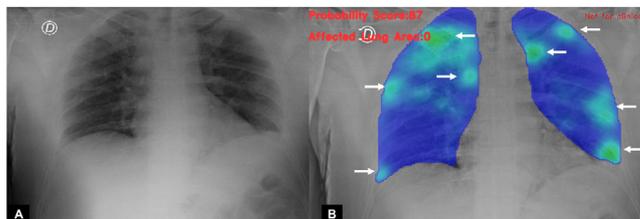
Se realizó una clasificación de los hallazgos por un médico radiólogo experimentado, así como por los algoritmos de IA, para establecer probabilidad de neumonía por COVID-19. Cada análisis fue determinado según grados de probabilidad de diversos tipos de afección pulmonar, entre otras, probabilidad de neumonía por



**Figura 1.** Estudio de Rx de tórax portátil en la cual el algoritmo de IA identifica algunas áreas de densidad parcheada (flechas en panel B) no sugestivas de proceso pulmonar viral.



**Figura 2.** A) Análisis de Rx de tórax portátil por algoritmo de IA en una paciente de 77 años, con diagnóstico confirmado de infección por SARS-CoV-2 mediante prueba de PCR. El algoritmo arrojó una probabilidad del 100% de infección por COVID-19, mostrando con color verde (flechas). B) Las opacidades identificadas en ambos campos pulmonares.



**Figura 3.** Paciente con diagnóstico confirmado de infección por SARS-CoV-2 mediante prueba de PCR. A) El resultado brindado por algoritmo de IA a partir del análisis de Rx de tórax portátil de probabilidad de neumonía por COVID-19 del 87%. B) Resultado y análisis del algoritmo de IA, mostrando en color verde (flechas) opacidades en ambos campos pulmonares.

COVID-19 (figs. 1-3). De forma independiente, el médico especialista categorizó los hallazgos entre estudios normales, neumonía por COVID u otros hallazgos.

Los principios de clasificación de cada estudio pueden ser visualizados en las páginas de acceso de cada algoritmo, para el ingreso de las imágenes.

## Análisis estadístico

Las variables continuas se reportaron como medias  $\pm$  desviación estándar mientras que las variables categóricas se reportaron como frecuencias y porcentajes. Se evaluó la sensibilidad, la especificidad, el valor predictivo positivo (VPP) y el valor predictivo negativo (VPN) para la detección de compromiso pulmonar por SARS-CoV-2. También comparamos la precisión diagnóstica de los distintos algoritmos (reas bajo la curva ROC, entendándose un rea de 0,50 como el azar, y un rea de 1,0 como la precisión perfecta) utilizando el método DeLong. Los análisis fueron realizados utilizando software SPSS® versión 22.0 (Armonk, NY, EE. UU.) y MedCalc® Statistical software versión 13.3.3 (MedCalc software bvba, Ostend, Bélgica).

## Resultados

Se incluyeron a un total de 493 pacientes evaluados en nuestra institución entre julio y noviembre de 2020 por sospecha de infección por COVID-19 a los que se le realizó Rx de tórax frente. La mediana de edad de los pacientes incluidos fue de 47 años (rango intercuartil 34; 71 años), 55% mujeres. Los pacientes fueron estudiados por presentar síntomas de COVID-19 en los días previos, tanto de manera ambulatoria, como aquellos que cursaban internación por otro motivo.

Los síntomas referidos fueron mayormente, disnea, tos seca o productiva, odinofagia, ageusia, anosmia y fiebre. Los pacientes por protocolo fueron estudiados con hisopado y PCR, laboratorio completo y Rx portátil inicial.

La prueba de PCR fue positiva en 140 (28%) pacientes, y 32 pacientes (7%) presentaron neumonía por SARS-CoV-2. Las tasas de detección de neumonía por SARS-CoV-2 fueron del 115 (23%) para el especialista; de 132 (27%) y 76 (15%) para el algoritmo AI-A con umbrales de probabilidad del 50 y 70%; de 334 (70%) y 157 (33%) para AI-B con umbrales de probabilidad del 50 y 70%; y de 143 (44%) y 45 (14%) para AI-C con umbrales de probabilidad del 50 y 70%.

