



UNIVERSITAT DE
BARCELONA



Revista de Bioética y Derecho

Perspectivas Bioéticas

www.bioeticayderecho.ub.edu - ISSN 1886-5887

DOSSIER CUESTIONES BIOÉTICAS DE LA PANDEMIA COVID-19

Inteligencia artificial y *Big Data* como soluciones frente a la COVID-19

Artificial Intelligence and Big Data as solutions to COVID-19

Intel·ligència Artificial i *Big Data* per fer front a la COVID-19

JAIRO MÁRQUEZ DÍAZ *

* Jairo Márquez Díaz. Doctor en Educación. Ingeniero de Sistemas, Especialista en docencia universitaria, en Bioética, en Actuarial y Ciberseguridad, Universidad de Cundinamarca (Colombia). Email: nanotechrd@gmail.com.



Copyright (c) 2020 Jairo Márquez Díaz

Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional.

Resumen

La inteligencia artificial y el *Big Data* se articulan para poder lidiar con diferentes problemas relacionados con el análisis de datos masivos, en particular información de la COVID-19. En el presente artículo se muestran algunos proyectos de investigación relacionados con el aprendizaje profundo, el aprendizaje automático, el *Big Data* y la ciencia de datos, tendientes a dar soluciones plausibles bien en el monitoreo, detección, diagnóstico y tratamiento de las enfermedades asociadas con el virus. Con esto en mente, se muestra la correspondencia entre las tecnologías disruptivas y la información crítica, creando sinergias que permiten elaborar sistemas más avanzados de estudio y análisis facilitando la obtención de datos relevantes para la toma de decisiones sanitarias.

Palabras clave: algoritmos de aprendizaje; analítica avanzada; aprendizaje automático; aprendizaje profundo; ciencia de datos; pandemia; representación de datos; COVID-19.

Abstract

Artificial intelligence and Big Data are articulated to be able to deal with different problems related to the analysis of big data, in particular, information from the COVID-19. In this sense, this article shows some research projects related to deep learning, machine learning, Big Data and data science, aimed to provide plausible solutions in monitoring, detection, diagnosis and treatment of diseases associated with the virus. The correspondence between disruptive technologies and critical information is shown, creating synergies that allow the development of more advanced systems of study and analysis, facilitating the obtaining of relevant data for health decision-making.

Keywords: learning algorithms; advanced analytics; machine learning; deep learning; data science; pandemic; data representation; COVID-19.

Resum

La Intel·ligència Artificial i el *Big Data* s'articulen per poder fer front a diferents problemes relacionats amb l'anàlisi de dades massiu, concretament, informació relativa a la COVID-19. En aquest sentit, en el present article es mostren alguns projectes d'investigació relacionats amb l'aprenentatge profund, l'aprenentatge automàtic, el *Big Data* i la ciència de dades, capaços de donar solucions plausibles en el monitoratge, detecció, diagnòstic i tractament de les malalties associades amb el virus. Amb això en ment, es mostra la correspondència entre les tecnologies disruptives i la informació crítica, creant sinergies que permeten elaborar sistemes més avançats d'estudi i anàlisi facilitant l'obtenció de dades rellevants per a la presa de decisions sanitàries.

Paraules clau: algorismes d'aprenentatge; analítica avançada; aprenentatge automàtic; aprenentatge profund; ciència de dades; pandèmia; representació de dades; COVID-19.

1. Introducción

La pandemia ocasionada por la COVID-19 en el año 2020 (Sohrabi, *et al.*, 2020) desencadenó un efecto devastador en la economía y salud en la población mundial, cuyas implicaciones sociales para los próximos años es aún incierta. Para la detección del COVID-19 se emplean dos tipos de pruebas estándar: la prueba viral que indica si el paciente está infectado y la prueba de anticuerpos que permite observar si el paciente ha tenido previamente una infección. Estas pruebas emplean técnicas como la transcripción inversa y reacción en cadena de la polimerasa (RT-PCR), el flujo lateral inmunocromatográfico o test rápido y el inmunoensayo de tipo ELISA.

Aunque las técnicas son refinadas, su sensibilidad en cuanto a las muestras es relativa. El problema surge en el procedimiento de recolección de las pruebas que se realiza manualmente, a lo que se suma la ingente cantidad de información que se requiere procesar, tales como datos relacionados con la rápida trasmisión, dinámica molecular y celular del virus, trazabilidad acerca de la susceptibilidad poblacional y étnica asociadas con la pandemia, incluso el monitoreo del nivel de riesgo de empleados en una empresa, entre otros aspectos.

Una alternativa complementaria a las técnicas mencionadas es el uso de la inteligencia artificial (IA), el *Big Data* y otras tecnologías disruptivas relacionadas con el análisis de datos masivos, que permiten realizar estudios pormenorizados en diferentes escalas estadísticas, imagenológicas y probabilísticas de información, condensadas en los denominados sistemas de representación de datos o *dataset*. En consecuencia, se vislumbra un sinnúmero de aplicaciones en el área de la salud en sus diferentes niveles optimizado los procesos de diagnóstico temprano minimizando los riesgos asociados a una nueva pandemia a nivel global por causa del COVID-19 y/o cualquier otro virus.

