



**Universidad de Valladolid**  
**Grado en Enfermería**  
**Facultad de Enfermería de Valladolid**

**UVa**

**Curso 2019-2020**

**Trabajo de Fin de Grado**

**USOS Y PERSPECTIVAS DE FUTURO  
DEL BIG DATA EN LA ASISTENCIA  
SANITARIA**

**Alicia Insua Alijas**

**Tutor/a: Antonio Orduña**

**Cotutor: Carlos Durántez Fernández**

## RESUMEN

La presente revisión expone el concepto Big Data y las posibilidades que ofrece dentro del ámbito sanitario en una época donde la tecnología está en auge. La implantación generalizada de Registros Sanitarios Electrónicos incrementa de forma colosal la cantidad de información obtenida, iniciando el camino de la Inteligencia Artificial para la investigación de nuevos datos explotables que acrecientan su valor. Se incide en sus aplicaciones, así como su relación con la enfermería. Se presentan varios estudios a favor del beneficio de su utilidad dentro de las Unidades de Cuidados Intensivos, y su función en el marco de la prevención de enfermedades cardiovasculares. Se pretende mostrar una perspectiva realista acerca del estado actual de su desarrollo, aludiendo a proyectos iniciados en España, e incluyendo otros internacionales, así como una visión futura dentro del entorno sanitario. Valorando tanto las ventajas, como los problemas técnicos, legales, y éticos que plantea utilizar una herramienta de esta magnitud.

**Palabras clave:** Big Data, Registros Sanitarios Electrónicos, Inteligencia Artificial, Aprendizaje Automático, Unidades de Cuidados Intensivos, cardiología.

## ABSTRACT

This review presents the Big Data concept and the possibilities it offers within the healthcare environment at a time when technology is booming. The generalised implementation of Electronic Health Records increases colossally the amount of information obtained, initiating the path of Artificial Intelligence for the investigation of new exploitable data that increase its value. Its applications are emphasized, as well as its relationship with nursing. Several studies are presented in favour of the benefit of its utility within Intensive Care Units, and its function in the framework of the prevention of cardiovascular diseases. The aim is to show a realistic perspective on the current state of their development, referring to projects initiated in Spain, and including other international ones, as

well as a future vision within the health environment. It values both the advantages and the technical, legal and ethical problems of entailing a tool of this magnitude.

**Keywords:** Big Data, Electronic Health Records, Artificial Intelligence, Machine Learning, Intensive Care Units, cardiology.

# ÍNDICE

<b>I.</b>	<b>INTRODUCCIÓN</b> .....	<b>1</b>
A.	BIG DATA .....	1
B.	LAS “V” DEL BIG DATA.....	2
C.	JUSTIFICACIÓN .....	4
<b>II.</b>	<b>OBJETIVOS</b> .....	<b>5</b>
A.	OBJETIVOS GENERALES .....	5
B.	OBJETIVOS ESPECÍFICOS .....	5
<b>III.</b>	<b>METODOLOGÍA</b> .....	<b>6</b>
A.	ESTRATEGIAS DE BÚSQUEDA .....	6
1.	<i>Criterios de inclusión y exclusión</i> .....	7
2.	<i>Extracción de los datos:</i> .....	8
B.	ANÁLISIS DE LOS DATOS .....	8
<b>IV.</b>	<b>RESULTADOS</b> .....	<b>9</b>
A.	BIG DATA Y ATENCIÓN SANITARIA .....	9
1.	<i>Registros Sanitarios Electrónicos (EHR)</i> .....	9
2.	<i>Aplicaciones del Big Data</i> .....	10
3.	<i>Enfermería y Big Data</i> .....	11
B.	BIG DATA EN UNIDADES DE CUIDADOS INTENSIVOS .....	12
1.	<i>Iniciativas en España</i> .....	14
C.	BIG DATA EN CARDIOLOGÍA .....	16
D.	BIG DATA Y ACTUALIDAD EN LA ATENCIÓN SANITARIA .....	18
<b>V.</b>	<b>DISCUSIÓN</b> .....	<b>22</b>
A.	PERSPECTIVAS DE FUTURO.....	22
B.	DIFICULTADES TÉCNICAS .....	23
C.	INCONVENIENTES ÉTICOS .....	27
<b>VI.</b>	<b>CONCLUSIONES</b> .....	<b>28</b>
<b>VII.</b>	<b>BIBLIOGRAFÍA</b> .....	<b>29</b>

## I. INTRODUCCIÓN

En la actualidad, los análisis de grandes datos, cobran gran valor a la hora de predecir la salud en la sociedad, razón por la cual, cada vez son más las investigaciones que se realizan acerca de la extracción real de conocimientos mediante algoritmos de aprendizaje automático, utilizando las tecnologías de la información que facilitan un procesamiento de los datos más eficiente, a objeto de favorecer la toma de decisiones, y, fomentar a partir de parámetros de salud el estado futuro de la misma, permitiendo un diagnóstico más preciso y multimodal, pues tales avances posibilitan una visión cada vez más detallada de los procesos de las enfermedades, además de analizar la carga económica de la atención sanitaria <sup>(1, 2)</sup>.

### A. Big Data

A lo largo de los últimos años, el término Big Data ha estado cada vez más presente, siendo más estudiado y considerado. Desde sus inicios, se ha intentado aclarar la complejidad de este término, pues eran diversas las interpretaciones, centrándose algunas en lo que es, mientras otras trataban de responder a lo que hace. Actualmente, puede definirse como conjuntos de datos extremadamente grandes y complejos que no pueden ser almacenados, buscados, analizados, o interpretados, mediante los métodos tradicionales de procesamiento de datos <sup>(3)</sup>.

La proliferación de dispositivos digitales en las últimas décadas favorece el aumento exponencial de datos, y, en consecuencia, los avances en este campo, viéndose impulsado el desarrollo de innovadoras técnicas y tecnologías de procesamiento de la información, que posibilitan disponer de inteligencia en tiempo real a partir de grandes volúmenes de datos perecederos <sup>(3)</sup>.

Entre sus características, se encuentran las denominadas “Vs” del Big Data, manteniendo en común coherencia y relación entre cada una de ellas. Se describen tres originales y principales, (volumen, variedad y velocidad), y en adición, dos dimensiones agregadas posteriormente por diversos autores,

(veracidad y valor), que, a pesar de no ser medidas de magnitud, son aspectos considerados intrínsecos al Big Data.

#### B. Las “V” del Big Data

*Volumen:* concierne a la magnitud de información. Se prevé que los datos en medicina y sanidad aumenten considerablemente en los años venideros, almacenados y medidos en terabytes (1024 GB), petabytes (1.048.576 GB) e incluso llegar a yottabytes (931.322.574.615.480 GB) <sup>(4)</sup>. Las definiciones de los grandes volúmenes son relativas y varían con el tiempo y el tipo de datos <sup>(3)</sup>.

*Variedad:* hace referencia a la heterogeneidad de un conjunto de datos. Los avances tecnológicos y los sistemas Big Data permiten la integración de diversos tipos de datos estructurados, semiestructurados y no estructurados <sup>(3)</sup>.

*Velocidad:* es el tiempo en el que se originan y recogen los datos, y en el que deben ser analizados <sup>(3)</sup>. La velocidad de los grandes datos, es una oportunidad para el análisis en tiempo real, con bucles de retroalimentación inteligente para optimizar los recursos y el funcionamiento de los sistemas de salud <sup>(5)</sup>.

*Veracidad:* representa la carencia de fiabilidad inherente a algunas fuentes de datos. La necesidad de manejar datos imprecisos que se aborda utilizando instrumentos y análisis para la gestión y la extracción de datos inciertos <sup>(3)</sup>.

*Valor:* potencial de la información recibida para impulsar la toma de decisiones basada en la evidencia, resultado de analizar grandes volúmenes de datos con un valor inicial bajo en relación con su volumen <sup>(3)</sup>.

