¿Por qué las técnicas de aprendizaje profundo superan a las técnicas de aprendizaje automático tradicionales?

Why deep learning performs better than classical machine learning?

Artzai Picon[1, 2], Aitor Alvarez-Gila[1], Unai Irusta[2], Jone Echazarra[1]

[1] TECNALIA, Basque Research and Technology Alliance (BRTA). (España)
[2] Universidad del País Vasco (España)

DOI: http://dx.doi.org/10.6036/9574

En los últimos años, las técnicas de aprendizaje profundo (*deep learning*) han superado a los métodos clásicos basados en aprendizaje máquina (*machine learning*) tradicional a la hora de realizar tareas complejas de reconocimiento de patrones. En este artículo trataremos de explicar las razones.

Aprendizaje máquina frente a aprendizaje profundo

El paradigma tradicional de aprendizaje máquina o aprendizaje automático se basa en la extracción y selección de aquellas características de una imagen o señal de entrada que resulten relevantes para resolver la tarea encomendada (por ejemplo, su clasificación en una de las categorías predefinidas) para después aplicar un modelo de inteligencia artificial sobre las características extraídas. Las operaciones matemáticas que realizan esta extracción de características (*descriptores*) son normalmente diseñadas de forma manual por expertos con un buen conocimiento del dominio de aplicación. Los sistemas de visión artificial basados en aprendizaje máquina tradicional utilizaban descriptores visuales cuya función era transformar cada imagen de entrada en una representación compacta de la misma, siendo ésta descriptiva de la forma, textura o color (entre otros aspectos) de los diferentes elementos presentes en la imagen. Los bancos de filtros de Gabor o los Local Binary Patterns (LBP), por ejemplo, eran de uso común para extraer características de textura de la imagen; asimismo, características extraídas de un histograma de la imagen sobre modelos de color transformados se utilizaban para describir el color de la misma.

En nuestro trabajo inicial para la identificación de enfermedades en plantas en ambientes reales a partir de imágenes [1], extraíamos características de textura (LBP) y de color (media y varianza del modelo de color *Lab*) de *blobs* y regiones de la imagen sospechosas de presentar una enfermedad. Este algoritmo requería definir manualmente descriptores visuales que permitieran transformar la alta dimensionalidad de la imagen en un conjunto reducido de números que describieran cada región de la imagen que fuera susceptible de presentar enfermedad.

Tal y como imaginará el lector, añadir nuevas enfermedades sin que la precisión se resintiera comenzó a ser gradualmente más complicado y, de forma adicional, la capacidad de discriminación de los descriptores manualmente diseñados no era suficiente para discernir con éxito enfermedades más complicadas o sutiles. Esto se debe a que, al tratar de optimizar dicha propiedad para un descriptor, se tiende a sacrificar la robustez del mismo frente a variaciones en aquellos aspectos de la apariencia de la imagen que resultan irrelevantes o un obstáculo para realizar la clasificación con éxito (tales como la iluminación, fondos complejos, deformaciones de los objetos a identificar, etc.). Tras la inclusión de múltiples módulos para abordar algunos de esos retos, dicho algoritmo lograba una precisión de 80% (*Balanced Accuracy*) para un conjunto de tres enfermedades, a costa de una gran complejidad del proceso resultante. Sin embargo, tras su migración a un diseño de algoritmo basado en aprendizaje profundo [2] el diagrama de diseño del algoritmo se simplificó de forma muy significativa, y la precisión del mismo se incrementó hasta el 91%. En estos momentos alcanzamos una precisión del 99% gracias a otra de las ventajas inherentes al esquema de aprendizaje profundo: su carácter de marco de trabajo común y flexible que permite integrar y fusionar múltiples entradas de diferente tipo, tales como la información relativa a la especie de planta analizada, así como realizar varias predicciones de salida simultáneas [3].

