Artificial intelligence finds disease-related genes

February 14 2020

Credit: CC0 Public Domain

An artificial neural network can reveal patterns in huge amounts of gene expression data and discover groups of disease-related genes. This has been shown by a new study led by researchers at Linköping University,

published in *Nature Communications*. The scientists hope that the method can eventually be applied within precision medicine and individualized treatment.

It's common when using social media that the platform suggests people whom you may want to add as friends. The suggestion is based on you and the other person having common contacts, which indicates that you may know each other. In a similar manner, scientists are creating maps of biological networks based on how different proteins or genes interact with each other. The researchers behind a new study have used artificial intelligence, AI, to investigate whether it is possible to discover biological networks using deep learning, in which entities known as "artificial neural networks" are trained by experimental data. Since artificial neural networks are excellent at learning how to find patterns in enormous amounts of complex data, they are used in applications such as image recognition. However, this machine learning method has until now seldom been used in biological research.

"We have for the first time used deep learning to find disease-related genes. This is a very powerful method in the analysis of huge amounts of biological information, or Big Data," says Sanjiv Dwivedi, postdoc in the Department of Physics, Chemistry and Biology (IFM) at Linköping University.

The scientists used a large database with information about the expression patterns of 20,000 genes in a large number of people. The information was "unsorted," in the sense that the researchers did not give the artificial neural network information about which gene expression patterns were from people with diseases, and which were from healthy people. The AI model was then trained to find patterns of gene expression.

One of the challenges of machine learning is that it is not possible to see

exactly how an artificial neural network solves a task. AI is sometimes described as a "black box"—we see only the information that we put into the box and the result that it produces. We cannot see the steps between. Artificial neural networks consist of several layers in which information is mathematically processed. The network comprises an input layer and an output layer that delivers the result of the information processing carried out by the system. Between these two layers are several hidden layers in which calculations are carried out. When the scientists had trained the artificial neural network, they wondered whether it was possible to lift the lid of the black box, in a manner of speaking, and understand how it works. Are the designs of the neural network and the familiar biological networks similar?

"When we analysed our neural network, it turned out that the first hidden layer represented to a large extent interactions between various proteins. Deeper in the model, in contrast, on the third level, we found groups of different cell types. It's extremely interesting that this type of biologically relevant grouping is automatically produced, given that our network has started from unclassified gene expression data," says Mika Gustafsson, senior lecturer at IFM and leader of the study.

The scientists then investigated whether their model of gene expression could be used to determine which gene expression patterns are associated with disease and which is normal. They confirmed that the model finds relevant patterns that agree well with biological mechanisms in the body. Since the model has been trained using unclassified data, it is possible that the artificial neural network has found totally new patterns. The researchers plan now to investigate whether such, previously unknown patterns, are relevant from a biological perspective.

"We believe that the key to progress in the field is to understand the neural network. This can teach us new things about biological contexts, such as diseases in which many factors interact. And we believe that our

method gives models that are easier to generalize and that can be used for many different types of biological information," says Mika Gustafsson.

Mika Gustafsson hopes that close collaboration with medical researchers will enable him to apply the method developed in the study in precision medicine. It may be possible, for example, to determine which groups of patients should receive a certain type of medicine, or identify the patients who are most severely affected.

More information: Sanjiv K. Dwivedi et al, Deriving disease modules from the compressed transcriptional space embedded in a deep autoencoder, *Nature Communications* (2020). DOI: 10.1038/s41467-020-14666-6

Provided by Linköping University