Prediction of the eco-neurotoxic potential of chemicals

Project: **MLTox: Enhancing toxicological testing through machine learning**

Eawag: Christoph Schür, Marco Baity-Jesi, Kristin Schirmer
SDSC, ETH Zürich: Lilian Gasser, Fernando Perez-Cruz

Background

Chemicals can only be put on the market if they are deemed safe both for humans and the environment. This registration process involves extensive animal testing. For instance, in ecotoxicology, the tests are mainly carried out on fish, crustaceans and algae. Given the evident ethical and economical concerns of animal testing, several approaches are taken to reduce it, summarized under the term New Approach Methods (NAMs). Some NAMs are based on Quantitative Structure Activity Relationship (QSAR) models, which are widely used in cheminformatics to predict activities or properties of chemicals based on their structure. Other alternatives are, for example, based on performing tests on isolated cells [3] or fish early life stages (fish embryo acute toxicity (FET) test) [2] instead of on the whole adult organism.

In the MLTox project, we focus on predicting toxicity, more specifically acute mortality of fish, crustaceans, and algae using *in vivo* experimental data. To foster comparability across studies that predict ecotoxicological outcomes a common dataset was needed. Hence, we created the ADORE dataset, which is based on the ECOTOX database and also contains species-specific and taxonomic information [4]. The response variable for mortality is the lethal concentration 50 (LC50), *i.e.*, the concentration at which half of a population dies. Currently, we are developing machine learning models to predict LC50 that integrate chemical, species-related and experimental features.

A blind spot that NAMs are currently not able to cover better than testing on adult fish is neurotoxicity, *i.e.*, chemicals that specifically act on the nervous system and can induce effects ranging from behavioral alterations to death. Accordingly, neurotoxicity can be studied in different ways (see Figure 1 from Legradi et al. [1]) and can be focused on humans, animals, or specifically interfere with developing nervous systems (developmental neurotoxicity, DNT). Our research focus is on ecological neurotoxicity (eco-neurotoxicity), *i.e.*, "*neurotoxicity resulting from exposure to environmental chemicals in species other than humans (e.g., fish, birds, invertebrates). It is important to distinguish between human and non-human neurotoxicity as the effects of exposure to compounds, both in terms of levels and pathways, as well as the structure and function of the nervous system itself, can differ widely between species.*" [1].

This variety of approaches to characterize neurotoxicity makes it difficult to conclusively define a substance as potentially neurotoxic. In this project, the goal is to establish a machine learning approach that predicts a given substance as either potentially neurotoxic or not. For this, firstly, several neurotoxicity categorization schemes need to be evaluated, which includes the compilation of the corresponding datasets from existing databases such as Tox21/ToxCast. In the second step, machine learning models need to be trained for neurotoxicity prediction using chemical properties and molecular representations as input features. To better understand the model, feature importances will then be analyzed (Figure 2). Ideally, molecular representations and chemical properties would be sufficiently informative for such a model, but expansion of the feature space may be necessary to improve prediction performance.

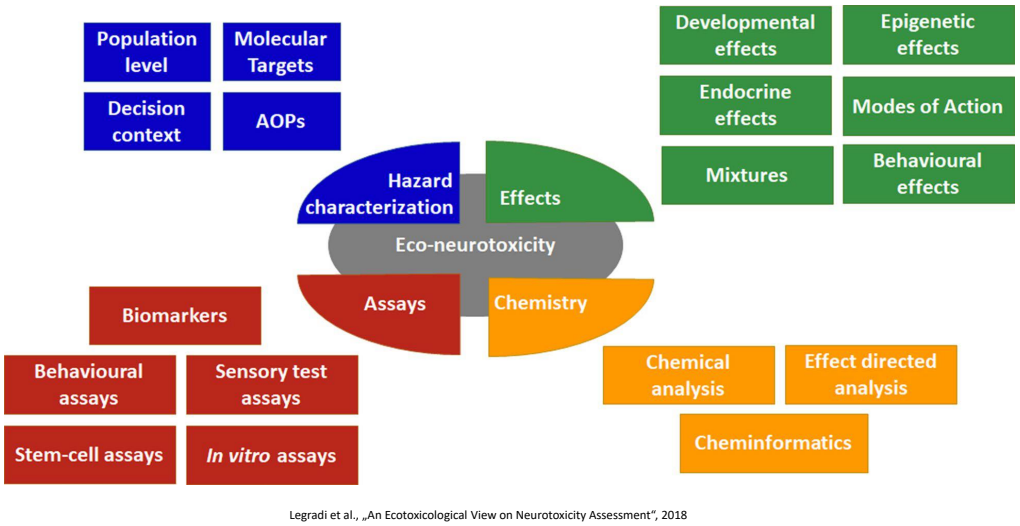

Figure 1: Different ways to study and characterize eco-neurotoxicity, subdivided in hazard characterization, effects, chemistry and assays.

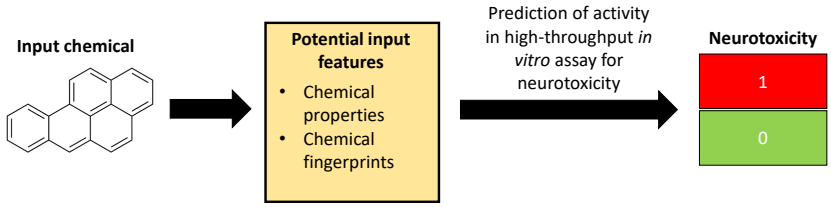

Figure 2: Schematic of a potential modeling pipeline for neurotoxicity.

In this MSc project, the aim is to

- generate an overview of different classification schemes of neurotoxicity and the related literature;
- compile a list of chemicals that are potentially neurotoxic and chemicals that are not (*i.e.*, establish a ground truth for modeling). This step requires ecotoxicological knowledge which will be provided;
- train machine learning models to predict neurotoxicity.

Additional Information

- **What will you learn?**
  - Assessment of different ways to measure and identify neurotoxicity
  - Integration of different data sources that cover cheminformatics and, potentially, high-throughput *in vitro* data
- **Requirements:**
  - Good knowledge of Python and/or R, and git
  - Experience with classification is an advantage
  - A strong interest in chemistry and cheminformatics
  - A basic understanding of biology is helpful
- **Supervisors and collaborators:**
  - Eawag: Christoph Schür, Marco Baity-Jesi, Kristin Schirmer
  - ETHZ: Lilian Gasser, Fernando Perez-Cruz
- Please contact Lili Gasser (lilian.gasser@sdsc.ethz.ch) or Christoph Schür (christoph.schuer@eawag.ch) for further information

References

[1] J. B. Legradi et al. "An Ecotoxicological View on Neurotoxicity Assessment". In: *Environmental Sciences Europe* 30.1 (Dec. 2018), p. 46. ISSN: 2190-4707, 2190-4715. DOI: 10.1186/s12302-018-0173-x. (Visited on 12/01/2021).

[2] OECD. *Test No. 236: Fish Embryo Acute Toxicity (FET) Test*. OECD Guidelines for the Testing of Chemicals, Section 2. OECD Publishing, July 2013. ISBN: 978-92-64-20370-9. DOI: 10.1787/9789264203709-en. (Visited on 01/11/2018).

[3] OECD. *Test No. 249: Fish Cell Line Acute Toxicity: The RTgill-W1 Cell Line Assay*. OECD, 2021.

[4] Christoph Schür et al. "A Benchmark Dataset for Machine Learning in Ecotoxicology". In: *Scientific Data* 10.1 (Oct. 2023), p. 718. ISSN: 2052-4463. DOI: 10.1038/s41597-023-02612-2.