# Transcript

Paul

Hello and welcome to another edition of To Immunity and Beyond. And today I've got with me Yang Hu, and Yang is going to introduce himself now.

Yang

OK. Thank you to have me here and I'm a post doc on computation pathology at Big Data Institute of Oxford and I'm working with Professor Jens Rittscher. In our group we basically do some pathology image analyzing, and a little part of endoscopy analyzing, to try to understand using AI power to understand something of the mechanism of the disease progression of the cancer.

Paul

Great. Thanks. So thanks for getting in touch. So we're always looking for volunteers to talk and I really appreciate it. And your paper is out in April, so it's very recent, so it's fantastic and it's about self-interactive learning. So it's a long title, available to everybody on the podcast. We'll share the link. But perhaps before we start on the details of the paper, you can explain to us a little bit about your background, and how you've come from your background in computing to address this particular question.

Yang

Yeah. So, basically in our group most people are working on medical image analyzing, and most of the people are coming from the background of computer science and artificial intelligence. And so basically the task we are going to accomplish is to use AI to understand the pathology images and help the pathologists to explore some new mechanisms which haven't been discovered yet to help us have a better understanding of the cancer or some chronic disease. So specifically for this paper we are going to solve the question that is - how the pathology images can help us to do some molecular subtyping problem. So molecular subtyping problem is that, for example, when we would like to decide how to make a treatment for different patient in cancer, we need to figure out what kind of specific subtype of patient it is so that we can provide the precision treatment plan for them, to improve their survival rate. But usually the subtyping actually is they rely on the golden standard, the gene expression test, and that is quite expensive. So it's not really realistic to do that for everyone. So we would like to think about, oh, if we can use the digital pathology images to help us, the training model, and can automatically do that from the perspective of the image analyzing part.

Paul

Great. So, and you’ve tested four different types of cancer here, but does it generally work like this in the clinic that the pathologist would, if they didn't have the molecular diagnostics with them, they would make some, essentially, estimate of which one of the molecular subtypes it was more likely to be, based on the appearance of the cancer, is that is that the case?

Yang

Yeah, that's the case, so, for training AI model, we always have some collected cohort, to collect the cohort with the annotation about specific subtype, has already been labelled by the pathologist, but the problem we are facing in this paper is that because the annotation and the data collection in clinical side is quite expensive, so usually we can we spend a lot, we invest a lot of money, we just collect the very tiny cohort, and that's conflicting to the demand of the AI model. Because usually when we are trying to train an AI model, we rely on a very big cohort training it, so we need a very data-efficient mechanism or algorithm to efficiently leverage those kinds of small cohort to do that. Yeah, that's the main problem we need to solve in this paper.

Paul

So for people who haven't yet read the paper and are just trying to understand the overall approach, so fundamentally you're going to take samples that are taken at surgery or biopsy and are just presented on a conventional clinical slide. That's the initial template of what you're given, is that right?

Yang

That's it. Yeah. I should introduce the fundamental workflow of the whole pipeline. So usually when we do the biopsy from the patient and we got the biopsy, and we scan it to make it digitalized, which can make us sharing and reviewing them in the computer side. And then we can programme in some AI model to try to use the medical imaging analyzing algorithm to deal with that. So that's the basic pipeline we are going to get through. But the problem is the whole slice image. Usually they are super big. Usually each slice will have two or three gigabytes. So that means we cannot deal with that directly with the computer, with the GPU we have. So we have to split the whole slide into many, many small patches. And extract the feature for each patch. Then we just send it to another kind of AI model based on the neural network to train in that part of small neural network. Then that gives us the prediction result. So that's the basic pipeline of the whole approach.

Paul

OK. So as I understand it, and this is how it's laid out in the paper, you break the bigger image up into little smaller tiles and then you're analyzing the patterns within each tile and also you have higher order patterns between the tiles as well. And what struck me as really interesting was you're also looking at, not just the patterns of how cells are aggregated, but also within the cell, at cytopathic features within the cells. So you're looking at all sorts of different scales. Is that, maybe you could explain a little bit in a bit more detail how you did that?

Yang

So the problem we are facing is that as we just separated the whole slide image into small patches right, that means we are facing two levels of, two scales of features. One scale is on the tile level. That means the very fine-grained cellular level features. And another scale is on the whole slide level. That means overall morphological level feature scales. So the problem is the original algorithm doesn't have any communication mechanism between these two scale features. So in this paper we just proposed a communication mechanism can help these two scales of features communicate with each other so that they can optimize each other further.

