Conceptos básicos utilizados en la inteligencia artificial
Basic concepts used in artificial intelligence

M. Baget Bernaldiz, P. Romero Aroca

Resumen

La inteligencia artificial (IA) generará cambios significativos en la manera de cribar, estudiar, diagnosticar y tratar las enfermedades oculares. Es por este motivo que los oftalmólogos debemos conocer estas nuevas herramientas que en breve vamos a utilizar.

Palabras clave: *Machine Learning. Deep Learning*. Algoritmo. Red neuronal. *Big Data.*

Resum

La intel·ligència artificial provocarà canvis significatius en la manera que tenim de fer el cribratge, estudiar, diagnosticar i tractar les malalties oculars. És per això, que els oftalmòlegs hem de conèixer de primera mà aquestes noves eines que ben aviat estaran al nostre abast.

Paraules clau: *Machine Learning. Deep Learning.* Algorisme. Xarxa neuronal. *Big Data.*

Abstract

Artificial intelligence will significantly change the way eye diseases are screened, studied, diagnosed, and treated. Consequently, ophthalmologists must learn first-hand about these new tools, since sooner rather than later we will use them in our daily practice.

Key words: Machine Learning. Deep Learning. Algorithm. Neural Network. Big Data.

1.2. Conceptos básicos utilizados en la inteligencia artificial

Basic concepts used in artificial intelligence

M. Baget Bernaldiz[1], P. Romero Aroca[2]

[1]Médico adjunto. Servicio de Oftalmología. Hospital Universitari Sant Joan. Reus. Universitat Rovira i Virgili. Reus. Institut d'Investigació Sanitària Pere Virgili (IISPV). Reus. Tarragona. [2]Jefe de Servicio. Servicio de Oftalmología. Hospital Universitari Sant Joan. Reus. Universitat Rovira i Virgili. Reus. Institut d'Investigació Sanitària Pere Virgili (IISPV). Reus. Tarragona.

Correspondencia:
Marc Baget Bernaldiz
E-mail: mbaget@gmail.com

Origen del término inteligencia artificial

Aunque el término IA parece muy moderno, tiene ya casi 70 años. Fue John McCarthy (Figura 1), considerado uno de los padres de la IA, quien lo acuñó en el año 1956. McCarthy, que era ingeniero informático, estaba muy interesado por entonces en desarrollar máquinas que hicieran tareas complejas. Con este objetivo, se propuso reunir a los ingenieros informáticos mundiales más relevantes del momento en una escuela de verano en la Universidad de Dartmouth (Hanover, New Hampshire, Estados Unidos). Y como necesitaba recursos para organizar el evento, fue a pedir dinero al Instituto Rockefeller, que finalmente accedió a subvencionarlo. Cuando le preguntaron qué título iba a llevar el evento, McCarthy, sin pensarlo mucho, respondió que escuela de verano sobre la IA[1]. En aquel momento, el término no estuvo exento de críticas ya que, a unos cuantos colegas no les pareció acertado el adjetivo de artificial, por transmitir falta de autenticidad. A pesar de ello, el término acabó cuajando, y se ha mantenido hasta nuestros días con más fuerza que nunca.

El *boom* de la inteligencia artificial

Podemos situar en el año 2014 el origen del interés global por la IA. En enero de este año, Google adquiere DeepMind, una pequeña *start-up* (empresa emergente) dedicada a la IA por un valor de 650 millones de dólares. Se generan grandes titulares

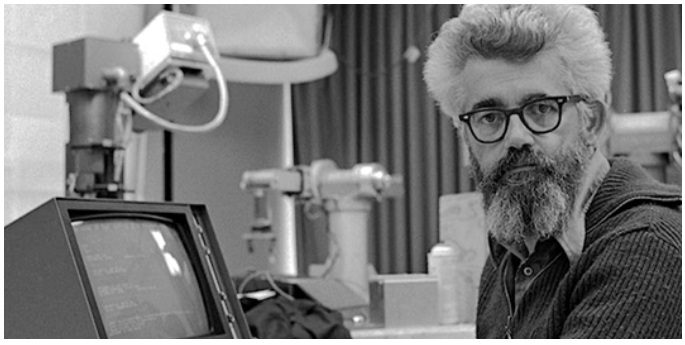

Figura 1. John McCarthy fue un ingeniero informático muy relevante, creador de lenguajes de programación que todavía se utilizan en la actualidad. Se le atribuye el origen del término inteligencia artificial en el año 1956. (Imagen cedida por el autor).

