

DOI: 10.25237/revchilanestv5129061641

Inteligencia artificial en medicina: Selección de métodos, aplicaciones y consideraciones (Parte II)

Artificial Intelligence in medicine: Methods selection, applications and considerations (Part II)

Daniela Biggs², Matías Vargas³, Trinidad Larraín⁴, Alexis Alvear⁵, Juan C. Pedemonte^{1,*}.

¹ Profesor Clínico Asistente, División de Anestesiología, Escuela de Medicina, Pontificia Universidad Católica de Chile. Programa de Farmacología y Toxicología, Facultad de Medicina, Pontificia Universidad Católica de Chile. Santiago, Chile.

² Residente de Anestesiología, División de Anestesiología, Escuela de Medicina, Pontificia Universidad Católica de Chile. Santiago, Chile.

³ Interno de Medicina, Escuela de Medicina, Pontificia Universidad Católica de Chile. Santiago, Chile.

⁴ Alumna de Medicina, Escuela de Medicina, Pontificia Universidad Católica de Chile. Santiago, Chile.

⁵ Profesor Asistente, Facultad de Matemáticas, Pontificia Universidad Católica de Chile. Director Ejecutivo de DATA UC, Director del Diplomado académico en Data Science UC, Pontificia Universidad Católica de Chile. Santiago, Chile.

Financiamiento: Los autores no recibieron financiamiento al realizar este trabajo. Los autores declaran no poseer conflictos de interés.

Fecha de recepción: 05 de marzo de 2022 / Fecha de aceptación: 12 de marzo de 2022

ABSTRACT

The spreading of artificial intelligence and machine learning (ML) methods in different healthcare areas is common. The second part of this review article describes the foresights when selecting different ML methods. It also presents an updated review of examples and the precautions or challenges that we will face in the future when using these technologies. We will describe how to know whether to use a descriptive or predictive approach, the characteristics of these methods and their potential applications. Later, we will discuss how the irruption of digital data, together with freely available algorithms and greater computational power, has made it possible to enhance the implementation of these models in medicine. We will review how ML has contributed to the development of diagnostic imaging, as well as the prediction of monitoring and clinical outcomes. Finally, we will analyze the challenges and ethical considerations associated with the implementation of ML in clinical practice.

Key words: Artificial intelligence, machine learning, medicine, algorithms.

RESUMEN

La difusión de los métodos de inteligencia artificial y *machine learning* (ML) en diversas áreas de la salud es transversal. La segunda parte de este artículo de revisión describe las consideraciones que se deben tener al seleccionar distintos métodos de ML. Además, presenta una revisión actualizada de ejemplos de su uso y de las precauciones o desafíos a los que nos veremos enfrentados en el futuro al utilizar estas tecnologías. Describiremos cómo saber si utilizar un enfoque descriptivo o predictivo, las características de estas aproximaciones y sus potenciales aplicaciones. Posteriormente, discutiremos cómo la irrupción de datos digitales, en conjunto con algoritmos de libre disposición y mayor poder computacional, ha permitido potenciar la implementación de estos modelos en medicina. Revisaremos como el ML ha contribuido en el desarrollo del diagnóstico por imágenes, como también en la predicción de monitorización y desenlaces clínicos. Finalmente, analizaremos los desafíos y consideraciones éticas asociadas a la implementación del ML en la práctica clínica.

Palabras clave: Inteligencia artificial, *machine learning*, medicina, algoritmos.

jcpedemo@gmail.com

*ORCID: <https://orcid.org/0000-0001-5729-065X>

Introducción

En la primera sección de estos artículos de revisión exploramos los principales métodos utilizados para modelar algoritmos de *machine learning* (ML). En esta segunda parte revisaremos las consideraciones que se deben tener al seleccionar distintos métodos de ML. Además, discutiremos ejemplos de su uso en medicina y analizaremos las precauciones o desafíos a los que nos veremos enfrentados al utilizar estas tecnologías en el futuro.

Criterios de selección de modelos de *Machine Learning*

La técnica de ML a seleccionar para determinado análisis va a depender del tipo de problema que deseamos analizar, de la estructura de los datos, de la complejidad y de la escalabilidad que deba tener el modelo. Desde un punto de vista teórico, se puede simplificar el criterio de selección clasificando la tarea que queremos realizar en dos enfoques, dependiendo si el problema es de descripción o de predicción. El enfoque descriptivo corresponde al análisis de datos buscando patrones, relaciones, tendencias y valores atípicos. El enfoque predictivo, en cambio, busca la construcción de modelos de regresión, clasificación, reconocimiento de patrones y evaluación de la precisión predictiva.

En términos de aprendizaje, estos enfoques se dividen en aprendizaje no supervisado (descriptivo) y supervisado (predictivo), descritos en la primera parte de esta revisión. De esta manera, las técnicas de ML se pueden comprender bajo una lógica de "caja de herramientas", donde, dependiendo del objetivo (describir o predecir), se recurre a los diferentes modelos existentes, seleccionándolos de acuerdo con el nivel de complejidad del análisis, estructura de datos, escalabilidad, entre otras consideraciones.

Incorporación de modelos de *Machine Learning* a medicina

En Chile, se han implementado sistemas de fichas clínicas electrónicas durante los últimos 15 años. Estos permiten aumentar la velocidad de obtención de datos clínicos y mejorar la integración de los sistemas de Inteligencia Artificial (IA) en la labor clínica. En EE. UU., aumentó desde 13% de acceso a sistemas básicos de registro electrónicos de salud en el año 2008 a 72% a fines del año 2012[1],[2]. Estos sistemas permiten centralizar la información clínica de los pacientes, incluyendo historias clínicas, valores de laboratorio y estudios de imágenes. Esto, sumado a los avances en los métodos de ML, la existencia de paquetes de código libre y mayor poder computacional, han permitido un crecimiento exponencial en la IA en medicina. Hasta la fecha, diversos estudios han demostrado que la IA puede tener niveles similares o superiores a médicos especialistas en diagnósticos de ciertas patologías, pueden predecir mejor el pronóstico de pacientes que clínicos y pueden asistir en intervenciones quirúrgicas[3]. Debido a que ML va a seguir progresando, se espera que la IA revolucione la práctica clínica redefiniendo los roles de los clínicos en este proceso.

