

INTELIGENCIA ARTIFICIAL EN ESTUDIOS DE GENOTOXICIDAD Y ONCOEPIDEMIOLOGÍA

Al enfoque tradicional epidemiológico, se ha sumado el de la inteligencia artificial que dispone de las herramientas para el análisis de una gran cantidad de datos y que puede determinar cuáles son las variables que más se asocian al estado de paciente con enfermedad neoplásica



Por Elio A. Prieto González

Los procesos considerados complejos son el resultado de muchos factores que pueden además variar de intensidad en el tiempo y en el espacio. Las interacciones que se producen entre seres vivos y el entorno en el que se encuentran efectores físicos, químicos y biológicos generan efectos que pueden ser antagónicos entre sí o ser por el contrario sinérgicos, de sumación o multiplicación, tales interacciones son complejas y en ellas, el número de variables involucradas es muy grande (1). Muy grande significa que pueden ser descritas con números de varios ceros, y que las condiciones intermedias entre los efectores y el o los resultados varían con velocidades e intensidades diversas y a su vez inestables.

El genotipo es el conjunto de genes de un individuo, es la expresión funcional del genoma. El genotipo se concibe de manera reduccionista al referirlo a una característica del sujeto, de ahí que podemos hablar del genotipo asociado a la celiaquía, al cáncer hereditario no polipósico o a la susceptibilidad a la obesidad. Por otra parte, el fenotipo (lo externo o *phaenomenon*) es el resultado de las interacciones entre el genotipo y el ambiente (2).

Es una definición clásica de genética, aparentemente simple: Genes + Ambiente = Fenotipo. Sin embargo, eso no describe lo que realmente ocurre: el fenotipo es el resultado de las interacciones entre los efectos de los genes en un sistema biológico y los efectos de los determinantes ambientales entre sí y con el sistema biológico. Reducido, por supuesto a un conjunto limitado de genes que en

la realidad esta interconectados con otros y así sus efectos son el resultado de las interacciones entre ellos (3). Un ejemplo son las variantes clínicas en la Celiaquía que pueden ser la consecuencia de genes modificadores no HLA, aunque cuando se habla de genotipo asociado a la Celiaquía, la mayoría de las veces solamente se consideran los genes HLA DQ2 y DQ8 (4). El cáncer puede ser considerado como un fenotipo cuyas principales características son las que expresan las interacciones entre las mutaciones en diversos genes supresores y oncogenes con el ambiente (5). Es un fenotipo en el que se observa la proliferación descontrolada, la indiferenciación, los cambios en la expresión de receptores hormonales, la angiogénesis y la capacidad de metastizar entre otras (6).

El cáncer (o los cánceres), en sus diferentes localizaciones, sus variables histológicas y su respuesta a la terapia onco específica, es la expresión de fenotipos. La incidencia y prevalencia de los cánceres, guarda relación con agentes ambientales y cambios del entorno metabólico individual, tales como los relacionados con la obesidad (7).

En tal sentido la cooperación de los carcinógenos ambientales y su efecto sobre individuos de una población heterogénea en cuanto a sus geno-

El autor es profesor en la Licenciatura de Nutrición ISALUD. Médico Especialista en Genética Clínica. Graduado en la Universidad Médica de la Habana y el Instituto Nacional de Oncología y Radiobiología de Cuba. Investigador en Toxicología Genética, Citogenética y Biología Celular. Aplicación de Inteligencia Artificial a la investigación médica. Director de proyectos en el Centro de Altos Estudios en Ciencias Humanas y de la Salud (CAECIHS), relacionados con el papel de la obesidad y los hábitos nutricionales con las lesiones en el ADN y el cáncer.

tipos, convierten la tarea de definir aquellas exposiciones influyentes sobre la incidencia de cánceres específicos, en algo muy difícil (8).

La prevención del cáncer requiere de la elucidación de cuáles son las interacciones que determinan el aumento de la incidencia de los diferentes cánceres en poblaciones específicas.

Los estudios en los que se definen las redes de control bioquímico, las variaciones en la expresión genética y sus interacciones son necesariamente reduccionistas, puesto que es imposible investigar todos los genes candidatos, con todos los carcinógenos sospechados, sin dejar de lado, aquellos alimentos o conductas que funcionan como anti carcinógenos, tal es el caso de las frutas y el ejercicio físico.

