

Oportunidades y limitaciones para el estudio estadístico de datos de historias clínicas en investigación y salud.

Stefano Poggi¹,
Natalia Garay² and Franco Simini²

¹Estudiante del XXXIII Seminario de Ingeniería Biomédica 2024.

²Docentes del XXXIII Seminario de Ingeniería Biomédica 2024.

Abstract— In an increasingly technical world, where technology and science advance at a dizzying pace, health sciences are not an exception. The advance in molecular biology techniques, imaging and medical procedures, with the consequent transition towards electronic medical records, have revolutionized medicine and propose a fertile field for the incorporation of data science and the massive use of these, in what we know as “Big Data”. At the same time, the recent advance of artificial intelligence opens an unprecedented door for the use of this data in clinical practice. However, the emergence of these new possibilities also brings with it some threats, limitations and challenges, since it involves the use of sensitive data such as those contained within medical records. With no doubt, the use of these tools can offer endless benefits for patients, as well as structural and logistical benefits for the health system, as long as we can have research ethics committees formed in this area and enough infrastructure to give guarantees to users regarding the processing of their data, anonymization and fines thereof.

Resumen— En un mundo cada vez más tecnificado, donde la tecnología y la ciencia avanzan a un ritmo vertiginoso, las ciencias de la salud no son una excepción. El avance en técnicas de biología molecular, imagenología y procedimientos médicos, con la consecuente transición hacia la historia clínica electrónica, han revolucionado la medicina y proponen un campo fértil para la incorporación de la ciencia de datos y la utilización masiva de estos, en lo que conocemos como “Big Data”. A su vez, el reciente avance de la inteligencia artificial, abre una puerta sin precedentes para la utilización de estos datos en la práctica clínica. Sin embargo, la aparición de estas nuevas posibilidades trae consigo también algunas amenazas, limitaciones y retos, ya que se trata de la utilización de datos sensibles como son los contenidos dentro de las historias clínicas. Sin dudas el uso de estas herramientas puede ofrecer un sinnúmero de beneficios para los pacientes, así como también beneficios estructurales y logísticos para el sistema sanitario, siempre y cuando, podamos contar con comités de ética de la investigación formados en esta área, y la infraestructura suficiente para dar garantías a los usuarios en cuanto a el tratamiento de sus datos, la anonimización y los fines de los mismos.

Palabras clave— Historia clínica, Big Data, Inteligencia artificial, Bioética, Comités de ética de la investigación (CEI), Datificación, Sistema sanitario.

INTRODUCCIÓN

La tecnificación y la datificación del mundo han influido significativamente en diversas áreas de nuestra vida y el conocimiento humano. La posibilidad de almacenar y utilizar grandes cantidades de datos plantea un nuevo paradigma en lo que refiere a la investigación médica. Ya sea desde el diagnóstico a través de imágenes con la utilización de inteligencia artificial, hasta la medicina de precisión; la utilización de Big Data e inteligencia artificial es un tema cada vez más presente en las discusiones sobre el futuro de la medicina.

Es importante, por lo tanto, saber cuales son las oportunidades que el uso de Big Data puede ofrecernos, así como también debatir sobre los diferentes aspectos éticos que atraviesan la utilización masiva de datos personales en el área de la salud.

METODOLOGÍA

En cuanto a la metodología utilizada para la elaboración de la monografía, realice una búsqueda bibliográfica de trabajos académicos relacionados con la utilización de Big Data en medicina, la integración con tecnologías de inteligencia artificial, así como también artículos de revistas de bioética sobre los aspectos teóricos y metodológicos pertinentes para trabajar en proyectos que involucren datos de historias clínicas. Para ello utilice plataformas como

Comentado [1]: Reducir el lenguaje superabundante, redundante, En el abstract tiene que haber alguna oportunidad y alguna limitación claramente establecida.

Comentado [2]: Se realizó

SciELO y Google Scholar. Además realice una búsqueda sobre los aspectos legislativos que rigen en nuestro país para la utilización de datos sensibles como lo son los contenidos en las historias clínicas.

