



Prédire la toxicité des produits chimiques avec l'IA

18 juillet 2024 | Ori Schipper

Catégories: Polluants

Des chercheuses et chercheurs de l'Eawag et du Swiss Data Science Center ont entraîné les algorithmes d'une IA avec un jeu de données écotoxicologiques complet. Leurs modèles d'apprentissage automatique peuvent désormais prédire la toxicité des produits chimiques sur les poissons.

Les produits chimiques sont omniprésents dans notre quotidien, notamment dans la production de denrées alimentaires, de médicaments et de biens de première nécessité. Divers mécanismes de contrôle surveillent de près leur influence sur la santé humaine et l'environnement. Dans son règlement REACH par exemple, l'UE impose que des tests soient effectués pour évaluer la toxicité pour les poissons de tous les produits chimiques dont les volumes de production atteignent au moins 10 tonnes par an. Ces analyses sont chères et nécessitent environ 50'000 poissons par an en Europe.

La science travaille depuis plusieurs décennies à des méthodes alternatives moins coûteuses et pouvant surtout se passer d'expérimentation animale. Ces approches fondées sur l'informatique et capables de prédire les effets des produits chimiques sur les poissons suscitent de grands espoirs.

Des modèles au pouvoir prédictif prometteur

L'institut de recherche sur l'eau Eawag et le Swiss Data Science Center (SDSC) ont réalisé ensemble la curation d'un jeu de données écotoxicologiques complet, afin de le mettre à disposition de la communauté scientifique en tant que benchmark et base de développement pour de nouveaux algorithmes d'IA en écotoxicologie. Baptisé «ADORE», ce jeu se compose d'environ 26'000 points de données qui décrivent les effets de près de 2'000 produits chimiques sur 140 espèces de poissons. Il

contient en outre un grand nombre de caractéristiques à la fois des produits chimiques et des espèces.

Comme l'expliquent les chercheuses et chercheurs dans un article qui vient de paraître, les modèles d'apprentissage automatique réussissent bien à prédire la toxicité de ces substances. «Les écarts constatés se situent dans la fourchette normale des fluctuations biologiques», déclarent les deux auteurs principaux de la publication, Lilian Gasser, experte en sciences des données au SDSC, et Christoph Schür, postdoctorant à l'Eawag. Les chercheuses et chercheurs estiment donc que les méthodes étudiées sont «prometteuses pour prédire la mortalité aiguë des poissons». En outre, ces méthodes pourraient être appliquées à d'autres groupes d'espèces sur la base de données appropriées.

«Néanmoins, il existe encore quelques restrictions dont il faut tenir compte», constate l'équipe dans un esprit autocritique. En effet, bien qu'ils fournissent en moyenne des prévisions exploitables, les algorithmes se trompent parfois clairement sur certains poissons. Ils estiment par exemple la toxicité d'un produit chimique est trop élevée pour des espèces données et trop faible pour d'autres. «Apparemment, les modèles sont principalement influencés par quelques propriétés chimiques et ne tiennent pas suffisamment compte des sensibilités spécifiques des espèces», déclare L. Gasser.

Des résultats pertinents grâce à une procédure de test appropriée

Lors de leurs travaux, L. Gasser et C. Schür ont tenu compte du fait que la manière de répartir les données dans un jeu d'entraînement et de test exerce une influence décisive sur l'applicabilité des modèles d'apprentissage automatique. « Il est indispensable de tester l'algorithme uniquement sur des produits chimiques qui ne sont pas présents dans le jeu de données d'entraînement, afin de prouver qu'il est capable d'identifier des caractéristiques chimiques et de prédire effectivement la toxicité », commentent L. Gasser et C. Schür.

L'avenir de la sûreté des produits chimiques

Selon L. Gasser et C. Schür et leurs co-auteurs, il est peu probable que les modèles d'apprentissage automatique et l'intelligence artificielle rendent bientôt superflus les tests de toxicité sur les poissons, mais ils contribuent à leur réduction à long terme. Les scientifiques souhaitent une évaluation plus ciblée de la sûreté des produits chimiques pour qu'elle tienne compte à l'avenir non seulement des propriétés physico-chimiques des substances et des données relatives à la mortalité, mais aussi d'autres facteurs biologiques.

Les prédictions des modèles pourraient ainsi être combinées à des évaluations d'une série d'autres tests, réalisés sans expérimentation animale, qui sont actuellement développés et validés à l'Eawag sur diverses lignées cellulaires de poissons. Pour développer un tel système de sûreté des produits chimiques à forte pertinence, les chercheuses et chercheurs encouragent une étroite collaboration avec les autorités de régulation afin de faire avancer en commun la transposition de la recherche dans la pratique.

Photo de couverture: Les poissons servent souvent d'animaux de laboratoire.
L'apprentissage automatique devrait remplacer les expérimentations animales (Photo: AdobeStock).