Actualmente, se reúnen constantemente grandes datos durante la atención sanitaria en forma de Registros Sanitarios Electrónicos (EHR) “Electronic Health Records”, estos son grandes debido a su heterogeneidad y complejidad, y la cantidad de información que abarca, se recoge con el objetivo de permitir una integración eficaz y fiable de la informática biomédica. Investigaciones relacionadas al respecto, enuncian ser una oportunidad para el descubrimiento de medicamentos, la innovación de tratamientos, la medicina personalizada, el cuidado óptimo del paciente, la supervisión de calidad, mejora en la gestión económica, etc., en vistas a optimizar los resultados, por ejemplo, reduciendo

el número de reingresos hospitalarios y las visitas a los servicios de urgencias.  
(6-8).

Diferentes análisis también valoran la repercusión del uso del big data en el entorno de los cuidados intensivos, pues brinda la posibilidad de almacenar una ingente cantidad de datos clínicos en formato electrónico durante el ingreso de los pacientes, pudiendo ser utilizados posteriormente para obtener nuevas respuestas a preguntas clínicas <sup>(9)</sup>, entre otros beneficios enunciados en los resultados del presente documento, así como también se muestran iniciativas en desarrollo dentro de España.

La detección temprana de una enfermedad puede llegar a ser esencial, diversos estudios se centran en la aplicación del big data y las enfermedades cardiovasculares, por ser una de las causas de muerte más frecuentes en todo el mundo. La disponibilidad de estos recursos de inteligencia artificial, como la comunicación digital, y los sistemas de apoyo a la toma de decisiones, ofrecen beneficios económicos, además de mejorar la calidad, y, acelerar el tratamiento ambulatorio y hospitalario <sup>(2, 10)</sup>.

El empleo del Big Data requiere un trabajo multidisciplinar y colaborativo con otros profesionales, como, analistas de datos, estadísticos, epidemiólogos, especialistas en derecho aplicado a grandes bases de datos, ..., permitiendo ampliar los horizontes en aspectos, como, el control de calidad de las labores sanitarias cotidianas, la comparación de resultados entre unidades, o la elaboración de nuevas líneas de investigación clínica <sup>(7, 9)</sup>.

Conseguir efectuar lo anterior citado, implica superar una serie de retos tecnológicos y organizativos fundamentales, como, la interoperabilidad y la confusión en la propiedad de los datos, además de existentes cuestiones de índole legal, entre ellas, la privacidad y el consentimiento de los pacientes, precisando ser abordados para poder hacer uso de esta potente herramienta. Esto dificulta la disponibilidad de pruebas de sus beneficios prácticos en la asistencia sanitaria, suponiendo un gran desafío <sup>(8)</sup>.

### C. Justificación

El adentramiento a la era tecnológica, y el aumento en consecuencia de la digitalización de los datos, está abriendo las puertas a una reforma en la atención de la salud y prestación de los cuidados que desempeña enfermería.

Big Data puede ser clave para explorar y avanzar en objetivos compartidos con diversos campos, sobre temas relacionados con los registros sanitarios electrónicos (EHR), como análisis, interoperabilidad, usabilidad, terminología, flujo de trabajo, calidad y resultados, que afectan a enfermería. Dispone, además, de potencial para impulsar la innovación médica, reducir los costes, y, mejorar la calidad de la atención sanitaria, razones por lo que está evolucionando cada vez más rápido.

Comprender las principales barreras que supone el uso de los grandes datos es crucial, pues llevar a cabo un proyecto de tal magnitud, conlleva grandes retos y obstáculos, destacando la necesidad de proteger la confidencialidad, seguridad, y privacidad de la población, además de los ingentes sacrificios económicos que suponen las mediciones experimentales necesarias.



## II. OBJETIVOS

### A. Objetivos generales

- ❖ Conocer Big Data y los principales campos de aplicación en el ámbito sanitario y su relación con enfermería.
- ❖ Analizar las posibilidades que ofrece la adopción de Big Data en diferentes áreas del sector sanitario.
- ❖ Identificar utilidades y aplicaciones del Big Data en la salud digital.

### B. Objetivos específicos

- ❖ Determinar las principales oportunidades que ofrece Big Data en relación con la salud.
- ❖ Exponer las ideas obtenidas tras la investigación de los estudios realizados.
- ❖ Hallar la relación del Big Data con las Unidades de Cuidados Intensivos y con la salud cardiológica.
- ❖ Procurar mostrar una visión realista acerca del estado actual de la introducción de Big Data en el sector sanitario.
- ❖ Enunciar tendencias de evolución futuras en un contexto socio sanitario.
- ❖ Identificar los principales retos y obstáculos que se deben afrontar para lograr la implantación de Big Data.
- ❖ Formular beneficios y limitaciones presentes en el sistema de salud a la hora de implementar esta tecnología.

### III. METODOLOGÍA

El trabajo planteado consiste en una revisión narrativa mediante la que se localiza y sintetiza información de diversos estudios acerca del Big Data, y su relación y utilidad en la atención sanitaria.

Las investigaciones iniciales se realizaron en plataformas de búsquedas convencionales con el fin de cimentar el conocimiento general sobre el objeto de análisis. Una vez contextualizado el tema a tratar y asentado las bases, se han incorporado nuevas búsquedas de carácter más específico y en plataformas científicas, con el propósito de obtener estudios científicos validados por profesionales, e indagar los campos dentro de la salud donde predominan los estudios realizados al respecto, a objeto de modelar un esquema que fragmentara los temas a tratar, y de esta manera, realizar un estudio más profundo de algunas áreas del sistema de salud donde la repercusión del Big Data puede ser importante.

Tras las búsquedas realizadas, se ha llevado a cabo un proceso de selección. Entre los artículos localizados, se encontraron varios de interés para la realización del trabajo, que, al no estar disponibles de forma completa, fueron buscados en otras fuentes con la intención de encontrarlos a texto completo, o en su defecto, ser descartados, continuando con la búsqueda de otros que ofrecieran contenido de utilidad, así como de entera disponibilidad de la información.

La investigación ha sido llevada a efecto en inglés dado que es el idioma vehicular en el campo médico, además de la insuficiente disposición de estudios con base científica presentados en castellano acerca del tema. Los artículos se seleccionaron mediante el título y el resumen disponibles, y posteriormente, aquellos que resultaban ser de interés, se analizaron por su lectura completa.

#### A. Estrategias de búsqueda

La plataforma utilizada en las búsquedas sobre las que se ha fundado el estudio ha sido predominantemente Pubmed y NCBI (National Center for Biotechnology Information), a pesar de haberse realizado búsquedas también

desde JANE (Journal, Author, Name Estimator), útil en un primer momento para orientar la realización de búsquedas, Google Académico, Elsevier, y Scopus, aunque de este último no se eligió ningún artículo debido a la falta de acceso completo a ellos.

Las palabras clave empleadas sobre las que se obtuvieron resultados fueron las siguientes: “Big Data”, “Nursing” “intensive care unit”, “UCI”, “Health services”, “Primary Care Center”, “Medical care”, “Future”, “Medicine”, “Covid”. Utilizándose combinadas entre sí, siendo “Big Data” el denominador común en todas las búsquedas, por medio del operador booleano “AND”, a razón de aspirar localizar los términos clave empleados, juntos en el mismo estudio, y, junto con “OR” en aquellas ocasiones que se pretendía obtener resultados de al menos uno de los términos en adición al otro. Procurando de esta manera efectuar una búsqueda más específica, al descartar aquellos que no serían de utilidad para esta revisión.

En Google Académico las búsquedas se realizaron empleado el lenguaje libre, conteniendo los términos “Big Data”, “Características”, “Enfermería”, “Medicina”, y “Covid”.

Fueron analizadas también las referencias bibliográficas de los artículos seleccionados con el propósito de descubrir otros estudios potencialmente incluíbles en el proceso de la revisión. Así como también fueron favorables las recomendaciones que Pubmed ofrecía sobre artículos similares, ubicando algunos de interés, o bien de enlace a encontrar otros también relacionados. Adicionalmente, se realizó la búsqueda de las páginas oficiales de los proyectos mencionados en artículos utilizados en el presente documento, a objeto de corroborar su actividad e indagar acerca del estado actual de los mismos, y ampliar la información obtenida al respecto.

