



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

Elaboración de modelos predictivos de la gravedad y la mortalidad en pacientes con COVID-19 que acuden al servicio de urgencias, incluida la radiografía torácica



P. Calvillo-Batlles^{a,*}, L. Cerdá-Alberich^b, C. Fonfría-Esparcia^a, A. Carreres-Ortega^a, C.F. Muñoz-Núñez^a, L. Trilles-Olaso^a y L. Martí-Bonmatí^{a,b}

^a Servicio de Radiología, Hospital Universitario y Politécnico La Fe, Valencia, España

^b Grupo de Investigación Biomédica en Imagen (GIBI2³⁰), Instituto de Investigación Sanitaria La Fe, Valencia, España

Recibido el 27 de marzo de 2021; aceptado el 15 de septiembre de 2021

Disponible en Internet el 9 de noviembre de 2021

PALABRAS CLAVE

COVID-19;
Radiografía torácica;
Pronóstico;
Mortalidad;
Modelos predictivos;
Inteligencia artificial

Resumen

Objetivos: Desarrollar modelos de predicción de pronóstico para pacientes con COVID-19 que acuden a urgencias, basados en la radiografía de tórax inicial (RXT), parámetros demográficos, clínicos y de laboratorio.

Métodos: Se reclutaron todos los pacientes sintomáticos con COVID-19 confirmada, que ingresaron en urgencias de nuestro hospital entre el 24 de febrero y el 24 de abril de 2020. Los parámetros de la RXT, las variables clínicas y de laboratorio y los índices de hallazgos en RXT extraídos por una herramienta diagnóstica de inteligencia artificial en esta primera visita se consideraron potenciales predictores. El desenlace individual más grave definió los tres niveles de gravedad: 0) alta domiciliaria u hospitalización de 3 días o inferior, 1) hospitalización más de 3 días y 2) necesidad de cuidados intensivos o muerte. Se desarrollaron y validaron internamente modelos de predicción multivariable de gravedad y mortalidad hospitalaria. El índice de Youden se utilizó para la selección del umbral óptimo del modelo de clasificación.

Resultados: Se registraron 440 pacientes (mediana de 64 años; 55,9% hombres); el 13,6% de los pacientes fueron dados de alta, el 64% estuvo hospitalizado más de 3 días, el 6,6% requirió cuidados intensivos y un 15,7% falleció. El modelo de predicción de gravedad incluyó saturación de oxígeno/fracción de oxígeno inspirado (SatO₂/FiO₂), edad, proteína C reactiva (PCR), linfocitos, puntuación de la extensión de la afectación pulmonar en la RXT (ExtScoreRXT), lactato deshidrogenasa (LDH), dímero D y plaquetas, con AUC-ROC = 0,94 y AUC-PRC = 0,88. El modelo de predicción de mortalidad incluyó edad, SatO₂/FiO₂, PCR, LDH, ExtScoreRXT, linfocitos y dímero D, con AUC-ROC = 0,97 y AUC-PRC = 0,78. La adición de índices radiológicos obtenidos por inteligencia artificial no mejoró significativamente las métricas predictivas.

* Autor para correspondencia.

Correo electrónico: calvillo.mar@gva.es (P. Calvillo-Batlles).

Conclusión: Los modelos de predicción de pronóstico desarrollados podrían ser útiles para clasificar en urgencias a los pacientes con COVID-19 u otras infecciones víricas con comportamiento similar.

© 2021 SERAM. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

KEYWORDS

COVID-19;
Chest X-Ray;
Prognosis;
Mortality;
Predictive models;
Artificial intelligence

Development of severity and mortality prediction models for covid-19 patients at emergency department including the chest x-ray

Abstract

Objectives: To develop prognosis prediction models for COVID-19 patients attending an emergency department (ED) based on initial chest X-ray (CXR), demographics, clinical and laboratory parameters.

Methods: All symptomatic confirmed COVID-19 patients admitted to our hospital ED between February 24th and April 24th 2020 were recruited. CXR features, clinical and laboratory variables and CXR abnormality indices extracted by a convolutional neural network (CNN) diagnostic tool were considered potential predictors on this first visit. The most serious individual outcome defined the three severity level: 0) home discharge or hospitalization \leq 3 days, 1) hospital stay $>$ 3 days and 2) intensive care requirement or death. Severity and in-hospital mortality multivariable prediction models were developed and internally validated. The Youden index was used for the optimal threshold selection of the classification model.

Results: A total of 440 patients were enrolled (median 64 years; 55.9% male); 13.6% patients were discharged, 64% hospitalized, 6.6% required intensive care and 15.7% died. The severity prediction model included oxygen saturation/inspired oxygen fraction (SatO₂/FiO₂), age, C-reactive protein (CRP), lymphocyte count, extent score of lung involvement on CXR (ExtScoreCXR), lactate dehydrogenase (LDH), D-dimer level and platelets count, with AUC-ROC = 0.94 and AUC-PRC = 0.88. The mortality prediction model included age, SatO₂/FiO₂, CRP, LDH, CXR extent score, lymphocyte count and D-dimer level, with AUC-ROC = 0.97 and AUC-PRC = 0.78. The addition of CXR CNN-based indices did not improve significantly the predictive metrics.

Conclusion: The developed and internally validated severity and mortality prediction models could be useful as triage tools in ED for patients with COVID-19 or other virus infections with similar behaviour.

© 2021 SERAM. Published by Elsevier España, S.L.U. All rights reserved.

Introducción

La pandemia de COVID-19 plantea un gran problema para los sistemas sanitarios, que se ven forzados a encontrar un equilibrio entre la gestión de los recursos y una toma de decisiones segura, sin disponer de toda la información científica necesaria. El cuadro clínico inicial de la enfermedad abarca desde síntomas leves a una bronconeumonía bilateral con hipoxemia. Un síndrome de dificultad respiratoria aguda y la afectación de otros órganos vitales pueden complicar la evolución clínica y poner en peligro la vida del paciente. La incertidumbre hace que sea necesaria la elaboración de modelos específicos de la enfermedad para identificar a los pacientes en función del pronóstico y la gravedad, y de si precisan hospitalización o incluso ingreso en la unidad de cuidados intensivos. Las técnicas de diagnóstico por la imagen torácicas se han utilizado como herramienta diagnóstica en el servicio de urgencias (SU), ya que pueden revelar características de afectación pulmonar indicativas de COVID-19¹⁻⁵. Sin embargo, los estudios sobre la utilidad de la radiografía torácica (RXT) para predecir resultados sanitarios son escasos⁶⁻⁹ y los estudios pronósticos se han

basado principalmente en el uso de la tomografía computarizada (TC) torácica¹⁰⁻¹².

Teniendo en cuenta el uso y la disponibilidad más amplios de la RXT, así como su uso más seguro para evitar la propagación del virus en comparación con la TC, nuestro objetivo era elaborar dos modelos predictivos multivariables para calcular la gravedad y la mortalidad en la COVID-19 teniendo en cuenta las variables radiológicas, demográficas, clínicas y analíticas registradas en la valoración inicial en los servicios de urgencias.

Materiales y métodos

Este estudio retrospectivo fue aprobado por el comité de ética de la investigación. Esta investigación no ha recibido ninguna subvención específica de organismos de financiación del sector público, comercial o sin ánimo de lucro.

Se han seguido las recomendaciones de presentación transparente de un modelo de predicción multivariable para el pronóstico o diagnóstico individual: la declaración TRIPOD (*Transparent Reporting of a multivariable prediction model*

for Individual Prognosis or Diagnosis)¹³. El riesgo de sesgo y pertinencia se evaluó con la herramienta elaborada para la evaluación del riesgo de sesgo de estudios de modelos predictivos (*Prediction model Risk of Bias Assessment Tool, PROBAST*)¹⁴.

Pacientes

Fueron incluidos en el estudio todos los pacientes adultos sintomáticos consecutivos que acudieron al SU de nuestro hospital universitario entre el 24 de febrero y el 24 de abril de 2020 si se les había realizado una RXT y se había detectado ARN del coronavirus del síndrome respiratorio agudo grave de tipo 2 (SARS-CoV-2) en hisopos nasofaríngeos o esputo/lavado broncoalveolar. Los pacientes con diagnóstico final simultáneo distinto de COVID-19 fueron excluidos. Los médicos de urgencias realizaron el triaje de estos pacientes. Los pacientes oligosintomáticos con RXT y parámetros analíticos normales, saturación de oxígeno mayor del 95%, ausencia de enfermedades crónicas y menores de 65 años recibieron el alta domiciliaria. Los pacientes ingresados en el hospital recibieron el tratamiento de referencia vigente en aquel momento y se les dio el alta hospitalaria después de un mínimo de 3 días sin fiebre y mejoría de los síntomas respiratorios y parámetros analíticos (diagrama de flujo del estudio en la figura 1).