Pero volvamos a la cuestión principal de este artículo: por qué las características extraídas mediante la aplicación de descriptores diseñados manualmente con el conocimiento del dominio, conjuntamente con potentes algoritmos de aprendizaje máquina como XGBoost [4], Support Vector Machines o Random Forests [5], no pueden competir con la capacidad de las técnicas de *deep learning* para aprender características con gran capacidad discriminativa? La respuesta la encontramos directamente en la descripción de aprendizaje profundo. De acuerdo con la definición dada en [6] por

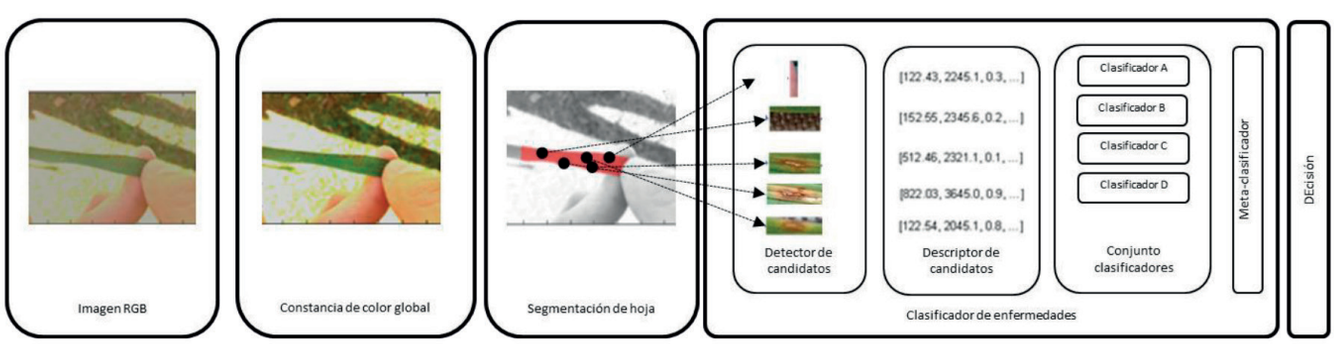


Figura 1: Algoritmo de visión artificial clásico basado en características diseñadas manualmente sobre las que se aplica un algoritmo de aprendizaje máquina

