



Retos clínicos

Reto clínico 4. Y en el paciente crítico... ¿Puedo utilizar estrategias de inteligencia artificial para mejorar los resultados clínicos?

Clinical challenge 4. And in the critical patient... Can I use artificial intelligence strategies to improve clinical outcomes?

Joaquín Álvarez Rodríguez

Jefe del Servicio de Medicina Intensiva. Hospital Universitario de Fuenlabrada. Fuenlabrada, Madrid

INTRODUCCIÓN

Actualmente se plantea si las estrategias de inteligencia artificial (IA) se pueden usar para mejorar los resultados clínicos, con un creciente número de profesionales a favor de este enfoque. Esta herramienta se está integrando de forma rápida en la sociedad en todos los aspectos de la vida. En la Medicina también se está introduciendo, siendo probable que forme parte cada vez más de nuestra vida cotidiana en los próximos años.

Por todo ello, es importante familiarizarse con tres conceptos:

1. *Inteligencia artificial (IA)*. Técnicas por las que los ordenadores u otras máquinas son capaces de emular capacidades humanas en la resolución de problemas y en la toma de decisiones. Existen múltiples ramas derivadas.
2. *Machine learning* (aprendizaje máquina, ML). Aplicación de IA a través de la cual los sistemas son capaces de aprender automáticamente y mejorar los resultados sin ser explícitamente programados.
3. *Deep learning* (aprendizaje profundo, DL). Parte del ML que imita la forma de procesar los datos del cerebro humano y analiza patrones para su uso en la toma de decisiones.

ALGUNOS EJEMPLOS

El aprendizaje automático y la IA se están introduciendo de forma rápida en el mundo de la Medicina, existiendo múltiples ejemplos, como:

- Analítica predictiva e identificación de enfermedades. Existen programas que en base a unas determinadas analíticas o pruebas diagnósticas son capaces de predecir si los pacientes van a desarrollar, con el paso del tiempo, enfermedades como la diabetes o una cardiopatía.
- Recomendaciones de diagnóstico y tratamiento. Hay programas para el diagnóstico de cáncer que utilizan conjuntamente la información clínica y el genotipo del paciente, además de otros datos, proporcionando sugerencias de recomendación de diagnóstico y tratamiento.
- Diagnóstico por imagen. Esta aplicación ya está introducida en muchos hospitales, utilizándose por ejemplo para la lectura de las radiografías.
- Cirugía robótica. El robot DaVinci, basado en IA, fue el primero que se introdujo para realizar cirugías poco invasivas.
- Apoyo a la decisión clínica.
- Medicina personalizada. Se pueden encontrar programas que ayudan a analizar una gran cantidad de datos sobre el paciente.
- Modificación del comportamiento basado en el aprendizaje automático. En pacientes con diabetes, por ejemplo, hay programas que regulan, conforme a otros patrones de conducta, el comportamiento del paciente y ofrecen recomendaciones en función de los datos recopilados.
- Gestión de la salud de la población. Actualmente se dispone de programas para control de epidemias y de herramientas que proporcionan consejos según la información de salud poblacional general.

Conflicto de intereses: el autor declara no tener conflicto de interés.

Inteligencia artificial: el autor declara no haber usado inteligencia artificial (IA) ni ninguna herramienta que use IA para la redacción del artículo

Álvarez Rodríguez J. Reto clínico 4. Y en el paciente crítico... ¿Puedo utilizar estrategias de inteligencia artificial para mejorar los resultados clínicos?. *Nutr Hosp* 2024;41(N.º Extra 4):51-54

DOI: <http://dx.doi.org/10.20960/nh.5775>

Correspondencia:

Joaquín Álvarez Rodríguez. Servicio de Medicina Interna. Hospital Universitario de Fuenlabrada. Cam. del Molino, 2. 28942 Fuenlabrada, Madrid
e-mail: joaquin.alvarez@urjc.es

- Registros sanitarios inteligentes. Casi todos los hospitales utilizan la historia clínica electrónica y ahora, con la aplicación de la IA, es posible emplear esos datos para que produzcan beneficios.
- Descubrimiento y fabricación de fármacos.
- Ensayos clínicos e investigación.