Evaluación de la radiografía torácica

Las imágenes de la RXT inicial en el momento del ingreso fueron distribuidas a cinco radiólogos con un promedio de 11 años de experiencia en el diagnóstico por la imagen torácica, en condiciones de enmascaramiento con respecto al resto

de parámetros y al desenlace. Se describieron las siguientes características (fig. 2):

- Ausencia (nivel 0) o presencia y densidad de las opacidades: de baja densidad únicamente (nivel 1) o consolidación (\pm de baja densidad) (nivel 2). Las opacidades pulmonares se consideraron de "baja densidad" si la atenuación no ocultaba los vasos subyacentes y "consolidación" si la opacificación del parénquima no permitía ver los vasos subyacentes.
- Distribución de opacidades: predominio periférico (en el tercio exterior de los pulmones)/central (en los dos tercios interiores de los pulmones)/ambas sin predominio claro; unilateral/bilateral; campos superior (suprahiliar)/medio (hiliar)/inferior (infrahiliar). Para determinar la distribución y el grado de afectación, se dividió cada pulmón en campo superior, medio e inferior, con un máximo de seis campos. Los lóbulos afectados no se registraron, ya que gran parte de las radiografías se realizaron con un equipo portátil (27,7%).
- Grado de extensión y puntuación de la afectación pulmonar: la extensión se clasificó como leve (si el tamaño de la opacidad era inferior a 1 campo); moderado (1-2 campos afectados); extenso (3-4 campos afectados), y muy extenso (5-6 campos afectados). Se asignó un valor numérico a cada campo en función del porcentaje con aumento de atenuación: 0 (0%), 1 ($\leq 50\%$) y 2 ($>50\%$). Se alcanzó una puntuación total del grado de afectación pulmonar (ExtScoreRXT) sumando las puntuaciones de los seis campos, con lo que se obtenía un valor entre 0 y 12. Los autores crearon ExtScoreRXT después de considerar, por consenso, que se trataba de un método óptimo, reproducible y rápido para semicuantificar el grado de afectación pulmonar. Otros autores también han usado la puntuación del

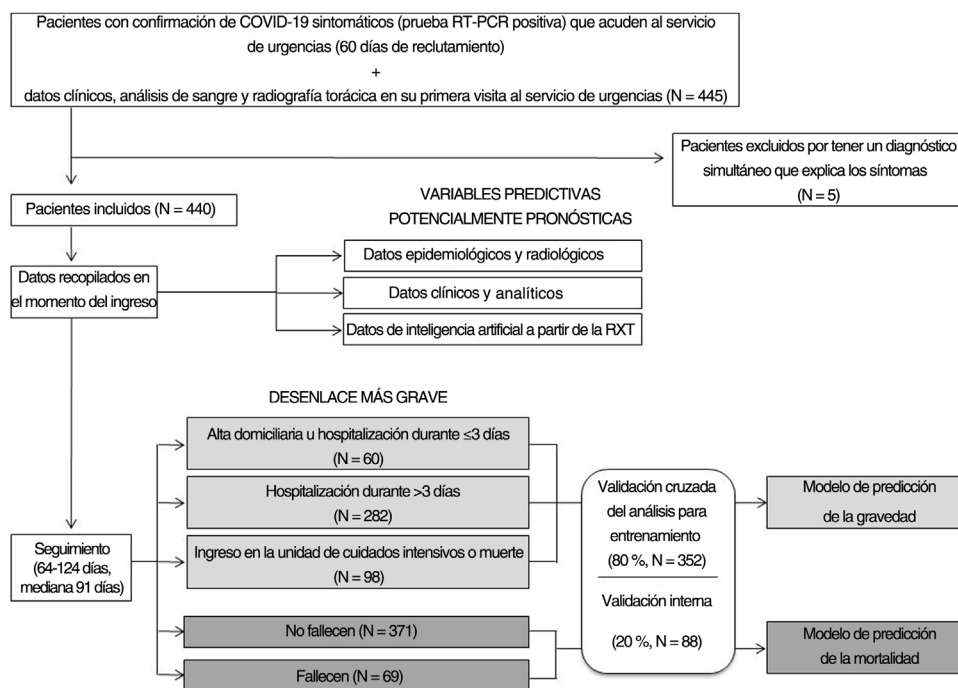


Figura 1 Diagrama de flujo del estudio.

RT-PCR: reacción en cadena de la polimerasa en tiempo real con retrotranscripción.

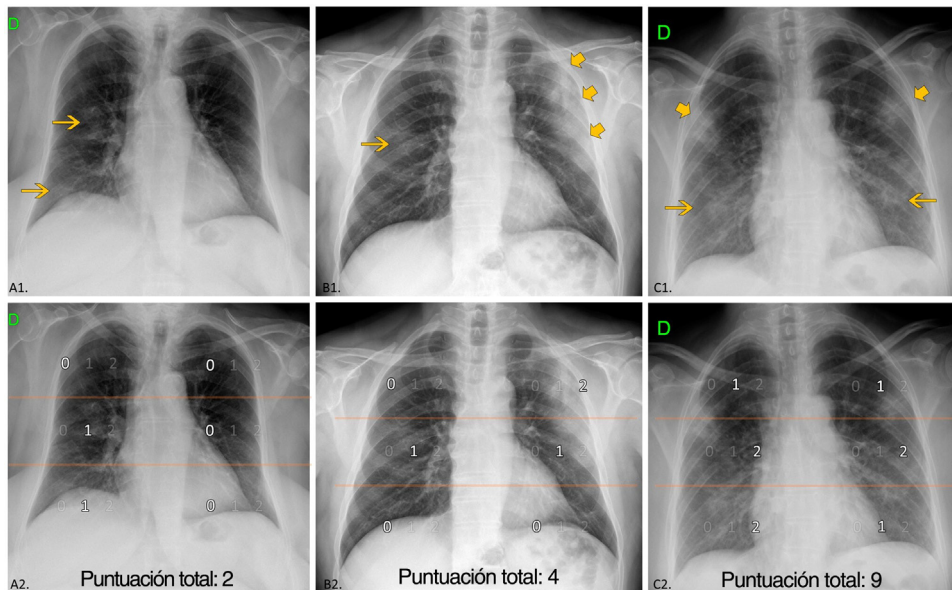


Figura 2 Manifestaciones en la radiografía torácica (RXT). Arriba: ejemplos de distribución, densidad de opacidades y puntuación de la extensión. Abajo (la misma RXT): división de los campos pulmonares. Campos superiores (área suprahiliar) limitados por la línea que pasa por debajo del cayado aórtico; campos medios (área hiliar) y campos inferiores (área infrahiliar) separados por la línea que divide el resto del pulmón en dos mitades (con frecuencia, esta línea cruza la bifurcación de la arteria lobar inferior derecha). A1) Opacidades de baja densidad centrales y periférica unilaterales (flecha), sin predominio. A2) Campos medio e inferior del lado derecho afectados $\leq 50\%$; ExtScoreRXT = 2. B1) Opacidad de baja densidad derecha central (flecha) y consolidación izquierda periférica (flechas más grandes). Predominio periférico. B2) campos medios derecho e izquierdo con un 50% o menos de afectación y campo izquierdo superior con más del 50% de afectación; ExtScoreRXT = 4. C1) opacidades de baja densidad bilaterales (flechas) y consolidaciones (flechas más grandes) sin predominio. C2) campos superiores y campo izquierdo inferior con 50% o menos de afectación, campos medios y campo derecho inferior con más del 50% de afectación; ExtScoreRXT = 9.

grado de afectación basada en la división del pulmón en campos pulmonares en los pacientes con COVID-19^{6-8,15}.

- Los radiólogos acordaron las variables del diagnóstico por la imagen a partir de la radiografía de los primeros 80 casos detectados. Se utilizó un *software* interno de almacenamiento de datos para registrar todas las variables en una base de datos compartida estructurada, con recordatorios para las descripciones y las imágenes con el fin de reducir la variabilidad entre evaluadores, y con campos de cumplimentación obligatoria para optimizar la recopilación de datos.

Variables demográficas, clínicas y analíticas recopiladas en la primera visita al servicio de urgencias

Se registraron: datos demográficos, internamiento, enfermedades concomitantes, manifestaciones clínicas, saturación periférica de oxígeno (SatO₂), datos analíticos, proteína C-reactiva (PCR), lactato-deshidrogenasa (LDH), recuento de linfocitos, recuento de plaquetas y dímero D. Calculamos SatO₂/FiO₂ para evitar la pérdida de datos de pacientes con SatO₂ obtenida durante la oxigenoterapia. La FiO₂ es la fracción inspirada de oxígeno y cambia en función del flujo de oxígeno que se administra a cada paciente; en el caso del aire ambiente es de 0,21.