La incorporación del ML a la práctica clínica puede desarrollarse acorde a diferentes modelos[4]:

- *Modelo clásico*: el clínico consigue información de los pacientes, toma decisiones clínicas usando sus juicios y escribe sus hallazgos en las fichas clínicas electrónicas.
- *Sistemas convencionales de apoyo a la toma de decisiones*: se ejecutan algoritmos utilizando los hallazgos de las fichas clínicas electrónicas (basado en reglas o ML) y se sugiere conducta al clínico, quien realiza la decisión final.
- *Sistemas integrativos de apoyo a la toma de decisiones*: consisten en algoritmos que pueden solicitar información clínica relevante u obtener información de fichas clínicas electrónicas. Muestran resultados a los clínicos y lo registran directamente en la ficha, pero son los clínicos quienes realizan la decisión final respecto a cómo se utiliza esta información.
- *Sistema clínico totalmente automatizado*: el sistema automáticamente recolecta información de los pacientes, toma decisiones y escribe los resultados en fichas clínicas electrónicas.

Además, existen diversos grados de integración práctica en los que se puede utilizar la IA[4]:

- IA es más confiable que experto (clínico), por lo que se debería relegar la tarea a IA. En este grupo se encuentran tareas como analizar suero y sistemas de alertas de interacción de drogas.
- IA se desempeña al mismo nivel que expertos/especialistas, por lo que IA debería desarrollar la mayoría de la tarea, pero los clínicos confirman su ejecución.
- IA posee desempeño razonable, por lo que desarrolla parte de la tarea, pero los clínicos deben completar y confirmar.
- IA no posee buen desempeño, por lo que clínicos deben desarrollar la tarea, con IA asistiendo en tareas menores en segundo plano.
- IA no debería participar por naturaleza de la relación médico-paciente, por lo que se mantiene la relación clínica actual.

Desarrollo de *Machine Learning* en diagnóstico por imágenes

El uso de IA ha presentado mayor éxito en los diagnósticos por imágenes, donde destaca la implementación de técnicas de aprendizaje profundo. A continuación, presentamos algunos ejemplos aplicados a diversas especialidades médicas (Tabla 1).

Cabe destacar que varios de los modelos de ML mencionados tienen rendimientos comparables a médicos especialistas. Esto permitiría solucionar la demanda de atención médica especializada, como por ejemplo la población de diabéticos en riesgo de desarrollar retinopatía diabética, ya que la fotografía de fondo de ojo suele ser interpretadas por médicos oftalmólogos. Tanto los modelos mencionados, como otros sistemas que se encuentran en el mercado, pueden ser posibles soluciones a esta dificultad, ya que son considerados alternativas rentables a la calificación manual[27].

Múltiples estudios señalan que los clínicos se ven beneficiados al utilizar algoritmos de ML, especialmente aquellos basados en aprendizaje profundo y redes neuronales convolucionales, como un segundo lector de imágenes, debido que se acelera el proceso diagnóstico y aumenta la sensibilidad[8]-[10],[14],[17],[18],[20],[28]. Además, médicos no especialistas pueden beneficiarse de contar con un sistema que

Tabla 1. Diagnóstico por imágenes mediante *Machine Learning* en algunas especialidades médicas

Especialidad	Métodos de ML	Ejemplo
Radiología	Aprendizaje profundo	Radiografía de tórax posteroanterior: búsqueda de nódulos pulmonares, tuberculosis pulmonar, fibrosis quística, neumoconiosis y localización de catéteres centrales insertados periféricamente[1] y diagnóstico de neumoconiosis[2], tuberculosis o nódulos pulmonares[3],[4] Radiografía: diagnóstico de fracturas[5]-[7] Tomografía computada: tamizaje de hemorragia intracerebral y cáncer pulmonar[8], predicción de oclusión de grandes vasos en pacientes con sospecha de accidente cerebrovascular[9].
Dermatología	Aprendizaje profundo (redes neuronales convolucionales)	Diagnóstico de carcinomas de queratinocitos, queratosis seborreicas benignas, melanomas malignos y nevos benignos[10] Detección de cáncer de piel no melanocíticos y no pigmentados[11]
Oftalmología	Aprendizaje profundo	Fondo de ojo: detección de retinopatía diabética no proliferativa moderada o de peor grado en pacientes adultos con diabetes[12],[13] y de degeneración macular asociada a la edad[14],[14] Escáneres de tomografía de coherencia óptica: detección de degeneración macular asociada a la óptica[15]
Anatomía patológica	Aprendizaje profundo (redes neuronales convolucionales)	Diagnóstico de metástasis de cáncer de mama en biopsias de linfonodos[16],[17]
Gastroenterología	Aprendizaje profundo (redes neuronales convolucionales)	Imágenes de endoscopia: diagnóstico de infección por <i>Helicobacter pylori</i> [18], de cáncer gástrico[19] y tamizaje de carcinoma esofágico celular escamoso[20] Colonoscopia: diferenciar en tiempo real entre pólipos colorrectales adenomatosos e hiperplásicos[21]
	Máquinas de vector soporte	Colonoscopia: evaluar pólipos de pequeño tamaño en tiempo real[22]

Referencias Tabla 1.

