Using machine learning to derive different causes from the same symptoms

May 27 2022

Head of the LipiTUM research group Dr. Josch Konstantin Pauling (left) and PhD student Nikolai Köhler (right) interpret the disease-related changes in lipid metabolism using a newly developed network. Credit: LipiTUM

Machine learning is playing an ever-increasing role in biomedical research. Scientists at the Technical University of Munich (TUM) have now developed a new method of using molecular data to extract subtypes of illnesses. In the future, this method can help to support the study of larger patient groups.

Nowadays doctors define and diagnose most diseases on the basis of

symptoms. However, that does not necessarily mean that the illnesses of patients with similar symptoms will have identical causes or demonstrate the same molecular changes. In biomedicine, one often speaks of the molecular mechanisms of a disease. This refers to changes in the regulation of genes, proteins or metabolic pathways at the onset of illness. The goal of stratified medicine is to classify patients into various subtypes at the molecular level in order to provide more targeted treatments.

To extract disease subtypes from large pools of patient data, new machine learning algorithms can help. They are designed to independently recognize patterns and correlations in extensive clinical measurements. The LipiTUM junior research group, headed by Dr. Josch Konstantin Pauling of the Chair for Experimental Bioinformatics has developed an algorithm for this purpose.

Complex analysis via automated web tool

Their method combines the results of existing algorithms to obtain more precise and robust predictions of clinical subtypes. This unifies the characteristics and advantages of each algorithm and eliminates their time-consuming adjustment. "This makes it much easier to apply the analysis in clinical research," reports Dr. Pauling. "For that reason, we have developed a web-based tool that permits online analysis of molecular clinical data by practitioners without prior knowledge of bioinformatics."

On the website, researchers can submit their data for automated analysis and use the results to interpret their studies. "Another important aspect for us was the visualization of the results. Previous approaches were not capable of generating intuitive visualizations of relationships between patient groups, clinical factors and molecular signatures. This will change with the web-based visualization produced by our MoSBi tool,"

says Tim Rose, a scientist at the TUM School of Life Sciences. MoSBi stands for molecular signatures using biclustering. Biclustering is the name of the technology used by the algorithm.

Application for clinically relevant questions

With the tool, researchers can now, for example, represent data from cancer studies and simulations for various scenarios. They have already demonstrated the potential of their method in a large-scale clinical study. In a cooperative study conducted with researchers from the Max Planck Institute in Dresden, the Technical University of Dresden and the Kiel University Clinic, they studied the change in lipid metabolism in the liver of patients with non-alcoholic fatty liver disease (NAFLD).

This widespread disease is associated with obesity and diabetes. It develops from the non-alcoholic fatty liver (NAFL), in which lipids are deposited in liver cells, to non-alcoholic steatohepatitis (NASH), in which the liver becomes further inflamed, to liver cirrhosis and the formation of tumors. Apart from dietary adjustments, no treatments have been found to date. Because the disease is characterized and diagnosed by the accumulation of various lipids in the liver, it is important to understand their molecular composition.

Biomarkers for liver disease

Using the MoSBi methods, the researchers were able to demonstrate the heterogeneity of the livers of patients in the NAFL stage at the molecular level. "From a molecular standpoint, the liver cells of many NAFL patients were almost identical to those of NASH patients, while others were still largely similar to healthy patients. We could also confirm our predictions using clinical data," says Dr. Pauling. "We were then able to identify two potential lipid biomarkers for disease

progression." This is important for early recognition of the disease and its progression and the development of targeted treatments.

The research group is already working on further applications of their method to gain a better understanding of other diseases. "In the future algorithms will play an even greater role in biomedical research than they already do today. They can make it significantly easier to detect complex mechanisms and find more targeted treatment approaches," says Dr. Pauling.

The research was published in *Proceedings of the National Academy of Sciences*.

More information: Tim Daniel Rose et al, MoSBi: Automated signature mining for molecular stratification and subtyping, *Proceedings of the National Academy of Sciences* (2022). DOI: 10.1073/pnas.2118210119

Olga Vvedenskaya et al, Nonalcoholic fatty liver disease stratification by liver lipidomics, *Journal of Lipid Research* (2021). DOI: 10.1016/j.jlr.2021.100104

MoSBi website: exbio.wzw.tum.de/mosbi/

Provided by Technical University Munich

study or research, no part may be reproduced without the written permission. The content is
provided for information purposes only.