

Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín



Surveillance of the application of artificial intelligence models for the detection of pneumonia in the radiology services of third and fourth level hospitals in Medellín

Proyecto: Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín

Elaborado por: Emmanuel Salinas Miranda. M.D. Radiólogo

Modificado por: Jose Julián Garcés M.D. MSc.

Revisado por: Juan Guillermo Paniagua. Dr. MSc. Ing

Aprobado por: O. L. Quintero M. Prof. Dr. Ing

Reporte Final para ser presentado ante el Ministerio de Ciencia y Tecnología de la Republica de Colombia

Introducción

La Universidad EAFIT y las instituciones universitarias ITM y Universidad de Antioquia han decidido unir sus capacidades y las de los grupos de investigación para responder al llamado del Ministerio de Ciencia y Tecnología en la “Mincienciación” debido a la pandemia de COVID-19 en 2020. Este es un estudio multicéntrico y multidisciplinario que busca generar modelos validados de inteligencia artificial para el diagnóstico y cuantificación de Coronavirus en la radiografía de tórax y tomografías. Nuestra iniciativa incluye instituciones de salud de nivel 3 y 4 de Medellín como la IPS universitaria, el hospital San Vicente de Paul, el hospital Pablo Tobón Uribe y la clínica CES. Estas instituciones colaborarán y trabajarán sobre el tiempo para desarrollar una solución robusta y clínicamente validada. Tradicionalmente el diagnóstico de nuevo coronavirus (SARS-CoV-2 / COVID-19) se obtiene a través de la reacción en cadena de la polimerasa de transcripción inversa (RT-PCR). Sin embargo, su sensibilidad es limitada para la detección temprana de la enfermedad llegando a ser tan baja como del 31 al 71% (1) (Se asume una especificidad mayor al 99%) (2). Múltiples factores como el procedimiento para la recolección de la muestra, la severidad de la infección de los pacientes, el transporte y el desempeño de los instrumentos diagnósticos pueden afectar el rendimiento de la RT-PCR.

En pacientes con sospecha de COVID-19 con riesgo de progresión, pacientes que se presenten con síntomas moderados-severos, pacientes con empeoramiento de su estado respiratorio y pacientes con altas probabilidades pretest de sufrir la enfermedad (3) las imágenes diagnósticas se encuentran indicadas. Particularmente, la tomografía axial computarizada (TAC) de tórax es un estudio de moderada-alta sensibilidad para la detección de COVID -19. La TAC se encuentra indicada para detectar enfermedad leve del parénquima pulmonar, la progresión de la enfermedad y descartar otros diagnósticos alternativos (4). También se ha recomendado como un marcador de imagen en las principales zonas epidémicas (5). A pesar de su capacidad diagnóstica (sensibilidad y especificidad de aproximadamente del 97% y el 68% respectivamente (6)), la decisión final de utilizar CT va a depender de los protocolos institucionales individuales, los recursos locales y la experticia. Estos dos últimos factores son particularmente relevantes en países latinoamericanos como Colombia.

Aunque la radiografía de tórax es menos sensible que la tomografía para la demostración de hallazgos relacionados con COVID-19, este es el estudio de primera línea en la mayoría de los pacientes con sospecha de COVID-19 (7). Dada su portabilidad y la facilidad para la descontaminación el colegio americano de radiología la recomienda como primera modalidad de imagen a considerar (8). Se estima, que de los pacientes que requieren hospitalización, 69% de los pacientes tendrán una radiografía anormal de tórax en el momento de

Reporte Final para ser presentado ante el Ministerio de Ciencia y Tecnología de la Republica de Colombia

Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín la admisión y hasta el 80% de los pacientes demostrará alguna alteración durante su hospitalización (7). A medida que transcurre el tiempo de la infección los hallazgos progresan, observándose de manera extensa luego de 10-12 una vez iniciados los síntomas (7). A pesar de que la mayoría de los pacientes demostrarán alguna alteración, la radiografía de tórax puede ser normal en las etapas tempranas de la enfermedad o en la enfermedad leve. La sensibilidad de la radiografía de tórax puede llegar a ser del 59-69% (9) comparada con la sensibilidad variable de la RT-PCR del 31-91% (7).

El impacto de la tomografía en el manejo del COVID-19 va a depender de su disponibilidad y de su adecuada interpretación. La tomografía de tórax es el tercer estudio tomográfico que se realiza con mayor frecuencia en el país (10), sin embargo, el número de tomografías realizadas no es proporcional al número de radiólogos disponibles para su lectura y esta proporción es menor en las regiones más periféricas del país como Meta, San Andrés o Caquetá (11). Estas cifras sugieren la dependencia de estas regiones de los servicios de tele- Radiología y en consecuencia una menor cantidad de servicios de lectura dedicados a las urgencias. Para profundizar la dificultad en la implementación de la tomografía en el diagnóstico COVID-19 se agrega la ausencia de personal con una formación suficiente y en un número óptimo en estas regiones. Refiriéndonos a la disponibilidad de servicios de salud, regiones como Guaviare, Amazonas y Caquetá no suman más de 50 camas de UCI (12). Para empeorar la crisis, medidas de salud pública como el distanciamiento social no están siendo acogidas por gran parte de la población. De esta forma la ausencia de adopción de las medidas de contención, una cantidad insuficiente de capital humano y la penetración insuficiente de la RT-PCR en regiones más alejadas del país, contribuyen en conjunto a empeorar los efectos de la pandemia en estas regiones más periféricas e históricamente con menores recursos.

En la declaración del consenso multinacional de Sociedad Fleischner (3) sobre el rol de las imágenes en los pacientes con COVID-19 se recomienda la radiografía de tórax en las regiones con escasos recursos donde el acceso a estudios tomográficos es restringido (3). En pacientes hospitalizados la radiografía de tórax podría ser un recurso adicional para evaluar progresión de la enfermedad y diagnósticos alternativos como neumonía lobar, sobreinfección bacteriana, neumotórax y derrame pleural (3). El valor de la radiografía de tórax en la sospecha de infección por COVID-19 va a estar dado por el riesgo pretest de la enfermedad, la severidad de la dificultad respiratoria y la disponibilidad de los recursos. En este sentido la radiografía de tórax permite: revelar diagnósticos alternativos en pacientes con signos clínicos moderados o severos, sugerir falsos negativos de la RT-PCR en pacientes con alta probabilidad pretest y hallazgos compatibles en imagen y servir de soporte para el triage rápido de pacientes con síntomas moderados o severos cuando la RT-

Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín

PCR se encuentra negativa o no es disponible en un contexto de escasos de recursos (3). Los sistemas de apoyo y triage basados en inteligencia artificial (IA) han demostrado su incorporación exitosa en la práctica clínica para la detección de enfermedades agudas del tórax (e.g. neumotórax) (13) y podrían tener un rol significativo en la detección de la enfermedad bajo los escenarios descritos en COVID-19. Estudios previos con métodos de aprendizaje profundo han reportado algunos logros de aprendizaje en la enfermedad y proclaman su utilidad para detectar COVID-19 y distinguirlo de neumonía adquirida en la comunidad y otras enfermedades del pulmón (14). Pese a que la literatura sugiere que este enfoque logró un desempeño óptimo (Área bajo la curva [AUC] 0.96, sensibilidad 87% y especificidad del 92%, respectivamente) comparable a la RT-PCR (14), todavía no se llega a la obtención de un tipo de modelo que sea usado como dispositivo médico y que sea además respaldado por un estudio clínico.

En esas aproximaciones, resulta particular que los autores usan redes de aprendizaje profundo sobre conjuntos de datos obtenidos previamente y en algunos casos con pacientes pediátricos, pero sobre todo sin cumplir las cotas de aprendizaje de los modelos matemáticos de tal complejidad. Aunque otras arquitecturas diferentes a ResNet50 y una segmentación pulmonar óptima (Usando U-Net de base) (15) podrían mejorar este desempeño, estos resultados dan cuenta del potencial de la inteligencia artificial para generar una prueba de detección rápida y precisa para casos de COVID-19.

Adicionalmente, estudios previos, sugieren un adecuado desempeño en el uso de la radiografía de tórax para la discriminación adecuada de neumonía adquirida en la comunidad y otras enfermedades del pulmón (16). Este enfoque logró una sensibilidad del 80% y una especificidad del 99% comparable con la RT-PCR (14). Sin embargo, la baja prevalencia de pacientes con COVID-19 (8 pacientes) limita aplicabilidad en el ámbito clínico. Otras desventajas de la RT-PCR como la pobre disponibilidad de sus pruebas rápidas, su alta tasa de falsos negativos y las demoras en el despliegue de las pruebas rápidas justifican la evaluación de estos modelos como métodos alternativos para el diagnóstico de la enfermedad.

En Colombia, la ley de la radiología califica solo al médico radiólogo para la lectura de estudios de tomografía (17). Sin embargo, dada la ausencia del personal suficiente, la baja sensibilidad de la RT-PCR, el insuficiente despliegue de la prueba a lo largo del país y la presencia de estudios preliminares que sugieren una sensibilidad de la radiografía de tórax y de la tomografía comparables con la RT-PCR en ciertos escenarios clínicos y abre la posibilidad de que un sistema dedicado de IA en para la evaluación de la radiografía de tórax contribuya en el manejo clínico de estos casos.

Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín

El objetivo de este estudio multiinstitucional es desarrollar y validar de forma prospectiva un modelo de aprendizaje profundo para la detección de las manifestaciones radiográficas y tomográficas de COVID-19. La inclusión de estudios normales y con otras enfermedades infecciosas y no infecciosas no relacionadas con COVID-19 demostrarán robustez de este modelo.

Este documento es un compendio de los métodos matemáticos, médicos y detalle de los estudios clínicos que acompañan la realización de principalmente tres resultados:

- vlvA-Med Rural una aplicación móvil para el apoyo a tamizaje de radiografía de tórax pensado para el uso de médicos sin acceso a conectividad en el territorio nacional que tengan acceso a una placa de tomografía.
- vlvA-Med Conrad una aplicación de telemedicina que posee un sistema de inteligencia artificial que puede clasificar imágenes de rayos x que estén en formatos DICOM del sistema RIS de instituciones donde no haya acceso a un médico radiólogo especializado.
- vlvA-Med Godfrey una aplicación de telemedicina que clasifica TAC brindando una clasificación previa de los tipos de patología, y que tiene como meta la cuantificación de compromiso pulmonar de los pacientes.

Asociados a esta investigación se tendrán seis (6) artículos científicos que validen cada uno de los métodos y materiales del estudio clínico así como los modelos matemáticos y despliegues de ingeniería, la formación de especialistas en radiología y magíster en ingeniería, y el avance y posicionamiento de vlvA-Med como alianza a nivel nacional.

Todo lo anterior en el marco de los más altos estándares de calidad científica y ética buscando proponer un marco regulatorio para el uso de sistemas de inteligencia artificial en salud en el país.

Resumen general del protocolo

El equipo de investigación de las Universidades está liderado por investigadores senior reconocidos por sus contribuciones en el campo de la inteligencia artificial aplicada a la biomedicina con más de 20 artículos científicos y 5 patentes en total verificables en sus CVLAC. El equipo de investigación médico de las instituciones es reconocido internacionalmente por sus contribuciones y sus innovaciones en la práctica médica y científica.

Con relación a las tecnologías, Quintero & Paniagua, 2020² describen la experiencia previa de las tecnologías de aprendizaje de máquina e

Reporte Final para ser presentado ante el Ministerio de Ciencia y Tecnología de la Republica de Colombia

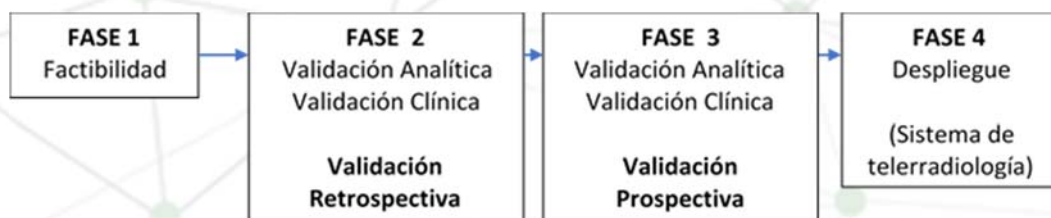
Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín

inteligencia artificial en apoyo médico a la toma de decisiones, haciendo una fuerte revisión de los logros en diferentes áreas de trabajo médico. En la última década, los sistemas de inteligencia artificial aplicados a la radiología han demostrado su potencial clínico. Varios modelos han sido validados para la detección de neumonías virales y bacterianas en imágenes convencionales y de tomografía. En condiciones ideales de lectura, el desempeño de estos algoritmos llega a ser comparable con el desempeño del radiólogo.

La adecuada validación de los desarrollos del grupo en el contexto colombiano y su aplicación exitosa dentro del flujo de trabajo clínico resultaría en herramientas adicionales susceptibles de ser distribuidas en todo el país para ayudar a combatir la enfermedad. La implementación, funcionamiento y validación de algoritmos de IA en la radiología requiere del trabajo mancomunado de los desarrolladores de los algoritmos y radiólogos, que, a través de la experticia y la preparación de conjuntos de datos y casos, éstos puedan ser evaluados en cuanto su rendimiento de tal manera mejorados y ajustados a los datos reales.

El siguiente protocolo de investigación se desarrollará en 4 fases y se realizará en el transcurso de 9 meses (Ver: Cronograma).

Estas etapas se superponen entre sí. La primera fase tendrá una duración de 3 meses, la segunda fase: 3 meses, la tercera fase: 4 meses y la última fase 1 mes. En este tiempo se espera lograr los objetivos primarios y secundarios.



Pregunta de investigación

¿Modelos de aprendizaje de máquina aplicados a las radiografías y tomografías de tórax en pacientes con diagnóstico de SARS-CoV-2 tienen un desempeño comparable a la RT-PCR para el diagnóstico y cuantificación de la enfermedad?

OBJETIVOS

Objetivo general:

Reporte Final para ser presentado ante el Ministerio de Ciencia y Tecnología de la Republica de Colombia

Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín
Validar retrospectiva y prospectivamente el desempeño de dos modelos de inteligencia artificial aplicados a la radiografía y tomografía de tórax para la detección de signos típicos de COVID-19.

Objetivos específicos:

1. Validación de modelos preexistentes de aprendizaje de máquina para detección de signos de COVID-19 en radiografía de tórax y tomografía. Verificar la viabilidad y factibilidad de las capacidades de las instituciones para contribuir a la validación de los modelos.
2. Identificar el desempeño de los modelos de aprendizaje de máquina en un estudio retrospectivo.
3. Diseñar un sistema de tele-radiología para apoyo a la toma de decisiones en el diagnóstico de COVID-19.
4. Validación de los modelos diseñados y entrenados con imágenes médicas validadas por profesionales médicos con la población colombiana bajo los protocolos establecidos en los departamentos de radiología de las instituciones de salud.

Particulares de este documento:

- Validar el desempeño de dos modelos de inteligencia artificial aplicados a la radiografía y tomografía de tórax para la cuantificación de compromiso por COVID-19.
- Desplegar un sistema de tele-radiología a nivel nacional que incorpore los modelos clínicamente validados.

FASE 1: FACTIBILIDAD

Es la primera etapa de generación del proyecto. Esta etapa se centra en conocer los recursos de los que disponen las instituciones de salud y las instituciones académicas para llevar a cabo el proyecto. Tiene los siguientes objetivos:

- Establecer los requisitos, protocolos y formatos de las bases de datos de imágenes y clínicas.
- Establecer las guías de privacidad y anonimización de las imágenes y los datos clínicos.
- Conocer y discutir las guías de las instituciones de salud para el diagnóstico en imagen de COVID-19.
- Lograr la aprobación del estudio clínico en los comités de ética de cada una de las instituciones.
- Generar de bases de datos de imágenes (CR/CT) y datos clínicos de los pacientes con COVID-19.
- Generar modelos pre-entrenados para la detección y cuantificación de COVID-19 en CR y CT.

Reporte Final para ser presentado ante el Ministerio de Ciencia y Tecnología de la Republica de Colombia

Durante esta fase se adelantarán experimentos con los modelos de aprendizaje de máquina de RX y CT que se encuentren disponibles en internet y que dispongan de bases de datos de acceso libre. Para esta finalidad se incluirán bases de datos dentro de las que se encuentran CHEXNET y MIT/MIMIC.

ESTUDIO CLINICO

Descripción del estudio clínico

Métodos:

Pacientes:

Este estudio se realizará en 4 instituciones de alta complejidad de la ciudad de Medellín, los cuales sumaron para el 2019 una admisión a servicios de urgencias de más de 248.811 pacientes (28.129 San Vicente Fundación, 100.000 HPTU, 78.244 IPS universitaria, 42.438 Clínica CES), cuentan con más de 1.311 camas hospitalarias (237 San Vicente Fundación, 371 HPTU, 531 IPS universitaria, 172 Clínica CES), y 184 camas de cuidados intensivos (59 San Vicente Fundación, 71 HPTU, 44 IPS universitaria, 10 Clínica CES). Se estima que el pico de la pandemia se presente entre finales de junio y comienzos de septiembre, alcanzando un número de infectados que varía entre 10.000 – 15.000 casos (20).

Los datos clínicos se obtendrán de cada uno de los sistemas de registro médico electrónico. Estas instituciones se seleccionaron dado que son instituciones de referencia para el manejo de COVID en la ciudad. Nuestro estudio se adhiere a las guías TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) (21) para estudios diagnósticos y pronósticos.

Proceso de recolección de la información

Se realizará un muestreo por conveniencia en el que se incluirán las imágenes según su disponibilidad en el sistema PACS y que cumplan los criterios de inclusión y exclusión hasta alcanzar el tamaño de muestra. Se recolectará los estudios y la información clínica de 1600 pacientes los cuales serán admitidos a las instituciones participantes entre el 28 de abril de 2020 y 31 de diciembre de 2020 en la ciudad de Medellín. Para la recolección de la información se utilizarán 3 instrumentos:

Reporte Final para ser presentado ante el Ministerio de Ciencia y Tecnología de la Republica de Colombia

Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín

1. Hoja de cálculo con las cohortes de pacientes, 2. Hoja de cálculo con los datos clínicos de pacientes COVID positivo, 3. Hoja de cálculo para la clasificación por imágenes de los pacientes COVID positivo/negativo y el porcentaje de compromiso pulmonar de los mismos.

