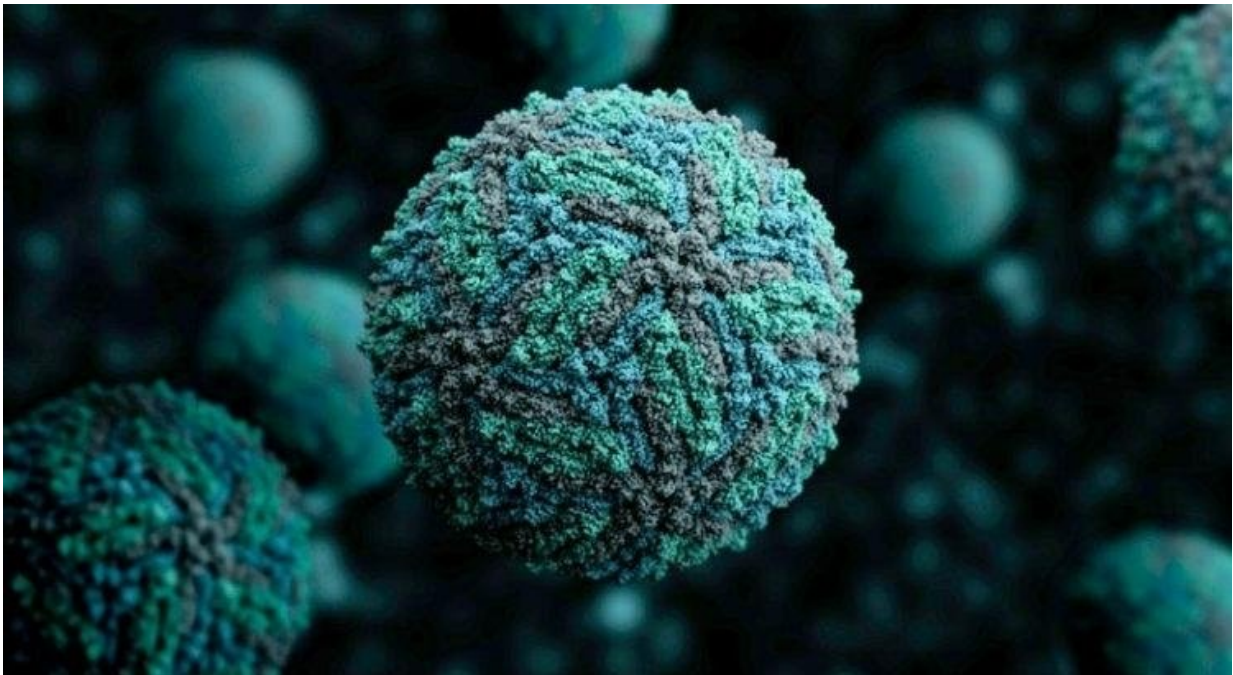# Platform uses artificial intelligence to diagnose Zika and other pathogens

June 13 2018



Method created in Brazil combines mass spectrometry analysis of blood serum with an algorithm that recognizes patterns associated with viral diseases as well as diseases of bacterial, fungal and even genetic origin. Credit: Wikimedia Commons/FAPESP

Scientists at the University of Campinas (UNICAMP) in Brazil have developed a platform that can diagnose several diseases with a high degree of precision using metabolic markers found in blood.

The method combines mass spectrometry, which can identify tens of thousands of molecules present in blood serum, with an [artificial intelligence algorithm](link) capable of finding patterns associated with diseases of viral, bacterial, fungal and even genetic origin. The research was supported by the São Paulo Research Foundation—FAPESP and conducted as part of Carlos Fernando Odir Rodrigues Melo's Ph.D. The results have been published in *Frontiers in Bioengineering and Biotechnology*.

"We used infection by Zika virus as a model to develop the platform and showed that in this case, diagnostic accuracy exceeded 95 percent. One of the main advantages is that the method doesn't lose sensitivity, even if the virus mutates," said Melo's supervisor Rodrigo Ramos Catharino, principal investigator for the project. Catharino is a professor at UNICAMP's School of Pharmaceutical Sciences (FCF) and head of its Innovare Biomarker Laboratory.

Another strength of the platform, he added, is the capacity to identify positive cases of Zika even in [blood serum](link) analyzed 30 days after the start of infection, when the acute phase of the disease is over.

"None of the currently available diagnostic kits has the sensitivity to detect infection by Zika after the end of the acute phase. The method we developed could be useful to analyze transfusion blood bags, for example," Catharino said.

## Machine learning

Development and validation of the platform involved analysis of blood samples from 203 patients treated at UNICAMP's general and teaching hospital. Of these, 82 were diagnosed with Zika by the method currently considered the gold standard in this field: real-time polymerase chain reaction (RT-PCR), which detects viral RNA in body fluids during the

acute phase of the infection.

The other 121 patients were the control group. Approximately half had the same symptoms as the group that tested positive for Zika, such as fever, joint pain, conjunctivitis and rash, but had negative RT-PCR results for Zika. The rest had no symptoms and also tested negative or were diagnosed with dengue.

All collected samples were analyzed in a mass spectrometer, a device that acts as a kind of molecular weighing scale, sorting molecules according to their mass.

"We identified some 10,000 different molecules in the patients' serum, including lipids, peptides, and fragments of DNA and RNA. Among these metabolites, there were particles produced both by Zika and by the patient's immune system in response to the infection," said the FAPESP scholarship supervisor.

All the data obtained in the spectrometry analysis of both the group that tested positive for Zika and the control group were then fed into a computer program running a random-forest machine learning algorithm. This type of artificial intelligence tool is capable of analyzing a large amount of data by specific statistical methods in search of patterns that can be used as a basis for classification, prediction, decision making, modeling and so on.

"The algorithm separates samples randomly, determines which one will be the training group and the blind group, and then carries out testing and validation. At the end, it tells us whether with that number of samples it was possible to obtain a set of metabolic markers capable of identifying patients infected by Zika," Catharino explained.

Each new set of patient data fed into the program enhances its learning

capacity and makes it more sensitive, he went on. In the case of Zika, the FAPESP-funded study established a panel of 42 biomarkers as a specific key to identifying the virus. Twelve of these were found by the algorithm to be highly prevalent in the blood of patients who tested positive for the disease.

"In this platform, it isn't important to know a lot individually about each of the molecules that serve as markers of the infection. It's the set that matters and that will tell us with a high level of accuracy whether we're looking at Zika. Moreover, even if the virus mutates, the program adapts and changes too. It's not a static methodology," Catharino said.

The UNICAMP group is currently performing tests to evaluate the platform's capacity to diagnose systemic diseases caused by fungi. They also plan to test how well it detects bacterial and genetic diseases. Anderson de Rezende Rocha, a professor at the same university's Institute of Computing (IC-UNICAMP), is collaborating on the research.

In theory, any laboratory equipped with a mass spectrometer could use the new diagnostic platform developed at UNICAMP. Mass spectrometers are routinely used in procedures such as measuring vitamin D and screening blood spots from newborns to detect metabolic diseases via the heel prick test.

"Our proposal is to make the platform available in the cloud, so that it can be downloaded to any mass spectrometer anywhere in the world, data analysis can be performed online. Whether it would be free or paid is yet to be defined," Catharino said.

  **More information:** Carlos Fernando Odir Rodrigues Melo et al, A Machine Learning Application Based in Random Forest for Integrating Mass Spectrometry-Based Metabolomic Data: A Simple Screening Method for Patients With Zika Virus, *Frontiers in Bioengineering and*

Provided by FAPESP