

Nuevos retos basados en machine learning para la mejora del procesamiento y la exploración de datos biomédicos

Autor: Sanchez-Cardenas, Ivon

UNIVERSITAT INTERNACIONAL DE CATALUNYA (UIC)

Resumen

En la actualidad, la investigación biomédica necesita que los procesos biológicos y enfermedades que deben ser investigados se basen en modelos sanitarios eficientes. Para ello, se ofrecen diversas soluciones biomédicas digitales claves para impulsar el ecosistema de salud del futuro, que usan procesos computacionales, entre los cuales se encuentra el aprendizaje automático o *machine learning*.

Introducción

Actualmente, todos los datos clínicos están pasando a convertirse en información digital, cuyo volumen resulta muy difícil de analizar e interpretar por parte de los investigadores biomédicos. Además, si nos preguntamos: *¿Qué genes son los más relevantes para cierto tipo de cáncer?* o *¿Qué medicamentos serían los más adecuados para este tipo de tratamiento?*, el uso de procesos computacionales complejos basados en machine learning (ML) serán los que nos ayuden a resolver estas preguntas. El ML es una rama de la inteligencia artificial en la que las máquinas se programan para aprender patrones a partir de los datos y tomar decisiones con una mínima intervención humana.

Este artículo ofrece una visión de cómo el ML puede transformar amplias áreas de la biomedicina, concretando su estudio en uno de los modelos más usados: las redes neuronales.

Las redes neuronales son una familia de algoritmos muy potentes con los que se pueden modelar comportamientos inteligentes y estructuras avanzadas además de codificar cualquier tipo de dato (figura 1).

Es aquí donde el ML aporta su capacidad de interpretar un gran volumen de datos potenciando la metodología científica, mejorando el uso de protocolos y ayudando a la toma de decisiones por parte del personal especialista (figura 2), además de potenciar la medicina personalizada y preventiva.

Grado de Biomedicina 2do curso

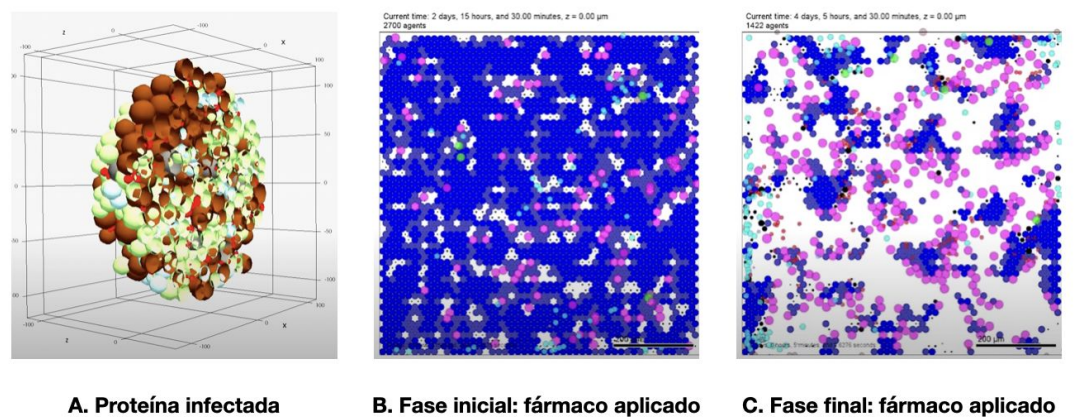


Figura 2. Interacción de un fármaco aplicado a una proteína infectada por un virus. Mediante el uso de la simulación a través de procesos computacionales, se puede visualizar cómo interacciona un fármaco (B,C) que se requiere como diana terapéutica para cierta proteína (A) que ha sido infectada por un virus. Para ello, si se quieren extraer características de los datos del fármaco, se usa el ML para así poder interpretar los resultados y mejorar la toma de decisiones por parte del especialista.

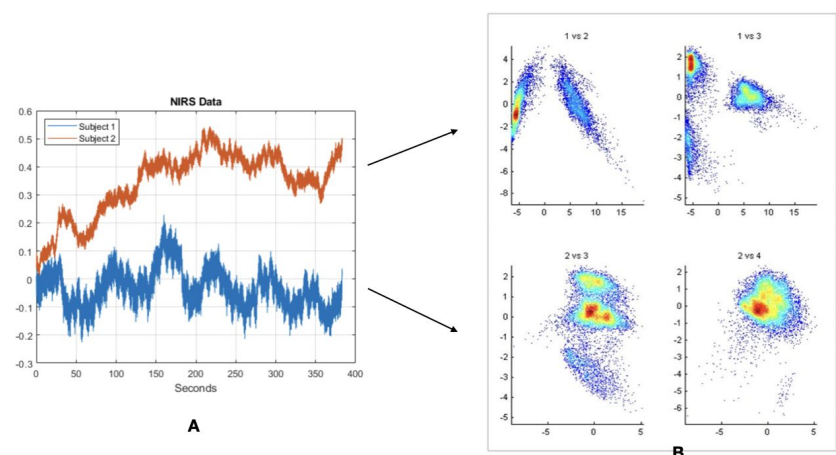


Figura 3. Visualización de datos biomédicos: métodos convencionales vs métodos basados en ML. En A se observa la gráfica de dos señales eléctricas cerebrales correspondientes a dos sujetos investigados por especialistas biomédicos. Estas señales son complejas de analizar ya que contienen información clínica oculta en la gráfica. Por el contrario, en B se pueden estudiar las imágenes con mayor claridad, ya que se han aplicado métodos de ML para agrupar y clasificar los patrones detectados en esas señales, aportando una mayor información clínica.