RESULTADOS

OPORTUNIDADES DEL USO DE BIG DATA EN SALUD

Para aproximarnos a las oportunidades que la utilización de Big Data puede tener para las ciencias de la salud, es importante definir a que nos referimos cuando hablamos de Big data. Según la RAE el término “Big Data” hace referencia al “Conjunto de técnicas que permiten analizar, procesar y gestionar conjuntos de datos extremadamente grandes que pueden ser analizados informáticamente para revelar patrones, tendencias y asociaciones...” [1]. El diagnóstico a través de imágenes médicas, la medicina de precisión y la incorporación del internet de las cosas a la medicina, son algunos de los múltiples campos en los que la utilización de estas tecnologías podría ser útil en el área sanitaria [2]. En cuanto al uso de Big Data e inteligencia artificial para el diagnóstico imagenológico, uno de los ejemplos más recientes en la escena global es el proyecto aDELAIDE [3], que cuenta con apoyo de la Unión Europea y tiene como objetivo la utilización de Big Data e inteligencia artificial para el diagnóstico precoz de enfermedades, sobre todo en cáncer, donde la posibilidad de realizar un diagnóstico precoz puede mejorar la supervivencia de los pacientes [4]. En este caso se busca unificar la información de millones de imágenes médicas y la información de los informes clínicos de esas mismas imágenes, para entrenar algoritmos que puedan realizar diagnósticos más precisos y precoces [5].

Por otro lado, la “medicina de precisión” es probablemente el área más prometedora en cuanto al uso de estas tecnologías. Esta “nueva medicina” nació junto con el fin del proyecto genoma humano en el año 2003 [6], el cual tenía como objetivo secuenciar el genoma humano completo por primera vez. El avance en técnicas de secuenciación del ADN, y la enorme reducción de los costos de estas técnicas en los siguientes años, impulsaron aún más este concepto, en el cual se busca individualizar a cada paciente ya no solo a partir de su estilo de vida y su ambiente, sino a partir de su genoma. El principal campo de aplicación de la genómica dentro de la medicina es la oncología. El cáncer es una enfermedad predominantemente genética, en la cual suceden alteraciones en la forma en que las células crecen y proliferan. Con los avances en la secuenciación genómica se han descubierto aberraciones y mutaciones involucradas en el desarrollo del cáncer, y de esta forma se busca la elaboración y detección de perfiles genómicos del cáncer. Sin embargo, con la incorporación de Big Data e inteligencia artificial se puede adquirir información potencialmente relevante a partir de grandes datos de secuenciación de ADN, con algoritmos de aprendizaje automático. Estas herramientas se pueden usar no solo para el diagnóstico, sino también para la clasificación y estadificación de tumores, y la evaluación de posibles dianas terapéuticas [7].

Comentado [3]: Que esta oportunidad quede claramente definida como El uso de esta la big data constituida por las historias clínicas

ASPECTOS LEGALES Y ÉTICOS DEL USO DE BIG DATA EN SALUD

A – Aspectos legales

Para acercarnos a los aspectos que refieren al tratamiento de datos masivos en salud, es importante comprender aspectos básicos de la normativa nacional al respecto. En Uruguay los datos comprendidos en las historias clínicas forman parte de lo que se denominan “datos sensibles” y están legislados por la ley 18.331. Los datos sensibles forman parte de un conjunto de datos especialmente protegidos, y como tales solamente pueden ser utilizados con el consentimiento escrito y expreso del titular de esos datos. En el caso de las historias clínicas, cada paciente es dueño de la información que contiene, y por lo tanto su utilización solamente es posible mediante el consentimiento informado de el mismo, el cual para ser válido debe cumplir con tres condiciones: información adecuada, ausencia de coerción y la plena competencia de la persona para consentir [8]. La ley también prevé la utilización de estos datos con fines estadísticos y científicos, siempre que disocien los titulares de sus datos [8].

B- Aspectos éticos

Los aportes que la ciencia de datos y la inteligencia artificial pueden realizar en el campo de la salud son muy alentadores, sin embargo, debemos considerar algunos aspectos éticos de relevancia dada la sensibilidad de los datos que se utilizan en este tipo de investigaciones, que, si bien pueden ser utilizados de manera noble y con fines científicos, también existe la posibilidad de que sean utilizados con otros fines, y que no representen un valor social.