La selección de búsqueda para la revisión bibliográfica se ordenó y llevó a cabo aplicando una serie de criterios de inclusión y exclusión.

#### 1. *Criterios de inclusión y exclusión*

Una de las condiciones que se tuvo en cuenta para la extracción de artículos, fue que estuvieran publicados en los últimos 5 años (2015-2020), también

debían ser libres de pago, o en su defecto, pedidos directamente del autor, realizado en humanos, y artículos en castellano o inglés.

En las primeras búsquedas, se incluyeron artículos que trataran acerca del Big Data describiendo sus características principales, y de su relación con el sistema sanitario y enfermería. Posteriormente, se especificó sobre aquellos dedicados a su posible función dentro de las unidades de cuidados intensivos y en la salud cardiológica, y, más adelante, se incluyeron aquellos más recientes acerca de su uso en la actualidad y/o futuro.

El principal criterio de exclusión, en adición a los filtros empleados, fue descartar aquellos artículos que trataran sobre Big Data, pero, enfocados a otros campos no intervenidos en esta revisión, o aquellos eliminados por repetición.

## 2. *Extracción de los datos:*

Del total de 2610 referencias encontradas, fueron seleccionados 39 artículos, de los cuales se excluyeron más adelante por repetición de contenido 6, y por no resultar válidos finalmente para esta revisión 11, resultando un total de 23 para la realización del trabajo, añadiendo 3 referencias obtenidas posteriormente de las páginas oficiales de los proyectos en desarrollo mencionados, siendo 25 el sumatorio total empleado.

### B. Análisis de los datos

La información se estructuró en un primer momento, acerca del Big Data, su significado y características principales, asociado seguidamente dentro del contexto socio-sanitario, afectación en la atención de la salud, y en el personal de enfermería, que, posteriormente, se disgrega especificando estudios realizados sobre la repercusión del Big Data, por un lado, dentro de las unidades de cuidados intensivos, y por otro, en la predicción y prevención de las enfermedades cardíacas. Así como sus aplicaciones actuales y proyectos en desarrollo, perspectivas futuras en el terreno de la atención sanitaria, y, desafíos encontrados en alusión a cada una de las partes mencionadas.

## IV. RESULTADOS

### A. Big Data y atención sanitaria

Los datos masivos en el ámbito de la salud, engloban conjuntos de datos significativos demasiado grandes, rápidos y complejos para poder ser procesados e interpretados con las herramientas existentes. La iniciativa de introducir el estudio y manejo del Big Data, es impulsada por el esfuerzo continuo de intentar que los servicios de salud aumenten en eficiencia y sostenibilidad, en respuesta a las demandas de una población en constante expansión, y con una pirámide de edad invertida, así como el cambio de paradigma en la prestación de servicios de salud hacia la prevención, la intervención temprana, y la gestión óptima de recursos <sup>(11)</sup>.

Big Data en la asistencia sanitaria, proviene de fuentes internas, dentro de los hospitales o clínicas, así como de fuentes externas, incluyendo el gobierno, laboratorios, compañías farmacéuticas, revistas médicas, etc. <sup>(8)</sup>.

Este instrumento, facilita el análisis de los resultados de estudios longitudinales a gran escala basados en la población y en miles de pacientes, identificando la correlación entre los conjuntos de datos, además de captar las tendencias, contribuyendo al desarrollo de modelos predictivos, utilizando técnicas de minería de datos a partir de registros sanitarios electrónicos (EHR), permitiendo abordar tanto las enfermedades agudas, como las crónicas <sup>(4, 11)</sup>.

#### 1. *Registros Sanitarios Electrónicos (EHR)*

Los EHR preservan información rica pero poco utilizada. Los centros de datos de salud, captan y almacenan una enorme cantidad de datos heterogéneos (estructurados, semiestructurados o no estructurados) como datos personales de los pacientes, notas clínicas, diagnósticos, datos administrativos, gráficos, prescripciones, procedimientos, pruebas de laboratorio, imágenes de resonancia magnética, ultrasonido, datos de tomografía computarizada, etc. <sup>(4)</sup>. Resultando el análisis de los mismos, una herramienta valiosa para la obtención de muestras representativas a grandes dimensiones, posibilitando mejorar el conocimiento clínico y apoyar la investigación clínica. Asimismo, la

extracción de la información aportada por estos datos, favorece el apoyo al manejo de enfermedades, la farmacovigilancia, la construcción de modelos para predecir la evaluación de riesgos para la salud, la mejora del conocimiento sobre las tasas de supervivencia, la recomendación terapéutica y el descubrimiento de comorbilidades, permitiendo por ejemplo, detectar mediante la aplicación de algoritmos a grandes bases de datos, nuevas interacciones, efectos adversos de medicamentos, y otros problemas de seguridad que pueden desencadenar la retirada de fármacos o la modificación de sus indicaciones. Además, incentiva la construcción de sistemas de apoyo, imprescindibles para el reclutamiento de pacientes para nuevos ensayos clínicos <sup>(11)</sup>.

## 2. *Aplicaciones del Big Data*

La usanza de los grandes análisis de datos, también pueden mejorar el servicio basado en el paciente para detectar antes la propagación de las enfermedades, generar nuevos conocimientos sobre los mecanismos de la enfermedad, supervisar la calidad de las instituciones médicas y de atención sanitaria, así como proporcionar mejores métodos de tratamiento <sup>(4)</sup>. En la gestión de una pandemia de gripe, por ejemplo, la información heterogénea procedente de fuentes gestionadas y no gestionadas (como los medios de comunicación social o los viajes aéreos), puede procesarse, extraerse y convertirse en medidas decisivas para controlar el brote <sup>(11)</sup>. Sin ir más lejos, la emergencia actual de COVID-19, se está produciendo en un mundo mucho más digitalizado y conectado. La cantidad de datos producidos durante años atrás, se genera hoy en día en unos pocos minutos, y los modelos computacionales avanzados, han demostrado un gran potencial para rastrear la fuente o predecir la propagación futura de enfermedades infecciosas. Por lo tanto, es ineludible aprovechar los grandes datos y los análisis inteligentes para hacer un buen uso de ellos en beneficio de la salud pública <sup>(12)</sup>.

Los datos de los teléfonos móviles, ya mostraron potencial para predecir la propagación del cólera durante la epidemia de 2010 en Haití, así como el aprovechamiento de los análisis de grandes datos demostró su eficacia durante la crisis del Ébola en África occidental de 2014-2016 <sup>(12)</sup>.

Las técnicas de minería de datos empleadas en los EHR, los datos de la web y los medios sociales, permiten identificar reglas de asociación entre los EHR, y así hallar el seguimiento de la enfermedad y las tendencias basadas en la salud. Además, la integración y el análisis de los datos de diferente naturaleza, como los sociales y científicos, pueden conducir a nuevos conocimientos, explorando nuevas hipótesis para localizar patrones ocultos <sup>(4)</sup>.

### 3. *Enfermería y Big Data*

Enfermería se encuentra en una reforma de la atención de salud, compartiendo responsabilidades, recursos, competencias, riesgos, y beneficios, con el fin de lograr los objetivos. Se incrementa la necesidad de capturar, administrar y compartir información de salud, para producir un marco longitudinal holístico centrado en el paciente, modificando la forma en que son realizadas estas acciones, y, ocupándose de complejos entornos y redes con múltiples sistemas clínicos donde los datos deben ser integrados dentro de estos EHR <sup>(13)</sup>.

Para poder compartir datos entre sistemas y organizaciones, es necesario que exista una alineación de los mismos, siendo necesario un método de recopilación junto con las terminologías necesarias para apoyarlos. Para ello, se recomienda la promoción de una serie de normas de interoperabilidad para el avance de las medidas electrónicas de calidad, estructurando los datos mediante patrones que promuevan el uso de terminologías aceptadas y estandarizadas que aborden las necesidades de documentación, independientemente del entorno de atención. Ajustes en la prestación de la atención sanitaria para crear un plan que implemente términos conocidos, y sean normalizados a través de un organismo de consenso internacional, pues la falta de la misma hace que la comparación de los datos se dificulte. Dichas terminologías de enfermería, deberían ser actualizadas constantemente y puestas a disposición de las organizaciones internacionales para su traducción. Asimismo, minimizando la documentación de texto libre, y debiendo ser almacenados los datos dentro de los EHR para permitir el apoyo a la investigación, el análisis y la toma de decisiones <sup>(13)</sup>.