El rendimiento diagnóstico del especialista para la detección de neumonía por SARS-CoV-2 mostró baja sensibilidad (16%, IC 95%: 5-34%) y moderada especificidad (76%, IC 95%: 72-80%); con un VPP del 4% (IC 95%: 1-10%) y VPN del 93% (IC 95%: 90-95%).

Utilizando un umbral de probabilidad superior al 50%, el algoritmo AI-A tuvo una sensibilidad del 25% (IC 95%: 11-43%), especificidad del 73% (IC 95%: 69-77%), VPP del 6% (IC 95%: 3-12%) y VPN del 93% (IC 95%: 90-96%); AI-B tuvo una sensibilidad del 97% (IC 95%: 84-100%), especificidad del 32% (IC 95%: 28-36%), VPP del 9% (IC 95%: 6-13%) y VPN del 99% (IC 95%: 96-100%); y AI-C una sensibilidad del 63% (IC 95%: 35-85%), especificidad del 57% (IC 95%: 52-63%), VPP del 7% (IC 95%: 3-12%) y VPN del 97% (IC 95%: 93-99%).

Utilizando un umbral de probabilidad superior al 70%, el algoritmo AI-A tuvo una sensibilidad del 15% (IC 95%: 5-33%), especificidad del 85% (IC 95%: 81-88%), VPP del 7% (IC 95%: 2-15%) y VPN del 94% (IC 95%: 91-96%); AI-B tuvo una sensibilidad del 50% (IC 95%: 32-68%), especificidad del 68% (IC 95%: 64-73%), VPP del 10% (IC 95%: 6-16%) y VPN del 95% (IC 95%: 92-97%); y AI-C una sensibilidad del 13% (IC 95%: 2-38%), especificidad del 86% (IC 95%: 82-90%), VPP del 4% (IC 95%: 1-15%) y VPN del 95% (IC 95%: 92-97%).

Mediante el análisis del conjunto de datos de los algoritmos de forma continua, el algoritmo AI-B tuvo el mejor rendimiento diagnóstico para la identificación de neumonía por SARS-CoV-2 (rea bajo la curva ROC AI-B 0,73 [IC 95%: 0,68-0,78] vs. AI-A 0,51 [IC 95%: 0,45-0,57]; vs. AI-C 0,57 [IC 95%: 0,51-0,62]). El mejor umbral de probabilidad para la identificación de neumonía por SARS-CoV-2 mediante el algoritmo AI-B fue del 55%, con una sensibilidad del 94% y una especificidad del 42%. Utilizando dicho umbral, el algoritmo AI-B presentó una sensibilidad del 94% (IC 95%: 79-99%), una especificidad del 42% (IC 95%: 38-47%), VPP del 10% (IC 95%: 7-15%) y VPN del 99% (IC 95%: 96-100%), con un área bajo la curva ROC superior a la del especialista (0,68 [IC 95%: 0,64-0,72] vs. 0,54 [IC 95%: 0,49-0,59]).

## Discusión

En el presente estudio, los algoritmos de IA basados en Rx portátiles permitieron una precisión diagnóstica comparable a la de la evaluación humana para la identificación de neumonía COVID-19. Si bien presentaron una baja sensibilidad y moderada especificidad, se asociaron a un elevado VPN que permitiría descartarla en la mayor parte de los casos. Según nuestro conocimiento, nuestro trabajo es el primero en evaluar el rendimiento diagnóstico de la

Rx portátil en este tipo de población. Estos hallazgos cobran relevancia en el contexto de grandes limitaciones de infraestructura y recursos humanos, particularmente en las economías emergentes; así como en la importancia de reducir los traslados intrahospitalarios, a fin de disminuir los contactos<sup>4</sup>. Por estos motivos, la Rx de tórax es considerada como el método por imagen de primera línea para evaluar anomalías en pacientes con síntomas pulmonares según la ACR<sup>5</sup>.