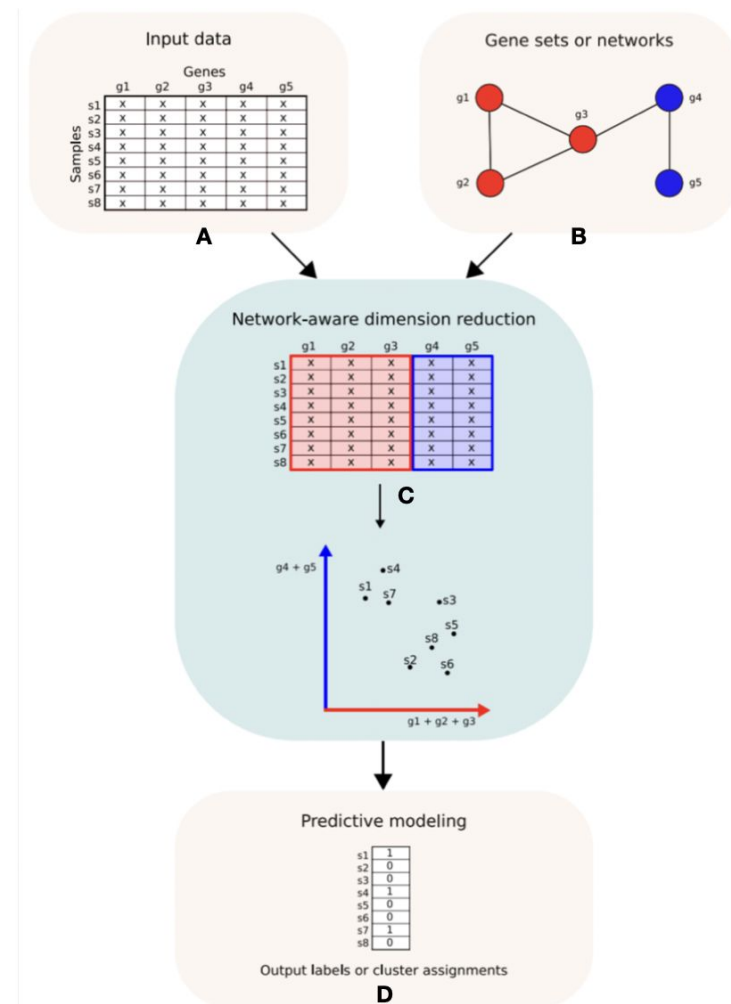


Figura 1. Clasificación de genes a través del ML basado en redes neuronales. Esquema donde se muestra el uso de un algoritmo de machine learning (ML) basado en redes neuronales en el cual se clasifican ciertos tipos de genes, siguiendo unos criterios establecidos por los especialistas. En general, el ML se basa en la idea de que los sistemas pueden aprender de los datos de entrada, identificar patrones y tomar decisiones con una mínima intervención humana. En el esquema, A muestra los datos de entrada que consisten en secuencias genéticas. B muestra el grado de similitud entre estos genes los cuales se agrupan por colores (rojo y azul). En C, los datos de entrada con mayor similitud ya agrupados se representan en 2D mediante procesos computacionales. Por último, en D se ofrece el modelo predictivo, que etiqueta los genes en 0 o 1 según la asignación de cada gen a su grupo más similar.

Conclusión

Podemos concluir que la integración del ML en el diagnóstico médico es muy prometedora para lograr la máxima precisión en el diagnóstico y facilitarlo por completo. No obstante, la recopilación de masivos conjuntos de datos de naturaleza heterogénea y compleja, para el entrenamiento de los algoritmos de ML sigue siendo el principal reto en la actualidad.