La medición de los datos de calidad es un proceso complejo. Las pautas de calidad clínica forman parte de los esfuerzos para el avance de las medidas que afectan a la enfermería. Estas incluyen, determinar cuáles son de calidad y apoyan la atención basada en la evidencia que sigue las pautas clínicas prácticas y minimiza el impacto negativo en el flujo de trabajo <sup>(13)</sup>.

Big Data está evolucionando de forma veloz, acelerando el crecimiento y la síntesis de nuevos conocimientos para tener un impacto positivo en las enfermeras y las poblaciones a las que sirven. Comprender las principales barreras, los desafíos y las implicaciones del Big Data en la enfermería, ayudará a alcanzar el objetivo de mejorar la experiencia de atención del paciente, mejorar la salud de la población, y reducir el costo per cápita de la atención sanitaria <sup>(13)</sup>.

Es decir, Big data puede ayudar a avanzar en la prestación de servicios de salud, mejorando la calidad y la eficiencia, ampliando considerablemente la capacidad de generar nuevos conocimientos, ayudando en la difusión de los mismos compartiendo y trasladando análisis e iniciativas de medicina personalizada a la práctica clínica a través de la red, y, transfiriendo la información de forma rápida y directa a los pacientes, por un lado, asociando datos clínicos a otros datos personales (hábitos, cultura, educación, tipo de trabajo), y por otro, facilitando su acceso a dicha información <sup>(14)</sup>.

Big Data puede ser un recurso valioso para proporcionar importantes conocimientos que mejoran los servicios de salud y mejorar la economía de la asistencia sanitaria, sin embargo, también plantea importantes desafíos sociales y legales en términos de privacidad, reidentificación, gobernanza, propiedad y administración de los datos, tratados más adelante <sup>(11)</sup>.

## B. Big Data en Unidades de Cuidados Intensivos

Las Unidades de Cuidados Intensivos (UCI) equipadas con sistemas de información electrónica, generan a diario cientos de elementos de información estructurada, como los signos vitales, los registros de vigilancia básicos y avanzados, los datos de laboratorio, las recetas de medicamentos y su



administración, así como los registros de enfermería, entre otros. También se genera una gran cantidad de información textual, como los comentarios y notas de progreso de enfermeras y médicos, así como imágenes de diferentes formatos y estructura heterogénea. Siendo uno de los desafíos fundamentales, lograr el manejo de esta enorme cantidad de información, ya que es imposible que el cerebro humano asimile plenamente esta abundancia de datos. El crecimiento exponencial de los sistemas de almacenamiento y procesamiento de información computarizada, posibilita tanto el almacenamiento de una ingente cantidad de datos clínicos en formato electrónico durante el ingreso de los pacientes, como emplearlos posteriormente para obtener respuestas a preguntas clínicas, mejorar la gestión de recursos, o sugerir líneas de investigación que luego pueden ser explotadas mediante ensayos clínicos aleatorizados <sup>(9)</sup>.

La disponibilidad de estas grandes cantidades de datos en las UCI, unido a la necesidad de obtener una atención óptima basada en pruebas, y la complejidad de las enfermedades críticas, incentiva el uso del Big Data y el desarrollo de nuevos estudios al respecto. Sin embargo, a pesar del creciente número de estudios y publicaciones en la materia, hasta el momento han sido pocos los proyectos que han dado lugar a la aplicación satisfactoria de sistemas basados en datos procedentes de estas unidades de cuidados críticos.

Los proyectos y aplicaciones más comunes, son los modelos de predicción de una condición, evento, o respuesta, y de pronóstico, entrenados específicamente para predecir la probabilidad de un criterio de valoración o resultado relacionado con una condición, como la mortalidad, utilizando ambos modelos, algoritmos de aprendizaje supervisado. Otras aplicaciones, surgen del procesamiento del lenguaje natural por métodos de análisis de los datos narrativos, generalmente conocidos como procesamiento del lenguaje natural, concebidos para extraer características de los textos que luego pueden utilizarse en algoritmos específicos <sup>(15)</sup>.

El Análisis Big Data (BDA), va de la mano con otro concepto, el "aprendizaje automático" o "machine learning" (ML), que sustenta lo que se conoce como

Inteligencia Artificial (IA). Esta tecnología utiliza diferentes metodologías para identificar patrones a partir de los datos, implicando la clasificación o la predicción de eventos, enfocando la salida del software a una tarea muy específica, como la interpretación automática de imágenes médicas. Los cuidados intensivos son particularmente adecuados para la ciencia del aprendizaje automático, dado que la tecnología es un gran componente de la atención que se presta en estas unidades <sup>(9)</sup>.

El empleo del BDA y el ML, supone la colaboración con profesionales científicos de datos, incluidos estadísticos y epidemiólogos, clínicos y otros investigadores biomédicos. Contribuyendo a ampliar los horizontes en aspectos como el control de calidad de las funciones cotidianas del sector, la comparación de resultados entre unidades, o la elaboración de nuevas líneas de investigación clínica <sup>(9)</sup>.

Su empleo puede ser relevante teniendo en cuenta que un paciente de la UCI genera miles de datos en un día, pudiendo añadir a ello, la fatiga del personal cualificado en momentos determinados por la carga de trabajo, sumado a las interrupciones y los propios sesgos cognitivos de los clínicos, existiendo la posibilidad de resultar desacertadas algunas de las decisiones clínicas tomadas. Por el contrario, los sistemas informáticos pueden cribar sin tener en cuenta esta problemática decenas de miles de datos, y analizar fácilmente complejas interacciones no lineales entre variables. No obstante, debe considerarse la diversidad de factores que influyen en la adopción de decisiones clínicas, como sociales y personales, que no son reflejados necesariamente en los registros digitales, por lo que cualquier resultado de un sistema basado en datos deberá ser evaluado, interpretado y enriquecido por los clínicos antes de adoptar cualquier medida <sup>(15)</sup>.

### 1. *Iniciativas en España*

Aunque la acogida de los EHR y las bases de datos de pacientes críticos se encuentra aún en desarrollo y no todas las unidades disponen de un sistema de información digital integrado, ya hay algunas iniciativas en curso.

La Sociedad Española de Medicina Intensiva, Crítica y Unidades Coronarias

(SEMICYUC) está en proceso de construcción del registro electrónico nacional del conjunto mínimo de datos básicos (CMBD), para facilitar la medición de la calidad, la evaluación comparativa, y la investigación clínica <sup>(9)</sup>.

En algunos hospitales ya se ha iniciado el proceso de creación de bases de datos anónimas para la investigación en colaboración. El Servicio de Cuidados Críticos del Hospital Clínico San Carlos de Madrid, en cooperación con Philips, ya incluye 12.000 ingresos en su base de datos tras la implantación del sistema de información clínica IntelliSpace Critical Care and Anesthesia (ICCA™, Philips). Otras UCI de España, como la Unidad del Hospital Joan XXIII de Tarragona, trabajan en iniciativas similares, empleando los datos obtenidos automáticamente de su Comité Ético de Investigación (CEI) <sup>(9)</sup>.