- Singh R, Kalra MK, Nitiwarangkul C y cols. Deep learning in chest radiography: Detection of findings and presence of change. PLOS ONE. 2018;13(10):e0204155.
- Zhang L, Rong R, Li Q y cols. A deep learning-based model for screening and staging pneumoconiosis. Sci Rep. 2021;11(1):2201.
- Rajpurkar P, Irvin J, Ball RL y cols. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLOS Medicine. 2018;15(11):e1002686.
- Nam JG, Park S, Hwang EJ y cols. Development and Validation of Deep Learning-based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs. Radiology. 2019;290(1):218-28.
- Cheng C-T, Chen C-C, Cheng F-J y cols. A Human-Algorithm Integration System for Hip Fracture Detection on Plain Radiography: System Development and Validation Study. JMIR medical informatics. 2020;8(11):e19416-e.
- Kitamura G. Deep learning evaluation of pelvic radiographs for position, hardware presence, and fracture detection. Eur J Radiol. 2020;130:109139.
- Lindsey R, Daluiski A, Chopra S y cols. Deep neural network improves fracture detection by clinicians. Proc Natl Acad Sci U S A. 2018;115(45):11591-6.
- Ardila D, Kiraly AP, Bharadwaj S y cols. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med. 2019;25(6):954-61.
- Olive-Gadea M, Crespo C, Granes C y cols. Deep Learning Based Software to Identify Large Vessel Occlusion on Noncontrast Computed Tomography. Stroke. 2020;51(10):3133-7.
- Esteva A, Kuprel B, Novoa RA y cols. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542(7639):115-8.
- Tschandl P, Rosendahl C, Akay BN y cols. Expert-Level Diagnosis of Nonpigmented Skin Cancer by Combined Convolutional Neural Networks. JAMA dermatology. 2019;155(1):58-65.
- Ting DSW, Cheung CY, Lim G y cols. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. Jama. 2017;318(22):2211-23.
- Gulshan V, Peng L, Coram M y cols. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. Jama. 2016;316(22):2402-10.
- Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM. Automated Grading of Age-Related Macular Degeneration From Color Fundus Images Using Deep Convolutional Neural Networks. JAMA Ophthalmol. 2017;135(11):1170-6.
- Lee CS, Baughman DM, Lee AY. Deep Learning Is Effective for Classifying Normal versus Age-Related Macular Degeneration OCT Images. Ophthalmology Retina. 2017;1(4):322-7.
- Steiner DF, MacDonald R, Liu Y y cols. Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes for Metastatic Breast Cancer. Am J Surg Pathol. 2018;42(12):1636-46.
- Ehteshami Bejnordi B, Veta M, Johannes van Diest P y cols. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA. 2017;318(22):2199-210.
- Shichijo S, Nomura S, Aoyama K y cols. Application of Convolutional Neural Networks in the Diagnosis of Helicobacter pylori Infection Based on Endoscopic Images. EBioMedicine. 2017;25:106-11.
- Hirasawa T, Aoyama K, Tanimoto T y cols. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. Gastric Cancer. 2018;21(4):653-60.
- Cai SL, Li B, Tan WM y cols. Using a deep learning system in endoscopy for screening of early esophageal squamous cell carcinoma (with video). Gastrointest Endosc. 2019;90(5):745-53.e2.
- Byrne MF, Chapados N, Soudan F y cols. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. Gut. 2019;68(1):94.
- Mori Y, Kudo SE, Misawa M y cols. Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy: A Prospective Study. Ann Intern Med. 2018;169(6):357-66.

tenga rendimiento similar al de especialistas para corroborar diagnósticos. Sin embargo, el desempeño de estos modelos depende de la calidad de las imágenes utilizadas para su entrenamiento y a la prolijidad de sus etiquetas, tal como se demuestra cuando la información inicial no es de buena calidad, por lo que se consiguen resultados inferiores a los especialistas[29].

Además, diversas áreas en las que imágenes o videos registrados en tiempo real como cirugías laparoscópicas, endoscopías o similares podrían ser evaluados mediante *Multiple Instance Learning*, una forma de ML supervisada. Permite detectar automáticamente patrones de objetivos localmente en imágenes o vídeos y proponer diagnósticos automáticos[30]. Además, se podrían complementar con otros modelos de ML, como algunos basados en aprendizaje profundo, con el fin de remover el ruido de las imágenes en tiempo real para así obtener resultados más precisos, como ya se han obtenido para angiografía[31] o tomografías de coherencia óptica[32].

Predicción de monitorización y desenlaces

Los algoritmos de ML se caracterizan por presentar una excelente capacidad predictiva. Esta característica ha sido aplicada en diversas áreas médicas (Tabla 2).

Desafíos y consideraciones asociadas al desarrollo del *Machine Learning*

Dificultad en la reproducibilidad

Los modelos de ML tienen un enorme número de parámetros que deben ser aprendidos a partir de los datos o fijados manualmente por el analista. Muchas decisiones se toman "silenciosamente" a través de los parámetros por defecto que una determinada biblioteca de *software* ha preseleccionado. Estos valores por defecto pueden diferir entre bibliotecas y pueden variar de una versión a otra. Así, dos investigadores utilizando el mismo código, pero diferentes versiones de una biblioteca

de *software* podrían llegar a conclusiones sustancialmente diferentes si los parámetros más relevantes tienen valores diferentes[49]. El entrenamiento de los modelos de ML hace uso de la aleatoriedad[49],[50]. La única manera de garantizar que los resultados de estos modelos sean reproducibles es establecer una cantidad o "semilla aleatoria", que controla cómo se generan los números aleatorios[49]. Un estudio demostró que el cambio de este único número, aparentemente inocuo, podía inflar el rendimiento del modelo estimado hasta 2 veces con respecto a lo que produciría un conjunto diferente de semillas aleatoria[51]. La capacidad predictiva de un modelo va a depender de los datos que hayan sido utilizados para generarlos, así un modelo que funciona en un ambiente puede no resultar útil en otro. Por esto, es fundamental realizar una validación externa de los modelos en la población objetivo previo a su uso clínico[52].

