Revista Latinoamericana de Hipertensión
ISSN: 1856-4550
latinoamericanadehipertension@gmail.com
Sociedad Latinoamericana de Hipertensión
Venezuela

Una introducción a las aplicaciones de la inteligencia artificial en Medicina: Aspectos históricos

Arias, Víctor; Salazar, Juan; Garicano, Carlos; Contreras, Julio; Chacón, Gerardo; Chacín-González, Maricarmen; Añez, Roberto; Rojas, Joselyn; Bermúdez-Pirela, Valmore

Una introducción a las aplicaciones de la inteligencia artificial en Medicina: Aspectos históricos
Revista Latinoamericana de Hipertensión, vol. 14, núm. 5, 2019
Sociedad Latinoamericana de Hipertensión, Venezuela
Disponible en: https://www.redalyc.org/articulo.oa?id=170262877013

Artículos

Una introducción a las aplicaciones de la inteligencia artificial en Medicina: Aspectos históricos

An introduction to artificial intelligence applications in medicine: Historical aspects

Redalyc: https://www.redalyc.org/articulo.oa?id=170262877013

Víctor Arias
Universidad Nacional de Colombia, Colombia
http://orcid.org/0000-0002-2358-5908

Juan Salazar
Centro Investigaciones Endocrino-Metabólicas "Dr. Félix Gómez, Venezuela
http://orcid.org/0000-0003-4211-528X

Carlos Garicano
ESE Metrosalud, Antioquia, Colombia

Julio Contreras
Tecnológico de Antioquia, Colombia
http://orcid.org/0000-0002-5179-5400

Gerardo Chacón
Universidad Simón Bolívar, Colombia
http://orcid.org/0000-0003-3615-5787

Maricarmen Chacín-González
Universidad Simón Bolívar, Colombia
http://orcid.org/0000-0002-5208-9401

Roberto Añez
Centro Investigaciones Endocrino-Metabólicas "Dr. Félix Gómez, Venezuela
http://orcid.org/0000-0001-6363-2767

Joselyn Rojas
Pulmonary and Critical Care Medicine Department, Estados Unidos
http://orcid.org/0000-0003-4994-075X

Valmore Bermúdez-Pirela
Universidad Nacional de Colombia, Colombia
http://orcid.org/0000-0003-1880-8887

Resumen:

En un sentido amplio la inteligencia artificial y el aprendizaje automático se ha aplicado a los datos médicos desde los inicios de la informática dado el profundo arraigo de esta área en la innovación, pero los últimos años han sido testigo de una generación cada vez mayor de datos relacionados con las ciencias de la salud, cuestión que ha dado nacimiento a un nuevo campo de las ciencias de la computación llamado *big data*. Los datos médicos a gran escala (en forma de bases de datos estructuradas y no

estructuradas) si son apropiadamente adquiridos e interpretados pueden generar grandes beneficios al reducir los costos y los tiempos del servicio de salud, pero también podrían servir para predecir epidemias, mejorar los esquemas terapéuticos, asesorar a médicos en lugares remotos y mejorar la calidad de vida. Los algoritmos de *deep learning* son especialmente útiles para manejar esta gran cantidad de datos complejos, poco documentados y generalmente no estructurados; todo esto debido a que el *deep learning* puede irrumpir al crear modelos que descubren de forma automática las características principales, así como las que mejor predicen el comportamiento de otras variables dentro de una gran cantidad de datos complejos. En el futuro, la relación hombre-máquina en biomedicina será más estrecha; mientras que la máquina se encargará de tareas de extracción, limpieza y búsquedas de correlaciones, el médico se concentraría en interpretar estas correlaciones y buscar nuevos tratamientos que mejoren la atención y en última instancia la calidad de vida del paciente.

Palabras clave: Inteligencia artificial, innovación, registros médicos, bases de.

Abstract:

In a broad sense, artificial intelligence and machine learning have been applied to medical data since the beginning of computing given the deep roots of this area in innovation, but recent years have witnessed an increasing generation of data related to health sciences, an issue that has given birth to a new field of computer science called big data. Large-scale medical data (in the form of structured and unstructured databases) if properly acquired and interpreted can generate great benefits by reducing costs and times of health service, but could also serve to predict epidemics, improve therapeutic schemes, advise doctors in remote places and improve the quality of life. The deep learning algorithms are especially useful to deal with this large amount of complex, poorly documented and generally unstructured data, all this because deep learning can break when creating models that automatically discover the predictive characteristics of a large amount of complex data. In the future, the human-machine relationship in the medical evaluation will be narrower and complex; while the machine would be responsible for extraction, cleaning and assisted searches, the physician will be concentrate on both, data interpretation and the best treatment option, improving the patient´s attention and ultimately, quality of life.

Keywords: Artificial intelligence, innovation, medical records, databases.

INTRODUCCIÓN

El cerebro humano ha evolucionado a lo largo de decenas de miles de años para percibir e interpretar estímulos visuales. Sin embargo, cuando un ordenador intenta ejecutar tareas de reconocimiento de imágenes, los algoritmos secuenciales utilizados en otras áreas de la computación carecen de precisión en sus resultados [1]. Esto se debe a la incapacidad de tales algoritmos de transformar la información de bajo nivel (matriz tridimensional de números) a conceptos de alta complejidad tal como lo hace el humano (**Figura 1**); este problema se conoce como vacío semántico [2].

Figura 1. Una sección amplificada de una fotografía muestra como realmente una computadora percibe una imagen, en este caso la oreja de un perro descrita por una matriz de números.

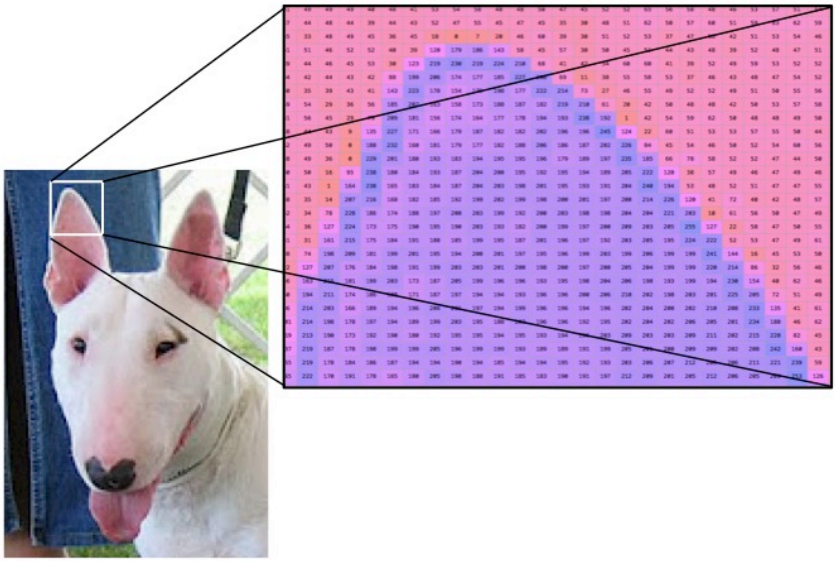

FIGURA 1.
Una sección amplificada de una fotografía muestra como realmente una computadora percibe una imagen, en este caso la oreja de un perro descrita por una matriz de números.

Una computadora percibe las imágenes mediante un Chip CMOS, como sucedáneo de nuestra retina, traduce la intensidad de luz a señales eléctricas que posteriormente se procesan y cuantifican en unidades llamadas pixeles [3]. En la Figura 1 por ejemplo, aunque fácilmente un ser humano puede identificar un perro en la imagen, para la computadora simplemente es una matriz numérica de 301x301x3 pixeles, es decir algo más de medio millón de números que oscilan entre 0 (negro) y 255 (blanco), El problema consiste en averiguar como el ordenador puede traducir esta matriz a un único concepto: "perro".

Para dar solución a este problema, nuestro cerebro no convierte las imágenes en pixeles, sino que las descompone en características como el color, la textura, forma, bordes, tamaño, ubicación, relación con otras imágenes, movimiento, continuidad, experiencias previas; integrando esta información en diferentes zonas de la corteza visual. Según la teoría más aceptada hasta la fecha, el modelo ventral-dorsal [1,4] propuesto por Milner y Goodale en 1992 y luego refinada por Norman en el 2000, plantea que existen dos vías bien definidas que transmiten información por las áreas visuales: a) La ruta dorsal: que comienza en el área V1 cruza a través de las áreas V2, dorso-medial (V6) y el área visual V5 y llega a la corteza parietal posterior. Esta vía se conoce como la "ruta del dónde" o "ruta del cómo", y está asociada al movimiento, representación y ubicación de los objetos; el control de los ojos (movimientos del ojo de tipo sacádico) y los brazos; b) La ruta ventral comienza en V1 y sigue a través de las áreas V2 y V4, y de allí a la corteza temporal inferior. La ruta ventral es llamada la "ruta del qué" y está asociada al reconocimiento y representación de los objetos, también está asociada con el almacenamiento de la memoria de largo término y el reconocimiento de patrones. Todas las áreas en la ruta ventral están influenciadas por factores extra-retinianos. Estos factores incluyen atención, memoria de trabajo y relevancia del estímulo. Por lo tanto, esta vía no solo proporciona una descripción de los elementos en el mundo visual, sino que también juega un papel crucial al juzgar la importancia de estos elementos.