También con el fin de promover el BDA y el ML en la sanidad del país, está en activo el evento “Critical Care Datathon”, organizado por el Grupo de Datos Críticos del Instituto Tecnológico de Massachussets (MIT), el grupo de investigación Life Supporting Technologies (LifeSTech) de la Universidad Politécnica de Madrid (UPM) y la Unidad de Innovación y el Departamento de Cuidados Críticos del Hospital Clínico de San Carlos (HCSC) de Madrid, bajo los auspicios de la SEMICYUC. Este reúne un equipo de clínicos, analistas de datos, estadísticos y expertos en seguridad informática de datos, para realizar sesiones de trabajo colaborativo en grupos, utilizando una base de datos reales anonimizada (MIMIC-III), y analizar preguntas de investigación de interés clínico, como la elección de los antibióticos empíricos en la UCI sobre la mortalidad hospitalaria, los factores de riesgo para la estancia prolongada en la UCI, la predicción del mal pronóstico después del ingreso por paro cardíaco reanimado fuera del hospital, la predicción de la mortalidad en pacientes con ingreso no electivo en la UCI, las consecuencias pronósticas de la hiperglucemia en pacientes críticos con y sin diabetes, la epidemiología de la transfusión masiva de células sanguíneas empaquetadas, entre otras. Los estudios han establecido resultados relevantes acerca de cómo evaluar datos, utilizando también técnicas de procesamiento de lenguaje natural para extraer información de los informes de los clínicos escritos en texto libre, además de identificarse variables potencialmente relacionadas con el pronóstico en MIMIC-

III, generando una tabla de más de 40.000 pacientes con todas las variables seleccionadas <sup>(9)</sup>.

El encuentro ayuda a su vez a forjar la unión entre las UCI españolas que trabajan en el campo del BDA, fortaleciendo las asociaciones entre ellas para promover el avance de esta iniciativa, planificando futuras reuniones, y contando con los datos de unidades españolas con el propósito conjunto de esbozar varias líneas de investigación clínica a desarrollar en el futuro. Asimismo, se asentaron las bases para poder utilizar las plataformas de datos de las UCI con las que se cuenta en España actualmente <sup>(9)</sup>.

La creación y el uso de estas bases de datos, conllevan importantes implicaciones legales que deben ser abordadas para cumplir con las políticas reglamentarias, en particular las de la Ley de Protección de Datos española <sup>(9)</sup>.

### C. Big Data en cardiología

Durante los últimos años, el auge y el desarrollo en técnicas de la Inteligencia Artificial (IA), se ha comenzado a labrar un camino en el campo de la salud cardiaca. Las enfermedades cardiovasculares según la Organización Mundial de la Salud (OMS), ocasionan cerca de 18 millones de muertes al año, posicionándolas en una de las causas de muerte más frecuentes en todo el mundo. Se denominan enfermedades "epidémicas silenciosas" debido a la manifestación paulatina de los síntomas, suponen un alto coste sanitario, y el aumento de la esperanza de vida se debe al éxito de la asistencia sanitaria recibida. Mejorar el diagnóstico y la predicción de estas enfermedades, aumentando en consecuencia la calidad de la atención al paciente y reduciendo las tasas de readmisión y mortalidad, son razones por las que el desarrollo del Big Data puede ser primordial <sup>(1, 2, 16)</sup>.

Las enfermedades cardiovasculares (ECV), son complejas y heterogéneas por naturaleza, ya que están causadas por múltiples factores <sup>(16)</sup>, produciendo multitud de datos variados, ya sean fisiológicos (ECG, presión arterial, niveles de edemas de pulmón, actividad del paciente), o basados en imágenes (ecocardiograma, resonancia magnética, etc.). La interpretación de los mismos

y la identificación de anomalías, puede mejorar gracias al uso de tecnologías digitales que incorporen el Big Data y la IA <sup>(2)</sup>, ya que poseen potencial para analizar los inherentemente diversos volúmenes de datos de los EHR, permitiendo explotar estos datos y enfocarlos a la identificación de síndromes como el coronario agudo, la cardiomiopatía hipertrófica, el fallo del corazón con fracción de eyección conservada, la cardiomiopatía de Takotsubo, entre otras, además de hallar factores de riesgo desconocidos <sup>(16)</sup>. Es decir, ayudan en la comprensión de la causa de las alteraciones de la salud, y a clasificar la información relevante, contribuyendo al conjunto de análisis que darán pie a la detección temprana de posibles anomalías causantes de estas patologías, modificando el modo de investigación y atención de la salud cardiovascular, y, ayudando a respaldar importantes decisiones clínicas <sup>(5)</sup>.

Aunque la utilización de Big Data se encuentra en sus inicios, se han implementado grandes análisis de datos en la medicina cardiovascular <sup>(16)</sup>.

Se han efectuado diferentes experiencias con el uso del Big Data y presentado estudios en pacientes con disfunciones cardíacas.

Por medio del estudio *“Development of Big Data Predictive Analytics Model for Disease Prediction using Machine learning Technique”* <sup>(1)</sup>, analizaron el funcionamiento del aprendizaje automático para pre-diagnosticar insuficiencia cardíaca. Las fuentes de datos correspondían a diagnósticos, medicaciones, datos de laboratorio, hospitalizaciones, datos demográficos, hábitos, imágenes médicas, signos vitales, y signos y síntomas generalmente relacionados con la insuficiencia cardíaca.

Emplearon el modelo clasificador de Bayes Ingenuo para elaborar una técnica de predicción del estado futuro de salud a partir de los datos. El sistema propuesto, utiliza una clasificación probabilística basada en el teorema de Bayes para analizar los datos mediante el aprendizaje automático. Este método de clasificación por agrupaciones, filtra los datos innecesarios de manera que el resultado de la predicción aumente su eficacia.

Para llevarlo a cabo, diseñaron un algoritmo tratando de clasificar el variado conjunto de datos a objeto de predecir de manera efectiva futuras anomalías en

los parámetros relacionados con el desarrollo de enfermedades cardíacas. Para efectuarlo, manejaron el conjunto de datos del repositorio de máquinas de la UCI, y los resultados en la predicción futura aseguraron que el sistema propuesto mejoraba significativamente en términos de precisión <sup>(1)</sup>.

En otro estudio, *“Artificial intelligence algorithm for predicting mortality of patients with acute heart failure”* <sup>(18)</sup>, desarrollaron y validaron un nuevo modelo de inteligencia artificial de predicción de la mortalidad de la insuficiencia cardíaca aguda (ICA) basado en el enfoque de aprendizaje profundo. Para ello, realizaron un estudio retrospectivo de cohorte de observación utilizando gran cantidad de datos de pacientes con ICA de 12 hospitales a través de los EHR.

El propósito de la práctica, trataba de evaluar la exactitud de varios modelos de predicción, analizando y comparando la diferencia de precisión y rendimiento entre métodos de aprendizaje profundo, de aprendizaje automático, y otros modelos convencionales.

El estudio reveló que la precisión del rendimiento del modelo de inteligencia artificial basado en el algoritmo del aprendizaje profundo, era excelente para predecir la mortalidad intrahospitalaria y a largo plazo de los pacientes con ICA en el momento del ingreso, prediciendo la mortalidad intrahospitalaria de 12 y 36 meses de pacientes con ICA con mayor precisión que las puntuaciones del modelo convencional, y, además, superando el modelo de aprendizaje profundo a otros modelos de predicción de aprendizaje automático. Demostrando la viabilidad y la eficacia del modelo de algoritmo de inteligencia artificial basado en el aprendizaje profundo para la cardiología, y posibilitando ser una herramienta útil para la toma de decisiones precisas <sup>(17)</sup>.

#### D. Big Data y actualidad en la atención sanitaria

Hoy en día, la evolución de fuentes digitales y su consiguiente aumento de posesión y uso de las mismas, como los teléfonos móviles y otros dispositivos, está abriendo las puertas a una época de cambios y adaptaciones en vistas a mejorar y amoldar los requerimientos actuales con las posibilidades adquiridas.

El Análisis Big Data (BDA) ya se emplea de forma habitual en otras esferas como la comercialización, la adopción de decisiones empresariales estratégicas, los seguros, la banca, la detección de fraudes en el comercio electrónico, etc. <sup>(9)</sup>. Sin embargo, en el ámbito de la salud y la medicina, aún son escasos los estudios y ensayos realizados.