Existen tecnologías disruptivas como la inteligencia artificial y la computación inteligente, manifiestas a través del aprendizaje máquina (*Machine Learning*) y el aprendizaje profundo (*Deep Learning*). Combinados con la ciencia de datos, el *Big Data* y la analítica de datos avanzada, entre otros que presentan diversas opciones de investigación y desarrollo, se puede ayudar a la detección temprana del COVID-19 a través de la búsqueda de características relevantes que permitan a la comunidad científica identificar factores bioquímicos, moleculares y celulares que faciliten la detección temprana del virus en sus diferentes estados de infección, incubación, propagación y tratamientos a emplear.

2. COVID-19

Los coronavirus (CoV) pertenecen a la subfamilia *Orthocoronavirinae* de la familia *Coronaviridae* en el orden *Nidovirales*, y esta subfamilia incluye α -coronavirus, β -coronavirus, γ -coronavirus y delta-coronavirus (Banerjee *et al.*, 2019). Entre los agentes causantes de las infecciones del tracto respiratorio humano se encuentran los coronavirus (CoV), que son virus de ARN de cadena positiva única envueltos, que pertenecen a la gran subfamilia *Coronavirinae* que infectan a aves y mamíferos (Raoult *et al.*, 2020). En cuanto al nuevo coronavirus-2 del SARS (SARS-CoV-2) Guan *et al.* (2020) señalan que parece altamente transmisible de un patógeno humano a humano, lo que causa un amplio espectro de manifestaciones clínicas en pacientes con COVID-19.

Sobre la transmisión del virus, Yang y Wang (2020, 2710) afirman que, en la revisión de 22 tipos de coronavirus, tanto el SARS-CoV, el MERS-CoV y coronavirus humanos endémicos pueden persistir en superficies inanimadas como metal, vidrio o plástico por hasta nueve días, proporcionando evidencias sólidas de la supervivencia ambiental del patógeno. A estas evidencias se suma la contaminación del agua por heces de personas infectadas, ampliando otra posible vía de transmisión de esta enfermedad. En cuanto a la inactivación de coronavirus por agentes desinfectantes en pruebas de suspensión se puede consultar a Kampf (2020) en la que se exponen evidencias al respecto.

Basada en estos hechos, la COVID-19 ha puesto en la palestra la fragilidad de la humanidad ante un virus que ha demostrado ser altamente resistente y peligroso, con un factor de propagación sin precedente alguno a escala global en tiempos modernos, exponiendo la lasitud del sector sanitario en atender una oleada creciente de infectados que asciende a millones y de víctimas fatales que suma cientos de miles. Aunque se están desarrollando técnicas de diagnóstico más rápidas (Pang, *et al.*, 2020) y terapias y medicamentos antivirales experimentales, quedan desafíos a superar como acelerar los procesos de fabricación y superar los problemas relacionados con la cadena de suministro y abastecimiento (Ledford, 2020).

Entre tanto, el diagnóstico temprano y tratamiento específico de esta enfermedad presentan una serie de dificultades que han impulsado a las comunidades académicas e investigativas a formular propuestas enmarcadas desde la óptica de las tecnologías disruptivas como la inteligencia artificial y tratamiento de datos masivos, que permitan encontrar soluciones alternativas sobre prevención, control, replicación y propagación del virus, que muchos científicos afirman que puede convertirse en endémico (Abdessater *et al.*, 2020; Kampf *et al.*, 2020).

Ante el panorama expuesto, la tecnología ha venido demostrar su valía en ayudar a encontrar y establecer determinados factores relacionados con la COVID-19, bien en establecer

su comportamiento molecular y celular, como en la manera de propagarse e infectar a un individuo, entre otros factores. Es por ello que diversos grupos de investigación, instituciones, empresas y gobiernos han planteado iniciativas en pro de buscar soluciones a corto plazo para contener la pandemia generada por el virus, al igual que encontrar un cura contra el mismo.

3. Inteligencia artificial

La inteligencia artificial o IA puede entenderse como una disciplina perteneciente a las ciencias de la computación, que plantea modelos computacionales de aprendizaje basado en redes neuronales biológicas humanas. En este sentido, se han planteado diversos modelos de IA, que gracias a los avances en la tecnología computacional ha permitido desarrollos de sistemas “inteligentes” que facilitan procesar mayor cantidad de datos en un tiempo menor, agilizando la toma de decisiones.

Un aspecto a mencionar acerca de la IA, es que abarca diversos campos como el reconocimiento de voz, procesamiento de lenguaje natural, visión por computador, robótica avanzada, captura de conocimiento, planificación y optimización, entre otros, en la que se busca que un sistema tenga la capacidad para sentir, razonar, participar y aprender. Como complemento a estos campos, se encuentran especialidades como el aprendizaje automático (*Machine Learning*) que involucra diferentes tipos de redes neuronales codificadas bajo el aprendizaje reforzado, el aprendizaje supervisado, el aprendizaje no supervisado y el aprendizaje semi-supervisado, que convergen en sistemas que trabajan con técnicas algorítmicas avanzadas para procesar diversos tipos de señales, optimizar modelos, procesar información masiva mediante *Big Data* y *Clusters*; todo ello mediante el uso algorítmicos de regresión, clasificación y *clustering*, entre otros.

Otra especialidad de la IA es el aprendizaje profundo (*Deep Learning*) para evaluar datos de tipo imagen, video y audio empleando redes neuronales convolucionales con sus diversas variantes (Tang *et al.*, 2020) y redes neuronales de memoria a corto y largo plazo. También permite la simulación de sistemas dinámicos mediante el uso del aprendizaje por refuerzo.