Estas bases fisiológicas constituyen los elementos fundamentales del aprendizaje automático, el cual representa una potencial herramienta bioinformática para el procesamiento de datos en la actualidad. Por ello, en esta revisión se describe la evolución historia del deep learning, como un componente de la inteligencia artificial y su posible utilización en la medicina moderna.

Enfoque basado en datos o aprendizaje automático

Con el enfoque del aprendizaje automático se intenta emular la característica humana de experiencias previas, para esto se le proporciona al ordenador gran cantidad de imágenes cada una de estas con la etiqueta correspondiente al concepto que se desea reconocer (datos de entrenamiento) con el objetivo de entrenar un conjuntos de hipótesis o posibles técnicas que aprendan la apariencia visual de cada concepto, seleccionando aquella técnica que de la mejor solución, es decir que pueda predecir con la mayor precisión las etiquetas de nuevas imágenes no usadas en la etapa de entrenamiento (datos de prueba), este enfoque de acumulación de datos se representa en la Figura 2.

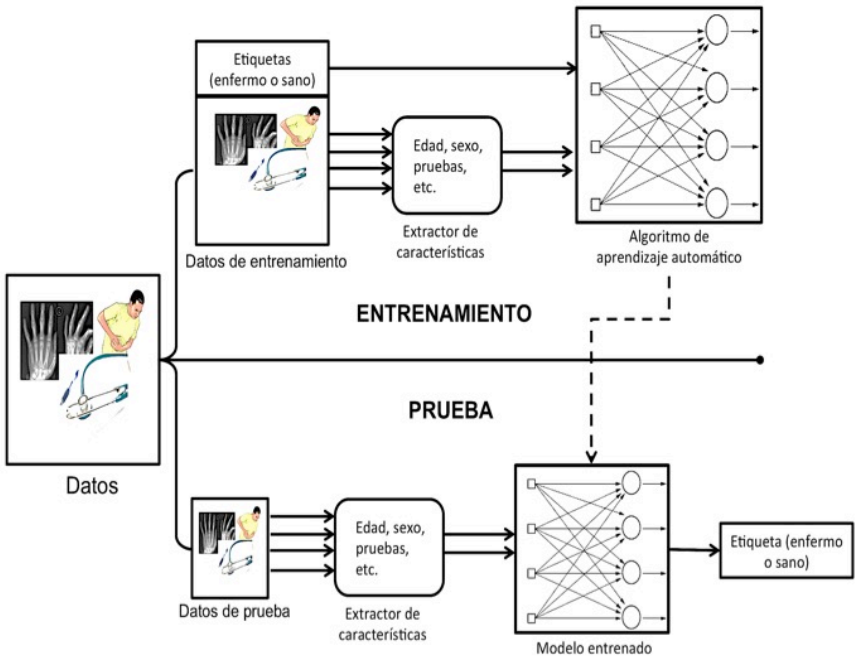

FIGURA 2

Para reconocer si un paciente tiene cierta enfermedad o no, el ordenador obtiene un conjunto de datos mixtos el cual divide entre datos de entrenamiento y de prueba, una vez el algoritmo seleccionado es entrenado, se evalúa la precisión del modelo al predecir etiquetas de datos de prueba.

Aprendizaje automático basado en el funcionamiento del cerebro

El término deep learning se refiere a un conjunto de métodos que permiten a una computadora descubrir automáticamente las características de alto nivel necesarias para la clasificación a partir de los datos en su estado natural utilizando múltiples capas de representación. El método intenta imitar la actividad en capas de neuronas en el neocórtex, sector que agrupa el ochenta por ciento del cerebro y donde ocurre el pensamiento. El software aprende, en un sentido muy real, a reconocer patrones en las representaciones digitales de sonidos, imágenes y otros datos [5].

El deep learning tiene el potencial de transformar una gama de sectores, no menos importante, el relacionado con la salud. Donde se le conoce como medicina de caja negra, porque aunque el algoritmo es capaz de diagnosticar lesiones cutáneas potencialmente malignas con la misma precisión que un dermatólogo certificado [6], las reglas para diagnosticar si es benigna o no son definidas por él mismo, y a menudo sin dejar un registro claro que explique sus decisiones [7]. Aun así, las ventajas son mucho mayores, el deep learning junto a los especialistas del sector pueden hacer del diagnóstico médico una tarea más rápida y precisa (inmensas posibilidades para mejorar los diagnósticos, la creación de vías de atención y la reproducibilidad en los procedimientos quirúrgicos para, en última instancia, lograr mejores resultados clínicos.), por ejemplo, Arterys, un sistema de imagenología cardíaca asistido por Inteligencia Artificial (IA) y entrenado con 3.000

casos cardiacos en los que se analizó el corazón y el flujo sanguíneo, al estar conectado a una máquina de resonancia magnética, puede examinar el flujo sanguíneo y las imágenes obtenidas para generar contornos editables proporcionando una imagen precisa de un corazón en segundos, un proceso que dura normalmente una hora de trabajo manual y que no requiere un pensamiento creativo, esto redunda en que el especialista tendrá más tiempo para dedicarse a idear otros tratamientos potenciales [8].

Breve historia del deep learning

El perceptrón

Una red neuronal biológica se considera el sistema mejor organizado para procesar información de diferentes sentidos tales como la vista, el oído, el tacto, el gusto y el olfato de una manera eficiente. Uno de los mecanismos clave para el procesamiento de la información en un cerebro humano es la complicada información de alto nivel que se procesa mediante las conexiones (llamadas sinapsis), de un gran número de elementos estructuralmente simples (llamados neuronas). En el aprendizaje automático, las redes neuronales artificiales son una familia de modelos que imitan la elegancia estructural del sistema neural y aprenden patrones inherentes a las observaciones [9], la similitud puede apreciarse en la Figura 3.

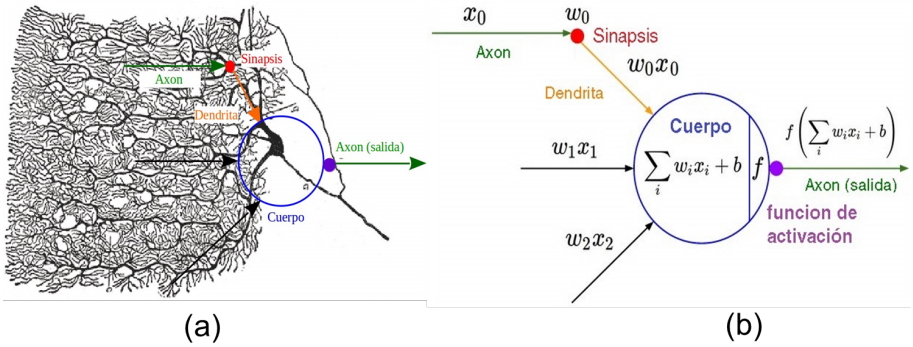

FIGURA 3

(a) Diagrama que muestra como el funcionamiento y la estructura de las redes neuronales biológicas inspiraron el funcionamiento netamente matemático del primer modelo de neurona artificial, el perceptrón (b).

La idea nace de los trabajos del psicólogo Frank Rosenblatt quien desarrolló el perceptrón [10], el primer algoritmo para entrenar una red neuronal basado en los trabajos de McCulloch y Pitts [11], que aunque no sigue exactamente el funcionamiento de las redes biológicas, produce un resultado simple, una neurona suma las múltiples entradas ponderadas por la medida de la fuerza de conexión sináptica (parámetros), la suma se hace pasar por una función de activación no lineal o estado de la neurona (las funciones no lineales se usan para modelar dinámicas relativamente complejas), múltiples neuronas pueden apilarse en una capa para generar múltiples salidas, los parámetros de esas neuronas son recalculados con el fin de disminuir la medida de la diferencia entre la salida deseada y la salida real, conocida como error (Figura 4).