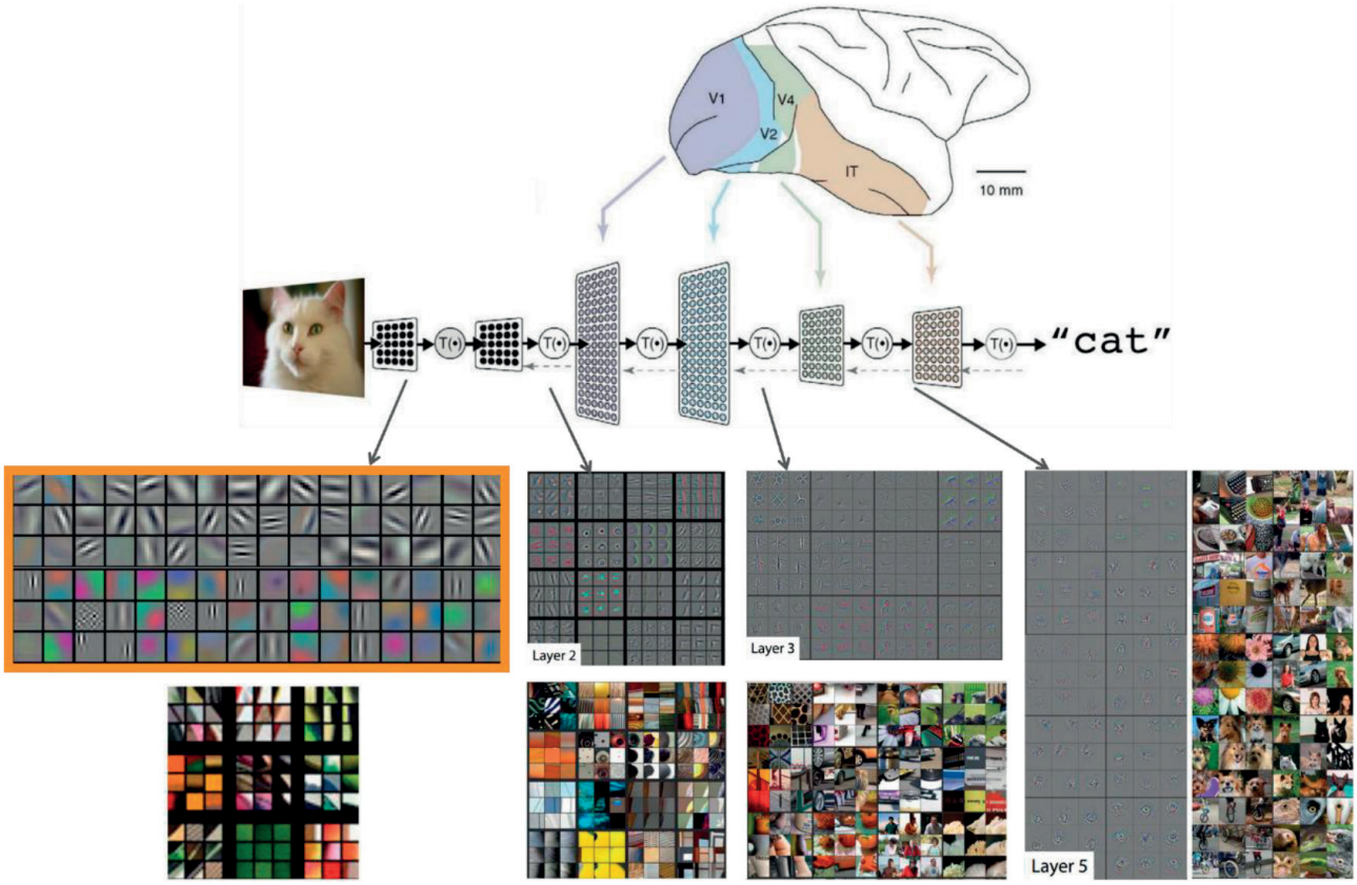


Figura 2: Filtros extraídos por una red convolucional (CNN-Convolutional Neural Network) cuando es entrenada sobre ImageNet. Se puede observar que las primeras capas aprenden filtros asociados a conceptos de bajo nivel como oponencias de color o detección de bordes, mientras que los filtros de más alto nivel se corresponden con conceptos más abstractos asociados a la tarea. Imágenes: [8], [9]

el reciente premio Turing Yoshua Bengio, "*Los métodos de aprendizaje profundo tratan de aprender jerarquías de características en las que las características de los niveles altos de la jerarquía se forman por composición de las características extraídas a niveles inferiores. El aprendizaje automático de características a múltiples niveles de abstracción permite a un sistema aprender funciones complejas que mapean la entrada con la salida esperada directamente a partir del dato en bruto sin depender del diseño manual de características por parte de un experto*". Así pues, los métodos basados en aprendizaje profundo son capaces de aprender a extraer las características más relevantes para la tarea objetivo, y resolver así el problema de forma óptima. Se trata, por tanto, de extractores de características jerárquicas óptimas, lo cual resulta especialmente relevante en el caso de las características de las capas más altas de la jerarquía (i.e. las de mayor nivel de abstracción), dado que los humanos no conocemos una buena forma de diseñarlas.