Con base en la dinámica comportamental del COVID-19, se requieren soluciones prontas para el monitoreo, detección y diagnóstico de las enfermedades generadas por su causa (Law, Leung & Xu, 2020), la IA plantea diversas opciones de *hardware* y *software* encaminadas para tal fin (Mei *et al.*, 2020). Bajo este panorama, ha venido tomando fuerza el desarrollo de *software* de código abierto, donde la inteligencia colectiva es el engranaje principal para obtener un programa de altas prestaciones, multipropósito en la mayoría de los casos.

Una particularidad de trabajar con la inteligencia colectiva radica en el hecho que está dirigida a investigadores interesados en contribuir a mejorar el sistema, perfeccionándolo hasta llegar a un producto funcional, fiable y cien por ciento operativo. Un ejemplo es el sistema *Covid-Net*, cuyo objetivo es la detección de la infección viral por COVID-19 mediante el examen de radiografías o tomografías de tórax (Wang & Wong, 2020), donde diversos estudios permiten identificar anomalías radiográficas en las imágenes con un alto grado de confiabilidad, tales como la neumonía viral y otros problemas respiratorios que pueden tener o no relación con el virus.

De igual manera, la IA se está empleando para descubrir compuestos que deriven en fármacos, no solo para lidiar con el coronavirus, sino para optimizar los procesos y procedimiento llevados en las industrias farmacéuticas y laboratorios, donde el uso de los *Dataset* es fundamental, ya que contienen ingentes colecciones de información actualizadas por lo general, expresadas mediante *Data Tables* y *Data Relations*, que permiten filtrar información masiva de manera más rápida. Con esto en mente, los costos de investigación y desarrollo se reducen ostensiblemente, viéndose reflejados en el valor final del medicamento para el paciente y sector salud en general. Las herramientas a emplear para este tipo de caso son los modelos generativos, que se sustentan en el aprendizaje automático y en el aprendizaje profundo, que permiten trabajar gran cantidad de datos relacionados con la dinámica molecular y propiedades fisicoquímicas de un fármaco, facilitando la identificación de miles de nuevas moléculas candidatas a ser sintetizadas y probadas.

3.1 Aprendizaje automático

El aprendizaje automático es una disciplina de la IA que se vale de algoritmos que permiten la identificación de patrones, efectuar predicciones, aprender de los datos y toma de decisiones. Para el caso del COVID-19, el aprendizaje automático se emplea para el diagnóstico e identificación de población que está en mayor riesgo de contagio. También se emplea para el desarrollo más rápido de medicamentos, incluyendo el estudio de reutilización de medicamentos que han sido probados para el tratamiento de otras enfermedades. Para ello, se construyen gráficos de conocimiento y se realizan análisis predictivos de interacción entre fármaco y proteínas virales (Zhou, Park, Choi & Han, 2018) e interactomas virus-huésped (Yang *et al.*, 2019), plegamiento predictivo de proteínas (Ivankov & Finkelstein, 2020), comprensión de la dinámica molecular y celular del virus, predicción y propagación de una enfermedad basado en patrones, e incluso predecir una próxima pandemia zoonótica.

Para realizar estudios de IA mediante aprendizaje automático (que incluye el aprendizaje profundo en algunos casos), se requieren de determinados algoritmos, como árboles de decisión,

regresión para análisis estadístico y predictivo, redes adversarias generativas, *Clustering* basado en instancias, bayesianos, redes neuronales, etc. Estos algoritmos se valen de la ciencia de datos en la que se ejecutan diversos cálculos matemáticos, donde la densidad de información es amplia, compleja y variada. Por ejemplo, encontrar moléculas antivirales (Ahuja, Reddy & Marques, 2020) que combatan la COVID-19 e identifiquen millones de anticuerpos para tratamiento de infecciones secundarias (Ciliberto & Cardone, 2020).

Otro tipo de aplicación del aprendizaje automático ronda entorno a la predicción de riesgos de infección, basado en características específicas de una persona, tales como edad, ubicación geográfica, nivel socioeconómico, hábitos sociales y de higiene, condiciones preexistentes e interacción humana, entre otros. Con estos datos se puede establecer un modelo predictivo sobre el riesgo que puede traer consigo un individuo o grupo de personas de contraer COVID-19 y factores asociados a desarrollar complicaciones (Jiang *et al.*, 2020) e incluso, predecir los resultados de un tratamiento. Con este tipo de proyecciones, literalmente se podría predecir si un paciente vive o muere.

El aprendizaje automático abre un sinnúmero de posibilidades de investigación en diversos campos clínicos, donde la COVID-19 ha sido el impulsor de ello. Esto involucra desde los escáneres faciales para identificación de síntomas como la fiebre, *wearables* para medición y detección de anomalías cardíacas o respiratorias, hasta *chatbots* que evalúan a un paciente cuando este menciona sus síntomas y, basado en las respuestas dadas, el sistema le indica si debe permanecer en casa, llamar al médico o ir al hospital.

La ventaja de emplear el aprendizaje automático frente a otras técnicas estándar que duran años es que el proceso de identificación puede realizarse en cuestión semanas, con una reducción de costos considerable, sumado a una probabilidad muy alta de éxito. Por ejemplo, Smith y Smith (2020, 2) afirman que, el diseño futuro de fármacos antivirales SARS-CoV-2, ya está a cargo de un equipo europeo que utiliza el supercomputador de IBM equipado con el sistema AI SUMMIT para ser utilizado en tratamientos por COVID-19.