Paul

OK, great. So then you're going to integrate all these separate pieces of information that are telling you something about the piece of tissue that's been taken. And then how is that information then integrated to give you a sort of answer to your question of what sort of cancer you're looking at and what sort of processes are going on?

Yang

So basically, once we got the optimized features from the tile level, and we are feeding it to the slide that will aggregate. And the aggregate - we are using a kind of attention mechanism to indicate which area is more important to help us decide, help us output the prediction result. So that means the other areas we just assume is not so important for the specific task. So then the attention mechanism can help us, can give us the result about: Oh, maybe this part of the area is very important for giving us the final prediction result. So usually we just expect that area will be the tumour area, but sometimes it will also contain some stroma areas surrounding the tumour. That means that means the stroma area also supports the growing up of the tumour, right? So we will ask pathology to look at few of the attention maps, and ask them if that makes sense. And we did that, and most of the result, pathology think it makes sense to them.

Paul

OK, so the pathologist has had years and years of training to become expert at essentially identifying patterns. And your machine learning has taught itself the same thing through your programming. And this was a challenge for me to look at all the programming elements in the in the paper, but essentially you've designed a toolkit for it to kind of self-learn from data sets. So once you got these answers and you say well, this pattern looks like this, and this one looks like that. So in terms of the cancer, you looking for KRAS cancers in the colon. You're looking at particular types of breast tumour. How do you, just a general question, how do you test whether it's doing what you think it's doing, and you're not just sort of modelling it on itself. So you have validation, and how does that actually work, and tell us how well validated your approach is?

Yang

So basically we also consider that problem - to avoid the observation, it just reflects some overfitting problem. We just separated one cohort into two parts. One is for training, another one, about 20%, is for testing. So we just train the model on the training set, about 80%, and directly test on the other 20% of testing dataset. And we do five floored testing, which means we pick up the 80% for training dataset five times and differently. So that we can calculate the average performance finally and we also do some external validation, that means we only train the model on one cohort, and test it on a completely different cohort. So that validates the performance of our algorithm.

Paul

So I’ve got some general questions about the approach, because I've sort of seen similar things going on with inflammatory disease and one of the problems there was everybody’s slides were just slightly different. So they're all H&E, so you'd think it's going to be the same, but, but every lab is just, you know, they're cut slightly differently. They're stained slightly differently. How easy is it to use the same approach in different labs, or does your approach actually cut through all that and mean that it will always be accurate even if people are doing slightly different preps?

Yang

Actually, before I talk about that, I would like to say, uh, we are we are in a very remarkably changing era of AI/pathology analyzing research. Because in the recent two years we have a new technology called foundation models. That means the foundation model is based on this big company. Just invests a lot of money and collect a very big, big cohort without any annotation. They just asked the model to pre train on that, to help the model have a very strong background knowledge about pathology images. So they assume that kind of model can give us a much better feature extractor. So as you just mentioned, when different teams are going to apply some AI model to deal with their specific cohort. And different staining and different background, different disease, right? They just assume that foundation model, so the seller of this foundation model just say - oh our foundation model is superpower and you can directly use our foundation model based feature extractor to directly use on your cohort, and they just assume their foundation model has already seen every type, most of the types of pathological images including IHC and H&E staining.

Paul

So does that mean that there’d be some kind of convergence of people doing this, because another feature would be that, I mean you've done this very nice job on these particular cancers. But there'll be other people reproducing this in different universities, or companies, across the globe. So it sounds like we don't need hundreds of different models. We probably just need, you know, one really good one.

Yang

One really good one. But the realistic problem is that although they’re just testing their foundation model on about 10 sub downstream tasks, and to prove their foundation model's performance is super good. But when we tested them on our private dataset, they still have some problem, they still have some bias. Because all these foundation models, although they just collect a very big cohort, but this data from that cohort is all based on one hospital, or one location and all these samples are focused on very few types of the cancer. So that means they have some very strong potential bias behind them. The real situation is not so optimised as they sell.

Paul

OK. I can imagine that happening, because whoever is doing it, they can't be completely comprehensive. But the community, I suppose, is, you know, there'll be people that are interested in, in particular rare cancers or different populations.

Yang

Yeah. Then they can fine tune their model.

Paul

OK. And so, I guess the other question is, can people share these models in a way that allows it to be developed as a community? Or does everybody have to develop their own model?

