TecnoLógicas

Editorial

Avances en modelos de aprendizaje automático para la predicción de toxicidad en sistemas acuáticos: tendencias y perspectivas

Karel Diéguez-Santana[1]

[1]Universidad Regional Amazónica Ikiam, Tena - Ecuador, karel.dieguez@ikiam.edu.ec

La evaluación del riesgo ecotoxicológico en sistemas acuáticos constituye un pilar fundamental de la gestión ambiental contemporánea. La presencia de compuestos de origen industrial, agrícola y farmacéutico en cuerpos de agua plantea desafíos urgentes para la regulación y el monitoreo ambiental. Entidades como la Organización para la Cooperación y el Desarrollo Económicos (OCDE), la Agencia de Protección Ambiental de los Estados Unidos (US-EPA, por sus siglas en inglés) y la Unión Europea, a través del reglamento sobre el registro, la evaluación, la autorización y la restricción de sustancias químicas (REACH), exigen la evaluación toxicológica de miles de sustancias en organismos acuáticos. Sin embargo, los ensayos experimentales son costosos, consumen tiempo y plantean dilemas éticos cuando involucran organismos vivos [1], [2].

En este contexto, los modelos de relación cuantitativa estructura-actividad (QSAR, por sus siglas en inglés) y estructura-toxicidad (QSTR, por sus siglas en inglés) han emergido como herramientas computacionales clave para estimar el potencial tóxico de compuestos a partir de su estructura molecular. En las últimas décadas, la integración de algoritmos de aprendizaje automático (machine learning, ML) ha transformado este campo, al ampliar la capacidad predictiva, la cobertura química y la transparencia interpretativa de los modelos, colocándolos como alternativas robustas a los métodos *in vivo* [3]-[5].

Los modelos QSTR basados en ML han evolucionado recientemente en tres frentes interconectados. El primero es la diversificación de organismos modelo: los estudios han ampliado su alcance más allá de la trucha arcoíris (*Oncorhynchus mykiss*) y *Daphnia magna* hacia especies como el rotífero *Brachionus calyciflorus* y el misidáceo *Americamysis bahia*, organismos de relevancia ecológica en ecosistemas tropicales y subtropicales [1], [2]. El segundo frente es la adopción de algoritmos de ensamble, como random forest, gradient boosting, redes neuronales profundas (deep learning), que han incrementado la precisión predictiva con valores de R^2 que superan 0,85 en conjuntos de prueba independientes para múltiples familias de contaminantes [3], [6].

El tercer frente, el más prometedor desde una perspectiva regulatoria, es el de la inteligencia artificial explicable (Explainable AI, XAI). Históricamente, los modelos de alta precisión operaban como "cajas negras", dificultando la identificación de los mecanismos moleculares subyacentes a la toxicidad. Herramientas como SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-Agnostic Explanations) y los modelos basados en la teoría de la perturbación con ML (PTML) han abierto la posibilidad de cuantificar la contribución de descriptores moleculares específicos a las predicciones, con implicaciones directas para el diseño de químicos más seguros y para la aceptación regulatoria de estos modelos [3], [7].

Un avance de gran impacto práctico ha sido la transición desde modelos de uso exclusivo científico hacia plataformas web de acceso abierto. Aplicaciones como T.E.S.T. (Toxicity Estimation Software Tool, por sus siglas en inglés) de la US-EPA, VEGA-QSAR y nuevas

herramientas como AQUA Tox, diseñada para predecir la toxicidad en rotíferos mediante modelos interpretables integrados, representan un avance importante hacia la democratización del análisis toxicológico computacional [3]. Estas plataformas permiten que laboratorios con recursos limitados, reguladores y la industria química accedan a predicciones confiables sin requerir experiencia en programación ni en quimioinformática.

La incorporación de módulos de incertidumbre y dominio de aplicabilidad en estas herramientas resulta esencial para garantizar que las predicciones sean científicamente sólidas y regulatoriamente aceptables. Por ello, el cumplimiento de los Principios de la OCDE para modelos QSAR/QSTR es un requisito indispensable para su uso en contextos normativos [4].

A pesar de los notables avances, aún persisten desafíos importantes. La disponibilidad y calidad de los datos de toxicidad acuática continúa siendo la limitación más crítica, las bases de datos existentes presentan sesgos hacia contaminantes orgánicos clásicos, con escasa cobertura de contaminantes emergentes como nanomateriales, microplásticos, fármacos y sus productos de transformación [5], [7]. La transferibilidad de los modelos entre especies y entornos ambientales distintos representa otro obstáculo por resolver, principalmente para ecosistemas tropicales, que tienen escasa representación en las bases de datos globales.