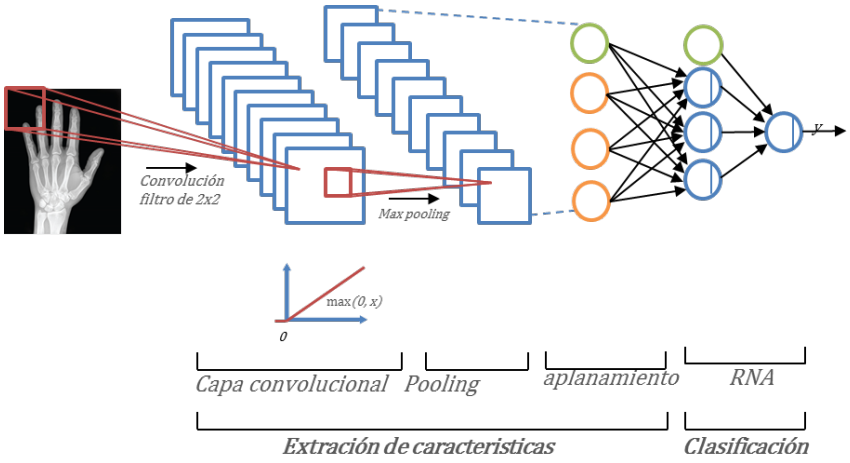

FIGURA 4.
Cada neurona aprende a identificar un número, la medida de la diferencia de las salidas se conoce como error que ajusta los parámetros de forma optima

Posteriormente en 1969 Minsky publicó junto con Papert la obra "Perceptrons" [12], donde mostraron todas las limitaciones de los perceptrones, en particular el aprendizaje en clases que no son linealmente separables como la función XOR (función lógica OR exclusiva), el argumento principal de Minsky en contra de los perceptrones, era que el algoritmo de aprendizaje de Rosenblatt no trabajaba para múltiples capas, ya que solo especifica la salida correcta para la capa de salida, así que no era posible calcular los parámetros correctos para las capas intermedias, esto llevo a las redes neuronales a un periodo de pocos progresos.

Las redes neuronales pueden organizarse en múltiples capas

Durante varios años las perspectivas no fueron buenas para las redes neuronales pero, ¿Por qué?, La idea del conexionismo era usar muchas capas de neuronas matemáticamente simples para resolver problemas complejos. La solución fue el desarrollo de un procedimiento de aprendizaje para redes neuronales más sofisticadas que los simples perceptrones, creando representaciones internas o subprocesos de una tarea en particular, estas representaciones son modeladas por capas "ocultas" con estados no especificados por la tarea tal como se muestra en la Figura 4, el nivel de sofisticación de estas redes permite al menos teóricamente implementar cualquier función computable [13].

Propagación hacia atrás

Ahora que se conocían las ventajas y utilidad de una arquitectura multicapa en una red neuronal, se debía buscar una forma práctica para el ajuste de los parámetros de las capas ocultas con los ejemplos de entrenamiento, con la finalidad de minimizar el error a un valor cercano a cero. Una manera elegante de resolver este problema fue el diseño de un algoritmo llamado "propagación hacia atrás", que utiliza la regla de cadena (ecuación utilizada en cálculo para hallar la derivada de funciones anidadas), es decir dado que la salida depende de la salida de las neuronas en la capa oculta, el cálculo asigna parte de la "culpa" del error a las neuronas de la capa oculta inmediatamente anterior, esta a su vez a la anterior si existe, así hasta propagar el error hacia atrás. De esta forma se conoce cuanto cambia el error en la salida si se cambia cualquier parámetro de la red, incluidos aquellos en las capas ocultas, por último para encontrar los parámetros óptimos se recurre por lo general, a una técnica llamada gradiente descendente estocástico (SGD).

Aunque varias investigaciones usaron esta técnica en computadoras [14], no fue hasta la tesis doctoral de Paul Werbos que se estudió en profundidad el entrenamiento de redes neuronales multicapas usando la propagación hacia atrás [15], aunque a su aplicación no se le prestó atención debido a los efectos adversos del periodo oscuro de las redes neuronales. Sin embargo, su trabajo fue descubierto y popularizado una década después por Rumelhart, Hinton y William1 [6], quienes siguiendo esta línea publicaron otro artículo

donde presentaban el nuevo algoritmo de aprendizaje y abordaron los problemas discutidos por Minsky en "Perceptrons" [12]. Esta formulación hizo comprender cómo las redes de neuronas multicapa podían ser entrenadas para abordar complejos problemas de aprendizaje [17].

Las redes neuronales aprenden a ver

Las redes neuronales resultan ser efectivas al entrenarse con data estructurada, por ejemplo: base de datos de precios de casas en función de sus características, la relación entre fenotipos metabólicos y los niveles de insulina plasmática. Pero se les dificulta manejar datos no estructurados como una imagen médica, una pieza de sonido o un historial médico electrónico [18].

En el caso de datos no estructurados, por ejemplo una imagen, está compuesta de una matriz de datos, donde cada dato representa la intensidad de un pixel, cada pixel se convierte en una entrada a la red, esto causa que la neurona solo puede recibir un valor por cada entrada, por lo tanto la imagen debe ser aplanada, es decir cada fila de la matriz se apila al lado de la siguiente, convirtiéndose en un vector muy largo; si se trabaja con una imagen de 40x40 pixeles y suponiendo que se encuentra en escala de grises, se tendrá 1600 variables de entrada. El aplanamiento causa que se pierda toda relación espacial de un pixel con sus pixeles vecinos, en esta relación está la clave de muchos algoritmos de pre-procesamiento de imágenes [19-21].

Las Redes Neuronales Convolucionales (CNN, pos sus siglas en inglés) funcionan de forma diferente pues en lugar de aplanar la imagen y usar un vector de pixeles usan una matriz más pequeña conocida como filtro o detector de características, en la Figura 5 se muestran filtros como una matriz de 2x2, la cual pondera a cada subconjunto de 2x2 en la matriz de la imagen, la suma de todas las ponderaciones se almacena como un dato de una nueva matriz conocida como mapa de características, se crean distintos mapas de características usando valores distintos en la matriz de filtros para construir la primera capa convolucional.

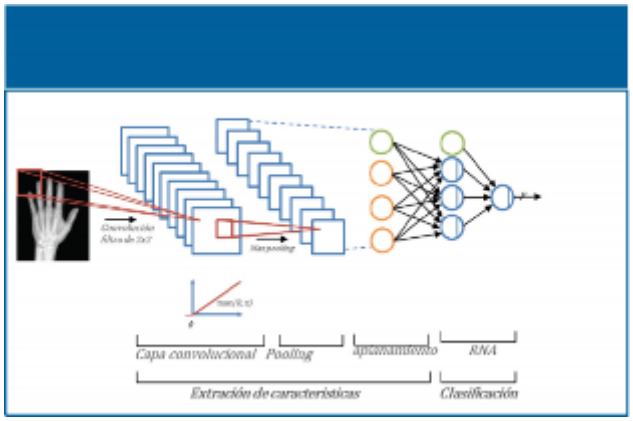

FIGURA 5
Arquitectura de una Redes Neuronales Convolucionales

n no lineal (ReLU) al mapa de características y al igual que las redes neuronales es necesaria una función no lineal a la salida, dada la naturaleza de los datos de entrada, en general se considerara al mapa de características junto a la función no lineal ReLU como una única capa convolucional. La siguiente capa se conoce como pooling o muestreo, la cual básicamente toma una sección del mapa de características y extrae un valor representativo, en el caso del max pooling el máximo valor en la sección muestreada. La capa pooling es una versión reducida del mapa de características, esta capa es importante dado que dota al modelo de invariabilidad, es decir si el objetivo es detectar una microfractura, el tipo de microfractura, el tamaño de la microfractura, los artefactos usados para la radiografía, la posición de la microfractura, entre otros; es irrelevante para el modelo y siempre intentará detectarla como una microfractura.

La red mostrada en la Figura 5 exhibe el esquema básico de una secuencia convolucional-ReLU-muestreo, donde la última capa pooling se aplana y sirve como entrada a una red neuronal que se encarga de la

clasificación. En este caso, el entrenamiento de los parámetros que forman los filtros CNN es similar al de las redes neuronales. El desarrollo de esta topología vino en 1989, cuando LeCun et al., en los laboratorios Bell de AT&T demostró una muy significativa aplicación en la detección de números escritos a mano con el fin de reconocer el código postal en una carta [22] . Las CNN se inspiraron en la naturaleza jerárquica de las redes neuronales en la corteza visual descubierta por Huber y Wiesel [23] donde una célula en una etapa superior tiende a responder selectivamente a una característica más complicada del patrón de estimulo y, al mismo tiempo, tiene un campo receptivo más grande, y es más insensible al cambio de posición del patrón de estímulo. A mediados de los 90's los trabajos de LeCun resultaron en la mayor aplicación comercial de las CNN hasta esa fecha, la lectura automática de cheques [24] que a finales de los 90's procesaban cerca del 20% de los cheques en los EEUU.