Ejemplos adversos

Cada modelo de ML estudiado es susceptible a la manipulación de sus entradas de información. Los ejemplos adversos son entradas de una máquina de ML que han sido creados intencionadamente para forzar al modelo a cometer un error. Los ataques adversos (*adversarial attacks*) es uno de tantos posibles modos de falla de los sistemas basados en ML aplicados en medicina. Por ejemplo, para un modelo basado en diagnósticos con imágenes, la aplicación de mínimos ajustes como rotar la imagen a un ángulo específico (Figura 1) puede ocasionar el diagnóstico equivocado[53].

Déficits en fichas electrónicas de salud

La información que obtienen los sistemas de ML basados en las fichas electrónicas de salud (*electronic health records, EHR*) puede ser muy heterogénea, debido a que información clínica tiende a ser desordenada, incompleta y potencialmente sesgada. Si los resultados se predicen sobre la base de mediciones, pueden surgir problemas cuando las variables cambian considerablemente. Por ejemplo, cambios en la población de

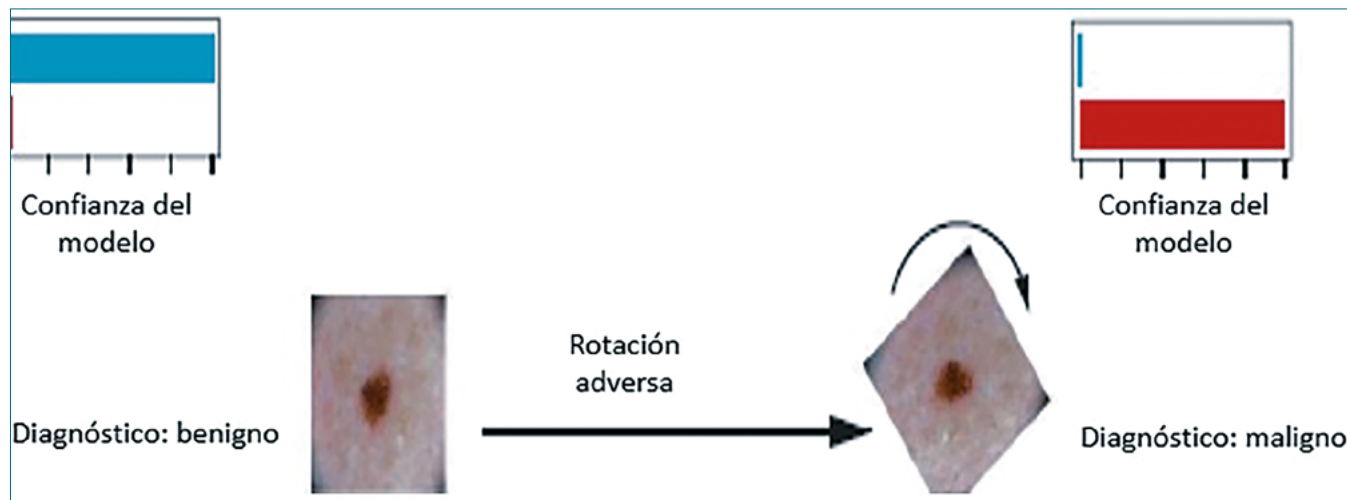


Figura 1. Ejemplo adversario. Descripción: Representación de ejemplo adversario, en que el diagnóstico de un nevo cambia de benigno a maligno mediante rotación de la imagen.

Tabla 2. Predicción mediante *Machine Learning* en algunas especialidades médicas

Especialidad	Métodos de ML	Ejemplo
Psiquiatría	Máquinas de vector soporte, aprendizaje profundo, regresión logística Bosques aleatorios, árboles de decisiones y regresión logística	Depresión: predicción de desarrollo, detección de ideación suicida y respuesta al tratamiento[1] Autismo: herramienta de tamizaje, mediante cuestionarios y videos realizados con teléfonos celulares[2],[3]
Neurología	Máquinas de vector soporte, k-vecinos más próximos y aprendizaje profundo (redes neuronales convolucionales) Árboles de decisiones	Epilepsia: detección de crisis en videos y electroencefalogramas y predecir resultados del tratamiento médico y quirúrgico epilepsia 4. Accidente cerebro vascular: predicción de mortalidad tras la rehabilitación en casos graves[5]
Ginecología y obstetricia	Bosques aleatorios y árboles de decisiones con impulso (<i>boosting</i>) Aprendizaje profundo	Predicción del peso fetal[6] Posibilidad de embarazo con corazón fetal a partir de un video de cámara rápida de la incubación y transferencia de un blastocisto[7]
Urgencias y medicina intensiva	Máquinas de vectores soporte, métodos de conjunto, redes neuronales y árboles de decisión	Detección temprana de sepsis[8],[9]
Anestesia y cirugía	Aprendizaje automático: algoritmo de escalada estocástica y redes neuronales Regresión logística, árboles de decisiones, bosques aleatorios, máquinas de vector soporte, impulso (<i>boosting</i>), combinación de bosques aleatorios con impulso y k-vecinos más próximos	Desenlaces postoperatorios: morbilidad, mortalidad y readmisión en cirugías electivas 10 y complicaciones postquirúrgicas[11] Complicaciones postoperatorias específicas: lesión renal aguda tras cirugía cardiaca[12], reingreso posoperatorio[13], hipoglicemia postprandial[14], requerimientos de opioides postoperatorios en pacientes ambulatorios[15] y prescripción prolongada de opioides después de cirugía por hernia discal lumbar[16]

