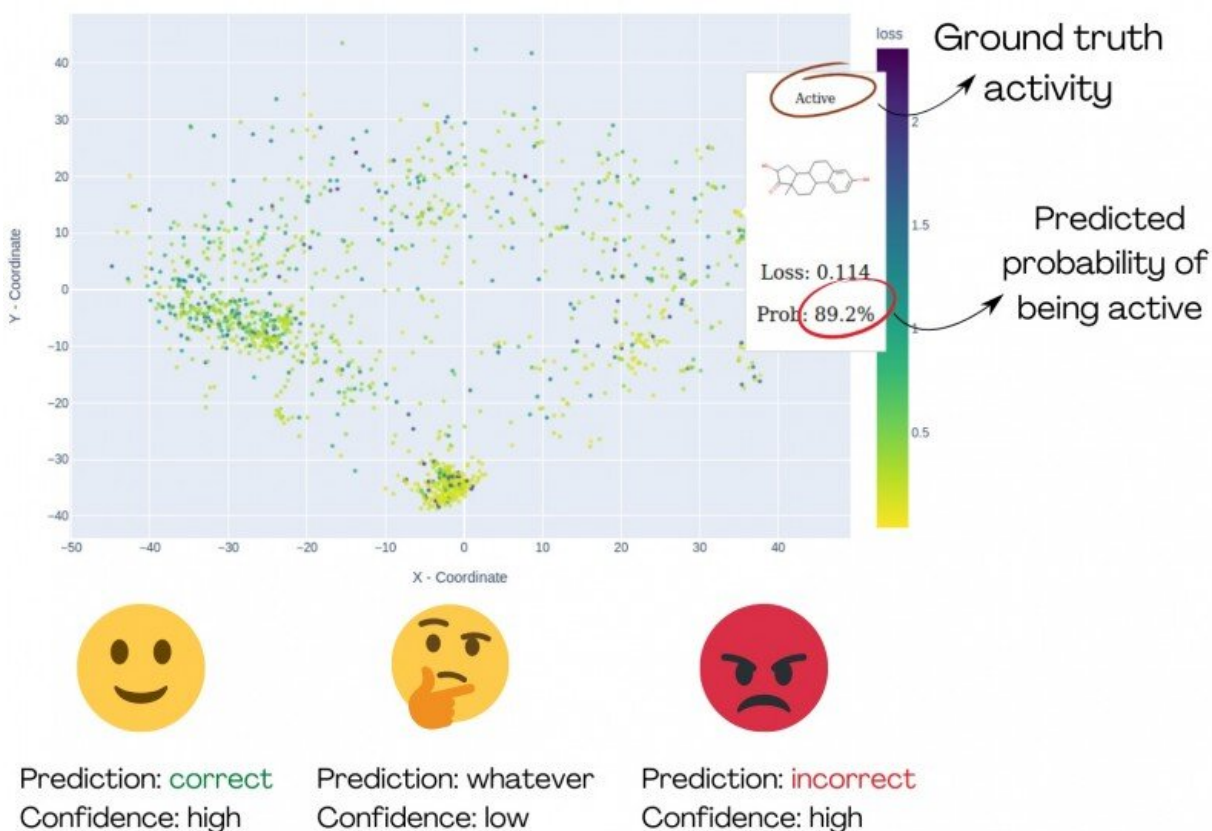


New software tool aims to reduce reliance on animal testing

August 20 2024



The demonstration of MolCompass illustrates how a computational toxicologist can identify concerning areas of chemical space. Using our software, the toxicologist can pinpoint regions where the model under investigation incorrectly predicts activity with high confidence. Credit: Sergey Sosnin

In recent years, machine learning models have become increasingly popular for risk assessment of chemical compounds. However, they are often considered "black boxes" due to their lack of transparency, leading to skepticism among toxicologists and regulatory authorities.