3.2 Aprendizaje profundo

El aprendizaje profundo es un subcampo del aprendizaje máquina, que busca clasificar datos mediante algoritmos correlacionales. Se basa en ciertas arquitecturas de redes neuronales, que le permiten jerarquizar la información (visual, auditiva y escrita) mediante una segmentación de patrones categorizados por niveles. Bajo este criterio, el aprendizaje se efectúa por etapas, de manera equivalente a lo que sucede en un humano. Es decir, se parte de datos básicos y conforme se van escalando niveles más complejos de los mismos, se va aprendiendo.

Con respecto a la pandemia del COVID-19 el sistema sanitario a nivel mundial demostró no dar abasto para realizar pruebas diagnósticas a corto plazo, sumado a problemas económicos, logísticos, de infraestructura tecnológica y falta de personal hospitalario. La IA está ayudado a minimizar estos problemas a través del uso de técnicas de aprendizaje profundo, mediante el reconocimiento de imágenes para pruebas de radiodiagnóstico que, a diferencia de las pruebas clínicas estándar, arroja resultados en pocos minutos, y en ellos se infiere si los pulmones de un paciente están o no enfermos por neumonía asociada específicamente con la COVID-19 (Öner, 2020). Esta especificidad radica en el hecho que mediante procesos de extracción de características morfológicas sobre el objeto de estudio (Márquez, 2019), se validan con respecto a patrones propios de otras enfermedades, estableciéndose factores de similitud, bien en cuanto a la geometría del tejido, lesiones en el mismo y variaciones de intensidad, entre otros.

El aprendizaje profundo se caracteriza porque emplea información gráfica para efectuar su análisis. En el caso del COVID-19 se han desarrollado diferentes tipos de algoritmos tendientes a la detección temprana de problemas pulmonares, bien a través de imágenes de rayos X (Sánchez, Torres & Martínez, 2020), tomografías o ultrasonido (Fraile, 2020). Las investigaciones en curso basan su desarrollo en la ciencia de datos, debido a que esta proporciona herramientas de análisis que permiten plantear soluciones plausibles bien para lidiar con futuras pandemias como con tratamientos actuales de frente al COVID-19. Por ejemplo, en todo el mundo han surgido comunidades que han creado plataformas que trabajan bajo el modelo de inteligencia colectiva, donde desarrolladores, científicos de datos, investigadores, médicos, entre otros voluntarios, formulan proyectos relacionados con la COVID-19. Una plataforma que centraliza esta información es *Helpwithcovid*, que cuenta con bibliotecas como el *Covid Healthcare Coalition*, que permite a cientos de miles de investigadores tener acceso a información clave para trabajar con aprendizaje profundo.

Es importante mencionar que la República de Corea proporciona acceso a información anónima de pacientes con COVID-19, incluyendo el historial médico de cinco años de cada paciente, con el fin de que al realizar un estudio de trazabilidad mediante aprendizaje profundo se cuente con una alta cantidad de información útil a estos propósitos. Para proteger la privacidad del paciente, quienes ejecutan el código son un grupo especializado, y luego una vez procesado devuelven los resultados a los investigadores. Este estudio es de especial interés porque permite realizar monitoreos epidemiológicos, caracterizando aquella población que se enferma, pero no ingresa a un hospital, o de aquellos que nunca muestran síntomas. Bajo estas circunstancias, es muy difícil rastrear cómo se propaga la enfermedad y cuántas personas realmente mata. Así, con la investigación empleando IA, se logra obtener información predictiva que ayuda a las autoridades sanitarias a tomar cartas sobre el asunto. Por ejemplo, mediante un modelo de

ramificación para estimar cuántas personas han sido infectadas, se analiza ADN viral extraído de cada paciente conocido, luego, el modelo utiliza la tasa de mutación para interpolar a cuántas otras personas pasaron el virus en el camino (Li & Ayscue, 2020). Otro tipo de investigación de IA utilizando aprendizaje profundo es mediante el uso del lenguaje natural, que permite crear nuevos medicamentos a partir de la búsqueda de compuestos existentes (Freedman, 2019). Para ello se analiza cómo el virus entrega su material genético a una célula infectada, con el fin de crear un modelo predictivo acerca de la estructura proteínica que permita *a posteriori* buscar aquellas que puedan inhibir este proceso (Réda, Kaufmann & Delahaye, 2020).

Otros tipos de desarrollo de aprendizaje profundo, basado en los sistemas de reconocimiento de rostro empleados comúnmente en seguridad física, han sido modificados para detectar si la comunidad está cumpliendo la distancia social reglamentaria. Este *software* emplea cámaras de video estándar o aquellas dispuestas en una ciudad para videovigilancia, permitiendo monitorear el flujo peatonal en zonas críticas, realizando un reconocimiento sobre la distancia mínima, indicando una alerta a las autoridades si alguien no cumple con la norma. En algunos casos este sistema se combina con cámaras de medición térmica con el fin de establecer si alguna persona tiene fiebre y emitir una alerta para que sea aislada de inmediato por personal sanitario.