Referencias Tabla 2

- Graham S, Depp C, Lee EE y cols. Artificial Intelligence for Mental Health and Mental Illnesses: an Overview. *Curr Psychiatry Rep.* 2019;21(11):116.
- Abbas H, Garberson F, Glover E, Wall DP. Machine learning approach for early detection of autism by combining questionnaire and home video screening. *J Am Med Inform Assoc.* 2018;25(8):1000-7.
- Tariq Q, Daniels J, Schwartz JN, Washington P, Kalantarian H, Wall DP. Mobile detection of autism through machine learning on home video: A development and prospective validation study. *PLOS Medicine.* 2018;15(11):e1002705.
- Abbasi B, Goldenholz DM. Machine learning applications in epilepsy. *Epilepsia.* 2019;60(10):2037-47.
- Scrutinio D, Ricciardi C, Donisi L y cols. Machine learning to predict mortality after rehabilitation among patients with severe stroke. *Scientific Reports.* 2020;10(1):20127.
- Naimi AI, Platt RW, Larkin JC. Machine Learning for Fetal Growth Prediction. *Epidemiology.* 2018;29(2):290-8.
- Tran D, Cooke S, Illingworth PJ, Gardner DK. Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer. *Hum Reprod.* 2019;34(6):1011-8.
- Goh KH, Wang L, Yeow AYK y cols. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. *Nature Communications.* 2021;12(1):711.
- Fleuren LM, Klausch TLT, Zwager CL y cols. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med.* 2020;46(3):383-400.
- Hyer JM, White S, Cloyd J y cols. Can We Improve Prediction of Adverse Surgical Outcomes? Development of a Surgical Complexity Score Using a Novel Machine Learning Technique. *Journal of the American College of Surgeons.* 2020;230(1):43-52.e1.
- El Hechi MW, Nour Eddine SA, Maurer LR, Kaafarani HMA. Leveraging interpretable machine learning algorithms to predict postoperative patient outcomes on mobile devices. *Surgery.* 2021;169(4):750-4.
- Tseng PY, Chen YT, Wang CH y cols. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit Care.* 2020;24(1):478.
- Mišić VV, Gabel E, Hofer I, Rajaram K, Mahajan A. Machine Learning Prediction of Postoperative Emergency Department Hospital Readmission. *Anesthesiology.* 2020;132(5):968-80.
- Seo W, Lee YB, Lee S, Jin SM, Park SM. A machine-learning approach to predict postprandial hypoglycemia. *BMC Med Inform Decis Mak.* 2019;19(1):210.
- Nair AA, Velagapudi MA, Lang JA y cols. Machine learning approach to predict postoperative opioid requirements in ambulatory surgery patients. *PLOS ONE.* 2020;15(7):e0236833.
- Karhade AV, Ogink PT, Thio Q y cols. Development of machine learning algorithms for prediction of prolonged opioid prescription after surgery for lumbar disc herniation. *Spine J.* 2019;19(11):1764-71.

pacientes pueden determinar predicciones distintas a las observadas previamente. Además, las etiquetas como progresión de una enfermedad son asignadas por expertos, pero no todos los expertos estarán de acuerdo en la misma etiqueta[54].

Modelos de *Machine Learning* sin incertidumbre

La incertidumbre es algo propio de la medicina. Sin embargo, múltiples modelos de ML desarrollados en medicina han sido diseñados, principalmente, para entregar resultados dicotómicos, por ejemplo, si una imagen incluye un melanoma

maligno o un nevo benigno. Esto podría ser útil en ciertos casos, pero para otros (como modelos que predicen condiciones críticas como sepsis) las estimaciones de incertidumbre serán vitales para la clasificación de los pacientes. Así, los médicos podrían centrarse en los pacientes con estimaciones con alta probabilidad de presentar condiciones críticas y examinar más a los pacientes para los que el modelo describe mayor incertidumbre con respecto a su condición actual. Los modelos de ML deberían ser diseñados con la capacidad de decir “no lo sé” y abstenerse potencialmente de dar un diagnóstico o una predicción cuando hay un exceso de incertidumbre[55].

Disminución de las habilidades de los clínicos

El fenómeno de la dependencia excesiva de la tecnología podría ser tentadora para los usuarios a corto plazo por la comodidad y la eficacia de la ayuda automatizada. Sin embargo, a largo plazo, estas herramientas pueden conducir al fenómeno de la descalificación. Esto afecta a la capacidad de los médicos en formar opiniones sobre la base de signos y síntomas detectables y datos disponibles. Una actitud prudente puede ayudar a prevenir los efectos negativos que puede tener la implementación de ML en la práctica clínica[56].

Errores de los sistemas informáticos

Dado que los modernos algoritmos de aprendizaje automático realizan complejas transformaciones matemáticas de los datos de entrada, los errores cometidos por los sistemas informáticos requerirán una vigilancia adicional para su detección e interpretación[57].

Consideraciones éticas

La IA se enfrenta a los desafíos básicos de la práctica médica: justicia, autonomía, respeto, beneficencia y no maleficencia. En todos los aspectos de ML surgen cuestiones éticas, en el diseño, la obtención de datos, la aplicación clínica y el control de calidad, los cuales deben discutirse y abordarse apropiadamente[58]. Se debe considerar que los algoritmos introducidos pueden, inadvertidamente, incorporar prejuicios discriminatorios, que reflejan sesgos propios de los datos usados para entrenarlos. Un algoritmo diseñado para predecir resultados de hallazgos genéticos, estará sesgado si los estudios genéticos han sido insuficientes en determinadas poblaciones[59]. Es por esto que si bien parte de los sesgos pueden ser corregidos durante el aprendizaje del modelo, debe haber siempre conciencia para investigar cuándo estos se pueden estar escondiendo en los datos y qué se puede hacer para mitigar su efecto[60]. La incorporación de datos únicos que tienen el mismo desenlace puede llevar a que ML no discrimine situaciones en las que variables no consideradas influyen en el resultado. Si se le indica a la herramienta que los médicos siempre retiran medidas de soporte vital en ciertas circunstancias (por ejemplo, daño cerebral o prematuridad extrema), el sistema puede concluir que tales hallazgos siempre son fatales, sin considerar todas las variables[59]. Es por esto que durante el proceso de desarrollo de ML debe haber tanto diversidad de pensamiento como de datos[58]. Por otro lado, si una herramienta está entrenada para dar recomendaciones médicas, éstas deben ser apropiadas y disponibles para un entorno específico. Dado que el uso de ML es cada vez más común en la práctica clínica, el manejo seguro de los datos de los pacientes se vuelve un tema primordial[58]. Además, con el uso de estas tecnologías, se corre el riesgo de divulgación inadvertida de información personal que atenta contra la confidencialidad médico-paciente. Se deben implementar fuertes medidas de anonimización y seguridad de información privada, incluyendo las columnas individuales de datos que pueden generar una huella digital tal, que los pacientes de-identificados en determinada base de datos, pueden ser re-identificados, proceso llamado divulgación deductiva[61]. Estos hechos demandan que, al momento de usar ML, se deben

considerar de manera primordial las implicancias éticas de la herramienta, para promover privacidad, seguridad y trato justo a los pacientes y a todas las partes afectadas.