Las redes neuronales comprenden el lenguaje hablado y escrito

Al igual que la escritura, la compresión del lenguaje (en cualquier idioma) por una máquina es otro problema difícil de resolver debido a las casi infinitas variaciones que podemos obtener al pronunciar palabras (acento, tono de voz, velocidad, entonación, etc.), sumado a todo esto, se sobre-impone otro desafío que resolver: largas secuencias de entrada generadas en tiempo real. En el caso de las imágenes, resulta relativamente fácil usar una red neuronal que sustraiga e identifique una letra dentro una imagen, sin embargo, el proceso que debe realizarse con un archivo de audio o una conversación no es tan simple ya que: primero separar el habla en caracteres es completamente impráctico e incluso encontrar palabras individuales dentro de una conversación no es fácil y segundo el contexto de la conversación (si se comprende) hace mucho más fácil encontrar el significado de cada palabra en contraposición a reconocer y definir cada palabra individualmente.

Si bien la estructura de una CNN funciona bastante bien para las imágenes, no es en absoluto adecuado para largos flujos de información continua con significado en tiempo y espacio (como el audio o el texto), pues la red neural carece de "memoria" que permita que una entrada pueda afectar a otra entrada procesada posteriormente. En otras palabras, una red capaz de procesar lenguaje debe ser capaz de manejar una cadena de palabras o de sonidos que se suman en el tiempo y que globalmente entregan un significado en lugar de una sola entrada de gran información que se entrega completa de una sola vez.

Para abordar el problema de la comprensión del habla, los investigadores trataron de modificar las redes neuronales para procesar la entrada como un flujo de datos continuos (en lugar de grandes lotes como en las imágenes). Para ello crearon una versión aplanada de la red neuronal (Figura 6), a la vez que la capa oculta tiene una bifurcación que se redirige al cuerpo de la neurona. La Figura 7 muestra la forma común de representar una Red Neuronal Recurrente (RNN, por sus siglas en inglés), que al desenrollarse exhibe múltiples copias de la misma red que pasan el mensaje de una copia a otra. Las bases de este tipo de arquitectura de red neuronal se remontan a los trabajos de Waibel et., al en redes neuronales con retardos de tiempo [25] .

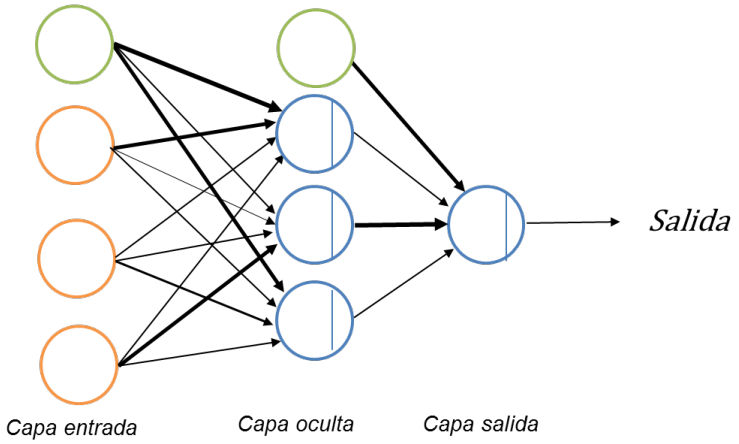

FIGURA 6.
Arquitectura simple de una red neuronal multicapa

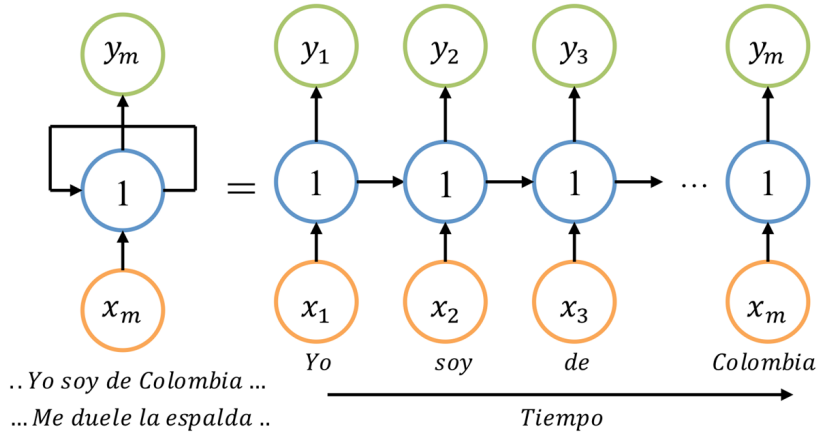

FIGURA 7.
Arquitectura de una Red Neuronal Recurrente

En este sentido, las RNN son bastantes similares a las CNN, pero en lugar de adquirir todos los datos al mismo tiempo, cada neurona observa solo un subconjunto de la entrada por ejemplo en la Figura 7: "Yo soy de Colombia", sería un subconjunto visto por la red en un tiempo determinado, a cada uno de estos subconjuntos se le aplica el mismo cálculo, pero hasta allí llegan las similitudes. En las CNN no existe el concepto de redes que se bifurcan sobre si mismas en el tiempo, además en las RNN hay entradas y salidas secuenciales de datos en lugar de un recorrido por toda la data de entrada para generar un resultado como es el caso de las capas de las CNN. Dada la naturaleza cíclica de la red, el concepto de propagar el error hacia atrás no aplicaría, la solución es desenrollar la red tal como se mostró en la Figura 8, tratando cada bucle a través de la red neuronal como una entrada a otra red neuronal por un número limitado de veces, proceso conocido como propagación hacia atrás a través del tiempo [26].

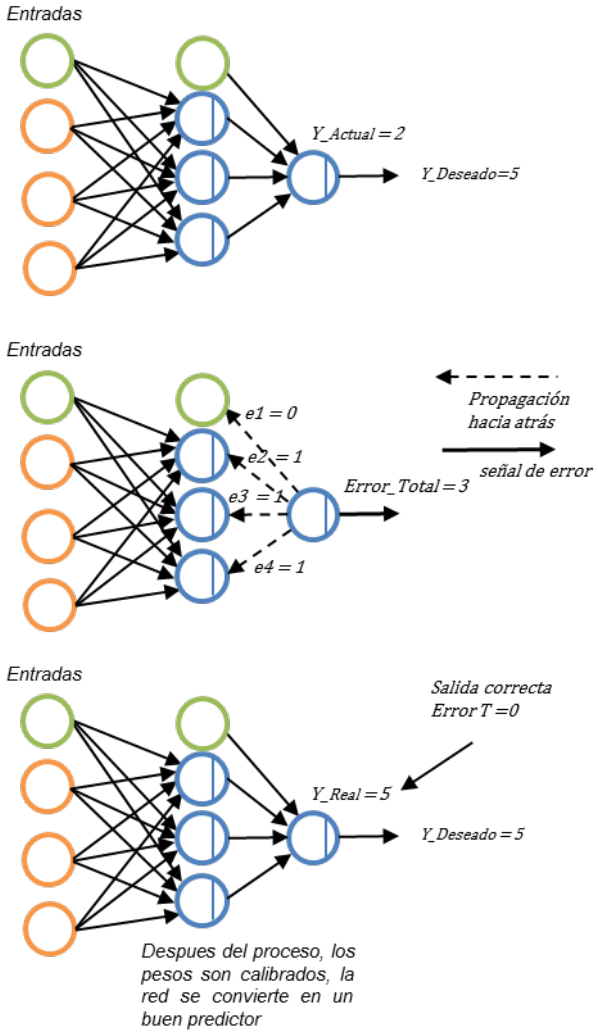

FIGURA 8.
Algoritmo de propagación hacia atrás del error

En experimentos posteriores, Bengio demostraría la dificultad de entrenar una RNN de forma óptima [27], debido a que los parámetros de la red solo tienen en cuenta dependencias a corto plazo. Por ejemplo, considerando un modelo de lenguaje que intenta la predicción de una palabra basada en las palabras anteriores en una frase: si el objetivo es predecir la última palabra en "Jorge está durmiendo en su cama" no se necesita otro contexto para saber que la última palabra es cama. En este caso cuando la brecha de información relevante y el lugar que se necesita es pequeño, las RNN pueden aprender a usar la información pasada. Pero en el caso donde la brecha entre la información relevante y el punto donde se necesita puede llegar a ser grande, como en la frase "Soy Colombiano... hablo con fluidez español", el contexto se encuentra en Colombiano, pero dada la distancia es posible que la red no trabaje correctamente. Este fenómeno se puede observar detalladamente en la Figura 9, pues el algoritmo de propagación hacia atrás no trabaja adecuadamente para RNN con largas dependencias (en realidad, la propagación hacia atrás no trabaja bien para redes muy largas).