Imágenes informáticas (datos de inteligencia artificial)

Los índices de probabilidad de obtener resultados pulmonares anómalos se consiguieron a partir de la RXT mediante una herramienta diagnóstica basada en una red neuronal convolucional (RNC), QUIBIM Precision RXT v2.0.0 (QUIBIM S.L) con marcado CE de clase IIa. El algoritmo incluye un conjunto de modelos de aprendizaje profundo que calculan la probabilidad de obtener diferentes resultados torácicos y la probabilidad de anomalía en la RXT. Un valor de 0 significaría que no hay ninguna probabilidad de presencia de ese hallazgo patológico y un valor de 1 significaría la certeza total de presencia de ese hallazgo patológico. *Consúltese, en el anexo 1, material complementario sobre la herramienta basada en una RNC.* Los índices de "consolidación", "opacidad pulmonar" y "RXT anómala" se incorporaron al modelo final para evaluar si mejoraban su exactitud predictiva.

Criterios de valoración

Se definieron tres niveles de gravedad: alta domiciliaria u hospitalización durante 3 días o menos (nivel 0), hospitalización durante más de 3 días (nivel 1), necesidad de ingreso en la unidad de cuidados intensivos (UCI) o muerte por COVID-19 (nivel 2). Se registraron tanto los días de hospitalización

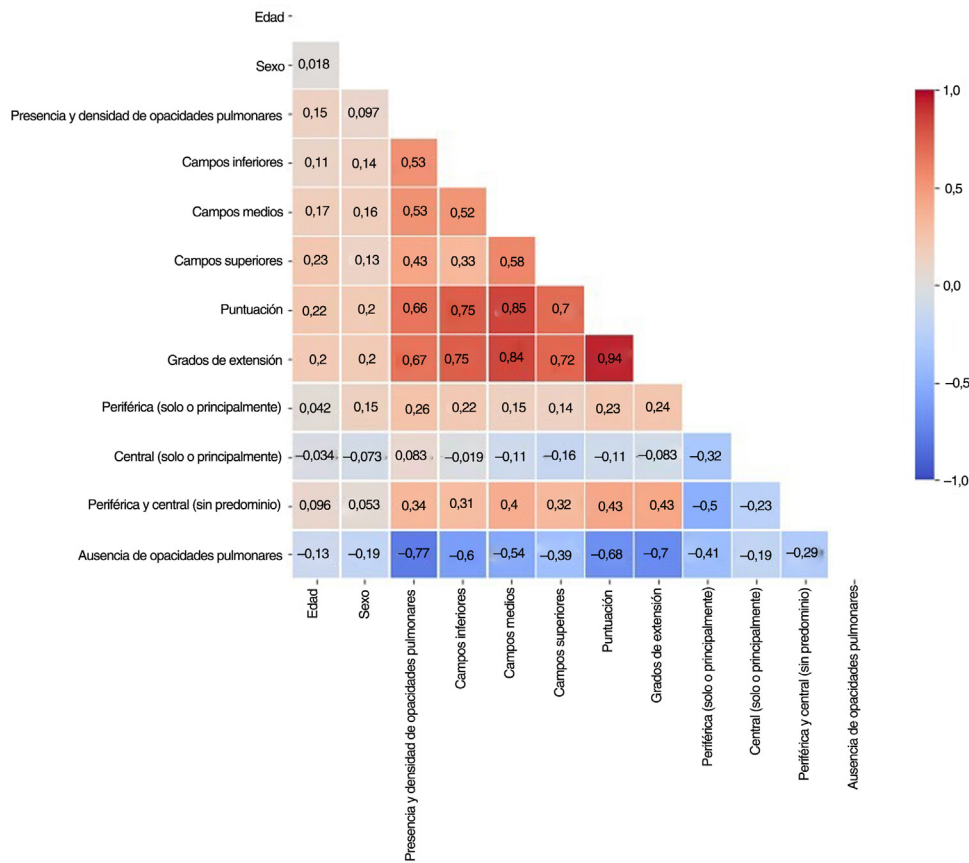


Figura 3 Matriz de correlación de rangos de Spearman para las manifestaciones epidemiológicas y radiológicas.

como los días transcurridos hasta la muerte. La mediana de seguimiento fue de 91 días (intervalo 64-124 días).

Análisis univariable

Se investigaron las correlaciones entre el alcance de la afectación pulmonar en la RXT (grados y puntuación) y los días con síntomas, $\text{SatO}_2/\text{FiO}_2$ y los criterios de valoración. Se utilizaron las pruebas biserialadas puntuales o por rangos, de Spearman o de Kendal, en función del tipo de variables estudiadas. Para interpretar la solidez de una relación en función de su valor de r (usando el valor absoluto del valor de r para que todos los valores fueran positivos), aplicamos la regla general siguiente: $r < 0,1$ ninguna; $0,1 < r < 0,3$ débil; $0,3 < r < 0,5$ moderada; $0,5 < r < 0,8$ fuerte y $r > 0,8$ muy fuerte.

Modelos predictivos pronósticos

Se elaboraron distintos modelos predictivos pronósticos con tres tipos de clasificadores o métodos de agrupación (potenciación del gradiente [*Gradient Boosting*], bosque aleatorio [*Random Forest*] y máquina de vectores de soporte [*Support Vector Machine*]) y aplicando una validación cruzada estratificada con el 80% de la población. Estos clasificadores se han elegido porque intrínsecamente son perfectos para resolver los problemas de clasificación con dos o más clases. Por ejemplo, el método de máquina de vectores de soporte

es un modelo lineal que escala relativamente bien hasta datos dimensionales elevados y es menos probable que se produzca un sobreajuste. Tanto el bosque aleatorio como la potenciación del gradiente son métodos de agrupación que consisten en entrenar múltiples clasificadores débiles y fusionar sus resultados para construir un "clasificador potente". Difieren en dos puntos clave: la forma de definir los grupos de entrenamiento para cada modelo de base y el orden de entrenamiento de los clasificadores débiles. En particular, los bosques aleatorios crean muestras de entrenamiento aleatorias a partir del grupo completo de entrenamiento en función de una selección aleatoria de las observaciones y las características (*bootstrapping* o método de muestreo con reposición). Un clasificador débil se entrena en paralelo en cada uno de los grupos de entrenamiento derivados. En cuanto al método de potenciación del gradiente, consiste en ajustar varios clasificadores débiles secuencialmente y, en cada repetición, añadir más peso a las observaciones con la peor predicción en la repetición previa. Puesto que cada clasificador débil se construye a partir de los resultados del anterior, el cálculo no se puede realizar de forma paralela y puede ser más largo.

Se realizó una validación interna con una base de datos nueva correspondiente al 20% restante para valorar la capacidad de generalizar el método y su robustez.

Los hiperparámetros del modelo se obtuvieron mediante una estrategia de búsqueda en cuadrícula, un método para optimizar hiperparámetros que construye y evalúa un

modelo de forma metódica para cada combinación de los parámetros del algoritmo especificados en una cuadrícula.

Con el fin de evitar información redundante, impedir que los modelos se volvieran inestables en presencia de dependencias de características fuertes y mejorar su interpretabilidad, se identificaron las características de correlación elevada (>80%) con una matriz de correlación de rangos Spearman (fig. 3). De cada par de características con correlación elevada se excluyó de los modelos aquella con el mayor valor de p en la prueba estadística univariable. Según este criterio, se descartaron los grados de extensión y la distribución de las opacidades en el campo pulmonar medio.

Se elaboraron tres modelos con diferentes variables predictivas; el primero contenía las características epidemiológicas (edad, sexo, internamiento y enfermedades concomitantes) y todas las características radiológicas mencionadas anteriormente. Los parámetros clínicos (síntomas y $\text{SatO}_2/\text{FiO}_2$) y todos los parámetros analíticos mencionados anteriormente se incorporaron al primer modelo para crear el segundo modelo y, finalmente, los datos derivados de la RNC se incorporaron al segundo modelo para crear el tercer modelo. Se utilizó una metodología de submuestreo seguida de una técnica de sobremuestreo minoritario sintético (*synthetic minority over-sampling technique*, SMOTE) para abordar el problema de desequilibrio de datos, muy habitual en los entornos de aprendizaje automático¹⁶. Las características se ajustaron en consecuencia.

Como método de reducción de la dimensionalidad y para evaluar el impacto de cada característica, se calculó la importancia variable. Esta importancia es una medida para calcular la magnitud de reducción y aumento de la exactitud al extraer o introducir una variable, respectivamente. El método por defecto para calcular la importancia variable es la reducción media del mecanismo de impureza (o importancia Gini): a cada división en cada árbol, la mejora del criterio de división es la medida de importancia atribuida a la variable de división, y se acumula en los tres árboles del bosque de forma separada para cada variable. Se debe tener en cuenta que esta medición es bastante parecida a R^2 en la regresión del grupo de entrenamiento. Si una variable tiene muy poca potencia predictiva, eliminarla puede incrementar la exactitud debido a ruido aleatorio.