El repositorio de imágenes naturales Imagenet [7] se compone de más de un millón de imágenes pertenecientes a 1000 clases diferentes. La tarea de *clasificación de imágenes* consiste en identificar, para una imagen de entrada dada, a cuál de esas categorías pertenece dicha imagen. En la Figura 2 se muestran los distintos filtros (extractores de características) aprendidos por una red neuronal profunda (DNN-Deep Neural Network) en sus diferentes capas durante su entrenamiento para resolver esta tarea, sobre la base de datos Imagenet. Cabe destacar que los filtros extraídos en la parte más baja de la jerarquía funcionan bien como detectores de oposiciones de color o como extractores de bordes de diverso tipo, y se asemejan, por un lado, a la codificación de la información cromática en canales oponentes de color y, por otro, a la funcionalidad de detección de bordes en diferentes orientaciones y frecuencias propias de las capas más bajas del córtex visual humano. Además, estos filtros nos recuerdan a los bancos de filtros de Gabor para detección de texturas y al espacio de color *CIE L*a*b** que se han venido usando de forma habitual para el diseño manual de extractores de características visuales para alimentar modelos clásicos de aprendizaje máquina.

La capacidad de las capas convolucionales que se apilan en una DNN para extraer filtros que puedan describir de forma jerárquica una señal de entrada para cumplir con una determinada tarea puede ser aplicado a otros tipos de datos diferentes a la imagen. Recientemente hemos publicado un trabajo [10] donde analizamos los beneficios de aplicar técnicas de aprendizaje profundo en vez de las clásicas de aprendizaje máquina para analizar la señal de electrocardiograma (ECG) de pacientes en parada cardiorrespiratoria extrahospitalaria (PCREH) para identificar si es necesario desfibrilar usando un Desfibrilador Externo Automático (DEA).

Algoritmo para identificación de arritmias ventriculares letales

La PCREH es una de las principales causas de muerte en el mundo, con una incidencia anual media de 55 casos por cada 100.000 habitantes. Las arritmias ventriculares letales producen la parada cardiorrespiratoria, y la única terapia efectiva para revertirlas es la desfibrilación por descarga eléctrica. Los DEA son dispositivos sencillos diseñados para que personal sin entrenamiento previo pueda usarlos, e incorporan un algoritmo de decisión que determina si el sistema debe desfibrilar o no a partir del análisis de la señal ECG del paciente.

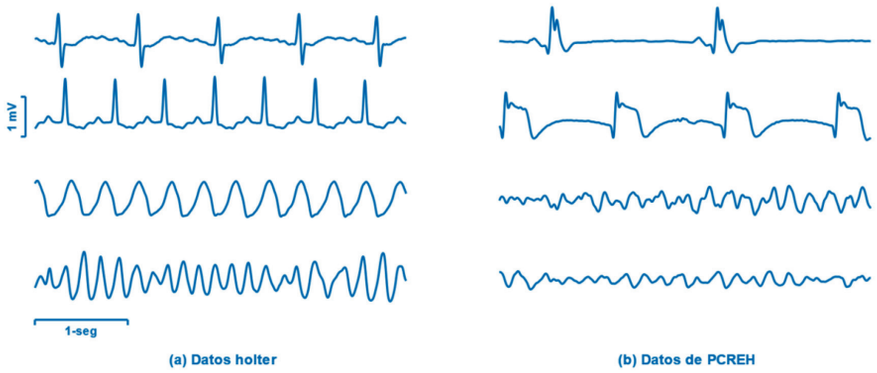


Figure 3: Ejemplos de ritmos ECG desfibrilables y no desfibrilables de un Holter (izquierda) y de pacientes en PCREH (derecha)

