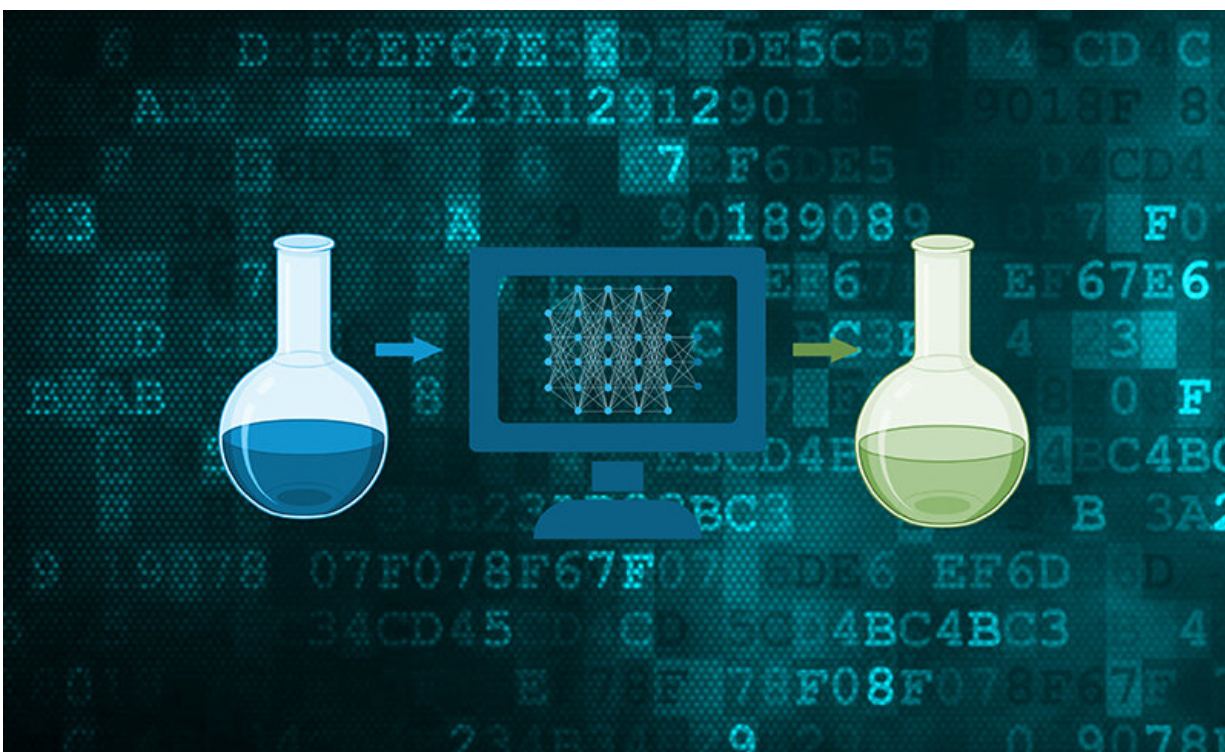# Machine learning innovation to develop chemical library for drug discovery

November 18 2020, by Chris Adam

Purdue University scientists are using machine learning models to create new options for drug discovery pipelines. Credit: Purdue University/Gaurav Chopra

Machine learning has been used widely in the chemical sciences for drug design and other processes.

The models that are prospectively tested for new reaction outcomes and

used to enhance human understanding to interpret [chemical](link) reactivity decisions made by such models are extremely limited.

Purdue University innovators have introduced chemical reactivity flowcharts to help chemists interpret reaction outcomes using statistically robust [machine learning](link) models trained on a small number of reactions. The work is published in *Organic Letters*.

"Developing new and fast reactions is essential for chemical library design in [drug discovery](link)," said Gaurav Chopra, an assistant professor of analytical and [physical chemistry](link) in Purdue's College of Science. "We have developed a new, fast and one-pot multicomponent reaction (MCR) of N-sulfonylimines that was used as a representative case for generating [training data](link) for machine learning models, predicting reaction outcomes and testing new reactions in a blind prospective manner.

"We expect this work to pave the way in changing the current paradigm by developing accurate, human understandable machine learning models to interpret reaction outcomes that will augment the creativity and efficiency of human chemists to discover new chemical reactions and enhance organic and process chemistry pipelines."

Chopra said the Purdue team's human-interpretable machine learning approach, introduced as chemical reactivity flowcharts, can be extended to explore the reactivity of any MCR or any chemical reaction. It does not need large-scale robotics since these methods can be used by the chemists while doing reaction screening in their laboratories.

"We provide the first report of a framework to combine fast synthetic chemistry experiments and quantum chemical calculations for understanding reaction mechanism and human-interpretable statistically robust machine learning models to identify chemical patterns for predicting and experimentally testing heterogeneous reactivity of N-

sulfonylimines," Chopra said.

"The unprecedented use of a machine learning model in generating chemical reactivity flowcharts helped us to understand the reactivity of traditionally used different N-sulfonylimines in MCRs," said Krupal Jethava, a postdoctoral fellow in Chopra's laboratory, who co-authored the work. "We believe that working hand-to-hand with organic and computational chemists will open up a new avenue for solving complex chemical reactivity problems for other reactions in the future."

Chopra said the Purdue researchers hope their work will pave the way to become one of many examples that will showcase the power of machine learning for new synthetic methodology development for drug design and beyond in the future.

"In this work, we strived to ensure that our machine learning model can be easily understood by chemists not well versed in this field," said Jonathan Fine, a former Purdue graduate student, who co-authored the work. "We believe that these models have the ability not only be used to predict reactions but also be used to better understand when a given reaction will occur. To demonstrate this, we used our model to guide additional substrates to test whether a reaction will occur."

Provided by Purdue University

Citation: Machine learning innovation to develop chemical library for drug discovery (2020,