En cada modelo se obtuvieron la sensibilidad, la especificidad, el valor predictivo de un resultado positivo (VPP), el valor predictivo de un resultado negativo (VPN), el área bajo la curva ROC (AUC-ROC) y las curvas de precisión-sensibilidad (AUC-CPS). El índice de Youden se utilizó para la selección del umbral óptimo del modelo de clasificación, maximizando la sensibilidad más elevada y el VPN para los pacientes muy graves (o fallecidos), y la especificidad más elevada y el VPP para los pacientes leves (o vivos). Los umbrales óptimos se definieron en el grupo de datos de entrenamiento. Se utilizó un abordaje estadístico de micropromedio ponderado para obtener los valores en función del nivel de gravedad, después de optimizar el umbral del modelo de clasificación con el índice de Youden. Un micropromedio calculará el parámetro de forma independiente para cada clase y, después, utilizará el promedio (por consiguiente, se tratarán todas las clases de forma equitativa), mientras que un micropromedio agrupará las contribuciones

de todas las clases para calcular el promedio del parámetro, que es especialmente útil si el conjunto de datos cambia de tamaño. Para analizar una posible sobrestimación del rendimiento del modelo diagnóstico, los parámetros se obtuvieron mediante la evaluación de los mismos umbrales en el conjunto de datos de validación interna. La prueba de DeLong de dos curvas ROC correlacionadas¹⁷ se utilizó para comparar el rendimiento de los modelos.

Se utilizaron las siguientes bibliotecas de Python y de aprendizaje automático para visualizar los datos y realizar el análisis estadístico del estudio: Pandas, Numpy, SciPy, Matplotlib y Scikit Learn.

Resultados

Pacientes

De los 445 pacientes inscritos, 5 fueron excluidos (1 apendicitis aguda, 1 colangitis, 1 diverticulitis, 1 ictus y 1 insuficiencia cardíaca). Se incluyó una población final total de 440 pacientes en el estudio.

Datos demográficos, enfermedades concomitantes

La mediana de edad era de 64 años (intervalo 17-100) y el 55,9% eran varones. El 79% sufría una o más enfermedades concomitantes; las más frecuentes eran hipertensión, dislipidemia y diabetes (tabla 1).

Variables clínicas y analíticas, $\text{SatO}_2/\text{FiO}_2$

A su llegada al SU, el promedio de días con síntomas era de 6,8 (intervalo 0-30) y los síntomas más habituales eran, en este orden, fiebre y tos. El promedio de saturación de oxígeno era del 93,7% (intervalo 55-100%). Había pérdida de uno o más parámetros analíticos en 67 pacientes por no haberse solicitado (tabla 2).

Criterios de valoración

El 13,6% de los pacientes recibieron el alta domiciliaria o estuvieron hospitalizados durante 3 días o menos; el 64% de los pacientes estuvieron hospitalizados (4-54 días, promedio de 17 días); el 6,6% precisó ingreso en la unidad de cuidados intensivos (UCI) (2-65 días, promedio de 18 días en la UCI) y el 15,7% falleció (0-51 días después del ingreso, promedio de 10 días).

Evaluación de la radiografía torácica

La mediana de tiempo entre la RXT y la reacción en cadena de la polimerasa en tiempo real con retrotranscripción (RT-PCR) fue de 1 día (intervalo 0-30). El 65,9% de los pacientes con resultado de RT-PCR pendiente presentaron afectación pulmonar indicativa de COVID-19 en la RXT, lo que adelantó el diagnóstico definitivo. La puntuación ExtScoreRXT fue de $3,3 \pm 3,07$ (promedio \pm DE) (tabla 3).

De los 76 pacientes que recibieron el alta domiciliaria inicialmente, el 24% fue ingresado en una segunda visita al SU. Siete de estos pacientes tuvieron una primera RXT normal y

Tabla 1 Características de los pacientes con COVID-19. Variables demográficas y enfermedades concomitantes investigadas como posibles factores pronósticos

Información demográfica	
Edad* Mediana (amplitud intercuartílica), (intervalo), años	64 (51-79), (17-100)
Sexo	N.º (% de 440)
Varones	246 (55,9)
Mujeres	194 (44)
Enfermedades concomitantes	
Hipertensión	191 (43,4)
Dislipidemia	125 (28,4)
Diabetes	93 (21,1)
Internamiento	75 (17)
Enfermedad cardiovascular	56 (12,7)
Inmunodepresión	46 (10,5)
Corticoterapia y otros inmunosupresores	16 (3,6)
Enfermedad renal crónica avanzada/diálisis	13 (2,9)
Enfermedad/neoplasia hematológica	10 (2,3)
Trasplante de víscera sólida	8 (1,8)
Virus de la inmunodeficiencia humana	1 (0,2)
Enfermedad renal crónica (total)	43 (9,8)
Obesidad	42 (9,5)
Enfermedad cerebrovascular	36 (8,2)
Neoplasia	32 (7,3)
Asma	25 (5,7)
Demencia	19 (4,3)
Fumador activo	16 (3,6)
Apnea obstructiva del sueño	16 (3,6)
Exfumador	15 (3,4)
Hipotiroidismo	13 (2,9)
Fibrilación auricular	13 (2,9)
Enfermedad pulmonar obstructiva crónica	9 (2)
Número de enfermedades concomitantes	
Ninguna	92 (20,9)
1	87 (19,8)
2	68 (15,4)
3	54 (12,3)
≥4 (4-8)	139 (31,5)

* Factores pronósticos en los modelos predictivos finales de pronóstico.

11 presentaron opacidades muy escasas o difíciles de interpretar; todos tuvieron progresión de la afectación pulmonar en la segunda visita.

Imágenes informáticas (datos de inteligencia artificial)

Los índices de probabilidad (promedio ± DE [intervalo]) para "consolidación", "opacidad pulmonar" y "RXT anómala" obtenidos a partir de la RXT de la población estudiada fueron: $0,39 \pm 0,19$ (0-0,84); $0,47 \pm 0,25$ (0-0,98), y $0,98 \pm 0,13$ (0-1), respectivamente.

Análisis univariable. Extensión de la afectación pulmonar en la RXT

El grado de afectación pulmonar (grados y puntuación) mostró una escasa correlación con los días de duración de los

síntomas ($r=0,198$ y $r=0,176$, respectivamente, valor de $p < 0,001$), una correlación negativa fuerte con $\text{SatO}_2/\text{FiO}_2$ ($r = -0,53$ y $r = -0,57$, respectivamente, valor de $p < 0,001$), una correlación fuerte-moderada con el nivel de gravedad ($r=0,536$ y $r=0,491$, respectivamente, valor de $p < 0,001$), una escasa correlación con los días de hospitalización ($r=0,240$ y $r=0,246$, respectivamente, valor de $p < 0,001$), una correlación no significativa con los días de ingreso en la UCI, y una correlación escasa-moderada con la mortalidad ($r=0,277$ y $r=0,310$, respectivamente, valor de $p < 0,001$).

Modelos predictivos pronósticos

La $\text{SatO}_2/\text{FiO}_2$ (33%), el índice basado en la RNC para la consolidación pulmonar (13%), la LDH (12%), la ExtScoreRXT (9%), la edad (9%), el recuento de linfocitos (9%), la PCR (7%),

Tabla 2 Cuadro clínico inicial de los pacientes con COVID-19. Variables clínicas y analíticas investigadas como posibles factores pronósticos. Síntomas gastrointestinales: diarrea, vómitos o dolor abdominal

Intervalo desde la aparición de los síntomas	Promedio ± DE (intervalo)
Días	6,8 ± 4,5 (0-30)
Síntomas N.º total 441	N.º (%)
Fiebre	361 (82)
Tos	287 (65,2)
Disnea	178 (40,4)
Molestia general/astenia	171 (38,9)
Síntomas gastrointestinales	103 (23,4)
Mialgias	72 (16,4)
Cefalea	47 (10,7)
Hiposmia/disgeusia	32 (7,3)
Saturación de oxígeno	Promedio ± DE (intervalo)
SatO ₂ (aire ambiente)	93,7 ± 5,6 (55-100)
SatO ₂ /FiO ₂ *	443,5 ± 34,4
Datos analíticos	Promedio/Mín./25%/Mediana/75%/Máx.
Lactato-deshidrogenasa (U/l) *	306/16/218,2/265,5/349,7/2146
Proteína C-reactiva (mg/l) *	81/0,3/15,5/44,6/117,2/655,5
Recuento de linfocitos (× 103 µ/l) *	1,2/0,06/0,7/1/1,4/12,3
Recuento de plaquetas (× 103 µ/l) *	203,7/32/150/193/242/716
Dímero D (ng/l) *	1390/83/350/608/1015/38282

* Factores pronósticos en los modelos predictivos definitivos del pronóstico.

DE: desviación estándar; SatO₂/FiO₂: saturación periférica de oxígeno/fracción inspirada de oxígeno (aire ambiente u oxigenoterapia).