intentando comprender cómo una compañía que tan solo cuenta con 20 empleados puede tener un valor tan alto en el mercado. Las empresas tecnológicas empiezan a invertir grandes cantidades de dinero para no quedarse atrás, y los diferentes gobiernos de los países desarrollados se preguntan cómo responder a este desafío. DeepMind había logrado crear una máquina capaz de jugar al videojuego Atari mucho mejor y de manera más creativa que los jugadores más expertos. Pero lo relevante de esta historia es que la máquina había aprendido a jugar sin que nadie le hubiera enseñado cómo hacerlo: ¡lo había aprendido por sí misma! DeepMind acababa de mostrar el enorme potencial de las redes neuronales o *neural networks*.

La inteligencia artificial

La IA es la ciencia computacional que tiene como objetivo construir máquinas inteligentes. El primer problema que se plantea entonces es definir qué se considera una máquina inteligente. Por suerte, Alan Turing (Figura 2), considerado por muchos como el verdadero padre de la IA, diseñó ya en el año 1950 un sistema sencillo y elegante, con el objeto de diferenciar claramente una máquina inteligente de la que no lo es.

Alan Turing se inspiró en *The imitation game*, un juego de la época victoriana. En este, un jugador debe adivinar si un determinado personaje es hombre o mujer, a través de la realización de una serie de preguntas a otro jugador. De manera similar, Turing propuso un test en el cual la persona interroga a través de un teclado y una pantalla a un sujeto, que puede ser una persona o una computadora (el interrogador desconoce de antemano si realiza las preguntas a una persona o a un programa). La interacción es en forma de preguntas y respuestas de texto: el interrogador plantea preguntas y las respuestas son mostradas en la pantalla. La tarea del interrogador es adivinar si el objeto interrogado es una persona o una máquina. En el caso de que las respuestas sean tan creativas que el interrogador crea que está interrogando a una persona, cuando en realidad se trata de una máquina, se considera que se ha alcanzado el umbral de la IA. Turing publicó su test en el artículo *Computing Machinery and Intelligence*, en la prestigiosa revista *Mind*, en el año 1950, siendo considerada la primera publicación mundial sobre IA. En la actualidad y en su honor, a esta prueba se la denomina test de Turing[2].

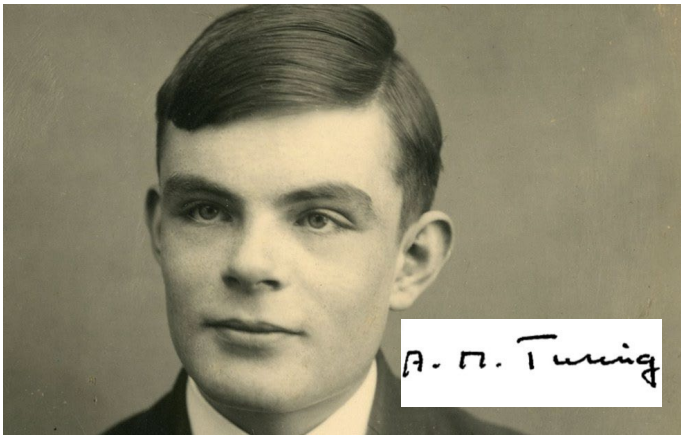

Figura 2. Alan Turing es el autor de la primera publicación mundial sobre inteligencia artificial, en la revista Mind, en el año 1950, donde explica cuándo se debe considerar a una máquina como inteligente. (Imagen cedida por el autor).

A partir del test de Turing se consideran máquinas inteligentes a aquellas con capacidad de sustituir a los cuerpos y/o cerebros humanos. En la actualidad, ya disponemos de algunos ejemplos de máquinas que se encuentran en el camino de ser realmente inteligentes. Recientemente todos nos hemos sorprendido ante *chatbots* que ya no ofrecen respuestas enlatadas, sino conversaciones creativas, fruto de la mejora en el procesamiento del lenguaje natural. O la existencia de redes neuronales que pueden clasificar las retinografías de pacientes diabéticos en los diferentes tipos de retinopatía diabética (RD) con gran precisión.