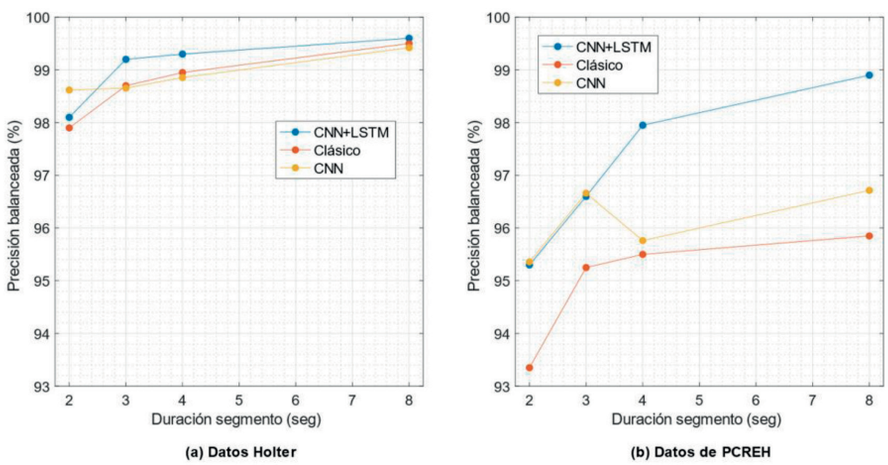


Figura 4: Resultados comparativos de precisión de algoritmos de aprendizaje máquina tradicional, red neuronal convolucional y red neuronal convolucional + capas recurrentes

Cuando el algoritmo detecta un ritmo desfibrilable el DEA proporciona una descarga eléctrica con el fin de restaurar el ritmo cardiaco normal. En caso contrario, el desfibrilador automático recomienda continuar con la reanimación cardiopulmonar (RCP). La American Heart Association (AHA) ha determinado una precisión mínima para los algoritmos de detección de ritmos desfibrilables, de forma que se garantice un uso seguro del dispositivo y que se aplique al paciente la terapia adecuada. Los ritmos no desfibrilables deben detectarse con especificidad (clase negativa) superior al 95%, para evitar descargas innecesarias, y los ritmos desfibrilables con una sensibilidad (clase positiva) superior al 90%.

En nuestro estudio, replicábamos un estudio anterior que analizaba la mejor combinación de características y descriptores de la señal de ECG que han sido diseñados durante las últimas décadas con el fin de identificar ritmos desfibrilables. Estos descriptores fueron diseñados manualmente para detectar rasgos distintivos de las arritmias ventriculares letales en la señal ECG, tales como irregularidad en la forma de onda, ausencia, estrechamiento o ampliación de los complejos QRS (las deflexiones en la forma de onda del ECG asociadas a la contracción ventricular), anchos de banda más pequeños o ritmos ventriculares elevados, entre otros.

Decidimos que, en vez de utilizar las características clásicas preexistentes, diseñaríamos una DNN compuesta, por un lado, por capas convolucionales que permitieran aprender aquellos filtros que eran óptimos para describir nuestra señal y, por otro, por capas recurrentes que permitieran el modelado de las dependencias temporales de la señal. Además, desarrollamos y probamos tanto métodos basados en deep learning como en aprendizaje máquina tradicional, para clasificar ECG grabados mediante Holter nada más producirse la parada, pero también para clasificar datos de PCREH grabados varios minutos después de producirse la parada (ver Figura 3).

La primera conclusión interesante fue que el método basado en aprendizaje profundo mejoraba la precisión de los métodos basados en aprendizaje máquina tradicional (ver Figura 4). Además, requería intervalos de señal más cortos para obtener dicha precisión. Esto se cumplía tanto para las muestras tomadas ya en el hospital como, de forma mucho más notable ya que la mejora de precisión era mucho mayor, para los casos de PCREH que constituyen el escenario más difícil para la clasificación.

Pero el beneficio del método basado en *deep learning* se reveló de forma más notoria cuando proyectamos las características diseñadas manualmente (seleccionadas como las mejores características para un abordaje del problema desde el aprendizaje máquina clásico) sobre un mapa bidimensional [11] e hicimos lo propio con las características de más alto nivel (*embedding*) aprendidas por la red DNN (Figura 5). Pudimos observar que las características extraídas por la DNN presentan una evidente mejora de la separabilidad con respecto de las diseñadas manualmente, facilitando así la labor de un clasificador simple actuando sobre éstas.