Una de las principales amenazas para este tipo de investigaciones tiene que ver con la utilización de los datos de historias clínicas con fines económicos, así como también su comercialización. Cuando se investiga con Big Data en salud, coexisten diversos agentes con distintos intereses en un mismo conjunto de datos, y generalmente estos datos son externalizados en algún momento, por lo que también puede existir las transferencias de estos datos. Por esta razón es muy importante al principio de una investigación, la identificación de los componentes que van a utilizar los datos, así como también los objetivos de los investigadores [9].

Otro aspecto importante, es lograr identificar el fin de la investigación, y que el mismo conste con un valor social. Desde un punto de vista ético, es imprescindible que la investigación cuente con este valor, y van a ser los comités de ética de la investigación los que lo evalúen. Es decir, de determinar que el valor social de la investigación sea tal, y que no sea una forma de utilizar la investigación con otros fines [9].

Por último, como vimos anteriormente, la disociación de los titulares de los datos es una condición muy importante en este tipo de investigación. Esto quiere decir que los datos deben ser anonimizados para salvaguardar la identidad de los pacientes, y evitar la reidentificación de los mismos [9]. Una teoría ética de relevancia para aplicar en estos tipos de investigación es el de “Bioética de la protección”. La Bioética de protección es una teoría de origen latinoamericano, que tiene como origen la realidad de un continente marcado por grandes desigualdades sociales y económicas, en el que la asimetría en las relaciones de poder hace necesario tener en cuenta la inequidad en los procesos de toma de decisiones en salud pública. Esta teoría parte de la base de que la protección es el fin principal del *ethos*, y por lo tanto se deben salvaguardar los derechos y los intereses de los individuos, más aún cuando la asimetría de poder puede amenazar la autonomía de los individuos.

Debemos tener en cuenta que muchas de las investigaciones que se realizan con datos sensibles pueden involucrar poblaciones vulnerables, y por lo tanto el análisis bioético por parte de los comités de ética de la investigación es fundamental, y debe admitir una perspectiva desde la bioética de la protección [10], [11].

Comentado [4]: Como se detalló anteriormente

EVALUACIÓN Y MÉTODOS. CEI.

En cualquier investigación en la que participen seres humanos, los aspectos metodológicos, legales y éticos deben estar a cargo de un Comité de Ética de la Investigación (CEI), con el objetivo de proteger los derechos de los participantes, así como sus datos. Es importante, por lo tanto, tener en cuenta algunas nociones básicas acerca de los CEI en relación con el uso de Big Data.

Los CEI van a exigir a los investigadores, pero también deben procurar contar con los recursos humanos y técnicos para dar seguimiento a este tipo de proyectos. Una de las principales exigencias para con los investigadores es la explicitación de las políticas de privacidad que van a regir durante la investigación, las cuales el CEI debe estudiar para analizar de qué forma se van a cumplir esas políticas a lo largo de todo el proceso. Esto incluye saber que datos se van a utilizar específicamente, dónde y por cuánto tiempo se van a almacenar esos datos, y sobre todo es importante saber quienes van a tener acceso a esos datos para su procesamiento. Además, se debe explicitar cuál es el objetivo de la investigación, para analizar los riesgos y beneficios de la misma.

Un desafío que tienen los CEI en estos casos es el de asegurar la trazabilidad y seguridad de los datos de los participantes, y estos deben saber que existe el riesgo de reidentificación, aunque el objetivo es que ese riesgo sea muy bajo. Para ello, los CEI deben estar integrados por expertos en ciencia de datos que analicen los mecanismos de anonimización y todo lo referente a la protección de los datos personales.

Para realizar investigaciones en las que se lleva a cabo el análisis masivo de datos, se debe contar con nubes y sistemas de supercomputación, que idealmente no deberían externalizarse, es decir que el cruce de datos referentes a historias clínicas de pacientes se den al margen de las instituciones sanitarias en las que estos datos se encuentran. Por último, cabe mencionar que los profesionales que trabajan con datos deben cumplir con la confidencialidad [12].

CONCLUSIONES

La irrupción del Big Data y la inteligencia artificial, así como la reducción de costos en técnicas de secuenciación de ADN e imágenes médicas, han generado la posibilidad de producir y analizar grandes cantidades de datos de relevancia clínica, y han llevado a la generación de la medicina de precisión, creando un nuevo paradigma en el campo de la medicina.