Publication originale

.extbase-debugger-tree{position:relative}.extbase-debugger-tree input{position:absolute

property{color:#f1f1f1}.extbase-debugger-center .extbase-debug-closure{color:#9BA223;}Extbase Variable Dumparray(3 items) publications => '33052,32142' (11 chars) libraryUrl => '' (0 chars) layout => '0' (1 chars) Extbase Variable Dumparray(2 items) 0 => Snowflake\Publications\Domain\Model\Publicationprototypepersistent entity (uid=33052, pid=124) originalId => protected33052 (integer) authors => protected'Gasser, L.; Schür, C.; Perez-Cruz, F.; Schirmer, K.; Baity-Jesi, M.' (93 chars) title => protected'Machine learning-based prediction of fish acute mortality: implementation, interpretation, and regulatory relevance' (115 chars) journal => protected'Environmental Science: Advances' (31 chars) year => protected2024 (integer) volume => protected3 (integer) issue => protected'8' (1 chars) startpage => protected'1124' (4 chars) otherpage => protected'1138' (4 chars) categories => protected'' (0 chars) description => protected'Regulation of chemicals requires knowledge of their toxicological effects on a large number of species, which has traditionally been acquired through in vivo testing. The recent effort to find alternatives based on machine learning, however, has not focused on guaranteeing transparency, comparability and reproducibility, which makes it difficult to assess advantages and disadvantages of these methods. Also, comparable baseline performances are needed. In this study, we trained regression models on the ADORE "t-F2F" challenge proposed in [Schür et al., Nature Scientific data, 2023] to predict acute mortality, measured as LC50 (lethal concentration 50), of organic compounds on fishes. We trained LASSO, random forest (RF), XGBoost, Gaussian process (GP) regression models, and found a series of aspects that are stable across models: (i) using mass or molar concentrations does not affect performances; (ii) the performances are only weakly dependent on the molecular representations of the chemicals, but (iii) strongly on how the data is split. Overall, the tree-based models RF and XGBoost performed best and we were able to predict the log10-transformed LC50 with a root mean square error of 0.90, which corresponds to an order of magnitude on the original LC50 scale. On a local level, on the other hand, the models are not able to consistently predict the toxicity of individual chemicals accurately enough. Predictions for single chemicals are mostly influenced by a few chemical properties while taxonomic traits are not captured sufficiently by the models. We discuss technical and conceptual improvements for these challenges to enhance the suitability of in silico methods to environmental hazard assessment. Accordingly, this work showcases state-of-the-art models and contributes to the ongoing discussion on regulatory integration.' (1853 chars) serialnumber => protected'' (0 chars) doi => protected'10.1039/d4va00072b' (18 chars) uid => protected33052 (integer) _localizedUid => protected33052 (integer)modified _languageUid => protectedNULL _versionedUid => protected33052 (integer)modified pid => protected124 (integer) 1 => Snowflake\Publications\Domain\Model\Publicationprototypepersistent entity (uid=32142, pid=124) originalId => protected32142 (integer) authors => protected'Schür, C.; Gasser, L.; Perez-Cruz, F.; Schirmer, K.; Baity-Jesi, M.' (93 chars) title => protected'A benchmark dataset for machine learning in ecotoxicology' (57 chars) journal => protected'Scientific Data' (15 chars) year => protected2023 (integer) volume => protected10 (integer) issue => protected'1' (1 chars) startpage => protected'718 (20 pp.)' (12 chars) otherpage => protected'' (0 chars) categories => protected'' (0 chars) description => protected'The use of machine learning for predicting

ecotoxicological outcomes is promising, but underutilized. The curation of data with informative features requires both expertise in machine learning as well as a strong biological and ecotoxicological background, which we consider a barrier of entry for this kind of research. Additionally, model performances can only be compared across studies when the same dataset, cleaning, and splittings were used. Therefore, we provide **ADORE**, an extensive and well-described dataset on acute aquatic toxicity in three relevant taxonomic groups (fish, crustaceans, and algae). The core dataset describes ecotoxicological experiments and is expanded with phylogenetic and species-specific data on the species as well as its chemical properties and molecular representations. Apart from challenging other researchers to try and achieve the best model performances across the whole dataset, we propose specific relevant challenges on subsets of the data and include datasets and splittings corresponding to each of these challenges as well as in-depth characterization and discussion of train-test splitting approaches.

(1154 chars) serialnumber => protected" (0 chars) doi => protected'10.1038/s41597-023-02612-2' (26 chars) uid => protected32142 (integer) _localizedUid => protected32142 (integer)modified _languageUid => protectedNULL _versionedUid => protected32142 (integer)modified pid => protected124 (integer) Gasser, L.; Schür, C.; Perez-Cruz, F.; Schirmer, K.; Baity-Jesi, M. (2024) Machine learning-based prediction of fish acute mortality: implementation, interpretation, and regulatory relevance, *Environmental Science: Advances*, 3(8), 1124-1138, [doi:10.1039/d4va00072b](https://doi.org/10.1039/d4va00072b), [Institutional Repository](#)

Schür, C.; Gasser, L.; Perez-Cruz, F.; Schirmer, K.; Baity-Jesi, M. (2023) A benchmark dataset for machine learning in ecotoxicology, *Scientific Data*, 10(1), 718 (20 pp.), [doi:10.1038/s41597-023-02612-2](https://doi.org/10.1038/s41597-023-02612-2), [Institutional Repository](#)

Financement / Coopération

Eawag Swiss Data Science Center (SDSC) ETH Zürich EPFL European Partnership for the Assessment of Risks from Chemicals (PARC) Horizon Europe

Links

Projekt «ML Tox»

Contact



Christoph Schuer

Chercheur postdoctoral

Tel. +41 58 765 5684

christoph.schuer@eawag.ch



Annette Ryser

Rédactrice scientifique

Tel. +41 58 765 6711

annette.ryser@eawag.ch

<https://www.eawag.ch/fr/portail/dinfo/actualites/news-archives/detail-de-larchive/predire-la-toxicite-des-produits-chimiques-avec-lia>