En el campo clínico, el aprendizaje profundo al valerse de imágenes demuestra resultados sólidos y permite identificar diversos tipos de enfermedades como el cáncer, glaucoma y neumonía; esta última asociada con la COVID-19 bajo ciertos patrones de reconocimiento celular, con un nivel de falsos positivos muy bajo. Por ejemplo, existen aplicaciones presentes en el mercado que emplean imágenes de rayos X para la detección temprana del Coronavirus, en la que se identifican anomalías pulmonares asociadas con la infección viral (Ozturk *et al.*, 2020) basado en la tasa de virulencia y transferencia detectadas, que con otras técnicas se tardaría más tiempo en precisar. En general, al crear un sistema automático combinando modelos de aprendizaje profundo se facilita la detección de diversas anomalías pulmonares y adicionalmente se puede calcular el riesgo de enfermedad mediante modelos probabilísticos a partir de una radiografía. Una particularidad de las tecnologías asociadas con la IA, es que precisan de algoritmos cada vez más potentes que faciliten el procesamiento con una mayor cantidad de información; lo que demanda una potencia computacional con servicios centralizados en la nube. Basado en este requerimiento, existen nuevos algoritmos que optimizan ciertos modelos relacionados con el aprendizaje profundo que se encuentran en chips diseñados para tal fin. Por ejemplo, los chips de Intel (Nervana (NNP) y el NNP-I), la familia de chips de IA Nvidia, chips de *Google* y *Amazon*, que prometen la ejecución de la IA con menos energía sin sacrificar su eficiencia, incluso poder realizar inferencias. A estos desarrollos de los conoce como IA diminuta, cuya particularidad subyace en el hecho que se pueden realizar diversos tipos de análisis de imágenes clínicas directamente desde

un dispositivo móvil. Esto permite realizar estudios *in situ* sobre algún tipo de dinámica histológica relacionada con la COVID-19 o cualquier otro agente infeccioso mediante el uso de patrones imagenológicos.

Para el caso de análisis de datos robustos se requiere de una gran infraestructura tecnológica, como superordenadores dedicados específicamente para trabajar con IA, donde empresas como *Microsoft* e *IBM* están apostando con el objetivo de encontrar tratamientos que permitan combatir la COVID-19 en un tiempo relativamente corto. La particularidad de trabajar con este tipo de tecnología es que puede ejecutarse desde diversos lugares del mundo, facilitando el acceso al código abierto a grupos de investigadores que trabajan en diversas áreas de la IA, el *Big Data* y ciencia de datos.

4. *Big Data*

El *Big Data* se entiende como el conjunto de procedimientos computacionales aplicados para analizar gran cantidad de datos con el fin de extraer información que presente ciertos patrones, relaciones y asociaciones relevantes para una organización. Debido a que progresivamente se crea más información en diferentes campos del saber, la tendencia para los próximos años es la integración masiva y recurrente del *Big Data* con la Internet de las cosas en un entorno urbano e industrial, donde no sólo se monitorearán variables como la calidad del aire, temperatura y humedad relativas, sino también la integración con sistemas biométricos, cámaras de monitoreo y vigilancia que buscan ayudar al sector sanitario mediante el estudio del estado de salud de las personas *in situ* (sanidad preventiva) empleando sensores y cámaras térmicas, al igual que la identificación temprana de posibles focos infecciosos mediante el uso de biosensores y nanosensores dispuestos estratégicamente en la ciudad.

Cabe señalar que el *Big Data* se caracteriza por tres adjetivos propios para el desarrollo de un proyecto de investigación, que son “Volumen”, “Variedad” y “Velocidad”, conocidos como las 3 V del *Big Data*; otros autores (Ishwarappa y Anuradha, 2015) adicionan la “Veracidad” y “Valor”. El volumen en el campo del *Big Data* demanda grandes recursos de procesamiento y almacenamiento de información, que están representados en la “Variedad” de los datos, que pueden ser de tipo estructurados y no estructurados. Con respecto a la “Velocidad”, hace referencia a la cantidad de datos que se generan periódicamente y requieren de una infraestructura tecnológica escalable que permita su disponibilidad y acceso en cualquier momento. Sobre la “Veracidad” y “Valor”, es imprescindible que los datos almacenados sean veraces, de lo contrario se estarían dilapidando recursos computacionales valiosos en

información poco confiable o inservible, que derivan en resultados y toma de decisiones incorrectas. Sobre el “valor”, se entiende en el sentido de extraer información relevante para definir estrategias y toma de decisiones. Para el caso particular del COVID-19, se genera gran cantidad de datos que, al emplear la IA para analizarlos permite diferenciar familias, tratamientos, riesgos, etc., que confluye a disminuir costos en el diagnóstico y tratamiento de un paciente, salvando en el proceso miles de vidas. Por consiguiente, el emplear de manera asertiva los cinco adjetivos mencionados deben garantizar información veraz y confiable para poderlas implementar en sistemas de IA de aprendizaje profundo, aprendizaje máquina o ambos.

Con base en lo anterior, el estudio del COVID-19 empleando *Big Data* puede valerse de la analítica retrospectiva y descriptiva avanzadas (Mohamed *et al.*, 2019) expresas en la inteligencia de negocios (*Business intelligence*) (Chahal, Jyoti y Wirtz, 2019); ya que esta permite focalizar el estudio mediante indicadores y tendencias en el tiempo, lo que incluye predicciones a futuro. Bajo estos criterios, entra a formar parte la ciencia de datos empleando técnicas estadísticas y matemáticas caracterizadas en las variables de estudio que permiten ampliar las técnicas y modelos representados como *clusters* (Mohebi *et al.*, 2016) a través de patrones o correlaciones de datos, que a la vez pueden ser integrados con la IA para mejorar los resultados.