Conclusiones

La IA y en especial, los métodos de ML, se están desarrollando cada vez más en el campo de la medicina. La obtención y almacenamiento de grandes cantidades de datos ha sido fundamental en su desarrollo. En la actualidad existen ejemplos en múltiples áreas médicas donde se demuestra la utilidad de los métodos de ML en el diagnóstico y predicción de resultados clínicos. Sin embargo, como cualquier herramienta tecnológica, estos métodos no son infalibles. Aún es necesario determinar si el uso de ML en medicina mejora los resultados y desenlaces de los pacientes. Resulta imprescindible que los médicos estén preparados e informados para poder interactuar con estos sistemas en el futuro. Independiente del desarrollo que logren estos sistemas, el conocimiento médico, asociado a una adecuada relación médico-paciente, serán las bases fundamentales del acto médico humano y compasivo.

Referencias

1. Hsiao CJ, Jha AK, King J, Patel V, Furukawa MF, Mostashari F. Office-based physicians are responding to incentives and assistance by adopting and using electronic health records. *Health Aff (Millwood)*. 2013 Aug;32(8):1470–7. <https://doi.org/10.1377/hlthaff.2013.0323> PMID:23840050
2. DesRoches CM, Campbell EG, Rao SR y cols. Electronic Health Records in Ambulatory Care - A National Survey of Physicians. *N Engl J Med*. 2008;359(1):50–60. <https://doi.org/10.1056/NEJMs0802005> PMID:18565855
3. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *N Engl J Med*. 2019 Apr;380(14):1347–58. <https://doi.org/10.1056/NEJMra1814259> PMID:30943338
4. Yu KH, Beam AL, Kohane IS. Artificial intelligence in health-care. *Nat Biomed Eng*. 2018 Oct;2(10):719–31. <https://doi.org/10.1038/s41551-018-0305-z> PMID:31015651
5. Singh R, Kalra MK, Nitiwarangkul C, Patti JA, Homayounieh F, Padole A, et al. Deep learning in chest radiography: detection of findings and presence of change. *PLoS One*. 2018 Oct;13(10):e0204155. <https://doi.org/10.1371/journal.pone.0204155> PMID:30286097
6. Zhang L, Rong R, Li Q, Yang DM, Yao B, Luo D, et al. A deep learning-based model for screening and staging pneumoconiosis. *Sci Rep*. 2021 Jan;11(1):2201. <https://doi.org/10.1038/s41598-020-77924-z> PMID:33500426
7. Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med*. 2018 Nov;15(11):e1002686. <https://doi.org/10.1371/journal.pmed.1002686> PMID:30457988
8. Nam JG, Park S, Hwang EJ, Lee JH, Jin KN, Lim KY, et al. Development and Validation of Deep Learning-based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs. *Radiology*. 2019 Jan;290(1):218–28. <https://doi.org/10.1148/radiol.2018180237> PMID:30251934