Dentro del denominado aprendizaje automático, se distinguen 4 tipos, siendo los dos principales el aprendizaje supervisado y el aprendizaje no supervisado. El aprendizaje supervisado se desarrolla cuando, por ejemplo, se dispone de miles de datos acerca de personas enfermas y se conoce cuáles van a ingresar en la unidad de cuidados intensivos (UCI) y cuáles no. Sobre esos datos se aplican técnicas de IA para buscar patrones que se repitan y diferencien a pacientes que ingresan en la UCI; comparando estos patrones con las características de los enfermos que se hallan ingresados en las plantas de hospitalización, se buscan, antes de que se produzca el deterioro clínico evidente, aquellos que tienen más probabilidad de necesitar ingreso en UCI. En el aprendizaje no supervisado es el propio algoritmo el que busca los patrones que se están repitiendo sin conocer el resultado final, es decir, la necesidad o no del ingreso en UCI.

Algunas de las técnicas más conocidas en el mundo de la salud de aprendizaje supervisado son:

1. *Algoritmos de clasificación:*
 - Árboles de decisión.
 - Refuerzo de gradiente.
2. *Regresión:*
 - Regresión lineal.
 - Regresión logística.

El DL trata de emular el procesamiento de la información por el cerebro. En 1949, Donald Hebb estableció la teoría de los ensamblajes neuronales; según esta, la información recibida en el cerebro llega a través de unos nodos de entrada formados por neuronas, los cuales van a distribuir la información a otros nodos intermedios que procesan dicha información produciendo una respuesta en un nodo de salida. El DL emula esa arquitectura recibiendo la información en unos nodos de entrada que es transmitida a unos nodos intermedios que la procesan y emiten una respuesta en los nodos de salida. Se desconoce con exactitud de qué factores depende el resultado generado. Son ejemplos de DL el procesamiento del lenguaje natural de los comentarios de los sanitarios incluidos en la historia clínica de los pacientes. El procesamiento de imagen también se incluye en el DL.

Para poder generar herramientas predictivas, los componentes requieren datos, tantos como sea posible. De esta forma se pueden buscar patrones repetitivos. Es necesario determinar las variables y el algoritmo que se va a usar para generar probabilidades.

La eficacia del algoritmo depende, principalmente, de la cantidad de datos, tanto en el ML como en el DL. En el caso del ML, se ha observado que el rendimiento puede alcanzar una meseta, es decir, aunque se introduzcan más datos, no aumenta el rendimiento. En el DL, la cantidad de datos influye en todo momento en el rendimiento. El resultado del conjunto de datos, variable y algoritmo genera un resultado de probabilidad.

APORTACIONES DEL PROGRAMA PRE-CRIT

En el Hospital Universitario de Fuenlabrada hemos desarrollado un algoritmo, PRE-CRIT, diseñado para la predicción de pacientes que van a deteriorarse en las plantas de hospitalización y van a necesitar ingreso en UCI. El objetivo es detectar a estos pacientes antes de que entren en fase crítica, para poder iniciar cuanto antes el tratamiento y tomar las medidas necesarias de forma que se obtengan mejores resultados.

PRE-CRIT permite estudiar los pacientes que hay en planta y detectar a aquellos susceptibles de empeorar y requerir ingreso en UCI. El programa utiliza un algoritmo ML (Light GBM), un algoritmo de potenciación de gradiente de código abierto, distribuido y de alto rendimiento basado en un árbol de decisiones que se utiliza principalmente para clasificar. Inicialmente trabajamos con otros algoritmos, pero este fue elegido debido a los resultados obtenidos.

Para probar el programa se usaron dos cohortes de enfermos del hospital. La primera cohorte incluyó a 68 000 pacientes ingresados entre 2013 y 2018. La segunda cohorte recoge el periodo de tiempo desde la pandemia hasta el presente, incorporando 30 000 historias clínicas. Tras aplicar una serie de criterios de exclusión, el número total final de historias clínicas usadas fue de 92 000. De todos ellos conocíamos cuáles ingresaron en la UCI y cuáles no. Se trabajó con la información de esos pacientes y una serie de variables establecidas, actualmente 34. La salida del programa se produce con el ingreso en la unidad de vigilancia intensiva (UVI), una variable poco frecuente, ya que solo acaban ingresando el 2-3 % de los pacientes.