Basado en estas características, en un estudio de bases de datos estructuradas y no estructuradas, se buscan correlaciones recurrentes por medio del cruce de información hasta llegar a una solución plausible, que por lo general permite anticiparse a un evento, como caso particular la COVID-19. Es decir, puede perfilarse en diversos contextos como la diseminación del virus en una determinada zona bajo parámetros climatológicos, factores de densidad demográfica, patrones de movilidad, características fenotípicas del virus con respecto a sus otras familias y medioambiente, detección de anomalías a nivel celular y/o molecular, entre muchos otros aspectos. Para este tipo de desarrollo, el *Big Data* se vale de la IA a través de diversos algoritmos aplicados bien en el aprendizaje profundo como en el aprendizaje automático mencionados anteriormente.

5. Aspectos bioéticos a considerar

Aunque el trabajo de la IA con el *Big Data* y demás tecnologías disruptivas que les acompañan están haciendo su aporte significativo al sector sanitario, no puede dejarse de lado que este tipo de desarrollo tecnológico deja una brecha en materia bioética y seguridad de la información, porque se manipulan datos de pacientes y víctimas del COVID-19 de forma masiva, cuyo historial clínico y trazabilidad sobre seguimiento de personas y comunidades en general subyace en

repositorios tanto públicos como privados y, aunque el acceso a estos es restringido en algunos casos, incluso con datos almacenados y compartidos de forma cifrada, se prestaría para que terceros la empleen en estudios discriminatorios, donde el anonimato no existe, tal como señalan Gené, Gallo y de Lecuona (2018, p. 4), ya que con la anonimización de los datos personales no se garantiza suficientemente la privacidad, específicamente porque la tecnología de datos masivos permite la reidentificación de las personas.

Lecuona y Villalobos (2018, p. 2) afirman que, al asignar a una persona o grupo características particulares, un individuo se convierte en un componente de un colectivo que genera preocupaciones sobre la discriminación consciente e inconsciente como resultado del uso de grandes datos en la toma de decisiones. La discriminación bajo este enfoque de las tecnologías disruptivas es un riesgo a tomar en cuenta, máxime cuando la toma de decisiones es cada vez más automatizada.

Por consiguiente, lo aconsejable es que tanto científicos de datos, ingenieros, Estado y sector salud, entre otros involucrados, evalúen el riesgo-beneficio del uso de estas tecnologías, no restringiéndolas, pero si creando los mecanismos legales, técnicos y tecnológicos que garanticen que no se vulnere ningún derecho, y un primer paso en esta dirección es hacer que la información clínica no se externalice sin ningún tipo de restricción (De Lecuona, 2018).

Con la continua incorporación de las tecnologías disruptivas en el campo de la salud, las normativas en cuanto a la preservación de los derechos del paciente deberán ir cambiando y adaptándose a pesar de la incertidumbre que trae consigo, que como señala López (2019), no solo subestimamos los efectos de la tecnología a corto plazo, sino que ni siquiera somos capaces de pronosticar, predecir o intuir sus efectos a largo plazo.

6. Conclusiones

Con el uso de tecnologías disruptivas como la inteligencia artificial y el *Big Data* se espera estar mejor preparados para una próxima pandemia, incluso prevenirla. Tecnologías 4.0 como el internet de las cosas, la computación inteligente y la computación en la nube aportarán lo suyo en cuanto al monitoreo permanente de las ciudades en busca de anomalías biológicas y químicas que impliquen algún riesgo para la sociedad o el medio ambiente. Este tipo de desarrollos pueden ser personalizados, pues al aplicar algoritmos de aprendizaje predictivo se minimizan los riesgos al formular tratamientos que pueden establecer si un paciente los tolera o no.

Lo que se puede rescatar de la pandemia del COVID-19 es que ha impulsado desarrollos tecnológicos sin precedentes en materia de inteligencia artificial en sus diferentes áreas del saber,

al igual que las ciencias de datos masivos. Bajo este panorama, el sector salud tendrá que incorporar rápidamente estos recursos a su sistema de análisis y diagnóstico, no solo de enfermedades infecciosas sino de cualquier otra, por lo que se espera mejorar el servicio prestado a un paciente o comunidad y preparar a la sociedad ante cualquier eventualidad de pandemia a futuro. Además, este tipo de desarrollos ayudan a los centros de salud a reducir costos operativos de diversa índole, donde el tiempo de diagnóstico juega un papel fundamental para detener un potencial foco pandémico.

La IA en conjunto con el *Big Data* han demostrado ser herramientas fundamentales para ayudar al sector salud a detectar y controlar este virus con cierto margen de éxito, permitiendo procesar grandes cantidades de datos estructurados y no estructurados con alto grado de complejidad, que al ser combinados con algoritmos propios de la IA permiten realizar predicciones basadas en patrones históricos y bucles de retroalimentación, entre otros. Lo importante de esta sinergia, es que ayuda de manera más eficaz a la atención médica, incluso una vez terminada la crisis. También, con el aprendizaje que se está gestando permanentemente ya hay desarrollos de algoritmos predictivos que permiten identificar poblaciones que son o serán más propensas a ser infectadas por la COVID-19, incluso determinar en términos probabilísticos quienes podrán padecer complicaciones graves con base en parámetros como edad, género, historial clínico, masa corporal, entre otros. Con este tipo de desarrollo, estos algoritmos pueden extenderse a ser aplicados a otro tipo de enfermedades, contribuyendo de esta manera a mejorar el servicio sanitario.