Los datos clínicos y exposiciones se recopilarán en hojas de cálculo para el análisis multivariado, protegidos por contraseña y en los computadores institucionales. La hoja de cálculo con contraseña será accesible solo por los investigadores clínicos. Las imágenes se almacenarán en los computadores institucionales de la misma forma. Una vez se complete el tamaño de muestra, las imágenes se anonimizarán y se transferirán a los equipos de entrenamiento

El software DICOM anonymizer Pro v. 2.0.9 (Neológica, Cairo Montenotte, Italy) se utilizó para anonimizar todos los casos. Un miembro del equipo (médico Investigador) fue el albacea de los datos y se encargó de la anonimización de la cohorte. Se utilizó un disco duro cifrado con contraseña (5 TB) para transferir las imágenes. Los computadores que se utilizaron para entrenar estuvieron protegidos por contraseña.

Criterios de inclusión - PACIENTES

Cohorte de pacientes con diagnóstico de COVID-19

Criterios de inclusión

- Pacientes mayores de 14 años, que consulten a las instituciones asociadas al estudio clínico, independiente del lugar de residencia.
- Con hallazgos imagenológicos (CT/CR) de compromiso pulmonar por COVID-19.
- Diagnostico confirmado por COVID-19 dado por: RT-PCR, test de anticuerpos IgM positivos.

Criterios de exclusión

- Pacientes con diagnóstico de COVID-19 sin hallazgos en imágenes.

Cohorte de pacientes con diagnóstico de "Otras" enfermedades

Criterios de inclusión

Reporte Final para ser presentado ante el Ministerio de Ciencia y Tecnología de la Republica de Colombia

Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín

- Pacientes con diagnóstico negativo confirmado y/o baja sospecha de COVID.
- Pacientes con estudios anteriores a la declaración del inicio de la pandemia por el ministerio de la protección social en el país (enero 2020).
- Pacientes con diagnóstico clínico, radiológico o histológico confirmado (ej.: baciloscopia, biopsia, lavado bronco alveolar) de otra enfermedad diferente a COVID-19 que genere hallazgos en el parénquima pulmonar.

Criterios de exclusión

- Pacientes con enfermedades diferentes a COVID-19 con alteraciones severas mediastinales o por fuera del parénquima pulmonar.

Cohorte de pacientes sin enfermedades: "Normal"

Criterios de inclusión

- Pacientes con diagnóstico negativo confirmado y/o baja sospecha de COVID.
- Pacientes con estudios anteriores a la declaración del inicio de la pandemia por el ministerio de la protección social en Colombia (anteriores a febrero de 2020).
- Ausencia de signos de enfermedad del parénquima pulmonar (Infecciosa, neoplásica, etc.).
- Hallazgos secuenciales leves en el parénquima pulmonar (ej.: Bandas fibroatelectásicas, granulomas, cicatrices del parénquima pulmonar, etc.)

Criterios de exclusión

- Hallazgos secuenciales moderados o severos (ej.: Bulas, cavitaciones, extensos cambios fibrocicatriciales, calcificación del parénquima pulmonar, etc.).
- Presencia de enfermedad extrapulmonar visible en la tomografía (ej.: masas, adenomegalias, timoma, etc.)

Criterios de calidad - MODALIDADES DIAGNÓSTICAS

Los criterios de calidad de las imágenes dependerán del tipo de modalidad de la siguiente forma:

Reporte Final para ser presentado ante el Ministerio de Ciencia y Tecnología de la Republica de Colombia

Radiografía de tórax:

Criterios de inclusión

- Radiografía anteroposterior.
- Ausencia de marcas de identificación. Las marcas de lateralidad son aceptables.
- Tráquea centrada, campos pulmonares simétricos, ausencia de las escápulas.
- Adecuada penetración.
- Ausencia de artefactos

Criterios de exclusión

- Radiografías rotadas.
- Presencia de elementos en campos pulmonares (ej.: tubos, catéteres, drenajes).

Tomografía de tórax:

Criterios de inclusión

- Disponibilidad de proyecciones axiales. Las proyecciones sagitales y coroneales serán opcionales.
- Disponibilidad de reconstrucciones para parénquima pulmonar o disponibilidad de imágenes "crudas" de base.
- Grosor de corte de las reconstrucciones del parénquima pulmonar 2.5 – 5.0 mm.

Criterios de exclusión

- Artefactos de respiración que impidan la adecuada interpretación del estudio.

Recolección de la información clínica

En la cohorte de pacientes con COVID-19 se obtendrán los siguientes datos clínicos, epidemiológicos y de tratamiento (Tabla 1).

Tabla 1. Información clínica

Información clínica	Información epidemiológica
<ul style="list-style-type: none"> • Sexo • Edad • Síntomas respiratorios agudos: Si/No <ul style="list-style-type: none"> ◦ Menos de 15 días. <ul style="list-style-type: none"> ▪ 2 o mas: <ul style="list-style-type: none"> • Tos • Dificultad respiratoria • Odinofagia • Fatiga/adinamia • Con presencia o no de fiebre mayor o igual a 38 grados. • Resultado PCR: Positiva, negativa, no disponible, indeterminada. • PaO2 por oximetría al ingreso • Condición clínica: Ambulatorio, hospitalizado, cuidados intermedios, cuidados intensivos(22) • Antecedentes personales: Diabetes, HTA, inmunosupresión, asma, cáncer. • Alta: Si/No • Muerte: Si/No • Ventilación mecánica: Si/No • Tratamientos relevantes: Inmunosupresión, antibióticoterapia. 	<p>Antecedente de exposición Si/No(23)</p> <ul style="list-style-type: none"> ◦ Viaje reciente a una ciudad o territorio endémico ◦ Contacto con persona sintomática de COVID-19 durante un periodo de 48 hrs antes del inicio de los síntomas <p>Información de tratamiento</p> <ul style="list-style-type: none"> • Dispositivo de oxigenoterapia • FIO2 • Pronación vigil • Antibióticos • Antivirales • Antimalaricos • Anticuerpos • AINES • Antiinflamatorios esteroideos • Otros antiinflamatorios • Immunomoduladores • Antifibrinolíticos • Tromboprolaxis • Anticoagulación

Recolección de los datos de las modalidades

Radiografía de tórax

Se obtendrán las proyecciones PA y AP disponibles. Dado el enfoque multiinstitucional, múltiples equipos de radiografías serán incluidos: De cada equipo se registrará proyección (AP, PA, portátil) el kilovoltaje (kv), la corriente del tubo (mA) y el fabricante.

Tomografía de tórax

Dado el enfoque multiinstitucional, múltiples equipos de tomografía serán incluidos: De cada equipo se registrará el KVP, la corriente del tubo (mA), pitch, colimación, número de detectores, tiempo de rotación (s/rot), campo de adquisición de imagen (Field of View (FOV)), espesor de corte, separación entre cortes, algoritmo de reconstrucción, fabricante.

Tamaño de muestra

El cálculo para el tamaño de muestra en estudios relacionados con aprendizaje de máquina es problemático, por lo que se realizará un muestreo por conveniencia ingresando todos los pacientes con criterios de inclusión. Sin embargo, en una aproximación inicial, el tamaño de la muestra se estimará para comparar el desempeño diagnóstico de pruebas pareadas con un resultado binario. Para esta tarea se comparará el desempeño diagnóstico del “Gold estándar” para COVID-19: RT-PCR, con el desempeño diagnóstico promedio Reporte Final para ser presentado ante el Ministerio de Ciencia y Tecnología de la Republica de Colombia

Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín reportado para los modelos de clasificación aplicados a radiografía y tomografía de tórax.

La exactitud de estas pruebas diagnósticas se determina por su AUC y se compara al someter los mismos sujetos a ambas pruebas. La hipótesis nula (H_0) será: $AUC_1 = AUC_2$ y la hipótesis alternativa (H_1) será: $AUC_1 \neq AUC_2$.

Deseamos conocer cuántos casos COVID-19 (casos) y cuantos casos NO-COVID-19 (controles) se requieren para detectar una diferencia entre las dos pruebas diagnósticas. Esta diferencia se define como $\delta = AUC_1 - AUC_2$ y junto con ello se asume un intervalo de confianza de $(1 - \delta) \%$ y un poder de $(1 - \beta)$ para la hipótesis alternativa (24).

Construyendo los intervalos de confianza para el parámetro de interés: $AUC_1 - AUC_2$ y utilizando una aproximación normal bajo la hipótesis nula y la hipótesis alternativa, el tamaño de la muestra para cada grupo se obtiene a partir de la siguiente fórmula (Fórmula 1):

$$n = \frac{[Z_{\alpha/2} \sqrt{V_{H_0}(\widehat{AUC}_1 - \widehat{AUC}_2)} + Z_{\beta} \sqrt{V_{H_1}(\widehat{AUC}_1 - \widehat{AUC}_2)}]^2}{[AUC_1 - AUC_2]^2} \quad (1)$$

Donde:

$$V_{H_0}(\widehat{AUC}_1 - \widehat{AUC}_2) = nVar(\widehat{AUC}_1) + nVar(\widehat{AUC}_2) = -2nCov(\widehat{AUC}_1, \widehat{AUC}_2)$$

Para realizar este cálculo se utilizó el paquete easyROC v.1.3(25) en el programa R v3.6 (26).

