Clinical Chemistry

Machine Learning to Assist in Large-Scale, Activity-Based Synthetic Cannabinoid Receptor Agonist Screening of Serum Samples

Article:
Liesl K Janssens, Dimitri Boeckaerts, Simon Hudson, Daria Morozova, Annelies Cannaert, David M Wood, Caitlin Wolfe, Bernard De Baets, Michiel Stock, Paul I Dargan, and Christophe P Stove.
Machine Learning to Assist in Large-Scale, Activity-Based Synthetic Cannabinoid Receptor Agonist Screening of Serum Samples.
Clin Chem 2022; 68: 906–16 https://doi.org/10.1093/clinchem/hvaa195.

Guest: Professor Christophe Stove and PhD candidate Liesl Janssens from the Laboratory of Toxicology at Ghent University.

Bob Barrett: This is a podcast from *Clinical Chemistry*, sponsored by the Department of Laboratory Medicine at Boston Children's Hospital. I am Bob Barrett.

Synthetic cannabinoid receptor agonists represent one of the largest groups of new psychoactive substances used worldwide and form the largest group monitored by the European Monitoring Centre for Drugs and Drug Addiction. Such monitoring is challenged by the rapid emergence of novel analogs that are often missed by traditional targeted detection strategies. A paper appearing in the July 2022 issue of *Clinical Chemistry* examined machine learning to assist in large-scale, activity-based synthetic cannabinoid receptor agonist screening of serum samples to help address that very issue.

We are pleased to have two of the authors of that paper with us as guest in this podcast. Dr. Christophe Stove is Professor at the Faculty of Pharmaceutical Sciences at Ghent University in Belgium, where he is head of the Laboratory of Toxicology. Liesl Janssens is a researcher in the Laboratory of Toxicology of Ghent University and a PhD candidate in Professor Christophe Stove's laboratory. And Dr. Stove, let's start with you, what are synthetic cannabinoid receptor agonists and why is it important to screen for them?

Christophe Stove: Synthetic cannabinoid receptor agonists, also SCRAs as we call them in short, are actually chemically synthesized small molecules, which act on a receptor, cannabinoids receptor. People might also know them as a class of designer drugs. From a drug perspective, you can actually look at them as a synthetic alternative of the active ingredient of cannabis, THC. An important difference though is that these chemical variants are usually way more active than cannabis or than THC, and this means that drug users can really experience way more intense, and also way more dangerous, effects as compared to smoking a regular joint.

Liesl Janssens: As to why we need to screen for them, their use is of course prohibited in most cases given their eminent danger to public health. It is however not uncommon that overdoses, intoxications, and even fatal cases are reported after the use of synthetic cannabinoids. Importantly, the chemical diversity of this class of compounds is enormous. And why is that? Because clandestine laboratories have for over a decade now been experimenting with slightly changing the chemical structures of known compounds to generate new SCRAs.

And these new SCRAs can be equally active or way more active than the ones that are already existing. Another result from those slight changes in chemical structures is that the new SCRAs can evade existing legislation. Then that makes them so called legal highs, but in fact that they aren't any less dangerous and they should in fact be illegal, as the others. Therefore, we first of course need to screen patient samples, so that clinicians know what they are dealing with. But we also need to screen them so that we can ID if there are new compounds or which compounds actually pose a danger to public health, and so that we can, based on that knowledge, prioritize legislation efforts accordingly, so that these so called legal highs become in fact rightfully illegal again.

Bob Barrett: Okay, seems pretty important to test for those compounds and intoxication cases. Now, I understand that the chemical diversity you mentioned poses a challenge for detection though. Why is that?

Christophe Stove: Indeed. Detection of small molecules like several medicines and drugs is typically performed by what we call targeted methodologies like using mass spectrometry. And this routine analyses analyze the blood or urine samples of patients to look for specific chemical structures or a list of specific chemical structures, which is in a library. However, only the structures which are in that library, which the structures we show on that list, will be detected. And this means that every compound which is not listed will in fact not be detected by these routine techniques. And given the continuous emergence of new synthetic cannabinoids with various chemical structures, we would have to continuously update that library of drugs in order to remain up to date.

Bob Barrett: But if newly emerging compounds synthesized by clandestine laboratories will be missed by the routine techniques, then how can you update this library?

Liesl Janssens: That is where untargeted techniques step in, as opposed to the targeted ones Professor Stove just mentioned. Untargeted means that you look at a sample without having an idea beforehand of what you are exactly looking for. So, you're not looking at a certain list. In fact, you look at

everything that is in there. This way, also new compounds, of which structures are not known or not included on that list, will also be detected. But those compounds will have to be identified later on because you see something in untargeted way, but you don't know what it is yet, so therefore identification afterwards is definitely necessary.

The current method for such untargeted analysis is high resolution mass spectrometry, a bit similar to normal mass spectrometry but more specified.

It's also structure-based, so it means that also this technique looks at specific chemical structures, which can come in handy as any technique can also be used for identification later on. The problem with this technique is that it's unfortunately quite expensive and also very labor intensive. So, I believe it cannot be expected to be implemented in just every clinical laboratory or at least not on a routine basis.

Bob Barrett: In your paper, you mentioned that your activity-based method is also untargeted. Would you suggest your activity-based detection method as an alternative for HRMS then, high resolution mass spectrometry?