La IA y el radiodiagnóstico están jugando un papel importante en la detección del COVID-19 con un porcentaje superior al 90%, lo que puede incrementarse cuando se entrena el sistema con mayor cantidad de datos, por lo que el *Big Data* en conjunto con otras disciplinas analíticas son un factor clave para llevar a feliz término un estudio. Cabe aclarar que la IA no sustituye al profesional sanitario, por el contrario, es un complemento a su quehacer médico, ayudándole a mejorar la precisión del diagnóstico en un tiempo menor y tomar decisiones mucho más rápido aligerando con ello su carga de trabajo.

Bibliografía

- ◆ Abdessater, M., Rouprêt, M., Misrai, V., Matillon, X., Tellier, B. G., Freton, L., ... Pradere, B. (2020) COVID19 pandemic impact on anxiety of French urologists in training: outcomes from a national survey. *Progrès en Urologie*. 1-12. <http://doi.org/10.1016/j.purol.2020.04.015>.
- ◆ Ahuja, A. S., Reddy, V. P., & Marques, O. (2020). Artificial Intelligence and COVID-19: A Multidisciplinary Approach. *Integrative Medicine Research* 100434. Advance online publication. <https://doi.org/10.1016/j.imr.2020.100434>.
- ◆ Banerjee, A., Kulcsar, K., Misra, V., Frieman, M., & Mossman, K. (2019). *Bats and Coronaviruses*. *Viruses*, 11(1), 41. <http://doi.org/10.3390/v11010041>.
- ◆ Chahal, H., Jyoti, J. & Wirtz, J. (2019). Business Analytics: Concepts and Applications. In Understanding the Role of Business Analytics; Springer: London, UK, 1–8.
- ◆ Ciliberto, G., & Cardone, L. (2020). Boosting the arsenal against COVID-19 through computational drug repurposing. *Drug discovery today*, S1359-6446(20)30152-5. Advance online publication. <https://doi.org/10.1016/j.drudis.2020.04.005>.
- ◆ De Lecuona, I. (2018). Evaluación de los aspectos metodológicos, éticos, legales y sociales de proyectos de investigación en salud con datos masivos (big data). *Gaceta Sanitaria* 1-3. <http://doi.org/10.1016/j.gaceta.2018.02.007>.
- ◆ De Lecuona, I., & Villalobos, Q. M. (2018). European perspectives on big data applied to health: The case of biobanks and human databases. *Developing World Bioethics*. 1-8. <http://doi.org/10.1111/dewb.12208>.
- ◆ Fraile, G. V. *et al.* (2020). Revisión narrativa de la ecografía en el manejo del paciente crítico con infección por SARS-CoV-2 (COVID-19): aplicaciones clínicas en Medicina Intensiva-una revisión narrativa. *Medicina Intensiva*. 1-15. <http://doi.org/10.1016/j.medin.2020.04.016>.
- ◆ Freedman, D. H. (2019). Hunting for New Drugs with AI. *Nature*, 576, 7787, S49–S53. doi:10.1038/d41586-019-03846-0.
- ◆ Gené, B. J., Gallo, de P. P., & de Lecuona, I. (2018). Big data y seguridad de la información. *Atención Primaria*, 50, 1, 3–5. <http://doi.org/10.1016/j.aprim.2017.10.004>.
- ◆ Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei CL, Hui DSC, Du B, Li LJ, Zeng G, *et al.* (2020) Características clínicas de la enfermedad por coronavirus 2019 en China. *The New England Journal of Medicine*. 1-13. <http://doi.org/10.1056/NEJMoa2002032>.

- ◆ Ishwarappa K. & Anuradha, J. (2015). A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology. *Procedia Computer Science* 48, 319-324. [10.1016/j.procs.2015.04.188](https://doi.org/10.1016/j.procs.2015.04.188).
- ◆ Ivankov, D. I. & Finkelstein, A. V. (2020). Solution of Levinthal's Paradox and a Physical Theory of Protein Folding Times. *Biomolecules*, 10, 2, 250. doi:10.3390/biom10020250.
- ◆ Jiang, X., Coffe, M., Bari, A., Wang, J., Jiang, X., Huang, J., *et al.* (2020). Towards an Artificial Intelligence Framework for Data-Driven Prediction of Coronavirus Clinical Severity. *Computers, Materials & Continua (CMC)*. 63, 1, 537-551. <http://doi.org/10.32604/cmc.2020.010691>.
- ◆ Kampf, G. (2020) Potential role of inanimate surfaces for the spread of coronaviruses and their inactivation with disinfectant agents. *Elsevier*. 2, 2, 1-2. <https://doi.org/10.1016/j.infpip.2020.100044>.
- ◆ Kampf, G., Todt, D., Pfaender, S., & Steinmann, E. (2020). Persistence of coronaviruses on inanimate surfaces and its inactivation with biocidal agents. *Journal of Hospital Infection*. 1-13. <http://doi.org/10.1016/j.jhin.2020.01.022>.
- ◆ Law, S., Leung, A. W., & Xu, C. (2020). Severe Acute Respiratory Syndrome (SARS) and Coronavirus disease-2019 (COVID-19): From Causes to Preventions in Hong Kong. *International Journal of Infectious Diseases*. <http://doi.org/10.1016/j.ijid.2020.03.059>.
- ◆ Li, L. & Ayscue, P. (2020). Using viral genomics to estimate undetected infections and extent of superspreading events for COVID-19. *medRxiv*, 1-17. <https://doi.org/10.1101/2020.05.05.20092098>.
- ◆ López, B. M. (2019). Las narrativas de la inteligencia artificial. *Revista de Bioética y Derecho*, 46, 5-28. <https://doi.org/10.1344/rbd2019.0.27280>.
- ◆ Márquez, D. J. E. (2019). Early identification of non-melanoma cancer and actinal keratosis through artificial vision. *Revista Compusoft. An international journal of advanced computer technology*, 8(3), 3079-3087. <http://dx.doi.org/10.6084/ijact.v8i3.786>.
- ◆ Mei, X., Lee, H., Diao, K. *et al.* (2020) Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. *Nat Med*. 1-11. <https://doi.org/10.1038/s41591-020-0931-3>.
- ◆ Mohamed, A., Nahafabadi, M.K., Wah, Y.B., Zaman, E.A.K. & Maskat, R. (2019). The state of the art and taxonomy of big data analytics: View from the new big data framework. *Artif. Intell. Rev.* 1-49. <https://doi.org/10.1007/s10462-019-09685-9>.