La mayoría de las empresas tecnológicas dedicadas a la IA están centradas en desarrollar tareas concretas, como por ejemplo, aprovechar la información recogida por los *Smart watches* o relojes inteligentes respecto diferentes parámetros vitales, con el objeto de predecir y anticiparse a determinadas enfermedades, como el infarto de miocardio o las demencias. Por el contrario, los escenarios distópicos alarmistas que proclaman que los humanos acabarán supeditados a los robots inteligentes forman parte del juego de ciertos medios sensacionalistas, pero que, hasta este momento, nada tienen que ver con la realidad.

Machine learning

El *machine learning* (aprendizaje automático) es la ciencia computacional dentro de la IA centrada en crear máquinas que realicen tareas complejas con la capacidad de aprender y mejorar por ellas mismas, sin que nadie las haya previamente programado de manera explícita. Imaginemos que una empresa quisiera desarrollar un lector inteligente para venderlo en farmacias con el objeto de

descifrar las recetas escritas a mano por cualquier médico. Nadie sabría cómo afrontar esta tarea mediante un sistema clásico de programación porque, sencillamente, sería imposible. Necesitamos algo diferente y esto se llama *machine learning*.

Deep learning

El *deep learning* (aprendizaje profundo) es el área dentro del *machine learning* dedicada al estudio de las *neural networks* o redes neuronales, que fueron conceptualizadas por McCulloch y Pitts en el año 1940, y una década más tarde, Rosenblatt diseñó el primer modelo de red neuronal basado en una única neurona, llamado perceptrón. Este modelo tan sencillo demostró tener una capacidad limitada respecto a las tareas que podía llevar a cabo, por lo que se fue perdiendo poco a poco el interés por las redes neuronales. Ya en 1980, el desarrollo tecnológico y la aparición de ciertos algoritmos permitió la creación de redes neuronales, formadas por varias capas de neuronas, que eran capaces de realizar tareas de cierta complejidad[3]. Por último, en el año 2005, se produce el impulso definitivo en el estudio y desarrollo de las redes neuronales, bajo el nombre de *deep learning*[4].

Si el objeto de la IA es desarrollar máquinas sustitutivas del cerebro humano, qué mejor que construir un modelo de IA basado en la anatomía y función del cerebro humano. Si este consta de 100 billones de neuronas, las redes neuronales actuales más potentes ya tienen más de 1 millón de neuronas, con un número de interconexiones neuronales similar al cerebro de un gato, lo que explica el salto de gigante en las tareas que están llevando a cabo. Desde un punto de vista funcional, si la neurona del ser vivo se despolariza cuando sobrepasa cierto umbral de estimulación, en la neurona artificial, se desencadena una función matemática denominada de activación, transmitiendo el "impuso matemático" a las neuronas subsiguientes con las que se relaciona.

En una red neuronal artificial se pueden distinguir tres tipos de *neural layers* o capas neuronales (Figura 3). Una primera capa, denominada *input layer*, que se relaciona con el exterior, a través de las variables independientes o explicativas asociadas a un de-