Tomografía de tórax

El tamaño de muestra mínimo de pacientes con COVID-19 es de 537 y pacientes NO-Covid 19 es de 537. Para este cálculo se asume un desempeño diagnóstico de la RT-PCR de 0.85 (AUC_1) (27,28) y un desempeño diagnóstico promedio de los modelos de inteligencia artificial aplicados a la tomografía de tórax con AUC de 0.80 (AUC_2).

Este cálculo se realiza con un enfoque de no-inferioridad entre pruebas diagnósticas pareadas binarias (29) y asume un error tipo I de 0,05, un poder estadístico de 0,8, una proporción de asignación de 1 para la detección de una diferencia mínima entre el desempeño de ambas pruebas de 0,1. Los pacientes No-COVID pueden ser pacientes normales (537) o pacientes con otras enfermedades (537) por lo que el número total de pacientes control es de 1074 pacientes.

Radiografía de tórax

Reporte Final para ser presentado ante el Ministerio de Ciencia y Tecnología de la Republica de Colombia

Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín

El tamaño de muestra mínimo de pacientes con COVID-19 es de 601 y pacientes NO-Covid 19 es de 601. Para este cálculo se asume un desempeño diagnóstico de la RT-PCR de 0.85 (AUC1) (27,28) y un desempeño diagnóstico promedio de los modelos de inteligencia artificial aplicados a la radiografía de tórax de 0.70 (AUC2). Este cálculo se realiza con un enfoque de no-inferioridad entre pruebas diagnósticas pareadas binarias (29) y asume un error tipo I de 0,05, un poder estadístico de 0,8, una proporción de asignación de 1 para la detección de una diferencia mínima entre el desempeño de ambas pruebas de 0,1. Los pacientes No-COVID pueden ser pacientes normales (601) o pacientes con otras enfermedades (601) por lo que el número total de pacientes control es de 1202 pacientes.

Homogeneización de la base de datos de imágenes

La nomenclatura de los pacientes incluidos en el estudio se homogeneizará y el nombre de sus series se estructurará para facilitar su análisis (Ver protocolo de recolección y adquisición de las imágenes). Una vez recopilada la base de datos inicial, se realizará una segunda verificación del dataset bajo los siguientes pasos:

- Confirmar criterios de inclusión y exclusión.
- Confirmar adecuada calidad de imagen.
- Confirmar hallazgos positivos sugestivos de COVID-19 en esta cohorte de pacientes.

FASE 2: VALIDACION RETROSPECTIVA

*Evaluación retrospectiva del desempeño en la **clasificación** de los modelos y del grupo de radiólogos para el diagnóstico de neumonía por COVID-19*

Para evaluar el desempeño retrospectivo comparativo entre los modelos y los radiólogos para el diagnóstico de COVID-19 en los estudios de radiografía y tomografía se utilizará como referencia el resultado de la RT-PCR. Las predicciones de los modelos de IA se compararán con 5 radiólogos (BM, TS, JC, JV, MA) BM. Radiólogo de cuerpo con 16 años de experiencia, TS. Radiólogo de tórax con 20 años de experiencia, JV. Radiólogo de abdomen con 6 años de experiencia, JC. Radiólogo de cuerpo con 10 años de experiencia, JV. Radiólogo de cuerpo con 10 años de experiencia, MA. Radiólogo de cuerpo con 8 años de experiencia. A cada uno de los radiólogos se le entregará las mismas imágenes que se ingresarán al modelo al igual que los datos clínicos de estos pacientes. Los radiólogos predecirán el estado de COVID-19 para estos pacientes. Estas predicciones se compararon con las predicciones de los algoritmos y los resultados de la RT-PCR.

Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín

*Evaluación retrospectiva del desempeño en la **cuantificación** de los modelos y del grupo de radiólogos para el diagnóstico de neumonía por COVID-19*

Radiografía de tórax

El compromiso pulmonar por COVID-19 se ha intentado cuantificar con el score de severidad RALE (Utilizado para la evaluación radiológica de edema pulmonar) (33). Este score ha demostrado correlación con el riesgo de ingreso a UCI en los pacientes con neumonía por COVID-19(34). Versiones modificadas del score RALE (35) han permitido su aplicación en pacientes con COVID. El score modificado difiere del score original en el hecho de que los campos pulmonares no se encuentran divididos en cuadrantes. En este score a cada pulmón se le asigna un porcentaje de compromiso en vidrio esmerilado o consolidación de la siguiente forma: 0=Ninguna, 1=<25%, 2=25-50%, 3=51-75%, 4=>75%. Cada pulmón se multiplica por un score de densidad general: 1=Opacidades ligeras, 2=Opacidades moderadas, 3=Opacidades densas. Un paciente normal recibirá un score de 0, mientras que un paciente con consolidaciones completas de ambos pulmones recibirá un score máximo de 24.

Tomografía de tórax

La estimación de compromiso de la tomografía de tórax se realizará de forma visual dado la facilidad y velocidad para su evaluación. Se dará un score de 1-5 de la siguiente forma: (1) Compromiso <5% del parénquima pulmonar, (2) Compromiso <5-25% del parénquima pulmonar, (3) Compromiso 26-49% del parénquima pulmonar, (4) Compromiso 50-75% del parénquima pulmonar y (5) Compromiso >75% del parénquima pulmonar. A cada radiólogo se le entregará las mismas imágenes que se ingresarán al modelo al igual que los datos clínicos de estos pacientes. Los radiólogos estimarán el porcentaje de compromiso pulmonar en estos pacientes. Estas estimaciones se compararon con los porcentajes de compromiso pulmonar arrojado por los algoritmos.

FASE 3: VALIDACION PROSPECTIVA

Referencias para avanzar en la validación prospectiva de los modelos

Para avanzar en la valoración prospectiva de los modelos de clasificación y segmentación, estos deben mostrar adecuado desempeño en estas tareas. Se considerará un adecuado desempeño diagnóstico: No inferioridad del desempeño de los modelos de clasificación comparado con el desempeño más bajo de los radiólogos.

Reporte Final para ser presentado ante el Ministerio de Ciencia y Tecnología de la Republica de Colombia

Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín

En caso de que el desempeño diagnóstico de todos los radiólogos supere el desempeño de los modelos, se avanzará en su evaluación prospectiva, si el desempeño es superior al reportado en la literatura (Tabla 3). Este mismo enfoque aplicará para evaluar el desempeño de los modelos de segmentación (Tabla 4).



Tabla 3. Modelos de referencia clasificación

Tipo de cohorte: Local/Externa Retrospectiva/prospectiva	Prevalencia de la condición en la población	Localización	Goldstandart	AUC	F1 AUC	PR AUC	Spec.	Sens.	Exactitud	Mod. (CT/CR)	Ref.
Retrospectiva. Validación externa	44/99 Covid/N.Viral	China	RT-PCR	0.78	0.63	-	0.67	0.74	0.73	CT	(36)
Retrospectiva. Validación externa	521/665 Covid/N.No Covid	China	RT-PCR	0.90	0.87	-	0.86	0.89	0.96	CT	(37)
Retrospectiva. Validación externa.	419/486 Covid/N.No Covid	China	RT-PCR	0.86	-	-	0.75	0.83	0.79	CT	(38)
Retrospectiva Validación externa.	155/36/17 Covid/N.No Covid/Normal	China	RT-PCR	0.98			0.85	0.92	0.90	CT	(30)
Prospectiva	261/329/200 Covid/N.No Covid/Normal	China	RT-PCR	0.91-0.97	-	-	0.80-0.84	0.84-0.88	0.84-0.91	CT	(30)
Retrospectiva. Validación externa	52/45/10 Covid/N.No Covid/Normal	Ecuador	RT-PCR	0.93	-	-	0.82	0.86	0.84	CT	(30)
Retrospectivo. Validación externa (MultiView)	368/127 Covid/N.No Covid/	China	RT-PCR	0.81	-	-	0.61	0.81	0.76	CT	(39)
Retrospectiva, Desempeño radiólogo. Ciego (n: 5)	521/665 Covid/N.No Covid	China	RT-PCR	0.80-0.92	-	-	0.88	0.79	0.88	CT	(37)
Retrospectiva, Desempeño radiólogo. Ciego (n: 2)	419/486 Covid/N.No Covid	China	RT-PCR	0.73-0.84	-	-	0.90-0.93	0.56-0.74	0.73-0.84	CT	(38)
Retrospectivo. Multicéntrico	223/231 Covid/ Normal	Holanda	RT-PCR	0.81	-	-	0.58-0.69	0.60-0.85	-	CR	(40)
Retrospectiva, Desempeño radiólogo. Ciego (n: 5)	223/231 Covid/ Normal	Holanda	RT-PCR	-	-	-	0.26-0.93	0.61-0.88	-	CR	(40)
Retrospectivo	284/330/327/310 Covid/ N.Bac /N.Vir/Normal	India	RT-PCR	-	0.83-0.95	0.98	0.95-0.97	0.87-0.92	0.89	CR	(41)
Retrospectivo	250/2753/3520 Covid/Otros/Normal	Italia	RT-PCR	-	0.89-0.94	-	0.94-0.98	0.87-0.96	0.97	CR	(42)
Retrospectiva	125/500/500 Covid/Otros/Normal	Turquia	RT-PCR	-	0.87	-	0.92	0.85	0.87	CR	(43)
Retrospectiva	106/5977/18619/18774 Covid/N.Viral/N.Bacte/Normal	China	RT-PCR	0.74-0.83	-	-	0.45-0.71	0.66-0.94	0.69-0.72	CR	(44)
Retrospectiva	266/5977/8066 Covid/N.NO covid/Normal	Mixto	RT-PCR	-	-	-	0.85	0.94	0.93	CR	(45)