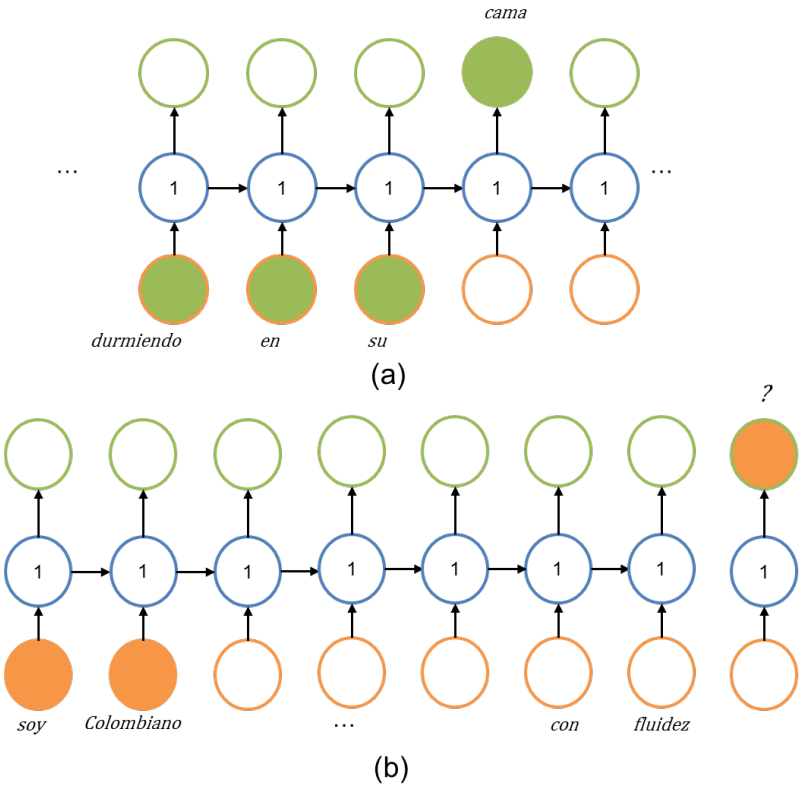

FIGURA 9.
(a) Red Neuronal Recurrente con dependencias cortas
(b) Red Neuronal Recurrente con dependencias largas

Arquitecturas profundas

Diverso resultados teóricos revisados sugieren que para aprender el tipo de funciones complicadas que pueden representar abstracciones de alto nivel (por ejemplo, visión y lenguaje), se requieren arquitecturas profundas, este tipo de arquitecturas requieren muchas menos neuronas en total, resultando en una reducción del número de parámetros, lo que permite entrenar redes con una base de datos relativamente pequeña29. Pero más importante aún, en comparación con la arquitectura superficial que necesita un extractor de características "bueno", diseñado principalmente por las habilidades de ingeniería o conocimiento de dominio experto, los modelos profundos son buenos para descubrir las características automáticamente de forma jerárquica [9].

Vanishing gradient problem

La clave del "deep learning" es tener muchas capas en los sistemas actuales de hasta veinte o más, pero ya a finales de los ochenta se sabía que las redes neuronales con muchas capas entrenadas con propagación hacia atrás simplemente no funcionaban bien, y en particular no funcionaba tan bien como las redes con menos capas [29]. Pues bien, teniendo en cuenta que el error en la salida se propaga hacia atrás, todas las neuronas en la red actualizan los parámetros de aprendizaje en función del valor de la culpa adjudicado a cada neurona. Ahora al actualizarse, esos parámetros son multiplicados por las salidas de una neurona para convertirse en la entrada de la próxima y así sucesivamente hasta llegar a la salida, el problema ocurre cuando el error empieza a disminuir, los valores del error que se propagan hacia las capas más cercanas a la entrada son muy pequeños, y por lo tanto los parámetros de aprendizaje en esas capas se actualizan de forma más lenta, el resultado es que luego de varias iteraciones las capas más próximas a la entrada no alcanzan a actualizarse hasta sus valores óptimos en un tiempo razonable, y más importante aún, son estas capas las responsables de extraer las

características básicas que luego serán usadas por las capas de orden superior, este es el problema que dificulta a las RNN tener memoria de largo plazo [30,31].

En el caso de las RNN. la solución llego usando un nuevo concepto conocido como Memoria de Corto y Largo Plazo (LSTM, por sus siglas en inglés) trabajo hecho por dos importantes investigadores en RNN Schmidhuber y Hochreiter [32], pero, esto hizo poco para arreglar el problema más grande, la percepción de que las redes neurales eran poco fiables y no trabajaron muy bien. Las computadoras además no tenían la suficiente velocidad de procesamiento y no había suficientes datos, esto hizo que los investigadores perdieran la fe en este enfoque de IA. De esta manera, mediados de los años 90 el estudio de las redes neuronales se estancó de nuevo y un método denominado Maquinas de Vectores de Soporte (SVM, por sus siglas en inglés), que en términos muy sencillos podría describirse como una forma matemáticamente óptima de formación de un equivalente a una red neuronal de dos capas, se desarrolló y empezó a considerarse superior a las complicadas redes neuronales [33].

La conspiración del "deep learning"

Con el ascenso de las SVM, el vanishing gradient problem, la falta de datos y una velocidad de procesamiento insuficiente, el nuevo milenio fue un tiempo oscuro para la investigación en redes neuronales. Pero todo empezó a cambiar cuando el Instituto Canadiense para Estudios Avanzados apoyó a un pequeño grupo de investigadores para que siguiera trabajando en el tema. Como lo diría Geoffrey Hinton, quien dirigía el proyecto, ellos tramaron una conspiración; dado el estigma de las redes neuronales se decidió "rebautizar" el campo con el nombre de deep learning.

Pero más importante que el nombre fue la idea, en su trabajo revolucionario de 2006 [34], el equipo demostró que las redes neuronales con muchas capas podría ser bien entrenadas si los parámetros se inician de una manera inteligente en lugar de aleatoria, que era la manera como se hacía hasta ese momento, pero ¿Cuál es la manera inteligente de inicializar los parámetros?, la solución era usar un nuevo algoritmo, las máquinas restrictivas de Boltzmann.

Máquinas Restrictivas de Boltzmann (RBM)

Las RBM son redes neuronales superficiales de dos capas, la primera capa se denomina capa visible (o, de entrada) y la segunda, denominada capa oculta. Las RBM se caracterizan por el hecho que no existe comunicación intra-capa entre nodos, en este punto los parámetros son inicializados aleatoriamente; si los datos son imágenes en escala de grises, por ejemplo, cada pixel es asignado a cada nodo de la capa visible, la multiplicación de cada uno de estos pixeles por los parámetros en la red, sirven como entrada a la capa oculta, alimentando la función no lineal de salida de cada una de las neuronas de esta capa, la salida de una RBM se conecta a la entrada de la siguiente, así sucesivamente hasta la capa final que genera la salida deseada (Figura 10).

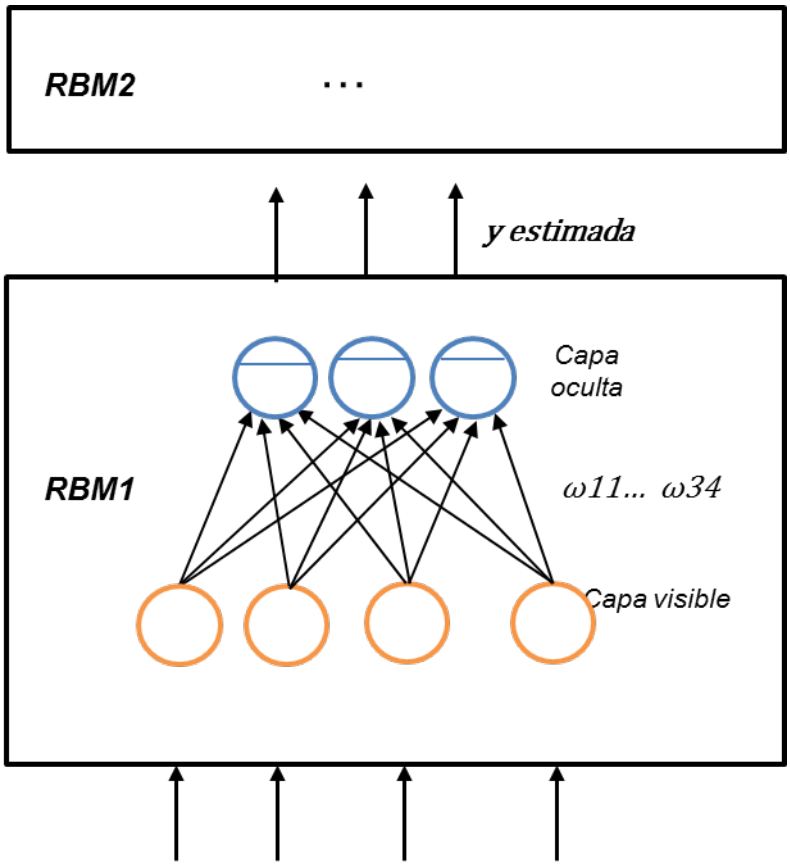

FIGURA 10
Figura 10

La siguiente fase es una reconstrucción de la base de datos de una forma no supervisada (no supervisada se refiere al hecho de que el modelo no conoce el resultado correcto en un ejemplo de entrenamiento, es decir que no se conoce a la salida). El procedimiento es usar el resultado de la RBM como entrada a la capa visible, una vez que cada salida es multiplicada por los parámetros intermedios, la sumatoria alimenta la función no lineal que genera una reconstrucción de la entrada, esto es, una aproximación al valor de los pixeles de la imagen, tal como se muestra en la Figura 11 [35].