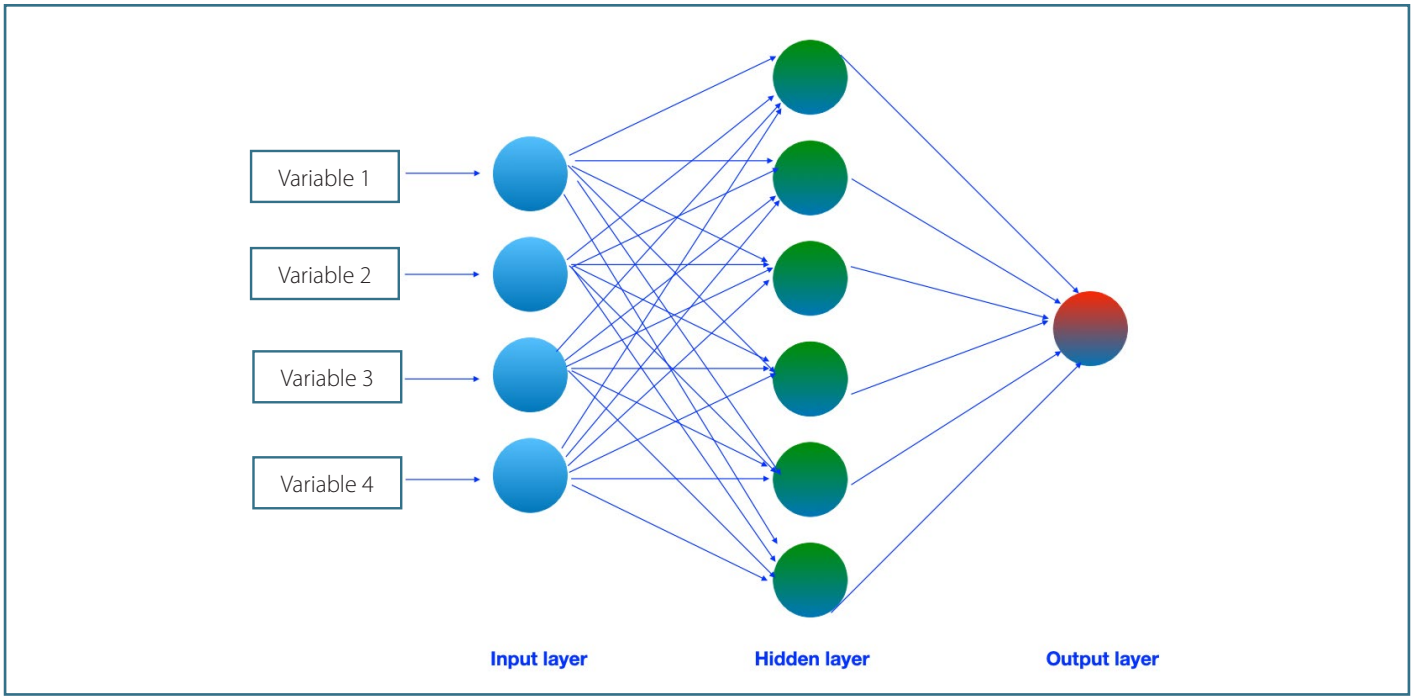

Figura 3. Esquema de una red neuronal simple. Una red neuronal consta de tres grupos de neuronas dispuestas en columnas. La *input layer* (primera capa) mantiene contacto con el exterior y se relaciona con las variables independientes. Las *hidden layers* (capas que están en medio) no mantienen contacto con el exterior, y es donde ocurre el aprendizaje de la red neuronal, razón por la cual se denomina *deep learning* (aprendizaje profundo) a la ciencia que investiga las redes neuronales. Por último, la *output layer* (última capa) es responsable de generar el resultado de la red. (Imagen cedida por el autor).

terminado resultado, que se encargaran de activarlas. Una última capa llamada *output layer*, que es la responsable de generar el resultado de la red neuronal. Y en medio, y por lo tanto, sin contacto con el exterior, se encuentran las *hidden layers*, que reciben los impulsos matemáticos de la *input layer*, para ser procesados y transmitidos finalmente a la *output layer*. Las *hidden layers* son el grupo de neuronas más importantes, ya que en ellas se produce el aprendizaje de la red neuronal, motivo por el cual a las redes neuronales actuales se las conoce con el nombre de *deep learning*.

El principal objetivo que se persigue con la construcción de las redes neuronales es la predicción de algún tipo de evento. Pronosticar si un determinado paciente desarrollará RD o un paciente con COVID-19 acabará ingresando en la UCI, por ejemplo. Para que una red neuronal o cualquier otro modelo de IA sea buen predictor, se necesitan muchos datos *(big data)* y de alta calidad, con el objeto de entrenar el modelo.