Abreviaciones: N: Neumonía

Tabla 4. Modelos de referencia segmentación

Modelo	Parénquima pulmonar		Lesiones pulmonares		Modalidad (CT/CR)	Ref.
	Coefficiente de Dice \pm SD	Exactitud por pixel \pm SD	Coefficiente de Dice \pm SD	Exactitud por pixel \pm SD		
U-Net	0.959 \pm 0.003	0.967 \pm 0.002	0.547 \pm 0.021	0.587 \pm 0.012	CT	(30)
DRUNET	0.959 \pm 0.003	0.966 \pm 0.003	0.562 \pm 0.015	0.587 \pm 0.012	CT	(30)
FCN	0.957 \pm 0.002	0.966 \pm 0.002	0.553 \pm 0.011	0.637 \pm 0.016	CT	(30)
SegNet	0.958 \pm 0.004	0.967 \pm 0.003	0.555 \pm 0.018	0.610 \pm 0.020	CT	(30)
DeepLabv3	0.958 \pm 0.003	0.965 \pm 0.004	0.587 \pm 0.012	0.662 \pm 0.011	CT	(30)

Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín



Two CNNs	0.93-0.94	-	-	-	CR	(46)
CNN	0.96	-	-	-	CR	(47)
Digital Chest CNN	0.96-0.97	-	-	-	CR	(48)
Deformable	0.80 – 0.88	-	-	-	CR	(49)

*Evaluación PROSPECTIVA del desempeño en la **clasificación** de los modelos y del grupo de radiólogos para el diagnóstico de neumonía por COVID-19.*

Para evaluar el desempeño de los radiólogos en el diagnóstico de neumonía de COVID se incluirán pacientes de forma prospectiva desde septiembre 2020 a octubre 2020. Estos pacientes serán reclutados en cada una de las instituciones. Para esta evaluación se obtendrán todas las radiografías y tomografías de tórax realizadas bajo la indicación de síntomas respiratorios. Estas imágenes serán sometidas a la evaluación de los algoritmos previamente entrenados. El resultado de estas predicciones será evaluado a la luz de la práctica clínica diaria. El análisis de los resultados se realizará de la misma manera que el estudio retrospectivo.

*Evaluación PROSPECTIVA del desempeño en la **cuantificación** de los modelos en neumonía por COVID-19.*

La evaluación prospectiva de la cuantificación de los estudios solo se realizará en la tomografía de tórax ya que la información sobre el compromiso pulmonar se puede obtener a partir de los reportes usuales de los radiólogos. El desempeño prospectivo de los algoritmos de segmentación también se realizará de forma prospectiva siguiendo la misma metodología que el estudio retrospectivo. Se utilizará el mismo score que se utilizó en la evaluación retrospectiva (1-5). Solamente se compararán los pacientes que tanto el radiólogo como el modelo consideren positivos. Se comparará el porcentaje de compromiso pulmonar asignado por el radiólogo con el porcentaje de compromiso asignado por el modelo.

*Evaluación prospectiva **externa** – Sistema de tele-radiología*

Para esta evaluación se utilizará una pequeña cohorte externa de pacientes (200 casos positivos para COVID-19 y 200 casos negativos por cada modalidad) la cual se ingresará a través del sistema de tele-radiología. Se evaluará los modelos de segmentación y los modelos de cuantificación en ambas modalidades. Esta evaluación seguirá la misma metodología que la valoración prospectiva.

*Evaluación prospectiva - Implementación e **interpretabilidad***

Con el objetivo de incrementar la interpretabilidad de los modelos generados se adoptará el método de Asignación de activación de clase ponderada por gradiente (Grad-CAM). Con este método se visualizarán las regiones de mayor importancia que guían las decisiones del modelo de aprendizaje profundo (Ver documento: Arquitectura de los modelos).

Reporte Final para ser presentado ante el Ministerio de Ciencia y Tecnología de la Republica de Colombia

Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín



Un subgrupo de estos mapas de gradiente de pacientes positivos para COVID-19 (n=100) serán presentados a los radiólogos durante la evaluación prospectiva quienes clasificarán los mapas de calor como adecuadas o inadecuadas para la predicción de los pacientes con COVID-19.

ASPECTOS ETICOS Y DE PRIVACIDAD

Este protocolo se rige por la ética de la investigación en seres humanos establecida por la Asociación Médica Mundial, acorde con la Declaración de Helsinki (Asamblea General WMA 64, Fortaleza, Brasil, octubre 2013). Esta declaración establece que la investigación biomédica en seres humanos solo puede ser realizada por personas calificadas bajo la supervisión de un profesional médico competente (art. 395). La investigación no puede realizarse legítimamente a menos que la importancia de su objetivo este en proporción al riesgo que corre el sujeto de experimentación (art 396). En esta investigación se preservarán con exactitud los datos de los resultados obtenidos, y se garantizará la protección de los datos recolectados en concordancia con los principios reconocidos científicamente.

El presente protocolo se acoge y sigue las “Normas Científicas, Técnicas y Administrativas para la Investigación en Salud” establecidas en la Resolución No. 008430 de 1993 del Ministerio de Salud. Es un estudio que empleará técnicas y métodos de investigación documental retrospectivo, en el que se utilizarán exámenes radiológicos realizados y estudiados de pacientes en quienes ha estado indicado. No se realizará estudios sobre nuevos pacientes. Consideramos que se trata de un estudio sin riesgos para los pacientes incluidos.

Este estudio fue sometido a los comités de ética de cada hospital y de la universidad EAFIT en la ciudad de Medellín. No requirió diligenciamiento de consentimiento informado dado que se trata de un estudio retrospectivo, el cual evaluará información que no incluye datos de identificación de pacientes, y no representará un riesgo potencial para los pacientes.

Anticipando potenciales brechas de confidencialidad se elaboró una guía para la pseudo - anonimización de los datos clínicos y la anonimización de las imágenes. Estas guías se realizaron de acuerdo con las recomendaciones de la HIPAA y GDPR (18,19) e incluyeron listas de chequeo para verificar la adecuada realización de la anonimización/pseudo-anonimización.

Este estudio generará dos modelos de inteligencia artificial validados clínicamente y una plataforma para su despliegue en el territorio nacional. Estos productos serán distribuidos de forma gratuita y su distribución, creación y despliegue serán patrocinados por el ministerio de ciencias y los hospitales e instituciones educativas involucrados en su creación. Cualquier otro producto que se genere a partir del uso de estas imágenes y los datos clínicos debe ser notificado, discutido y autorizado por cada una de las instituciones implicadas en su generación previo a su distribución. Para la

Reporte Final para ser presentado ante el Ministerio de Ciencia y Tecnología de la Republica de Colombia

Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín
elaboración del presente estudio se diseñó un protocolo de anonimización y de privacidad del paciente.

La información clínica y las imágenes tiene una fecha de caducidad. Los vacíos en el marco legal colombiano obligan a limitar el tiempo en el uso de las imágenes. Un año después (Tiempo arbitrario) de que se logre el producto de este estudio, todas las instituciones involucradas deben asegurar la eliminación completa de las imágenes. No se autoriza copias parciales o completas de las imágenes ni de los datos clínicos a menos que se requiera para avanzar en el estudio.

Las imágenes y los datos clínicos anonimizados no se compartirán a terceros por fuera de las instituciones aprobadas en este estudio. Los datos clínicos y las imágenes no serán sujetos de comercialización. La participación del paciente en el estudio puede proporcionar información que ayude a los investigadores a determinar la utilidad del método.

En este estudio se hará una valoración de los hallazgos de los pacientes y no existe un riesgo en el desenlace final del diagnóstico patológico o en la historia natural de la enfermedad.

Los resultados de este estudio serán publicados en revistas científicas. Todos los datos incluidos en las publicaciones serán anónimos para preservar la confidencialidad.

PRESENTACIÓN DE LOS RESULTADOS

A continuación, se describen los instrumentos que se utilizaran para presentar los resultados en el estudio clínico. Estos instrumentos serán reportados en los artículos y reportes del proyecto.