Tabla 3 Afectación pulmonar en la radiografía de tórax (TXT), distribución y extensión. Manifestaciones de la RXT investigadas como posibles factores pronósticos

Presencia y densidad de opacidades pulmonares	N.º (%) Total 440
Ausencia de opacidades pulmonares (± otros hallazgos) - nivel 0-	86 (19,5)
Opacidad(es) de baja densidad - nivel 1 -	254 (57,7)
Consolidaciones ± opacidad(es) de baja densidad - nivel 2-	100 (22,7)
Distribución/localización de las opacidades pulmonares	N.º (%) Total 354
Bilateral	212 (59,9)
Unilateral	142 (39,8)
Periférica (solo o principalmente)	182 (51,1)
Periférica y central (sin predominio)	116 (32,5)
Central (solo o principalmente)	56 (15,7)
Campos inferiores	290 (81,4)
Campos medios	259 (72,7)
Campos superiores	178 (50)
Grados de extensión de las opacidades pulmonares	N.º (%) Total 440
Ausencia de opacidades pulmonares (± otros hallazgos) - nivel 0-	86 (19,5)
Leve	66 (15)
Moderado	111 (25,2)
Extendido	101 (22,9)
Muy extendido	76 (17,3)
Puntuación de la extensión de la afectación pulmonar*	Promedio ± DE (intervalo)
Puntuación (ExtScoreRXT)	3,3 ± 3,07 (0-12)

* Factor pronóstico en los modelos predictivos de pronóstico finales.

ExtScoreRXT: puntuación del grado de afectación pulmonar en la RXT.

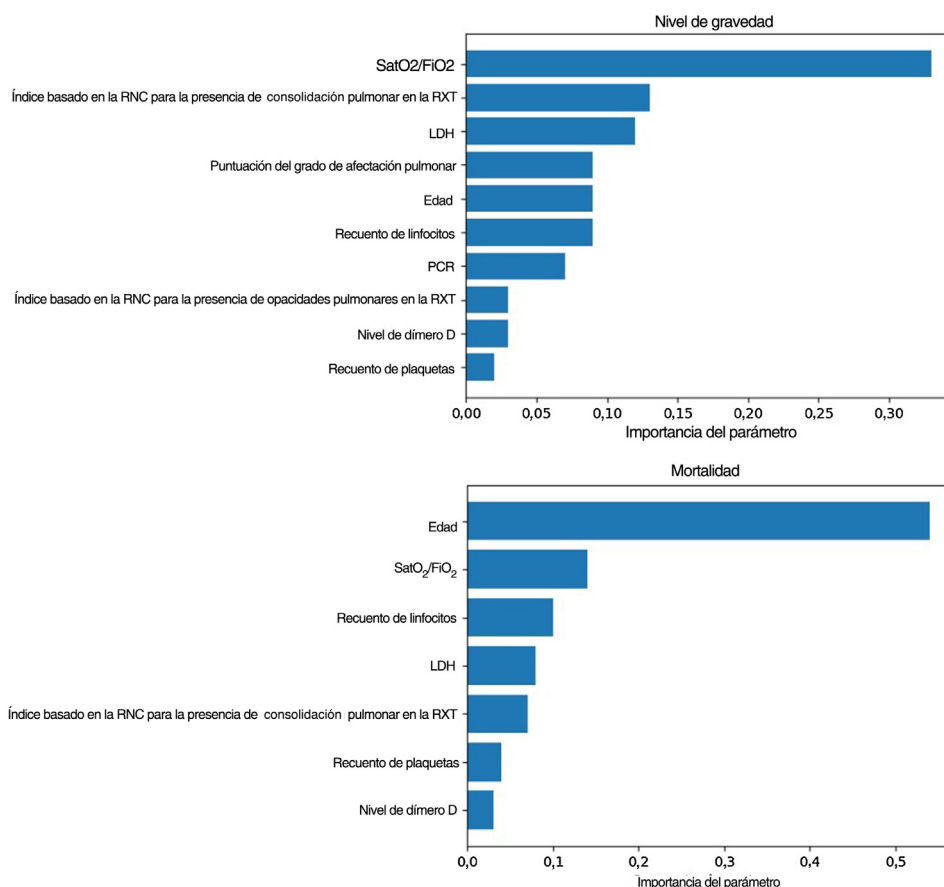


Figura 4 Importancia de los factores pronósticos del modelo obtenidos para el *nivel de gravedad* (arriba): saturación de oxígeno/fracción inspirada de oxígeno (SatO₂/FiO₂) (33%), índice basado en la red neuronal convolucional (RNC) para la presencia de consolidación pulmonar en la radiografía torácica (RXT) (13%), lactato-deshidrogenasa (LDH) (12%), puntuación del grado de afectación pulmonar en la RXT (ExtScoreRXT) (9%), edad (9%), recuento de linfocitos (9%), proteína C reactiva (PCR) (7%), índice basado en la RNC para la presencia de opacidades pulmonares en la RXT (3%), nivel de dímero D (3%) y recuento de plaquetas (2%); y *la mortalidad* (abajo): edad (54%), SatO₂/FiO₂ (14%), recuento de linfocitos (10%), LDH (8%), índice basado en la RNC para la presencia de consolidación pulmonar en la RXT (7%), recuento de plaquetas (4%) y nivel de dímero D (3%). Aquellos con una importancia menor a 0,01 fueron excluidos del modelo.

el índice basado en la RNC para las opacidades pulmonares (3%), el nivel de dímero D (3%) y el recuento plaquetario (2%) fueron, en este orden, los factores pronósticos más importantes del criterio de valoración del nivel de gravedad para el grupo de pacientes más graves (fig. 4). Los valores entre paréntesis corresponden a la importancia variable en el modelo elaborado.

Las curvas ROC y CPS de la validación interna realizada con un conjunto de datos nuevo para los modelos predictivos del pronóstico del nivel de gravedad construidas con tres combinaciones distintas de características se presentan en la figura 5. Las curvas basadas en el nivel de gravedad se obtienen después de aplicar una metodología de clasificación de uno frente a todos. Se observa una mejora del AUC-ROC y del AUC-CPS tanto para los pacientes más graves como para los menos graves, cuando se incluyen más características en el modelo. En particular, el efecto mayor se obtiene al añadir los parámetros clínicos y analíticos (micropromedio de AUC-ROC = 0,94, micropromedio de AUC-CPS = 0,88). La adición de los índices basados en la RNC aumenta el valor del AUC-CPS de los pacientes correspondientes a los niveles de

gravedad extrema, pero tiene el efecto contrario en los niveles de gravedad media, lo que da lugar a un empeoramiento de los parámetros predictivos (micropromedio de AUC-ROC = 0,91, micropromedio de AUC-CPS = 0,82) (tabla 4).

Con respecto al criterio clínico de valoración de la mortalidad, el mejor modelo se consigue con un clasificador de potenciación del gradiente, con la inclusión de los parámetros epidemiológicos, radiológicos, clínicos y analíticos seleccionados y los índices basados en la RNC (AUC-ROC = 0,97, AUC-CPS = 0,83). La edad (54%), la SatO₂/FiO₂ (14%), el recuento de linfocitos (10%), la LDH (8%), el índice basado en la RNC para la consolidación pulmonar en la RXT (7%), el recuento plaquetario (4%) y el nivel de dímero D (3%) fueron, en este orden, los factores pronósticos más ponderados de la mortalidad hospitalaria (fig. 4). Se observa una disminución en estos parámetros al eliminar del modelo los índices basados en la RNC (AUC-ROC = 0,97, AUC-CPS = 0,78), pero se alcanza una mejora del VPP y la especificidad del modelo mediante la optimización del umbral con el índice de Youden (tabla 4). En este caso, la edad (43%), la SatO₂/FiO₂ (20%), la PCR (15%), la LDH (7%), la