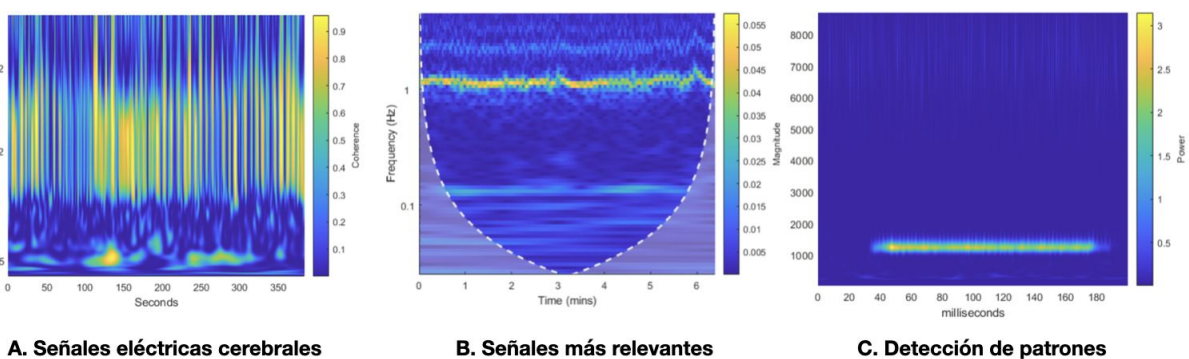


Figura 4. Uso de ML para la visualización personalizada en señales eléctricas cerebrales. Existen varios métodos de ML para la selección de características relevantes en señales eléctricas cerebrales comúnmente llamadas EEG: El método visualizado en A extrae información temporal de la señal. El método visualizado en B aprovecha tanto la información temporal como la frecuencial para detectar las señales más relevantes. Por último, en C se realiza un filtrado espacial para luego extraer las características que conllevarán a la detección de patrones en los datos.

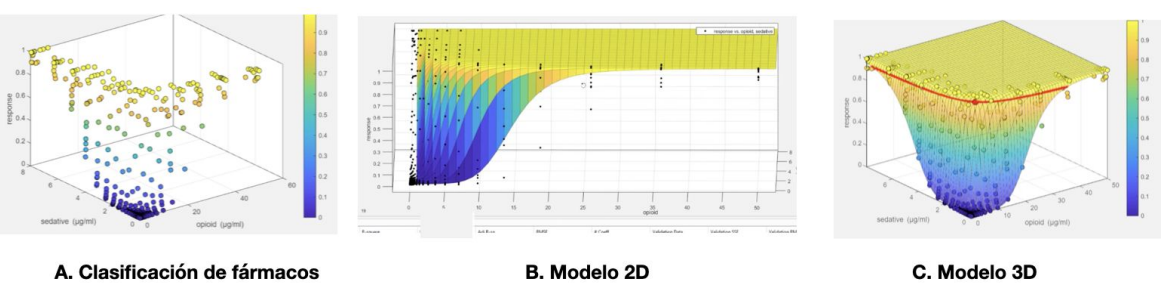


Figura 5. Visualización de fármacos: cuál es el más adecuado? Cual es la diana terapéutica de cada uno? El uso del diseño de fármacos basado en el ML contribuye a la mejora en su precisión y eficiencia ya que en el modelo de entrenamiento de las redes neuronales utilizadas, se tienen también en cuenta propiedades farmacológicas como la absorción, distribución, metabolismo, excreción y la toxicidad. En A se muestran 3 tipos de fármacos agrupados según el grado de respuesta ante un virus. B y C muestran los modelos 2D y 3D respectivamente, de los fármacos para una mejor visualización de los datos.

Discusión

En este estudio se muestran diferentes herramientas basadas en ML para mejorar los modelos sanitarios usados en el estudio de datos biomédicos, siendo este un enfoque innovador e inteligente que ayuda a resolver problemas específicos y aumenta el rendimiento de la investigación biomédica. Además, la detección e identificación de patrones significativos a partir de la visualización de grandes conjuntos de datos heterogéneos y complejos (figuras 3, 4 y 5) contribuye al desarrollo de estrategias de medicina de precisión, así como la aceleración del estudio de ciertas patologías de gran impacto mundial.

Bibliografía

- [1] Mallat, S. "A Wavelet Tour of Signal Processing: The Sparse Way", Academic Press, 2009.
- [2] Jeremy Goecks, Vahid Jalili, Laura M. Heiser, and Joe W. Gray. "How Machine Learning Will Transform Biomedicine", Cell, 181, April 2, (2020) 101.
- [3] Yinghao Li, Qiumei Pu, Shuheng Li, Hong Zhang, Xiaofeng Wang, Haodong Yao, Lina Zhao. "Machine learning methods for research highlight prediction in biomedical effects of nanomaterial application", Pattern Recognition Letters, 117 (2019) 111-118.
- [4] Shuihua Wang, M. Emre Celebi, Yu-Dong Zhang, Xiang Yu, Siyuan Lu, Xujing Yao, Qinghua Zhou, Martínez-García Miguel, Yingli Tian, Juan M Gorri, Ivan Tyukin. "Advances in Data Preprocessing for Biomedical Data Fusion: An Overview of the Methods, Challenges, and Prospects.", Information Fusion 76 (2021) 376-421.
- [5] Software de visualización figuras 2-5: MATLAB and Statistics Toolbox Release 2012b, The MathWorks, Inc., Natick, Massachusetts, United States.