Características de la cohorte

Tabla 1a. Características de la población COVID-19

	Entrenamiento	Validación	Total
Edad			
< 20			
20-39			
40-59			
>60			
Sexo			
Masculino			
Femenino			
Localización			
Ambulatorio			
Hospitalización			
Urgencias			
Hallazgos por imagen CT			
Típico			
Indeterminado			
Atípico			
Hallazgos por imagen CR			

Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín



Típico			
Indeterminado			

Tabla 1b. Características de la población COVID-19 Negativo

	Entrenamiento	Validación	Total
Edad			
< 20			
20-39			
40-59			
>60			
Sexo			
Masculino			
Femenino			
Localización			
Ambulatorio			
Hospitalización			
Urgencias			
Indicación clínica (Extraer DICOM tag)			
Neumonía			
Tuberculosis			
Malignidad			
Falla cardíaca			
Estudio de rutina			
Otros			
Hallazgos por imagen CR			
Sin alteraciones			
Neumonía			
Malignidad			
Derrame pleural			
Edema pulmonar			
Otros			
Hallazgos por imagen CT			
Sin alteraciones			
Neumonía			
Malignidad			
Derrame pleural			
Edema pulmonar			
Otros			

Desempeño de los modelos

Tabla 2a. Radiografía de tórax

Tipo de cohorte: Local/Externa Retrospectiva/prospectiva	Prevalencia de la enfermedad	Tamaño de la Muestra (Covid/Otros/Normal)	Sensibilidad	Especificidad	ROC AUC	PR AUC	Exactitud
Evaluación retrospectiva							
Por paciente							
Por imagen							
Evaluación prospectiva							
Por paciente							
Por imagen							

Tabla 2b. Tomografía de tórax

Tipo de cohorte: Local/Externa Retrospectiva/prospectiva	Prevalencia de la enfermedad	Tamaño de la Muestra (Covid/Otros/Normal)	Sensibilidad	Especificidad	ROC AUC	PR AUC	Exactitud
Evaluación retrospectiva							
Por paciente							

Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín



Evaluación prospectiva									
Por paciente									

Comparación de los modelos con el desempeño del radiólogo para el diagnóstico de COVID-19 (Valoración retrospectiva/prospectiva)

Tabla 3a. Valoración retrospectiva

	Medida	Desempeño				Diferencia			
		Desempeño Radiólogo CR (%)	Desempeño Radiólogo CT (%)	Desempeño Modelo CR (%)	Desempeño Modelo CT (%)	Diferencia Desempeño CR	Valor P	Diferencia Desempeño CT	Valor P
Radiólogo 1 (BM)	Exactitud								
	Sensibilidad								
	Especificidad								
Radiólogo 2 (TS)	Exactitud								
	Sensibilidad								
	Especificidad								
Radiólogo 3 (JC)	Exactitud								
	Sensibilidad								
	Especificidad								
Radiólogo 4 (JV)	Exactitud								
	Sensibilidad								
	Especificidad								
Radiólogo 5 (MA)	Exactitud								
	Sensibilidad								
	Especificidad								
Media radiólogos	Exactitud								
	Sensibilidad								
	Especificidad								

Tabla 3b. Valoración prospectiva.

	Medida	Desempeño				Diferencia			
		Desempeño Radiólogo CR (%)	Desempeño Radiólogo CT (%)	Desempeño Modelo CR (%)	Desempeño Modelo CT (%)	Diferencia Desempeño CR	Valor P	Diferencia Desempeño CT	Valor P
Radiólogo 1 (BM)	Exactitud								
	Sensibilidad								
	Especificidad								
Radiólogo X (TS)	Exactitud								
	Sensibilidad								
	Especificidad								

Tabla 4. Acuerdo en el porcentaje de compromiso pulmonar entre los modelos de segmentación y los radiólogos en la tomografía de tórax.

	Score	Modelo de IA – Cuantificación				
		1. n (%)	2. n (%)	3. n (%)	4. n (%)	5. n (%)
Radiólogo	1. n (%)					
	2. n (%)					
	3. n (%)					
	4. n (%)					
	5. n (%)					

Tablas suplementarias

Reporte Final para ser presentado ante el Ministerio de Ciencia y Tecnología de la Republica de Colombia

Tabla S1. Lista de los hospitales participantes y número de casos.

Hospital	Localización	Estudios positivos COVID-19	Estudios Negativos COVID-19

Tabla S2a. Lista de los hospitales participantes y número de casos utilizado para los modelos aplicados a la radiografía de tórax

Modelo radiografía de tórax	Set de entrenamiento	Set de validación	Set de test
Cohorte interna	X / x	X / x	X / x
Distribución simétrica	X / x	X / x	X / x

Tabla S2b. Lista de los hospitales participantes y número de casos utilizado para los modelos aplicados a la tomografía de tórax

Modelo tomografía de tórax	Set de entrenamiento	Set de validación	Set de test
Cohorte interna	X / x	X / x	X / x
Distribución simétrica	X / x	X / x	X / x

Tabla S3. Desempeño de los modelos en los sets de entrenamiento, validación y test

		Sensibilidad (%)	Especificidad (%)	ROC AUC	ROC PR
Cohorte interna	Set de entrenamiento (n = X)	X / x	X / x	X / x	
	Set de validación (n = X)				
	Set de test (n = X)				
Distribución simétrica	Set de entrenamiento (n = X)	X / x	X / x	X / x	
	Set de validación (n = X)				
	Set de test (n = X)				

PROTOCOLOS DE IMAGENES

A continuación, se describen los protocolos de radiografía y tomografía de tórax que se utilizaron para obtener las imágenes en el presente estudio.

Protocolos tomografía de tórax simple

	IPS universitaria		Clínica CES		San Vicente de Paul	HPTU		
Ubicación	bloque 2 piso 2	Bloque 1 piso 1 Urgencias	Primer piso	Primer piso	Segundo piso	Segundo piso torre B Tac 16 Siemens	Segundo piso Torre B PET Siemens	Segundo piso Torr A Urgencias Siemens
Marca	TOSHIBA	GENERAL ELECTRIC	Phillips	GE	GE			

Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín



Modelo	AQUILION 64	BRIVO	Ingenuity	LIGHTSPEED	OPTIMA PLUS 520	SOMATOM EMOTION	Biograph MCT	SOMATOM Definition AS
Serie	HCA0682313 - 1AA682352							
Numero de detectores	64	2	128	64	16	16	64	32
Kvp	120	120	100	120	140	130	100-110	100-110
Corriente del tubo (mAs)	Automatico	Automatico	Automatica	Automatico	Automatico	automático	automático	Automático
Pitch	0.8	1	1,4	0.9	1.3	1	1.2	1.2
Colimacion	1	3	Automatica	Automatica	Automatica	Automatica	Automatica	Automatica
Tiempo de rotación (s/rot)	0,5	1,5	0,5	0,8	0,8	0.6	0.3	0.3
Espesor de corte (mm)	1	3	0,6	0,6	0,6	1.2	0.6	0.6
Campo de adquisición de la imagen (Field of view) (mm)	400	400	500	360	360	320	395	395
Algoritmo de reconstrucción	Estándar	Estándar	Estándar	Estándar	Estándar	estandar	Estandar	Estandar
Sistema PACS	ClearCanvas	ClearCanvas	Carestream	IMPAX	IMPAX	ENTERPRISE	ENTERPRISE	ENTERPRISE
Sistema RIS	GHIPS	GHIPS	Sonda	SAP	SAP	SERVINTE	SERVINTE	SERVINTE

Protocolos de reconstrucción axial parénquima pulmonar

Algoritmo de reconstrucción	Estándar	Estándar	Estándar	Estándar	Estándar	Estándar	Estándar	Estándar
Espesor de corte de reconstrucción (mm)	5	3	2	1 - 5	1 - 5	1 - 3	1 - 3	1 - 3
Intervalo de reconstrucción (mm)	0,5	3	1	1,25	1,25	1	0.7	0.7

Equipos de radiografía

	IPS universitaria		Clínica CES	San Vicente de Paul	HPTU						
	Equipo en sala	Equipo portatil	Equipo en sala	Equipo en Sala	Equipo portatil	Equipo portatil	Equipo portatil	Equipo portatil	Equipo de sala	Equipo de sala	Equipo de sala
Marca	AMRAD	SHIMADZU	PHILLIPS	PHILLIPS	PHILLIPS	Siemens	Siemens	Siemens	Siemens	Siemens	Siemens
Modelo	SUMMIT INDUSTRIES	MOBILEART EVOLUTION	Duodiagnos	DIGITAL DIAGNOST	AQUILA 300 PLUS	Mobilet XP	Mobilet XP	Miramax	Axiom luminos TF	Luminos Fusion	Multix Fusion
Serie	AMA246-0712	41000330C001	-	-	-	-	-	-	-	-	-
Kvp	120	120	80	125	125	75	75	75	120	120	120
Corriente del tubo (mAs)	0,5-0,7	0,5-0,7	0,7	0,6	0,6	0.5	0.5	0.5	0.2-0.5	0.2-0.5	0.2-0.5
Distancia (mt)	1,8	1	1	1,8	1	1	1	1	1.5	1.5	1.5

FASE 4: DESPLIEGUE DEL SISTEMA DE TELERRADIOLOGIA

Esta fase consiste en el despliegue a nivel nacional de un Sistema de tele-radiología amigable, eficiente y estable. El cual permitirá seguir alimentando la base de datos actual. Se estima que el resultado final de este sistema logre clasificar y cuantificar adecuadamente el compromiso por COVID en las radiografías y tomografías de tórax. El resultado que proveerá este sistema debe permitir la clasificación de los estudios en neumonía por COVID, neumonía bacteriana, otras enfermedades y estudios normales.