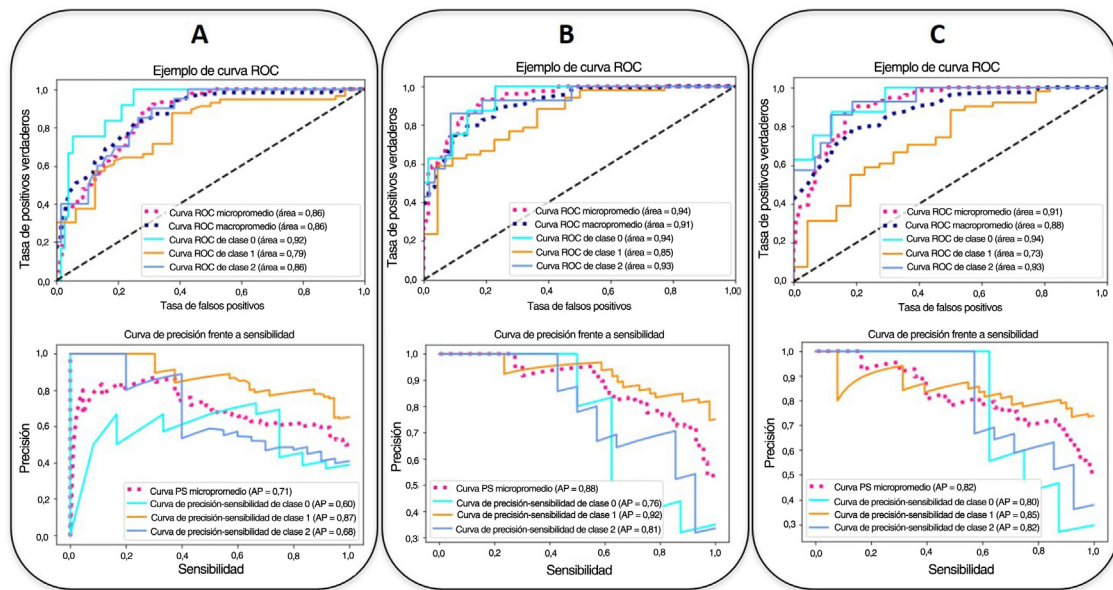


Figura 5 Curvas ROC (arriba) y curva de precisión-sensibilidad (CPS) (abajo) de la validación interna realizada con un conjunto de datos nuevo para los modelos predictivos de gravedad elaborados con tres combinaciones distintas de parámetros. (A: parámetros epidemiológicos y radiológicos, B: parámetros epidemiológicos, radiológicos, clínicos y analíticos, C: parámetros epidemiológicos, radiológicos, clínicos, analíticos y basados en la red neuronal convolucional (RNC). Las curvas según el nivel de gravedad se obtienen con una metodología de clasificación de uno frente a todos: alta domiciliaria u hospitalización durante 3 días o menos (nivel/clase 0, en azul claro), necesidad de hospitalización durante más de 3 días (nivel/clase 1, en naranja), necesidad de ingreso en la unidad de cuidados intensivos o muerte por COVID-19 (nivel/clase 2, en azul). Las líneas discontinuas representan los estadísticos de la curva de micropromedio (magenta) y macropromedio (azul oscuro) que tienen en cuenta las contribuciones de un único nivel. Se muestran los valores del AUC correspondientes. Precisión = positivo verdadero / (positivo verdadero + falso positivo). Sensibilidad = positivo verdadero / (positivo verdadero + falso positivo).

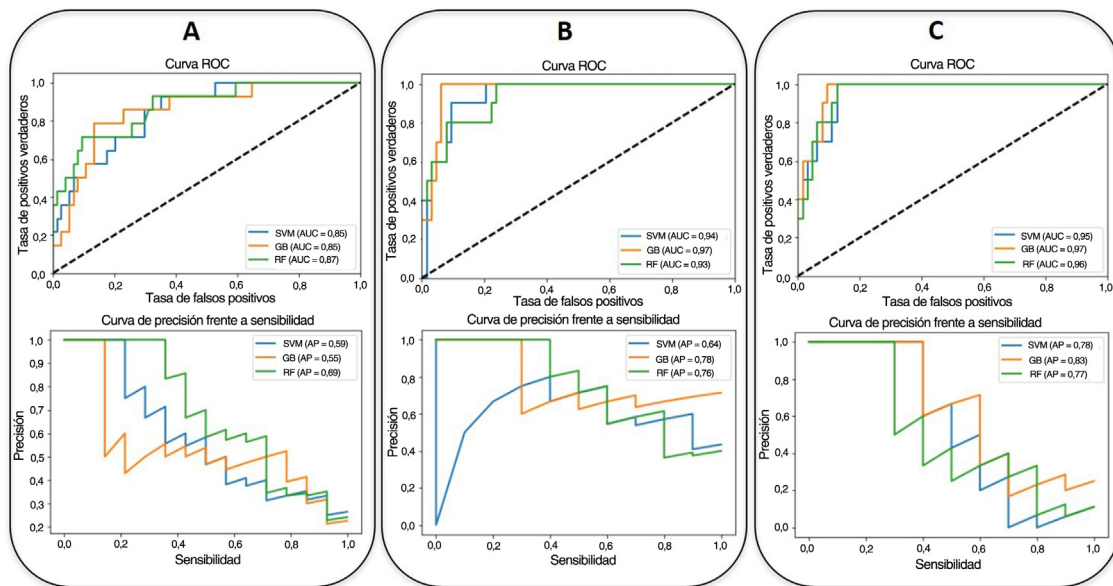


Figura 6 Las curvas ROC (arriba) y CPS (abajo) de la validación interna realizada con un conjunto de datos nuevo para los modelos predictivos de mortalidad hospitalaria (SVM: máquinas de vectores de soporte [*Support Vector Machine*, azul]; RF: bosque aleatorio [*Random Forest*, verde]; GB: potenciación del gradiente [*Gradient Boosting*, naranja]) elaborados con tres combinaciones distintas de parámetros (A: parámetros epidemiológicos y radiológicos; B: parámetros epidemiológicos, radiológicos, clínicos y analíticos; C: parámetros epidemiológicos, radiológicos, clínicos, analíticos y basados en la RNC). Se muestran los valores del AUC correspondientes.

Tabla 4 Parámetros de los modelos predictivos de la gravedad y la mortalidad*

Modelos predictivos de gravedad							
Combinación de parámetros	Sensibilidad (%)	Especificidad (%)	VPP (%)	VPN (%)	AUC-ROC (%)	AUC-CPS (%)	
Epidemiológicos y radiológicos	92,5	68,2	59,8	93,6	85,7	71,3	
Epidemiológicos, radiológicos, clínicos y analíticos	83,1	87,9	77,5	95,1	93,8	87,6	
Epidemiológicos, radiológicos, clínicos, analíticos y basados en la RNC	86,4	84,1	72,5	96,7	91,4	82,3	
Modelos predictivos de la mortalidad							
Combinación de parámetros	Estructura del modelo	Sensibilidad (%)	Especificidad (%)	VPP (%)	VPN (%)	AUC-ROC (%)	AUC-CPS (%)
Epidemiológicos y radiológicos	RF	71,8	90,3	59,5	92,2	86,7	69,3
Epidemiológicos, radiológicos, clínicos y analíticos	GB	90,0	93,7	69,2	98,3	96,5	77,9
Epidemiológicos, radiológicos, clínicos, analíticos y basados en la RNC	GB	90,0	92,1	64,3	98,3	97,1	83,4

* Parámetros de rendimiento de la validación interna realizada con un conjunto de datos nuevo para cada uno de los modelos predictivos de gravedad y mortalidad hospitalaria seleccionados, elaborados con tres combinaciones distintas de parámetros. El índice de Youden se utilizó para la selección del umbral óptimo de los modelos de clasificación.

AUC-ROC: área bajo la curva ROC; AUC-CPS: curva de precisión-sensibilidad; GB: potenciación del gradiente (*Gradient Boosting*); RF: bosque aleatorio (*Random Forest*); RNC: red neuronal convolucional; VPN: valor predictivo negativo; VPP: valor predictivo positivo.

ExtScoreRXT (6%), el recuento de linfocitos (6%) y el nivel de dímero D (3%) fueron, en este orden, los factores pronósticos más ponderados. No hubo diferencias estadísticamente significativas en términos del AUC-ROC entre los modelos con y sin índices basados en la RNC (valor de $p=0,315$), lo que indica que la adición de los parámetros de inteligencia artificial no ofrece una mejoría adicional significativa en el rendimiento del modelo. En la [figura 6](#) se pueden ver las curvas ROC y CPS de la validación interna realizada con un conjunto de datos nuevo para una selección de tres modelos de clasificación de la mortalidad elaborados con tres combinaciones de características distintas.

Discusión

En este estudio, la presencia y el grado de afectación pulmonar en la RXT inicial de los pacientes con COVID-19 tienen un valor pronóstico. En el análisis univariable, la ExtScoreRXT mostró una correlación moderada con el nivel de gravedad y la mortalidad y, en los primeros modelos multivariable elaborados basados en la edad, el sexo y las características radiológicas, la ExtScoreRXT fue el factor pronóstico más importante de la gravedad y el segundo de mortalidad hospitalaria, después de la edad. Sin embargo, la adición de otros parámetros registrados habitualmente en el momento del

ingreso mejoró significativamente la exactitud predictiva de los modelos. Estos resultados muestran la mayor utilidad de la puntuación de la RXT como herramienta pronóstica en la COVID-19, cuando se tiene en cuenta en un modelo junto con la $\text{SatO}_2/\text{FiO}_2$, la edad, la LDH, el recuento de linfocitos, la PCR, el nivel de dímero D y el recuento de plaquetas. La distribución y la densidad de las opacidades no fueron lo suficientemente importantes como para seguir siendo factores pronósticos en los modelos definitivos, y los índices de la herramienta diagnóstica basada en la RNC no mejoraron los parámetros predictivos de forma significativa, probablemente porque no se cuantificó la extensión. Por otro lado, a pesar del análisis minucioso de las imágenes, no es raro observar superposición de las estructuras, opacidades escasas o indeterminadas o RXT normal (31,8% en nuestra serie). Por lo tanto, con la integración de la ExtScoreRXT en un modelo pronóstico se trata de conseguir un proceso decisorio seguro, evitando la alta domiciliaria de pacientes que precisan atención hospitalaria, así como ingresos innecesarios o un uso excesivo de la TC.