CONCLUSIÓN

El poder descriptivo de las arquitecturas de aprendizaje profundo para describir jerárquicamente datos y señales complejas en niveles elevados de abstracción las habilita para extraer características optimizadas para realizar la tarea encomendada de forma automática.

Estos algoritmos son capaces de extraer aquellas características que mejor resuelven la tarea sin intervención humana, lo que puede resultar de importancia crítica. Por ejemplo en mayo de 2019, científicos del MIT publicaron un artículo [12] sobre detección de cáncer de mama, donde utilizaron imágenes que habían sido analizadas por radiólogos. Sin embargo, para entrenar la red neuronal no utilizaron como anotación el diagnóstico de los radiólogos, sino el estado de salud del paciente cinco años después. Los modelos generados extrajeron patrones y una descripción de alto nivel de la imagen que permite diagnosticar el riesgo de cáncer cinco años antes de que el diagnóstico pudiera ser realizado. Esto habría resultado imposible mediante el uso de descriptores de características prediseñados por humanos expertos que están basados en conocimiento existente. En un escenario totalmente diferente, los sistemas basados en aprendizaje por refuerzo (basado en aprendizaje profundo) que superan a jugadores humanos en el ajedrez, go o Starcraft aprenden estrategias totalmente desconocidas e inesperadas para los humanos.