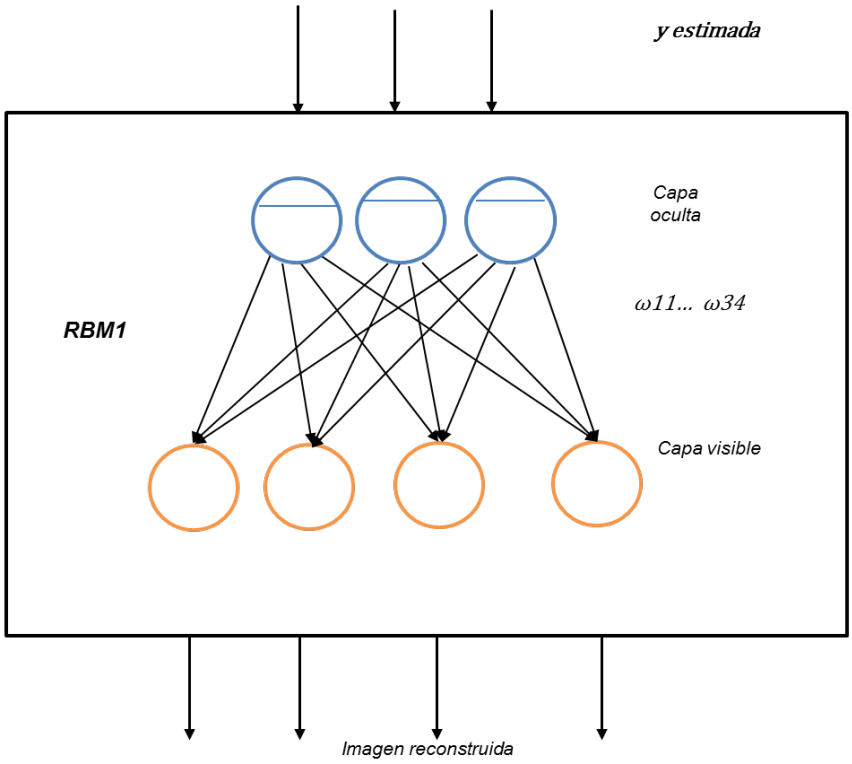

FIGURA 11.
Figura 11.

Dado que los parámetros de una RBM se inicializan aleatoriamente, la medida de la diferencia entre la imagen reconstruida y la original (conocida como error de reconstrucción) es grande, este error se propaga hacia atrás para ajustar los parámetros en un proceso iterativo igual al aplicado en las RNA estándar. La salida de la Figura 10 se puede definir como la probabilidad de que la salida sea dada parametrizado por : , en el caso de la Figura 11, el resultado es la probabilidad de que la salida sea dado parametrizada por : .

El proceso de reconstrucción estima la distribución de probabilidad de la entrada original; esto es, los valores de muchos pixeles a la vez, este tipo de algoritmos se conocen como modelos generativos [36] , que difieren al llamado aprendizaje discriminatorio realizado por la clasificación, que discrimina las entradas entre clases o etiquetas. Las RBM utilizan la divergencia de Kullback-Leibler para medir la distancia entre la distribución de probabilidad estimada y la distribución real de la entrada [37] , midiendo las áreas no superpuestas o divergentes bajo las dos curvas, tal como se muestra en la Figura 12.

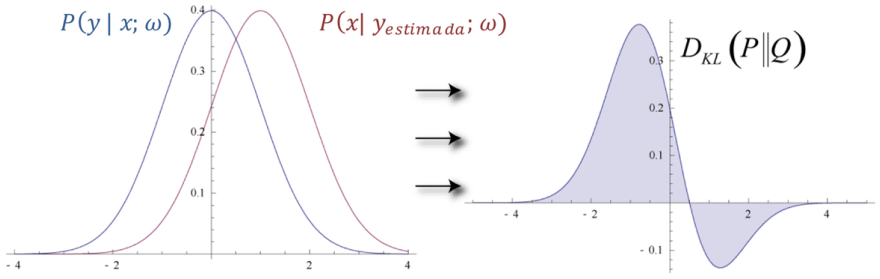

FIGURA 12.
Figura 12.

El algoritmo de propagación hacia atrás de las RBM intenta minimizar esas áreas de modo que las entradas ponderadas por los parámetros compartidos, produzcan una aproximación cercana a la entrada original. A la izquierda de la Figura 12 está la distribución de probabilidad de un conjunto de entrada original , yuxtapuestos a la distribución reconstruida ; mientras que a la derecha, la integración de sus diferencias. Mediante el ajuste iterativo de los pesos según el error que producen, un RBM aprende a aproximar los datos originales. Se podría decir que los pesos vienen lentamente a reflejar la estructura de la entrada, que se codifica en las activaciones de la primera capa oculta. El proceso de aprendizaje se parece a dos distribuciones de probabilidad que convergen paso a paso [38].

Bengio et al. [39], continuaron sus investigaciones y en un reporte en 2007 argumentaron que los métodos con muchos pasos de procesamiento, o de manera equivalente con representaciones jerárquicas de los datos, son más eficientes para problemas difíciles que los métodos superficiales como las SVM. También presentaron razones de la importancia del pre-entrenamiento no supervisado, y concluyeron que esto no sólo inicializa los pesos de una manera más óptima, sino lo que es más importante conduce a representaciones más útiles aprendidas de los datos.

Para el año 2012 los avances en el uso de Unidades de Procesamiento Gráfico (GPUs, por sus siglas en inglés) [40], la técnica de dropout para prevenir el sobre-entrenamiento [41] y el uso de la función de activación ReLU para prevenir el desvanecimiento del gradiente por saturación [42-45], fueron usados por una red convolucional profunda para batir el récord en 10,8% en el desafío de reconocimiento visual a gran escala en ImageNet con 1,3 millones de imágenes de alta resolución, cuyo objetivo es discriminar cerca de 1000 clases diferentes [46].

"Deep learning" y datos médicos

Es importante recordar que la IA (en el sentido amplio) y el aprendizaje automático se ha aplicado en el análisis de datos médicos, incluidos los estudios por imágenes desde los primeros días de la informática [47]. Los sistemas de diagnóstico asistido por ordenador han existido desde los años setenta [48], el procesamiento automatizado y el análisis de señales de tiempo unidimensionales (por ejemplo, electrocardiogramas) ha existido durante décadas [49], pero los últimos 10 años han sido testigo de una generación cada vez mayor de datos relacionados con las ciencias de la salud, cuestión que nos ha llevado a ingresar en un campo de las ciencias de la computación llamado big data y a una nueva disciplina llamada big data analysis [50].

Los datos médicos a gran escala (en forma de bases de datos estructuradas y no estructuradas) si son apropiadamente adquiridos e interpretados pueden generar grandes beneficios al reducir los costos del servicio de salud (lo cual ya en la actualidad se está utilizando), pero también podrían servir para predecir epidemias, mejorar los esquemas terapéuticos, asesorar a médicos en lugares remotos y mejorar la calidad de vida de pacientes. El objetivo ahora es entender la mayor información sobre un paciente y tan temprano en

su vida como sea posible, detectando señales de una enfermedad peligrosa en una etapa temprana en la cual que el tratamiento sea más simple (y menos costoso) que en una etapa mucho más avanzada [51].

Lamentablemente, la mayoría de los datos de salud están menos organizados y estandarizados que los datos de imágenes médicas. Por ejemplo, en muchos países aún las historias clínicas se encuentran con base analógica en papel y en aquellos países que cuentan con registros electrónicos, la información es altamente heterogénea y muchas veces, inconsistente, que incluye datos demográficos, diagnósticos, procedimientos, resultados de pruebas de laboratorio y medicamentos, así como notas clínicas no estructuradas en texto libre. En este contexto, es muy difícil al menos en la actualidad, para los modelos de aprendizaje profundo (como lo es para el cerebro humano) reconocer patrones confiables de información dispersa y ruidosa que de información estructurada.

De igual manera, resulta dispendioso el tiempo en que el médico debe ingresar información dentro del sistema (o en papel) que le suministra un paciente en vez de utilizar este tiempo en la valoración clínica del mismo o dedicar más tiempo al paciente en la educación o explicaciones sobre su patología. En este caso en particular, existe investigación muy activa en el desarrollo de un sistema basado en machine learning u otras técnicas de IA capaces de reconocer el lenguaje natural de la relación médico paciente y realizar las anotaciones pertinentes derivadas de la conversación entre ambos. Esto es de vital interés en el área de salud, pues ha sido consistentemente reportado que la documentación de datos clínicos, su búsqueda y posterior revisión es la causa principal de pérdidas en la productividad del médico en los EUA. De hecho, se estima que un médico en promedio invierte del 34 al 55% de su día laboral haciendo anotaciones y revisando registros médicos electrónicos en menoscabo de la interacción directa con sus pacientes [52].