Cuando se entrena a un algoritmo y modelo de predicción, es necesario dividir todos los pacientes en varios grupos. El primer grupo es el de entrenamiento, que incluye el 60-70 % de ellos. A este grupo de pacientes seleccionado al azar le aplicamos el algoritmo o algoritmos que estemos probando y, una vez se ha conseguido una precisión y sensibilidad que se consideran oportunas, aplicamos el algoritmo al grupo que llamamos de validación (10-20 % de los pacientes), pudiendo variar determinados parámetros. Se comparan los resultados con aquellos obtenidos en el grupo de entrenamiento y, si los resultados son adecuados, se somete al algoritmo el grupo de prueba, que incluye un 10-20 % de pacientes cuyos datos no han sido usados previamente. De esta forma se puede corroborar que los resultados tienen repetibilidad. Este es el momento de pasar a la fase clínica y aplicar el algoritmo a los pacientes en tiempo real.

Los resultados obtenidos con PRE-CRIT no son buenos hasta el momento, ya que, tras mucho trabajo, se obtuvo poco rendimiento. Es probable que sea necesario ajustar las variables para obtener los resultados deseados. Actualmente, el programa genera muchas alertas teniendo una precisión solo del 1 % y una sensibilidad del 46 %.

PRE-CRIT se encuentra en producción desde el 15 de enero aplicando el algoritmo 2 veces al día a todos los pacientes ingresados. Hemos analizado los resultados a partir de los datos disponibles desde el 15 de enero hasta el 15 de abril. Durante ese periodo, un médico a las 10 de la mañana y a las 6 de la tarde valoraba a los pacientes que el programa indicaba.

La media de avisos que se obtuvieron durante esos 3 meses fue de 55 al día con un número de falsos positivos elevado. Se trata de un número de avisos muy grande para la cantidad de pacientes ingresados. Otro problema que surgió fue que muchos de los avisos de la tarde ya se habían producido por la mañana en pacientes que no habían sufrido ningún cambio. Actualmente se está trabajado en solucionar el gran número de avisos, su repetitividad y la cantidad de falsos positivos (Fig. 1).

En total, durante los 3 meses se recibieron casi 5000 alertas, de las cuales fueron útiles el 6 % (94 % son no útiles). De las alertas útiles se llegaron a producir 31 ingresos, lo que supuso el 12 % de las alertas útiles. El resto de las alertas útiles sirvieron para que el médico responsable estuviera pendiente de hacer un seguimiento al paciente. También existen alertas que sirven para establecer límites de esfuerzo terapéutico, es decir, pacientes con patologías crónicas graves o terminales que no van a ser ingresados. El problema principal reside en el 94 % de falsos positivos, de los cuales el 13 % son avisos sobre pacientes mal detectados (al paciente no le sucede nada y no se sabe por qué ha generado una alerta en el sistema) y el 87 % restante son pacientes no graves cuya patología se está solventando sin incidencias.

También se registran falsos negativos, es decir, pacientes que no generan alerta, pero acaban ingresados en la UCI. Estos casos fueron 38. Además, las alertas tienen una media de 2,8 por paciente (se repiten mucho y aún se desconoce el motivo).

En total, de los 69 pacientes que finalmente sí ingresaron en UCI procedentes de las plantas de hospitalización, se detectaron 31 (el 45 %). Actualmente PRE-CRIT solo se utiliza para los pacientes que ingresan desde planta porque, debido al elevado nú-

mero de alertas, parece inviable aplicarlo por ahora a Urgencias u Observación. De los 31 pacientes ingresados, no todos fueron detectados cuando el médico fue a investigar la alerta, sino que en algunos casos sucedió más tarde, cuando aquella se había descartado y se pensaba que el paciente se encontraba estable. Se está investigando qué sucede en estos casos de pacientes mal valorados.