En España, varios proyectos han comenzado a desarrollar herramientas con datos masivos. En Girona, se está trabajando en el proyecto Savana, una plataforma de apoyo a la decisión médica que ya está siendo empleada en algunos centros sanitarios españoles <sup>(14)</sup>, diseñada para convertir Historias Clínicas Electrónicas en Big Data, desbloqueando mediante Inteligencia Artificial (IA) el valor clínico que existe en la información no estructurada <sup>(19)</sup>, provocando un cambio disruptivo en la atención sanitaria al aumentar la eficiencia de la gestión e investigación clínica.

Algunos de los usuarios que han formado parte del proyecto, lo describen de la siguiente manera:

*“Savana es una de las mejores soluciones en tecnologías de la información que hemos visto aparecer en España en años. Un aterrizaje real y concreto del big data que permite gestionar de forma totalmente distinto el conocimiento de los hospitales.”* Dr. Cristóbal Belda, Director de Innovación de HM Hospitales <sup>(18)</sup>.

*“Savana es un ejemplo tangible de cómo el big data mejora la forma en la que detectamos y anticipamos hechos clínicos en la práctica clínica, gracias a la reutilización de la información que hasta ahora teníamos infrautilizada. Para los profesionales sanitarios es una herramienta motivadora en la práctica asistencial para obtener respuestas ágiles a sus preguntas asistenciales, así como un incentivo potente para fomentar y desarrollar investigación.”* Dra. Carmen Pantoja, Gerente del Hospital Universitario Infanta Leonor <sup>(18)</sup>.

*“Savana concreta el uso de la inteligencia artificial en una herramienta que transforma completamente la forma de evaluar los resultados clínicos. Al interactuar con Savana, puede apreciarse su diseño completamente enfocado a la forma de trabajar en los sistemas sanitarios reales, con sus retos y*

*peculiaridades*” Dr. Julio Mayol. Director Médico del Hospital Universitario Clínico San Carlos <sup>(18)</sup>.

*“El análisis de millones de textos clínicos mediante técnicas big data permite inferir algoritmos predictivos que individualizan el riesgo del paciente y permiten tomar decisiones donde hasta ahora la evidencia era insuficiente”* Dra. Silvia García, Especialista en Neumología Complejo Universitario Hospital de León <sup>(18)</sup>.

Este tipo de sistemas que utilizan las grandes bases de datos junto con los propios datos de los pacientes a través de dispositivos, conducen a la denominada Medicina de Precisión, evolucionando hacia la Medicina Personalizada, y, conformando la Medicina Generadora de Evidencia, a diferencia de la actual Medicina Basada en la Evidencia <sup>(14)</sup>.

Otra de las aptitudes ofrecidas por las avanzadas tecnologías, es la posibilidad de estudiar el comportamiento humano, incluyendo sus acciones físicas, emociones observables, personalidad, y patrones de interacción social. Características difíciles de medir y cuantificar, sin embargo, la capacidad actual de su análisis, está facilitando el entendimiento de mecanismos de desarrollo y propagación de enfermedades, así como su afectación también a otras patologías a nivel comunitario <sup>(11)</sup>. Pues la utilización de estas fuentes de datos digitales, como los datos de los teléfonos móviles y otros dispositivos, es de particular valor en los brotes causados por patógenos recientemente descubiertos, para los cuales, los datos oficiales y las previsiones fiables son escasos <sup>(12)</sup>. Durante la presente pandemia originada por el COVID-19, enfermedad desarrollada en un contexto de salud digital, la importancia del uso de las tecnologías de la información emergentes, como las aplicaciones móviles (apps), y dispositivos de Internet para concentrar iniciativas gubernamentales y privadas de salud digital, ha tenido como propósito, lograr la geolocalización de casos diagnosticados y utilizar Big Data para afrontarla, además de brindar información, consejos prácticos, y consultas en línea y telefónicas <sup>(19)</sup>. Siendo la recopilación de datos a gran escala, fundamentales para gestionar la pandemia por COVID-19 mediante el uso de los mismos y junto a algoritmos disponibles digitalmente para la predicción y la vigilancia <sup>(12)</sup>.



En China, se desarrolló la aplicación “Close contact detector”, lanzado por la Comisión Nacional de Salud de China, y China Electronics Technology Group Corporation. La plataforma "detector de contacto cercano" utiliza de las autoridades públicas Big Data acerca del movimiento de las personas y de registros de casos de enfermedades para constatar si el usuario ha tenido algún contacto cercano con personas cuya infección se haya confirmado, o se sospeche que haya sido infectada en el pasado reciente, a través de las tres aplicaciones sociales y de pago móviles más estimadas en China; Alipay, WeChat y QQ <sup>(19, 20)</sup>.

El desarrollo de estos sistemas, puede apoyar la lucha no solo contra el actual coronavirus, sino también al conjunto de brotes y epidemias de enfermedades infecciosas <sup>(20)</sup>.

## V. DISCUSIÓN

Big Data en el marco sanitario, es un designio prometedor de exploración y análisis de una gran cantidad de datos heterogéneos complejos de distinta naturaleza. Su integración entrelaza varios campos, como la bioinformática y la estadística <sup>(4)</sup>, y puede suponer un cambio de paradigma en la práctica de la medicina.

Las ventajas principales de la reutilización de los datos recopilados de forma rutinaria para la investigación, es poder realizar estudios a mayor escala, más económicos, y en menor tiempo que los estudios convencionales, permitiendo analizar tamaños de cohorte muy grandes durante amplios periodos de seguimiento, optimizando la estimación fiable de las tasas de incidentes y de resultados extraños, además de los efectos a largo plazo de las intervenciones. También, el entorno real en el que se recogen los datos, aumenta la validez de los estudios en comparación con diseños más tradicionales, como los ensayos controlados, por el descenso de la tendencia de los individuos a modificar su comportamiento al ser conscientes de estar siendo observados <sup>(6)</sup>.

### A. Perspectivas de futuro

Las características que describen los grandes datos, proporcionan una base muy apropiada para utilizar plataformas de software prometedoras para el desarrollo futuro de aplicaciones destinadas a la medicina y la atención sanitaria, que permitan la aplicación de técnicas de minería de datos sobre gran cantidad de información heterogénea y compleja, aventajando a la ciencia e impulsándola a revelar nuevos conocimientos y patrones ocultos <sup>(4)</sup>. También, mediante la combinación de diferentes hospitales, empresas de tecnología de la información, fabricantes de equipo e instituciones médicas comunitarias, se espera crear una unificación de datos sanitarios, y establecer una biblioteca central de EHR para proporcionar datos médicos abiertos, recuperables, disponibles y de alta calidad, para mejorar el servicio y su desarrollo <sup>(21)</sup>.

Análisis realizados entre profesionales clínicos, líderes médicos y ejecutivos de áreas relativas a la salud, acerca de los datos masivos y su evolución frente a

años venideros, consideraron que, a pesar de ser aún escasas las aplicaciones del Big Data sobre la asistencia sanitaria, se verán aumentadas, de igual modo, el uso de información de los EHR incrementaría su efectividad, suponiendo un beneficio para la toma de decisiones clínicas <sup>(14)</sup>.

En relación a los cuidados intensivos, se prevé que grandes cantidades de datos clínicos sean cada vez más asistidos por sistemas informáticos que realizarían su integración y análisis, transformándolos en forma de conocimiento manejable, interpretable y procesable, de manera que pueda beneficiar la capacidad de decisión clínica. Las falsas alertas se reducirían, y los modelos predictivos realizarían recomendaciones diagnósticas y terapéuticas, mientras los profesionales contextualizarían dicha información, coordinando su aplicación. Se potenciaría el valor de los datos, y aumentaría la individualidad de la atención sanitaria. Los sistemas mejorarían continuamente mediante un enfoque colaborativo y científicamente riguroso con otros especialistas no sanitarios en el área de la ciencia de los datos, primordial para garantizar el buen funcionamiento del método implantado <sup>(9, 16)</sup>.

Todos los profesionales sanitarios, técnicos de laboratorio clínico, investigadores, etcétera, tendrían el compromiso y la responsabilidad de adquirir nuevos conocimientos, marcos mentales y perspectivas, que condujesen a un ecosistema de salud totalmente revisado, además de realizar un trabajo cooperativo entre los diversos integrantes, destinado a instituir una medicina personalizada, alterando también el modo de actuación de los pacientes, asumiendo en su atención sanitaria, mayores niveles de responsabilidad sobre su propia salud y bienestar <sup>(22)</sup>.