Inteligencia artificial

Enfrentar el problema de las interacciones entre los conjuntos de variables que caracterizan a los individuos (hereditarias, daño genético, metabólicas, ...) y las variables ambientales (nutricionales, ocupacionales, hábitat, hábitos tóxicos...) es una empresa que reviste mucha complejidad y que debe ser abordada con un enfoque dual. Este debe continuar el estudio en profundidad de cada uno de los aspectos biológicos que están incluidos en cada uno de los aspectos mencionados. Por ejemplo, hereditarios: genes de respuesta al estrés oxidativo, genes de enzimas del complejo de citocromo p450, genes de reparación del ADN, entre otros. Daño genético: significado para el proceso neoplásico de las lesiones de simple y doble cadena del ADN, mutaciones inducidas en genes de reparación, inestabilidad cromosómica por nombrar algunos. Mientras que en los estudios ambientales pueden mencionarse como

ejemplos factores nutricionales, el consumo de ácidos grasos omega 3 y 6, el consumo de antioxidantes, de fibras, de carne roja. (9,10).

Los genes se pueden secuenciar, determinar sus variantes asociadas al desarrollo neoplásico, rastrear su presencia en las poblaciones, definir su capacidad de producir cáncer al mutar, saber si es necesaria una o más mutaciones y que proteínas alteradas aparecen cuando se producen los cambios genéticos. En el ambiente pueden detectarse aquellos agentes capaces de producir aberraciones cromosómicas y mutaciones. También es posible evaluar a la obesidad como generadora de un entorno pro carcinogénico, los efectos de la inflamación crónica asociada a esa enfermedad, las alteraciones vinculadas a las adipocitoquinas, el estrés oxidativo, las proteínas de estrés (11). Esos participantes: genes, ambiente, entorno metabólico, sistema inmune son el terreno donde se desarrollan los cánceres y son investigados en conjunto para identificar aquellos factores ambientales e individuos que al vincularse pueden llevar al cáncer (12,13).

La complejidad de las interacciones genes ambiente, sus dimensiones espacio temporales orientan a pensar en la necesidad de hacer estudios donde se evalúen muchas variables para definir cuáles son las combinaciones que puedan ser las responsables mayores de la presencia de cáncer en las poblaciones en estudio. Este enfoque combinatorio es posible gracias a la inteligencia artificial (AI) (14).

La inteligencia artificial es el término utilizado para describir el desarrollo de procesos similares a los de la inteligencia humana mediante herramientas informáticas. Los procesos considerados son el aprendizaje (la adquisición de información y de las reglas para su utilización), el razonamiento (mediante el empleo de reglas o algoritmos que conducen a definiciones/conclusiones parciales o definitivas) y además la capacidad de autocorrección (14).

El término IA fue propuesto ya en 1956 por John McCarthy, un especialista en informática norteamericano (15). Actualmente es un concepto que incluye la automatización robótica, el desarrollo de sistemas con habilidades, como reconocer patrones visuales, jugar al ajedrez, realizar diagnósticos médicos, desarrollar vehículos autónomos que funcionan en entornos reales, procesamiento del lenguaje natural y el análisis de gran número de datos, aplicación que ha crecido mucho a la par de los sistemas de alta capacidad de producción de tales datos, que permiten encontrar patrones relacionales entre ellos con mucha mayor rapidez que los cerebros humanos y que en Biología avanzada, Medicina y Nutrición, pueden resumirse en el término *omics*. Este sufijo (ómica en español) abarca a la Genómica, la Proteómica, Metabolómica y Nutrigenómica como ejemplo de procedimientos que profundizan hasta el nivel molecular en la caracterización de los genes, proteínas, metabolitos o nutrientes en individuos o poblaciones (16,17).

El cáncer, en sus diferentes localizaciones, sus variables histológicas y su respuesta a la terapia oncoespecífica es la expresión de fenotipos. La incidencia y prevalencia de los cánceres, guarda relación con agentes ambientales y cambios del entorno metabólico individual, tales como los relacionados con la obesidad

Los estudios epidemiológicos se basan en el análisis de los efectos de muchos genes y factores ambientales en la aparición del cáncer (18). Al enfoque tradicional epidemiológico, se ha sumado el de la Inteligencia Artificial que dispone de las herramientas para el análisis de una gran cantidad de datos y que puede determinar cuáles son las variables que más se asocian al estado de paciente con enfermedad neoplásica (19).

