



Article:

Edmund H. Wilkes, Gill Rumsby, and Gary M. Woodward.

Using Machine Learning to Aid the Interpretation of Urine Steroid Profiles.

Clin Chem 2018; 64:1586-95.

<http://clinchem.aaccjnl.org/content/64/11/1586>

Guests: Dr. Edmund Wilkes is a Senior Clinical Scientist and Dr. Gary Woodward is a Principal Biochemist, both from the University College London Hospitals in England.

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.

Machine learning involves the study of algorithms and mathematical models so that computer systems can progressively improve their performance on solving a specific task. A recent application of machine learning through the interpretation of urine steroid profiles appears in the November 2018 issue of *Clinical Chemistry*. Two of the authors of that paper are our guests in this podcast. They are Dr. Edmund Wilkes, a Senior Clinical Scientist and Dr. Gary Woodward, a Principal Biochemist, both from the University College London Hospitals in London, England and Dr. Woodward we'll start with you. How did the idea to apply machine learning to the area of clinical chemistry, and steroid profiles in particular, arise?

Dr. Gary Woodward: Well, I guess the main thing to think about is the fact that machine learning is really about -- well, one of the things of machine learning is the fact that it's about pattern recognition. And through the process of doing the day-to-day job of a clinical biochemist, we spent a lot of our time actually really recognizing patterns and that's recognizing patterns in clinical biochemistry results, relating those results to clinical pictures, for example, and things like that. So, I've always been intimately aware of the fact that the job we perform has to do with pattern recognition as such.

And through working with a colleague of mine on some other projects that deal with imaging and pattern recognition imaging, I quickly saw that the same techniques could be applied particularly to the field of clinical biochemistry. I was very keen to do that. When I became involved with actual steroid biochemistry, the case was even more to point in that we look at profile data and this is pretty much a matter of recognizing different patterns of profiles to diagnose specific diseases. So, you could see there was a very clear use case for machine learning, and so that's really where the idea came from. We're really looking at steroid profiles and it's a pattern-based problem, can we

Using Machine Learning to Aid the Interpretation of Urine Steroid Profiles

do something with machine learning to help us perform that task?

Bob Barrett: So, Dr. Wilkes, we'll go to you. What sort of tools did you need to have available to you in order to carry out this study?

Dr. Edmund Wilkes: So actually, most of the tools that you can use for training machine learning models are actually freely available, and so the only tools we had available to us actually within the NHS was just a laptop and some freely available statistical software known as the R-statistical computing language. But there are many other freely available tools such as Weka and then Python which can also be used for the same thing. And most of the tools within those software packages are actually open source and a lot of the code is all available online and there are excellent tutorials online, both YouTube and blogs, of actually how to use these tools. So, there's a kind of wealth of resources online for tools to actually do machine learning with data you produce in your lab.

I think the main point is that these technologies and tools are widely available, and the amount of options of tools available are increasing with every step that we take, so there's many options.

Bob Barrett: You must have had some stumbling blocks and challenges. Can you tell us about some of those and how they were overcome?

Dr. Edmund Wilkes: I think the main challenges with machine learning are A, the data to train the models with. But I think throughout the process, it's much of a learning process for both of us and I think what we realized is, that there are lots of kind of "gotchas" as they're called in machine learning, where bias in the training set or bias in the way in which your training models can impact to the model you get at the end and then the results you get at the end.

So, you can often get results you think are too good to be true and they often are too good to be true. And I think the thing we learned most about applying machine learning into real world data is you have to be very, very careful about the way in which you train your models and test them. And that's why in the paper, we settled on a process called nested cross-validation, which for the size of the dataset we're using, is kind of the gold standard for making sure that you're not getting results that are too good to be true and you're not overfitting the models. So, I think throughout the process, those were the main stumbling blocks, just getting to grips with a lot of the statistical concepts behind training and testing machine learning models.

Dr. Gary Woodward: So, one of the things that I thought was particularly challenging as well and I've seen in many other fields now too, is this concept of what your ground truth is. So, the idea is that you have -- in conventional machine learning, you sort of are training an algorithm to perform a specific task against something that, and in this case, categorical variables that are based on a ground truth. Now, in many instances, that ground truth isn't necessarily valid or is something other than opinion.