Otro desafío relacionado con lo anterior es que los pacientes con un estatus socioeconómico bajo pueden tener datos particularmente no confiables, tales como información faltante o incorrecta dentro de los registros electrónicos de salud o más aún en los analógicos, debido a la recepción de atención fragmentada en múltiples instituciones que no están interconectadas. En estos casos, los algoritmos de aprendizaje profundo pueden ser de poca ayuda para estos pacientes reforzando las inequidades de salud ya existentes.

Los algoritmos de deep learning son especialmente útiles para tratar gran cantidad de datos, especialmente de naturaleza compleja, poco documentados y generalmente no estructurados, como por ejemplo imágenes, registros médicos electrónicos, datos de sensores, entre otros. El aprendizaje automático tradicional requiere la extracción de características de los datos antes de ser implementada sobre los modelos, esto adiciona el problema de la necesidad de un profundo conocimiento del área, y que, aun teniendo al personal idóneo, la gran cantidad de variables pueden desbordar la capacidad del profesional para encontrar nuevos patrones. En este escenario, el deep learning puede irrumpir al crear modelos que descubren de forma automática las características predictoras de una gran cantidad de datos complejos [53].

Aunque se encuentra en pleno desarrollo, la aplicación de estos algoritmos en el área médica ha dado resultados importantes a varios desafíos, tal es el caso de la detección de mitosis anormales en imágenes histológicas de cáncer de mama usando CNN profunda49. Otro avance importante es la clasificación de mutaciones con el fin conocer la probabilidad de que alguna alteración genética cause una enfermedad, esto se ha utilizado para descubrir nuevos determinantes genéticos del autismo, cánceres de tipo hereditarios y la atrofia muscular espinal [1,54]. El descubrimiento de compuestos bioactivos con los efectos farmacológicos y su modificación químico-estructural para mejorar su potencia y el diagnóstico precoz de la enfermedad de Alzheimer en su etapa prodrómica – el deterioro cognitivo leve- son aplicaciones que han despertado gran interés en la comunidad médica, que espera su rápido desarrollo y una mejor interoperabilidad en los modelos [55].

Son varias las áreas donde la inteligencia artificial ha logrado sus mejores aportes en el área de biomedicina, un área clásica de desarrollo ha sido la visión por computadora y el reconocimiento de objetos (OR). La primera se enfoca en la comprensión de imágenes y video, además de ocuparse de tareas como la clasificación,

detección y segmentación de objetos, que son útiles para determinar si un estudio por imágenes de un paciente contiene alguna estructura anormal como una tumoración. Muchos estudios han demostrado resultados prometedores en diagnósticos complejos que abarcan dermatología, radiología, oftalmología y anatomía patológica; los diagnósticos a nivel de imagen han tenido bastante éxito al emplear métodos basados en CNN, esto se debe en gran parte al hecho de que las CNN han logrado un desempeño igual al humano en tareas de clasificación de objetos, en las cuales aprenden a clasificar el objeto contenido en una imagen [56].

De manera similar, los algoritmos de detección y segmentación de objetos identifican partes específicas de una imagen que corresponden a objetos particulares, los métodos CNN toman los datos de la imagen como entrada y lo deforman iterativamente a través de una serie de operaciones convolucionales y no lineales hasta que la matriz de datos sin procesar original se transforma en una distribución de probabilidad sobre posibles clases de imagen. Sorprendentemente, los modelos de aprendizaje profundo han logrado una precisión a nivel médico en una amplia variedad de tareas de diagnóstico, incluida la identificación de lunares de melanomas, retinopatía diabética, riesgo cardiovascular y estudios de fondo de ojo y tomografía de coherencia óptica, la detección de lesiones mamarias en mamografías y la búsqueda de lesiones en la columna a partir de imágenes de resonancia magnética [57].

Una segunda área de gran interés es el procesamiento del lenguaje natural (PNL), el cual se centra en analizar el texto y el habla para inferir el significado de las palabras. En este caso, las RNN son efectivas en el procesamiento de entradas secuenciales, como el lenguaje, el habla y los datos de series de tiempo. Los éxitos más interesantes en las PNL incluyen la traducción automática, la generación de texto y la subtitulación de imágenes. En salud, las tecnologías de aprendizaje profundo y de lenguaje secuenciales potencian las aplicaciones en dominios como los registros de salud electrónicos ya mencionados. La historia clínica digitalizada de una gran organización médica puede capturar las transacciones médicas de más de 10 millones de pacientes a lo largo de una década [58,59], una sola hospitalización por sí sola generalmente genera unos 150.000 datos.

Es por esto que la aplicación de métodos de aprendizaje profundo a los datos registrados electrónicamente es un área en rápida expansión. En este momento, al hacer predicciones en medicina la mayoría de los estudios han utilizado aprendizaje supervisado en conjuntos limitados de datos estructurados, que incluyen resultados de laboratorio, signos vitales, códigos de diagnóstico y datos demográficos. Para dar cuenta de los datos estructurados y no estructurados que contienen las historias médicas digitales y analógicas (en papel), los investigadores están empezando a emplear enfoques de aprendizaje no supervisados en los cuales las redes se entrenan primero para aprender representaciones útiles mediante la compresión y luego la reconstrucción de datos sin etiquetar, para predecir diagnósticos específicos.

Conclusiones

El aprendizaje automático no es un dispositivo mágico que puede convertir datos en soluciones perfectas e inmediatas, más bien debe considerarse como una extensión de los procedimientos estadísticos que venimos utilizando desde hace décadas. Teniendo en cuenta la gran cantidad de información a la que un médico tiene acceso y probablemente tendrá que analizar, una decisión clínica puede ser una tarea abrumadora.

La calidad de los resultados que puede darnos un algoritmo depende de la cantidad y calidad de los datos con los que alimentamos nuestro sistema, por lo que el sistema de generación y recolección de datos es estratégico para resultados robustos y válidos. Esto es especialmente cierto en el cuidado de la salud ya que estos algoritmos tienen el potencial de afectar la vida de millones de pacientes. De seguro, el *deep learning* no sustituirá a los médicos en el futuro cercano. Pero al eliminar gran parte del trabajo aburrido y la memorización, facilitarán a los médicos a enfocarse en el cuidado de los pacientes y a tomar decisiones

dentro de un contexto de miles de resultados en el ámbito clínico que forman parte de toda nuestra evidencia científica acumulada en cientos de años de investigación.

REFERENCIAS

1. DiCarlo JJ, Zoccolan D, Rust NC. How does the brain solve visual object recognition? Neuron. el 9 de febrero de 2012;73(3):415–34.
2. Wan J, Wang D, Hoi SCH, Wu P, Zhu J, Zhang Y, et al. Deep Learning for Content-Based Image Retrieval: A Comprehensive Study. En: Proceedings of the 22Nd ACM International Conference on Multimedia [Internet]. New York, NY, USA: ACM; 2014 [citado el 18 de octubre de 2017]. p. 157–166. (MM '14). Disponible en: http://doi.acm.org/10.1145/2647868.2654948
3. Nilsson F. Intelligent network video: understanding modern video surveillance systems. Boca Raton: CRC Press; 2009. 389 p.
4. Leuba G, Kraftsik R. Changes in volume, surface estimate, three-dimensional shape and total number of neurons of the human primary visual cortex from midgestation until old age. Anat Embryol (Berl). el 1 de octubre de 1994;190(4):351–66.
5. Hof RD. Is Artificial Intelligence Finally Coming into Its Own? [Internet]. MIT Technology Review. [citado el 23 de octubre de 2017]. Disponible en: https://www.technologyreview.com/s/513696/deep-learning/
6. 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. el 2 de febrero de 2017;542(7639):115–8.
7. Brouillette M. AI diagnostics are coming [Internet]. MIT Technology Review. [citado el 23 de octubre de 2017]. Disponible en: https://www.technologyreview.com/s/604271/deep-learning-is-a-black-box-but-health-care-wont-mind/
8. Eastwood G. How deep learning is transforming healthcare [Internet]. Network World. 2017 [citado el 23 de octubre de 2017]. Disponible en: https://www.networkworld.com/article/3183745/health/how-deep-learning-is-transforming-healthcare.html
9. Suk H-I. An Introduction to Neural Networks and Deep Learning. En: Deep Learning for Medical Image Analysis [Internet]. Elsevier; 2017 [citado el 13 de agosto de 2017]. p. 3–24. Disponible en: http://linkinghub.elsevier.com/retrieve/pii/B978012810408800002X
10. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev. noviembre de 1958;65(6):386–408.
11. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys. el 1 de diciembre de 1943;5(4):115–33.
12. Minsky M, Papert S. Perceptrons. Oxford, England: M.I.T. Press; 1969.
13. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Netw. 1989;2(5):359–66.
14. Linnainmaa S. Taylor expansion of the accumulated rounding error. BIT Numer Math. el 1 de junio de 1976;16(2):146–60.
15. Werbos PJ. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. Harvard University; 1975.906 p.
16. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. el 9 de octubre de 1986;323(6088):533–6.
17. Rumelhart DE, Hinton GE, Williams RJ. Learning Internal Representations by Error Propagation. 1985 sep.
18. Bengio Y. Learning Deep Architectures for AI. Found Trends® Mach Learn. 2009;2(1):1–127.
19. Bay H, Ess A, Tuytelaars T, Van Gool L. Speeded-Up Robust Features (SURF). Comput Vis Image Underst. el 1 de junio de 2008;110(3):346–59.