Es relevante entender a los pacientes mal detectados, ya que sobrecargan el sistema y saturan a los profesionales. Es necesario eliminar del sistema a pacientes que, por ejemplo, han sido sometidos a una cirugía menor y saltan en el sistema. Se está trabajando también en la eliminación de las alertas repetidas, entrenando al algoritmo para que entienda que, si no ha habido cambios desde la primera alerta producida, no debe generar una nueva alerta sobre el mismo paciente.

En general, se está tratando de optimizar el algoritmo mediante una mejora de los factores que se introducen en él, con el fin de obtener una mayor precisión. Así se pretende disminuir la cantidad de falsos positivos para evitar sobrecargar el sistema. A pesar de los déficits detectados, se trata de una línea interesante de trabajo.

INTELIGENCIA ARTIFICIAL PARA LA NUTRICIÓN EN LA UCI

La IA también se está aplicando con éxito al campo de Nutrición en la UCI. En la bibliografía se pueden encontrar artículos sobre varios aspectos en este terreno.

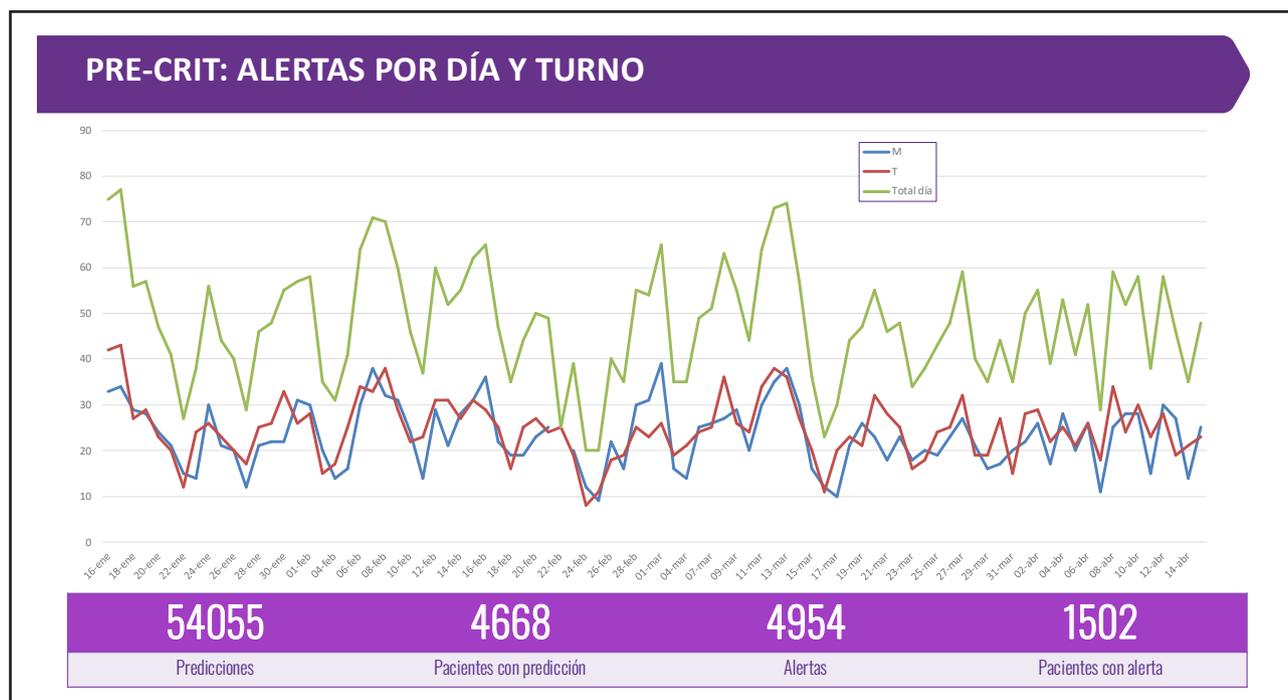


Figura 1. Alertas por día y turno recogidos por el programa CRIT (M: mañana; T: tarde).

En un trabajo realizado por Choi y cols., buscando modelos predictivos para el síndrome de realimentación (1), se incluyeron 806 pacientes ingresados con 2 o más días sin alimentación oral y con una medición de los niveles de fósforo en los 5 días siguientes a la supresión de la alimentación oral. Se utilizaron modelos de regresión, además de métodos de ML, y se encontró una serie de factores en relación con el citado síndrome.