Minería de datos

La identificación de relaciones entre factores dentro de un conjunto de información, que en conjunto se denomina Minería de Datos (MD) (*Data Mining*), puede aplicarse al estudio de las interacciones de elementos del ambiente con el ADN y al de los factores asociados al desarrollo del cáncer.

Los términos Minería de Datos y Descubrimiento del Conocimiento (KDD según sus siglas en inglés) a menudo se los trata como sinónimos. En general está aceptado que MD se refiera a un paso

particular en el proceso que consiste en la aplicación de algoritmos específicos para extraer patrones o modelos de los datos. Hay otros pasos en el proceso del conocimiento KDD además de la MD: la preparación de los datos, la selección y limpieza de estos, la incorporación de conocimiento previo, y la propia interpretación de los resultados de minería. Estos pasos aplicados de una manera iterativa e interactiva aseguran que se les extraiga conocimiento útil a los datos. Hoy es de suma importancia de incluir en la metodología el pre-proceso de los

datos, y la formalización del conocimiento descubierto. La Inteligencia Artificial es una disciplina que engloba a ambos conceptos (20).

Un ejemplo aplicado en un estudio realizado por nosotros es el de la Teoría de la Información de Shannon (21). El estudio tenía como objetivo evaluar el *Efecto de la obesidad sobre los niveles de daño en el ADN determinado por elec-*

troforesis alcalina de células aisladas en gel de agarosa o Ensayo Cometa. La muestra estaba constituida por un grupo de 54 mujeres residentes en la periferia de la Ciudad de la Plata. Al incluir las variables ambientales, relativas al hábitat, los patrones de consumo de alimentos, los hábitos tóxicos, los niveles de glucosa y lipoproteínas entre otras se alcanzó el número de 159 variables. El análisis de todas las combinaciones posibles, designado con la letra Q era igual a $8,57 \times 10^9$. El tiempo estimado para realizar todos los cálculos combinatorios podía llevar 1,33 años utilizando una computadora de última generación, algo que no era conveniente en

el contexto del estudio.

El empleo de algoritmos de Inteligencia Artificial era lo que permitiría explicar cual era la combinación de variables que mejor explicaba el nivel de daño en el ADN. El estudio mediante los métodos de IA podía hacerse en menos tiempo y sin perder el análisis de la complejidad de las interacciones. Es por lo que se aplicó la Teoría de la Información de Shannon (21) que hace posible el cálculo del aporte de las variables ambientales, metabólicas, antropométricas, nutricionales y de hábitos tóxicos. La aplicación de

los métodos de IA condujo al desarrollo de dos modelos de Regresión Lineal Múltiple explicativos que fueron analizados y discutidos por los investigadores médicos a fin de corroborar la pertinencia de las asociaciones descritas en los polinomios resultantes: el modelo más sencillo y el modelo mejor, que involucraron a 6 y 14 variables respectivamente. La *r* calculada en el mejor modelo fue de 55.72%. El tiempo necesario para hacer los cálculos fue de 2 horas y 4 minutos (Kamlofsky J. Comunicación Personal).

El estudio de la relación de las características ambientales, antropométricas, nutricionales y tóxicas con el nivel de daño al ADN en una muestra pequeña, pero con un número de variables elevado, permite ilustrar la potencia de uno de los procedimientos de IA. El daño en el ADN detectado por el *Ensayo Cometa* es un biomarcador de efecto que cada vez gana más relevancia en la evaluación y manejo de afecciones producidas por alteraciones genéticas y que en el caso del estudio mencionado, puede contribuir a comprender mejor las relaciones entre IMC y roturas de simple cadena en el ADN. Los resultados indican que los niveles de glucemia, la circunferencia abdominal, el porcentaje de grasa visceral, la dieta rica en vegetales o por el contrario en carnes, la residencia en áreas con mayor contaminación del aire, son factores que intervienen en la relación entre el IMC y el índice de daño al ADN.

La aplicación de la IA hizo posible encontrar una correlación positiva, aunque no significativa entre las variables. Una relación que expresa mejor la realidad que la que se obtiene haciendo evaluaciones que no incluyen todas las combinaciones posibles de variables.