20. Zhou H, Yuan Y, Shi C. Object tracking using SIFT features and mean shift. Comput Vis Image Underst. el 1 de marzo de 2009;113(3):345–52.

21. Dalal N, Triggs B. Histograms of oriented gradients for human detection. En: Computer Vision and Pattern Recognition, 2005 CVPR 2005 IEEE Computer Society Conference on [Internet]. IEEE; 2005 [citado el 11 de mayo de 2017]. p. 886–893. Disponible en: http://ieeexplore.ieee.org/abstract/document/1467360/

22. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation Applied to Handwritten Zip Code Recognition. Neural Comput. diciembre de 1989;1(4):541–51.

23. Hubel DH, Wiesel TN. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. J Neurophysiol. 1965;28(2):229–289.

24. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE. noviembre de 1998;86(11):2278–324.

25. Waibel A, Hanazawa T, Hinton G, Shikano K, Lang KJ. Phoneme recognition using time-delay neural networks. IEEE Trans Acoust Speech Signal Process. marzo de 1989;37(3):328–39.

26. Werbos PJ. Backpropagation through time: what it does and how to do it. Proc IEEE. octubre de 1990;78(10):1550–60.

27. Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Netw. marzo de 1994;5(2):157–66.

28. Schwarz G. Estimating the Dimension of a Model. Ann Stat. marzo de 1978;6(2):461–4.

29. Schmidhuber J. Deep Learning in Neural Networks: An Overview. Neural Netw. enero de 2015;61:85–117.

30. Bengio Y. A connectionist approach to speech recognition. Int J Pattern Recognit Artif Intell. el 1 de agosto de 1993;07(04):647–67.

31. Hochreiter S. {Untersuchungen zu dynamischen neuronalen Netzen. Diploma thesis, Institut f\\"{u}r Informatik, Lehrstuhl Prof. Brauer, Technische Universit\\"{a}t M\\"{u}nchen}. 1991;

32. Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Comput. el 1 de noviembre de 1997;9(8):1735–80.

33. LeCun Y, Jackel LD, Bottou L, Brunot A, Cortes C, Denker JS, et al. Comparison of learning algorithms for handwritten digit recognition. En: International conference on artificial neural networks [Internet]. Perth, Australia; 1995 [citado el 2 de mayo de 2017]. p. 53–60. Disponible en: https://pdfs.semanticscholar.org/d50d/ce749321301f0104689f2dc582303a83be65.pdf

34. Hinton GE, Osindero S, Teh Y-W. A Fast Learning Algorithm for Deep Belief Nets. Neural Comput. el 17 de mayo de 2006;18(7):1527–54.

35. DL4J. A Beginner's Tutorial for Restricted Boltzmann Machines - Deeplearning4j: Open-source, Distributed Deep Learning for the JVM [Internet]. [citado el 14 de agosto de 2017]. Disponible en: https://deeplearning4j.org/restrictedboltzmannmachine

36. Salakhutdinov R. Learning deep generative models [Internet]. University of Toronto; 2009 [citado el 14 de agosto de 2017]. Disponible en: http://www.cs.toronto.edu/~rsalakhu/papers/Russ_thesis.pdf

37. Kullback S, Leibler RA. On Information and Sufficiency. Ann Math Stat. 1951;22(1):79–86.

38. Hinton GE. Training Products of Experts by Minimizing Contrastive Divergence. Neural Comput. el 1 de agosto de 2002;14(8):1771–800.

39. Bengio Y, Lamblin P, Popovici D, Larochelle H, others. Greedy layer-wise training of deep networks. Adv Neural Inf Process Syst. 2007;19:153.

40. Raina R, Madhavan A, Ng AY. Large-scale Deep Unsupervised Learning Using Graphics Processors. En: Proceedings of the 26th Annual International Conference on Machine Learning [Internet]. New York, NY, USA: ACM; 2009 [citado el 2 de mayo de 2017]. p. 873–880. (ICML '09). Disponible en: http://doi.acm.org/10.1145/1553374.1553486

41. Srivastava N, Hinton GE, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res. 2014;15(1):1929–1958.

42. Jarrett K, Kavukcuoglu K, LeCun Y, others. What is the best multi-stage architecture for object recognition? En: Computer Vision, 2009 IEEE 12th International Conference on [Internet]. IEEE; 2009 [citado el 13 de mayo de 2017]. p.2146–2153. Disponible en: http://ieeexplore.ieee.org/abstract/document/5459469/

43. Dahl GE, Sainath TN, Hinton GE. Improving deep neural networks for LVCSR using rectified linear units and dropout. En: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. 2013. p. 8609–13.

44. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. En: Proceedings of the 27th international conference on machine learning (ICML-10) [Internet]. 2010 [citado el 13 de mayo de 2017]. p. 807–814. Disponible en: http://machinelearning.wustl.edu/mlpapers/paper_files/icml2010_NairH10.pdf

45. Glorot X. Apprentissage des réseaux de neurones profonds et applications en traitement automatique de la langue naturelle. 2015 [citado el 13 de mayo de 2017]; Disponible en: https://papyrus.bib.umontreal.ca/xmlui/handle/1866/11989

46. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. En: Pereira F, Burges CJC, Bottou L, Weinberger KQ, editores. Advances in Neural Information Processing Systems 25 [Internet]. Curran Associates, Inc.; 2012 [citado el 13 de mayo de 2017]. p. 1097–1105. Disponible en: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

47. Channin D. Deep Learning in Healthcare: Challenges and Opportunities [Internet]. The Mission. 2016 [citado el 14 de mayo de 2017]. Disponible en: https://themission.co/deep-learning-in-healthcare-challenges-and-opportunities-d2eee7e2545

48. Leaper DJ, Horrocks JC, Staniland JR, de Dombal FT. Computer-Assisted Diagnosis of Abdominal Pain using "Estimates" Provided by Clinicians. Br Med J. el 11 de noviembre de 1972;4(5836):350–4.

49. Reaz MBI, Hussain MS, Mohd-Yasin F. Techniques of EMG signal analysis: detection, processing, classification and applications. Biol Proced Online. diciembre de 2006;8(1):11–35.

50. Ristevski B, Chen M. Big Data Analytics in Medicine and Healthcare. J Integr Bioinforma [Internet]. el 25 de septiembre de 2018 [citado el 2 de agosto de 2019];15(3). Disponible en: http://www.degruyter.com/view/j/jib.2018.15.issue-3/jib-2017-0030/jib-2017-0030.xml

51. Salazar J, Espinoza C, Mindiola A, Bermudez V. Data Mining and Endocrine Diseases: A New Way to Classify? Arch Med Res. abril de 2018;49(3):213–5.

52. Gruber K. Is the future of medical diagnosis in computer algorithms? Lancet Digit Health. mayo de 2019;1(1):e15–6.

53. Miotto R, Li L, Kidd BA, Dudley JT. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. Sci Rep. el 17 de mayo de 2016;6:26094.

54. Xiong HY, Alipanahi B, Lee LJ, Bretschneider H, Merico D, Yuen RKC, et al. The human splicing code reveals new insights into the genetic determinants of disease. Science. el 9 de enero de 2015;347(6218):1254806.

55. Gawehn E, Hiss JA, Schneider G. Deep Learning in Drug Discovery. Mol Inform. el 1 de enero de 2016;35(1):3–14.

56. Cao C, Liu F, Tan H, Song D, Shu W, Li W, et al. Deep Learning and Its Applications in Biomedicine. Genomics Proteomics Bioinformatics. 2018; 16(1): 17–32.

57. Shen D, Wu G, Suk HI. Deep Learning in Medical Image Analysis. Annu Rev Biomed Eng. 2017;19:221-248.

58. Klann JG, Szolovits P. An intelligent listening framework for capturing encounter notes from a doctor-patient dialog. BMC Med Inform Decis Mak. 2009; 9(Suppl 1): S3.

59. Pang S, Du A, Orgun MA, Yu Z. A novel fused convolutional neural network for biomedical image classification. Med Biol Eng Comput. 2019;57(1):107-121.