## B. Dificultades técnicas

El hallazgo del Big Data, está proporcionando pruebas convincentes de que el uso inteligente de este instrumento podría proporcionar información inestimable. Esta nueva perspectiva, está dotada de oportunidades únicas, pero también se encuentra con una serie de problemas técnicos a la hora de su implantación que deben ser abordados con el fin de lograr las pretensiones aspiradas <sup>(22)</sup>.

Para empezar, es necesario identificar la disponibilidad de los datos de interés, que, a causa de la gobernanza de la información y los modelos de acceso a los datos, puedan no tenerla <sup>(5)</sup>. Para ello, deben superarse los obstáculos referidos a la integración sistemática de los datos, una traducción efectiva de la investigación y la atención de la salud a gran escala que posibilite la existencia de instrumentos fiables para integrar los conjuntos de datos y permitiendo su comparabilidad y coherencia, para dar lugar a una ciencia ampliable y reproducible en todas las instituciones y países gracias al equipamiento de la fuerza de trabajo clínica y científica con nuevos conocimientos interdisciplinarios. Aunque, aún combinando las diferencias en un modelo único, serían abundantes las controversias presentes, además de carecer de una metodología consensuada y sólida <sup>(5, 22)</sup>.

Los datos requeridos en este ámbito, suelen ser diversos, incompletos e imprecisos por naturaleza. Incluso la función de adquirir e incorporar los EHR en todos los hospitales dentro de un mismo país, ha demostrado ser mucho más complejo de lo previsto, hasta en los sistemas de atención de salud más avanzados. En Estados Unidos, más del 90% de los hospitales han adoptado las herramientas de EHR, obteniéndose de diversas empresas, y su capacidad de comunicación sigue siendo limitada. En Europa, también se están llevando a cabo iniciativas para establecer plataformas sólidas de recopilación e intercambio de datos normalizados <sup>(22)</sup>, como DIFUTURE (Data Integration for Future Medicine) en Alemania, comprometida con la colaboración internacional y con el objetivo de proporcionar datos de gran amplitud y profundidad trabajando en la aceleración de la innovación para la mejora en la toma de decisiones y de la atención sanitaria <sup>(23)</sup>, o Alleanza contro il Cancro en Italia, cuyo objetivo es favorecer una agregación multidisciplinaria y sinérgica entre el área de investigación y el área clínica con el fin de efectuar una investigación traslacional, intensificando la medicina personalizada y de precisión, adquiriendo importantes efectos clínicos en períodos de tiempo relativamente cortos <sup>(24)</sup>.

Otra de las cuestiones difíciles, se presenta con respecto a la obtención de datos de alto rendimiento, siendo necesario superar controles de calidad y

seguridad de los mismos, estando ligado todo ello al costo de las mediciones experimentales. Tanto las fuentes de información, como la gran diversidad de los datos, las técnicas experimentales, las condiciones ambientales y la naturaleza biológica, deben considerarse antes de integrarlos y de emplear los métodos de extracción de datos.

Se les suma a estos inconvenientes la influencia del personal que los introduce, como puede ser el caso de los EHR, considerando la posibilidad de introducirse valores perdidos o datos incorrectos, incrementándose este hecho cuando se emplean datos procedentes de dispositivos digitales personales, como resultado de mala comprensión o interpretación errónea de los datos originales. Incorporándose pequeños errores en uno o varios factores, y, en consecuencia, pudiendo tener un efecto desproporcionado en los modelos de predicción a gran escala.

La integración de los datos de diversas plataformas y la normalización de protocolos, siguen siendo cuestiones difíciles, pudiendo dar lugar estas deficiencias a la falta de fiabilidad en algunos de los puntos, como los valores perdidos o los valores atípicos <sup>(12, 25)</sup>.

Todo ello, promueve la escasa disponibilidad de evidencias de sus beneficios prácticos sobre el uso en el mundo real de la adopción de los análisis Big Data en el cuidado de la salud <sup>(8)</sup>, interfiriendo también en el proceso, la adopción y el mantenimiento de la confianza del público <sup>(5)</sup>, debiéndose cumplir en todo momento la necesidad de proteger la seguridad, privacidad y confidencialidad de los individuos. Cuestiones importantes hacia los grandes datos en el cuidado de la salud, ya que todos los datos clínicos son muy sensibles, y los diferentes países consideran que estos son legalmente posesión de los pacientes <sup>(25)</sup>.

Para abordar estos desafíos de seguridad y privacidad, las grandes soluciones de software de análisis de datos deben utilizar algoritmos avanzados de encriptación y pseudo-anonimato de los datos personales, proporcionando seguridad a nivel de red y autenticación para todos los usuarios involucrados, debiendo garantizar la privacidad y la seguridad, así como establecer normas y prácticas de buen gobierno <sup>(25)</sup>.

Las barreras referidas a la integración del Big Data a los que se enfrenta enfermería, además de las anteriores mencionadas, se enuncia la insuficiente definición de estándares universales de los elementos de datos discretos, así como la disminución de la capacidad de la documentación de texto libre de almacenar estos datos discretos en los EHR, pudiendo alterar el resultado, afectando a las comunicaciones entre usuarios y lugares, y limitando la interoperabilidad de los mismos <sup>(13)</sup>.

La reforma de las políticas sanitarias está suponiendo una carga adicional en los requisitos de documentación, lo que aumenta aún más la complejidad de los sistemas de salud y los procedimientos organizativos para el cumplimiento normativo. Además de la debida consideración al impacto en el flujo de trabajo en el momento de añadir nuevas medidas que garanticen la calidad clínica y sean de uso significativo <sup>(13)</sup>.

Es decir, se espera que los datos clínicos se conviertan en información analítica eficaz, sin embargo, para su empleo se encuentran barreras que incluyen la inexactitud, el carácter incompleto, el modo de conversión de los datos, la regulación desconocida, la inaccesibilidad de los datos de investigación, la granularidad insuficiente, y la incompatibilidad con las herramientas de investigación <sup>(21)</sup>.

Dentro del marco de las unidades de cuidados intensivos, uno de los primeros retos a los que se enfrenta Big Data, son los pocos ejemplos de proyectos que dan lugar a una aplicación satisfactoria de estos sistemas basados en la UCI, generando desconfianzas acerca de compartir responsabilidades en la toma de decisiones clínicas teniendo en cuenta los algoritmos del aprendizaje automático.

La introducción de la aplicación de estos sistemas, debe realizarse bajo rigurosos auspicios de estudios experimentales que incluyan suficientes ensayos controlados y aleatorios previos, ya que la facilidad de acceso a datos masivos, puede dar lugar a investigaciones de baja calidad debido a problemas clínicos mal enmarcados, datos erróneos, o métodos discutibles, que podrían dar lugar a una ciencia defectuosa, y, en ese caso, predominar la creación de problemas sobre la solución de los mismos <sup>(15)</sup>.

Por último, se enfrentan al reto de equilibrar la necesidad de apertura y reproducibilidad de los datos con la demanda de privacidad y seguridad de los mismos. El movimiento de apertura de esta ciencia, exige una investigación transparente y reproducible con un intercambio de datos sin fisuras entre las instituciones, pues se ha demostrado una alarmante falta de reproducibilidad en los estudios de ciencia de los datos que utilizan los mismos datos de la UCI, lo que sugiere que los algoritmos, los procedimientos de estudio, el código informático e incluso los conjuntos de datos, deberían estar disponibles abiertamente para garantizar la reproducibilidad <sup>(15)</sup>.

### C. Inconvenientes éticos

La Inteligencia Artificial (IA) plantea graves problemas de posibles ataques adversos con intenciones de alterar o corromper alguno de los procesos, por lo que sería necesario establecer límites éticos apropiados.