Liesl Janssens: No. I would not suggest it as an alternative really, because we need structure-based techniques for identification for one. So, I would rather suggest it as a compliment to high resolution mass spectrometry or HRMS. Because activity-based detection is indeed also untargeted, because you look at a certain characteristic that is shared amongst the whole class of compounds. Me, for example, look at the activation of the cannabinoid receptor, which means that ever compound that activates the receptor will be detected by the bioassay whether it's a known compound or it's a completely new one of which we are not aware yet. So, if we were to implement the activity-based detection method as a first line screening tool, then we would already get an idea of which samples are interesting or suspicious, depending on the way look at it. And therefore, we only have to send in fewer samples for the more specified technique. Only the positive samples in the first line screening tool, which would be activity-based, would need confirmation by HRMS later on.

Bob Barrett: So, you want to save time and cost by reducing the number of samples run on highly specific techniques, but your paper also focuses on saving time and cost for activity-based detection. You developed the machine learning method to analyze the data. In practical terms, how much time does this method save?

Christophe Stove: Let's say that a computer can do in a few minutes what we do in quite some hours. The original historical process that we use included really visual interpretational vault, time,

response, profiles for every sample. And this was to be done by two individuals, followed then by discussion between these individuals, Liesl, and myself, in those cases where there was a profile which wasn't very clear cut and was a bit doubtful whether it would be positive or negative. So, this is for a total of almost a thousand samples from our publication resulted in many, many hours of discussion and decision taking. And altogether, we can actually say that machine learning would allow us to do this way quicker.

Bob Barrett: Besides saving time, are there other advantages associated with the machine learning approach?

Liesl Janssens: Besides saving time, the machine learning method also objectifies the outcome. Because as we as individuals interpreted all the data separately, this is quite a subjective process and you have, in fact, to be a little bit experienced to know how to look at such data, especially for example, for samples only concerning lower concentrations of SCRA metabolites. Those are really difficult to distinguish, because metabolites also give a signal and are also impact positive samples. But that signal is always little bit reduced, because it's not the main compound you're looking at.

So, definitely in those cases, it's even more subjective how to distinguish a positive from a negative sample. But by implementing a machine learning model, you would remove that subjective aspect and just have a standardized outcome, which also makes it easier to transfer the methods to other laboratories for example, because you don't have to transfer expertise in data processing them and you don't have to train any personnel to do so. And of course, the staff doesn't have to spend many hours on data analysis, which would be a waste in fact.

So, I would summarize it as definitely saving time but also objectifying, standardizing, and also making it easier to transfer the methods. In fact, if everything goes as planned, we would reduce the data processing to one simple mouse click, some minutes for the computer to process, and then having the results ready in a simple Excel sheet.

Bob Barrett: Is your method ready for use in clinical labs?

Liesl Janssens: No. Unfortunately, I believe not yet. As the publication, and also the editorial on our publication mentioned, there is still some optimization to be done. The sensitivity and specificity we could achieve manually was very good actually, but it could not be completely matched by the machine learning model. We could match the sensitivity we gained, but this was at a cost of still too many false positives, which was not what we'd want, because this increases cost again.

Bob Barrett: And I presume those false positives would have to be run by high resolution mass spectrometry.

Liesl Janssens: Yes, indeed. So, that's why we currently prefer to still evaluate data following our manual process. So, the visual interpretation of all the data, so that we don't have to send that many samples to HRMS or high resolution mass spectrometry, but we intend to use all the data that will from now on be generated still, and to use that to further train the models, so that we can maybe in time get it smarter, so that we can still retain this high sensitivity, but that we can get that number of false positives to drop.

That will be the ideal case.

Bob Barrett: Okay. For people listening who are not familiar with artificial intelligence or machine learning, you use a supervised learning approach. In other words, this means that a computer was told to interpret the bioassay outcome exactly as the scoring individuals. You told the computer what a positive result or negative result looks like. What about deep learning as an alternative for this? Just let the computer learn for itself how a positive or negative sample should look. Have you tried this approach?

Liesl Janssens: No, we have not. We have explicitly chosen to use a supervised learning approach. Given the high sensitivity and specificity that we attained manually with the visual interpretation of data, we actually wanted to have the computer interpret the data just like we do. So, we explicitly chose for a supervised learning approach. In the study, it has also become clear that our opinion on the data can actually be of added value to the computer, because if we train the data on the way that we interpret the outcome of the bioassay instead of the fixed analytical results, we could increase the performance of the models so to say. So, I believe that we are somewhat more experienced interpreters of such data profiles, can still be useful in training the computer how to do so.

Bob Barrett: Finally, doctors, machine learning and bioassays are quite distant research topics. How did you manage to put these together?

Christophe Stove: I should say that this work was only possible because of such a multidisciplinary approach. Clinicians were responsible for patients and recruitments. The analytical toxicology was needed to get an idea of the concentrations present in the patient samples and we, ourselves, applied our cell-based bioassays for activity-based detection. And then also, importantly for the machine learning, we could rely on very experienced colleagues from the Bioengineering Department of Ghent University, with Dimitri Boeckaerts, equally

contributed first author with Liesl. So, I could say that for this publication in *Clinical Chemistry*, collaboration was key.

Bob Barrett: That was Dr. Christophe Stove from the Faculty of Pharmaceutical Sciences at Ghent University in Belgium, where he is head of the Laboratory of Toxicology. He was joined by Liesl Janssens, from the same laboratory, and they have been our guests in this podcast on Machine Learning to Assist in Large-Scale, Activity-Based Synthetic Cannabinoid Receptor Agonist Screening of Serum Samples. Their paper on that topic appears in the July 2022 issue of *Clinical Chemistry*. I'm Bob Barrett. Thanks for listening.