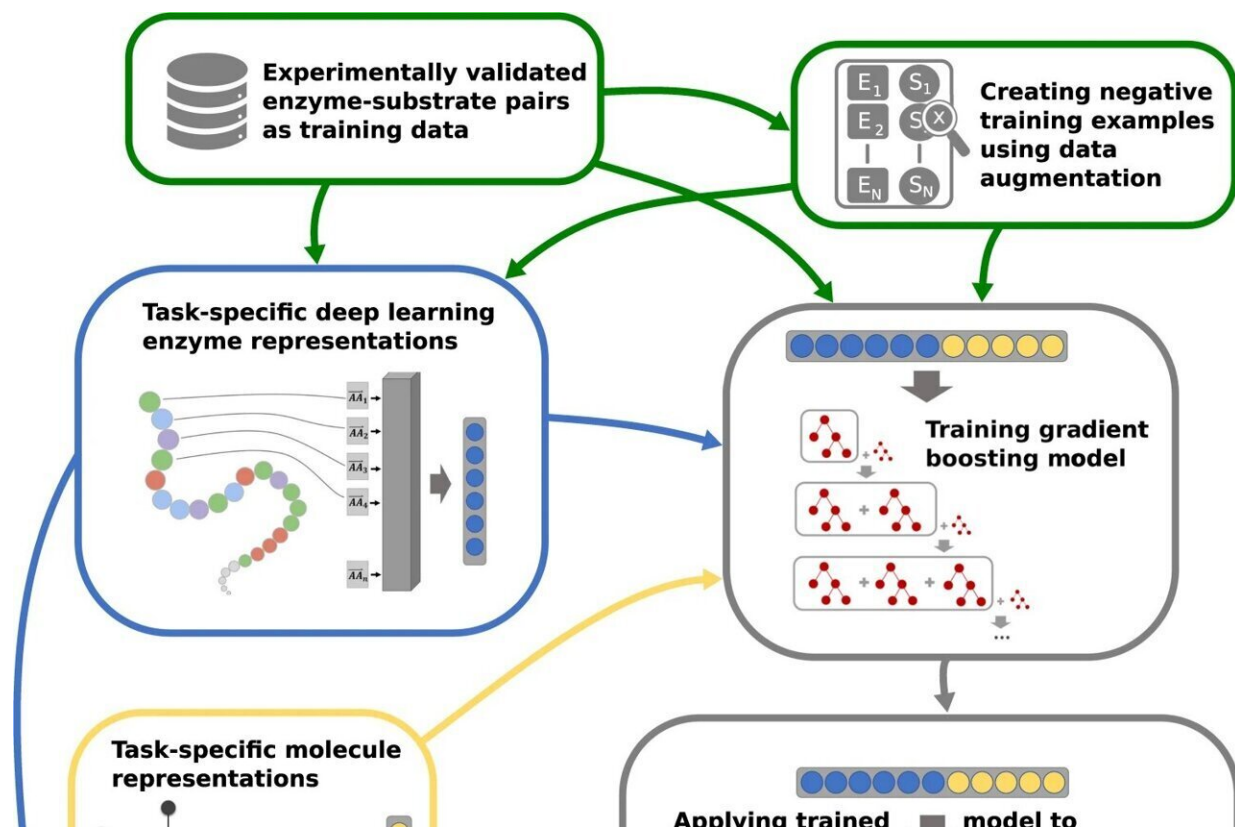# AI predicts the function of enzymes

May 22 2023

Model overview. Experimentally validated enzyme-substrate pairs and sampled negative enzyme-small metabolite pairs are numerically represented with task-specific enzyme and small molecule representations. Concatenated enzyme-small molecule representations are used to train a gradient-boosting model. After training, the fitted model can be used to predict promising candidate substrates for enzymes. Credit: *Nature Communications* (2023). DOI: 10.1038/s41467-023-38347-2

Enzymes are the molecule factories in biological cells. However, which basic molecular building blocks they use to assemble target molecules is often unknown and difficult to measure. An international team including bioinformaticians from Heinrich Heine University Düsseldorf (HHU) has now taken an important step forward in this regard: Their AI method predicts with a high degree of accuracy whether an enzyme can work with a specific substrate. They now present their results in the scientific journal *Nature Communications*.

Enzymes are important biocatalysts in all living cells: They facilitate chemical reactions, through which all molecules important for the organism are produced from basic substances (substrates). Most organisms possess thousands of different enzymes, with each one responsible for a very specific reaction. The collective function of all enzymes makes up the metabolism and thus provides the conditions for the life and survival of the organism.

Even though genes which encode enzymes can easily be identified as such, the exact function of the resultant enzyme is unknown in the vast majority—over 99%—of cases. This is because experimental characterizations of their function—i.e. which starting molecules a specific enzyme converts into which concrete end molecules—is extremely time-consuming.

Together with colleagues from Sweden and India, the research team headed by Professor Dr. Martin Lercher from the Computational Cell Biology research group at HHU has developed an AI-based method for predicting whether an enzyme can use a specific molecule as a substrate for the reaction it catalyzes.

Professor Lercher says, "The special feature of our ESP ('Enzyme Substrate Prediction') model is that we are not limited to individual, special enzymes and others closely related to them, as was the case with

previous models. Our general model can work with any combination of an enzyme and more than 1,000 different substrates."

Ph.D. student Alexander Kroll, lead author of the study, has developed a so-called Deep Learning model in which information about enzymes and substrates was encoded in mathematical structures known as numerical vectors. The vectors of around 18,000 experimentally validated enzyme-substrate pairs—where the enzyme and substrate are known to work together—were used as input to train the Deep Learning model.

Alexander Kroll says, "After training the model in this way, we then applied it to an independent test dataset where we already knew the correct answers. In 91% of cases, the model correctly predicted which substrates match which enzymes."

This method offers a wide range of potential applications. In both drug research and biotechnology it is of great importance to know which substances can be converted by enzymes. Professor Lercher says, "This will enable research and industry to narrow a large number of possible pairs down to the most promising, which they can then use for the enzymatic production of new drugs, chemicals or even biofuels."

Kroll adds, "It will also enable the creation of improved models to simulate the metabolism of cells. In addition, it will help us understand the physiology of various organisms—from bacteria to people."

Alongside Kroll and Lercher, Professor Dr. Martin Engqvist from the Chalmers University of Technology in Gothenburg, Sweden, and Sahasra Ranjan from the Indian Institute of Technology in Mumbai were also involved in the study. Engqvist helped design the study, while Ranjan implemented the model which encodes the enzyme information fed into the overall model developed by Kroll.

Provided by Heinrich-Heine University Duesseldorf