Este tipo de investigación requiere de una infraestructura y recursos humanos capacitados en ciencias de datos, para contar con CEI que ofrezcan garantías en el análisis ético y técnico de los proyectos, ponderando riesgos y beneficios, y procurando el valor social de la investigación y la protección de los datos de los participantes. De esta forma es posible utilizar la gran cantidad de datos de los que se dispone, ofreciendo garantías a los pacientes y generando ciencia con valor social.

REFERENCIAS

(ESTILO IEEE CON CORCHETES CUADROS, MANEJADO POR EL MENDELEY U OTRO "REFERENCE MANAGER")

- [1] RAE (2023) *Definición de big data - Diccionario Panhispánico del español jurídico - rae, Diccionario panhispánico del español jurídico - Real Academia Española*. Available at: <https://dpej.rae.es/lema/big-data> (Accessed: 07 July 2024).
- [2] Menasalvas, E., Consuelo, G. and Rodriguez-González, A. (2017) *Big Data en Salud: RETOS y oportunidades*. Available at: https://www.mintur.gob.es/Publicaciones/Publicacionesperiodicas/EconomiaIndustrial/RevistaEconomiaIndustrial/405/menasalvas_gonzalo_y_rodriguez.pdf (Accessed: 07 July 2024).
- [3] Radiología cognitiva para proyectos big data y ensayos clínicos 'in silico' (2022) Adelaide. Available at: <http://adelaide-project.ai/es/> (Accessed: 11 July 2024).
- [4] Fuentes-Rojas, J. (2019) Importancia de la Prevención y el diagnóstico precoz en el cáncer mamario, *Revista Archivo Médico de Camagüey*. Available at: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1025-02552019000100004 (Accessed: 11 July 2024).
- [5] *Inteligencia artificial y Big Data aplicados a la radiología, base del proyecto adelaide: Publicación de Sanidad Privada* (no date) *Inteligencia artificial y Big Data aplicados a la Radiología, base del proyecto aDELAIDE | Publicación de Sanidad Privada*. Available at: <https://sanidadprivada.publicacionmedica.com/noticia/inteligencia-artificial-y-big-data-aplicados-a-la-radiologia-base-del-proyecto-adelaide> (Accessed: 07 July 2024).
- [6] Reguero , M.T. (2013) 10 años de la secuenciación del genoma humano. Available at: <http://www.scielo.org.co/pdf/biote/v15n1/v15n1a1.pdf> (Accessed: 11 July 2024).
- [7] Dlamini, Z. et al. (2020) *Artificial Intelligence (AI) and Big Data in cancer and precision oncology, Computational and Structural Biotechnology Journal*. Available at: <https://www.sciencedirect.com/science/article/pii/S200103702030372X> (Accessed: 07 July 2024).
- [8] Vázquez , T. et al. (2008) *Ley de protección de datos personales N° 18.331, Impo*. Available at: <https://www.imo.com.uy/bases/leyes/18331-2008> (Accessed: 07 July 2024).
- [9] Leucona, I. (2018) *Evaluación de los Aspectos Metodológicos, éticos, legales y sociales de proyectos de investigación en salud con datos masivos (big data), Gaceta Sanitaria*. Available at: <https://www.sciencedirect.com/science/article/pii/S0213911118300864> (Accessed: 07 July 2024).
- [10] Rego, S. and Gorini, C. (2023) *Aspectos éticos del Uso del Big Data, EIDON. Revista española de bioética*. Available at: <https://www.revistaeidon.es/index.php/revistaeidon/article/view/206> (Accessed: 07 July 2024).
- [11] Rodrigues, C.A.B. and Schramm, F.R. (2022) *Bioética de Proteção: Fundamentos e perspectiva, Revista Bioética*. Available at: <https://www.scielo.br/j/bioet/a/nzCPK7SLsb9R9hNyyMzdfdf/> (Accessed: 07 July 2024).
- [12] Lecuona, I. de (2020) *Aspectos éticos, legales y sociales del uso de la inteligencia artificial y el big data en salud en un contexto de pandemia, Revista Internacional de Pensamiento Político*. Available at: <https://www.upo.es/revistas/index.php/ripp/article/view/5599> (Accessed: 07 July 2024).