Imaginemos que queremos desarrollar una red neuronal para que lea y clasifique las retinografías de los pacientes diabéticos en los diferentes tipos de RD. El primer paso consiste en subministrar a la red miles de retinografías con las lesiones convenientemente etiquetadas, para que pueda entrenarse y acabe por "entender" qué es un microaneurisma, qué es una hemorragia, qué es un exudado duro y qué es un neovaso retiniano. Cada retinografía se descompone en píxeles, y estos, en números enteros, que es el tipo de lenguaje que lee una red neuronal. Por lo tanto, la red deberá aprender que un microaneurisma o una hemorragia se puede presentar de muchísimas maneras (mediante una infinidad de números diferentes). Este modelo planteado sería un ejemplo de *supervised learning* o aprendizaje supervisado, ya que hemos ayudado al modelo a que aprenda a identificar las lesiones mediante su etiquetado previo.

Siguiendo con este mismo ejemplo, en el inicio de la fase de entrenamiento, observaríamos que la red comete muchos errores, porque todavía no se le han presentado suficientes retinografías para desempeñar su tarea de manera precisa. Posiblemente identificaría como normales retinografías con RD moderada o severa. Entonces, ¿cómo aprende? La red neuronal aprende aprovechando los errores que comete, computándolos. En cada set de retinografías que lee, se mide la distancia entre el valor dado por la red neuronal respecto el valor que se espera que dé (diferencia entre el valor real dado y el valor ideal esperado). Posteriormente y gracias a la función del algoritmo de *backpropagation* (propagación hacia atrás de errores o retropropagación), se van ajustando los parámetros o pesos de cada una de las neuronas de la red para que, al final, la lectura de las retinografías sea correcta. Por lo tanto y de manera general, un algoritmo de IA es una formulación matemática compleja cuya función es encontrar los pesos de cada una de las neuronas de la red. Dicho en otras palabras, la función del algoritmo es averiguar la importancia relativa o el peso de cada variable independiente en el resultado final.

Las redes neuronales son tan efectivas porque son capaces de encontrar cualquier relación existente, por compleja que sea, entre las variables independientes o explicativas y la variable dependiente o el resultado. Esta propiedad de aproximación universal permite a las redes neuronales ser buenas predictoras, siempre y cuando exista relación entre las variables independientes y el resultado, y la base de datos que se ha utilizado para su entrenamiento sea representativa de la población a la cual se pretenda utilizar.

Imaginemos ahora que nos proponemos averiguar cuáles de los 5.000 pacientes con degeneración macular asociada a la edad exudativa, que han entrado en un ensayo clínico, han respondido mejor a un nuevo fármaco antiangiogénico. Disponemos de los datos clínicos, las retinografías, las AGF, los OCT y los OCT-A de todos ellos. Si quisiéramos realizar un análisis pormenorizado de cada paciente de manera manual, midiendo las áreas de los neovasos coroideos, la presencia y cuantificación de fluido en cada compartimiento y la integridad anatómica de las capas neurosensoriales de la retina, sería una tarea casi imposible de realizar. Una aproximación más realista a este problema sería diseñar un modelo de aprendizaje no supervisado o *unsupervised learning* que aprendiera por él mismo a reconocer similitudes y diferencias en el comportamiento anatómico y funcional de los pacientes, con el objeto de encontrar diferentes patrones de pacientes o *clustering*.

Existen muchos tipos de redes neuronales. A los oftalmólogos, nos interesa particularmente las *convolutional neural networks* (CNN) o redes neuronales convolucionales, que son aquellas redes neuronales especialmente construidas para analizar imágenes como las del ojo (Figura 4). Las CNN no procesan las imágenes, sino los números enteros que resultan de la descomposición de los píxeles que contienen. De este modo, la imagen inicial se ha convertido ahora en una matriz de números enteros, pero que continúa albergando todas sus características de base. A esta