También confirmamos la fuerte correlación negativa entre el grado de afectación pulmonar y la $\text{SatO}_2/\text{FiO}_2$. Esto avala la indicación de angiografía pulmonar por TC ampliamente aceptada en caso de desaturación de oxígeno o disnea y afectación pulmonar normal o leve en la RXT¹⁸, en busca

de opacidades pulmonares no visibles en la RXT o de trombo-sis/embolia pulmonar¹⁹. De acuerdo con esto, la $\text{SatO}_2/\text{FiO}_2$ demostró ser un factor pronóstico fuerte en ambos modelos pronósticos elaborados.

En la bibliografía, las puntuaciones de clasificación de opacidades observadas en la RXT de pacientes con COVID-19 en el SU eran predictivas del riesgo de ingreso hospitalario y de intubación en pacientes de 21 a 50 años⁶, y del tiempo transcurrido hasta la intubación, independientemente de las enfermedades concomitantes⁸. El grado de afectación pulmonar también se asoció a peores desenlaces del síndrome respiratorio agudo grave^{20,21}.

Los factores pronósticos de gravedad descritos con más frecuencia en pacientes con COVID-19 fueron la edad, el sexo, las manifestaciones observadas en la TC, la PCR, la LDH y el recuento de linfocitos^{11,22}; y los factores pronósticos más publicados de mortalidad son la edad avanzada^{23–25} y el nivel de dímero D^{22,23}. Estos factores pronósticos coinciden con la mayoría de los que hemos observado en el análisis multivariable y hemos incluido en los modelos predictivos.

Los días con síntomas, el cuadro clínico inicial, el internamiento, las enfermedades concomitantes y el resto de las manifestaciones observadas en la RXT no presentaron una potencia predictiva suficiente (importancia variable < 1%) para su inclusión en el modelo. El número de días con síntomas en el momento de llegada al SU no estuvo relacionado con el grado de afectación pulmonar. En otra serie no se identificó una diferencia significativa entre los pacientes graves y no graves en cuanto a la mediana de días desde el inicio de los síntomas hasta el ingreso hospitalario²⁶. El tabaquismo, las enfermedades concomitantes como la obesidad, la hipertensión, la diabetes, enfermedades cardiovasculares o respiratorias y antecedentes de cáncer, así como la presencia de fiebre, disnea, hemoptisis y pérdida del conocimiento, también se asociaron a un peor pronóstico en algunas publicaciones^{25,27–29}, pero no en nuestro estudio. Probablemente haya un sesgo en la obtención de datos a partir de las historias clínicas, especialmente en el caso de la obesidad, que puede no haberse registrado en la totalidad de los casos. Sin embargo, parece ser que las enfermedades concomitantes y los síntomas tienen un menor peso pronóstico relativo con respecto a las variables definitivas de los modelos. El mayor impacto de la $\text{SatO}_2/\text{FiO}_2$, la ExtScoreRXT y los parámetros analíticos aumenta la necesidad de realizar estas pruebas a todos los pacientes con COVID-19 y síntomas respiratorios o síntomas sistémicos persistentes. Este abordaje pronóstico podría identificar inicialmente a los pacientes que podrían beneficiarse de un tratamiento más específico o de la hospitalización.

La escala nacional de alerta rápida 2 (*National Early Warning Score 2*, NEWS2), basada en las constantes vitales, es la puntuación más utilizada en el SU. Su exactitud predictiva en pacientes con COVID-19 es superior a la de otras puntuaciones de riesgo clínico^{30,31}. Aun así, los modelos elaborados en este estudio superaron esta exactitud con un AUC-ROC = 0,94 para la gravedad y un AUC-ROC = 0,97 para la mortalidad, tal como se esperaba como consecuencia de la adición de otras variables de interés. Se elaboró una puntuación de riesgo específica de la COVID-19 para predecir la enfermedad clínica en el momento del ingreso, con diez variables, que incluían la anomalía de la RXT como factor pronóstico, pero sin evaluación de la extensión; mostró un AUC-ROC de

0,88²⁹. Se construyó otro modelo multivariable que incluía la RXT en el momento del ingreso para predecir enfermedad grave en pacientes con COVID-19 hospitalizados⁷. Los factores pronósticos que se mantuvieron en el modelo fueron el sexo masculino, la enfermedad pulmonar obstructiva, una duración de los síntomas superior a 7 días, el recuento de neutrófilos, la PCR, la LDH, la distribución de la enfermedad pulmonar y la puntuación de la RXT, con un AUC-ROC = 0,77. La PCR, la LDH y la extensión de la afectación pulmonar también se incluyen en nuestro modelo final, pero no hay más coincidencias en el resto de factores pronósticos. Probablemente, esto se explique por las distintas metodologías de elaboración, incluida una estrategia distinta de selección de características, ya que ellos utilizaron una prueba estadística univariable y nosotros basamos nuestra selección en la correlación entre los parámetros y con respecto a la importancia variable. Otras discrepancias se refieren a las etapas previas al tratamiento de los datos, ya que nosotros incluimos una combinación de varias técnicas de sobremuestreo y submuestreo, así como la normalización de los datos; y la consideración de estructuras de modelo distintas, ya que ellos utilizaron una regresión logística multivariable que depende de transformaciones de características no lineales. Para superar este problema, analizamos tres estructuras de modelo distintas: máquina de vectores de soporte, bosque aleatorio y potenciación del gradiente, que pueden manejar características no lineales y sus interacciones, y tener un buen rendimiento en un gran espacio de características.

Como posibles fuentes de sesgo, el nivel de gravedad es un criterio clínico de valoración basado en decisiones, a diferencia de la mortalidad. Para reducir este sesgo, agrupamos en el nivel 0 el alta domiciliaria del SU y también una hospitalización durante 3 días o menos. Además, el seguimiento de un mínimo de 2 meses incluyó a pacientes que volvieron al hospital. En estos casos, todas las variables recopiladas también eran las obtenidas en la primera visita al SU, pero el episodio considerado como desenlace fue el más grave. La proporción de los pacientes más graves (22%) en nuestra población estuvo dentro del intervalo publicado en series más largas (15–36%)^{32,33}.

Con respecto al método propuesto para cuantificar el grado de afectación pulmonar (ExtScoreRXT), no hemos analizado el acuerdo entre observadores. Por otro lado, un buen acuerdo entre observadores se demostró con el uso de Brixia, una puntuación más compleja diseñada para pacientes con COVID-19⁹ y, casualmente, hemos utilizado la misma puntuación que otros autores, que han publicado recientemente una buena correlación con la puntuación Brixia¹⁵.

La validación interna se realizó con 88 casos. Sin embargo, se ha publicado que se recomienda un tamaño mínimo de muestra de 100 para alcanzar una validación sólida³⁴. Sería deseable llevar a cabo una validación externa con casos de otros hospitales para evaluar la posibilidad de generalización y de uso de los modelos elaborados en la práctica clínica diaria.

En conclusión, los modelos predictivos multivariables del pronóstico mostraron una exactitud muy predictiva que podría permitir el triaje de pacientes con COVID-19 sintomáticos en el SU para mejorar la toma de decisiones. La aplicación para calcular el nivel de gravedad y la mortalidad hospitalaria está disponible

en <http://upv.datahub.egi.eu:30054/hulafecovid19models>. Debería validarse en distintos SU tanto para la COVID-19 como para otras infecciones víricas con un comportamiento similar.

Autoría

1. Responsable de la integridad del estudio: LMB.
2. Concepción del estudio: PCB, LMB, LCA.
3. Diseño del estudio: PCB, LMB, LCA.
4. Obtención de datos: PCB, CFE, ACO, CFMN, LTO.
5. Análisis e interpretación de los datos: LCA.
6. Tratamiento estadístico: LCA.
7. Búsqueda bibliográfica: PCB, LCA.
8. Redacción del trabajo: PCB, LCA.
9. Revisión crítica del manuscrito con aportaciones intelectualmente relevantes: PCB, LCA, LMB.
10. Aprobación de la versión final: PCB, LCA, CFE, ACO, CFMN, LTO, LMB.

Conflictos de intereses

Los autores declaran no tener ningún conflicto de intereses.

Agradecimientos

Nuestro agradecimiento a todo el personal que trabaja en el Hospital Universitario y Politécnico La Fe por su dedicación durante la pandemia de COVID-19 y, especialmente, a Carlos Baeza, José Sánchez, Ignacio Gómez, Ana Gil, Rodrigo Blanco, Pau Lozano, Alberto García, Marta Montero, Pilar Estellés, Javier Millán y Cristina Aguado.