9. Cheng C-T, Chen C-C, Cheng F-J y cols. A Human-Algorithm Integration System for Hip Fracture Detection on Plain Radiography: System Development and Validation Study. *JMIR medical informatics*. 2020;8(11):e19416-e.
10. Kitamura G. Deep learning evaluation of pelvic radiographs for position, hardware presence, and fracture detection. *Eur J Radiol*. 2020 Sep;130:109139. <https://doi.org/10.1016/j.ejrad.2020.109139> PMID:32623269
11. Lindsey R, Daluiski A, Chopra S, Lachapelle A, Mozer M, Siclar S, et al. Deep neural network improves fracture detection by clinicians. *Proc Natl Acad Sci USA*. 2018 Nov;115(45):11591–6. <https://doi.org/10.1073/pnas.1806905115> PMID:30348771
12. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med*. 2019 Jun;25(6):954–61. <https://doi.org/10.1038/s41591-019-0447-x> PMID:31110349
13. Olive-Gadea M, Crespo C, Granes C, Hernandez-Perez M, Pérez de la Ossa N, Laredo C, et al. Deep Learning Based Software to Identify Large Vessel Occlusion on Noncontrast Computed Tomography. *Stroke*. 2020 Oct;51(10):3133–7. <https://doi.org/10.1161/STROKEAHA.120.030326> PMID:32842922
14. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017 Feb;542(7639):115–8. <https://doi.org/10.1038/nature21056> PMID:28117445
15. Tschandl P, Rosendahl C, Akay BN, Argenziano G, Blum A, Braun RP, et al. Expert-Level Diagnosis of Nonpigmented Skin Cancer by Combined Convolutional Neural Networks. *JAMA Dermatol*. 2019 Jan;155(1):58–65. <https://doi.org/10.1001/jamadermatol.2018.4378> PMID:30484822
16. Ting DS, Cheung CY, Lim G, Tan GS, Quang ND, Gan A, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA*. 2017 Dec;318(22):2211–23. <https://doi.org/10.1001/jama.2017.18152> PMID:29234807
17. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*. 2016 Dec;316(22):2402–10. <https://doi.org/10.1001/jama.2016.17216> PMID:27898976
18. Burlina PM, Joshi N, Pekala M, Pacheco KD, Freund DE, Bressler NM. Automated Grading of Age-Related Macular Degeneration From Color Fundus Images Using Deep Convolutional Neural Networks. *JAMA Ophthalmol*. 2017 Nov;135(11):1170–6. <https://doi.org/10.1001/jamaophthalmol.2017.3782> PMID:28973096
19. Lee CS, Baughman DM, Lee AY. Deep Learning Is Effective for Classifying Normal versus Age-Related Macular Degeneration OCT Images. *Ophthalmol Retina*. 2017;1(4):322–7. <https://doi.org/10.1016/j.oret.2016.12.009> PMID:30693348
20. Steiner DF, MacDonald R, Liu Y, Truszkowski P, Hipp JD, Gamme C, et al. Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes for Metastatic Breast Cancer. *Am J Surg Pathol*. 2018 Dec;42(12):1636–46. <https://doi.org/10.1097/PAS.0000000000001151> PMID:30312179
21. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, et al.; the CAMELYON16 Consortium. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA*. 2017 Dec;318(22):2199–210. <https://doi.org/10.1001/jama.2017.14585> PMID:29234806
22. Shichijo S, Nomura S, Aoyama K, Nishikawa Y, Miura M, Shinagawa T, et al. Application of Convolutional Neural Networks in the Diagnosis of Helicobacter pylori Infection Based on Endoscopic Images. *EBioMedicine*. 2017 Nov;25:106–11. <https://doi.org/10.1016/j.ebiom.2017.10.014> PMID:29056541
23. Hirasawa T, Aoyama K, Tanimoto T, Ishihara S, Shichijo S, Ozawa T, et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. *Gastric Cancer*. 2018 Jul;21(4):653–60. <https://doi.org/10.1007/s10120-018-0793-2> PMID:29335825
24. Cai SL, Li B, Tan WM, Niu XJ, Yu HH, Yao LQ, et al. Using a deep learning system in endoscopy for screening of early esophageal squamous cell carcinoma (with video). *Gastrointest Endosc*. 2019 Nov;90(5):745–753.e2. <https://doi.org/10.1016/j.gie.2019.06.044> PMID:31302091
25. Byrne MF, Chapados N, Soudan F, Oertel C, Linares Pérez M, Kelly R, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. *Gut*. 2019 Jan;68(1):94–100. <https://doi.org/10.1136/gutjnl-2017-314547> PMID:29066576
26. Mori Y, Kudo SE, Misawa M, Saito Y, Ikematsu H, Hotta K, et al. Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy: A Prospective Study. *Ann Intern Med*. 2018 Sep;169(6):357–66. <https://doi.org/10.7326/M18-0249> PMID:30105375
27. Tufail A, Rudisill C, Egan C, Kapetanakis VV, Salas-Vega S, Owen CG, et al. Automated Diabetic Retinopathy Image Assessment Software: Diagnostic Accuracy and Cost-Effectiveness Compared with Human Graders. *Ophthalmology*. 2017 Mar;124(3):343–51. <https://doi.org/10.1016/j.ophtha.2016.11.014> PMID:28024825
28. Wismüller A, Stockmaster L, editors. A prospective randomized clinical trial for measuring radiology study reporting time on Artificial Intelligence-based detection of intracranial hemorrhage in emergent care head CT. *Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging*. International Society for Optics and Photonics; 2020.
29. Voets M, Møllersen K, Bongo LA. Reproduction study using public data of: development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *PLoS One*. 2019 Jun;14(6):e0217541. <https://doi.org/10.1371/journal.pone.0217541> PMID:31170223
30. Quellec G, Cazuguel G, Cochener B, Lamard M. Multiple-Instance Learning for Medical Image and Video Analysis. *IEEE Rev Biomed Eng*. 2017;10:213–34. <https://doi.org/10.1109/RBME.2017.2651164> PMID:28092576
31. Sadda P, Qarni T. Real-Time Medical Video Denoising with Deep Learning: application to Angiography. *Int J Appl Inf Syst*. 2018 May;12(13):22–8. PMID:29877510
32. Dong Z, Liu G, Ni G, Jerwick J, Duan L, Zhou C. Optical coherence tomography image denoising using a generative adversarial network with speckle modulation. *J Biophotonics*. 2020 Apr;13(4):e201960135. <https://doi.org/10.1002/jbio.201960135> PMID:31970879
33. Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim HC, et al. Artificial Intelligence for Mental Health and Mental Illnesses: an Overview. *Curr Psychiatry Rep*. 2019 Nov;21(11):116. <https://doi.org/10.1007/s11920-019-1094-0> PMID:31701320