So, the idea is that A, based on that sort of opinion ground truth, it's how do you show that you perform better than something like that? And how do you compensate with things that don't have a very clear ground truth, and how do you build out a model that was quite (00:05:20). So, if you were to look at the study of lipids, perhaps something like a sort of consensus-based ground truth or some sort of scoring with the variation if that would be good. But I think that's probably one of the big challenges when we look at this kind of analysis, it's ultimately what are you comparing to and how do you know that that is true? And that goes back into how you're accounting for bias in your data because ground truth is biased, your analysis is going to be biased too.

Bob Barrett: So, if you had to do it all over again, what would you do different?

Dr. Gary Woodward: Yeah, so like I said, so in this particular paper, I suppose one of the things that we haven't assessed was -- so we've used a bunch of analysts or expert operators to decide on a categorical variable to which we're going to compare. And of course, that's distributed across the data. We haven't looked at the repeated variations for how likely an individual is to get the same answer over and over and between people too. And I think including that sort of number into the analysis would be really useful, because if you know how reproducible the ground truth is, you can then have a look at how your metric compares to that in a much more informed way.

I think in the paper, we do, do that to a small degree and that we were lucky enough to have sort of EQA data in which the sample have been distributed to a whole load of experts, and so we could get a flavor for what the sort of consensus and interpretation might be. But I think that that concept wasn't built into the model itself per se. So, I think probably it would be good to redo it, but kind of build that variation into the model rather than have to do a tester.

Dr. Edmund Wilkes: Yeah, I agree. I think for you to do it again, we'd get the kind of expert raters to interpret their kind of fixed set of

Using Machine Learning to Aid the Interpretation of Urine Steroid Profiles

urine steroid profile and get them to do that multiple times - - over say, kind of long time periods to assess the variation between interpreters and also giving them the same profile maybe five times to see if, how consistent their interpretation of background truth is. I think as Gary mentioned, the definition of your ground truth is kind of completely fundamental to training machine learning models, so that would help us, as Gary said, kind of get an idea how well we're performing compared to the average interpreter which I think in hindsight, adds a lot to the study.

Bob Barrett:

Dr. Woodward, what training or experience was required to perform such a study and could one of our listeners or readers of the article just start to apply machine learning in their lab tomorrow?

Dr. Gary Woodward: Yeah, so I think that's a good question and one that we were keen to express in the paper. To start with, well, his background has been quite heavily I guess on what you might call bioinformatics. And for myself, I had as much statistical background as any average scientist would have. But neither of us had any particular expertise, skills, or experience in machine learning itself. Yet, with what was available to us in the literature and online and with colleagues and such, we were able to assimilate what needed to be done and start to apply such things to the data we had.

And of course, the paper took a while to do because there are several iterations of the same thing as we did it. We thought we did well and then we sort of learned a bit more and realized actually what we did was wrong, and there were a lot of flaws in it and we'd do it again. So, one, then you have an iterative process of getting better, but the key thing is that really, we started from zero and got to where we are today.

I feel like we have a very good appreciation of it and sure, there's going to be a lot of new things to learn as there always are, and sort of we kind of really focused at least in the paper on a small set of possible techniques you could use in a machine learning environment. But I think it's positive and encouraging to say that any one of our clinical chemist colleagues could take this forward. However before further applying it in their clinical practice, many of the lessons we've learned and that we've described here need to be accounted for. You can quite easily go wrong if you're not careful, and if you don't understand the concepts of accuracy and precision and sensitivity and balance accuracy. You could quite easily build an algorithm that's telling you something that isn't true, but you won't know it. I think

Using Machine Learning to Aid the Interpretation of Urine Steroid Profiles

that's the only caveat. But the actual performance and stuff being involved with it is open to everybody.

Dr. Edmund Wilkes: Yeah, I agree. I think as Gary said, I had a kind of background in bioinformatics, but that's not to say that anyone who can learn a bit of programming can't use these freely available tools to do this kind of work. I think it's worthwhile highlighting that a lot of learning-to-program courses are available as well freely online, and so anyone can really pick it up and do it. But I think it also kind of highlights a need for training of clinical scientist, at least in the UK, in these kind of fundamental programming techniques and things like that to equip the new cohort of clinical scientists coming to in the UK with the kind of skills they need to do this because I think these kinds of tools are going to be much more prevalent and should be more prevalently used within our healthcare system.