Reporte Final para ser presentado ante el Ministerio de Ciencia y Tecnología de la Republica de Colombia

Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín



Los algoritmos que serán incorporados en el sistema de tele-radiología serán los algoritmos cuyo desempeño haya sido confirmado en los estudios clínicos retrospectivos y prospectivos descritos. Estos algoritmos fueron previamente validados en la población colombiana y fueron considerados robustos para su uso en las modalidades de radiografías y tomografías de tórax con un AUC mayor a 0.8 – 0.9 para la detección de signos radiográficos de coronavirus.

CONCLUSIONES

En la actualidad, el virus del COVID-19 tiene una distribución nacional y es responsable de más de 356.000 contagios en el país. Latinoamérica representa un foco tardío de COVID-19. Esto le permite la posibilidad de materializar los aprendizajes en el campo de la salud y la tecnología que otros países han alcanzado con anterioridad a raíz de la pandemia.

Nuestro estudio representa en el primer estudio de tipo prospectivo, multiinstitucional en un país latinoamericano abriendo la posibilidad de explorar el campo de la inteligencia artificial aplicado a las imágenes diagnósticas en múltiples problemas de salud pública. El perfil a gran escala y de distribución regional es aplicable a otra serie de problemas que incluyen enfermedades mucho más prevalentes como las enfermedades cerebrovasculares.

Una de las fortalezas del proyecto actual incluye la posibilidad de optimizar el desempeño de los modelos en la tarea de segmentación de las lesiones pulmonares dado que existe un muy bajo rendimiento de los modelos reportados en esta tarea específica. Adicionalmente la cantidad de estudios utilizados en la construcción de estos modelos es muy superior a los reportados en la literatura tanto para radiografía de tórax como para tomografía. Esto permite generalizar con mayor confianza sus resultados y desempeño.

Otra de las ventajas del proyecto actual es la posibilidad de validar el desempeño de los modelos en el contexto de la pandemia, permitiendo desplegar y evaluar en tiempo real los mismos durante el pico de la pandemia. El análisis temporal del despliegue de los modelos y el momento de la pandemia permite interpretar su impacto global. Sin embargo, es a través de la validación prospectiva de estos modelos que se entenderá su real impacto en la atención en salud.

LIMITACIONES

Existen múltiples limitaciones en el proyecto actual. La principal limitación es la sobreestimación de la sensibilidad y especificidad de los modelos en la práctica clínica. Adicionalmente los estudios que se incluyen se realizan en una población hospitalizada, lo que introduce sesgo de selección reclutando los pacientes más enfermos. También se realiza el estudio bajo la premisa de que todos los pacientes van a presentar alteraciones en sus estudios cuando

Reporte Final para ser presentado ante el Ministerio de Ciencia y Tecnología de la Republica de Colombia

Viabilidad y validación de la aplicación de modelos de inteligencia artificial para la detección de neumonía en los servicios de radiología de hospitales de tercer y cuarto nivel de Medellín hasta la mitad de los pacientes se pueden presentar con un estudio normal de radiografía o de tomografía.

Finalmente, la utilidad de las imágenes diagnósticas en pacientes con COVID sigue siendo incierta. Algunas guías recomiendan los estudios de imágenes solo en casos de pacientes con complicaciones, ausencia de certeza diagnóstica o enfermedades severas. La utilidad de los modelos implementados va a depender directamente de su aplicabilidad en la práctica clínica.

Nuestro estudio excluye el mediastino en el proceso de la segmentación y el énfasis de los modelos se realizará en el parénquima pulmonar por lo que las enfermedades del mediastino no son abordadas en el presente estudio. La heterogeneidad en la adquisición de las imágenes es otra de las limitaciones. Aunque el diseño multicéntrico del estudio puede incrementar la posibilidad de crear un algoritmo generalizable, una cantidad insuficiente de estudios con una gran heterogeneidad en el origen de estos puede disminuir el desempeño de los algoritmos de clasificación y de cuantificación de la enfermedad.

El desempeño de la plataforma nacional en el ámbito clínico también es desconocido.

CONTENIDO EN LINEA

Las referencias adicionales, los modelos presentados, la información extendida y la información suplementaria estarán disponible en línea sujeto a la aprobación de los autores principales.

Bibliografía

1. Yang Y, Yang M, Shen C, Wang F, Yuan J, Li J, et al. Evaluating the accuracy of different respiratory specimens in the laboratory diagnosis and monitoring the viral shedding of 2019-nCoV infections. medRxiv. 2020 Feb 17;2020.02.11.20021493.
2. Seaman CP, Tran LTT, Cowling BJ, Sullivan SG. Self-collected compared with professional-collected swabbing in the diagnosis of influenza in symptomatic individuals: A meta-analysis and assessment of validity. Vol. 118, Journal of Clinical Virology. Elsevier B.V.; 2019. p. 28–35.
3. Rubin GD, Haramati LB, Kanne JP, Schluger NW, Yim J-J, Anderson DJ, et al. The Role of Chest Imaging in Patient Management during the COVID-19 Pandemic: A Multinational Consensus Statement from the Fleischner Society. Radiology [Internet]. 2020 Apr 7 [cited 2020 Apr 9];201365. Available from: <http://pubs.rsna.org/doi/10.1148/radiol.2020201365>

4. Driggin E, Madhavan M V., Bikdeli B, Chuich T, Laracy J, Bondi-Zoccai G, et al. Cardiovascular Considerations for Patients, Health Care Workers, and Health Systems During the Coronavirus Disease 2019 (COVID-19) Pandemic. *J Am Coll Cardiol*. 2020 Mar;
5. Lin L, Li TS. [Interpretation of “Guidelines for the Diagnosis and Treatment of Novel Coronavirus (2019-nCoV) Infection by the National Health Commission (Trial Version 5)”]. *Zhonghua Yi Xue Za Zhi* [Internet]. 2020 Feb 7 [cited 2020 Apr 10];100(0):E001. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/32033513>
6. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology* [Internet]. 2020 Feb 26 [cited 2020 Apr 9];200642. Available from: <http://pubs.rsna.org/doi/10.1148/radiol.2020200642>
7. Wong HYF, Lam HYS, Fong AH-T, Leung ST, Chin TW-Y, Lo CSY, et al. Frequency and Distribution of Chest Radiographic Findings in COVID-19 Positive Patients. *Radiology* [Internet]. 2019 Mar 27 [cited 2020 Apr 12];201160. Available from: <http://pubs.rsna.org/doi/10.1148/radiol.2020201160>
8. ACR. ACR Recommendations for the use of Chest Radiography and Computed Tomography (CT) for Suspected COVID-19 Infection [Internet]. 2020 [cited 2020 Apr 12]. p. 18–9. Available from: <https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection>
9. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med*. 2020 Feb 28;
10. Ruíz Gómez F, BURGOS BERNAL Secretario General G, Gutiérrez Álvarez O, Vega Figueroa A. ASEGURAMIENTO EN CIFRAS Exámenes por imágenes en el sistema general de Seguridad Social en Salud Régimen Contributivo 2012-2014 Boletín mensual No.13 Mayo 2016 Dirección de Regulación de Beneficios, Costos y Tarifas del Aseguramiento en salud [Internet]. Bogotá; 2016 [cited 2020 Apr 9]. Available from: <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VP/RBC/examen-es-por-imagen-es-en-el-sistema-general-de-seguridad-social-en-salud-regimen-contributivo-2012-2014.pdf>
11. Ministerio de Salud y Protección Social. Estudio 2015- Datos de 2013. Atlas de Variaciones Geográficas en Salud de Colombia 2015- Estudio piloto - Resultados de uso de Tomografía Axial Computarizada TAC [Internet]. Bogotá; 2017 [cited 2020 Apr 9]. Available from:

<https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/PES/Resultados-generales-atlas-salud-tac-2015.pdf>

12. Colombia rural, sin hospitales para atender la crisis [Internet]. [cited 2020 Apr 9]. Available from: <https://www.semana.com/nacion/articulo/colombia-rural-sin-hospitales-para-atender-la-criisis/659730>

13. Rachh P, Levey AO, Lemmon A, Marinescu A, Auffermann WF, Haycock D, et al. Reducing STAT Portable Chest Radiograph Turnaround Times: A Pilot Study [Internet]. Vol. 47, Current Problems in Diagnostic Radiology. Mosby Inc.; 2018 [cited 2020 Apr 9]. p. 156–60. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28705527>

14. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT. Radiology [Internet]. 2020 Mar 19 [cited 2020 Apr 9];200905. Available from: <http://pubs.rsna.org/doi/10.1148/radiol.2020200905>

15. Dadário AMV, Paiva JPQ, Chate RC, Machado BS, Szarf G. Regarding “Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT”. Radiology [Internet]. 2020 Apr 3 [cited 2020 Apr 9];201178. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/32243239>

16. Wang L, Wong A. COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest Radiography Images. 2020 Mar 22 [cited 2020 Apr 12]; Available from: <http://arxiv.org/abs/2003.09871>

17. Ministerio de educación. Ley 657 de Junio 7 de 2001. 2001 https://www.mineducacion.gov.co/1621/articles-86095_archivo_pdf.pdf; 2019.