34. Abbas H, Garberson F, Glover E, Wall DP. Machine learning approach for early detection of autism by combining questionnaire and home video screening. *J Am Med Inform Assoc*. 2018 Aug;25(8):1000–7. <https://doi.org/10.1093/jamia/ocy039> PMID:29741630
35. Tariq Q, Daniels J, Schwartz JN, Washington P, Kalantarian H, Wall DP. Mobile detection of autism through machine learning on home video: A development and prospective validation study. *PLoS Med*. 2018 Nov;15(11):e1002705. <https://doi.org/10.1371/journal.pmed.1002705> PMID:30481180
36. Abbasi B, Goldenholz DM. Machine learning applications in epilepsy. *Epilepsia*. 2019 Oct;60(10):2037–47. <https://doi.org/10.1111/epi.16333> PMID:31478577
37. Scrutinio D, Ricciardi C, Donisi L, Losavio E, Battista P, Guida P, et al. Machine learning to predict mortality after rehabilitation among patients with severe stroke. *Sci Rep*. 2020 Nov;10(1):20127. <https://doi.org/10.1038/s41598-020-77243-3> PMID:33208913
38. Naimi AI, Platt RW, Larkin JC. Machine Learning for Fetal Growth Prediction. *Epidemiology*. 2018 Mar;29(2):290–8. <https://doi.org/10.1097/EDE.0000000000000788> PMID:29199998
39. Tran D, Cooke S, Illingworth PJ, Gardner DK. Deep learning as a predictive tool for fetal heart pregnancy following time-lapse incubation and blastocyst transfer. *Hum Reprod*. 2019 Jun;34(6):1011–8. <https://doi.org/10.1093/humrep/dez064> PMID:31111884
40. Goh KH, Wang L, Yeow AY, Poh H, Li K, Yeow JJ, et al. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. *Nat Commun*. 2021 Jan;12(1):711. <https://doi.org/10.1038/s41467-021-20910-4> PMID:33514699
41. Fleuren LM, Klausch TL, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med*. 2020 Mar;46(3):383–400. <https://doi.org/10.1007/s00134-019-05872-y> PMID:31965266
42. Hyer JM, White S, Cloyd J, Dillhoff M, Tsung A, Pawlik TM, et al. Can We Improve Prediction of Adverse Surgical Outcomes? Development of a Surgical Complexity Score Using a Novel Machine Learning Technique. *J Am Coll Surg*. 2020 Jan;230(1):43–52.e1. <https://doi.org/10.1016/j.jamcollsurg.2019.09.015> PMID:31672674
43. El Hechi MW, Nour Eddine SA, Maurer LR, Kaafarani HM. Leveraging interpretable machine learning algorithms to predict postoperative patient outcomes on mobile devices. *Surgery*. 2021 Apr;169(4):750–4. <https://doi.org/10.1016/j.surg.2020.06.049> PMID:32919784
44. Tseng PY, Chen YT, Wang CH, Chiu KM, Peng YS, Hsu SP, et al. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit Care*. 2020 Jul;24(1):478. <https://doi.org/10.1186/s13054-020-03179-9> PMID:32736589
45. Mišić VV, Gabel E, Hofer I, Rajaram K, Mahajan A. Machine Learning Prediction of Postoperative Emergency Department Hospital Readmission. *Anesthesiology*. 2020 May;132(5):968–80. <https://doi.org/10.1097/ALN.0000000000003140> PMID:32011336
46. Seo W, Lee YB, Lee S, Jin SM, Park SM. A machine-learning approach to predict postprandial hypoglycemia. *BMC Med Inform Decis Mak*. 2019 Nov;19(1):210. <https://doi.org/10.1186/s12911-019-0943-4> PMID:31694629
47. Nair AA, Velagapudi MA, Lang JA, Behara L, Venigandla R, Velagapudi N, et al. Machine learning approach to predict postoperative opioid requirements in ambulatory surgery patients. *PLoS One*. 2020 Jul;15(7):e0236833. <https://doi.org/10.1371/journal.pone.0236833> PMID:32735604
48. Karhade AV, Ogink PT, Thio QC, Cha TD, Gormley WB, Hershman SH, et al. Development of machine learning algorithms for prediction of prolonged opioid prescription after surgery for lumbar disc herniation. *Spine J*. 2019 Nov;19(11):1764–71. <https://doi.org/10.1016/j.spinee.2019.06.002> PMID:31185292
49. Beam AL, Manrai AK, Ghassemi M. Challenges to the Reproducibility of Machine Learning Models in Health Care. *JAMA*. 2020 Jan;323(4):305–6. <https://doi.org/10.1001/jama.2019.20866> PMID:31904799
50. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. *JAMA*. 2018 Apr;319(13):1317–8. <https://doi.org/10.1001/jama.2017.18391> PMID:29532063
51. Henderson P, Islam R, Bachman P, Pineau J, Precup D, Meger D. Deep Reinforcement Learning That Matters. *Proc Conf AAAI Artif Intell*. 2018;32(1). <https://doi.org/10.1609/aaai.v32i1.11694>.
52. Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KG. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol*. 2015 Mar;68(3):279–89. <https://doi.org/10.1016/j.jclinepi.2014.06.018> PMID:25179855
53. Finlayson SG, Bowers JD, Ito J, Zittrain JL, Beam AL, Kohane IS. Adversarial attacks on medical machine learning. *Science*. 2019 Mar;363(6433):1287–9. <https://doi.org/10.1126/science.aaw4399> PMID:30898923
54. Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R. Practical guidance on artificial intelligence for health-care data. *Lancet Digit Health*. 2019 Aug;1(4):e157–9. [https://doi.org/10.1016/S2589-7500\(19\)30084-6](https://doi.org/10.1016/S2589-7500(19)30084-6) PMID:33323184
55. Kompa B, Snoek J, Beam AL. Second opinion needed: communicating uncertainty in medical machine learning. *NPJ Digit Med*. 2021 Jan;4(1):4. <https://doi.org/10.1038/s41746-020-00367-3> PMID:33402680
56. Cabitza F, Rasoini R, Gensini GF. Unintended Consequences of Machine Learning in Medicine. *JAMA*. 2017 Aug;318(6):517–8. <https://doi.org/10.1001/jama.2017.7797> PMID:28727867
57. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. *J Am Med Inform Assoc*. 2004 Mar-Apr;11(2):104–12. <https://doi.org/10.1197/jamia.M1471> PMID:14633936
58. Currie G, Hawk KE, Rohren E, Vial A, Klein R. Machine Learning and Deep Learning in Medical Imaging: intelligent Imaging. *J Med Imaging Radiat Sci*. 2019 Dec;50(4):477–87. <https://doi.org/10.1016/j.jmir.2019.09.005> PMID:31601480
59. Char DS, Shah NH, Magnus D. Implementing Machine Learning in Health Care - Addressing Ethical Challenges. *N Engl J Med*. 2018 Mar;378(11):981–3. <https://doi.org/10.1056/NEJM1714229> PMID:29539284
60. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med*. 2019 Sep;25(9):1337–40. <https://doi.org/10.1038/s41591-019-0548-6> PMID:31427808
61. Mooney SJ, Pejaver V. Big Data in Public Health: Terminology, Machine Learning, and Privacy. *Annu Rev Public Health*. 2018 Apr;39(1):95–112. <https://doi.org/10.1146/annurev-publ-health-040617-014208> PMID:29261408