La reciente introducción en la comunidad europea del Reglamento General de Protección de Datos de la UE (GDPR) es un intento global de proteger los derechos de privacidad de las personas, al mismo tiempo que se fomenta la investigación y, más concretamente, el intercambio gratuito de datos científicos. A pesar de su considerable dureza sancionadora en caso de incumplimiento, la filosofía general en la que se basa, gira en torno a la descentralización, mediante la delegación de la responsabilidad a los controladores de datos. Asimismo, la GDPR aumenta el cometido de las juntas de examen interno y los comités de ética, con una función más importante en la elaboración de políticas.

Aunque es demasiado pronto para evaluar adecuadamente el impacto y la función de la GDPR en la gestión de Big Data, se prevé que surgirán disputas en torno a la gestión y a la regulación de los datos, y de quién podría evaluarlos. Por consiguiente, surgen dos cuestiones; la necesidad de volver a preguntar a los pacientes por el uso de sus datos en proyectos de investigación que puedan ir más allá del alcance específico del consentimiento inicial, y, si las normas impuestas a los círculos académicos en el manejo de los datos, podrían ampliarse a los agentes comerciales, siempre que los datos sean buscados y adquiridos con fines de investigación científica <sup>(22)</sup>.

## **VI. CONCLUSIONES**

Dentro del análisis expuesto, es posible vislumbrar la importancia de fomentar el conocimiento sobre Big Data, puesto que revela tener un inmenso potencial para optimizar resultados de la calidad de la atención sanitaria. Tal como esta revisión demuestra, entre sus aplicaciones se encuentra reutilizar los EHR para obtener mayor eficacia y rentabilidad en la investigación clínica, implementando nuevos sistemas de aprendizaje automático y análisis predictivo de gran interés para el diseño de métodos de tratamiento y cuidados clínicos.

En el mismo sentido, se han observado favorables resultados en la mejora de pronóstico y apoyo a las decisiones clínicas, mediante Inteligencia Artificial, dentro de las UCI y frente a patologías cardíacas, ganando relevancia ya en la actualidad, tanto a nivel internacional como nacional, exprimiendo los recursos tecnológicos disponibles para transformar la prestación de la atención sanitaria actual, en busca de reducir el error y el gasto sanitario.

Aunque no constan grandes resultados de su vínculo con enfermería, sí se ha incrementado la evidencia del potencial de esta herramienta para lograr una mejora en el cuidado de los pacientes, repercutiendo a poblaciones completas.

Se visualiza un futuro donde se logra la unificación de datos, aumentando la efectividad de los mismos, fortaleciendo la confianza tanto de los investigadores, como de la población, para continuar fomentando su uso y con ello, la mejora de la misma. Sin embargo, la revolución del Big Data aún no se ha realizado en todo su potencial, aún quedan retos formidables que solventar, predominando la importancia de recopilar datos de alta calidad, requiriendo procesos eficientes para convertir grandes volúmenes de complejos datos, en conocimientos significativos, pues la mera disponibilidad de los datos no implica validez de los mismos. Sin olvidar los problemas de interoperabilidad de estos, así como las implicaciones legales y cuestiones éticas que supone el uso de datos personales. No obstante, ya se manifiesta que esta tecnología a pesar de no estar completamente asentada, se encuentra en el inicio de un largo y prometedor camino.



## VII. BIBLIOGRAFÍA

1. Venkatesh R, Balasubramanian C, Kaliappan M. Development of Big Data Predictive Analytics Model for Disease Prediction using Machine learning Technique. *J Med Syst.* 2019;43(8):272.
2. Okuyan E. Big Data in Cardiology. *Anatol J Cardiol.* 2019;22(2):23-24.
3. Gandomi A. Beyond the hype: Big data concepts, methods, and analytics. *Int J Inf Manag.* 2015;35(2):137-144.
4. Ristevski B, Chen M. Big Data Analytics in Medicine and Healthcare. *J Integr Bioinform.* 2018;15(3):20170030.
5. Hemingway H, Asselbergs FW, Danesh J, Dobson R, Maniadakis N, Maggioni A, et al. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. *Eur Heart J.* 2018;39(16):1481-1495.
6. Martin-Sanchez FJ, Aguiar-Pulido V, Lopez-Campos GH, Peek N, Sacchi L. Secondary Use and Analysis of Big Data Collected for Patient Care. *Yearb Med Inform.* 2017;26(1):28-37.
7. Adibuzzaman M, DeLaurentis P, Hill J, Benneyworth BD. Big data in healthcare - the promises, challenges and opportunities from a research perspective: A case study with a model database. *AMIA Annu Symp Proc.* 2018;2017(16):384-392.
8. Mehta N, Pandit A. Concurrence of big data analytics and healthcare: A systematic review. *Int J Med Inform.* 2018;114:57-65.
9. Organizing Committee of the Madrid 2017 Critical Care Datathon, Núñez-Reiz A, Martínez-Sagasti F, Álvarez-González M, Blesa-Malpica A, Martín-Benítez JC, et al. Big data and machine learning in critical care: Opportunities for collaborative research. *Med Intensiva.* 2019;43(1):52-57.
10. Venkatesh R, Balasubramanian C, Kaliappan M. Development of Big Data Predictive Analytics Model for Disease Prediction using Machine learning Technique. *J Med Syst.* 2019;43(8):272.
11. Andreu-Perez J, Poon CC, Merrifield RD, Wong ST, Yang GZ. Big data for health. *IEEE J Biomed Health Inform.* 2015;19(4):1193-208.
12. Ienca M, Vayena E. On the responsible use of digital data to tackle the COVID-19 pandemic. *Nat Med.* 2020;26(4):463-464.
13. Harper EM, Parkerson S. Powering Big Data for Nursing Through Partnership. *Nurs Adm Q.* 2015;39(4):319-24.
14. López-Messa JB. Big data en la asistencia sanitaria (II). REMI. [Internet]. 2017 [citado

- 27 mayo 2020];233:1578-7710. Recuperado a partir de: <http://medicina-intensiva.com>.
15. Sanchez-Pinto LN, Luo Y, Churpek MM. Big Data and Data Science in Critical Care. *Chest*. 2018;154(5):1239-1248.
  16. Krittanawong C, Zhang H, Wang Z, Aydar M, Kitai T. Artificial Intelligence in Precision Cardiovascular Medicine. *JACC*. 2017;69(21):2657-2664.
  17. Kwon JM, Kim KH, Jeon KH, Lee SE, Lee HY, Cho HJ, et al. Kwon, Joon-Myoung et al. Artificial intelligence algorithm for predicting mortality of patients with acute heart failure. *PloS one*.2019;14(7):e0219302.
  18. Savana - Transform the free text of your clinical records into Big Data [Internet]. 2020 [citado 27 Mayo 2020]. Recuperado a partir de: <https://www.savanamed.com/>.
  19. Quispe-Juli C, Vela-Anton P, Meza-Rodriguez M, Moquillaza-Alcántara V. COVID-19: A pandemic in the era of digital health. *Scientific Electronic Library Online*. 2020.
  20. Kamel Boulos MN, Geraghty EM. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. *Int J Health Geogr*. 2020;19(1):8.
  21. Li B, Li J, Jiang Y, Lan X. Experience and reflection from China's Xiangya medical big data Project. *J Biomed Inform*. 2019;93:103149.
  22. Riba M, Sala C, Toniolo D, Tonon G. Big Data in Medicine, the Present and Hopefully the Future. *Front Med (Lausanne)*. 2019;6:263.
  23. DIFUTURE – Data Integration for Future Medicine [Internet]. 2020 [citado 27 Mayo 2020]. Recuperado a partir de: <https://difuture.de/>.
  24. Ciliberto G. Alleanza contro il cancro. [Internet]. Istituto Nazionale Tumori Regina Elena.Regina. 2020 [citado 27 Mayo 2020]. Recuperado a partir de:<https://www.alleanzacontroilcancro.it/istituto/istituto-nazionale-tumori-regina-elena/>.
  25. Ristevski B, Chen M. Big Data Analytics in Medicine and Healthcare. *J Integr Bioinform*. 2018;15(3):20170030.