Yang

Yeah, so basically all these foundation model, most of them are have already been shared - their weight, they are not shared the original images, they just shared the trained weight of all these models and to help other people to use for extractor features on their own cohort. But as you just mentioned for some rare cancer, for some rare cases, we still need to fine tune the models on our specific tasks. So that means these papers, the algorithm proposed by these papers, can play the very big role in that, because in this paper we can fine tune the features from both slide level and tile level.

Paul

Great. Well, it sounds like a really big step forward and I had a few follow up questions really about its use in terms of cancer. So you've clearly got a kind of method for addressing molecular diversity and that's manifest as structural changes or ultra cellular changes, ultrastructural changes rather. But at least somehow the molecular diagnostics map onto what you're seeing on the slide. In an inflammatory disease which, you know, a lot of people are interested in around here, where it's a pre cancerous state potentially but there's no transformation. Do you think that would also be a good template for this sort of approach where you've got, you know, a range of different features that you could explore and come up with essentially something more like a molecular diagnosis? I think this is something that people are also trying to understand in some detail.

Yang

Yeah, I think it's applicable because this algorithm just proposed is very general and it can be used into every circumstance and include something like you just mentioned, try to figure out how is the progress of the inflammatory disease. What kind of mechanism may cause some more serious disease. What kind of mechanism may cause the healing of that disease. We, for example, in the next step maybe for this paper, we actually think the performance on the accuracy side is already good enough for us, actually from the perspective of the statistic, is even better than human sometimes. But the problem always appeared in the interpretability side. So that means, could we just always get the stable and reliable interpretable output from all these AI models and what kind of interpretability technology is the best for us? For now, we only rely on the mechanism called attention. The attention is just very simply, tell us which region is more important. Actually, they cannot describe the interaction between different cells, or some behaviour like immunity behaviour or some inflammatory behaviour, or it cannot describe more complicated mechanisms in the microenvironment, for example some immune cells just trying to attack the tumour cells, something like that. It cannot detect that. It just tells us which area is more important for prediction. So maybe for the next step, maybe we should think about how to develop a better interpretability technology.

Paul

That sounds brilliant, so that you could not only define areas which are essentially giving you maximum value. But also essentially, what is the information in there and what that's telling you about the underlying process. Yeah. So that sounds very exciting. So what is your practical next step? So, you got the thing published. Congratulations. What are you going to do now?

Yang

So, just as I said, so a very recently new work I have conducted is about to test the stable of the output of these foundation models. We just find a very interesting phenomenon. For example, we tested two foundation model. They just show us very similar accuracy performance. But actually the attention moving behaviour, so for example, when they just trying to transfer some knowledge from their portraying the state to the specific tasks, the attention will moving from some area to another area. That means, for that foundation model in their understanding, if they want to accomplish that downstream task, they think that area is much more important than the than the others, but the attention moving behaviour actually shows completely different in different foundation models. So although they just, although they even can give us the same performance on accuracy. So that means when we trying to check the result, which one are you willing to believe? Right, that's just raised a very…

Paul

I think this would be people's general concern about these AI models, is because it's kind of a black box, you don't really know what's going on. And so it's quite hard as a clinician where you want to make important diagnostic decisions to, to rely on something where you don't really know how the data is being generated.

Yang

Yeah. So that's the thing. Actually so in the recent work I just have that proof of concept and have that kind of test and we just find that dangerous situation. But we still don't know where is the reason to cause that heterogeneous of the of the interpretable results from different foundational models and we also don't know how to try to solve that. We still need to invest more time, maybe in future to do that.

Paul

Well, that's the usual conclusion for many papers, whether it's computational or experimental, is that more work is needed. So that's probably a good place to stop. But thank you very much for that explanation and it's certainly an exciting area, and I know Jens’s group is doing amazing work, so people need to watch this space and we'll continue through the network to try and promote spatial biology. Because it's really got so much to offer to such a range of immunologists around the place. So we're really grateful for you to come along.

Yang

OK. Thank you very much to have me here. Yeah, it's quite an honour to explain my work here. Thank you.

**Follow up note from Yang after the recording:**

‘In my view, pathology remains a “blue ocean” domain for AI, especially when we try to probe the tissue micro environment in greater depth. With continued progress, pathology AI could illuminate disease mechanisms at a detailed cellular level and even track how new drugs affect specific micro environmental activities or finely defined cell populations. Yet today’s AI tools are still far from delivering that degree of granularity. Even in my own work, cross scale information transfer can suffer substantial data loss when we zoom in so closely, and I am not fully confident in the fidelity of the results. Despite AI’s impressive performance on many “low hanging fruit” tasks, we must be certain that any subtle mechanisms we identify are truly reliable and not merely statistically “significant” artefacts.’