El estudio siguiente fue diseñado para poder aplicar la IA. El cáncer es considerado como un fenotipo

Es necesario mencionar que la posibilidad de encontrar asociaciones novedosas que deban ser explicadas, lo que implica que los procedimientos de IA en nuestro campo tienen un carácter heurístico que justifica el apelativo de inteligencia a estas funciones de los sistemas informáticos.



**Cada desafío es
La oportunidad
de crear
un mundo mejor**


relacionado a un conjunto de interacciones.

El estudio de corte transversal, tipo encuesta se ha venido aplicando en una muestra de la población que incluye a 1600 personas sanas al momento de la entrevista y 300 pacientes con diagnóstico positivo de cáncer. Se cumplió con los requerimientos de Bioética, habiendo sido aprobado por 2 comités de Ética. Se tomaron en conjunto 52 preguntas cuyas respuestas posibles están entre 6 y 20. Se elaboró una base de datos que permite exportar fácilmente los resultados para el procesamiento de la información.

En el estudio de Hábitos y Enferme-

dad Neoplásica se espera encontrar una mayor potencia en las asociaciones con una muestra que es grande en comparación con estudios como el arriba comentado sobre la relación entre la obesidad y el daño al ADN, pero pequeña en comparación con los estudios epidemiológicos clásicos. La posibilidad de encontrar asociaciones biológicamente sustentables después de combinar todas las variables que no sean excluidas por la IA en un primer paso, abre el camino a una mejor comprensión de cuáles son las variables, cual la magnitud de éstas, que puedan asociarse a la presencia de cáncer en la población de una región particular del país.

Es necesario mencionar que la posibilidad de encontrar asociaciones novedosas que deban ser explicadas, lo que implica que los procedimientos de IA en nuestro campo tienen un carácter heurístico que justifica el apelativo de inteligencia a estas funciones de los sistemas informáticos.

Es un primer paso para el desarrollo de estudios multicéntricos en los que se obtengan datos aún más precisos y en una muestra aun mayor, condiciones que mejoren la capacidad de generar descripciones matemáticas del aporte de los antecedentes heredo familiares, hábitos y conductas de la población al desarrollo de la enfermedad neoplásica. 