Dr. Gary Woodward: And you can also see that the -- I mean in the AI bubble, I think it is fair to say there's a bubble, it's growing quite immensely and you can see it everywhere now and in every scientific space. And so, I feel that as a biochemist, you would probably be doing this out for the service and not know about it and not have some fundamental understanding of it. However, I think we should be aware that while everything seems to be AI-based at the moment, it is really just at a fundamental level, another tool we can use to solve the problem. And as long as we always keep it within context, I think it would be very much in the interest of all biochemists to equip themselves with this tool to take problems forward in their lab because in fact, a lot of that, of problems we find in the labs are very amenable to these kinds of analytic applications. They can just make our lives easier as it has in this case.

Bob Barrett: Well, do you foresee machine learning ever replacing the clinical chemist or pathologist?

Dr. Gary Woodward: No, I think it's bit of a misnomer in the fact that that is the case. Nothing is going to replace the clinical chemist and you've always got to be aware of, at the fundamental level, machine learning algorithms or any, in a sense, to some extent supervise and only designed to perform a task in a very specific instance for very specific reason.

So, the oversight of someone who understands what it's doing is always going to be required. The only thing that it's going to do is help us do our job easier and more efficiently.

Dr. Edmund Wilkes: Yeah, I agree, I think it's always going to be a synergistic relationship between technology and a human. I think there will be instances where a human gets interpretation right or task correct, where a machine doesn't and vice versa, when

Using Machine Learning to Aid the Interpretation of Urine Steroid Profiles

the machine gets it right and human doesn't. And I think by combining those two complimentary skill sets if you like, the combination of a human and the machine will always perform better than a human or machine in isolation.

So, I really don't think this is about replacing clinical chemist or pathologists, but rather as Gary said, it's making our lives easier and especially given the increased workload we're seeing and through laboratories in the UK, making all our process more efficient, and hopefully at the end of the day, safer and better.

Bob Barrett:

Laboratories, of course, are highly regulated areas of healthcare. What do you see as the regulatory or quality-related issues in applying machine learning to the clinical lab?

Dr. Gary Woodward: That's an interesting question and I don't think anyone really has a good handle on this at the moment. So, for example I can speak in terms of the UK and I'm sure that labs are regulated, UKAS and I haven't seen any provision for something like machine learning to be used there. But I think the key and important thing to bear in mind is that in the context of which we're using the tools at the moment when they're not diagnostics. They're not performing the diagnosis or interpretations themselves, they're really providing information that a human still uses to advance reputation. So, in that sense, it's a tool.

I know in the imaging space for example, there's a lot of machine learning is being used, the well-defined steps for validating those tools, and I think the FDA have even started a couple of clinical track pathways for regulating devices in the machine learning diagnostics for these purposes. But I think it always just comes down to the particular validations and context of use, but I'm seeing in labs at the moment, they're really just precision support tools rather than diagnostics and as long as they're rigorously validated, it should conform to what is currently in scope for regulatory approval, although that might change in the future as things become more ubiquitous.

Bob Barrett:

Well, finally, are you continuing this research, and which areas of clinical chemistry do you see having the most opportunities for machine learning applications? We'll start with you Dr. Wilkes.

Dr. Edmund Wilkes:

So, yeah, we are continuing to use the tools we developed for this paper and because that's where some readers I'm sure can appreciate, there are number of profiling assays used within clinical chemistry laboratories that are amenable to this approach where we're kind of measuring a panel of analytes and then coming up with some sort of

Using Machine Learning to Aid the Interpretation of Urine Steroid Profiles

interpretation as to what the panel represents. So, we're looking at applying these tools to more profiling assays which I think will be the kind of low hanging fruit in clinical chemistry and the kind of assays that this will be most applicable to in the short term.

But I think what we'd like to start doing is applying this to clinical outcome data, to changing our ground truth from a person's interpretation of a profile to an actual hard clinical outcome, I don't know, histopathology result in the box receive the treatment and see if these kinds of assays can be predictive in that as well.

But I think that there's been some great work and some other labs in the world, Jason Brown's lab is a good example of using general routine biochemistry and hematology to actually predict the results, a full diagnosis from it. So, I don't think it's just specialist profiling assays. I think these kinds of machine learning techniques will be more widely applied to routine biochemistry as well.

Bob Barrett:

That was Dr. Edmund Wilkes, a Senior Clinical Scientist and he was joined by Dr. Gary Woodward, a Principal Biochemist both from the University College London Hospitals in London, England. They have been our guests in this podcast from *Clinical Chemistry*, about machine learning and urinary steroid profiles. Their article appears in the November 2018 issue. I'm Bob Barrett. Thanks for listening!