- ◆ Mohebi, A.; Aghabozorgi, S.; Wah, T.Y.; Herawan, T.; Yayapour, R. (2016). Iterative big data clustering algorithms: A review. *Softw. Pract. Exp.* 46, 107–129. <https://doi.org/10.1002/spe.2341>.
- ◆ Öner, O. (2020). Coronavirus Disease 2019 (COVID-19): Diagnosis and Management (Narrative Review). *Erciyes Med J.* 42(3): 1-6.
- ◆ Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., & Rajendra, A. U. (2020). Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in Biology and Medicine*, 103792. Advance online publication. <https://doi.org/10.1016/j.combiomed.2020.103792>.
- ◆ Pang, J, Wang MX, Ang IYH, Tan SHX, Lewis RF, Chen JI-P, *et al.* (2020) Potential Rapid Diagnostics, Vaccine and Therapeutics for 2019 Novel Coronavirus (2019-nCoV). A Systematic Review. *J Clin Med.* 9, 3,623.
- ◆ Raoult, D., Zumla, A., Locatelli, F., Ippolito, G., & Kroemer, G. (2020). Coronavirus infections: Epidemiological, clinical and immunological features and hypotheses. *Cell Stress*, 1-10. <https://doi.org/10.15698/cst2020.04.216>.
- ◆ Réda, C. Kaufmann, E. & Delahaye, D. A. (2020). Machine learning applications in drug development. *Computational and Structural Biotechnology Journal*, 18, 241-252. <https://doi.org/10.1016/j.csbj.2019.12.006>.
- ◆ Sánchez, O. R., Torres, N. J., & Martínez, S. G. (2020). Radiological findings for diagnosis of SARS-CoV-2 pneumonia (COVID-19). La radiología en el diagnóstico de la neumonía por SARS-CoV-2 (COVID-19). *Medicina clínica*, S0025-7753(20) 30185-8. Advance online publication. <https://doi.org/10.1016/j.medcli.2020.03.004>.
- ◆ Smith, M. & Smith, J. C. (2020). Repurposing Therapeutics for COVID-19: Supercomputer-Based Docking to the SARS-CoV-2 Viral Spike Protein and Viral Spike Protein-Human ACE2 Interface. *ChemRxiv*. 1-28. <https://doi.org/10.26434/chemrxiv.11871402.v4>.
- ◆ Sohrabi, C, Alsafi, Z, O'Neill, N, *et al.* (2020). World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *Int J Surg.* 76:71-76. <http://doi.org/10.1016/j.ijsu.2020.02.034>.
- ◆ Tang, Y., Tang, Y., Peng, Y. *et al.* (2020). Automated abnormality classification of chest radiographs using deep convolutional neural networks. *npj Digit. Med.* 3, 70. <https://doi.org/10.1038/s41746-020-0273-z>.
- ◆ Wang, L., y Wong, A. (2020). COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images, 1-12.

- ◆ Wang, Y. Zhang, D., Du, G., Zhao, J., Jin, Y., Fu, S. *et al.* (2020) Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *The Lancet*, 1-10 [https://doi.org/10.1016/S0140-6736\(20\)31022-9](https://doi.org/10.1016/S0140-6736(20)31022-9).
- ◆ Yang, C. y Wang, J. (2020) A mathematical model for the novel coronavirus epidemic in Wuhan, China. *Mathematical Biosciences and Engineering*, 17, 3, 2708-2724. <http://doi.org/10.3934/mbe.2020148>.
- ◆ Yang, S., Fu, C., Lian, X., Dong, X., & Zhang, Z. (2019). Understanding Human-Virus Protein-Protein Interactions Using a Human Protein Complex-Based Analysis Framework. *mSystems*, 4(2), e00303-18. <https://doi.org/10.1128/mSystems.00303-18>.
- ◆ Zhou, X., Park, B., Choi, D., & Han, K. (2018). A generalized approach to predicting protein-protein interactions between virus and host. *BMC Genomics*, 19(S6). <http://doi.org/10.1186/s12864-018-4924-2>.

Fecha de recepción: 8 de junio de 2020

Fecha de aceptación: 29 de junio de 2020