Anexo. Material adicional

Se puede consultar material adicional a este artículo en su versión electrónica disponible en [doi:10.1016/j.rx.2021.09.011](https://doi.org/10.1016/j.rx.2021.09.011).

Bibliografía

1. Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X, et al. CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). *Radiology*. 2020;295:202–7, <http://dx.doi.org/10.1148/radiol.20200230>.
2. Bernheim A, Mei X, Huang M, Yang Y, Fayad ZA, Zhang N, et al. Chest CT Findings in Coronavirus Disease-19 (COVID-19): Relationship to Duration of Infection. *Radiology*. 2020;295, <http://dx.doi.org/10.1148/radiol.20200463>, 200463.
3. Pan F, Ye T, Sun P, Gui S, Liang B, Li L, et al. C. Time Course of Lung Changes at Chest CT during Recovery from Coronavirus Disease 2019 (COVID-19). *Radiology*. 2020;295:715–21, <http://dx.doi.org/10.1148/radiol.20200370>.
4. Bai HX, Hsieh B, Xiong Z, Halsey K, Choi JW, Tran TML, et al. Performance of Radiologists in Differentiating COVID-19 from Non-COVID-19 Viral Pneumonia at Chest CT. *Radiology*. 2020;296:E46–54, <http://dx.doi.org/10.1148/radiol.20200823>.
5. Salehi S, Abedi A, Balakrishnan S, Gholamrezanezhad A. Coronavirus Disease 2019 (COVID-19): A Systematic Review of Imaging Findings in 919 Patients. *AJR Am J Roentgenol*. 2020;215:87–93, <http://dx.doi.org/10.2214/AJR.20.23034>.
6. Toussie D, Voutsinas N, Finkelstein M, Cedillo MA, Manna S, Maron SZ, et al. Clinical and Chest Radiography Features Determine Patient Outcomes in Young and Middle-aged Adults with COVID-19. *Radiology*. 2020;297:E197–206, <http://dx.doi.org/10.1148/radiol.202001754>.
7. Schalekamp S, Huisman M, van Dijk RA, Boomsma MF, Freire Jorge PJ, de Boer WS, et al. Model-based Prediction of Critical Illness in Hospitalized Patients with COVID-19. *Radiology*. 2021;298:E46–54, <http://dx.doi.org/10.1148/radiol.202002723>.
8. Xiao N, Cooper JG, Godbe JM, Bechel MA, Scott MB, Nguyen E, et al. Chest radiograph at admission predicts early intubation among inpatient COVID-19 patients. *Eur Radiol*. 2021;31:2825–32, <http://dx.doi.org/10.1007/s00330-020-07354-y>.
9. Borghesi A, Maroldi R. COVID-19 outbreak in Italy: experimental chest X-ray scoring system for quantifying and monitoring disease progression. *Radiol Med*. 2020;125:509–13, <http://dx.doi.org/10.1007/s11547-020-01200-3>.
10. Yuan M, Yin W, Tao Z, Tan W, Hu Y. Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China. *PLoS One*. 2020;15, e0230548.
11. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. M. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ*. 2020;369:m1328. doi: 10.1136/bmj.m1328. Update in: *BMJ*.;1; 2021;372:n236. Erratum in: *BMJ*.;1; 2020;369:m2204.
12. Colombi D, Bodini FC, Petrini M, Maffi G, Morelli N, Milanese G, et al. Well-aerated Lung on Admitting Chest CT to Predict Adverse Outcome in COVID-19 Pneumonia. *Radiology*. 2020;296:E86–96, <http://dx.doi.org/10.1148/radiol.202001433>.
13. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med*. 2015;162:55–63.
14. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins, et al. PROBAST Group†. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med*. 2019;170:51–8, <http://dx.doi.org/10.7326/M18-1376>.
15. Setiawati R, Widyaningroem A, Handarini T, Hayati F, Basja AT, Putri ARDS, et al. Modified Chest X-Ray Scoring System in Evaluating Severity of COVID-19 Patient in Dr. Soetomo General Hospital Surabaya, Indonesia. *Int J Gen Med*. 2021;14:2407–12, <http://dx.doi.org/10.2147/IJGM.S310577>.
16. Leevy JL, Khoshgoftaar TM, Bauder RA, Seliya N. A survey on addressing high-class imbalance in big data. *J Big Data*. 2018;5:42.
17. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837–45.
18. Revel MP, Parkar AP, Prosch H, Silva M, Sverzellati N, Gleeson F, et al. European Society of Radiology (ESR) and the European Society of Thoracic Imaging (ESTI). COVID-19 patients and the radiology department - advice from the European Society of Radiology (ESR) and the European Society of Thoracic Imaging (ESTI). *Eur Radiol*. 2020;30:4903–9, <http://dx.doi.org/10.1007/s00330-020-06865-y>.
19. Rosovsky RP, Grodzin C, Channick R, Davis GA, Giri JS, Horowitz J, et al. PERT Consortium. Diagnosis and Treatment of Pulmonary Embolism During the Coronavirus Disease 2019 Pandemic: A Position Paper From the National PERT Consortium. *Chest*. 2020;158:2590–601, <http://dx.doi.org/10.1016/j.chest.2020.08.2064>.
20. Chau TN, Lee PO, Choi KW, Lee CM, Ma KF, Tsang TY, et al. Value of initial chest radiographs for predicting

- clinical outcomes in patients with severe acute respiratory syndrome. *Am J Med.* 2004;117:249–54, <http://dx.doi.org/10.1016/j.amjmed.2004.03.020>.
21. Wan YL, Tsay PK, Cheung YC, Chiang PC, Wang CH, Tsai YH, et al. A correlation between the severity of lung lesions on radiographs and clinical findings in patients with severe acute respiratory syndrome. *Korean J Radiol.* 2007;8:466–74, <http://dx.doi.org/10.3348/kjr.2007.8.6.466>.
 22. Velavan TP, Meyer CG. Mild versus severe COVID-19: Laboratory markers. *Int J Infect Dis.* 2020;95:304–7, <http://dx.doi.org/10.1016/j.ijid.2020.04.061>.
 23. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet.* 2020;395:1054–62. doi: 10.1016/S0140-6736(20)30566-3. Erratum in: *Lancet.* 2020;395:1038.
 24. Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, et al. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Intern Med.* 2020;180:934–43. doi: 10.1001/jamainternmed.2020.0994. Erratum in: *JAMA Intern Med.* 2020;180:1031.
 25. Zheng Z, Peng F, Xu B, Zhao J, Liu H, Peng J, et al. Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis. *J Infect.* 2020;81:e16–25, <http://dx.doi.org/10.1016/j.jinf.2020.04.021>.
 26. Zhang JJ, Dong X, Cao YY, Yuan YD, Yang YB, Yan YQ, et al. Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. *Allergy.* 2020 Jul;75:1730–41, <http://dx.doi.org/10.1111/all.14238>.
 27. Ruan Q, Yang K, Wang W, Jiang L, Song J. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Med.* 2020;46:846–8, <http://dx.doi.org/10.1007/s00134-020-05991-x>, 3. Erratum in: *Intensive Care Med.* 2020 Apr 6.
 28. Huang Y, Lu Y, Huang YM, Wang M, Ling W, Sui Y. Obesity in patients with COVID-19: a systematic review and meta-analysis. *Metabolism.* 2020;113, <http://dx.doi.org/10.1016/j.metabol.2020.154378>, 154378.
 29. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, et al. China Medical Treatment Expert Group for COVID-19. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. *JAMA Intern Med.* 2020;180:1081–9, <http://dx.doi.org/10.1001/jamainternmed.2020.2033>.
 30. Jang JG, Hur J, Hong KS, Lee W, Ahn JH. Prognostic Accuracy of the SIRS, qSOFA, and NEWS for Early Detection of Clinical Deterioration in SARS-CoV-2 Infected Patients. *J Korean Med Sci.* 2020;35:e234, <http://dx.doi.org/10.3346/jkms.2020.35.e234>. PMID: 32597046.
 31. Myrstad M, Ihle-Hansen H, Tveita AA, Andersen EL, Nygård S, Tveit A, et al. National Early Warning Score 2 (NEWS2) on admission predicts severe disease and in-hospital mortality from Covid-19 - a prospective cohort study. *Scand J Trauma Resusc Emerg Med.* 2020;28:66, <http://dx.doi.org/10.1186/s13049-020-00764-3>.
 32. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA.* 2020;323:1239–42, <http://dx.doi.org/10.1001/jama.2020.2648>.
 33. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA.* 2020; 323:2052–9. doi: 10.1001/jama.2020.6775. Erratum in: *JAMA.* 2020; 323:2098.
 34. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med.* 2016;35:214–26, <http://dx.doi.org/10.1002/sim.6787>.