Referencias

1. Ladyman J., Lambert J., Wiesner K. (2013). What is a complex system? *Eur. J. Philos. Sci.*, 3 (1):33-67
2. Jones R, Pembrey M, Golding J, Herrick D. (2005). The search for genotype/phenotype associations and the phenome scan. *Paediatr Perinat Epidemiol.* 19(4):264-75. Review. PubMed PMID: 15958149.
3. Lehner B. (2007). Modelling genotype-phenotype relationships and human disease with genetic interaction networks. *J Exp Biol*;210(Pt 9):1559-66. Review. PubMed PMID: 17449820.
4. Coleman C, Quinn EM, Ryan AW, Conroy J, Trimble V, Mahmud N, Kennedy N, Corvin AP, Morris DW, Donohoe G, O'Morain C, MacMathuna P, Byrnes V, Kiat C, Trynka G, Wijmenga C, Kelleher D, Ennis S, Anney RJ, McManus R. (2016). Common polygenic variation in coeliac disease and confirmation of ZNF335 and NIFA as disease susceptibility loci. *Eur J Hum Genet.* 24(2):291-7. doi: 10.1038/ejhg.2015.87. Epub 2015 Apr 29. PubMed PMID: 25920553; PubMed Central PMCID: PMC4717209.
5. Dingli D, Chalub FA, Santos FC, Van Segbroeck S, Pacheco JM. (2009). Cancer phenotype as the outcome of an evolutionary game between normal and malignant cells. *Br J Cancer*;101(7):1130-6. doi: 10.1038/sj.bjc.6605288. Epub 2009 Sep 1. PubMed PMID: 19724279; PubMed Central PMCID: PMC2768082.
6. Gupta SC, Kim JH, Prasad S, Aggarwal BB. (2010). Regulation of survival, proliferation, invasion, angiogenesis, and metastasis of tumor cells through modulation of inflammatory pathways by nutraceuticals. *Cancer Metastasis Rev.*;29(3):405-34. doi: 10.1007/s10555-010-9235-2. Review. PubMed PMID: 20737283
7. Hu Z, Brooks SA, Dormoy V, Hsu CW, Hsu HY, Lin LT, Massfelder T, Rathmell WK, Xia M, Al-Mulla F, Al-Temaimi R, Amedei A, Brown DG, Prudhomme KR, Colacci A, Hamid RA, Mondello C, Raju J, Ryan EP, Woodrick J, Scovassi AI, Singh N, Vaccari M, Roy R, Forte S, Memeo L, Salem HK, Lowe L, Jensen L, Bisson WH, Kleinstreuer N. (2015). Assessing the carcinogenic potential of low-dose exposures to chemical mixtures in the environment: focus on the cancer hallmark of tumor angiogenesis. *Carcinogenesis.* 36 Suppl 1:S184-202. doi: 10.1093/carcin/bgv036. Review. PubMed PMID: 26106137; PubMed Central PMCID: PMC4492067.
8. Domingo JL, Nadal M. (2017). Carcinogenicity of consumption of red meat and processed meat: A review of scientific news since the IARC decision. *Food Chem Toxicol.* 105:256-261. doi: 10.1016/j.fct.2017.04.028. Epub 2017 Apr 24. Review. PubMed PMID: 28450127.
9. Davies AE, Albeck JG. (2018) Microenvironmental Signals and Biochemical Information Processing: Cooperative Determinants of Intratumoral Plasticity and Heterogeneity. *Front Cell Dev Biol.* 20:6:44. doi: 10.3389/fcell.2018.00044. eCollection 2018. Review. PubMed PMID: 29732370; PubMed Central PMCID: PMC5921997.
10. Eastmond DA. (2017). Recommendations for the evaluation of complex genetic toxicity data sets when assessing carcinogenic risks to humans. *Environ Mol Mutagen.* ;58(5):380-385. doi: 10.1002/em.22078. Epub 2017 Mar 7. PubMed PMID: 28266084.
11. Brinke A, Buchinger S. (2017). Toxicogenomics in Environmental Science. *Adv Biochem Eng Biotechnol.*;157:159-186. doi: 10.1007/10_2016_15.
12. Umbuzeiro GA, Heringa M, Zeiger E. (2017). In Vitro Genotoxicity Testing: Significance and Use in Environmental Monitoring. *Adv Biochem Eng Biotechnol.*;157:59-80. doi: 10.1007/10_2015_5018
13. Afanasieva K, Sivolob A. (2018) Physical principles and new applications of comet assay. *Biophys Chem.*;238:1-7. doi: 10.1016/j.bpc.2018.04.003. Epub 2018 Apr 20.
14. McCorduck, P. (2004). *Machines Who Think. A personal inquiry into the history and prospects of artificial intelligence.* New York: A K Peters / CRC Press
15. The History of Artificial Intelligence History of Computing CSEP 590A University of Washington December 2006. <https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf>
16. Fenech, M., El-Sohehy, A., Cahill, L., Ferguson, L. R., French, T.-A. C., Tai, E. S., ... Head, R. (2011). Nutrigenetics and Nutrigenomics: Viewpoints on the Current Status and Applications in Nutrition Research and Practice. *Journal of Nutrigenetics and Nutrigenomics*, 4(2), 69–89. <http://doi.org/10.1159/000327772>
17. Clish CB. (2015). Metabolomics: an emerging but powerful tool for precision medicine. *Cold Spring Harbor Molecular Case Studies*, 1(1), a000588. <http://doi.org/10.1101/mcs.a000588>
18. Graggruber P, Hrazdira E, Sebera M, Kalina T. (2018). Cancer Incidence in Europe: An Ecological Analysis of Nutritional and Other Environmental Factors. *Front. Oncol*, 13 June 2018 | <https://doi.org/10.3389/fonc.2018.00151>
19. Liu Z, Chen D, Chen X, Jia H. (2009). Computational Data Mining in Cancer Bioinformatics and Cancer Epidemiology. *Journal of Biomedicine and Biotechnology*, 2009, 582697. <http://doi.org/10.1155/2009/582697>
20. Riquelme J, Ruiz R, Gilbert K. (2006). Minería de datos: Conceptos y tendencias. *Revista Iberoamericana de Inteligencia Artificial* 10 (29):11-18.
21. Shannon CE. "A mathematical theory of communication." *ACM SIGMOBILE Mobile Computing and Communications Review* 5.1, 1948, pp. 3-55.