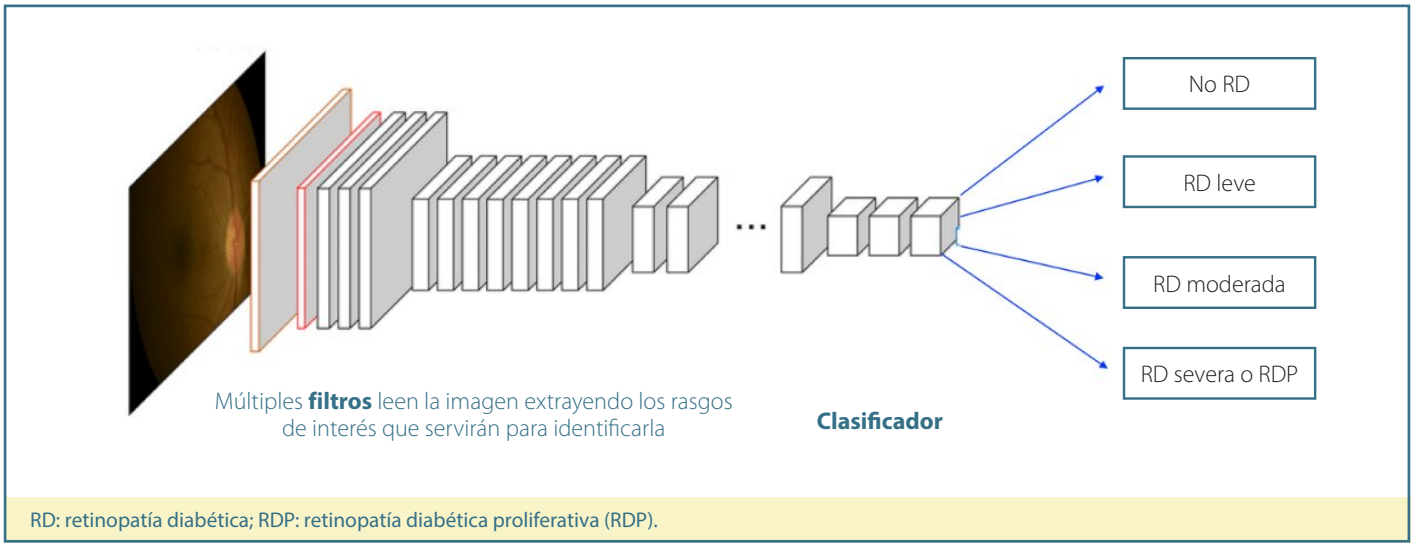

Figura 4. *Convolutional neural networks* o redes neuronales convolucionales. Son las redes neuronales diseñadas para la interpretación de imágenes. Constan de dos partes: una primera, donde unos filtros leen la imagen extrayendo los rasgos de interés mediante una operación matemática denominada convolución; y una segunda, en la que una vez creado el mapa de rasgos, clasifica la imagen en una de las categorías posibles. (Imagen cedida por el autor).

matriz numérica, se le aplican unos filtros (también numéricos), con el objeto de extraer las características de la imagen a través de una operación matemática llamada convolución, de ahí el nombre de este tipo de red neuronal. Por último, la matriz resultante, después de haber aplicado todos los filtros, que contiene todas las características de la imagen inicial (denominada mapa de rasgos o *feature map*), se someterá al clasificador de la red que decidirá de qué tipo de imagen se trata. Las CNN se utilizan, por ejemplo, para clasificar las retinografías de los pacientes diabéticos en los distintos tipos de RD.

Big data

Para construir un modelo de IA, se necesitan muchos datos con el objeto de entrenarlo, validarlo y, finalmente, testarlo. No obstante, al igual que para lograr un plato de alta cocina se necesitan los mejores ingredientes, para construir un modelo de IA se necesitan datos muy fiables. De lo contrario, se obtendrá un modelo con nula capacidad predictiva.

La primera dificultad que encontramos respecto a los datos de interés es que están compartimentados. Los datos clínicos se encuentran en la historia clínica; los parámetros clínicos, en medio de los análisis; y los datos radiológicos, en diferentes pruebas de imagen. Se encuentran desestructurados, al no poder ser utilizados directamente por los modelos algorítmicos de IA. Deben estructurarse previamente, lo que significa que ha de convertirse toda la información de interés en datos numéricos, que es el lenguaje que entienden los modelos de IA. Este proceso se denomina *feature engineering* (procesamiento de las variables independientes o atributos), y es la parte más costosa en la construcción de un modelo de IA.