Estudios previos incluyendo distintas poblaciones, reportaron hallazgos variables respecto a la precisión diagnóstica de la Rx de tórax para la detección de neumonía por SARS-CoV-2<sup>6</sup>. Murphy K et al. reportaron un área bajo la curva ROC del 0,81 para la detección de neumonía por COVID-19, siendo en casi todos los segmentos mejor que el lector humano. Si bien no de forma concluyente debido a los moderados resultados de la Rx, identificamos un mayor rendimiento diagnóstico en uno de los algoritmos (utilizando un umbral de probabilidad mayor del 55%) comparado con el especialista. Cabe destacar que, si bien los 3 algoritmos fueron desarrollados sobre adquisiciones convencionales, nuestra experiencia fue aplicada sobre Rx portátiles, con la configuración original de los programas.

En nuestro trabajo, la relativamente similar detección entre el observador humano y el algoritmo, alienta a suponer que en un futuro, este proceso será cada día más eficiente y simple de implementar.

Los confundidores más frecuentes que pueden dificultar el diagnóstico de COVID-19 son las atelectasias, hemorragias, edema o neoplasias, entre otros, siendo en nuestra experiencia elementos que dificultan tanto el diagnóstico o exclusión de la enfermedad por COVID-19, como el entrenamiento de los algoritmos de IA.

El diagnóstico de neumonía por COVID-19 debe sustentarse en la clínica, el análisis de PCR, y al menos un estudio de imágenes de Rx. Nuestro trabajo adopta como estándar de referencia la presencia simultánea de PCR positiva y síntomas respiratorios pulmonares, en línea con lo expuesto por Albahri et al., que afirman que la asociación con datos clínicos optimiza la detección de verdaderos positivos<sup>6</sup>.

Deben destacarse las limitaciones propias del estándar de referencia utilizado, posiblemente afectando los resultados. De todas maneras, el rendimiento de la IA fue comparable al humano, siendo ambas estrategias afectadas de forma similar por dicha limitante. En línea con esto, escasos pacientes de la muestra se realizaron una Tc para confirmar o descartar los hallazgos.

## Conclusión

En nuestro estudio, los algoritmos de IA basados en Rx portátiles permitieron una precisión diagnóstica comparable a la de la evaluación humana para la detección de neumonía por SARS-CoV-2. Estos hallazgos son relevantes en el contexto de importantes limitaciones de recursos humanos, particularmente en las economías emergentes.

## Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

## Bibliografía

- Adly AS, Adly AS, Adly MS. Approaches Based on Artificial Intelligence and the Internet of Intelligent Things to Prevent the Spread of COVID-19: Scoping Review. *J Med Internet Res*. 2020;22:e19104. <http://dx.doi.org/10.2196/19104>.
- Shi F, Wang J, Shi J, Wu Z, Wang Q, Tang Z, et al. Review of Artificial Intelligence Techniques in Imaging Data Acquisition Segmentation, and Diagnosis for COVID-19. *IEEE Rev Biomed Eng*. 2021;14:4-15. <http://dx.doi.org/10.1109/RBME.2020.2987975>.

3. Chowdhury NK, Kabir MA, Rahman MM, Rezoana N. ECOVNet: a highly effective ensemble based deep learning model for detecting COVID-19. *PeerJ Comput Sci.* 2021;7:e551, <http://dx.doi.org/10.7717/peerj-cs.551>.
4. Monaco CG, Zaottini F, Schiaffino S, Villa A, Della Pepa G, Carbonaro LA, et al. Chest x-ray severity score in COVID-19 patients on emergency department admission: A two-centre study. *Eur Radiol Exp.* 2020;4:68, <http://dx.doi.org/10.1186/s41747-020-00195-w>.
5. Murphy K, Smits H, Knoops AJG, Korst MBJM, Samson T, Scholten ET, et al. COVID-19 on Chest Radiographs: A Multireader Evaluation of an Artificial Intelligence System. *Radiology.* 2020;296:E166–72, <http://dx.doi.org/10.1148/radiol.2020201874>.
6. Albahri OS, Zaidan AA, Albahri AS, Zaidan BB, Abdulkareem KH, Al-Qaysi ZT, et al. Systematic review of artificial intelligence techniques in the detection and classification of COVID-19 medical images in terms of evaluation and benchmarking: Taxonomy analysis, challenges, future solutions and methodological aspects. *J Infect Public Health.* 2020;13:1381–96, <http://dx.doi.org/10.1016/j.jiph.2020.06.028>.