Desde la perspectiva metodológica, la integración de datos ómicos (transcriptómica, metabolómica) con modelos QSTR multi-objetivo ofrece una vía prometedora para capturar mecanismos de acción a nivel molecular, que pueden superar la aproximación fenotípica tradicional. De la misma manera, la convergencia entre ML y modelado de dinámica molecular podría proporcionar modelos mecanísticos más robustos para la evaluación del riesgo ecológico acumulativo [8]. En términos regulatorios, la aceptación formal de los modelos ML por parte de la OCDE y la Agencia Europea de Sustancias y Mezclas Químicas (ECHA) avanza de manera gradual, en gran parte por el rigor creciente en la documentación de la transparencia mecanística [4], [9], [10].

En conclusión, el aprendizaje automático ha consolidado su papel como herramienta indispensable en la predicción de toxicidad para sistemas acuáticos ya que ofrece velocidad, cobertura y, cada vez más, interpretabilidad mecanística. El reto inmediato para la comunidad científica iberoamericana radica en ampliar la representación de sus ecosistemas en las bases de datos globales y en desarrollar modelos adaptados a la biodiversidad acuática regional, que puedan contribuir tanto a la generación de conocimiento de frontera como al fortalecimiento de marcos regulatorios basados en evidencia computacional.

REFERENCIAS

[1] K. Diéguez-Santana, G. M. Casanola-Martin, R. Torres-Gutiérrez, B. Rasulev, and H. González-Díaz, “First report on Quantitative Structure-Toxicity Relationship modeling approaches for the prediction of acute toxicity of various organic chemicals against rotifer species,” *Sci. Total Environ.*, vol. 977, p. 179350, May. 2025. https://doi.org/10.1016/j.scitotenv.2025.179350

[2] K. Diéguez Santana, M. M. Nachimba-Mayanchi, A. Puris, R. Torres Gutiérrez, and H. González-Díaz, “Prediction of acute toxicity of pesticides for *Americamysis bahia* using linear and nonlinear QSTR modelling approaches,” *Environ. Res.*, vol. 214, no. Part 3, p. 113984, Nov. 2022. https://doi.org/10.1016/j.envres.2022.113984

[3] K. Diéguez-Santana, G. M. Casanola-Martin, R. Torres-Gutiérrez, B. Rasulev y H. González-Díaz, “AQUA Tox: A web tool for predicting aquatic toxicity in rotifer species using intrinsic explainable models,” *J. Hazard. Mater.*, vol. 492, p. 138050, Jul. 2025. https://doi.org/10.1016/j.jhazmat.2025.138050

[4] K. Diéguez-Santana, H. Pham-The, P. J. Villegas-Aguilar, H. Le-Thi-Thu, J. A. Castillo-Garit, and G. M. Casañola-Martin, “Prediction of acute toxicity of phenol derivatives using multiple linear regression approach for *Tetrahymena pyriformis* contaminant identification in a median-size database,” *Chemosphere*, vol. 165, pp. 434-441, Dec. 2016. https://doi.org/10.1016/j.chemosphere.2016.09.041

[5] E. N. Muratov *et al.*, “QSAR without borders,” *Chem. Soc. Rev.*, vol. 49, no. 11, pp. 3525–3564, Jun. 2020. https://doi.org/10.1039/D0CS00098A

[6] K. Diéguez-Santana, B. Rasulev, and H. González-Díaz, "Towards Rational Nanomaterial Design by Prediction of Drug-Nanoparticle Systems Interaction vs. Bacteria Metabolic Networks," *Environ. Sci.: Nano*, vol. 9, pp. 1391-1413, Apr. 2022. https://doi.org/10.1039/D1EN00967B

[7] J. A. Castillo-Garit *et al.*, "Aplicaciones y potencialidades de los métodos de diseño computacional en estudios ambientales y farmacocinéticos," *Anales de la ACC*, vol. 11, no. 1, Aug. 2021. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2304-01062021000100013

[8] H. Tan *et al.*, "Deep learning in environmental toxicology: Current progress and open challenges," *ACS ES&T Water*, vol. 4, no. 3, pp. 805–819, Jun. 2024. https://doi.org/10.1021/acsestwater.3c00152

[9] L. D. Burgoon, F. M. Kluxen, and M. Frericks. "Understanding and overcoming the technical challenges in using in silico predictions in regulatory decisions of complex toxicological endpoints–a pesticide perspective for regulatory toxicologists with a focus on machine learning models," *Regulatory Toxicology and Pharmacology,* vol. 137, p. 105311. Jan. 2023. https://doi.org/10.1016/j.yrtph.2022.105311

[10] A. Gajewicz-Skretna, A. Furuhama, H. Yamamoto, and N. Suzuki, "Generating accurate in silico predictions of acute aquatic toxicity for a range of organic chemicals: Towards similarity-based machine learning methods," *Chemosphere*, vol. 280, p. 130681, Oct. 2021. https://doi.org/10.1016/j.chemosphere.2021.130681