El segundo aspecto a tener en cuenta es que las diferentes subpoblaciones de la base de datos deben estar balanceadas, de lo contrario se pueden generar sesgos. Si quisiéramos entrenar un modelo de IA a leer retinografías de pacientes diabéticos, en la fase de entrenamiento, deberíamos suministrarle aproximadamente el mismo número de retinografías de cada clase de RD. Además, procuraríamos que existiera un equilibrio en cuanto a sexo y grupo étnico de los pacientes. No obstante, puede ocurrir que no sea posible obtener subpoblaciones balanceadas. Delante de este escenario, cabe decir que existen técnicas de IA para equilibrar subpoblaciones mediante técnicas de aumentación o de reducción de los subgrupos minoritarios o dominantes, respectivamente.

Aspectos no resueltos de la inteligencia artificial

Algoritmos sesgados

Siempre que se trabaja con muestras poblacionales, se cometen sesgos, porque la muestra escogida, aunque sea al azar, no es del todo representativa de la población de la cual procede. Este tipo de sesgos son involuntarios e inherentes al muestreo. Por lo tanto, no son evitables. El segundo tipo de sesgos que sufren los algoritmos proceden de los datos que se utilizan para entrenarlos. Si los datos son sexistas o racistas, los algoritmos también lo serán. Existen varios ejemplos notorios de ello.

Falta de explicabilidad del resultado

En muchas situaciones, sabemos que el modelo de IA funciona, pero no es posible justificar el resultado para un caso concreto. Imaginemos que tenemos un *fuzzy random forest* (árboles de decisión *fuzzy*), que es un modelo de IA apto para estimar el riesgo de una persona diabética a desarrollar RD basado en sus factores de riesgo. Si dicho modelo considera que el riesgo de desarrollar RD para un determinado paciente es bajo, propondrá incrementar el intervalo de tiempo para el siguiente chequeo, pero no será posible explicarle al paciente exactamente el porqué de la decisión del modelo.

Nadie sabe muy bien qué ocurre en las *hidden layers* de las redes neuronales y, en consecuencia, nadie sabe cómo toma el algoritmo sus decisiones; es lo que viene llamándose el *black box* (caja negra). Este punto puede llegar a incomodarnos como oftalmólogos cuando tengamos que tomar decisiones importantes basándonos en algoritmos de IA. Del mismo modo, los pacientes pueden mostrarse reticentes a poner su salud en manos de algoritmos que no expliquen sus decisiones de manera entendible.

Siendo conscientes de la importancia de poder explicarnos y explicar a nuestros pacientes en qué se basan los algoritmos para tomar las decisiones, las empresas desarrolladoras de modelos de IA están destinando muchos recursos a investigar esta cuestión tan crucial con el objeto de entender las recomendaciones de los algoritmos para cada caso concreto y, por lo tanto, poder aceptar mejor sus recomendaciones en un futuro próximo.

No es suficientemente robusta

La mayoría de los modelos de IA que están operativos en el mercado son bastante fiables, pero no lo suficiente como para que funcionen de manera completamente autónoma. Todavía se exige su control por parte de los seres humanos.

Responsabilidad del funcionamiento de los modelos algorítmicos

No existe por el momento una legislación específica que regule aspectos legales relacionados con la IA. Y este es un aspecto crucial, ya que pronto los modelos algorítmicos estarán presentes en todas nuestras facetas cotidianas. Establecer el marco de responsabilidad es básico si queremos confiar en ellos. Y tan importante como este punto es definir los procesos de monitorización que se van a exigir a los modelos para detectar cualquier anomalía, y de este modo, evitar perjudicar a nuestros pacientes.

Conclusiones

Desde el año 2014, existe un interés exponencial por la IA, al descubrir el potencial de las redes neuronales. La oftalmología es la primera especialidad médica en beneficiarse de la IA. En un breve plazo de tiempo, habrá algoritmos que nos ayudarán a diagnosticar, monitorizar y tratar a los pacientes de una manera más individualizada. No obstante, es necesario asegurar su eficacia y seguridad, su trasparencia e imparcialidad, antes de ser implementados de manera global.

Bibliografía

1. McCarthy J, Minsky ML, Rochester N, Shannon CE. A proposal for the Dartmouth summer research project on artificial intelligence. *AI Mag.* 2006;27(4).
2. Turing AM. Computing Machinery and Intelligence. *Mind.* 1950;49:433-60.
3. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature.* 1986;323:533-6.
4. Goodfellow I, Bengio Y, Courville A. Deep Learning. *Google Books.* London: MIT Press; 2016.