18. G3P - Good Privacy Protection Practice in Clinical Research: Principles of Pseudonymization and Anonymization / Edition 1 by Karl-Heinz Schriever, Markus Schroder | | 9783110367645 | Hardcover | Barnes & Noble® [Internet]. [cited 2020 Jun 22]. Available from: <https://www.barnesandnoble.com/w/g3p-good-privacy-protection-practice-in-clinical-research-karl-heinz-schriever/1134092497>

19. Official 2020 HIPAA Compliance Checklist [Internet]. [cited 2020 Jun 22]. Available from: <https://www.hipaajournal.com/hipaa-compliance-checklist/>

20. COVID19_Medellin - MEData | Tableau Public [Internet]. [cited 2020 Jul 20]. Available from:

https://public.tableau.com/profile/medata#!/vizhome/COVID19_Medellin/COVID19_Medellin

21. 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 Jan 6;162(1):55–63.
22. 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 - J Am Med Assoc*. 2020 Apr 7;323(13):1239–42.
23. CDC. Public Health Recommendations for Community-Related Exposure [Internet]. Centers for Disease Control and Prevention. 2020 [cited 2020 Apr 10]. Available from: <https://www.cdc.gov/coronavirus/2019-ncov/php/public-health-recommendations.html><http://files/15949/public-health-recommendations.html>
24. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. Vol. 48, *Journal of Biomedical Informatics*. Academic Press Inc.; 2014. p. 193–204.
25. Goksuluk D, Korkmaz S, Zararsiz G, Karaagaoglu A, Ergu. easyROC: An Interactive Web-tool for ROC Curve Analysis Using R Language Environment. *R J* [Internet]. 2016 [cited 2020 Jul 20];8(2):213. Available from: <https://journal.r-project.org/archive/2016/RJ-2016-042/index.html>
26. Team RC. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria; 2013.
27. Yang H, Lan Y, Yao X, Lin S, Xie B. Evaluation on the diagnostic efficiency of different methods in detecting COVID-19. *medRxiv* [Internet]. 2020 Jun 26 [cited 2020 Jul 20];2020.06.25.20139931. Available from: <https://www.medrxiv.org/content/10.1101/2020.06.25.20139931v1>
28. Hasab AA. COVID-19 Screening by RT-PCR: An Epidemiological Modelling. 2020 Jun 18 [cited 2020 Jul 20]; Available from: <https://doi.org/10.21203/rs.3.rs-35816/v1>
29. Liu JP, Hsueh HM, Hsieh E, Chen JJ. Tests for equivalence or non-inferiority for paired binary data. *Stat Med* [Internet]. 2002 Jan 30 [cited 2020 Apr 9];21(2):231–45. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11782062>
30. Zhang K, Liu X, Shen J, Li Z, Sang Y, Wu X, et al. Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of
Reporte Final para ser presentado ante el Ministerio de Ciencia y Tecnología de la Republica de Colombia

COVID-19 Pneumonia Using Computed Tomography. Cell. 2020 Jun 11;181(6):1423-1433.e11.

31. Simpson S, Kay FU, Abbara S, Bhalla S, Chung JH, Chung M, et al. Radiological Society of North America Expert Consensus Statement on Reporting Chest CT Findings Related to COVID-19. Endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA. Radiol Cardiothorac Imaging [Internet]. 2020 Apr 1 [cited 2020 Jul 12];2(2):e200152. Available from: <http://pubs.rsna.org/doi/10.1148/ryct.2020200152>

32. The British Society of Thoracic Imaging. COVID-19 BSTI Reporting templates | The British Society of Thoracic Imaging [Internet]. Bsti. 2020 [cited 2020 Jul 12]. Available from: <https://www.bsti.org.uk/covid-19-resources/covid-19-bsti-reporting-templates/>

33. Warren MA, Zhao Z, Koyama T, Bastarache JA, Shaver CM, Semler MW, et al. Severity scoring of lung oedema on the chest radiograph is associated with clinical outcomes in ARDS. Thorax [Internet]. 2018 Sep 1 [cited 2020 Jul 20];73(9):840–6. Available from: <https://thorax.bmj.com/content/73/9/840>

34. Cozzi D, Albanesi M, Cavigli E, Moroni C, Bindi A, Luvarà S, et al. Chest X-ray in new Coronavirus Disease 2019 (COVID-19) infection: findings and correlation with clinical outcome. Radiol Medica [Internet]. 2020 Aug 1 [cited 2020 Jul 20];125(8). Available from: <https://pubmed.ncbi.nlm.nih.gov/32519256/>

35. Li MD, Arun NT, Gidwani M, Chang K, Deng F, Little BP, et al. Automated Assessment and Tracking of COVID-19 Pulmonary Disease Severity on Chest Radiographs using Convolutional Siamese Neural Networks. Radiol Artif Intell [Internet]. 2020 Jul 1 [cited 2020 Aug 5];2(4):e200079. Available from: <http://pubs.rsna.org/doi/10.1148/ryai.2020200079>

36. Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, et al. A deep learning algorithm using CT images to screen for corona virus disease (COVID-19). medRxiv [Internet]. 2020 Apr 24 [cited 2020 Jul 19];1–19. Available from: <https://www.medrxiv.org/content/medrxiv/early/2020/02/17/2020.02.14.20023028.full.pdf%0Ahttp://medrxiv.org/content/early/2020/02/17/2020.02.14.20023028.abstract%0Ahttps://www.medrxiv.org/content/10.1101/2020.02.14.20023028v1%0Ahttp://medrxiv.org/content/e>

37. Bai HX, Wang R, Xiong Z, Hsieh B, Chang K, Halsey K, et al. AI Augmentation of Radiologist Performance in Distinguishing COVID-19 from Pneumonia of Other Etiology on Chest CT. Radiology [Internet]. 2020 Apr 27 [cited 2020 Jul 19];201491. Available from: <http://pubs.rsna.org/doi/10.1148/radiol.2020201491>

38. Mei X, Lee HC, Diao K yue, Huang M, Lin B, Liu C, et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. Nat Med [Internet]. 2020 May 19 [cited 2020 Jul 19];1–5. Available from: <https://doi.org/10.1038/s41591-020-0931-3>
39. Wu X, Hui H, Niu M, Li L, Wang L, He B, et al. Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: A multicentre study. Eur J Radiol [Internet]. 2020 Jul 1 [cited 2020 Jul 19];128. Available from: <https://doi.org/10.1016/j.ejrad.2020.109041>
40. Murphy K, Smits H, Knoop AJG, Korst MBJM, Samson T, Scholten ET, et al. COVID-19 on the Chest Radiograph: A Multi-Reader Evaluation of an AI System. Radiology [Internet]. 2020 May 8 [cited 2020 Jul 19];201874. Available from: <http://pubs.rsna.org/doi/10.1148/radiol.2020201874>
41. Khan AI, Shah JL, Bhat MM. CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. Comput Methods Programs Biomed. 2020 Nov 1;196:105581.
42. Brunese L, Mercaldo F, Reginelli A, Santone A. Explainable Deep Learning for Pulmonary Disease and Coronavirus COVID-19 Detection from X-rays. Comput Methods Programs Biomed [Internet]. 2020 Nov 1 [cited 2020 Jul 20];196:105608. Available from: www.elsevier.com/locate/cmpb
43. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra Acharya U. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med. 2020 Jun 1;121:103792.
44. Zhang J, Xie Y, Liao Z, Pang G, Verjans J, Li W, et al. Viral Pneumonia Screening on Chest X-ray Images Using Confidence-Aware Anomaly Detection. 2020.
45. Wang L, Lin ZQ, Wong A. COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images.
46. Souza JC, Bandeira Diniz JO, Ferreira JL, França da Silva GL, Corrêa Silva A, de Paiva AC. An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks. Comput Methods Programs Biomed. 2019 Aug 1;177:285–96.
47. Kalinovsky A, Kovalev V. Lung Image Segmentation Using Deep Learning Methods and Convolutional Neural Networks [Internet]. [cited 2020 Jul 20]. Available from: <http://imlab.grid.by/>
48. Kaur S, Hooda R, Mittal A, Akashdeep, Sofat S. Deep CNN-based method for segmenting lung fields in digital chest radiographs. In: Reporte Final para ser presentado ante el Ministerio de Ciencia y Tecnología de la Republica de Colombia

49. Annangi P, Thiruvankadam S, Raja A, Xu H, Sun XW, Mao L. Region based active contour method for x-ray lung segmentation using prior shape and lowlevel features. In: 2010 7th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2010 - Proceedings [Internet]. IEEE; 2010 [cited 2020 Jul 20]. p. 892–5. Available from: <http://ieeexplore.ieee.org/document/5490130/>
50. Saito T, Rehmsmeier M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. Brock G, editor. PLoS One [Internet]. 2015 Mar 4 [cited 2020 Apr 10];10(3):e0118432. Available from: <http://dx.plos.org/10.1371/journal.pone.0118432>
51. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: An open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics [Internet]. 2011 Mar 17 [cited 2020 Jul 12];12(1):77. Available from: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-77>
52. Daniel D. Sjoberg. dca: Decision Curve Analysis. 2020.