Las técnicas de aprendizaje profundo

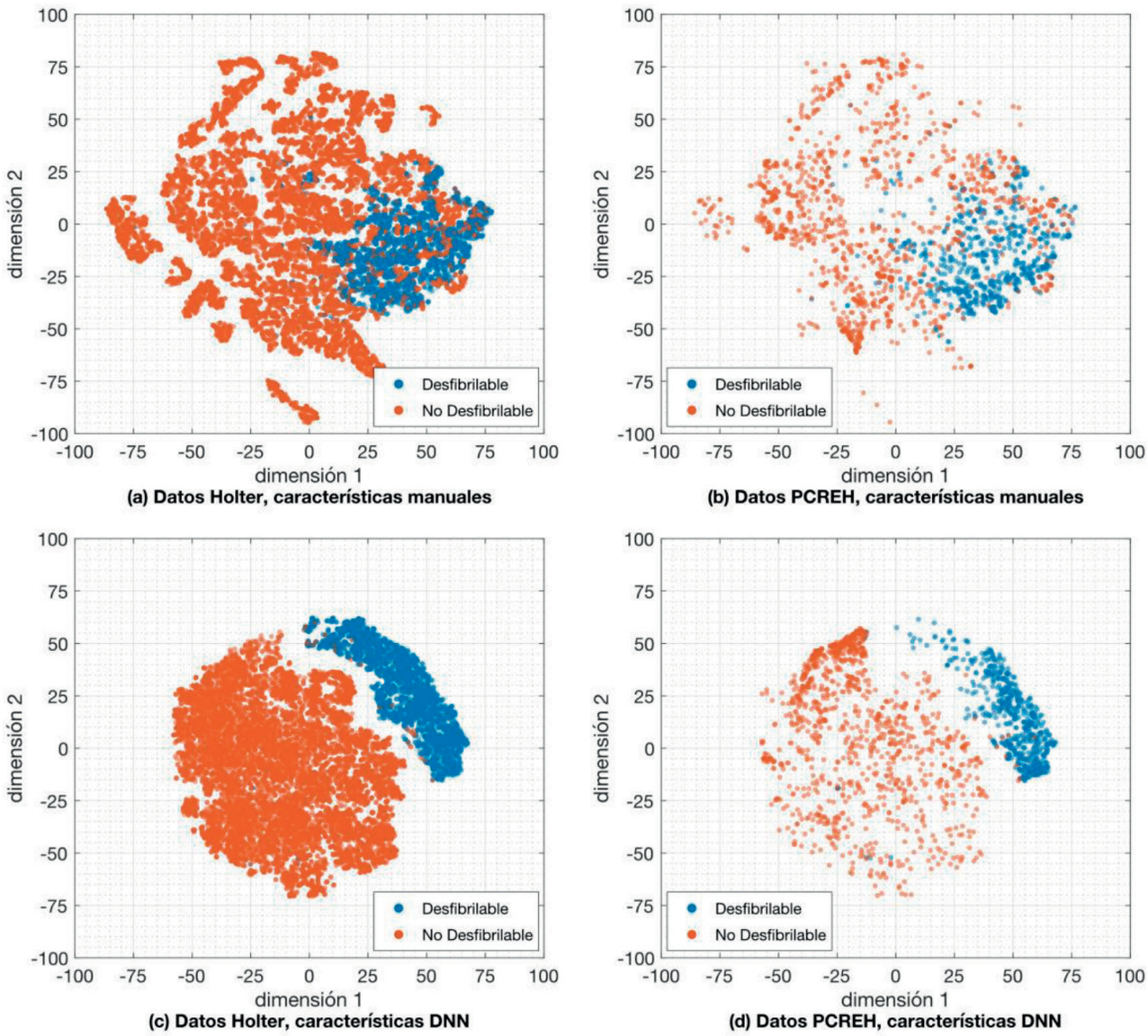


Figure 5: Análisis de la separabilidad de las características extraídas manualmente y de las extraídas por nuestra red neuronal [10]

permiten extraer de forma automática la mejor representación interna de los datos para la tarea objetivo, permitiendo incluso aflorar información actualmente desconocida o ignorada por el ser humano con gran flexibilidad y eficacia. Todo ello da lugar a un marco de trabajo enormemente versátil y aplicable de forma masiva y transversal en dominios más allá de la imagen agro o señales médicas, tales como la industria productiva y el control de calidad.

REFERENCIAS

[1] A. Johannes, A. Picon, A. Alvarez-Gila, J. Echazarra, S. Rodriguez-Vaamonde, A. D. Navajas, and A. Ortiz-Barredo, "Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case," Comput. Electron. Agric., vol. 138, pp. 200–209, Jun. 2017.

[2] A. Picon, A. Alvarez-Gila, M. Seitz, A. Ortiz-Barredo, J. Echazarra, and A. Johannes, "Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild," Comput. Electron. Agric., vol. 161, pp. 280–290, Jun. 2019.

[3] A. Picon, A. O. Barredo, A. Alvarez-Gila, J. Echazarra, M. Seitz, and P. Mohnke, "Crop conditional Convolutional Neural Networks for massive multi-crop plant disease classification over cell phone acquired images taken on real field condition," Comput. Electron. Agric., Nov. 2019.

[4] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 2016, pp. 785–794.

[5] L. Breiman, "Random Forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, Oct. 2001.

[6] Y. Bengio, "Learning Deep Architectures for AI," Found Trends Mach Learn, vol. 2, no. 1, pp. 1–127, Jan. 2009.

[7] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," Int. J. Comput. Vis. IJCV, vol. 115, no. 3, pp. 211–252, 2015.

[8] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," in Computer Vision – ECCV 2014, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Springer International Publishing, 2014, pp. 818–833.

[9] L. Masuch, "Deep learning. A visual introduction," 06-Dec-2015.

[10] A. Picon, Unai Irusta, A. Alvarez-Gila, E. Aramendi, F. Alonso-Atienza, C. Figuera, U. Ayala, E. Garrote, L. Wik, J. Kramer-Johansen, and T. Estestol, "Mixed Convolutional and Long Short-Term Memory Network for the Detection of Lethal Ventricular Arrhythmia," PLOS ONE 2019;14(5):e0216756.

[11] L. V. der Maaten and G. Hinton, "Visualizing data using t-SNE," J. Mach. Learn. Res., vol. 9, no. 2579–2605, p. 85, 2008.

[12] A. Yala, C. Lehman, T. Schuster, T. Portnoi, and R. Barzilay, "A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction," Radiology, vol. 292, no. 1, pp. 60–66, May 2019.