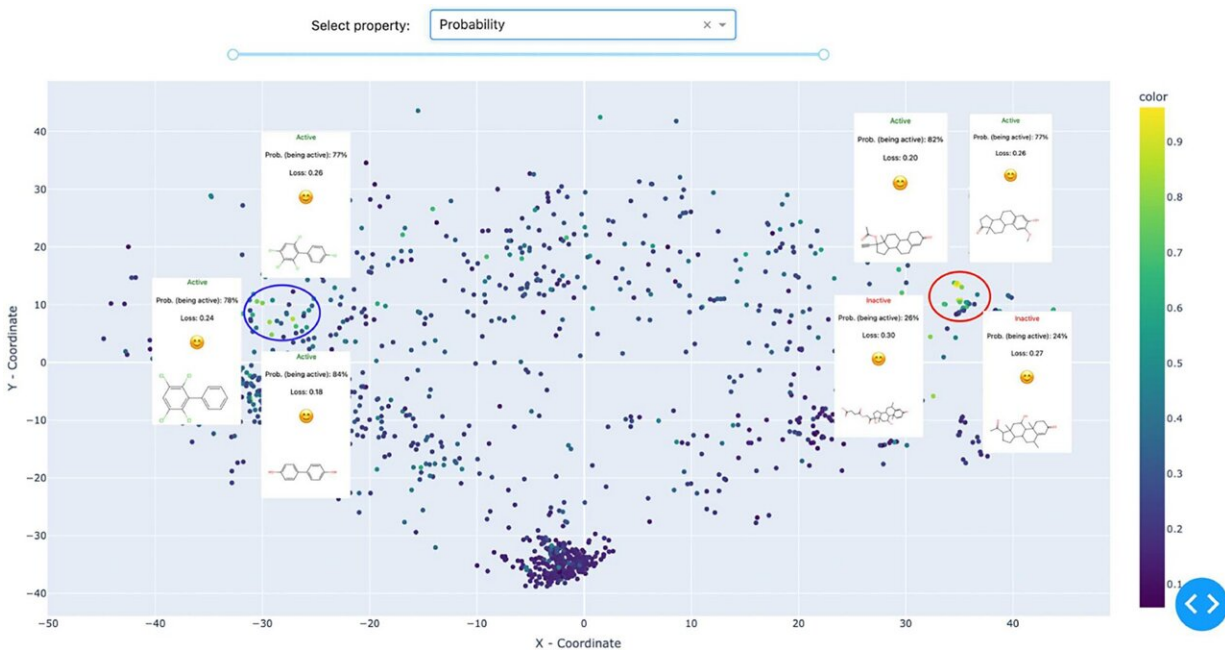
To increase confidence in these models, researchers at the University of Vienna proposed to carefully identify the areas of chemical space where these models are weak. They developed an innovative software tool ("MolCompass") for this purpose and the results of this research approach have just been [published](#) in the *Journal of Cheminformatics*.

Over the years, new pharmaceuticals and cosmetics have been tested on animals. These tests are expensive, raise [ethical concerns](#), and often fail to accurately predict human reactions. Recently, the European Union supported the RISK-HUNT3R project to develop the next generation of non-animal risk assessment methods.

The University of Vienna is a member of the project consortium. Computational methods now allow the toxicological and environmental risks of new chemicals to be assessed entirely by computer, without the need to synthesize the chemical compounds. But one question remains: How confident are these computer models?

It's all about reliable prediction

To address this issue, Sergey Sosnin, a senior scientist of the Pharmacoinformatics Research Group at the University of Vienna, focused on binary classification. In this context, a [machine learning model](#) provides a probability score from 0% to 100%, indicating whether a chemical compound is active or not (e.g., toxic or non-toxic, bioaccumulative or non-bioaccumulative, a binder or non-binder to a specific human protein). This probability reflects the confidence of the model in its prediction.



Two clusters have been attributed with high confidence by the reference model. The denser cluster on the left represents steroid derivatives, while the right, less defined cluster includes polychlorinated biphenyls and polyphenols. The visualized data is: Estrogen Binders dataset (see sec. Reference model)

Ideally, the model should be confident only in its correct predictions. If the model is uncertain, giving a confidence score around 51%, these predictions can be disregarded in favor of alternative methods. A challenge arises, however, when the model is fully confident in incorrect predictions.

"This is the real nightmare scenario for a computational toxicologist," says Sosnin. "If a model predicts that a compound is non-toxic with 99% confidence, but the compound is actually toxic, there is no way to know that something was wrong."

The only solution is to identify areas of "chemical space"—encompassing possible classes of organic compounds—where the model has 'blind spots' in advance and avoid them. To do this, a researcher evaluating the model must check the predicted results for thousands of chemical compounds one by one—a tedious and error-prone task.

Overcoming this significant hurdle

"To assist these researchers," Sosnin continues. "We developed interactive graphical tools that display chemical compounds onto a 2D plane, like geographical maps. Using colors, we highlight the compounds that were predicted incorrectly with high confidence, allowing users to identify them as clusters of red dots. The map is interactive, enabling users to investigate the chemical space and explore regions of concern."

The methodology was proven using an estrogen receptor binding model. After visual analysis of the chemical space, it became clear that the model works well for e.g. steroids and polychlorinated biphenyls, but fails completely for small non-cyclic compounds and should not be used for them.

The software developed in this project is freely available to the community on GitHub. Sosnin hopes that MolCompass will lead chemists and toxicologists to a better understanding of the limitations of computational models. This study is a step toward a future where [animal testing](#) is no longer necessary and the only workplace for a toxicologist is a computer desk.

More information: Sergey Sosnin, MolCompass: multi-tool for the navigation in chemical space and visual validation of QSAR/QSPR models, *Journal of Cheminformatics* (2024). [DOI: 10.1186/s13321-024-00888-z](https://doi.org/10.1186/s13321-024-00888-z)

Provided by University of Vienna

Citation: New software tool aims to reduce reliance on animal testing (2024, August 20)
retrieved 20 August 2024 from <https://phys.org/news/2024-08-molecular-compass-reduction-animal.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.