En otro estudio se analizó el inicio de la nutrición enteral en pacientes de UCI (2). Se tuvieron en cuenta 56 factores. El problema principal al que se enfrentaba el estudio era el bajo número de pacientes, que combatieron utilizando los datos depositados en la base MIMIC-IV (*Medical Information Mart for Intensive Care IV*), logrando así aplicar su algoritmo de ML sobre más de 76 000 pacientes de UCI. Finalmente se detectó que en 7210 pacientes el inicio de la nutrición enteral fue más precoz. Se trataba de pacientes con sepsis, fallo orgánico secuencial, fallo renal o temperatura corporal elevada.

También se han realizado estudios para evaluar la intolerancia a la nutrición enteral en pacientes de UCI (3,4). En estos estudios el número de pacientes se encontraba alrededor de 200 (una cohorte muy pequeña para este tipo de técnicas). Se observó una tasa de intolerancia del 35-45 %.

Se han llevado a cabo otros estudios acerca del volumen residual gástrico y el pronóstico de los pacientes (5,6). El número de pacientes fue inferior a 3000, y se observó que la mortalidad se relacionaba con un residuo superior a 250 ml, mientras que la intolerancia se asociaba con el índice de masa corporal, la ratio creatina/urea, la acidosis y el género masculino. Se observó que el retraso del inicio de la nutrición enteral estaba vinculado con la estancia prolongada en la UCI.

Estos estudios pueden mejorarse de varias formas. La principal sería incrementar el número de pacientes incluidos, ya que

cuando se trabaja con técnicas de ML es necesaria una gran cantidad de información para que los resultados sean fiables. Por último, la IA predictiva debe tener una función práctica. Ha de ser empleada en el sistema informático, con el objetivo de que el modelo generado sea aplicable en la práctica clínica diaria.

Actualmente se están empezando a usar estas herramientas y, desde el punto de vista de la Medicina y la clínica, el mayor problema reside en el bajo número de pacientes incluidos. Es necesario que se generen más bases de datos colaborativas para poder aplicar la IA. En un futuro cercano, muchos de estos modelos serán de gran utilidad en cualquier aspecto de la Medicina.

BIBLIOGRAFÍA

1. Choi TY, Chang M-Y, Heo S, Jang JY. Explainable machine learning model to predict refeeding hypophosphatemia. *Clin Nutr ESPEN* 2021;45:213-9. DOI: 10.1016/j.clnesp.2021.08.022
2. Wang Y-X, Li X-L, Zhang L-H, Li H-N, Liu X-M, Song W, et al. Machine learning algorithms assist early evaluation of enteral nutrition in ICU patients. *Front Nutr* 2023;10:1060398. DOI: 10.3389/fnut.2023.1060398
3. Hu K, Deng XL, Han L, Xiang S, Xiong B, Pinhu L. Development and validation of a predictive model for feeding intolerance in intensive care unit patients with sepsis. *Saudi J Gastroenterol* 2022;28(1):32-8. DOI: 10.4103/sjg.sjg_286_21
4. Lu X-M, Jia D-S, Wang R, Yang Q, Jin S-S, Chen L. Development of a prediction model for enteral feeding intolerance in intensive care unit patients: A prospective cohort study. *World J Gastrointest Surg* 2022;14(12):1363-74. DOI: 10.4240/wjgs.v14.i12.1363
5. Raphaeli O, Statlander L, Hajaj C, Bendavid I, Goldstein A, Robinson E, et al. Using Machine-Learning to Assess the Prognostic Value of Early Enteral Feeding Intolerance in Critically Ill Patients: A Retrospective Study. *Nutrients* 2023;15(12):2705. DOI: 10.3390/nu15122705
6. Raphaeli O, Yechiel LB, Statlander L, Kagan I, Hiesmayr M, Tarantino S, et al. Gastric residual volume as predictor of clinical outcomes in critically ill patients at nutritionday: A machine learning (ML) approach. *Clin Nutr ESPEN* 2023;58:452-3. DOI: 10.1016/j.clnesp.2023.09.083